INTERROGATORY CELL-BASED ASSAYS FOR IDENTIFYING DRUG-INDUCED TOXICITY MARKERS

Cross-Reference to Related Applications

This application claims priority to U.S. Provisional Application Serial No., 61/650462 filed May 22, 2012, the entire content of which is incorporated herein.

Background of the Invention

The pharmaceutical industry is currently witnessing a 90% attrition of potential compounds entering clinical development, 30% of which is owing to poor clinical safety (*Kola et al.(2004) Nat Rev Drug Discovery:3 711-715*) . In the U.S., fatal adverse drug reactions (ADRs) are the 4th to 6th leading causes of death. Costs directly attributable to ADRs may lead to an additional $1.56 to $4 billion in direct hospital costs per year in the U.S. (*Lazarou J et al.(1998) JAMA; 279(15):1200-1225*). The cost of drug discovery and development has increased to about $1 billion, partly due to increased attrition of compounds and NME late in clinical development *(Adams CP, Brantner VV (2010) "Spending on New Drug Development" Health Econ. 19: 130–141*). The lack of reliable tools that can help with predicting toxicity early in drug development is partly to blame for increasing costs and lower return on investment. Further, drug safety issues are the leading cause of increased litigation and settlements in the pharmaceutical industry. Between January 2009 and May 2011 the industry has spent over USD 8 billion on litigation cases related to drug safety issues.

In order to augment a "kill early policy" of compounds in early clinical trials and drug development, the FDA is now encouraging the drug industry and the community to adopt a very innovative strategy. FDA white paper *Innovation or Stagnation: Challenges and Opportunity on the Critical Path to New Medical Projects* states, "A new product development toolkit containing powerful new scientific and technical methods such as animal or computer-based predictive models, biomarkers for safety and effectiveness, and new clinical evaluation techniques—is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product" (FDA, 2005). The FDA declaration clearly underscores the lack of innovative technologies that can aid in efficient decision making in drug development.

Cardiotoxicity refers to a broad range of adverse effects on heart function induced by therapeutic molecules. Cardiotoxicity may emerge early in pre-clinical studies or become apparent later in the clinical setting. It is a leading cause of drug withdrawal, accounting for over 45% of all drugs withdrawn since 1994, which results in significant financial burden for drug development. Cardiovascular toxicity includes increased QT duration, arrhythmias, myocardial ischemia, hypertension and thromboembolic complications, and myocardial dysfunction.

Cardiac safety biomarkers currently used by the FDA are QTc prolongation – lectrophysiological arrhythmias, circulating troponin c, heart rate, blood pressure, lipids, troponin, C-reactive protein (CRP), brain ot B-type natriuretic peptide (BNP), ex vivo platelet aggregation, and imaging biomarkers (cardiac magnetic resonance imaging). The QTc prolongation is a very robust but complex marker. However, a decision on whether to kill or sustain a drug in early development is hard to make based on QTc alone. In addition, QTc is subjective and is dependent upon underlying pathologies that can lead to tachyarrythmias.

In view of the foregoing, it is evident that new cardiac safety biomarkers, such as molecular cardiac safety biomarkers, are needed in the art.

Summary of the Invention

The platform technology described herein is useful for identifying markers associated with drug-induced toxicity. This platform technology integrates molecular interactions within and across a hierarchy of models starting from primary human cell based model to human clinical samples. This approach leads to the identification of biomarkers that reflect an underlying toxicity caused by a compound or NME that is a potential drug, such as a drug candidate ready to enter phase I clinical trials. Drug induced toxicities can include cardiac, renal, hepatic and other tissue toxicity. The instant application provides several novel biomarkers associated with drug-induced toxicity, and which are useful in methods for predicting potential toxicity of a molecule or drug candidate, and as potential therapeutic targets for treating, preventing or counteracting drug-induced toxicity.

The invention described herein is based, at least in part, on a novel, collaborative utilization of network biology, genomic, proteomic, metabolomic, transcriptomic, and bioinformatics tools and methodologies, which, when combined, may be used to study any biological system of interest, such as obtaining insight into the molecular mechanisms associated with or causal for drug-induced toxicity. The platform technology is further described in international PCT Application PCT/US2012/027615, the entire contents of which are hereby expressly incorporated herein. Additional embodiments of the platform technology, including a description of how to carry out platform technology methods involving incorporation of enzyme (e.g., kinase) activity data, are described in U.S. Application Serial No. 13/607,587, filed on September 7, 2012, the entire contents of which are expressly incorporated herein by reference. In a first step, cellular modeling systems are developed to probe a drug-induced toxicities, such as cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity or myotoxicity . A cellular system modeling drug-induced toxicity can comprise toxicity-related cells subjected to various -relevant environment stimuli (e.g., hyperglycemia, hypoxia, immuno-stress, and lipid peroxidation, or exposure to a test molecule or drug candidate). In some embodiments, the cellular modeling system involves cellular cross-talk mechanisms between various interacting cell types related to specific drug-induced toxicity, such as cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neuronal cells, renal cells, or myoblasts. High throughput biological readouts from the cell model system are obtained by using a combination of techniques, including, for example, cutting edge mass spectrometry (LC/MSMS), flow cytometry, cell-based assays, and functional assays. The high throughput biological readouts are then subjected to a bioinformatic analysis to study congruent data trends by *in vitro*, *in vivo*, and *in silico* modeling. The resulting matrices allow for cross-related data mining where linear and non-linear regression analysis were developed to reach conclusive pressure points (or "hubs"). These "hubs", as presented herein, are candidates for drug discovery. In particular, these hubs represent potential drug targets for reducing or alleviating drug-induced toxicity and/or drug-induced toxicity markers.

The molecular signatures of the differentials allow for insight into the mechanisms that dictate the alterations in the tissue microenvironment that lead to drug-induced toxicity. Taken together, the combination of the aforementioned technology platform with strategic cellular modeling allows for robust intelligence that can be

employed to further establish an understanding of the underlying mechanisms and molecular drivers contributing to drug-induced toxicity, e.g., cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renal toxicity or myotoxicity while creating biomarker libraries that may allow early identification of drug candidates at risk for causing drug-induced toxic effects, as well as drug targets that may reduce or alleviate drug-induced toxicity.

A significant feature of the platform of the invention is that the AI-based system is based on the data sets obtained from the drug-induced toxicity cell model system, without resorting to or taking into consideration any existing knowledge in the art, such as known biological relationships (i.e., no data points are artificial), concerning the drug-induced toxicity. Accordingly, the resulting statistical models generated from the platform are unbiased. Another significant feature of the platform of the invention and its components, e.g., the cell model systems and data sets obtained therefrom, is that it allows for continual building on the drug-induced toxicity cell models over time (e.g., by the introduction of new cells and/or conditions), such that an initial, "first generation" consensus causal relationship network generated from a cell model for a drug-induced toxicity can evolve along with the evolution of the cell model itself to a multiple generation causal relationship network (and delta or delta-delta networks obtained therefrom). In this way, both the drug-induced toxicity cell models, the data sets from the drug-induced toxicity cell models, and the causal relationship networks generated from the drug-induced toxicity cell models by using the Platform Technology methods can constantly evolve and build upon previous knowledge obtained from the Platform Technology.

The present invention is based, at least in part, on the identification of novel biomarkers that are associated with drug-induced cardiotoxicity. The invention is further based, at least in part, on the discovery that Coenzyme Q10 is capable of reducing or preventing drug-induced cardiotoxicity.

Accordingly, the invention provides methods for identifying an agent that causes or is at risk for causing cardiotoxicity. In one embodiment, the agent is a drug or drug candidate. In one embodiment, the toxicity is drug-induced toxicity, e.g., cardiotoxicity. In one embodiment, the agent is a drug or drug candidate for treating diabetes, obesity, a cardiovascular disorder, cancer, a neurological disorder, or an inflammatory disorder. In these methods, the amount of one or more biomarkers/proteins in a pair of samples (a

first sample not subject to the drug treatment, and a second sample subjected to the drug treatment) is assessed. A modulation in the level, expression level, or activity of the one or more biomarkers in the second sample as compared to the level of expression of the one or more biomarkers in the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity. In one embodiment, the one or more biomarkers is selected from the markers listed in table 2. The methods of the present invention can be practiced in conjunction with any other method used by the skilled practitioner to identify a drug at risk for causing drug-induced cardiotoxocity.

In one embodiment, a drug that may be used in the methods of the invention includes, but is not limited to, Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab, Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, and TNF antagonists.

Accordingly, in one aspect, the invention provides a method for identifying a drug that causes or is at risk for causing drug-induced cardiotoxicity, comprising: comparing (i) the level of expression of one or more biomarkers present in a first cell sample obtained prior to the treatment with the drug; with (ii) the level of expression of the one or more biomarkers present in a second cell sample obtained following the treatment with the drug; wherein the one or more biomarkers is selected from the markers listed in table 2; wherein a modulation in the level of expression of the one or more biomarkers in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.
In one embodiment, the cells are cells of the cardiovascular system, e.g., cardiomyocytes. In one embodiment, the cells are diabetic cardiomyocytes. In one embodiment, the drug is a drug or candidate drug for treating diabetes, obesity,cardiovascular disease, cancer, neurological disorder, or inflammatory disorder. In one embodiment, the drug is any one of Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab, Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, and TNF antagonists.

In one embodiment, a modulation (e.g., an increase or a decrease) in the level of expression of one, two, three, four, five, six, seven, eight, nine,ten, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the biomarkers selected from the markers listed in table 2 in the second

sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.

In one embodiment, a modulation (e.g., an increase or a decrease) in the level of expression of a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4 in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.

Methods for identifying a rescue agent that can reduce or prevent drug-induced cardiotoxicity are also provided by the invention. In these methods, the amount of one or more biomarkers in three samples (a first sample not subjected to the drug treatment, a second sample subjected to the drug treatment, and a third sample subjected both to the drug treatment and the agent) is assessed. A normalized level of expression of the one or more biomarkers in the third sample as compared to the first sample, with a change of expression in the second example treated with the drug, is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity. In one embodiment, the one or more biomarkers is selected from the markers listed in table 2.

Using the methods described herein, a variety of molecules, particularly including molecules sufficiently small to be able to cross the cell membrane, may be screened in order to identify molecules which modulate, *e.g.,* increase or decrease the expression and/or activity of a marker of the invention. Compounds so identified can be provided to a subject in order to reduce, alleviate or prevent drug-induced toxicity in the subject.

Accordingly, in another aspect, the invention provides a method for identifying a rescue agent that can reduce or prevent drug-induced cardiotoxicity comprising: (i) determining a normal level of expression of one or more biomarkers present in a first cell sample obtained prior to the treatment with a cardiotoxicity inducing drug; (ii) determining a treated level of expression of the one or more biomarkers present in a second cell sample obtained following the treatment with the cardiotoxicity inducing drug to identify one or more biomarkers with a change of expression in the treated cell sample; (iii) determining the level of expression of the one or more biomarkers with a changed level of expression in the cardiotoxicity inducing drug treated sample present in a third cell sample obtained following the treatment with the cardiotoxicity inducing

drug and the rescue agent; and (iv) comparing the level of expression of the one or more biomarkers determined in the third sample with the level of expression of the one or more biomarkers present in the first sample; wherein the one or more biomarkers is selected from the markers listed in table 2; and wherein a normalized level of expression of the one or more biomarkers in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity.

In one embodiment, the cells are cells of the cardiovascular system, e.g., cardiomyocytes. In one embodiment, the cells are diabetic cardiomyocytes. In one embodiment, the drug is a drug or candidate drug for treating diabetes, obesity,cardiovascular disease, cancer, neurological disorder, or inflammatory disorder. In one embodiment, the drug is is any one of Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab, Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, and TNF antagonists. In one embodiment, about the same level of expression of one, two, three, four, five, six, seven, eight, nine, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the biomarkers selected from the markers listed in table 2 in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity.

In one embodiment, a normalized level of expression of a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen,markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4, in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity.

The invention further provides methods for alleviating, reducing or preventing drug-induced cardiotoxicity in a subject in need thereof, comprising administering to a subject (e.g., a mammal, a human, or a non-human animal) an agent identified by the screening methods provided herein, thereby reducing or preventing drug-induced cardiotoxicity in the subject. In one embodiment, the agent is administered to a subject that has already been treated with a cardiotoxicity-inducing drug. In one embodiment, the agent is administered to a subject at the same time as treatment of the subject with a

cardiotoxicity-inducing drug. In one embodiment, the agent is administered to a subject prior to treatment of the subject with a cardiotoxicity-inducing drug.

The invention further provides methods for alleviating, reducing or preventing drug-induced cardiotoxicity in a subject in need thereof, comprising administering Coenzyme Q10 to the subject (e.g., a mammal, a human, or a non-human animal), thereby reducing or preventing drug-induced cardiotoxicity in the subject. In one embodiment, the Coenzyme Q10 is administered to a subject that has already been treated with a cardiotoxicity-inducing drug. In one embodiment, the Coenzyme Q10 is administered to a subject at the same time as treatment of the subject with a cardiotoxicity-inducing drug. In one embodiment, the Coenzyme Q10 is administered to a subject prior to treatment of the subject with a cardiotoxicity-inducing drug. In one embodiment, the drug-induced cardiotoxicity is associated with modulation of expression of one, two, three, four, five, six, seven, eight, nine, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the biomarkers selected from the markers listed in table 2. All values presented in the foregoing list can also be the upper or lower limit of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 2 and 5, 2 and 10, or 5 and 10 of the foregoing genes (or proteins).

In one embodiment, the drug-induced cardiotoxicity is cardiomyopathy, heart failure, atrial fibrillation, cardiomyopathy and heart failure, heart failure and LV dysfunction, atrial flutter and fibrillation, or heart valve damage and heart failure.

The invention further provides biomarkers (e.g, genes and/or proteins) that are useful as predictive markers for drug-induced cardiotoxicity. These biomarkers include the markers listed in table 2.

In one embodiment, the drug-induced cardiotoxicity is associated with modulation of a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen, markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4.

In one embodiment, the predictive markers for drug-induced cardiotoxicity is a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4.

The ordinary skilled artisan would, however, be able to identify additional biomarkers predictive of drug-induced cardiotoxicity by employing the methods described herein, e.g., by carrying out the methods described in Example 3 but by using a different drug known to induce cardiotoxicity. Exemplary drug-induced cardiotoxicity biomarkers of the invention are further described below.

In one aspect, the invention relates to a method for identifying a modulator of adrug-induced toxicity, said method comprising: (1) establishing a model for drug-induced toxicity, using cells associated with drug-induced toxicity, to represents a characteristic aspect of drug-induced toxicity; (2) obtaining a first data set from the model for drug-induced toxicity, wherein the first data set represents one or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data characterizing the cells associated with drug-induced toxicity; (3) obtaining a second data set from the model for drug-induced toxicity, wherein the second data set represents a functional activity or a cellular response of the cells associated with drug-induced toxicity; (4) generating a consensus causal relationship network among the expression levels of the one or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data and the functional activity or cellular response based solely on the first data set and the second data set using a programmed computing device, wherein the generation of the consensus causal relationship network is not based on any known biological relationships other than the first data set and the second data set; (5) identifying, from the consensus causal relationship network, a causal relationship unique in drug-induced toxicity, wherein a gene, lipid, protein, metabolite, transcript, or SNP associated with the unique causal relationship is identified as a modulator of drug-induced toxicity.

In certain embodiments, the modulator stimulates or promotes the drug-induced toxicity.

In certain embodiments, the modulator inhibits the drug-induced toxicity.

In certain embodiments, the model of the drug-induced toxicity comprises an in vitro culture of cells associated with the drug-induced toxicity, optionally further comprising a matching in vitro culture of control cells.

In certain embodiments, the in vitro culture of the cells is subject to an environmental perturbation, and the in vitro culture of the matching control cells is identical cells not subject to the environmental perturbation.

In certain embodiments, the environmental perturbation comprises one or more of a contact with an agent, a change in culture condition, an introduced genetic modification / mutation, and a vehicle (*e.g.*, vector) that causes a genetic modification / mutation.

In certain embodiments, the first data set comprises protein and/or mRNA expression levels of the plurality of genes.

In certain embodiments, the first data set further comprises two or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data. In certain embodiments, the first data set further comprises three or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data.

In certain embodiments, the second data set representing the functional activity or cellular response of the cells comprises one or more of bioenergetics, cell proliferation, apoptosis, organellar function, a genotype-phenotype association actualized by functional models selected from ATP, ROS, OXPHOS, and Seahorse assays, global enzyme activity, and an effect of global enzyme activity on the enzyme metabolic substrates of cells associated with drug-induced toxicity. In one embodiment, the global enzyme activity is global kinase activity. In one embodiment, the effect of global enzyme activity on the enzyme metabolic substrates is the phospho proteome of the cell.

In certain embodiments, step (4) is carried out by an artificial intelligence (AI) - based informatics platform.

In certain embodiments, the AI-based informatics platform comprises REFS(TM).

In certain embodiments, the AI-based informatics platform receives all data input from the first data set and the second data set without applying a statistical cut-off point.

In certain embodiments, the consensus causal relationship network established in step (4) is further refined to a simulation causal relationship network, before step (5), by *in silico* simulation based on input data, to provide a confidence level of prediction for one or more causal relationships within the consensus causal relationship network.

In certain embodiments, the unique causal relationship is identified as part of a differential causal relationship network that is uniquely present in cells, and absent in the matching control cells.

In one embodiment, the unique causal relationship identified is a relationship between at least one pair selected from the group consisting of expression of a gene and level of a lipid; expression of a gene and level of a transcript; expression of a gene and level of a metabolite; expression of a first gene and a second gene; expression of a gene and presence of a SNP; expression of a gene and a functional activity; level of a lipid and level of a transcript; level of a lipid and level of a metabolite; level of a first lipid and a second lipid; level of a lipid and presence of a SNP; level of a lipid and a functional activity; level of a first transcript and level of a second transcript; level of a transcript and level of a metabolite; level of a transcript and presence of a SNP; level of a first transcript and a functional activity; level of a first metabolite and level of a second metabolite; level of a metabolite and presence of a SNP; level of a metabolite and a functional activity; level of a first SNP and presence of a second SNP; and presence of a SNP and a functional activity.

In one embodiment, the functional activity is selected from the group consisting of bioenergetics, cell proliferation, apoptosis, organellar function, kinase activity, protease activity, and a genotype-phenotype association actualized by functional models selected from ATP, ROS, OXPHOS, and Seahorse assays. In certain embodiments, the method further comprising validating the identified unique causal relationship in a drug-indiced toxicity model.

In one embodiment, the drug-induced toxicity is drug-induced cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity or myotoxicity.

In one embodiment, the drug-induced cardiotoxicity is cardiomyopathy, heart failure, atrial fibrillation, cardiomyopathy and heart failure, heart failure and LV dysfunction, atrial flutter and fibrillation, or, heart valve damage and heart failure.

In one embodiment, the model for drug-induced toxicity comprises cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neuronal cells, renal cells, or myoblasts.

In one embodiment, the model for drug-induced toxicity comprises a toxicity inducing drug, cancer drug, diabetic drug, neurological drug, or anti-inflammatory drug. In one embodiment, the drug is Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab,

Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, or TNF antagonists.

In one aspect, the invention provides a method for identifying a drug that causes or is at risk for causing drug-induced toxicity, comprising: comparing (i) a level of one or more biomarkers present in a first cell sample obtained prior to the treatment with the drug; with (ii) a level of the one or more biomarkers present in a second cell sample obtained following the treatment with the drug; wherein the one or more biomarkers is selected from the modulators identified by the methods described above; wherein a modulation in the level of the one or more biomarkers in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced toxicity.

In one aspect, the invention provides a method for identifying a rescue agent that can reduce or prevent drug-induced toxicity comprising: (i) determining a normal level of one or more biomarkers present in a first cell sample obtained prior to the treatment with a toxicity inducing drug; (ii) determining a treated level of the one or more biomarkers present in a second cell sample obtained following the treatment with the toxicity inducing drug to identify one or more biomarkers with a change of level in the treated cell sample; (iii) determining the level of the one or more biomarkers with a changed level in the toxicity inducing drug treated sample present in a third cell sample obtained following the treatment with the toxicity inducing drug and the rescue agent; and (iv) comparing the level of the one or more biomarkers determined in the third sample with the level of the one or more biomarkers present in the first sample; wherein the one or more biomarkers is selected from the modulators identified by the methods described above and wherein a normalized level of the one or more biomarkers in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced toxicity.

In another aspect, the invention relates to a method for alleviating, reducing or preventing drug-induced toxicity, comprising administering to a subject the rescue agent identified by the methods described above, thereby reducing or preventing drug-induced toxicity in the subject.

In another aspect, the invention relates to a method for providing a model for drug-induced toxicity for use in a platform method, comprising: establishing a drug-induced toxicity model, using cells associated with the drug-induced toxicity, to represent a characteristic aspect of the drug-induced toxicity, wherein the model for the drug-induced toxicity is useful for generating data sets used in the platform method; thereby providing a model for drug-induced toxicity for use in a platform method.

In one embodiment, the model for drug-induced toxicity comprises cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neuronal cells, renal cells, or myoblasts.

In another aspect, the invention relates to a method for obtaining a first data set and second data set from a model for drug-induced toxicity for use in a platform method, comprising: (1) obtaining a first data set from the model for drug-induced toxicity for use in a platform method, wherein the model for the drug-induced toxicity comprises cells associated with the drug-induced toxicity, and wherein the first data set represents expression levels of a plurality of genes in the cells associated with the drug-induced toxicity; (2) obtaining a second data set from the model for drug-induced toxicity for use in the platform method, wherein the second data set represents a functional activity or a cellular response of the cells associated with the drug-induced toxicity; thereby obtaining a first data set and second data set from the model for the drug-induced toxicity for use in a platform method.

In another aspect, the invention relates to a method for identifying a modulator of drug-induced toxicity, said method comprising: (1) generating a consensus causal relationship network among a first data set and second data set obtained from a model for the drug-induced toxicity, wherein the model comprises cells associated with the drug-induced toxicity, and wherein the first data set represents expression levels of a plurality of genes in the cells and the second data set represents a functional activity or a cellular response of the cells, using a programmed computing device, wherein the generation of the consensus causal relationship network is not based on any known biological relationships other than the first data set and the second data set; (2) identifying, from the consensus causal relationship network, a causal relationship unique in the drug-induced toxicity, wherein a gene associated with the unique causal relationship is identified as a modulator of the drug-induced toxicity; thereby identifying a modulator of drug-induced toxicity.

In another aspect, the invention relates to a method for identifying a modulator of a drug-induced toxicity, said method comprising: 1) providing a consensus causal relationship network generated from a model for the drug-induced toxicity; 2) identifying, from the consensus causal relationship network, a causal relationship unique in the drug-induced toxicity, wherein a gene associated with the unique causal relationship is identified as a modulator of the drug-induced toxicity; thereby identifying a modulator of a drug-induced toxicity.

In certain embodiments of the various methods, the consensus causal relationship network is generated among a first data set and second data set obtained from the model for the drug-induced toxicity, wherein the model comprises cells associated with the drug-induced toxicity, and wherein the first data set represents expression levels of a plurality of genes in the cells and the second data set represents a functional activity or a cellular response of the cells, using a programmed computing device, wherein the generation of the consensus causal relationship network is not based on any known biological relationships other than the first data set and the second data set.

In certain embodiments, the "environmental perturbation", also referred to herein as "external stimulus component", is a therapeutic agent. In certain embodiments, the external stimulus component is a small molecule (*e.g.*, a small molecule of no more than 5 kDa, 4 kDa, 3 kDa, 2 kDa, 1 kDa, 500 Dalton, or 250 Dalton). In certain embodiments, the external stimulus component is a biologic. In certain embodiments, the external stimulus component is a chemical. In certain embodiments, the external stimulus component is endogenous or exogenous to cells. In certain embodiments, the external stimulus component is a MIM or epishifter. In certain embodiments, the external stimulus component is a stress factor for the cell system, such as hypoxia, hyperglycemia, hyperlipidemia, hyperinsulinemia, and/or lactic acid rich conditions.

In certain embodiments, the external stimulus component may include a therapeutic agent or a candidate therapeutic agent for treating a drug-induced toxicity, including chemotherapeutic agent, protein-based biological drugs, antibodies, fusion proteins, small molecule drugs, lipids, polysaccharides, nucleic acids, *etc.*

In certain embodiments, the external stimulus component may be one or more stress factors, such as those typically encountered *in vivo* under the various drug-induced

toxicities, including hypoxia, hyperglycemic conditions, acidic environment (that may be mimicked by lactic acid treatment), *etc*.

In other embodiments, the external stimulus component may include one or more MIMs and/or epishifters, as defined herein below. Exemplary MIMs include Coenzyme Q10 (also referred to herein as CoQ10) and compounds in the Vitamin B family, or nucleosides, mononucleotides or dinucleotides that comprise a compound in the Vitamin B family.

In making cellular output measurements (such as protein expression), either absolute amount (e.g., expression amount) or relative level (e.g., relative expression level) may be used. In one embodiment, absolute amounts (e.g., expression amounts) are used. In one embodiment, relative levels or amounts (e.g., relative expression levels) are used. For example, to determine the relative protein expression level of a cell system, the amount of any given protein in the cell system, with or without the external stimulus to the cell system, may be compared to a suitable control cell line or mixture of cell lines (such as all cells used in the same experiment) and given a fold-increase or fold-decrease value. The skilled person will appreciate that absolute amounts or relative amounts can be employed in any cellular output measurement, such as gene and/or RNA transcription level, level of lipid, level of metabolite, or any functional output, *e.g.*, level of apoptosis, level of toxicity, level of enzyme (e.g., kinase) activity, or ECAR or OCR as described herein. A pre-determined threshold level for a fold-increase (*e.g.*, at least 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 or 100 or more fold increase) or fold-decrease (*e.g.*, at least a decrease to 0.9, 0.8, 0.75, 0.7, 0.6, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1 or 0.05 fold, or a decrease to 90%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10% or 5% or less) may be used to select significant differentials, and the cellular output data for the significant differentials may then be included in the data sets (e.g., first and second data sets) utilized in the platform technology methods of the invention. All values presented in the foregoing list can also be the upper or lower limit of ranges, *e.g.*, between 1.5 and 5 fold, 5 and 10 fold, 2 and 5 fold, or between 0.9 and 0.7, 0.9 and 0.5, or 0.7 and 0.3 fold, are intended to be a part of this invention.

Throughout the present application, all values presented in a list, e.g., such as those above, can also be the upper or lower limit of ranges that are intended to be a part of this invention.

In one embodiment of the methods of the invention, not every observed causal relationship in a causal relationship network may be of biological significance. With respect to any given drug-induced toxicity for which the subject interrogative biological assessment is applied, some (or maybe all) of the causal relationships (and the genes associated therewith) may be "determinative" with respect to the specific biological problem at issue, *e.g.*, either responsible for causing a drug-induced toxicity (a potential target for therapeutic intervention) or is a biomarker for the drug-induced toxicity (a potential diagnostic or prognostic factor). In one embodiment, an observed causal relationship unique in the drug-induced toxicity is determinative with respect to the specific biological problem at issue. In one embodiment, not every observed causal relationship unique in the drug-induced toxicity is determinative with respect to the specific problem at issue.

Such determinative causal relationships may be selected by an end user of the subject method, or it may be selected by a bioinformatics software program, such as REFS, DAVID-enabled comparative pathway analysis program, or the KEGG pathway analysis program. In certain embodiments, more than one bioinformatics software program is used, and consensus results from two or more bioinformatics software programs are preferred.

As used herein, "differentials" of cellular outputs include differences (*e.g.*, increased or decreased levels) in any one or more parameters of the cellular outputs. In certain embodiments, the differentials are each independently selected from the group consisting of differentials in mRNA transcription, protein expression, lipid expression, protein activity, kinase activity, metabolite / intermediate level, and/or ligand-target interaction. For example, in terms of protein expression level, differentials between two cellular outputs, such as the outputs associated with a cell system before and after the treatment by an external stimulus component, can be measured and quantitated by using art-recognized technologies, such as mass-spectrometry based assays (*e.g.*, iTRAQ, 2D-LC-MSMS, *etc.*).

In one aspect, the cell model for a drug-induced toxicity comprises a cellular cross-talking system, wherein a first cell system having a first cellular environment with an external stimulus component generates a first modified cellular environment; such that a cross-talking cell system is established by exposing a second cell system having a second cellular environment to the first modified cellular environment.

In one embodiment, at least one significant cellular cross-talking differential from the cross-talking cell system is generated; and at least one determinative cellular cross-talking differential is identified such that an interrogative biological assessment occurs. In certain embodiments, the at least one significant cellular cross-talking differential is a plurality of differentials.

In certain embodiments, the at least one determinative cellular cross-talking differential is selected by the end user. Alternatively, in another embodiment, the at least one determinative cellular cross-talking differential is selected by a bioinformatics software program (such as, e.g., REFS, KEGG pathway analysis or DAVID-enabled comparative pathway analysis) based on the quantitative proteomics data.

In certain embodiments, the method further comprises generating a significant cellular output differential for the first cell system.

In certain embodiments, the differentials are each independently selected from the group consisting of differentials in mRNA transcription, protein expression, lipid expression, protein activity, metabolite / intermediate level, and/or ligand-target interaction.

In certain embodiments, the first cell system and the second cell system are independently selected from: a homogeneous population of primary cells, a drug-induced toxicity related cell line, or a normal cell line.

In certain embodiments, the first modified cellular environment comprises factors secreted by the first cell system into the first cellular environment, as a result of contacting the first cell system with the external stimulus component. The factors may comprise secreted proteins or other signaling molecules. In certain embodiments, the first modified cellular environment is substantially free of the original external stimulus component.

In certain embodiments, the cross-talking cell system comprises a transwell having an insert compartment and a well compartment separated by a membrane. For example, the first cell system may grow in the insert compartment (or the well compartment), and the second cell system may grow in the well compartment (or the insert compartment).

In certain embodiments, the cross-talking cell system comprises a first culture for growing the first cell system, and a second culture for growing the second cell system.

In this case, the first modified cellular environment may be a conditioned medium from the first cell system.

In certain embodiments, the first cellular environment and the second cellular environment can be identical. In certain embodiments, the first cellular environment and the second cellular environment can be different.

In certain embodiments, the cross-talking cell system comprises a co-culture of the first cell system and the second cell system.

The methods of the invention may be used for, or applied to, any number of "interrogative biological assessments." Application of the methods of the invention to an interrogative biological assessment allows for the identification of one or more modulators of a drug-induced toxicity or determinative cellular process "drivers" of a drug-induced toxicity.

In one embodiment, the interrogative biological assessment is the assessment of the toxicological profile of an agent, *e.g.*, a drug, on a cell, tissue, organ or organism, wherein the identified modulators of drug-induced toxicity, e.g., determinative cellular process driver (e.g., cellular cross-talk differentials or causal relationships unique in drug-induced toxicity) may be indicators of toxicity, *e.g.*, cytotoxicity, cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity or myotoxicity, and may in turn be used to predict or identify the toxicological profile of the agent. In one embodiment, the identified modulators of a drug-induced toxicity, e.g., determinative cellular process driver (e.g., cellular cross-talk differentials or causal relationships unique in a drug-induced toxicity) is an indicator of cardiotoxicity of a drug or drug candidate, and may in turn be used to predict or identify the cardiotoxicological profile of the drug or drug candidate.

In another aspect, the invention provides a kit for conducting an interrogative biological assessment using a discovery Platform Technology, comprising one or more reagents for detecting the presence of, and/or for quantitating the amount of, an analyte that is the subject of a causal relationship network generated from the methods of the invention. In one embodiment, said analyte is the subject of a unique causal relationship in the drug-induced toxicity, e.g., a gene associated with a unique causal relationhip in the drug-induced toxicity. In certain embodiments, the analyte is a protein, and the reagents comprise an antibody against the protein, a label for the protein, and/or one or

more agents for preparing the protein for high throughput analysis (*e.g.*, mass spectrometry based sequencing).

It should be understood that all embodiments described herein, including those described only in examples, are parts of the general description of the invention, and can be combined with any other embodiments of the invention unless explicitly disclaimed or inapplicable.

Brief Description of the Drawings

Various embodiments of the present disclosure will be described herein below with reference to the figures wherein:

Figure 1: Illustration of approach to identify therapeutics.

Figure 2: Illustration of systems biology of cancer and consequence of integrated multi-physiological interactive output regulation.

Figure 3: Illustration of systematic interrogation of biological relevance using MIMS.

Figure 4: Illustration of modeling cancer network to enable interrogative biological query.

Figure 5: Illustration of the interrogative biology platform technology.

Figure 6: Illustration of technologies employed in the platform technology.

Figure 7: Schematic representation of the components of the platform including data collection, data integration, and data mining.

Figure 8: Schematic representation of the systematic interrogation using MIMS and collection of response data from the "omics" cascade.

Figure 9: Sketch of the components employed to build the *in vitro* models representing normal and diabetic states.

Figure 10: Schematic representation of the informatics platform REFS™ used to generate causal networks of the protein as they relate to disease pathophysiology.

Figure 11: Schematic representation of the approach towards generation of differential network in diabetic versus normal states and diabetic nodes that are restored to normal states by treatment with MIMS.

Figure 12: A representative differential network in diabetic versus normal states.

Figure 13: A schematic representation of a node and associated edges of interest (Node1in the center). The cellular functionality associated with each edge is represented.

Figure 14: High level flow chart of an exemplary method, in accordance with some embodiments.

Figure 15A-15D: High level schematic illustration of the components and process for an AI-based informatics system that may be used with exemplary embodiments.

Figure 16: Flow chart of process in AI-based informatics system that may be used with some exemplary embodiments.

Figure 17: Schematically depicts an exemplary computing environment suitable for practicing exemplary embodiments taught herein.

Figure 18: Illustration of the mathematical approach towards generation of delta-delta networks.

Figure 19: A schematic representing experimental design and modeling parameters used to study drug induced toxicity in diabetic cardiomyocytes.

Figure 20: Dysregulation of transcriptional network and expression of human mitochondrial energy metabolism genes in diabetic cardiomyocytes by drug treatment (T): rescue molecule (R) normalizes gene expression.

Figure 21: A. Drug treatment (T) induced expression of GPAT1 and TAZ in mitochondria from cardiomyocytes conditioned in hyerglycemia. In combination with the rescue molecule (T+R) the levels of GPAT1 and TAZ were normalized. B. Synthesis of TAG from G3P.

Figure 22: A. Drug treatment (T) decreases mitochondrial OCR (oxygen consumption rate) in cardiomyocytes conditioned in hyperglycemia. The rescue molecule (T+R) normalizes OCR. B. Drug treatment (T) represses mitochondrial ATP synthesis in cardiomyocytes conditioned in hyperglycemia.

Figure 23: GO Annotation of proteins down regulated by drug treatment. Proteins involved in mitochondrial energy metabolism were down regulated with drug treatment.

Figure 24: Illustration of the mathematical approach towards generation of delta networks. Compare unique edges from T versus UT both the models being in diabetic environment.

Figure 25: A schematic representing potential protein hubs and networks that drive pathophysiology of drug induced toxicity.

Figure 26: Schematic representation of the Interrogative biology platform.

Figure 27: Illustration of cellular functional models, data integration and mathematical model Building.

Figure 28: Causal molecular interaction network that drives pathophysiology of drug-induced toxicity.

Figure 29: Causal molecular interaction sub-network of PTX3 as the central hub that drives pathophysiology of drug-induced toxicity.

Figure 30: Mitochondria ATP synthesis capacity of cardiomyocutes in normal glucose and high glucose conditions.

Figure 31: Causal molecular interaction network of ATP drivers.

Figure 32: Causal molecular interaction sub-network of ATP drivers with P4HB as the central hub.

Figure 33: Unique edges of causal molecular interaction sub-network of ATP drivers with P4HB as the central hub.

Figure 34: Illustration of functional toxicomics: multi-omics integration.

Attached herewith, as in Appendix A, are the sequences of all biomarkers referenced herein. All of the information associated with the Gene Bank accession numbers listed in Appendix A and through this application are incorporated herein by reference in the verions available on the filing date of this application.

Detailed Description of the Invention

I. Overview

Exemplary embodiments of the present invention incorporate methods that may be performed using an interrogative biology platform ("the Platform") that is a tool for understanding a wide variety of drug-induced toxicities, such as cardiotoxicity,

hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity or myotoxicity, and the key molecular drivers underlying such drug-induced toxicities, including factors that enable a drug-induced toxicity. Some exemplary embodiments include systems that may incorporate at least a portion of, or all of, the Platform. Some exemplary methods may employ at least some of, or all of the Platform. Goals and objectives of some exemplary embodiments involving the platform are generally outlined below for illustrative purposes:

i) to create specific molecular signatures as drivers of critical components of the drug-induced toxicity as they relate to overall pathophysiology of the relevant cells, tissues, and/or organs;

ii) to generate molecular signatures or differential maps pertaining to the drug-induced toxicity, which may help to identify differential molecular signatures that distinguishes one biological state (e.g., a drug-induced toxicity state) versus a different biological stage (e.g., a normal state), and develop understanding of signatures or molecular entities as they arbitrate mechanisms of change between the two biological states (e.g., from normal to drug-induced toxicity state); and,

iii) to investigate the role of "hubs" of molecular activity as potential intervention targets for external control of the drug-induced toxicity (e.g., to use the hub as a potential therapeutic target), or as potential bio-markers for the drug-induced toxicity in question (e.g., drug-induced toxicity specific biomarkers, in prognostic and/or theranostics uses).

Some exemplary methods involving the Drug-induced Toxicity Platform may include one or more of the following features:

1) modeling the drug-induced toxicities (e.g., cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity) and/or components of the drug-induced toxicity (e.g., physiology & pathophysiology associated with toxicities) in one or more models, preferably in vitro models, using cells associated with the drug-induced toxicity. For example, the cells may be human derived cells which normally participate in the drug-induced toxicity in question (e.g., heat muscle cells involved in cardiotoxicity). The model may include various cellular cues / conditions / perturbations that are specific to the drug-induced toxicity. Ideally, the model represents various drug-induced toxicity states and flux components, instead of a static assessment of the drug-induced toxicity condition.

2) profiling mRNA and/or protein signatures using any art-recognized means. For example, quantitative polymerase chain reaction (qPCR) & proteomics analysis tools such as Mass Spectrometry (MS). Such mRNA and protein data sets represent biological reaction to environment / perturbation. Where applicable and possible, lipidomics, metabolomics, and transcriptomics data may also be integrated as supplemental or alternative measures for the drug-induced toxicity in question. SNP analysis is another component that may be used at times in the process. It may be helpful for investigating, for example, whether the SNP or a specific mutation has any effect on the drug-induced toxicity. These variables may be used to describe the drug-induced toxicity, either as a static "snapshot," or as a representation of a dynamic process.

3) assaying for one or more functional activities or cellular responses to cues and perturbations, including but not limited to bioenergetics, cell proliferation, apoptosis, and organellar function. True genotype-phenotype association is actualized by employment of functional models, such as ATP, ROS, OXPHOS, Seahorse assays, etc. Such functional activities can involve global enzyme activity, such as kinase activity, and/or effects of global enzyme activity or the enzyme metabolites or substrates in the cells, e.g., the phosphor proteome of the cells. Such cellular responses represent the reaction of the cells in the drug-induced toxicity process (or models thereof) in response to the corresponding drug-induced toxicity state(s) of the mRNA / protein expression, and any other related states in 2) above.

4) integrating functional assay data thus obtained in 3) with proteomics and other data obtained in 2), and determining protein, gene, lipid, enzyme activity and other functional acitivity associations as driven by causality, by employing artificial intelligence based (AI-based) informatics system or platform. Such an AI-based system is based on, and preferably based only on, the data sets obtained in 2) and/or 3), without resorting to existing knowledge concerning the drug-induced toxicity process. Preferably, no data points are statistically or artificially cut-off. Instead, all obtained data is fed into the AI-system for determining protein, gene, lipid, enzyme activity and other functional acitivity associations. One goal or output of the integration process is one or more differential networks (otherwise may be referred to herein as "delta networks," or, in some cases, "delta-delta networks" as the case may be) between the different biological states (e.g., drug-induced toxicity vs. normal states).

5) profiling the outputs from the AI-based informatics platform to explore each hub of activity as a potential therapeutic target and/or biomarker. Such profiling can be done entirely *in silico* based on the obtained data sets, without resorting to any actual wet-lab experiments.

6) validating hub of activity by employing molecular and cellular techniques. Such post-informatic validation of output with wet-lab cell-based experiments may be optional, but they help to create a full-circle of interrogation.

Any or all of the approaches outlined above may be used in any specific application concerning any drug-induced toxicity, depending, at least in part, on the nature of the specific application. That is, one or more approaches outlined above may be omitted or modified, and one or more additional approaches may be employed, depending on specific application.

Various schematics illustrating the platform are provided. In particular, an illustration of an exemplary approach to identify therapeutics using the platform is depicted in Figure 1. An illustration of systems biology of cancer and the consequence of integrated multi-physiological interactive output regulation is depicted in Figure 2. An illustration of a systematic interrogation of biological relevance using MIMS is depicted in Figure 3. An illustration of modeling a cancer network to enable an interrogative biological query is depicted in Figure 4.
Illustrations of the interrogative biology platform and technologies employed in the platform are depicted in Figures 5 and 6. A schematic representation of the components of the platform including data collection, data integration, and data mining is depicted in Figure 7. A schematic representation of a systematic interrogation using MIMS and collection of response data from the "omics" cascade is depicted in Figure 8.

Figure 14 is a high level flow chart of an exemplary method 10, in which components of an exemplary system that may be used to perform the exemplary method are indicated. Initially, a model (e.g., an in vitro model) is established for a biological process (e.g., a drug-induced toxicityprocess) and/or components of the biological process (e.g., drug-induced toxicity physiology and pathophysiology) using cells normally associated with the process (step 12). For example, the cells may be human-derived cells that normally participate in the biological process (e.g., drug-induced toxicity). The cell model may include various cellular cues, conditions, and/or perturbations that are specific to the biological process (e.g., drug-induced toxicity).

Ideally, the cell model represents various (drug-induced toxicity) states and flux components of the biological process (e.g., drug-induced toxicity), instead of a static assessment of the biological process. The comparison cell model may include control cells or normal cells, e.g., cells not exposed to a drug which induces toxicity. Additional description of the cell models appears below in sections III.A and IV.

A first data set is obtained from the cell model for the biological process (e.g. drug-induced toxicity), which includes information representing, by way of example, expression levels of a plurality of genes (e.g., mRNA and/or protein signatures) (step 16) using any known process or system (e.g., quantitative polymerase chain reaction (qPCR) & proteomics analysis tools such as Mass Spectrometry (MS)).

A third data set is obtained from the comparison cell model for the biological process (e.g. drug-induced toxicity) (step 18). The third data set includes information representing, e.g., expression levels of a plurality of genes in the comparison cells from the comparison cell model.

In certain embodiments of the methods of the invention, these first and third data sets are collectively referred to herein as a "first data set" that represents, e.g., expression levels of a plurality of genes in the cells (all cells including comparison cells) associated with the biological system (e.g. drug-induced toxicity model).

The first data set and third data set may be obtained from one or more mRNA and/or Protein Signature Analysis System(s). The mRNA and protein data in the first and third data sets may represent biological reactions to environment and/or perturbation. Where applicable and possible, lipidomics, metabolomics, and transcriptomics data may also be integrated into the first data set as supplemental or alternative measures for the biological process (e.g. drug-induced toxicity). The SNP analysis is another component that may be used at times in the process. It may be helpful for investigating, for example, whether a single-nucleotide polymorphism (SNP) or a specific mutation has any effect on the biological process (e.g. drug-induced toxicity). The data variables may be used to describe the biological process (e.g. drug-induced toxicity) either as a static "snapshot," or as a representation of a dynamic process. Additional description regarding obtaining information representing expression levels of a plurality of genes in cells appears below in section III.B.

A second data set is obtained from the cell model for the biological process (e.g. drug-induced toxicity), which includes information representing a functional activity or

response of cells (step 20). Similarly, a fourth data set is obtained from the comparison cell model for the biological process (e.g. drug-induced toxicity), which includes information representing a functional activity or response of the comparison cells (step 22).

In certain embodiments of the methods of the invention, these second and fourth data sets are collectively referred to herein as a "second data set" that represents a functional activity or a cellular response of the cells (all cells including comparison cells) associated with the biological system (e.g. drug-induced toxicity).

One or more functional assay systems may be used to obtain information regarding the functional activity or response of cells or of comparison cells. The information regarding functional cellular responses to cues and perturbations may include, but is not limited to, bioenergetics profiling, cell proliferation, apoptosis, and organellar function. Functional models for processes and pathways (e.g., adenosine triphosphate (ATP), reactive oxygen species (ROS), oxidative phosphorylation (OXPHOS), Seahorse assays, etc.,) may be employed to obtain true genotype-phenotype association. Such functional activities can involve global enzyme activity, such as kinase activity, and/or effects of global enzyme activity, or the enzyme metabolites or substrates in the cells, e.g., the phosphor proteome of the cells. The functional activity or cellular responses represent the reaction of the cells in the biological process (or models thereof) in response to the corresponding state(s) of the mRNA / protein expression, and any other related applied conditions or perturbations. Additional information regarding obtaining information representing functional activity or response of cells is provided below in section III.B.

The method also includes generating computer-implemented models of the biological processes (e.g. drug-induced toxicity) in the cells and in the control cells. For example, one or more (e.g., an ensemble of) Bayesian networks of causal relationships between the expression level of the plurality of genes and the functional activity or cellular response may be generated for the cell model (the "generated cell model networks") from the first data set and the second data set (step 24). The generated cell model networks, individually or collectively, include quantitative probabilistic directional information regarding relationships. The generated cell model networks are not based on known biological relationships between gene expression and/or functional activity or cellular response, other than information from the first data

set and second data set. The one or more generated cell model networks may collectively be referred to as a consensus cell model network.

One or more (e.g., an ensemble of) Bayesian networks of causal relationships between the expression level of the plurality of genes and the functional activity or cellular response may be generated for the comparison cell model (the "generated comparison cell model networks") from the first data set and the second data set (step 26). The generated comparison cell model networks, individually or collectively, include quantitative probabilistic directional information regarding relationships. The generated cell networks are not based on known biological relationships between gene expression and/or functional activity or cellular response, other than the information in the first data set and the second data set. The one or more generated comparison model networks may collectively be refered to as a consensus cell model network.

The generated cell model networks and the generated comparison cell model networks may be created using an artificial intelligence based (AI-based) informatics platform. Further details regarding the creation of the generated cell model networks, the creation of the generated comparison cell model networks and the AI-based informatics system appear below in section III.C and in the description of Figures 2A-3.

It should be noted that many different AI-based platforms or systems may be employed to generate the Bayesian networks of causal relationships including quantitative probabilistic directional information. Although certain examples described herein employ one specific commercially available system, i.e., REFS™ (Reverse Engineering/Forward Simulation) from GNS (Cambridge, MA), embodiments are not limited. AI-Based Systems or Platforms suitable to implement some embodiments employ mathematical algorithms to establish causal relationships among the input variables (e.g., the first and second data sets), based only on the input data without taking into consideration prior existing knowledge about any potential, established, and/or verified biological relationships.

For example, the REFS™ AI-based informatics platform utilizes experimentally derived raw (original) or minimally processed input biological data (e.g., genetic, genomic, epigenetic, proteomic, metabolomic, and clinical data), and rapidly performs trillions of calculations to determine how molecules interact with one another in a complete system. The REFS™ AI-based informatics platform performs a reverse engineering process aimed at creating an *in silico* computer-implemented cell model

(e.g., generated cell model networks), based on the input data, that quantitatively represents the underlying biological system (e.g. drug-induced toxicity). Further, hypotheses about the underlying biological system can be developed and rapidly simulated based on the computer-implemented cell model, in order to obtain predictions, accompanied by associated confidence levels, regarding the hypotheses.

With this approach, biological systems are represented by quantitative computer-implemented cell models in which "interventions" are simulated to learn detailed mechanisms of the biological system (e.g., drug-induced toxicity), effective intervention strategies, and/or clinical biomarkers that determine which patients will respond to a given treatment regimen. Conventional bioinformatics and statistical approaches, as well as approaches based on the modeling of known biology, are typically unable to provide these types of insights.

After the generated cell model networks and the generated comparison cell model networks are created, they are compared. One or more causal relationships present in at least some of the generated cell model networks, and absent from, or having at least one significantly different parameter in, the generated comparison cell model networks are identified (step 28). Such a comparison may result in the creation of a differential network. The comparison, identification, and/or differential (delta) network creation may be conducted using a differential network creation module, which is described in further detail below in section III.D and with respect to the description of Figure 18.

In some embodiments, input data sets are from one cell type and one comparison cell type, which creates an ensemble of cell model networks based on the one cell type and another ensemble of comparison cell model networks based on the one comparison control cell type. A differential may be performed between the ensemble of networks of the one cell type and the ensemble of networks of the comparison cell type(s).

In other embodiments, input data sets are from multiple cell types (e.g., two or more cell types that are normally associated with the particular type of drug-induced toxicity and multiple comparison cell types (e.g., two or more normal cell types, e.g., same cells which are not exposed to the drug). An ensemble of cell model networks may be generated for each cell types and each comparison cell type individually, and/or data from the multiple cell types and the multiple comparison cell types may be combined into respective composite data sets. The composite data sets produce an

ensemble of networks corresponding to the multiple cell types (composite data) and another ensemble of networks corresponding to the multiple comparison cell types (comparison composite data). A differential may be performed on the ensemble of networks for the composite data as compared to the ensemble of networks for the comparison composite data.

In some embodiments, a differential may be performed between two different differential networks. This output may be referred to as a delta-delta network, and is described below with respect to Figure 18.

Quantitative relationship information may be identified for each relationship in the generated cell model networks (step 30). Similarly, quantitative relationship information for each relationship in the generated comparison cell model networks may be identified (step 32). The quantitative information regarding the relationship may include a direction indicating causality, a measure of the statistical uncertainty regarding the relationship (e.g., an Area Under the Curve (AUC) statistical measurement), and/or an expression of the quantitative magnitude of the strength of the relationship (e.g., a fold). The various relationships in the generated cell model networks may be profiled using the quantitative relationship information to explore each hub of activity in the networks as a potential therapeutic target and/or biomarker. Such profiling can be done entirely *in silico* based on the results from the generated cell model networks, without resorting to any actual wet-lab experiments.

In some embodiments, a hub of activity in the networks may be validated by employing molecular and cellular techniques. Such post-informatic validation of output with wet-lab cell based experiments need not be performed, but it may help to create a full-circle of interrogation.Figure 15 schematically depicts a simplified high level representation of the functionality of an exemplary AI-based informatics system (e.g., REFS™ AI-based informatics system) and interactions between the AI-based system and other elements or portions of an interrogative biology platform ("the Platform"). In Figure 15A, various data sets obtained from a model for a biological process (e.g., a drug-induced toxicity model), such as drug dosage, treatment dosage, protein expression, mRNA expression, lipid levels, metabolite levels, kinase activity and any of many other associated functional measures (such as OCR, ECAR) are fed into an AI-based system. As shown in Figure 15B, from the input data sets, the AI-system creates a library of "network fragments" that includes variables (e.g., proteins, lipids, kinases and

metabolites) that drive molecular mechanisms in the biological process (e.g., drug-induced toxicity), in a process referred to as Bayesian Fragment Enumeration (Figure 15B).

In Figure 15C, the AI-based system selects a subset of the network fragments in the library and constructs an initial trial network from the fragments. The AI-based system also selects a different subset of the network fragments in the library to construct another initial trial network. Eventually an ensemble of initial trial networks are created (e.g., 1000 networks) from different subsets of network fragments in the library. This process may be termed parallel ensemble sampling. Each trial network in the ensemble is evolved or optimized by adding, subtracting and/or substitution additional network fragments from the library. If additional data is obtained, the additional data may be incorporated into the network fragments in the library and may be incorporated into the ensemble of trial networks through the evolution of each trial network. After completion of the optimization/evolution process, the ensemble of trial networks may be described as the generated cell model networks.

As shown in Figure 15D, the ensemble of generated cell model networks may be used to simulate the behavior of the biological system (e.g. drug-induced toxicity). The simulation may be used to predict behavior of the biological system (e.g. drug-induced toxicity) to changes in conditions, which may be experimentally verified using wet-lab cell-based, or animal-based, experiments. Also, quantitative parameters of relationships in the generated cell model networks may be extracted using the simulation functionality by applying simulated perturbations to each node individually while observing the effects on the other nodes in the generated cell model neworks. Further detail is provided below in section III.C.

The automated reverse engineering process of the AI-based informatics system, which is depicted in Figures 2A-2D, creates an ensemble of generated cell model networks networks that is an unbiased and systematic computer-based model of the cells.

The reverse engineering determines the probabilistic directional network connections between the molecular measurements in the data, and the phenotypic outcomes of interest. The variation in the molecular measurements enables learning of the probabilistic cause and effect relationships between these entities and changes in

endpoints. The machine learning nature of the platform also enables cross training and predictions based on a data set that is constantly evolving.

The network connections between the molecular measurements in the data are "probabilistic," partly because the connection may be based on correlations between the observed data sets "learned" by the computer algorithm. For example, if the expression level of protein X and that of protein Y are positively or negatively correlated, based on statistical analysis of the data set, a causal relationship may be assigned to establish a network connection between proteins X and Y. The reliability of such a putative causal relationship may be further defined by a likelihood of the connection, which can be measured by p-value (e.g., $p < 0.1$, 0.05, 0.01, etc).

The network connections between the molecular measurements in the data are "directional," partly because the network connections between the molecular measurements, as determined by the reverse-engineering process, reflects the cause and effect of the relationship between the connected gene / protein, such that raising the expression level of one protein may cause the expression level of the other to rise or fall, depending on whether the connection is stimulatory or inhibitory.

The network connections between the molecular measurements in the data are "quantitative," partly because the network connections between the molecular measurements, as determined by the process, may be simulated *in silico*, based on the existing data set and the probabilistic measures associated therewith. For example, in the established network connections between the molecular measurements, it may be possible to theoretically increase or decrease (e.g., by 1, 2, 3, 5, 10, 20, 30, 50,100-fold or more) the expression level of a given protein (or a "node" in the network), and quantitatively simulate its effects on other connected proteins in the network.

The network connections between the molecular measurements in the data are "unbiased," at least partly because no data points are statistically or artificially cut-off, and partly because the network connections are based on input data alone, without referring to pre-existing knowledge about the biological process in question.

The network connections between the molecular measurements in the data are "systemic" and (unbiased), partly because all potential connections among all input variables have been systemically explored, for example, in a pair-wise fashion. The reliance on computing power to execute such systemic probing exponentially increases as the number of input variables increases.

In general, an ensemble of ~1,000 networks is usually sufficient to predict probabilistic causal quantitative relationships among all of the measured entities. The ensemble of networks captures uncertainty in the data and enables the calculation of confidence metrics for each model prediction. Predictions generated using the ensemble of networks together, where differences in the predictions from individual networks in the ensemble represent the degree of uncertainty in the prediction. This feature enables the assignment of confidence metrics for predictions of clinical response generated from the model.

Once the models are reverse-engineered, further simulation queries may be conducted on the ensemble of models to determine key molecular drivers for the biological process in question, such as a drug-induced toxicity condition.

Sketch of components employed to build examplary *In vitro* models representing normal and diabetic statesis is depicted in Figure 9. Schematic representation of an examplary informatics platform REFS™ used to generate causal networks of the protein as they relate to disease pathophysiology is depicted in Figure 10. Schematic representation of examplary approach towards generation of differential network in diabetic versus normal states and diabetic nodes that are restored to normal states by treatment with MIMS is depicted in Figure 11. A representative differential network in diabetic versus normal states is depicted in Figure 12. A schematic representation of a node and associated edges of interest (Node1 in the center) and the cellular functionality associated with each edge is depicted in Figure 13.

The invention having been generally described above, the sections below provide more detailed description for various aspects or elements of the general invention, in conjunction with one or more specific biological systems (e.g. drug-induced toxicity) that can be analyzed using the methods herein. It should be noted, however, the specific drug-induced toxicity used for illustration purpose below are not limiting. To the contrary, it is intended that other distinct drug-induced toxicities, including any alternatives, modifications, and equivalents thereof, may be analyzed similarly using the subject Platform technology.

II. Definitions

As used herein, certain terms intended to be specifically defined, but are not already defined in other sections of the specification, are defined herein.

The articles "a" and "an" are used herein to refer to one or to more than one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The term "including" is used herein to mean, and is used interchangeably with, the phrase "including but not limited to."

The term "or" is used herein to mean, and is used interchangeably with, the term "and/or," unless context clearly indicates otherwise.

The term "such as" is used herein to mean, and is used interchangeably, with the phrase "such as but not limited to."

"Metabolic pathway" refers to a sequence of enzyme-mediated reactions that transform one compound to another and provide intermediates and energy for cellular functions. The metabolic pathway can be linear or cyclic or branched.

"Metabolic state" refers to the molecular content of a particular cellular, multicellular or tissue environment at a given point in time as measured by various chemical and biological indicators as they relate to a state of health or disease.

The term "microarray" refers to an array of distinct polynucleotides, oligonucleotides, polypeptides (e.g., antibodies) or peptides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms "disorders" and "diseases" are used inclusively and refer to any deviation from the normal structure or function of any part, organ or system of the body (or any combination thereof). A specific disease is manifested by characteristic symptoms and signs, including biological, chemical and physical changes, and is often associated with a variety of other factors including, but not limited to, demographic, environmental, employment, genetic and medically historical factors. Certain characteristic signs, symptoms, and related factors can be quantitated through a variety of methods to yield important diagnostic information.

The term "drug-induced toxicity" includes but is not limited to cardiotoxicity, hepatotoxicity, hephrotoxicity, neurotoxicity, renaltoxicity or myotoxicity.

The term "cardiotoxicity" refers to a broad range of adverse effects on heart function induced by therapeutic molecules. It may emerge early in pre-clinical studies or become apparent later in the clinical setting. Cardiovascular toxicity described herein includes, but is not limited to, any one or more of increased QT duration, arrhythmias, myocardial ischemia, hypertension and thromboembolic complications, myocardial dysfunction, cardiomyopathy, heart failure, atrial fibrillation, cardiomyopathy and heart failure, heart failure and LV dysfunction, atrial flutter and fibrillation, and, heart valve damage and heart failure.

The term "expression" includes the process by which a polypeptide is produced from polynucleotides, such as DNA. The process may involves the transcription of a gene into mRNA and the translation of this mRNA into a polypeptide. Depending on the context in which it is used, "expression" may refer to the production of RNA, protein or both.

The terms "level of expression of a gene" or "gene expression level" refer to the level of mRNA, as well as pre-mRNA nascent transcript(s), transcript processing intermediates, mature mRNA(s) and degradation products, or the level of protein, encoded by the gene in the cell.

The term "modulation" refers to upregulation (i.e., activation or stimulation), downregulation (i.e., inhibition or suppression) of a response, or the two in combination or apart. A "modulator" is a compound or molecule that modulates, and may be, e.g., an agonist, antagonist, activator, stimulator, suppressor, or inhibitor.

"Normal level" of a protein, a lipid, a transcript, a metabolite, or gene expression refers to the level of the protein, lipid, transcript, metabolite, or gene expression prior to contacting the cells with the drug with the potentially toxic drug. A "normal level" can be determined in cells grown under various conditions, e.g., hyperglycemia, hypoxia, if the toxicity of the drug is to be tested under the same conditions.

"Modulated level" refers to a changed value relative to the normal level which is based on historical normal control samples or preferably normal control samples tested in the same experiment. The specific "normal" value will depend, for example, on the type of assay (e.g., ELISA, enzyme activity, immunohistochemistry, PCR), the sample to be tested (e.g., cell type and culture conditions), and other considerations known to

those of skill in the art. Control samples can be used to define cut-offs between normal and abnormal.

A drug is considered to be toxic if treatment of cells with the drug results in a statistically significant change in the level of at least one marker relative to a "normal" or appropriate control level. It is understood that not all concentrations of a drug must result in a statistically significant change in the level of the at least one marker. In a preferred embodiment, a drug is considered to potentially have toxicities if a therapeutically relevant concentration of the drug results in a statistically significant change in the level of at least on marker.

A "rescue agent" is considered to be effective in reducing toxicity if the level of the marker is modulated in a statistically significant manner towards the marker level in the "normal cells" when the rescue agent is present at a therapeutically relevant concentration. In a preferred embodiment, the rescue agent returns the marker to a level that is not statistically different from the level of the marker in the control cells.

The term "control level" refers to an accepted or pre-determined level of a marker, or preferably the marker level determined in a control sample tested in parallel with the test sample, which is used to compare with the level of a marker in a sample derived from cells not treated with the potentially toxic drug or rescue agent. A "control level" is obtained from cells that are cultured under the same conditions, e.g., hypoxia, hyperglycemia, lactic acid, etc.

The term "Trolamine," as used herein, refers to Trolamine NF, Triethanolamine, TEALAN®, TEAlan 99%, Triethanolamine, 99%, Triethanolamine, NF or Triethanolamine, 99%, NF. These terms may be used interchangeably herein.

The term "genome" refers to the entirety of a biological entity's (cell, tissue, organ, system, organism) genetic information. It is encoded either in DNA or RNA (in certain viruses, for example). The genome includes both the genes and the non-coding sequences of the DNA.

The term "proteome" refers to the entire set of proteins expressed by a genome, a cell, a tissue, or an organism at a given time. More specifically, it may refer to the entire set of expressed proteins in a given type of cells or an organism at a given time under defined conditions. Proteome may include protein variants due to, for example,

alternative splicing of genes and/or post-translational modifications (such as glycosylation or phosphorylation).

The term "transcriptome" refers to the entire set of transcribed RNA molecules, including mRNA, rRNA, tRNA, microRNA and other non-coding RNA produced in one or a population of cells at a given time. The term can be applied to the total set of transcripts in a given organism, or to the specific subset of transcripts present in a particular cell type. Unlike the genome, which is roughly fixed for a given cell line (excluding mutations), the transcriptome can vary with external environmental conditions. Because it includes all mRNA transcripts in the cell, the transcriptome reflects the genes that are being actively expressed at any given time, with the exception of mRNA degradation phenomena such as transcriptional attenuation.

The study of transcriptomics, also referred to as expression profiling, examines the expression level of mRNAs in a given cell population, often using high-throughput techniques based on DNA microarray technology.

The term "metabolome" refers to the complete set of small-molecule metabolites (such as metabolic intermediates, hormones and other signalling molecules, and secondary metabolites) to be found within a biological sample, such as a single organism, at a given time under a given condition. The metabolome is dynamic, and may change from second to second.

The term "lipidome" refers to the complete set of lipids to be found within a biological sample, such as a single organism, at a given time under a given condition. The lipidome is dynamic, and may change from second to second.

The term "interactome" refers to the whole set of molecular interactions in a biological system under study (*e.g.*, cells). It can be displayed as a directed graph. Molecular interactions can occur between molecules belonging to different biochemical families (proteins, nucleic acids, lipids, carbohydrates, *etc.*) and also within a given family. When spoken in terms of proteomics, interactome refers to protein-protein interaction network (PPI), or protein interaction network (PIN). Another extensively studied type of interactome is the protein-DNA interactome (network formed by transcription factors (and DNA or chromatin regulatory proteins) and their target genes.

The term "cellular output" includes a collection of parameters, preferably measurable parameters, relating to cellullar status, including (without limiting): level of transcription for one or more genes (*e.g.*, measurable by RT-PCR, qPCR, microarray,

etc.), level of expression for one or more proteins (*e.g.*, measurable by mass spectrometry or Western blot), absolute activity (*e.g.*, measurable as substrate conversion rates) or relative activity (*e.g.*, measurable as a % value compared to maximum activity) of one or more enzymes or proteins, level of one or more metabolites or intermediates, level of oxidative phosphorylation (*e.g.*, measurable by Oxygen Consumption Rate or OCR), level of glycolysis (*e.g.*, measurable by Extra Cellular Acidification Rate or ECAR), extent of ligand-target binding or interaction, activity of extracellular secreted molecules, *etc.* The cellular output may include data for a pre-determined number of target genes or proteins, *etc.*, or may include a global assessment for all detectable genes or proteins. For example, mass spectrometry may be used to identify and/or quantitate all detectable proteins expressed in a given sample or cell population, without prior knowledge as to whether any specific protein may be expressed in the sample or cell population.

As used herein, a "cell system" includes a population of homogeneous or heterogeneous cells. The cells within the system may be growing *in vivo*, under the natural or physiological environment, or may be growing *in vitro* in, for example, controlled tissue culture environments. The cells within the system may be relatively homogeneous (*e.g.*, no less than 70%, 80%, 90%, 95%, 99%, 99.5%, 99.9% homogeneous), or may contain two or more cell types, such as cell types usually found to grow in close proximity *in vivo*, or cell types that may interact with one another *in vivo* through, *e.g.*, paracrine or other long distance inter-cellular communication. The cells within the cell system may be derived from established cell lines, including cancer cell lines, immortal cell lines, or normal cell lines, or may be primary cells or cells freshly isolated from live tissues or organs.

Cells in the cell system are typically in contact with a "cellular environment" that may provide nutrients, gases (oxygen or CO_2, *etc.*), chemicals, or proteinaceous / non-proteinaceous stimulants that may define the conditions that affect cellular behavior. The cellular environment may be a chemical media with defined chemical components and/or less well-defined tissue extracts or serum components, and may include a specific pH, CO_2 content, pressure, and temperature under which the cells grow. Alternatively, the cellular environment may be the natural or physiological environment found *in vivo* for the specific cell system.

In certain embodiments, a cell environment comprises conditions that simulate an aspect of a biological system or process, e.g., simulate a disease state, process, or environment. Such culture conditions include, for example, hyperglycemia, hypoxia, or lactic-rich conditions. Numerous other such conditions are described herein.

In certain embodiments, a cellular environment for a specific cell system also include certain cell surface features of the cell system, such as the types of receptors or ligands on the cell surface and their respective activities, the structure of carbohydrate or lipid molecules, membrane polarity or fluidity, status of clustering of certain membrane proteins, *etc*. These cell surface features may affect the function of nearby cells, such as cells belonging to a different cell system. In certain other embodiments, however, the cellular environment of a cell system does not include cell surface features of the cell system.

The cellular environment may be altered to become a "modified cellular environment." Alterations may include changes (*e.g.*, increase or decrease) in any one or more component found in the cellular environment, including addition of one or more "external stimulus component" to the cellular environment. The environmental perturbation or external stimulus component may be endogenous to the cellular environment (*e.g.*, the cellular environment contains some levels of the stimulant, and more of the same is added to increase its level), or may be exogenous to the cellular environment (*e.g.*, the stimulant is largely absent from the cellular environment prior to the alteration). The cellular environment may further be altered by secondary changes resulting from adding the external stimulus component, since the external stimulus component may change the cellular output of the cell system, including molecules secreted into the cellular environment by the cell system.

As used herein, "external stimulus component", also referred to herein as "environmental perturbation", include any external physical and/or chemical stimulus that may affect cellular function. This may include any large or small organic or inorganic molecules, natural or synthetic chemicals, temperature shift, pH change, radiation, light (UVA, UVB *etc*.), microwave, sonic wave, electrical current, modulated or unmodulated magnetic fields, *etc*.

The term "**M**ultidimensional **I**ntracellular **M**olecule (**MIM**)", is an isolated version or synthetically produced version of an endogenous molecule that is naturally produced by the body and/or is present in at least one cell of a human. A MIM is

capable of entering a cell and the entry into the cell includes complete or partial entry into the cell as long as the biologically active portion of the molecule wholly enters the cell. MIMs are capable of inducing a signal transduction and/or gene expression mechanism within a cell. MIMs are multidimensional because the molecules have both a therapeutic and a carrier, e.g., drug delivery, effect. MIMs also are multidimensional because the molecules act one way in a disease state and a different way in a normal state. For example, in the case of CoQ-10, administration of CoQ-10 to a melanoma cell in the presence of VEGF leads to a decreased level of Bcl2 which, in turn, leads to a decreased oncogenic potential for the melanoma cell. In contrast, in a normal fibroblast, co-administration of CoQ-10 and VEFG has no effect on the levels of Bcl2.

In one embodiment, a MIM is also an epi-shifter In another embodiment, a MIM is not an epi-shifter. In another embodiment, a MIM is characterized by one or more of the foregoing functions. In another embodiment, a MIM is characterized by two or more of the foregoing functions. In a further embodiment, a MIM is characterized by three or more of the foregoing functions. In yet another embodiment, a MIM is characterized by all of the foregoing functions. The skilled artisan will appreciate that a MIM of the invention is also intended to encompass a mixture of two or more endogenous molecules, wherein the mixture is characterized by one or more of the foregoing functions. The endogenous molecules in the mixture are present at a ratio such that the mixture functions as a MIM.

MIMs can be lipid based or non-lipid based molecules. Examples of MIMs include, but are not limited to, CoQ10, acetyl Co-A, palmityl Co-A, L-carnitine, amino acids such as, for example, tyrosine, phenylalanine, and cysteine. In one embodiment, the MIM is a small molecule. In one embodiment of the invention, the MIM is not CoQ10. MIMs can be routinely identified by one of skill in the art using any of the assays described in detail herein. MIMs are described in further detail in US 12/777,902 (US 2011-0110914), the entire contents of which are expressly incorporated herein by reference.

As used herein, an "epimetabolic shifter" (epi-shifter) is a molecule that modulates the metabolic shift from a healthy (or normal) state to a disease state and vice versa, thereby maintaining or reestablishing cellular, tissue, organ, system and/or host health in a human. Epi-shifters are capable of effectuating normalization in a tissue microenvironment. For example, an epi-shifter includes any molecule which is capable,

when added to or depleted from a cell, of affecting the microenvironment (*e.g.*, the metabolic state) of a cell. The skilled artisan will appreciate that an epi-shifter of the invention is also intended to encompass a mixture of two or more molecules, wherein the mixture is characterized by one or more of the foregoing functions. The molecules in the mixture are present at a ratio such that the mixture functions as an epi-shifter. Examples of epi-shifters include, but are not limited to, CoQ-10; vitamin D3; ECM components such as fibronectin; immunomodulators, such as TNFa or any of the interleukins, e.g., IL-5, IL-12, IL-23; angiogenic factors; and apoptotic factors.

In one embodiment, the epi-shifter also is a MIM. In one embodiment, the epi-shifter is not CoQ10. Epi-shifters can be routinely identified by one of skill in the art using any of the assays described in detail herein. Epi-shifters are described in further detail in US 12/777,902 (US 2011-0110914), the entire contents of which are expressly incorporated herein by reference.

Other terms not explicitly defined in the instant application have meaning as would have been understood by one of ordinary skill in the art.

III. Exemplary Steps and Components of the Platform Technology

For illustration purpose only, the following steps of the subject Platform Technology may be described herein below as an exemplary utility for integrating data obtained from a custom built drug-induced toxicity model, and for identifying novel proteins / pathways driving the pathogenesis of drug-induced toxicity. Relational maps resulting from this analysis provides drug-induced toxicity treatment targets, as well as diagnostic / prognostic markers associated with drug-induced toxicity. However, the subject Platform Technology has general applicability for any drug-induced toxicity, and is not limited to any particular drug-induced toxicityor other specific drug-induced toxicity models.

In addition, although the description below is presented in some portions as discrete steps, it is for illustration purpose and simplicity, and thus, in reality, it does not imply such a rigid order and/or demarcation of steps. Moreover, the steps of the invention may be performed separately, and the invention provided herein is intended to encompass each of the individual steps separately, as well as combinations of one or

more (e.g., any one, two, three, four, five, six or all seven steps) steps of the subject Platform Technology, which may be carried out independently of the remaining steps.

The invention also is intended to include all aspects of the Drug-induced Toxicity Platform Technology as separate components and embodiments of the invention. For example, the generated data sets are intended to be embodiments of the invention. As further examples, the generated causal relationship networks, generated consensus causal relationship networks, and/or generated simulated causal relationship networks, are also intended to be embodiments of the invention. The causal relationships identified as being unique in the drug-induced toxicity system are intended to be embodiments of the invention. Further, the custom built models for a particular drug-induced toxicity system are also intended to be embodiments of the invention. For example, custom built models for a drug-induced toxicity state or process, such as, e.g., a custom built model for toxicity (e.g., cardiotoxicity) of a drug, are also intended to be embodiments of the invention.

A. Custom Model Building

The first step in the Platform Technology is the establishment of a model for a drug-induced toxicity system or process. An example of a drug-induced toxicity system or process is cardiotoxicity. As any other complicated biological process or system, cardiotoxicity is a complicated pathological condition characterized by multiple unique aspects. For example, chronic imbalance in uptake, utilization, organellar biogenesis and secretion in non-adipose tissue (heart and liver) is thought to be at the center of mitochondrial damage and dysfunction and a key player in drug induced cardiotoxicity. To this end, a custom cardiotoxicity model comprising diabetic and normal cardiomyocytes may be established to simulate the environment of cardiotoxicity, e.g., by creating cell culture conditions closely approximating the conditions of a cadiac cell experiencing cardiotoxicity. One or more relevant types of cells may be used in the model, such as, for example, cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neural cells, renal cells, or myoblasts.

One such "environment", or growth stress condition, is hypoxia, a condition typically found in a number of disease states and in late stage diabetes or in cardiovascular disease due to ischemia and poor circulation. Hypoxia can be induced in

cells in cells using art-recognized methods. For example, hypoxia can be induced by placing cell systems in a Modular Incubator Chamber (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), which can be flooded with an industrial gas mix containing 5% CO_2, 2% O_2 and 93% nitrogen. Effects can be measured after a pre-determined period, *e.g.*, at 24 hours after hypoxia treatment, with and without additional external stimulus components (*e.g.*, CoQ10 at 0, 50, or 100 μM).

Likewise, lactic acid treatment of cells mimics a cellular environment where glycolysis activity is high. Lactic acid induced stress can be investigated at a final lactic acid concentration of about 12.5 mM at a pre-determined time, *e.g.*, at 24 hours, with or without additional external stimulus components (*e.g.*, CoQ10 at 0, 50, or 100 μM).

Hyperglycemia is normally a condition found in diabetes. As high glucose is known to alter cellular metabolism, agents for the treatment of diabetes can be tested in cells cultured under hyperglycemic conditions. Exposing subject cells to a typical hyperglycemic condition may include adding 10% culture grade glucose to suitable media, such that the final concentration of glucose in the media is about 22 mM. However, as subjects with type 2 diabetes, are frequently overweight or obese, they are frequently treated for other diseases or conditions with other agents, e.g., arthritis with anti-inflammatory agents, cardiovascular disease with cholesterol lowering, blood pressure lowering, or blood thinning agents. Thus, custom built models can be used to assess drug toxicity in normal subjects as compared to subjects to be treated for a first condition with a first agent that also have other diseases or conditions. For example, cells not exposed or exposed to hyperglycemic conditions can be tested together to detect differential toxicities of agents in subjects with or without diabetes.

Hyperlipidemia is a condition found, for example, in obesity and cardiovascular disease. Hyperlipidemia is also a condition which mimics one aspect of cardiotoxicity. The hyperlipidemic conditions can be provided by culturing cells in media containing 0.15 mM sodium palmitate.

Individual conditions reflecting different aspects of toxicity may be investigated separately in the custom built toxicity model, and/or may be combined together. In one embodiment, combinations of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 or more conditions reflecting or simulating different aspects of toxicity conditions are investigated in the custom built toxicity model. In one embodiment, individual conditions and, in addition, combinations of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25,

30, 40, 50 or more of the conditions reflecting or simulating different aspects of toxicity conditions are investigated in the custom built toxicity model. All values presented in the foregoing list can also be the upper or lower limit of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 1 and 20, 1 and 30, 2 and 5, 2 and 10, 5 and 10, 1 and 20, 5 and 20, 10 and 20, 10 and 25, 10 and 30 or 10 and 50 different conditions.

Listed herein below are a few exemplary combinations of conditions that can be used to treat cells for building drug-induced toxicity models. Other combinations can be readily formulated depending on the specific interrogative biological assessment that is being conducted.

1. Media only
2. 50 μM CTL Coenzyme Q10 (CoQ10)
3. 100 μM CTL Coenzyme Q10
4. 12.5 mM Lactic Acid
5. 12.5 mM Lactic Acid + 50 μM CTL Coenzyme Q10
6. 12.5 mM Lactic Acid + 100 μM CTL Coenzyme Q10
7. Hypoxia
8. Hypoxia + 50 μM CTL Coenzyme Q10
9. Hypoxia + 100 μM CTL Coenzyme Q10
10. Hypoxia + 12.5 mM Lactic Acid
11. Hypoxia + 12.5 mM Lactic Acid + 50 μM CTL Coenzyme Q10
12. Hypoxia + 12.5 mM Lactic Acid + 100 μM CTL Coenzyme Q10
13. Media + 22 mM Glucose
14. 50 μM CTL Coenzyme Q10 + 22 mM Glucose
15. 100 μM CTL Coenzyme Q10 + 22 mM Glucose
16. 12.5 mM Lactic Acid + 22 mM Glucose
17. 12.5 mM Lactic Acid + 22 mM Glucose + 50 μM CTL Coenzyme Q10
18. 12.5 mM Lactic Acid + 22 mM Glucose +100 μM CTL Coenzyme Q10
19. Hypoxia + 22 mM Glucose
20. Hypoxia + 22 mM Glucose + 50 μM CTL Coenzyme Q10
21. Hypoxia + 22 mM Glucose + 100 μM CTL Coenzyme Q10
22. Hypoxia +12.5 mM Lactic Acid + 22 mM Glucose

23. Hypoxia +12.5 mM Lactic Acid + 22 mM Glucose + 50 μM CTL Coenzyme Q10
24. Hypoxia + 12.5 mM Lactic Acid + 22 mM Glucose +100 μM CTL Coenzyme Q10

As a control one or more cell lines (e.g.,cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neural cells, renal cells, or myoblasts) are cultured under control conditions in order to identify toxicity unique proteins or pathways (see below). The control may be the comparison cell model described above.

Multiple cells of the same or different origin (for example, cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neural cells, renal cells, or myoblasts), as opposed to a single cell type, may be included in the toxicity model. In certain situations, cross talk or ECS experiments between different cells (cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neuro cells, renal cells, or myoblasts) may be conducted for several inter-related purposes.

In some embodiments that involve cross talk, experiments conducted on the cell models are designed to determine modulation of cellular state or function of one cell system or population (*e.g.*,cardiomyocytes) by another cell system or population (*e.g.*, diabetic cardiomyocytes) under defined treatment conditions (*e.g.*, hyperglycemia, hypoxia (ischemia)). According to a typical setting, a first cell system / population is contacted by an external stimulus components, such as a candidate molecule (*e.g.*, a small drug molecule, a protein) or a candidate condition (*e.g.*, hypoxia, high glucose environment). In response, the first cell system / population changes its transcriptome, proteome, metabolome, and/or interactome, leading to changes that can be readily detected both inside and outside the cell. For example, changes in transcriptome can be measured by the transcription level of a plurality of target mRNAs; changes in proteome can be measured by the expression level of a plurality of target proteins; and changes in metabolome can be measured by the level of a plurality of target metabolites by assays designed specifically for given metabolites. Alternatively, the above referenced changes in metabolome and/or proteome, at least with respect to certain secreted metabolites or proteins, can also be measured by their effects on the second cell system / population, including the modulation of the transcriptome, proteome, metabolome, and interactome of the second cell system / population. Therefore, the experiments can be used to

identify the effects of the molecule(s) of interest secreted by the first cell system / population on a second cell system / population under different treatment conditions. The experiments can also be used to identify any proteins that are modulated as a result of signaling from the first cell system (in response to the external stimulus component treatment) to another cell system, by, for example, differential screening of proteomics. The same experimental setting can also be adapted for a reverse setting, such that reciprocal effects between the two cell systems can also be assessed. In general, for this type of experiment, the choice of cell line pairs is largely based on the factors such as origin, toxicity state and cellular function.

Although two-cell systems are typically involved in this type of experimental setting, similar experiments can also be designed for more than two cell systems by, for example, immobilizing each distinct cell system on a separate solid support.

Once the custom model is built, one or more "perturbations" may be applied to the system, such as genetic variation from patient to patient, or with / without treatment by certain drugs or pro-drugs. *See* Figure 15D. The effects of such perturbations to the system, including the effect on cells related to drug-induced toxicity, and normal control cells, can be measured using various art-recognized or proprietary means, as described in section III.B below.

In an exemplary experiment, cardiomyocytes are conditioned in hyperglycemia and hyperlipidemia conditions, and in addition with or without an environmental perturbation, specifically treatment by a diabetic drug known for inducing cardiotoxicity and/or a potential rescue agent CoenzymeQ10.

The custom built cell model may be established and used throughout the steps of the Platform Technology of the invention to ultimately identify a causal relationship unique in the drug-induced toxicity system, by carrying out the steps described herein. It will be understood by the skilled artisan, however, that a custom built cell model that is used to generate an initial, "first generation" consensus causal relationship network for a drug-induced toxicity can continually evolve or expand over time, e.g., by the introduction of additional drug-induced toxicity related cell lines and/or additional drug-induced toxicity related conditions. Additional data from the evolved cell model, i.e., data from the newly added portion(s) of the cell model, can be collected. The new data collected from an expanded or evolved cell model, i.e., from newly added portion(s) of the cell model, can then be introduced to the data sets previously used to generate the

"first generation" consensus causal relationship network in order to generate a more robust "second generation" consensus causal relationship network. New causal relationships unique to the drug-induced toxicity can then be identified from the "second generation" consensus causal relationship network. In this way, the evolution of the cell model provides an evolution of the consensus causal relationship networks, thereby providing new and/or more reliable insights into the modulators of the drug-induced toxicity.

Custom models can also be designed to assess toxicity of drugs used in combination. For example, therapeutic agents for the treatment of a number of conditions including cancer, auto-immune disease, or HIV are typically administered as cocktails of combinations of agents. Further, many subjects have multiple, unrelated conditions to be treated simultaneously (e.g., diabetes, arthritis, cardiovascular disease). Models can be built, either in normal cells or in cells subjected to various culture conditions, to identify combinations of agents that may result in toxicities when administered simultaneously. Thus, the methods provided include testing combinations of agents (e.g., 2, 3, 4, 5, 6, 7, 8 or more) together to determine if the combination results in drug related toxicities, including with agents that do not result in toxicities alone.

Models can also be built for "personalized medicine" applications in which the specific combination of drugs being administered or considered for administration can be tested using the methods provided herein to determine if the combination of drugs are likely to have unacceptable toxicities. Such combinations can be tested in various cell types (e.g., cardiac cells, kidney cells, nerve cells, muscle cells, liver cells; either cell lines or primary cells cultured from the subject) grown under various conditions to mimic the subject of interest (e.g., grown in high glucose for a subject with diabetes or hypoxia for a subject with ischemia).

Additional examples of custom built cell models are described in detail herein.

B. Data Collection

In general, two types of data may be collected from any custom built model systems. One type of data (e.g., the first set of data, the third set of data) usually relates to the level of certain macromolecules, such as DNA, RNA, protein, lipid, etc. An exemplary data set in this category is proteomic data (e.g., qualitative and quantitative

data concerning the expression of all or substantially all measurable proteins from a sample). The other type of data is generally functional data (e.g., the second set of data, the fourth set of data) that reflects the phenotypic changes resulting from the changes in the first type of data. Functional activity or cellular response of the cells can include any one or more of bioenergetics, cell proliferation, apoptosis, organellar function, a genotype-phenotype association actualized by functional models selected from ATP, ROS, OXPHOS, and Seahorse assays, global enzyme activity (e.g., global kinase activity), and an effect of global enzyme activity on the enzyme metabolic substrates of cells associated with drug-induced toxicity (e.g., phosphoproteomic data).

With respect to the first type of data, in some example embodiments, quantitative polymerase chain reaction (qPCR) and proteomics are performed to profile changes in cellular mRNA and protein expression by quantitative polymerase chain reaction (qPCR) and proteomics. Total RNA can be isolated using a commercial RNA isolation kit. Following cDNA synthesis, specific commercially available qPCR arrays (e.g., those from SA Biosciences) for disease area or cellular processes such as angiogenesis, apoptosis, and diabetes, may be employed to profile a predetermined set of genes by following a manufacturer's instructions. For example, the Biorad cfx-384 amplification system can be used for all transcriptional profiling experiments. Following data collection (Ct), the final fold change over control can be determined using the δCt method as outlined in manufacturer's protocol. Proteomic sample analysis can be performed as described in subsequent sections.

The subject method may employ large-scale high-throughput quantitative proteomic analysis of hundreds of samples of similar character, and provides the data necessary for identifying the cellular output differentials.

There are numerous art-recognized technologies suitable for this purpose. An exemplary technique, iTRAQ analysis in combination with mass spectrometry, is briefly described below.

The quantitative proteomics approach is based on stable isotope labeling with the 8-plex iTRAQ reagent and 2D-LC MALDI MS/MS for peptide identification and quantification. Quantification with this technique is relative: peptides and proteins are assigned abundance ratios relative to a reference sample. Common reference samples in multiple iTRAQ experiments facilitate the comparison of samples across multiple iTRAQ experiments.

For example, to implement this analysis scheme, six primary samples and two control pool samples can be combined into one 8-plex iTRAQ mix according to the manufacturer's suggestions. This mixture of eight samples then can be fractionated by two-dimensional liquid chromatography; strong cation exchange (SCX) in the first dimension, and reversed-phase HPLC in the second dimension, then can be subjected to mass spectrometric analysis.

A brief overview of exemplary laboratory procedures that can be employed is provided herein.

Protein extraction: Cells can be lysed with 8 M urea lysis buffer with protease inhibitors (Thermo Scientific Halt Protease inhibitor EDTA-free) and incubate on ice for 30 minutes with vertex for 5 seconds every 10 minutes. Lysis can be completed by ultrasonication in 5 seconds pulse. Cell lysates can be centrifuged at 14000 x g for 15 minutes (4 oC) to remove cellular debris. Bradford assay can be performed to determine the protein concentration. 100ug protein from each samples can be reduced (10mM Dithiothreitol (DTT), 55 °C, 1 h), alkylated (25 mM iodoacetamide, room temperature, 30 minutes) and digested with Trypsin (1:25 w/w, 200 mM triethylammonium bicarbonate (TEAB), 37 oC, 16 h).

Secretome sample preparation: 1) In one embodiment, the cells can be cultured in serum free medium: Conditioned media can be concentrated by freeze dryer, reduced (10mM Dithiothreitol (DTT), 55 °C, 1 h), alkylated (25 mM iodoacetamide, at room temperature, incubate for 30 minutes), and then desalted by actone precipitation. Equal amount of proteins from the concentrated conditioned media can be digested with Trypsin (1:25 w/w, 200 mM triethylammonium bicarbonate (TEAB), 37 oC, 16 h).

In one embodiment, the cells can be cultured in serum containing medium: The volume of the medium can be reduced using 3k MWCO Vivaspin columns (GE Healthcare Life Sciences), then can be reconstituted with1xPBS (Invitrogen). Serum albumin can be depleted from all samples using AlbuVoid column (Biotech Support Group, LLC) following the manufacturer's instructions with the modifications of buffer-exchange to optimize for condition medium application.

iTRAQ 8 Plex Labeling: Aliquot from each tryptic digests in each experimental set can be pooled together to create the pooled control sample. Equal aliquots from each sample and the pooled control sample can be labeled by iTRAQ 8 Plex reagents according to the manufacturer's protocols (AB Sciex). The reactions can be combined,

vacuumed to dryness, re-suspended by adding 0.1% formic acid, and analyzed by LC-MS/MS.

2D-NanoLC-MS/MS: All labeled peptides mixtures can be separated by online 2D-nanoLC and analysed by electrospray tandem mass spectrometry. The experiments can be carried out on an Eksigent 2D NanoLC Ultra system connected to an LTQ Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ion source (Thermo Electron, Bremen, Germany).

The peptides mixtures can be injected into a 5 cm SCX column (300µm ID, 5µm, PolySULFOETHYL Aspartamide column from PolyLC, Columbia, MD) with a flow of 4 µL / min and eluted in 10 ion exchange elution segments into a C18 trap column (2.5 cm, 100µm ID, 5µm, 300 Å ProteoPep II from New Objective, Woburn, MA) and washed for 5 min with H2O/0.1%FA. The separation then can be further carried out at 300 nL/min using a gradient of 2-45% B (H2O /0.1%FA (solvent A) and ACN /0.1%FA (solvent B)) for 120 minutes on a 15 cm fused silica column (75µm ID, 5µm, 300 Å ProteoPep II from New Objective, Woburn, MA).

Full scan MS spectra (m/z 300-2000) can be acquired in the Orbitrap with resolution of 30,000. The most intense ions (up to 10) can be sequentially isolated for fragmentation using High energy C-trap Dissociation (HCD) and dynamically exclude for 30 seconds. HCD can be conducted with an isolation width of 1.2 Da. The resulting fragment ions can be scanned in the orbitrap with resolution of 7500. The LTQ Orbitrap Velos can be controlled by Xcalibur 2.1 with foundation 1.0.1.

Peptides/proteins identification and quantification: Peptides and proteins can be identified by automated database searching using Proteome Discoverer software (Thermo Electron) with Mascot search engine against SwissProt database. Search parameters can include 10 ppm for MS tolerance, 0.02 Da for MS2 tolerance, and full trypsin digestion allowing for up to 2 missed cleavages. Carbamidomethylation (C) can be set as the fixed modification. Oxidation (M), TMT6, and deamidation (NQ) can be set as dynamic modifications. Peptides and protein identifications can be filtered with Mascot Significant Threshold ($p<0.05$). The filters can be allowed a 99% confidence level of protein identification (1% FDA).

The Proteome Discoverer software can apply correction factors on the reporter ions, and can reject all quantitation values if not all quantitation channels are present. Relative protein quantitation can be achieved by normalization at the mean intensity.

With respect to the second type of data, in some exemplary embodiments, bioenergetics profiling of cancer and normal models may employ the Seahorse™ XF24 analyzer to enable the understanding of glycolysis and oxidative phosphorylation components.

Specifically, cells can be plated on Seahorse culture plates at optimal densities. These cells can be plated in 100 µl of media or treatment and left in a 37°C incubator with 5% CO_2. Two hours later, when the cells are adhered to the 24 well plate, an additional 150 µl of either media or treatment solution can be added and the plates can be left in the culture incubator overnight. This two step seeding procedure allows for even distribution of cells in the culture plate. Seahorse cartridges that contain the oxygen and pH sensor can be hydrated overnight in the calibrating fluid in a non-CO_2 incubator at 37°C. Three mitochondrial drugs are typically loaded onto three ports in the cartridge. Oligomycin, a complex III inhibitor, FCCP, an uncoupler and Rotenone, a complex I inhibitor can be loaded into ports A, B and C respectively of the cartridge. All stock drugs can be prepared at a 10x concentration in an unbuffered DMEM media. The cartridges can be first incubated with the mitochondrial compounds in a non-CO_2 incubator for about 15 minutes prior to the assay. Seahorse culture plates can be washed in DMEM based unbuffered media that contains glucose at a concentration found in the normal growth media. The cells can be layered with 630 ul of the unbuffered media and can be equilibriated in a non-CO_2 incubator before placing in the Seahorse instrument with a precalibrated cartridge. The instrument can be run for three-four loops with a mix, wait and measure cycle for get a baseline, before injection of drugs through the port is initiated. There can be two loops before the next drug is introduced.

OCR (Oxygen consumption rate) and ECAR (Extracullular Acidification Rate) can be recorded by the electrodes in a 7 µl chamber and can be created with the cartridge pushing against the seahorse culture plate.

C. Data Integration and *in silico* Model Generation

Once relevant data sets have been obtained, integration of data sets and generation of computer-implemented statistical models may be performed using an AI-based informatics system or platform (e.g, the REFS™ platform). For example, an exemplary AI-based system may produce simulation-based networks of protein associations as key drivers of metabolic end points (ECAR/OCR). *See* Figure 15. Some

background details regarding the REFS™ system may be found in Xing et al., "Causal Modeling Using Network Ensemble Simulations of Genetic and Gene Expression Data Predicts Genes Involved in Rheumatoid Arthritis," *PLoS Computational Biology*, vol. 7, issue. 3, 1-19 (March 2011) (e100105) and U.S. Patent 7,512,497 to Periwal, the entire contents of each of which is expressly incorporated herein by reference in its entirety. In essence, as described earlier, the REFS™ system is an AI-based system that employs mathematical algorithms to establish causal relationships among the input variables (e.g., protein expression levels, mRNA expression levels, and the corresponding functional data, such as the OCR / ECAR values measured on Seahorse culture plates). This process is based only on the input data alone, without taking into consideration prior existing knowledge about any potential, established, and/or verified biological relationships.

In particular, a significant advantage of the platform of the invention is that the AI-based system is based on the data sets obtained from the cell model, without resorting to or taking into consideration any existing knowledge in the art concerning the biological process. Further, preferably, no data points are statistically or artificially cut-off and, instead, all obtained data is fed into the AI-system for determining protein associations. Accordingly, the resulting statistical models generated from the platform are unbiased, since they do not take into consideration any known biological relationships.

Specifically, data from the proteomics and ECAR/OCR can be input into the AI-based information system, which builds statistical models based on data associations, as described above. Simulation-based networks of protein associations are then derived for each disease versus normal scenario, including treatments and conditions using the following methods.

A detailed description of an exemplary process for building the generated (e.g., optimized or evolved) networks appears below with respect to Figure 16. As described above, data from the proteomics and functional cell data is input into the AI-based system (step 210). The input data, which may be raw data or minimally processed data, is pre-processed, which may include normalization (e.g., using a quantile function or internal standards) (step 212). The pre-processing may also include imputing missing data values (e.g., by using the K-nearest neighbor (K-NN) algorithm) (step 212).

The pre-processed data is used to construct a network fragment library (step 214). The network fragments define quantitative, continuous relationships among all possible small sets (e.g., 2-3 member sets or 2-4 member sets) of measured variables (input data). The relationships between the variables in a fragment may be linear, logistic, multinomial, dominant or recessive homozygous, etc. The relationship in each fragment is assigned a Bayesian probabilistic score that reflect how likely the candidate relationship is given the input data, and also penalizes the relationship for its mathematical complexity. By scoring all of the possible pairwise and three-way relationships (and in some embodiments also four-way relationships) inferred from the input data, the most likely fragments in the library can be identified (the likely fragments). Quantitative parameters of the relationship are also computed based on the input data and stored for each fragment. Various model types may be used in fragment enumeration including but not limited to linear regression, logistic regression, (Analysis of Variance) ANOVA models, (Analysis of Covariance) ANCOVA models, non-linear/polynomial regression models and even non-parametric regression. The prior assumptions on model parameters may assume Gull distributions or Bayesian Information Criterion (BIC) penalties related to the number of parameters used in the model. In a network inference process, each network in an ensemble of initial trial networks is constructed from a subset of fragments in the fragment library. Each initial trial network in the ensemble of initial trial networks is constructed with a different subset of the fragments from the fragment library (step 216).

An overview of the mathematical representations underlying the Bayesian networks and network fragments, which is based on Xing et al., "Causal Modeling Using Network Ensemble Simulations of Genetic and Gene Expression Data Predicts Genes Involved in Rheumatoid Arthritis," *PLoS Computational Biology*, vol. 7, issue. 3, 1-19 (March 2011) (e100105), is presented below.

A multivariate system with random variables $X = X_1,\ldots,X_n$ may be characterized by a multivariate probability distribution function $P(X_1,\ldots,X_n;\Theta)$, that includes a large number of parameters Θ. The multivariate probability distribution function may be factorized and represented by a product of local conditional probability distributions:

$$P(X_1,\ldots,X_n;\Theta) = \prod_{i-1}^{n} P_i\left(X_i \middle| Y_{j1},\ldots,Y_{jK_i};\Theta_i\right),$$

in which each variable X_i is independent from its non-descendent variables given its K_i parent variables, which are $Y_{j1},...,Y_{jK_i}$. After factorization, each local probability distribution has its own parameters Θ_i.

The multivariate probability distribution function may be factorized in different ways with each particular factorization and corresponding parameters being a distinct probabilistic model. Each particular factorization (model) can be represented by a Directed Acrylic Graph (DAC) having a vertex for each variable X_i and directed edges between vertices representing dependences between variables in the local conditional distributions $P_i\left(X_i\middle|Y_{j1},...,Y_{jK_i}\right)$. Subgraphs of a DAG, each including a vertex and associated directed edges are network fragments.

A model is evolved or optimized by determining the most likely factorization and the most likely parameters given the input data. This may be described as "learning a Bayesian network," or, in other words, given a training set of input data, finding a network that best matches the input data. This is accomplished by using a scoring function that evaluates each network with respect to the input data.

A Bayesian framework is used to determine the likelihood of a factorization given the input data. Bayes Law states that the posterior probability, $P(D|\mathrm{M})$, of a model M, given data D is proportional to the product of the product of the posterior probability of the data given the model assumptions, $P(D|\mathrm{M})$, multiplied by the prior probability of the model, $P(M)$, assuming that the probability of the data, P(D), is constant across models. This is expressed in the following equation:

$$P(M|\mathrm{D}) = \frac{P(D|\mathrm{M}) * P(M)}{P(D)}.$$

The posterior probability of the data assuming the model is the integral of the data likelihood over the prior distribution of parameters:

$$P(D|\mathrm{M}) = \int P(D|\mathrm{M}(\Theta))P(\Theta|M)d\Theta.$$

Assuming all models are equally likely (i.e., that $P(M)$ is a constant), the posterior probability of model M given the data D may be factored into the product of integrals over parameters for each local network fragment $\mathrm{M_i}$ as follows:

$$P(M|D) = \prod_{i=1}^{n} \int P_i\left(X_i \middle| Y_{j1},...,Y_{jK_i};\Theta_i\right)$$

Note that in the equation above, a leading constant term has been omitted. In some embodiments, a Bayesian Information Criterion (BIC), which takes a negative logarithm of the posterior probability of the model $P(D|\mathrm{M})$ may be used to "Score" each model as follows:

$$S_{tot}(M) = -\log P(M|D) = \sum_{i=1}^{n} S(M_i),$$

where the total score S_{tot} for a model M is a sum of the local scores S_i for each local network fragment. The BIC further gives an expression for determining a score each individual network fragment:

$$S(M_i) \approx S_{BIC}(M_i) = S_{MLE}(M_i) + \frac{\kappa(M_i)}{2}\log N$$

where $\kappa(M_i)$ is the number of fitting parameter in model M_i and N is the number of samples (data points). $S_{MLE}(M_i)$ is the negative logarithm of the likelihood function for a network fragment, which may be calculated from the functional relationships used for each network fragment. For a BIC score, the lower the score, the more likely a model fits the input data.

The ensemble of trial networks is globally optimized, which may be described as optimizing or evolving the networks (step 218). For example, the trial networks may be evolved and optimized according to a Metropolis Monte Carlo Sampling alogorithm. Simulated annealing may be used to optimize or evolve each trial network in the ensemble through local transformations. In an example simulated annealing processes, each trial network is changed by adding a network fragment from the library, by deleted a network fragment from the trial network, by substituting a network fragment or by otherwise changing network topology, and then a new score for the network is calculated. Generally speaking, if the score improves, the change is kept and if the score worsens the change is rejected. A "temperature" parameter allows some local changes which worsen the score to be kept, which aids the optimization process in avoiding some local minima. The "temperature" parameter is decreased over time to allow the optimization/evolution process to converge.

All or part of the network inference process may be conducted in parallel for the trial different networks. Each network may be optimized in parallel on a separate processor and/or on a separate computing device. In some embodiments, the optimization process may be conducted on a supercomputer incorporating hundreds to thousands of processors which operate in parallel. Information may be shared among the optimization processes conducted on parallel processors.

The optimization process may include a network filter that drops any networks from the ensemble that fail to meet a threshold standard for overall score. The dropped network may be replaced by a new initial network. Further any networks that are not "scale free" may be dropped from the ensemble. After the ensemble of networks has been optimized or evolved, the result may be termed an ensemble of generated cell model networks, which may be collectively referred to as the generated consensus network.

D. Simulation to Extract Quantitative Relationship Information and for Prediction

Simulation may be used to extract quantitative parameter information regarding each relationship in the generated cell model networks (step 220). For example, the simulation for quantitative information extraction may involve perturbing (increasing or decreasing) each node in the network by 10 fold and calculating the posterior distributions for the other nodes (e.g., proteins) in the models. The endpoints are compared by t-test with the assumption of 100 samples per group and the 0.01 significance cut-off. The t-test statistic is the median of 100 t-tests. Through use of this simulation technique, an AUC (area under the curve) representing the strength of prediction and fold change representing the *in silico* magnitude of a node driving an end point are generated for each relationship in the ensemble of networks.

A relationship quantification module of a local computer system may be employed to direct the AI-based system to perform the perturbations and to extract the AUC information and fold information. The extracted quantitative information may include fold change and AUC for each edge connecting a parent note to a child node. In some embodiments, a custom-built R program may be used to extract the quantitative information.

In some embodiments, the ensemble of generated cell model networks can be used through simulation to predict responses to changes in conditions, which may be later verified though wet-lab cell-based, or animal-based, experiments.

The output of the AI-based system may be quantitative relationship parameters and/or other simulation predictions (222).

E. Generation of Differential (Delta) Networks

A differential network creation module may be used to generate differential (delta) networks between generated cell model networks and generated comparison cell model networks. As described above, in some embodiments, the differential network compares all of the quantitative parameters of the relationships in the generated cell model networks and the generated comparison cell model network. The quantitative parameters for each relationship in the differential network are based on the comparison. In some embodiments, a differential may be performed between various differential networks, which may be termed a delta-delta network. An example of a delta-delta network is described below with respect to Figure 18 in the Examples section. The differential network creation module may be a program or script written in PERL.

F. Visualization of Networks

The relationship values for the ensemble of networks and for the differential networks may be visualized using a network visualization program (e.g., Cytoscape open source platform for complex network analysis and visualization from the Cytoscape consortium). In the visual depictions of the networks, the thickness of each edge (e.g., each line connecting the proteins) represents the strength of fold change. The edges are also directional indicating causality, and each edge has an associated prediction confidence level.

G. Exemplary Computer System

Figure 17 schematically depicts an exemplary computer system/environment that may be employed in some embodiments for communicating with the AI-based

informatics system, for generating differential networks, for visualizing networks, for saving and storing data, and/or for interacting with a user. As explained above, calculations for an AI-based informatics system may be performed on a separate supercomputer with hundreds or thousands of parallel processors that interacts, directly or indirectly, with the exemplary computer system. The environment includes a computing device 100 with associated peripheral devices. Computing device 100 is programmable to implement executable code 150 for performing various methods, or portions of methods, taught herein. Computing device 100 includes a storage device 116, such as a hard-drive, CD-ROM, or other non-transitory computer readable media. Storage device 116 may store an operating system 118 and other related software. Computing device 100 may further include memory 106. Memory 106 may comprise a computer system memory or random access memory, such as DRAM, SRAM, EDO RAM, etc. Memory 106 may comprise other types of memory as well, or combinations thereof. Computing device 100 may store, in storage device 116 and/or memory 106, instructions for implementing and processing each portion of the executable code 150.

The executable code 150 may include code for communicating with the AI-based informatics system 190, for generating differential networks (e.g., a differential network creation module), for extracting quantitative relationship information from the AI-based informatics system (e.g., a relationship quantification module) and for visualizing networks (e.g., Cytoscape).

In some embodiments, the computing device 100 may communicate directly or indirectly with the AI-based informatics system 190 (e.g., a system for executing REFS). For example, the computing device 100 may communicate with the AI-based informatics system 190 by transferring data files (e.g., data frames) to the AI-based informatics system 190 through a network. Further, the computing device 100 may have executable code 150 that provides an interface and instructions to the AI-based informatics system 190.

In some embodiments, the computing device 100 may communicate directly or indirectly with one or more experimental systems 180 that provide data for the input data set. Experimental systems 180 for generating data may include systems for mass spectrometry based proteomics, microarray gene expression, qPCR gene expression, mass spectrometry based metabolomics, and mass spectrometry based lipidomics, SNP

microarrays, a panel of functional assays, and other in-vitro biology platforms and technologies.

Computing device 100 also includes processor 102, and may include one or more additional processor(s) 102', for executing software stored in the memory 106 and other programs for controlling system hardware, peripheral devices and/or peripheral hardware. Processor 102 and processor(s) 102' each can be a single core processor or multiple core (104 and 104') processor. Virtualization may be employed in computing device 100 so that infrastructure and resources in the computing device can be shared dynamically. Virtualized processors may also be used with executable code 150 and other software in storage device 116. A virtual machine 114 may be provided to handle a process running on multiple processors so that the process appears to be using only one computing resource rather than multiple. Multiple virtual machines can also be used with one processor.

A user may interact with computing device 100 through a visual display device 122, such as a computer monitor, which may display a user interface 124 or any other interface. The user interface 124 of the display device 122 may be used to display raw data, visual representations of networks, etc. The visual display device 122 may also display other aspects or elements of exemplary embodiments (e.g., an icon for storage device 116). Computing device 100 may include other I/O devices such a keyboard or a multi-point touch interface (e.g., a touchscreen) 108 and a pointing device 110, (e.g., a mouse, trackball and/or trackpad) for receiving input from a user. The keyboard 108 and the pointing device 110 may be connected to the visual display device 122 and/or to the computing device 100 via a wired and/or a wireless connection.

Computing device 100 may include a network interface 112 to interface with a network device 126 via a Local Area Network (LAN), Wide Area Network (WAN) or the Internet through a variety of connections including, but not limited to, standard telephone lines, LAN or WAN links (e.g., 802.11, T1, T3, 56kb, X.25), broadband connections (e.g., ISDN, Frame Relay, ATM), wireless connections, controller area network (CAN), or some combination of any or all of the above. The network interface 112 may comprise a built-in network adapter, network interface card, PCMCIA network card, card bus network adapter, wireless network adapter, USB network adapter, modem or any other device suitable for enabling computing device 100 to interface with any

type of network capable of communication and performing the operations described herein.

Moreover, computing device 100 may be any computer system such as a workstation, desktop computer, server, laptop, handheld computer or other form of computing or telecommunications device that is capable of communication and that has sufficient processor power and memory capacity to perform the operations described herein.

Computing device 100 can be running any operating system 118 such as any of the versions of the MICROSOFT WINDOWS operating systems, the different releases of the Unix and Linux operating systems, any version of the MACOS for Macintosh computers, any embedded operating system, any real-time operating system, any open source operating system, any proprietary operating system, any operating systems for mobile computing devices, or any other operating system capable of running on the computing device and performing the operations described herein. The operating system may be running in native mode or emulated mode.

IV. Models for Drug-induced Toxicity and Uses Therefor

A. Establishing a Model for Drug-induced Toxicity

Virtually all drug-induced toxicity involves complicated interactions among different cell types and/or organ systems. Perturbation of critical functions in one cell type or organ may lead to secondary effects on other interacting cells types and organs, and such downstream changes may in turn feedback to the initial changes and cause further complications. Therefore, it is beneficial to dissect a given drug-induced toxicity to its components, such as interaction between pairs of cell types or organs, and systemically probe the interactions between these components in order to gain a more complete, global view of the drug-induced toxicity process.

Accordingly, the present invention provides cell models for drug-induced toxicity. To this end, Applicants have built cell models for an exemplary drug-induced toxicity (e.g., cardiotoxicity) which have been employed in the subject discovery Platform Technology. Applicants have conducted experiments with the cell models using the subject discovery Platform Technology to generate consensus causal

relationship networks, including causal relationships unique in the drug-induced toxicity, and thereby identify "modulators" or critical molecular "drivers" important for the particular drug-induced toxicity.

One significant advantage of the Platform Technology and its components, e.g., the custom built cell models and data sets obtained from the drug-induced toxicity cell models, is that an initial, "first generation" consensus causal relationship network generated for a drug-induced toxicity can continually evolve or expand over time, e.g., by the introduction of additional cell lines/types and/or additional conditions. Additional data from the evolved cell model, i.e., data from the newly added portion(s) of the cell model, can be collected. The new data collected from an expanded or evolved cell model, i.e., from newly added portion(s) of the cell model, can then be introduced to the data sets previously used to generate the "first generation" consensus causal relationship network in order to generate a more robust "second generation" consensus causal relationship network. New causal relationships unique to the drug-induced toxicity can then be identified from the "second generation" consensus causal relationship network. In this way, the evolution of the drug-induced toxicity cell model provides an evolution of the consensus causal relationship networks, thereby providing new and/or more reliable insights into the modulators of the drug-induced toxicity. In this way, both the drug-induced toxicity cell models, the data sets from the cell models, and the causal relationship networks generated from the drug-induced toxicity cell models by using the Platform Technology methods can constantly evolve and build upon previous knowledge obtained from the Platform Technology.

Accordingly, the invention provides consensus causal relationship networks generated from the drug-induced toxicity cell models employed in the Platform Technology. These consensus causal relationship networks may be first generation consensus causal relationship networks, or may be multiple generation consensus causal relationship networks, e.g., $2^{nd,}$ 3rd, 4th, 5th, 6th, 7th, 8th, 9th, 10th, 11th, 12th, 13th, 14th, 15th, 16th, 17th, 18th, 19th, 20th or greater generation consensus causal relationship networks. Further, the invention provides simulated consensus causal relationship networks generated from the drug-induced toxicity cell models employed in the Platform Technology. These simulated consensus causal relationship networks may be first generation simulated consensus causal relationship networks, or may be multiple generation simulated consensus causal relationship networks, e.g., $2^{nd,}$ 3rd, 4th, 5th, 6th, 7th,

8th, 9th, 10th, 11th, 12th, 13th, 14th, 15th, 16th, 17th, 18th, 19th, 20th or greater simulated generation consensus causal relationship networks. The invention further provides delta networks and delta-delta networks generated from any of the consensus causal relationship networks of the invention.

A custom built cell model for a drug-induced toxicity comprises one or more cells associated with the drug-induced toxicity. The model for a drug-induced toxicity may be established to simulate an environment of the drug-induced toxicity, e.g., environment of drug-induced cardiotoxicity i*n vivo*, by creating conditions (e.g., cell culture conditions) that mimic a characteristic aspect of the drug-induced toxicity.

Multiple cells of the same or different origin, as opposed to a single cell type, may be included in the cell model. In one embodiment, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50 or more different cell lines or cell types are included in the drug-induced toxicity cell model. In one embodiment, the cells are all of the same type, e.g., all cardiomyocytes, but are different established cell lines, e.g., different established cell lines of cardiomyocytes. All values presented in the foregoing list can also be the upper or lower limit of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 2 and 5, or 5 and 15 different cell lines or cell types.

Examples of cell types that may be included in the cell models of the invention include, without limitation, human cells, animal cells, mammalian cells, plant cells, yeast, bacteria, or fungae. In one embodiment, cells of the cell model can include diseased cells, such as cancer cells or bacterially or virally infected cells. In one embodiment, cells of the cell model can include drug-induced toxicity associated cells, such as cells involved in diabetes, obesity or cardiovascular drug-induced toxicity state, e.g., aortic smooth muscle cells or hepatocytes. The skilled person would recognize those cells that are involved in or associated with a particular drug-induced toxicity, e.g., cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity, and any such cells may be included in a cell model of the invention, e.g., cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neuro cells, renal cells, or myoblasts.

Cell models of the invention may include one or more "control cells." In one embodiment, a control cell may be an untreated or unperturbed cell. In another embodiment, a "control cell" may be a normalcell, e.g., a cell that has not been exposed

to a toxicity-causing agent or drug. In one embodiment, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 or more different control cells are included in the cell model. All values presented in the foregoing list can also be the upper or lower limit of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 2 and 5, or 5 and 15 different control cell lines or control cell types. In one embodiment, the control cells are all of the same type but are different established cell lines of that cell type. In one embodiment, as a control, one or more normal, e.g., non-diseased, cell lines are cultured under similar conditions, and/or are exposed to the same perturbation, as the primary cells of the cell model in order to identify proteins or pathways unique to the drug-induced toxicity.

A custom cell model of the invention may also comprise conditions that mimic a characteristic aspect of the drug-induced toxicity. For example, cell culture conditions may be selected that closely approximating the conditions of a cell in a diabetic environment in vivo for probing diabetic drug induced toxicity, or of an aortic smooth muscle cell of a patient suffering from drug-induced cardiotoxicity. In some instances, the conditions are stress conditions.Various conditions / stressors may be employed in the cell models of the invention. In one embodiment, these stressors / conditions may constitute the "perturbation", e.g., external stimulus, for the cell systems. One exemplary stress condition is hypoxia, a condition typically found, for example, within patients with advanced stage of diabetes. Hypoxia can be induced using art-recognized methods. For example, hypoxia can be induced by placing cell systems in a Modular Incubator Chamber (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), which can be flooded with an industrial gas mix containing 5% CO_2, 2% O_2 and 93% nitrogen. Effects can be measured after a pre-determined period, *e.g.*, at 24 hours after hypoxia treatment, with and without additional external stimulus components (*e.g.*, CoQ10 at 0, 50, or 100 μM). Likewise, lactic acid treatment mimics a cellular environment where glycolysis activity is high. Lactic acid induced stress can be investigated at a final lactic acid concentration of about 12.5 mM at a pre-determined time, *e.g.*, at 24 hours, with or without additional external stimulus components (*e.g.*, CoQ10 at 0, 50, or 100 μM). Hyperglycemia is a condition found in diabetes as well as in diabetic drug-induced toxicity. A typical hyperglycemic condition that can be used to treat the subject cells include 10% culture grade glucose added to suitable media to bring up the final concentration of glucose in the media to about 22 mM. Hyperlipidemia is a condition

found, for example, in obesity and cardiovascular disease, and can be used to simulate drug-induced cardiotoxicity. The hyperlipidemic conditions can be provided by culturing cells in media containing 0.15 mM sodium palmitate. Hyperinsulinemia is a condition found, for example, in diabetes, as well as in diabetic drug-induced toxicity. The hyperinsulinemic conditions may be induced by culturing the cells in media containing 1000 nM insulin.

Individual conditions may be investigated separately in the custom built cell models of the invention, and/or may be combined together. In one embodiment, a combination of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 or more conditions reflecting or simulating different characteristic aspects of the biological system are investigated in the custom built cell model. In one embodiment, individual conditions and, in addition, combinations of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50 or more of the conditions reflecting or simulating different characteristic aspects of the drug-induced toxicity are investigated in the custom built drug-induced toxicity cell model. All values presented in the foregoing list can also be the upper or lower limit of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 1 and 20, 1 and 30, 2 and 5, 2 and 10, 5 and 10, 1 and 20, 5 and 20, 10 and 20, 10 and 25, 10 and 30 or 10 and 50 different conditions.

Once the custom drug-induced toxicity cell model is built, one or more "perturbations" may be applied to the system, such as genetic variation from patient to patient, or with / without treatment by certain drugs or pro-drugs. *See* Figure 15D. The effects of such perturbations to the cell model system can be measured using various art-recognized or proprietary means, as described in section III.B below.

The custom built drug-induced toxicity cell model may be exposed to a perturbation, e.g., an "environmental perturbation" or "external stimulus component". The "environmental perturbation" or "external stimulus component" may be endogenous to the cellular environment (*e.g.*, the cellular environment contains some levels of the stimulant, and more of the same is added to increase its level), or may be exogenous to the cellular environment (*e.g.*, the stimulant/perturbation is largely absent from the cellular environment prior to the alteration). The cellular environment may further be altered by secondary changes resulting from adding the environmental perturbation or external stimulus component, since the external stimulus component may change the cellular output of the cell system, including molecules secreted into the cellular

environment by the cell system. The environmental perturbation or external stimulus component may include any external physical and/or chemical stimulus that may affect cellular function. This may include any large or small organic or inorganic molecules, natural or synthetic chemicals, temperature shift, pH change, radiation, light (UVA, UVB *etc.*), microwave, sonic wave, electrical current, modulated or unmodulated magnetic fields, *etc.* The environmental perturbation or external stimulus component may also include an introduced genetic modification or mutation or a vehicle (e.g., vector) that causes a genetic modification / mutation.

(i) Cross-talk cell systems

In certain situations, where interaction between two or more cell systems are desired to be investigated, a "cross-talking cell system" may be formed by, for example, bringing the modified cellular environment of a first cell system into contact with a second cell system to affect the cellular output of the second cell system.

As used herein, "cross-talk cell system" comprises two or more cell systems, in which the cellular environment of at least one cell system comes into contact with a second cell system, such that at least one cellular output in the second cell system is changed or affected. In certain embodiments, the cell systems within the cross-talk cell system may be in direct contact with one another. In other embodiments, none of the cell systems are in direct contact with one another.

For example, in certain embodiments, the cross-talk cell system may be in the form of a transwell, in which a first cell system is growing in an insert and a second cell system is growing in a corresponding well compartment. The two cell systems may be in contact with the same or different media, and may exchange some or all of the media components. External stimulus component added to one cell system may be substantially absorbed by one cell system and/or degraded before it has a chance to diffuse to the other cell system. Alternatively, the external stimulus component may eventually approach or reach an equilibrium within the two cell systems.

In certain embodiments, the cross-talk cell system may adopt the form of separately cultured cell systems, where each cell system may have its own medium and/or culture conditions (temperature, CO_2 content, pH, *etc.*), or similar or identical culture conditions. The two cell systems may come into contact by, for example, taking the conditioned medium from one cell system and bringing it into contact with another

cell system. Direct cell-cell contacts between the two cell systems can also be effected if desired. For example, the cells of the two cell systems may be co-cultured at any point if desired, and the co-cultured cell systems can later be separated by, for example, FACS sorting when cells in at least one cell system have a sortable marker or label (such as a stably expressed fluorescent marker protein GFP).

Similarly, in certain embodiments, the cross-talk cell system may simply be a co-culture. Selective treatment of cells in one cell system can be effected by first treating the cells in that cell system, before culturing the treated cells in co-culture with cells in another cell system. The co-culture cross-talk cell system setting may be helpful when it is desired to study, for example, effects on a second cell system caused by cell surface changes in a first cell system, after stimulation of the first cell system by an external stimulus component.

The cross-talk cell system of the invention is particularly suitable for exploring the effect of certain pre-determined external stimulus component on the cellular output of one or both cell systems. The primary effect of such a stimulus on the first cell system (with which the stimulus directly contact) may be determined by comparing cellular outputs (*e.g.*, protein expression level) before and after the first cell system's contact with the external stimulus, which, as used herein, may be referred to as "(significant) cellular output differentials." The secondary effect of such a stimulus on the second cell system, which is mediated through the modified cellular environment of the first cell system (such as its secretome), can also be similarly measured. There, a comparison in, for example, proteome of the second cell system can be made between the proteome of the second cell system *with* the external stimulus treatment on the first cell system, and the proteome of the second cell system *without* the external stimulus treatment on the first cell system. Any significant changes observed (in proteome or any other cellular outputs of interest) may be referred to as a "significant cellular cross-talk differential."

In making cellular output measurements (such as protein expression), either absolute expression amount or relative expression level may be used. For example, to determine the relative protein expression level of a second cell system, the amount of any given protein in the second cell system, with or without the external stimulus to the first cell system, may be compared to a suitable control cell line and mixture of cell lines and given a fold-increase or fold-decrease value. A pre-determined threshold level for

such fold-increase (*e.g.*, at least 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 or 100 or more fold increase) or fold-decrease (*e.g.*, at least a decrease to 0.95, 0.9, 0.8, 0.75, 0.7, 0.6, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1 or 0.05 fold, or 90%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10% or 5% or less) may be used to select significant cellular cross-talk differentials. All values presented in the foregoing list can also be the upper or lower limit of ranges, *e.g.*, between 1.5 and 5 fold, between 2 and 10 fold, between 1 and 2 fold, or between 0.9 and 0.7 fold, that are intended to be a part of this invention.

Throughout the present application, all values presented in a list, e.g., such as those above, can also be the upper or lower limit of ranges that are intended to be a part of this invention.

To illustrate, in one exemplary two-cell system established to imitate aspects of a drug-induced cardiotoxicity and nephrotoxicity model, a heart smooth muscle cell line (first cell system) may be treated with a hypoxia condition (an external stimulus component), and proteome changes in a kidney cell line (second cell system) resulting from contacting the kidney cells with conditioned medium of the heart smooth muscle may be measured using conventional quantitative mass spectrometry. Significant cellular cross-talking differentials in these kidney cells may be determined, based on comparison with a proper control (*e.g.*, similarly cultured kidney cells contacted with conditioned medium from similarly cultured heart smooth muscle cells not treated with hypoxia conditions).

Not every observed significant cellular cross-talking differentials may be of biological significance. With respect to any given drug-induced toxicity for which the subject interrogative biological assessment is applied, some (or maybe all) of the significant cellular cross-talking differentials may be "determinative" with respect to the specific biological problem at issue, *e.g.*, either responsible for causing a drug-induced toxicity (a potential target for therapeutic intervention) or is a biomarker for the drug-induced toxicity (a potential diagnostic or prognostic factor).

Such determinative cross-talking differentials may be selected by an end user of the subject method, or it may be selected by a bioinformatics software program, such as DAVID-enabled comparative pathway analysis program, or the KEGG pathway analysis program. In certain embodiments, more than one bioinformatics software program is

used, and consensus results from two or more bioinformatics software programs are preferred.

As used herein, "differentials" of cellular outputs include differences (*e.g.*, increased or decreased levels) in any one or more parameters of the cellular outputs. For example, in terms of protein expression level, differentials between two cellular outputs, such as the outputs associated with a cell system before and after the treatment by an external stimulus component, can be measured and quantitated by using art-recognized technologies, such as mass-spectrometry based assays (*e.g.*, iTRAQ, 2D-LC-MSMS, *etc.*).

B. Use of Cell Models for Interrogative Biological Assessments

The methods and cell models described herein, and further described in international Application No. PCT/US2012/027615, may be used for, or applied to, any number of "interrogative biological assessments." Use of the methods of the invention for an interrogative biological assessment facilitates the identification of "modulators" or determinative cellular process "drivers" of a drug-induced toxicity.

As used herein, an "interrogative biological assessment" may include the identification of one or more modulators of a biological system, e.g., determinative cellular process "drivers," (e.g., an increase or decrease in activity of a biological pathway, or key members of the pathway, or key regulators to members of the pathway) associated with the environmental perturbation or external stimulus component, or a unique causal relationship unique in a biological system or process. It may further include additional steps designed to test or verify whether the identified determinative cellular process drivers are necessary and/or sufficient for the downstream events associated with the environmental perturbation or external stimulus component, including in vivo animal models and/or in vitro tissue culture experiments.

In a preferred embodiment, the interrogative biological assessment is the assessment of the drug-induced toxicological profile of an agent, e.g., a drug, on a cell, tissue, organ or organism, wherein the identified modulators of a biological system, e.g., determinative cellular process driver (e.g., cellular cross-talk differentials or causal relationships unique in a biological system or process) may be indicators of drug-induced toxicities, e.g., cytotoxicity, cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity, and may in turn be used to predict or

identify the toxicological profile of the drug. In one embodiment, the identified modulators of a drug-induced toxicity, e.g., determinative cellular process driver (e.g., cellular cross-talk differentials or causal relationships unique in a drug-induced toxicity) is an indicator of cardiotoxicity of a drug or drug candidate, and may in turn be used to predict or identify the cardiotoxicological profile of the drug or drug candidate.

V. Proteomic Sample Analysis

In certain embodiments, the subject method employs large-scale high-throughput quantitative proteomic analysis of hundreds of samples of similar character, and provides the data necessary for identifying the cellular output differentials.

There are numerous art-recognized technologies suitable for this purpose. An exemplary technique, iTRAQ analysis in combination with mass spectrometry, is briefly described below.

To provide reference samples for relative quantification with the iTRAQ technique, multiple QC pools are created. Two separate QC pools, consisting of aliquots of each sample, were generated from the Cell #1 and Cell #2 samples - these samples are denoted as QCS1 and QCS2, and QCP1 and QCP2 for supernatants and pellets, respectively. In order to allow for protein concentration comparison across the two cell lines, cell pell*et al*iquots from the QC pools described above are combined in equal volumes to generate reference samples (QCP).

The quantitative proteomics approach is based on stable isotope labeling with the 8-plex iTRAQ reagent and 2D-LC MALDI MS/MS for peptide identification and quantification. Quantification with this technique is relative: peptides and proteins are assigned abundance ratios relative to a reference sample. Common reference samples in multiple iTRAQ experiments facilitate the comparison of samples across multiple iTRAQ experiments.

To implement this analysis scheme, six primary samples and two control pool samples are combined into one 8-plex iTRAQ mix, with the control pool samples labeled with 113 and 117 reagents according to the manufacturer's suggestions. This mixture of eight samples is then fractionated by two-dimensional liquid chromatography; strong cation exchange (SCX) in the first dimension, and reversed-phase HPLC in the second dimension. The HPLC eluent is directly fractionated onto

MALDI plates, and the plates are analyzed on an MDS SCIEX/AB 4800 MALDI TOF/TOF mass spectrometer.

In the absence of additional information, it is assumed that the most important changes in protein expression are those within the same cell types under different treatment conditions. For this reason, primary samples from Cell#1 and Cell#2 are analyzed in separate iTRAQ mixes. To facilitate comparison of protein expression in Cell#1 vs. Cell#2 samples, universal QCP samples are analyzed in the available "iTRAQ slots" not occupied by primary or cell line specific QC samples (QC1 and QC2).

A brief overview of the laboratory procedures employed is provided herein.

A. Protein Extraction From Cell Supernatant Samples

For cell supernatant samples (CSN), proteins from the culture medium are present in a large excess over proteins secreted by the cultured cells. In an attempt to reduce this background, upfront abundant protein depletion was implemented. As specific affinity columns are not available for bovine or horse serum proteins, an anti-human IgY14 column was used. While the antibodies are directed against human proteins, the broad specificity provided by the polyclonal nature of the antibodies was anticipated to accomplish depletion of both bovine and equine proteins present in the cell culture media that was used.

A 200-µl aliquot of the CSN QC material is loaded on a 10-mL IgY14 depletion column before the start of the study to determine the total protein concentration (Bicinchoninic acid (BCA) assay) in the flow-through material. The loading volume is then selected to achieve a depleted fraction containing approximately 40 µg total protein.

B. Protein Extraction From Cell Pellets

An aliquot of Cell #1 and Cell #2 is lysed in the "standard" lysis buffer used for the analysis of tissue samples at BGM, and total protein content is determined by the BCA assay. Having established the protein content of these representative cell lystates, all cell pellet samples (including QC samples described in Section 1.1) were processed to cell lysates. Lysate amounts of approximately 40 µg of total protein were carried forward in the processing workflow.

C. Sample Preparation for Mass Spectrometry

Sample preparation follows standard operating procedures and constitute of the following:

- Reduction and alkylation of proteins
- Protein clean-up on reversed-phase column (cell pellets only)
- Digestion with trypsin
- iTRAQ labeling
- Strong cation exchange chromatography – collection of six fractions (Agilent 1200 system)
- HPLC fractionation and spotting to MALDI plates (Dionex Ultimate3000/Probot system)

D. MALDI MS and MS/MS

HPLC-MS generally employs online ESI MS/MS strategies. BG Medicine uses an off-line LC-MALDI MS/MS platform that results in better concordance of observed protein sets across the primary samples without the need of injecting the same sample multiple times. Following first pass data collection across all iTRAQ mixes, since the peptide fractions are retained on the MALDI target plates, the samples can be analyzed a second time using a targeted MS/MS acquisition pattern derived from knowledge gained during the first acquisition. In this manner, maximum observation frequency for all of the identified proteins is accomplished (ideally, every protein should be measured in every iTRAQ mix).

E. Data Processing

The data processing process within the BGM Proteomics workflow can be separated into those procedures such as preliminary peptide identification and quantification that are completed for each iTRAQ mix individually (Section 1.5.1) and those processes (Section 1.5.2) such as final assignment of peptides to proteins and final quantification of proteins, which are not completed until data acquisition is completed for the project.

The main data processing steps within the BGM Proteomics workflow are:

- Peptide identification using the Mascot (Matrix Sciences) database search engine
- Automated in house validation of Mascot IDs
- Quantification of peptides and preliminary quantification of proteins

- Expert curation of final dataset
- Final assignment of peptides from each mix into a common set of proteins using the automated PVT tool
- Outlier elimination and final quantification of proteins

(i) Data Processing of Individual iTRAQ Mixes

As each iTRAQ mix is processed through the workflow the MS/MS spectra are analyzed using proprietary BGM software tools for peptide and protein identifications, as well as initial assessment of quantification information. Based on the results of this preliminary analysis, the quality of the workflow for each primary sample in the mix is judged against a set of BGM performance metrics. If a given sample (or mix) does not pass the specified minimal performance metrics, and additional material is available, that sample is repeated in its entirety and it is data from this second implementation of the workflow that is incorporated in the final dataset.

(ii) Peptide Identification

MS/MS spectra was searched against the Uniprot protein sequence database containing human, bovine, and horse sequences augmented by common contaminant sequences such as porcine trypsin. The details of the Mascot search parameters, including the complete list of modifications, are given in Table 1.

Table 1: Mascot Search Parameters

Precursor mass tolerance	100 ppm
Fragment mass tolerance	0.4 Da
Variable modifications	N-term iTRAQ8 Lysine iTRAQ8 Cys carbamidomethyl Pyro-Glu (N-term) Pyro-Carbamidomethyl Cys (N-term) Deamidation (N only) Oxidation (M)
Enzyme specificity	Fully Tryptic
Number of missed tryptic sites allowed	2
Peptide rank considered	1

After the Mascot search is complete, an auto-validation procedure is used to promote (*i.e.*, validate) specific Mascot peptide matches. Differentiation between valid and invalid matches is based on the attained Mascot score relative to the expected Mascot score and the difference between the Rank 1 peptides and Rank 2 peptide Mascot scores. The criteria required for validation are somewhat relaxed if the peptide is one of several matched to a single protein in the iTRAQ mix or if the peptide is present in a catalogue of previously validated peptides.

(iii) Peptide and Protein Quantification

The set of validated peptides for each mix is utilized to calculate preliminary protein quantification metrics for each mix. Peptide ratios are calculated by dividing the peak area from the iTRAQ label (*i.e.*, *m/z* 114, 115, 116, 118, 119, or 121) for each validated peptide by the best representation of the peak area of the reference pool (QC1 or QC2). This peak area is the average of the 113 and 117 peaks provided both samples pass QC acceptance criteria. Preliminary protein ratios are determined by calculating the median ratio of all "useful" validated peptides matching to that protein. "Useful" peptides are fully iTRAQ labeled (all N-terminal are labeled with either Lysine or PyroGlu) and fully Cysteine labeled (*i.e.*, all Cys residues are alkylated with Carbamidomethyl or N-terminal Pyro-cmc).

(iv) Post-acquisition Processing

Once all passes of MS/MS data acquisition are complete for every mix in the project, the data is collated using the three steps discussed below which are aimed at enabling the results from each primary sample to be simply and meaningfully compared to that of another.

(v) Global Assignment of Peptide Sequences to Proteins

Final assignment of peptide sequences to protein accession numbers is carried out through the proprietary Protein Validation Tool (PVT). The PVT procedure determines the best, minimum non-redundant protein set to describe the entire collection of peptides identified in the project. This is an automated procedure that has been optimized to handle data from a homogeneous taxonomy.

Protein assignments for the supernatant experiments were manually curated in order to deal with the complexities of mixed taxonomies in the database. Since the automated paradigm is not valid for cell cultures grown in bovine and horse serum

supplemented media, extensive manual curation is necessary to minimize the ambiguity of the source of any given protein.

(vi) Normalization of Peptide Ratios

The peptide ratios for each sample are normalized based on the method of Vandesompele *et al. Genome Biology*, **2002**, *3(7)*, research 0034.1-11. This procedure is applied to the cell pellet measurements only. For the supernatant samples, quantitative data are not normalized considering the largest contribution to peptide identifications coming from the media.

(vii) Final Calculation of Protein Ratios

A standard statistical outlier elimination procedure is used to remove outliers from around each protein median ratio, beyond the 1.96 σ level in the log-transformed data set. Following this elimination process, the final set of protein ratios are (re-)calculated.

VI. Markers of the Invention and Uses Thereof

The present invention is based, at least in part, on the identification of novel biomarkers that are associated with drug-induced toxicities, such as a drug-induced cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity, or response of a drug-induced toxicity to a perturbation, such as a therapeutic agent.

In particular, the invention relates to markers (hereinafter "markers" or "markers of the invention"), which are described in the examples. The invention provides nucleic acids and proteins that are encoded by or correspond to the markers (hereinafter "marker nucleic acids" and "marker proteins," respectively). These markers are particularly useful in diagnosing drug-induced toxicity states; prognosing drug-induced toxicity states; developing drug targets for varies drug-induced toxicity states; screening for the presence of toxicity, preferably drug-induced toxicities, e.g., cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity; identifying an agent that cause or is at risk for causing drug-induced toxicity; identifying an agent that can reduce or prevent drug-induced toxicity; alleviating, reducing or preventing drug-inducedtoxicity; and identifying markers predictive of drug-induced toxicity.

A "marker" is a gene whose altered level of expression in a tissue or cell from its expression level in normal or healthy tissue or cell is associated with a toxicity state, such as a drug-induced toxicity, e.g., cardiotoxicity. A “marker nucleic acid” is a nucleic acid (*e.g.*, mRNA, cDNA) encoded by or corresponding to a marker of the invention. Such marker nucleic acids include DNA (*e.g.*, cDNA) comprising the entire or a partial sequence of any of the genes that are markers of the invention or the complement of such a sequence. Such sequences are known to the one of skill in the art and can be found for example, on the NIH government pubmed website. The marker nucleic acids also include RNA comprising the entire or a partial sequence of any of the gene markers of the invention or the complement of such a sequence, wherein all thymidine residues are replaced with uridine residues. A “marker protein” is a protein encoded by or corresponding to a marker of the invention. A marker protein comprises the entire or a partial sequence of any of the marker proteins of the invention. Such sequences are known to the one of skill in the art and can be found for example, on the NIH government pubmed website. The terms “protein” and “polypeptide’ are used interchangeably.

A “toxic state associated" body fluid is a fluid which, when in the body of a patient, contacts or passes through sarcoma cells or into which cells or proteins shed from sarcoma cells are capable of passing. Exemplary disease state or toxic state associated body fluids include blood fluids (*e.g.* whole blood, blood serum, blood having platelets removed therefrom), and are described in more detail below. Disease state or toxic state associated body fluids are not limited to, whole blood, blood having platelets removed therefrom, lymph, prostatic fluid, urine and semen.

The "normal" level of expression of a marker is the level of expression of the marker in cells of a human subject or patient not afflicted with a toxicity state.

An “over-expression” or “higher level of expression” of a marker refers to an expression level in a test sample that is greater than the standard error of the assay employed to assess expression, and is preferably at least twice, and more preferably three, four, five, six, seven, eight, nine or ten times the expression level of the marker in a control sample (*e.g.*, sample from a healthy subject not having the marker associated a drug-induce toxicity state, *e.g.,* cardiotoxicit, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity) and preferably, the average expression level of the marker in several control samples.

A “lower level of expression” of a marker refers to an expression level in a test sample that is at least twice, and more preferably three, four, five, six, seven, eight, nine or ten times lower than the expression level of the marker in a control sample (*e.g.*, sample from a healthy subjects not having the marker associated a drug-induced toxicity state, *e.g.,* cardiotoxicity, cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity) and preferably, the average expression level of the marker in several control samples.

A "transcribed polynucleotide" or “nucleotide transcript” is a polynucleotide (*e.g.* an mRNA, hnRNA, a cDNA, or an analog of such RNA or cDNA) which is complementary to or homologous with all or a portion of a mature mRNA made by transcription of a marker of the invention and normal post-transcriptional processing (*e.g.* splicing), if any, of the RNA transcript, and reverse transcription of the RNA transcript.

"Complementary" refers to the broad concept of sequence complementarity between regions of two nucleic acid strands or between two regions of the same nucleic acid strand. It is known that an adenine residue of a first nucleic acid region is capable of forming specific hydrogen bonds ("base pairing") with a residue of a second nucleic acid region which is antiparallel to the first region if the residue is thymine or uracil. Similarly, it is known that a cytosine residue of a first nucleic acid strand is capable of base pairing with a residue of a second nucleic acid strand which is antiparallel to the first strand if the residue is guanine. A first region of a nucleic acid is complementary to a second region of the same or a different nucleic acid if, when the two regions are arranged in an antiparallel fashion, at least one nucleotide residue of the first region is capable of base pairing with a residue of the second region. Preferably, the first region comprises a first portion and the second region comprises a second portion, whereby, when the first and second portions are arranged in an antiparallel fashion, at least about 50%, and preferably at least about 75%, at least about 90%, or at least about 95% of the nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion. More preferably, all nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion.

"Homologous" as used herein, refers to nucleotide sequence similarity between two regions of the same nucleic acid strand or between regions of two different nucleic acid strands. When a nucleotide residue position in both regions is occupied by the

same nucleotide residue, then the regions are homologous at that position. A first region is homologous to a second region if at least one nucleotide residue position of each region is occupied by the same residue. Homology between two regions is expressed in terms of the proportion of nucleotide residue positions of the two regions that are occupied by the same nucleotide residue. By way of example, a region having the nucleotide sequence 5'-ATTGCC-3' and a region having the nucleotide sequence 5'-TATGGC-3' share 50% homology. Preferably, the first region comprises a first portion and the second region comprises a second portion, whereby, at least about 50%, and preferably at least about 75%, at least about 90%, or at least about 95% of the nucleotide residue positions of each of the portions are occupied by the same nucleotide residue. More preferably, all nucleotide residue positions of each of the portions are occupied by the same nucleotide residue.

"Proteins of the invention" encompass marker proteins and their fragments; variant marker proteins and their fragments; peptides and polypeptides comprising an at least 15 amino acid segment of a marker or variant marker protein; and fusion proteins comprising a marker or variant marker protein, or an at least 15 amino acid segment of a marker or variant marker protein.

The invention further provides antibodies, antibody derivatives and antibody fragments which specifically bind with the marker proteins and fragments of the marker proteins of the present invention. Unless otherwise specified herewithin, the terms "antibody" and "antibodies" broadly encompass naturally-occurring forms of antibodies (*e.g.*, IgG, IgA, IgM, IgE) and recombinant antibodies such as single-chain antibodies, chimeric and humanized antibodies and multi-specific antibodies, as well as fragments and derivatives of all of the foregoing, which fragments and derivatives have at least an antigenic binding site. Antibody derivatives may comprise a protein or chemical moiety conjugated to an antibody.

In one embodiment, the markers of the invention are genes or proteins associated with or involved in drug-induced toxicity. Such genes or proteins involved in drug-induced toxicity include, for example, the markers listed in table 2. In some embodiments, the markers of the invention are a combination of at least two, three, four, five, six, seven, eight, nine, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the foregoing genes (or proteins). All values presented in the foregoing list can also be the upper or lower limit

of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 1 and 20, 1 and 30, 2 and 5, 2 and 10, 5 and 10, 1 and 20, 5 and 20, 10 and 20, 10 and 25, 10 and 30 of the foregoing genes (or proteins).

A. Cardiotoxicity Associated Markers

The present invention is based, at least in part, on the identification of novel biomarkers that are associated with drug-induced cardiotoxicity. The invention is further based, at least in part, on the discovery that Coenzyme Q10 is capable of reducing or preventing drug-induced cardiotoxicity.

Accordingly, the invention provides methods for identifying an agent that causes or is at risk for causing drug-induced cardiotoxicity. In one embodiment, the agent is a drug or drug candidate. In these methods, the amount of one or more biomarkers/proteins in a pair of samples (a first sample not subject to the drug treatment, and a second sample subjected to the drug treatment) is assessed. A modulation in the level of expression of the one or more biomarkers in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity. In one embodiment, the one or more biomarkers is selected from the markers listed in table 2. The methods of the present invention can be practiced in conjunction with any other method used by the skilled practitioner to identify a drug at risk for causing drug-induced cardiotoxocity.

Accordingly, in one aspect, the invention provides a method for identifying a drug that causes or is at risk for causing drug-induced cardiotoxicity, comprising: comparing (i) the level of expression of one or more biomarkers present in a first cell sample obtained prior to the treatment with the drug; with (ii) the level of expression of the one or more biomarkers present in a second cell sample obtained following the treatment with the drug; wherein the one or more biomarkers is selected from the markers listed in table 2; wherein a modulation in the level of expression of the one or more biomarkers in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.

In one embodiment, the cells are cells of the cardiovascular system, e.g., cardiomyocytes. In one embodiment, the cells are diabetic cardiomyocytes. In one embodiment, the drug is a drug or candidate drug for treating diabetes, obesity or cardiovascular disease.

In one embodiment, a modulation (e.g., an increase or a decrease) in the level of expression of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, twenty-five, thirty, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160 or more of the biomarkers selected from the markers listed in table 2 in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.

In one embodiment, a modulation (e.g., an increase or a decrease) in the level of expression of a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen, markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4 in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.

Methods for identifying a rescue agent that can reduce or prevent drug-induced cardiotoxicity are also provided by the invention. In one embodiment, the drug is a drug or drug candidate for treating diabetes, obesity or a cardiovascular disorder. In these methods, the amount of one or more biomarkers in three samples (a first sample not subjected to the drug treatment, a second sample subjected to the drug treatment, and a third sample subjected both to the drug treatment and the agent) is assessed. Approximately a normalized level of expression of the one or more biomarkers, in the third sample as compared to the first sample, with a changed level of expression in the second sample, is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity. In one embodiment, the one or more biomarkers is selected from the markers listed in table 2.

Using the methods described herein, a variety of molecules, particularly including molecules sufficiently small to be able to cross the cell membrane, may be screened in order to identify molecules which modulate, *e.g.,* increase or decrease the expression and/or activity of a marker of the invention. Compounds so identified can be provided to a subject in order to reduce, alleviate or prevent drug-induced cardiotoxicity in the subject.

Accordingly, in another aspect, the invention provides a method for identifying an agent that can reduce or prevent drug-induced cardiotoxicity comprising: (i) determining a normal level of expression of one or more biomarkers present in a first

cell sample obtained prior to the treatment with a toxicity inducing drug; (ii) determining a treated level of expression of the one or more biomarkers present in a second cell sample obtained following the treatment with the toxicity inducing drug to identify one or more biomarkers with a change of expression in the treated cell sample; (iii) determining the level of expression of the one or more biomarkers with a changed level of expression in the toxicity inducing drug treated sample present in a third cell sample obtained following the treatment with the toxicity inducing drug and the rescue agent; and (iv) comparing the level of expression of the one or more biomarkers determined in the third sample with the level of expression of the one or more biomarkers determined in the first sample; and a normalized level of expression of the one or more biomarkers in the third sample as compared to the first sample is an indication that the agent can reduce or prevent drug-induced cardiotoxicity. In one embodiment, the one or more biomarkers is selected from the markers listed in table 2.

In one embodiment, the cells are cells of the cardiovascular system, e.g., cardiomyocytes. In one embodiment, the cells are diabetic cardiomyocytes. In one embodiment, the drug is a drug or candidate drug for treating diabetes, obesity or cardiovascular disease. In one embodiment, the drug is Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab, Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, or TNF antagonists.In one embodiment, a normalized level of expression of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, twenty-five, thirty, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the biomarkers selected from the markers listed in table 2 in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity.

In one embodiment, a normalized level of expression of a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4, in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity.

In one embodiment, the sample comprises a fluid obtained from the subject. In one embodiment, the fluid is selected from the group consisting of blood fluids, vomit,

saliva, lymph, cystic fluid, urine, fluids collected by bronchial lavage, fluids collected by peritoneal rinsing, and gynecological fluids. In one embodiment, the sample is a blood sample or a component thereof.

In another embodiment, the sample comprises a tissue or component thereof obtained from the subject. In one embodiment, the tissue is selected from the group consisting of bone, connective tissue, cartilage, lung, liver, kidney, muscle tissue, heart, pancreas, and skin.

In one embodiment, the subject is a human.

In one embodiment, the level of expression of the one or more markers in the biological sample is determined by assaying a transcribed polynucleotide or a portion thereof in the sample. In one embodiment, wherein assaying the transcribed polynucleotide comprises amplifying the transcribed polynucleotide.

In one embodiment, the level of expression of the marker in the subject sample is determined by assaying a protein or a portion thereof in the sample. In one embodiment, the protein is assayed using a reagent which specifically binds with the protein.

In one embodiment, the level of expression of the one or more markers in the sample is determined using a technique selected from the group consisting of polymerase chain reaction (PCR) amplification reaction, reverse-transcriptase PCR analysis, single-strand conformation polymorphism analysis (SSCP), mismatch cleavage detection, heteroduplex analysis, Southern blot analysis, Northern blot analysis, Western blot analysis, in situ hybridization, array analysis, deoxyribonucleic acid sequencing, restriction fragment length polymorphism analysis, and combinations or sub-combinations thereof, of said sample.

In one embodiment, the level of expression of the marker in the sample is determined using a technique selected from the group consisting of immunohistochemistry, immunocytochemistry, flow cytometry, ELISA and mass spectrometry.

In one embodiment, the level of expression of a plurality of markers is determined.

The invention further provides methods for alleviating, reducing or preventing drug-induced cardiotoxicity in a subject in need thereof, comprising administering to a subject (e.g., a mammal, a human, or a non-human animal) an agent identified by the screening methods provided herein, thereby reducing or preventing drug-induced

cardiotoxicity in the subject. In one embodiment, the agent is administered to a subject that has already been treated with a cardiotoxicity-inducing drug. In one embodiment, the agent is administered to a subject at the same time as treatment of the subject with a cardiotoxicity-inducing drug. In one embodiment, the agent is administered to a subject prior to treatment of the subject with a cardiotoxicity-inducing drug.

The invention further provides methods for alleviating, reducing or preventing drug-induced cardiotoxicity in a subject in need thereof, comprising administering Coenzyme Q10 to the subject (e.g., a mammal, a human, or a non-human animal), thereby reducing or preventing drug-induced cardiotoxicity in the subject. In one embodiment, the Coenzyme Q10 is administered to a subject that has already been treated with a cardiotoxicity-inducing drug. In one embodiment, the Coenzyme Q10 is administered to a subject at the same time as treatment of the subject with a cardiotoxicity-inducing drug. In one embodiment, the Coenzyme Q10 is administered to a subject prior to treatment of the subject with a cardiotoxicity-inducing drug. In one embodiment, the drug-induced cardiotoxicity is associated with modulation of expression of one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, twenty-five, thirty, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the biomarkers selected from the markers listed in table 2. All values presented in the foregoing list can also be the upper or lower limit of ranges, that are intended to be a part of this invention, *e.g.*, between 1 and 5, 1 and 10, 2 and 5, 2 and 10, or 5 and 10 of the foregoing genes (or proteins).

In one embodiment, the drug-induced cardiotoxicity is associated with modulation of a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4.

The invention further provides biomarkers (e.g, genes and/or proteins) that are useful as predictive markers for drug-induced cardiotoxicity. These biomarkers include the markers listed in table 2. In one embodiment, the predictive markers for drug-induced cardiotoxicity is a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen, markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4. The ordinary skilled artisan would, however, be able to identify additional

biomarkers predictive of drug-induced cardiotoxicity by employing the methods described herein, e.g., by carrying out the methods described in Example 3 but by using a different drug known to induce cardiotoxicity. Exemplary drug-induced cardiotoxicity biomarkers of the invention are further described below.

GRP78 and GRP75 are also referred to as glucose response proteins. These proteins are associated with endo/sarcoplasmic reticulum stress (ER stress) of cardiomyocytes. SERCA, or sarcoendoplasmic reticulum calcium ATPase, regulates Ca2+ homeostatsis in cardiac cells. Any disruption of these ATPase can lead to cardiac dysfunction and heart failure. Based upon the data provided herein, GRP75 and GRP78 and the edges around them are novel predictors of drug induced cardiotoxicity.

TIMP1, also referred to as TIMP metalloprotease inhibitor 1, is involved with remodeling of extra cellular matrix in association with MMPs. TIMP1 expression is correlated with fibrosis of the heart, and hypoxia of vascular endothelial cells also induces TIMP1 expression. Based upon the data provided herein, TIMP1 is a novel predictor of drug induced cardiactoxicity

PTX3, also referred to as Pentraxin 3, belongs to the family of C Reactive Proteins (CRP) and is a good marker of an inflammatory condition of the heart. However, plasma PTX3 could also be representative of systemic inflammatory response due to sepsis or other medical conditions. Based upon the data provided herein, PTX3 may be a novel marker of cardiac function or cardiotoxicity. Additionally, the edges associated with PTX 3 in the network could form a novel panel of biomarkers.

HSP76, also referred to as HSPA6, is only known to be expressed in endothelial cells and B lymphocytes. There is no known role for this protein in cardiac function. Based upon the data provided herein, HSP76 may be a novel predictor of drug induced cardiotoxicity

PDIA4, PDIA1, also referred to as protein disulphide isomerase family A proteins, are associated with ER stress response, like GRPs. There is no known role for these proteins in cardiac function. Based upon the data provided herein, these proteins may be novel predictors of drug induced cardiotoxicity.

CA2D1 is also referred to as calcium channel, voltage-dependent, alpha 2/delta subunit. The alpha-2/delta subunit of voltage-dependent calcium channel regulates calcium current density and activation/inactivation kinetics of the calcium channel. CA2D1 plays an important role in excitation-contraction coupling in the heart. There is

no known role for this protein in cardiac function. Based upon the data provided herein, CA2D1 is a novel predictor of drug induced cardiotoxicity

GPAT1 is one of four known glycerol-3-phosphate acyltransferase isoforms, and is located on the mitochondrial outer membrane, allowing reciprocal regulation with carnitine palmitoyltransferase-1. GPAT1 is upregulated transcriptionally by insulin and SREBP-1c and downregulated acutely by AMP-activated protein kinase, consistent with a role in triacylglycerol synthesis. Based upon the data provided herein, GPAT1 is a novel predictor of drug induced cardiotoxicity.

TAZ, also referred to as Tafazzin, is highly expressed in cardiac and skeletal muscle. TAZ is involved in the metabolism of cardiolipin and functions as a phospholipid-lysophospholipid transacylase. Tafazzin is responsible for remodeling of a phospholipid cardiolipin (CL), the signature lipid of the mitochondrial inner membrane. Based upon the data provided herein, TAZ is a novel predictor of drug induced cardiotoxicity

Various aspects of the invention are described in further detail in the following subsections.

B. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules, including nucleic acids which encode a marker protein or a portion thereof. Isolated nucleic acids of the invention also include nucleic acid molecules sufficient for use as hybridization probes to identify marker nucleic acid molecules, and fragments of marker nucleic acid molecules, *e.g.*, those suitable for use as PCR primers for the amplification or mutation of marker nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. In one embodiment, an "isolated" nucleic acid molecule is free of sequences (preferably protein-encoding sequences) which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism

from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. In another embodiment, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule that is substantially free of cellular material includes preparations having less than about 30%, 20%, 10%, or 5% of heterologous nucleic acid (also referred to herein as a "contaminating nucleic acid").

A nucleic acid molecule of the present invention can be isolated using standard molecular biology techniques and the sequence information in the database records described herein. Using all or a portion of such nucleic acid sequences, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, ed., *Molecular Cloning: A Laboratory Manual, 2nd ed.,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA, or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, nucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which has a nucleotide sequence complementary to the nucleotide sequence of a marker nucleic acid or to the nucleotide sequence of a nucleic acid encoding a marker protein. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence, wherein the full length nucleic acid sequence comprises a

marker nucleic acid or which encodes a marker protein. Such nucleic acids can be used, for example, as a probe or primer. The probe/primer typically is used as one or more substantially purified oligonucleotides. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, preferably about 15, more preferably about 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, or 400 or more consecutive nucleotides of a nucleic acid of the invention.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences corresponding to one or more markers of the invention. The probe comprises a label group attached thereto, *e.g.*, a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, *e.g.*, detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

The invention further encompasses nucleic acid molecules that differ, due to degeneracy of the genetic code, from the nucleotide sequence of nucleic acids encoding a marker protein, and thus encode the same protein.

It will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence can exist within a population (*e.g.*, the human population). Such genetic polymorphisms can exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. In addition, it will be appreciated that DNA polymorphisms that affect RNA expression levels can also exist that may affect the overall expression level of that gene (*e.g.*, by affecting regulation or degradation).

As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide corresponding to a marker of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a

variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 7, 15, 20, 25, 30, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or more nucleotides in length and hybridizes under stringent conditions to a marker nucleic acid or to a nucleic acid encoding a marker protein. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in sections 6.3.1-6.3.6 of *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989). A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50-65°C.

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention that can exist in the population, the skilled artisan will further appreciate that sequence changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein encoded thereby. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologs of various species (*e.g.*, murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a variant marker protein that contain changes in amino acid residues that are not essential for activity. Such variant marker proteins differ in amino acid sequence from the naturally-occurring marker proteins, yet retain biological activity. In one

embodiment, such a variant marker protein has an amino acid sequence that is at least about 40% identical, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the amino acid sequence of a marker protein.

An isolated nucleic acid molecule encoding a variant marker protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of marker nucleic acids, such that one or more amino acid residue substitutions, additions, or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

The present invention encompasses antisense nucleic acid molecules, *i.e.*, molecules which are complementary to a sense nucleic acid of the invention, *e.g.*, complementary to the coding strand of a double-stranded marker cDNA molecule or complementary to a marker mRNA sequence. Accordingly, an antisense nucleic acid of the invention can hydrogen bond to (*i.e.* anneal with) a sense nucleic acid of the invention. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can also be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a marker protein.

The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been sub-cloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a marker protein to thereby inhibit expression of the marker, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the

case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. Examples of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site or infusion of the antisense nucleic acid into toxicity state associated body fluid. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier *et al.,* 1987, *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.,* 1987, *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.,* 1987, *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach, 1988, *Nature* 334:585-591) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a marker protein can be designed based upon the nucleotide sequence of a cDNA corresponding to the marker. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved (see Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742). Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a

catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see, *e.g.*, Bartel and Szostak, 1993, *Science* 261:1411-1418).

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a marker of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the marker nucleic acid or protein (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. See generally Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.,* 1996, *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup (1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup, 1996, *supra*; Perry-O'Keefe *et al.,* 1996, *Proc. Natl. Acad. Sci. USA* 93:14670-675).

In another embodiment, PNAs can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery

known in the art. For example, PNA-DNA chimeras can be generated which can combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup, 1996, *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra,* and Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag *et al.,* 1989, *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a step-wise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.,* 1996, *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser *et al.,* 1975, *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide can include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.,* 1989, *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.,* 1987, *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, *e.g.,* Krol *et al.,* 1988, *Bio/Techniques* 6:958-976) or intercalating agents (see, *e.g.,* Zon, 1988, *Pharm. Res.* 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The invention also includes molecular beacon nucleic acids having at least one region which is complementary to a nucleic acid of the invention, such that the molecular beacon is useful for quantitating the presence of the nucleic acid of the invention in a sample. A "molecular beacon" nucleic acid is a nucleic acid comprising a pair of complementary regions and having a fluorophore and a fluorescent quencher

associated therewith. The fluorophore and quencher are associated with different portions of the nucleic acid in such an orientation that when the complementary regions are annealed with one another, fluorescence of the fluorophore is quenched by the quencher. When the complementary regions of the nucleic acid are not annealed with one another, fluorescence of the fluorophore is quenched to a lesser degree. Molecular beacon nucleic acids are described, for example, in U.S. Patent 5,876,930.

C. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated marker proteins and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a marker protein or a fragment thereof. In one embodiment, the native marker protein can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, a protein or peptide comprising the whole or a segment of the marker protein is produced by recombinant DNA techniques. Alternative to recombinant expression, such protein or peptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, *i.e.*, it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such

preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a marker protein include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the marker protein, which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding full-length protein. A biologically active portion of a marker protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the marker protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of the marker protein.

Preferred marker proteins are encoded by nucleotide sequences comprising the sequences encoding any of the genes described in the examples. Other useful proteins are substantially identical (*e.g.*, at least about 40%, preferably 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) to one of these sequences and retain the functional activity of the corresponding naturally-occurring marker protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. Preferably, the percent identity between the two sequences is calculated using a global alignment. Alternatively, the percent identity between the two sequences is calculated using a local alignment. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions) x100). In one embodiment the two sequences are the same length. In another embodiment, the two sequences are not the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the BLASTN and BLASTX programs of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the BLASTP program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, a newer version of the BLAST algorithm called Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402, which is able to perform gapped local alignments for the programs BLASTN, BLASTP and BLASTX. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, BLASTX and BLASTN) can be used. See http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Yet another useful algorithm for identifying regions of local sequence similarity and alignment is the FASTA algorithm as described in Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85:2444-2448. When using the FASTA algorithm for comparing nucleotide or amino acid sequences, a PAM120 weight residue table can, for example, be used with a *k*-tuple value of 2.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins comprising a marker protein or a segment thereof. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably a biologically active part) of a marker protein operably linked to a heterologous polypeptide (*i.e.*, a polypeptide other than the marker protein). Within the fusion protein, the term "operably linked" is intended to indicate that the marker protein or segment thereof and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the amino-terminus or the carboxyl-terminus of the marker protein or segment.

One useful fusion protein is a GST fusion protein in which a marker protein or segment is fused to the carboxyl terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its amino terminus. For example, the native signal sequence of a marker protein can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, NY, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook *et al., supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a marker protein is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a marker protein. Inhibition of ligand/receptor interaction can be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies

directed against a marker protein in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of the marker protein with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see, *e.g.*, Ausubel *et al., supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence can be used to facilitate secretion and isolation of marker proteins. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to marker proteins, fusion proteins or segments thereof having a signal sequence, as well as to such proteins from which the signal sequence has been proteolytically cleaved (*i.e.*, the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence can be operably linked in an expression vector to a protein of interest, such as a marker protein or a segment thereof. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

The present invention also pertains to variants of the marker proteins. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation. An agonist can retain substantially the same, or a subset,

of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Variants of a marker protein which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the marker proteins from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.,* Narang, 1983, *Tetrahedron* 39:3; Itakura *et al.,* 1984, *Annu. Rev. Biochem.* 53:323; Itakura *et al.,* 1984, *Science* 198:1056; Ike *et al.,* 1983 *Nucleic Acid Res.* 11:477).

In addition, libraries of segments of a marker protein can be used to generate a variegated population of polypeptides for screening and subsequent selection of variant marker proteins or segments thereof. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes amino terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan, 1992, *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.,* 1993, *Protein Engineering* 6(3):327- 331).

Another aspect of the invention pertains to antibodies directed against a protein of the invention. In preferred embodiments, the antibodies specifically bind a marker protein or a fragment thereof. The terms "antibody" and "antibodies" as used interchangeably herein refer to immunoglobulin molecules as well as fragments and derivatives thereof that comprise an immunologically active portion of an immunoglobulin molecule, (*i.e.*, such a portion contains an antigen binding site which specifically binds an antigen, such as a marker protein, *e.g.*, an epitope of a marker protein). An antibody which specifically binds to a protein of the invention is an antibody which binds the protein, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the protein. Examples of an immunologically active portion of an immunoglobulin molecule include, but are not limited to, single-chain antibodies (scAb), F(ab) and $F(ab')_2$ fragments.

An isolated protein of the invention or a fragment thereof can be used as an immunogen to generate antibodies. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30 or more) amino acid residues of the amino acid sequence of one of the proteins of the invention, and encompasses at least one epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein. Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, *e.g.*, hydrophilic regions. Hydrophobicity sequence analysis,

hydrophilicity sequence analysis, or similar analyses can be used to identify hydrophilic regions. In preferred embodiments, an isolated marker protein or fragment thereof is used as an immunogen.

An immunogen typically is used to prepare antibodies by immunizing a suitable (*i.e.* immunocompetent) subject such as a rabbit, goat, mouse, or other mammal or vertebrate. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed or chemically-synthesized protein or peptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or a similar immunostimulatory agent. Preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a protein of the invention. In such a manner, the resulting antibody compositions have reduced or no binding of human proteins other than a protein of the invention.

The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope. Preferred polyclonal and monoclonal antibody compositions are ones that have been selected for antibodies directed against a protein of the invention. Particularly preferred polyclonal and monoclonal antibody preparations are ones that contain only antibodies directed against a marker protein or fragment thereof.

Polyclonal antibodies can be prepared by immunizing a suitable subject with a protein of the invention as an immunogen. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies (mAb) by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (see Kozbor *et al.,* 1983, *Immunol. Today* 4:72), the EBV-hybridoma technique (see Cole *et al.*, pp. 77-96 In *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., 1985) or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology,* Coligan *et*

al. ed., John Wiley & Sons, New York, 1994). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a protein of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275- 1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734.

The invention also provides recombinant antibodies that specifically bind a protein of the invention. In preferred embodiments, the recombinant antibodies specifically binds a marker protein or fragment thereof. Recombinant antibodies include, but are not limited to, chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, single-chain antibodies and multi-specific antibodies. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.,* Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397, which are incorporated herein by reference in their entirety.) Single-chain antibodies have an antigen binding site and consist of a single polypeptide. They can be produced by techniques known in the art, for example using methods described in Ladner *et. al* U.S. Pat. No. 4,946,778 (which is incorporated herein by reference in its entirety); Bird *et al.*, (1988) *Science* 242:423-426; Whitlow *et al.*, (1991) *Methods in Enzymology* 2:1-9;

Whitlow *et al.*, (1991) *Methods in Enzymology* 2:97-105; and Huston *et al.*, (1991) *Methods in Enzymology Molecular Design and Modeling: Concepts and Applications* 203:46-88. Multi-specific antibodies are antibody molecules having at least two antigen-binding sites that specifically bind different antigens. Such molecules can be produced by techniques known in the art, for example using methods described in Segal, U.S. Patent No. 4,676,980 (the disclosure of which is incorporated herein by reference in its entirety); Holliger et al., (1993) *Proc. Natl. Acad. Sci. USA* 90:6444-6448; *Whitlow et al.*, (1994) *Protein Eng.* 7:1017-1026 and U.S. Pat. No. 6,121,424.

Humanized antibodies are antibody molecules from non-human species having one or more complementarity determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, *e.g.,* Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) Humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521- 3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

More particularly, humanized antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide corresponding to a marker of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically

useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, *e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope (Jespers *et al.*, 1994, *Bio/technology* 12:899-903).

The antibodies of the invention can be isolated after production (*e.g.*, from the blood or serum of the subject) or synthesis and further purified by well-known techniques. For example, IgG antibodies can be purified using protein A chromatography. Antibodies specific for a protein of the invention can be selected or (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those of the desired protein of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein of the invention.

In a preferred embodiment, the substantially purified antibodies of the invention may specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain or cytoplasmic membrane of a protein of the invention. In a particularly preferred embodiment, the substantially purified antibodies of the invention specifically bind to a secreted sequence or an extracellular domain of the amino acid sequences of a protein of the invention. In a more preferred embodiment, the substantially purified antibodies of the invention specifically bind to a secreted sequence or an extracellular domain of the amino acid sequences of a marker protein.

An antibody directed against a protein of the invention can be used to isolate the protein by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the marker protein or fragment thereof (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the level and pattern of expression of the marker. The antibodies can also be used diagnostically to monitor protein levels in tissues or body fluids (*e.g.* in toxicity state associated body fluid) as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by the use of an antibody derivative, which comprises an antibody of the invention coupled to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibodies of the invention may also be used as therapeutic agents in treating cancers. In a preferred embodiment, completely human antibodies of the invention are used for therapeutic treatment of human cancer patients, particularly those having a cancer. In another preferred embodiment, antibodies that bind specifically to a marker protein or fragment thereof are used for therapeutic treatment. Further, such therapeutic antibody may be an antibody derivative or immunotoxin comprising an antibody

conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugated antibodies of the invention can be used for modifying a given biological response, for the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as ribosome-inhibiting protein (see Better et al., U.S. Patent No. 6,146,631, the disclosure of which is incorporated herein in its entirety), abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer

Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Accordingly, in one aspect, the invention provides substantially purified antibodies, antibody fragments and derivatives, all of which specifically bind to a protein of the invention and preferably, a marker protein. In various embodiments, the substantially purified antibodies of the invention, or fragments or derivatives thereof, can be human, non-human, chimeric and/or humanized antibodies. In another aspect, the invention provides non-human antibodies, antibody fragments and derivatives, all of which specifically bind to a protein of the invention and preferably, a marker protein. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies. In still a further aspect, the invention provides monoclonal antibodies, antibody fragments and derivatives, all of which specifically bind to a protein of the invention and preferably, a marker protein. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention. In one embodiment, the pharmaceutical composition comprises an antibody of the invention and a pharmaceutically acceptable carrier.

D. Predictive Medicine

The present invention pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining the level of expression of one or more marker proteins or nucleic

acids, in order to determine whether an individual is at risk of developing drug-induced toxicity. Such assays can be used for prognostic or predictive purposes to thereby prophylactically treat an individual prior to the onset of the disorder.

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs or other compounds administered either to inhibit or to treat or prevent or drug-induced toxicity {*i.e.* in order to understand any drug-induced toxic effects that such treatment may have}) on the expression or activity of a marker of the invention in clinical trials. These and other agents are described in further detail in the following sections.

E. Diagnostic Assays

An exemplary method for detecting the presence or absence of a marker protein or nucleic acid in a biological sample involves obtaining a biological sample (*e.g.* toxicity-associated body fluid or tissue sample) from a test subject and contacting the biological sample with a compound or an agent capable of detecting the polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA, or cDNA). The detection methods of the invention can thus be used to detect mRNA, protein, cDNA, or genomic DNA, for example, in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a marker protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. *In vivo* techniques for detection of mRNA include polymerase chain reaction (PCR), Northern hybridizations and *in situ* hybridizations. Furthermore, *in vivo* techniques for detection of a marker protein include introducing into a subject a labeled antibody directed against the protein or fragment thereof. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

A general principle of such diagnostic and prognostic assays involves preparing a sample or reaction mixture that may contain a marker, and a probe, under appropriate conditions and for a time sufficient to allow the marker and probe to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways.

For example, one method to conduct such an assay would involve anchoring the marker or probe onto a solid phase support, also referred to as a substrate, and detecting target marker/probe complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, a sample from a subject, which is to be assayed for presence and/or concentration of marker, can be anchored onto a carrier or solid phase support. In another embodiment, the reverse situation is possible, in which the probe can be anchored to a solid phase and a sample from a subject can be allowed to react as an unanchored component of the assay.

There are many established methods for anchoring assay components to a solid phase. These include, without limitation, marker or probe molecules which are immobilized through conjugation of biotin and streptavidin. Such biotinylated assay components can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In certain embodiments, the surfaces with immobilized assay components can be prepared in advance and stored.

Other suitable carriers or solid phase supports for such assays include any material capable of binding the class of molecule to which the marker or probe belongs. Well-known supports or carriers include, but are not limited to, glass, polystyrene, nylon, polypropylene, nylon, polyethylene, dextran, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

In order to conduct assays with the above mentioned approaches, the non-immobilized component is added to the solid phase upon which the second component is anchored. After the reaction is complete, uncomplexed components may be removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized upon the solid phase. The detection of marker/probe complexes anchored to the solid phase can be accomplished in a number of methods outlined herein.

In a preferred embodiment, the probe, when it is the unanchored assay component, can be labeled for the purpose of detection and readout of the assay, either directly or indirectly, with detectable labels discussed herein and which are well-known to one skilled in the art.

It is also possible to directly detect marker/probe complex formation without further manipulation or labeling of either component (marker or probe), for example by

utilizing the technique of fluorescence energy transfer (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that, upon excitation with incident light of appropriate wavelength, its emitted fluorescent energy will be absorbed by a fluorescent label on a second 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, spatial relationships between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

In another embodiment, determination of the ability of a probe to recognize a marker can be accomplished without labeling either assay component (probe or marker) by utilizing a technology such as real-time Biomolecular Interaction Analysis (BIA) (see, *e.g.*, Sjolander, S. and Urbaniczky, C., 1991, *Anal. Chem.* 63:2338-2345 and Szabo *et al.*, 1995, *Curr. Opin. Struct. Biol.* 5:699-705). As used herein, "BIA" or "surface plasmon resonance" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

Alternatively, in another embodiment, analogous diagnostic and prognostic assays can be conducted with marker and probe as solutes in a liquid phase. In such an assay, the complexed marker and probe are separated from uncomplexed components by any of a number of standard techniques, including but not limited to: differential centrifugation, chromatography, electrophoresis and immunoprecipitation. In differential centrifugation, marker/probe complexes may be separated from uncomplexed assay components through a series of centrifugal steps, due to the different sedimentation equilibria of complexes based on their different sizes and densities (see,

for example, Rivas, G., and Minton, A.P., 1993, *Trends Biochem Sci.* 18(8):284-7). Standard chromatographic techniques may also be utilized to separate complexed molecules from uncomplexed ones. For example, gel filtration chromatography separates molecules based on size, and through the utilization of an appropriate gel filtration resin in a column format, for example, the relatively larger complex may be separated from the relatively smaller uncomplexed components. Similarly, the relatively different charge properties of the marker/probe complex as compared to the uncomplexed components may be exploited to differentiate the complex from uncomplexed components, for example through the utilization of ion-exchange chromatography resins. Such resins and chromatographic techniques are well known to one skilled in the art (see, *e.g.*, Heegaard, N.H., 1998, *J. Mol. Recognit.* Winter 11(1-6):141-8; Hage, D.S., and Tweed, S.A. *J Chromatogr B Biomed Sci Appl* 1997 Oct 10;699(1-2):499-525). Gel electrophoresis may also be employed to separate complexed assay components from unbound components (see, *e.g.*, Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1987-1999). In this technique, protein or nucleic acid complexes are separated based on size or charge, for example. In order to maintain the binding interaction during the electrophoretic process, non-denaturing gel matrix materials and conditions in the absence of reducing agent are typically preferred. Appropriate conditions to the particular assay and components thereof will be well known to one skilled in the art.

In a particular embodiment, the level of marker mRNA can be determined both by *in situ* and by *in vitro* formats in a biological sample using methods known in the art. The term "biological sample" is intended to include tissues, cells, biological fluids and isolates thereof, isolated from a subject, as well as tissues, cells and fluids present within a subject. Many expression detection methods use isolated RNA. For *in vitro* methods, any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from cells (see, *e.g.*, Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York 1987-1999). Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (1989, U.S. Patent No. 4,843,155).

The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction

analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a marker of the present invention. Other suitable probes for use in the diagnostic assays of the invention are described herein. Hybridization of an mRNA with the probe indicates that the marker in question is being expressed.

In one format, the mRNA is immobilized on a solid surface and contacted with a probe, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probe(s) are immobilized on a solid surface and the mRNA is contacted with the probe(s), for example, in an Affymetrix gene chip array. A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the markers of the present invention.

An alternative method for determining the level of mRNA marker in a sample involves the process of nucleic acid amplification, *e.g.*, by RT-PCR (the experimental embodiment set forth in Mullis, 1987, U.S. Patent No. 4,683,202), ligase chain reaction (Barany, 1991, *Proc. Natl. Acad. Sci. USA,* 88:189-193), self sustained sequence replication (Guatelli *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and

with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

For *in situ* methods, mRNA does not need to be isolated from the prior to detection. In such methods, a cell or tissue sample is prepared/processed using known histological methods. The sample is then immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the marker.

As an alternative to making determinations based on the absolute expression level of the marker, determinations may be based on the normalized expression level of the marker. Expression levels are normalized by correcting the absolute expression level of a marker by comparing its expression to the expression of a gene that is not a marker, *e.g.*, a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene, or epithelial cell-specific genes. This normalization allows the comparison of the expression level in one sample, *e.g.*, a patient sample, to another sample, *e.g.*, a non-disease or non-toxic sample, or between samples from different sources.

Alternatively, the expression level can be provided as a relative expression level. To determine a relative expression level of a marker, the level of expression of the marker is determined for 10 or more samples of normal versus disease or toxic cell isolates, preferably 50 or more samples, prior to the determination of the expression level for the sample in question. The mean expression level of each of the genes assayed in the larger number of samples is determined and this is used as a baseline expression level for the marker. The expression level of the marker determined for the test sample (absolute level of expression) is then divided by the mean expression value obtained for that marker. This provides a relative expression level.

Preferably, the samples used in the baseline determination will be from non-toxic cells. The choice of the cell source is dependent on the use of the relative expression level. Using expression found in normal tissues as a mean expression score aids in validating whether the marker assayed is toxicity specific (versus normal cells). In addition, as more data is accumulated, the mean expression value can be revised, providing improved relative expression values based on accumulated data. Expression data from disesase cells or toxic cells provides a means for grading the severity of the disease or toxic state.

In another embodiment of the present invention, a marker protein is detected. A preferred agent for detecting marker protein of the invention is an antibody capable of binding to such a protein or a fragment thereof, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment or derivative thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

Proteins from cells can be isolated using techniques that are well known to those of skill in the art. The protein isolation methods employed can, for example, be such as those described in Harlow and Lane (Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

A variety of formats can be employed to determine whether a sample contains a protein that binds to a given antibody. Examples of such formats include, but are not limited to, enzyme immunoassay (EIA), radioimmunoassay (RIA), Western blot analysis and enzyme linked immunoabsorbant assay (ELISA). A skilled artisan can readily adapt known protein/antibody detection methods for use in determining whether cells express a marker of the present invention.

In one format, antibodies, or antibody fragments or derivatives, can be used in methods such as Western blots or immunofluorescence techniques to detect the expressed proteins. In such uses, it is generally preferable to immobilize either the antibody or proteins on a solid support. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

One skilled in the art will know many other suitable carriers for binding antibody or antigen, and will be able to adapt such support for use with the present invention. For example, protein isolated from disease or toxic cells can be run on a polyacrylamide gel electrophoresis and immobilized onto a solid phase support such as nitrocellulose. The

support can then be washed with suitable buffers followed by treatment with the detectably labeled antibody. The solid phase support can then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support can then be detected by conventional means.

The invention also encompasses kits for detecting the presence of a marker protein or nucleic acid in a biological sample. Such kits can be used to determine if a subject is suffering from or is at increased risk of developing drug-induced toxicity. For example, the kit can comprise a labeled compound or agent capable of detecting a marker protein or nucleic acid in a biological sample and means for determining the amount of the protein or mRNA in the sample (*e.g.*, an antibody which binds the protein or a fragment thereof, or an oligonucleotide probe which binds to DNA or mRNA encoding the protein). Kits can also include instructions for interpreting the results obtained using the kit.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a marker protein; and, optionally, (2) a second, different antibody which binds to either the protein or the first antibody and is conjugated to a detectable label.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a marker protein or (2) a pair of primers useful for amplifying a marker nucleic acid molecule. The kit can also comprise, *e.g.*, a buffering agent, a preservative, or a protein stabilizing agent. The kit can further comprise components necessary for detecting the detectable label (*e.g.*, an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

F. Pharmacogenomics

The markers of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker whose expression level correlates with a specific clinical drug response or susceptibility in a patient (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The

presence or quantity of the pharmacogenomic marker expression is related to the predicted response of the patient and more particularly the patient's diseased or toxic cells to therapy with a specific drug or class of drugs. By assessing the presence or quantity of the expression of one or more pharmacogenomic markers in a patient, a drug therapy which is most appropriate for the patient, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA or protein encoded by specific tumor markers in a patient, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the patient. The use of pharmacogenomic markers therefore permits selecting or designing the most appropriate treatment for each cancer patient without trying different drugs or regimes.

Another aspect of pharmacogenomics deals with genetic conditions that alters the way the body acts on drugs. These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who

do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the level of expression of a marker of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of expression of a marker of the invention.

G. Monitoring Clinical Trials

Monitoring the influence of agents (*e.g.*, drug compounds) on the level of expression of a marker of the invention can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent to affect marker expression can be monitored in clinical trials of subjects receiving treatment for cardiotoxicity, or drug-induced toxicity. In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of one or more selected markers of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression of the marker(s) in the post-administration samples; (v) comparing the level of expression of the marker(s) in the pre-administration sample with the level of expression of the marker(s) in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased expression of the marker gene(s) during the course of treatment may indicate ineffective dosage and the desirability of increasing the dosage. Conversely, decreased expression of the marker gene(s) may indicate efficacious treatment and no need to change dosage.

H. Arrays

The invention also includes an array comprising a marker of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of drug-induced toxicity, progression of drug-induced toxicity, and processes, such a cellular transformation associated with drug-induced toxicity.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells. This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes that could serve as a molecular target for diagnosis or therapeutic intervention.

VII. Methods for Obtaining Samples

Samples useful in the methods of the invention include any tissue, cell, biopsy, or bodily fluid sample that expresses a marker of the invention. In one embodiment, a sample may be a tissue, a cell, whole blood, serum, plasma, buccal scrape, saliva, cerebrospinal fluid, urine, stool, or bronchoalveolar lavage. In preferred embodiments, the tissue sample is a toxicity state sample. In more preferred embodiments, the tissue sample is a a cardiovascular sample or a drug-induced toxicity sample.

Body samples may be obtained from a subject by a variety of techniques known in the art including, for example, by the use of a biopsy or by scraping or swabbing an area or by using a needle to aspirate bodily fluids. Methods for collecting various body samples are well known in the art.

Tissue samples suitable for detecting and quantitating a marker of the invention may be fresh, frozen, or fixed according to methods known to one of skill in the art. Suitable tissue samples are preferably sectioned and placed on a microscope slide for further analyses. Alternatively, solid samples, *i.e.,* tissue samples, may be solubilized and/or homogenized and subsequently analyzed as soluble extracts.

In one embodiment, a freshly obtained biopsy sample is frozen using, for example, liquid nitrogen or difluorodichloromethane. The frozen sample is mounted for sectioning using, for example, OCT, and serially sectioned in a cryostat. The serial sections are collected on a glass microscope slide. For immunohistochemical staining the slides may be coated with, for example, chrome-alum, gelatine or poly-L-lysine to ensure that the sections stick to the slides. In another embodiment, samples are fixed and embedded prior to sectioning. For example, a tissue sample may be fixed in, for example, formalin, serially dehydrated and embedded in, for example, paraffin.

Once the sample is obtained any method known in the art to be suitable for detecting and quantitating a marker of the invention may be used (either at the nucleic acid or at the protein level). Such methods are well known in the art and include but are not limited to western blots, northern blots, southern blots, immunohistochemistry, ELISA, *e.g.,* amplified ELISA, immunoprecipitation, immunofluorescence, flow

cytometry, immunocytochemistry, mass spectrometrometric analyses, *e.g.*, MALDI-TOF and SELDI-TOF, nucleic acid hybridization techniques, nucleic acid reverse transcription methods, and nucleic acid amplification methods. In particular embodiments, the expression of a marker of the invention is detected on a protein level using, for example, antibodies that specifically bind these proteins.

Samples may need to be modified in order to make a marker of the invention accessible to antibody binding. In a particular aspect of the immunocytochemistry or immunohistochemistry methods, slides may be transferred to a pretreatment buffer and optionally heated to increase antigen accessibility. Heating of the sample in the pretreatment buffer rapidly disrupts the lipid bi-layer of the cells and makes the antigens (may be the case in fresh specimens, but not typically what occurs in fixed specimens) more accessible for antibody binding. The terms "pretreatment buffer" and "preparation buffer" are used interchangeably herein to refer to a buffer that is used to prepare cytology or histology samples for immunostaining, particularly by increasing the accessibility of a marker of the invention for antibody binding. The pretreatment buffer may comprise a pH-specific salt solution, a polymer, a detergent, or a nonionic or anionic surfactant such as, for example, an ethyloxylated anionic or nonionic surfactant, an alkanoate or an alkoxylate or even blends of these surfactants or even the use of a bile salt. The pretreatment buffer may, for example, be a solution of 0.1% to 1% of deoxycholic acid, sodium salt, or a solution of sodium laureth-13-carboxylate (*e.g.*, Sandopan LS) or and ethoxylated anionic complex. In some embodiments, the pretreatment buffer may also be used as a slide storage buffer.

Any method for making marker proteins of the invention more accessible for antibody binding may be used in the practice of the invention, including the antigen retrieval methods known in the art. See, for example, Bibbo, *et al.* (2002) *Acta. Cytol.* 46:25-29; Saqi, *et al.* (2003) *Diagn. Cytopathol.* 27:365-370; Bibbo, *et al.* (2003) *Anal. Quant. Cytol. Histol.* 25:8-11, the entire contents of each of which are incorporated herein by reference.

Following pretreatment to increase marker protein accessibility, samples may be blocked using an appropriate blocking agent, *e.g.*, a peroxidase blocking reagent such as hydrogen peroxide. In some embodiments, the samples may be blocked using a protein blocking reagent to prevent non-specific binding of the antibody. The protein blocking reagent may comprise, for example, purified casein. An antibody, particularly a

monoclonal or polyclonal antibody that specifically binds to a marker of the invention is then incubated with the sample. One of skill in the art will appreciate that a more accurate prognosis or diagnosis may be obtained in some cases by detecting multiple epitopes on a marker protein of the invention in a patient sample. Therefore, in particular embodiments, at least two antibodies directed to different epitopes of a marker of the invention are used. Where more than one antibody is used, these antibodies may be added to a single sample sequentially as individual antibody reagents or simultaneously as an antibody cocktail. Alternatively, each individual antibody may be added to a separate sample from the same patient, and the resulting data pooled.

Techniques for detecting antibody binding are well known in the art. Antibody binding to a marker of the invention may be detected through the use of chemical reagents that generate a detectable signal that corresponds to the level of antibody binding and, accordingly, to the level of marker protein expression. In one of the immunohistochemistry or immunocytochemistry methods of the invention, antibody binding is detected through the use of a secondary antibody that is conjugated to a labeled polymer. Examples of labeled polymers include but are not limited to polymer-enzyme conjugates. The enzymes in these complexes are typically used to catalyze the deposition of a chromogen at the antigen-antibody binding site, thereby resulting in cell staining that corresponds to expression level of the biomarker of interest. Enzymes of particular interest include, but are not limited to, horseradish peroxidase (HRP) and alkaline phosphatase (AP).

In one particular immunohistochemistry or immunocytochemistry method of the invention, antibody binding to a marker of the invention is detected through the use of an HRP-labeled polymer that is conjugated to a secondary antibody. Antibody binding can also be detected through the use of a species-specific probe reagent, which binds to monoclonal or polyclonal antibodies, and a polymer conjugated to HRP, which binds to the species specific probe reagent. Slides are stained for antibody binding using any chromagen, *e.g.,* the chromagen 3,3-diaminobenzidine (DAB), and then counterstained with hematoxylin and, optionally, a bluing agent such as ammonium hydroxide or TBS/Tween-20. Other suitable chromagens include, for example, 3-amino-9-ethylcarbazole (AEC). In some aspects of the invention, slides are reviewed microscopically by a cytotechnologist and/or a pathologist to assess cell staining, *e.g.,* fluorescent staining (*i.e.*, marker expression). Alternatively, samples may be reviewed

via automated microscopy or by personnel with the assistance of computer software that facilitates the identification of positive staining cells.

Detection of antibody binding can be facilitated by coupling the anti-marker antibodies to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin; and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S, ^{14}C, or ^{3}H.

In one embodiment of the invention frozen samples are prepared as described above and subsequently stained with antibodies against a marker of the invention diluted to an appropriate concentration using, for example, Tris-buffered saline (TBS). Primary antibodies can be detected by incubating the slides in biotinylated anti-immunoglobulin. This signal can optionally be amplified and visualized using diaminobenzidine precipitation of the antigen. Furthermore, slides can be optionally counterstained with, for example, hematoxylin, to visualize the cells.

In another embodiment, fixed and embedded samples are stained with antibodies against a marker of the invention and counterstained as described above for frozen sections. In addition, samples may be optionally treated with agents to amplify the signal in order to visualize antibody staining. For example, a peroxidase-catalyzed deposition of biotinyl-tyramide, which in turn is reacted with peroxidase-conjugated streptavidin (Catalyzed Signal Amplification (CSA) System, DAKO, Carpinteria, CA) may be used.

Tissue-based assays (*i.e.*, immunohistochemistry) are the preferred methods of detecting and quantitating a marker of the invention. In one embodiment, the presence or absence of a marker of the invention may be determined by immunohistochemistry. In one embodiment, the immunohistochemical analysis uses low concentrations of an anti-marker antibody such that cells lacking the marker do not stain. In another embodiment, the presence or absence of a marker of the invention is determined using

an immunohistochemical method that uses high concentrations of an anti-marker antibody such that cells lacking the marker protein stain heavily. Cells that do not stain contain either mutated marker and fail to produce antigenically recognizable marker protein, or are cells in which the pathways that regulate marker levels are dysregulated, resulting in steady state expression of negligible marker protein.

One of skill in the art will recognize that the concentration of a particular antibody used to practice the methods of the invention will vary depending on such factors as time for binding, level of specificity of the antibody for a marker of the invention, and method of sample preparation. Moreover, when multiple antibodies are used, the required concentration may be affected by the order in which the antibodies are applied to the sample, *e.g.*, simultaneously as a cocktail or sequentially as individual antibody reagents. Furthermore, the detection chemistry used to visualize antibody binding to a marker of the invention must also be optimized to produce the desired signal to noise ratio.

In one embodiment of the invention, proteomic methods, *e.g.,* mass spectrometry, are used for detecting and quantitating the marker proteins of the invention. For example, matrix-associated laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) or surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) which involves the application of a biological sample, such as serum, to a protein-binding chip (Wright, G.L., Jr., *et al.* (2002) *Expert Rev Mol Diagn* 2:549; Li, J., *et al.* (2002) *Clin Chem* 48:1296; Laronga, C., *et al.* (2003) *Dis Markers* 19:229; Petricoin, E.F., *et al.* (2002) 359:572; Adam, B.L., *et al.* (2002) *Cancer Res* 62:3609; Tolson, J., *et al.* (2004) *Lab Invest* 84:845; Xiao, Z., *et al.* (2001) *Cancer Res* 61:6029) can be used to detect and quantitate the PY-Shc and/or p66-Shc proteins. Mass spectrometric methods are described in, for example, U.S. Patent Nos. 5,622,824, 5,605,798 and 5,547,835, the entire contents of each of which are incorporated herein by reference.

In other embodiments, the expression of a marker of the invention is detected at the nucleic acid level. Nucleic acid-based techniques for assessing expression are well known in the art and include, for example, determining the level of marker mRNA in a sample from a subject. Many expression detection methods use isolated RNA. Any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from cells that express a marker of the invention

(see, *e.g.*, Ausubel et al., ed., (1987-1999) Current Protocols in Molecular Biology (John Wiley & Sons, New York). Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (1989, U.S. Pat. No. 4,843,155).

The term "probe" refers to any molecule that is capable of selectively binding to a marker of the invention, for example, a nucleotide transcript and/or protein. Probes can be synthesized by one of skill in the art, or derived from appropriate biological preparations. Probes may be specifically designed to be labeled. Examples of molecules that can be utilized as probes include, but are not limited to, RNA, DNA, proteins, antibodies, and organic molecules.

Isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the marker mRNA. The nucleic acid probe can be, for example, a full-length cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to marker genomic DNA.

In one embodiment, the mRNA is immobilized on a solid surface and contacted with a probe, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative embodiment, the probe(s) are immobilized on a solid surface and the mRNA is contacted with the probe(s), for example, in an Affymetrix gene chip array. A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of marker mRNA.

An alternative method for determining the level of marker mRNA in a sample involves the process of nucleic acid amplification, *e.g.*, by RT-PCR (the experimental embodiment set forth in Mullis, 1987, U.S. Pat. No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi et al., U.S. Pat. No. 5,854,033) or any other nucleic acid

amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers. In particular aspects of the invention, marker expression is assessed by quantitative fluorogenic RT-PCR (*i.e.*, the TaqMan™ System). Such methods typically utilize pairs of oligonucleotide primers that are specific for a marker of the invention. Methods for designing oligonucleotide primers specific for a known sequence are well known in the art.

The expression levels of a marker of the invention may be monitored using a membrane blot (such as used in hybridization analysis such as Northern, Southern, dot, and the like), or microwells, sample tubes, gels, beads or fibers (or any solid support comprising bound nucleic acids). See U.S. Pat. Nos. 5,770,722, 5,874,219, 5,744,305, 5,677,195 and 5,445,934, which are incorporated herein by reference. The detection of marker expression may also comprise using nucleic acid probes in solution.

In one embodiment of the invention, microarrays are used to detect the expression of a marker of the invention. Microarrays are particularly well suited for this purpose because of the reproducibility between different experiments. DNA microarrays provide one method for the simultaneous measurement of the expression levels of large numbers of genes. Each array consists of a reproducible pattern of capture probes attached to a solid support. Labeled RNA or DNA is hybridized to complementary probes on the array and then detected by laser scanning. Hybridization intensities for each probe on the array are determined and converted to a quantitative value representing relative gene expression levels. See, U.S. Pat. Nos. 6,040,138, 5,800,992 and 6,020,135, 6,033,860, and 6,344,316, which are incorporated herein by reference. High-density oligonucleotide arrays are particularly useful for determining the gene expression profile for a large number of RNA's in a sample.

The amounts of marker, and/or a mathematical relationship of the amounts of a marker of the invention may be used to calculate the risk of a toxicity state, e.g., a drug-induced toxicity or cardiotoxicity, in a subject being treated with a drug, , the efficacy of a treatment regimen for treating, preventing or counteracting a toxicity state, and the like, using the methods of the invention, which may include methods of regression analysis known to one of skill in the art. For example, suitable regression models include, but are not limited to CART (*e.g.,* Hill, T, and Lewicki, P. (2006)

"STATISTICS Methods and Applications" StatSoft, Tulsa, OK), Cox (*e.g.*, www.evidence-based-medicine.co.uk), exponential, normal and log normal (*e.g.*, www.obgyn.cam.ac.uk/mrg/statsbook/stsurvan.html), logistic (*e.g.*, www.en.wikipedia.org/wiki/Logistic_regression), parametric, non-parametric, semi-parametric (*e.g.*, www.socserv.mcmaster.ca/jfox/Books/Companion), linear (*e.g.*, www.en.wikipedia.org/wiki/Linear_regression), or additive (*e.g.*, www.en.wikipedia.org/wiki/Generalized_additive_model).

In one embodiment, a regression analysis includes the amounts of marker. In another embodiment, a regression analysis includes a marker mathematical relationship. In yet another embodiment, a regression analysis of the amounts of marker, and/or a marker mathematical relationship may include additional clinical and/or molecular co-variates. Such clinical co-variates include, but are not limited to, nodal status, tumor stage, tumor grade, tumor size, treatment regime, *e.g.*, chemotherapy and/or radiation therapy, clinical outcome (*e.g.*, relapse, disease-specific survival, therapy failure), and/or clinical outcome as a function of time after diagnosis, time after initiation of therapy, and/or time after completion of treatment.

VIII. Kits

The invention also provides compositions and kits for identifying an agent at risk for causing drug-induced toxicity, e.g., cardiotoxicity, for prognosing a cardiotoxic state, e.g., a drug-induced cardiotoxicity, recurrence of cardiotoxicity, or survival of a subject being treated for cardiotoxicity. These kits include one or more of the following: a detectable antibody that specifically binds to a marker of the invention, a detectable antibody that specifically binds to a marker of the invention, reagents for obtaining and/or preparing subject tissue samples for staining, and instructions for use.

The kits of the invention may optionally comprise additional components useful for performing the methods of the invention. By way of example, the kits may comprise fluids (*e.g.*, SSC buffer) suitable for annealing complementary nucleic acids or for binding an antibody with a protein with which it specifically binds, one or more sample compartments, an instructional material which describes performance of a method of the invention and tissue specific controls/standards.

IX. Screening Assays

Targets of the invention include, but are not limited to, the genes and/or proteins listed herein. Based on the results of experiments described by Applicants herein, the key proteins modulated in a toxicity state are associated with or can be classified into different pathways or groups of molecules, including cytoskeletal components, transcription factors, apoptotic response, pentose phosphate pathway, biosynthetic pathway, oxidative stress (pro-oxidant), membrane alterations, and oxidative phosphorylation metabolism. Accordingly, in one embodiment of the invention, a marker may include one or more genes (or proteins) selected from the markers listed in table 2. In some embodiments, the markers are a combination of at least two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, twenty-five, thirty, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more of the foregoing genes (or proteins).

Screening assays useful for identifying modulators of identified markers are described below.

The invention also provides methods (also referred to herein as "screening assays") for identifying modulators, *i.e.,* candidate or test compounds or agents (*e.g.,* proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs), which are useful for treating or preventing a toxicity state by modulating the expression and/or activity of a marker of the invention. Such assays typically comprise a reaction between a marker of the invention and one or more assay components. The other components may be either the test compound itself, or a combination of test compounds and a natural binding partner of a marker of the invention. Compounds identified *via* assays such as those described herein may be useful, for example, for modulating, *e.g.,* inhibiting, ameliorating, treating, or preventing aggressiveness of a disease state or toxicity state.

The test compounds used in the screening assays of the present invention may be obtained from any available source, including systematic libraries of natural and/or synthetic compounds. Test compounds may also be obtained by any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, *e.g.,* Zuckermann *et al.*, 1994, *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries;

synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.,* Houghten, 1992, *Biotechniques* 13:412-421), or on beads (Lam, 1991, *Nature* 354:82-84), chips (Fodor, 1993, *Nature* 364:555-556), bacteria and/or spores, (Ladner, USP 5,223,409), plasmids (Cull *et al*, 1992, *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith, 1990, *Science* 249:386-390; Devlin, 1990, *Science* 249:404-406; Cwirla *et al,* 1990, *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici, 1991, *J. Mol. Biol.* 222:301-310; Ladner, *supra.*).

The screening methods of the invention comprise contacting a toxicity state cell with a test compound and determining the ability of the test compound to modulate the expression and/or activity of a marker of the invention in the cell. The expression and/or activity of a marker of the invention can be determined as described herein.

In another embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a marker of the invention or biologically active portions thereof. In yet another embodiment, the invention provides assays for screening candidate or test compounds which bind to a marker of the invention or biologically active portions thereof. Determining the ability of the test compound to directly bind to a marker can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to the marker can be determined by detecting the labeled marker compound in a complex. For example, compounds (*e.g.,* marker substrates) can be labeled with ^{131}I, ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, assay components can be

enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent capable of modulating the expression and/or activity of a marker of the invention identified as described herein can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatment as described above.

Exemplification of the Invention

EXAMPLE 1: Employing Platform Technology to Build Models of Drug Induced Cardiotoxicity

In this example, the platform technology described in detail in international PCT Application No. PCT/US2012/027615 was employed to integrate data obtained from a custom built drug-induced cardiotoxicity model, and to identify novel proteins/pathways driving the pathogenesis/ cardiotoxicity of drugs. Relational maps resulting from this analysis have provided drug-induced cardiotoxicity biomarkers.

In the healthy heart contractile function depends on a balance of fatty acid and carbohydrate oxidation. Chronic imbalance in uptake, utilization, organellar biogenesis and secretion in non-adipose tissue (heart and liver) is thought to be at the center of mitochondrial damage and dysfunction and a key player in drug induced cardiotoxicity. Here Applicants describe a systems approach combining protein and lipid signatures with functional end point assays specifically looking at cellular bioenergetics and mitochondrial membrane function. *In vitro* models comprising diabetic and normal cardiomyocytes supplemented with excessive fatty acid and hyperglycemia were treated with a panel of drugs to create signatures and potential mechanisms of toxicity. Applicants demonstrated the varied effects of drugs in destabilizing the mitochondria by disrupting the energy metabolism component at various levels including (i) Dysregulation of transcriptional networks that controls expression of mitochondrial energy metabolism genes; (ii) Induction of GPAT1 and taffazin in diabetic cardiomyocytes thereby initiating de novo phospholipid synthesis and remodeling in the mitochondrial membrane; and (iii) Altered fate of fatty acid in diabetic cardiomyocytes, influencing uptake, fatty acid oxidation and ATP synthesis. Further, Applicants combined the power of wet lab biology and AI based data mining platform to generate causal network based on bayesian models. Networks of proteins and lipids that are causal for loss of normal cell function were used to discern mechanisms of drug induced toxicity from cellular protective mechanisms. This novel approach will serve as a powerful new tool to understand mechanism of toxicity while allowing for development of safer therapeutics that correct an altered phenotype.

Human cardiomyocytes were subject to conditions simulating an diabetic environment experienced by the disease-relevant cells *in vivo*. Specifically, the cells were exposed to hyperglycemic conditions and hyperlipidemia conditions. The hyperglycemic condition was induced by culturing cells in media containing 22 mM glucose. The hyperlipidemia condition was induced by culturing the cells in media containing 1mM L-carnitine, 0.7mM Oleic acid and 0.7mM Linoleic acid.

The cell model comprising the above-mentioned cells, wherein the cells were exposed to each condition described above, was additionally "interrogated" by exposing the cells to an "environmental perturbation" by treating with a diabetic drug (T) which is known to cause cardiotoxicity, a rescue molecule (R) or both the diabetic drug and the rescue molecule (T+R). Specifically, the cells were treated with diabetic drug; or treated with rescue molecule Coenzyme Q10 at 0, 50μM, or 100μM; or treated with both of the diabetic drug and the rescue molecule Coenzyme Q10.

Cell samples from each condition with each perturbation treatment were collected at various times following treatment, including after 6 hours of treatment. For certain conditions, media samples were also collected and analyzed.

iProfiling of changes in total cellular protein expression by quantitative proteomics was performed for cell and media samples collected for each condition and with each "environmental perturbation", i.e, diabetic drug treatment, Coenzyme Q10 treatment or both, using the techniques described above in the detailed description. Transcriptional profiling experiments were carried out using the Biorad cfx-384 amplification system. Following data collection (Ct), the final fold change over control was determined using the δCt method as outlined in manufacturer's protocol. Lipidomics experiments were carried out using mass spectrometry. Functional assays such as Oxygen consumption rate OCR were measured by employing the Seahorse analyzer essentially as recommended by the manufacturer. OCR was recorded by the electrodes in a 7 μl chamber created with the cartridge pushing against the seahorse culture plate.

As shown in Figure 20, transcriptional network and expression of human mitochondrial energy metabolism genes in diabetic cardiomyocytes (cardiomyocytes conditioned in hyperglycemic and hyperlipidemia) were compared between perturbed and unperturbed treatments. Specifically, data of transcriptional network and expression of human mitochondrial energy metabolism genes were compared between diabetic

cardiomyocytes treated with diabetic drug (T) and untreated diabetic cardiomyocytes samples (UT). Data of Transcriptional network and expression of human mitochondrial energy metabolism genes were compared between diabetic cardiomyocytes treated with both diabetic drug and rescue molecule Coenzyme Q10 (T+R) and untreated diabetic cardiomyocytes samples (UT). Comparing to data from untreated diabetic cardiomyocytes, certain genes expression and transcription were altered when diabetic cardiomyocytes were treated with diabetic drug. Rescue molecule Coenzyme Q10 was demonstrated to reverse the toxic effect of diabetic drug and normalize gene expression and transcription.

As shown in Figure 21A, cardiomyocytes were cultured either in normoglycemia (NG) or hyperglygemia (HG) condition and treated with either diabetic drug alone (T) or with both diabetic drug and rescue molecule Coenzyme Q10 (T+R) . Protein expression levels of GPAT1 and TAZ for each condition and each treatment were tested with western blotting. Both GPAT1 and TAZ were upregulated in hyperglycemia conditioned and diabetic drug treated cardiomyocytes. When hyperglycemia conditioned cardiomyocytes were treated with both diabetic drug and rescue molecule Coenzyme Q10, the upregulated protein expression level of GPAT1 and TAZ were normalized.

As shown in Figure 22A, mitochondrial oxygen consumption rate (%) experiments were carried out for hyperglycemia conditioned cardiomyocytes samples. Hyperglycemia conditioned cardiomyocytes were either untreated (UT), treated with diabetic drug T1 which is known to cause cardiotoxicity, treated with diabetic drug T2 which is known to cause cardiotoxicity, treated with both diabetic drug T1 and rescue molecule Coenzyme Q10 (T1+R), or treated with both diabetic drug T2 and rescue molecule Coenzyme Q10 (T2+R). Comparing to untreated control samples, mitochondrial OCR was decreased when hyperglycemia conditioned cardiomyocytes were treated with diabetic drug T1 or T2. However, mitochondrial OCR was normalized when hyperglycemia conditioned cardiomyocytes were treated with both diabetic drug and rescue molecule Coenzyme Q10 (T1 + R, or T2 + R).

As shown in Figure 22B, mitochondria ATP synthesis experiments were carried out for hyperglycemia conditioned cardiomyocytes samples. Hyperglycemia conditioned cardiomyocytes were either untreated (UT), treated with a diabetic drug (T), or treated with both diabetic drug and rescue molecule Coenzyme Q10 (T+R).

Comparing to untreated control samples, mitochondrial ATP synthesis was repressed when hyperglycemia conditioned cardiomyocytes were treated with diabetic drug (T).

As shown in Figure 23, based on the collected proteomic data, proteins down regulated by drug treatment were annotated with GO terms. Proteins involved in mitochondrial energy metabolism were down regulated when hyperglycemia conditioned cardiomyocytes were treated with a diabetic drug which is known to cause cardiotoxicity.

Proteomics, lipidomics, transcriptional profiling, functional assays, and western blotting data collected for each condition and with each perturbation, were then processed by the REFS™ system. Composite perturbed networks were generated from combined data obtained from one specific condition (e.g., hyperglycemia, or hyperlipidemia) exposed to each perturbation (e.g., diabetic drug, CoQ10, or both). Composite unperturbed networks were generated from combined data obtained from the same one specific condition (e.g., hyperglycemia, or hyperlipidemia), without perturbation (untreated). Similarly, composite perturbed networks were generated from combined data obtained for a second, control condition (e.g., normal glycemia) exposed to each perturbation (e.g., diabetic drug, CoQ10, or both). Composite unperturbed networks were generated from combined data obtained from the same second, control condition (e.g., normal glycemia), without perturbation (untreated).

Each node in the consensus composite networks described above was simulated (by increasing or decreasing by 10-fold) to generate simulation networks using REFS™, as described in detail above in the detailed description.

The area under the curve and fold changes for each edge connecting a parent node to a child node in the simulation networks were extracted by a custom-built program using the R programming language, where the R programming language is an open source software environment for statistical computing and graphics.

Delta networks were generated from the simulated composite networks. To generate a drug induced cardiotoxicity condition vs. normal condition differential network in response to the diabetic drug (delt network), steps of comparison were performed as illustrated in Figure 24, by a custom built program using the PERL programming language.

Specifically, as shown in Figure 24, Untreated refers to protein expression networks of untreated control cardiomyocytes in hyperglycemia condition. Drug refers

to protein expression networks of diabetic drug treated cardiomyocytes in hyperglycemia condition. Unique edges from Drug in the Drug ∩ Untreated delta network are presented in Figure 25.

Specifically, a simulated composite map of untreated cardiomyocytes in hyperglycemia condition and a simulated composite map of diabetic drug treated cardiomyocytes in hyperglycemia condition were compared using a custom-made Perl program to generate unique edges of the diabetic drug treated cardiomyocytes in hyperglycemia condition. Output from the PERL and R programs were input into Cytoscape, an open source program, to generate a visual representation of the delta network. As shown in Figure 25, the network represents delta networks that are driven by the diabetic drug versus untreated in cardiomyocytes/ cardiotox models in hyperglycemia condition.

From the drug induced toxicity condition vs. normal condition differential network shown in Figure 25, proteins were identified which drive pathophysiology of drug induced cardiotoxicity, such as GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1. These proteins can function as biomarkers for identification of other cardiotoxicity inducing drugs. These proteins can also function as biomarkers for identification of agents which can alleviate cardiotoxicity.

The experiments described in this Example demonstrate that perturbed membrane biology and altered fate of free fatty acid in diabetic cardiomyocytes exposed to drug treatment represent the center piece of drug induced toxicity. Data integration and network biology have allowed for an enhanced understanding of cardiotoxicity, and identification of novel biomarkers predictive for cardiotoxicity.

EXAMPLE 2: Employing Models of Drug Induced Cardiotoxicity to Identify Additional Markers of Cardiotoxicity

The platform technology described above in Example 1 was similarly employed to integrate further data obtained from the same custom built cardiotoxicity model. Five patient cardiomyocyte lines were used to create a model of cardiotoxicity as explained in the above-detailed description. The five cardiomyocyte lines were then subjected to a mitochondrial ATP assay to assay for mitochondrial dysfunction imposed by drug treatment or absence there of (as indicated as + and -) under diabetic conditions

(hyperglycemia) and normal conditions (normoglycemia). A reduction of mitochondrial ATP was observed under diabetic conditions upon drug treatment in only 2 out of the 5 cardiotoxicity model (see Figure 30). The results of these further experiments lead to the identification of additional novel proteins/pathways driving the pathogenesis of cardiotoxicity of drugs, as summarized in Figures 26-34.

The causal interaction network identified several novel biomarkers and potential therapeutic targets for drug-induced cardiotoxicity. Relational maps resulting from this analysis as shown in Figures 28, 29, 31-33 have provided additional drug-induced cardiotoxicity biomarkers, which are listed below in Table 2. These biomarkers may be used for predicting drug-induced cardiotoxicity of a drug, for diagnosis/prognosis of drug-induced cardiotoxicity, and for identifying a rescue agent which can reduce or alleviate drug-induced cardiotoxicity.

Table 2: biomarkers identified by the Interrogative Biology Discovery Platform

1A69, 1C17, ACBD3, ACLY, ACTR2, ANXA6, ANXA7, AP2A1, ARCN1, ASNA1, ATAD3A, ATP5A, ATP5B, ATP5D, ATP5F1, ATP5H, ATPIF1, BSG, C14orf166, CA2D1, CAPN1, CAPZA2, CARS, CCDC22, CCDC47, CCT7, CLIC4, CMPK1, CNN2, CO1A2, CO6A1, COTL1, COX6B1, CRTAP, CS010, CTSA, CTSB, CYB5, DDX1, DDX17, DDX18, DLD, EDIL3, EHD2, EIF4A3, ENO2, EPHX1, ETFA, FERMT2, FINC, FKB10, FKBP2, FLNC, G3BP2, GOLGA3, GPAT1, GPSN2, GRP75, GRP78, HMOX1, HNRNPD, HNRNPH1, HNRPG, HPX, HSP76, HSP90AB1, HSPA1A, HSPA4, HSPA9, IBP7, IDH1, IQGAP1, ITB1, ITGB1, KARS, KIF5B, KPNA3, KPNB1, LAMC1, LGALS1, LMO7, M6PRBP1, MACF1, MAP1B, MARS, MDH1, MPR1, MTHFD1, MYH10, NCL, NHP2L1, NUCB1, OLA1, P08621, P3H1, P4HA2, P4HB, SEC61A1 (P61619), PAI1, PAPSS2, PCBP2, PDCD6, PDIA1, PDIA3, PDIA3, PDIA4, PDLIM7, PEBP1, PFKM, PH4B, PLIN2, POFUT1, PRKDC, PSMA1, PSMA7, PSMD12, PSMD3, PSMD4, PSMD6, PSME2, PTBP1, PTX3, Q9BQE5, Q9Y262, RAB1B, RP515A, RPL32, RPL7A, RPL8, RPS25, RPS6, RRAS2, RRP1, SAR1B, SDHA, SENP1, SEPT11, SEPT7, SERPH, SERPINE1, SFRS2, SH3BGRL, SNRPB, SNX12, SOD1, SPRC, ST13, SUB1,

SYNCRIP, TAGLN, TAZ, TGM2, TIMP1, TLN1, TPM4, TRAP1, TSP1, TTLL12, TXNDC12, UBA1C, UGDH, UGP2, UQCRH, VAMP3, VAPA

In one embodiment, a panel of two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen markers selected from a group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, HSPA4 can be used for predicting drug-induced cardiotoxicity of a drug, for diagnosis/prognosis of drug-induced cardiotoxicity, for identifying a rescue agent which can reduce or alleviate drug-induced cardiotoxicity.

Among the markers listed in Table 2, PTX3, PAI1, IBP7 (IGFBP7) have been reported as markers of cardiomyopathy previously. GRP78 and PDIA3 have been reported as serving important indications of ER stress and hypoxic insult. The fact that these markers have been identified by the above-descriped platform technology for drug-induced cardiotoxicity, have validated this platform technology for probing novel drug-induced cardiotoxicity biomarkers.

The sDNA sequences of the markers listed in Table 2 are set forth in Appendix A, and are known in the art.

Incorporation by Reference

The contents of all cited references (including literature references, patents, patent applications, and websites) that maybe cited throughout this application are hereby expressly incorporated by reference in their entirety, as are the references cited therein. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of protein formulation, which are well known in the art.

Equivalents

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting of the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are therefore intended to be embraced herein.

Appendix A

Grp78

Official Symbol: HSPA5

Official Name: heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)

Gene ID: 3309

Organism: Homo sapiens

Other Aliases: BIP; MIF2; GRP78

Other Designations: 78 kDa glucose-regulated protein; endoplasmic reticulum lumenal Ca(2+)-binding protein **grp78**; immunoglobulin heavy chain-binding protein

Nucleotide sequence:

NCBI Reference Sequence: NM_005347.4

LOCUS NM_005347

ACCESSION NM_005347

```
        1 gggctggggg agggtatata agccgagtag gcgacggtga ggtcgacgcc
ggccaagaca
       61 gcacagacag attgacctat tggggtgttt cgcgagtgtg agagggaagc
gccgcggcct
      121 gtatttctag acctgccctt cgcctggttc gtggcgcctt gtgaccccgg
gcccctgccg
      181 cctgcaagtc ggaaattgcg ctgtgctcct gtgctacggc ctgtggctgg
actgcctgct
      241 gctgcccaac tggctggcaa gatgaagctc tccctggtgg ccgcgatgct
gctgctgctc
      301 agcgcggcgc gggccgagga ggaggacaag aaggaggacg tgggcacggt
ggtcggcatc
      361 gacctgggga ccacctactc ctgcgtcggc gtgttcaaga acggccgcgt
ggagatcatc
      421 gccaacgatc agggcaaccg catcacgccg tcctatgtcg ccttcactcc
tgaaggggaa
      481 cgtctgattg gcgatgccgc caagaaccag ctcacctcca accccgagaa
cacggtcttt
      541 gacgccaagc ggctcatcgg ccgcacgtgg aatgacccgt ctgtgcagca
ggacatcaag
      601 ttcttgccgt tcaaggtggt tgaaaagaaa actaaaccat acattcaagt
tgatattgga
      661 ggtgggcaaa caaagacatt tgctcctgaa gaaatttctg ccatggttct
cactaaaatg
      721 aaagaaaccg ctgaggctta tttgggaaag aaggttaccc atgcagttgt
tactgtacca
      781 gcctatttta atgatgccca acgccaagca accaaagacg ctggaactat
tgctggccta
      841 aatgttatga ggatcatcaa cgagcctacg gcagctgcta ttgcttatgg
cctggataag
      901 agggaggggg agaagaacat cctggtgttt gacctgggtg gcggaacctt
cgatgtgtct
```

961 cttctcacca ttgacaatgg tgtcttcgaa gttgtggcca ctaatggaga
tactcatctg
1021 ggtggagaag actttgacca gcgtgtcatg gaacacttca tcaaactgta
caaaaagaag
1081 acgggcaaag atgtcaggaa agacaataga gctgtgcaga aactccggcg
cgaggtagaa
1141 aaggccaaac gggccctgtc ttctcagcat caagcaagaa ttgaaattga
gtccttctat
1201 gaaggagaag acttttctga gaccctgact cgggccaaat ttgaagagct
caacatggat
1261 ctgttccggt ctactatgaa gcccgtccag aaagtgttgg aagattctga
tttgaagaag
1321 tctgatattg atgaaattgt tcttgttggt ggctcgactc gaattccaaa
gattcagcaa
1381 ctggttaaag agttcttcaa tggcaaggaa ccatcccgtg gcataaaccc
agatgaagct
1441 gtagcgtatg gtgctgctgt ccaggctggt gtgctctctg gtgatcaaga
tacaggtgac
1501 ctggtactgc ttgatgtatg tccccttaca cttggtattg aaactgtggg
aggtgtcatg
1561 accaaactga ttccaaggaa cacagtggtg cctaccaaga agtctcagat
cttttctaca
1621 gcttctgata atcaaccaac tgttacaatc aaggtctatg aaggtgaaag
acccctgaca
1681 aaagacaatc atcttctggg tacatttgat ctgactggaa ttcctcctgc
tcctcgtggg
1741 gtcccacaga ttgaagtcac ctttgagata gatgtgaatg gtattcttcg
agtgacagct
1801 gaagacaagg gtacagggaa caaaaataag atcacaatca ccaatgacca
gaatcgcctg
1861 acacctgaag aaatcgaaag gatggttaat gatgctgaga agtttgctga
ggaagacaaa
1921 aagctcaagg agcgcattga tactagaaat gagttggaaa gctatgccta
ttctctaaag
1981 aatcagattg gagataaaga aaagctggga ggtaaacttt cctctgaaga
taaggagacc
2041 atggaaaaag ctgtagaaga aaagattgaa tggctggaaa gccaccaaga
tgctgacatt
2101 gaagacttca aagctaagaa gaaggaactg gaagaaattg ttcaaccaat
tatcagcaaa
2161 ctctatggaa gtgcaggccc tcccccaact ggtgaagagg atacagcaga
aaaagatgag
2221 ttgtagacac tgatctgcta gtgctgtaat attgtaaata ctggactcag
gaacttttgt
2281 taggaaaaaa ttgaaagaac ttaagtctcg aatgtaattg gaatcttcac
ctcagagtgg
2341 agttgaaact gctatagcct aagcggctgt ttactgcttt tcattagcag
ttgctcacat
2401 gtctttgggt gggggggaga agaagaattg gccatcttaa aaagcgggta
aaaaacctgg
2461 gttagggtgt gtgttcacct tcaaaatgtt ctatttaaca actgggtcat
gtgcatctgg
2521 tgtaggaagt tttttctacc ataagtgaca ccaataaatg tttgttattt
acactggtct
2581 aatgtttgtg agaagcttct aattagatca attacttatt ttaggaaatt
taagactaga
2641 tactcgtgtg tggggtgagg ggagggagta tttggtatgt tgggataagg
aaacacttct
2701 atttaatgct tccagggatt ttttttttt tttttaaccc tcctgggccc
aagtgatcct

```
     2761 tccacctcag tctcccagct aattgagacc acaggcttgt taccaccatg
ctcggctttt
     2821 gcattaatct aagaaaaggg gagagaagtt aatccacatc tttactcagg
caaggggcat
     2881 ttcacagtgc ccaagagtgg ggttttcttg aacatacttg gtttcctatt
tccccttatc
     2941 tttctaaaac tgcctttctg gtggcttttt ttaaaattat tactaatgat
gcttttatag
     3001 ctgcttggat tctctgagaa atgatgggga gtgagtgatc actggtatta
actttataca
     3061 cttggatttc atttgtaact ttaggatgta aaggtatatt gtgaacccta
gctgtgtcag
     3121 aatctccatc cctgaaattt ctcattagtg gtactggggt gggatcttgg
atggtgacat
     3181 tgaaactaca ctaaatcccc tcactatgaa tgggttgtta aaggcaatgg
tttgtgtcaa
     3241 aactggttta ggattactta gattgtgttc ctgaagaaaa gagtccaggt
aaatggtatg
     3301 atcaataaag gacaggctgg tgctaacata aaatccaata ttgtaatcct
agcactttgg
     3361 gaggccaagg cgggtggatc acaaggtcaa gagatagaga ccatctttgc
caacatggtg
     3421 aaactccatc tctactgaaa atacaaaaat tagctgggcg tggtagtgca
agctgaaggc
     3481 tgaggcagga gaatcactcg aacccgggag gcagaggttg cagtgagccg
agatcacacc
     3541 actgtactcc agcccggcac tccagcctgg cgacaagagt gagactccac
ctcaaaaaaa
     3601 aaaaaaagaa tccaatactg cccaaggata ggtattttat agatgggcaa
ctggctgaaa
     3661 ggttaattct ctagggctag tagaactgga tcccaacacc aaactcttaa
ttagacctag
     3721 gcctcagctg cactgcccga aaagcatttg ggcagaccct gagcagaata
ctggtctcag
     3781 gccaagccca atacagccat taaagatgac ctacagtgct gtgtaccctg
gggcaatagg
     3841 gttaaatggt agttagcaac tagggctagt cttcccttac ctcaaaggct
ctcactaccg
     3901 tggaccacct agtctgtaac tctttctgag gagctgttac tgaatattaa
aaagatagac
     3961 ttcaactatg aaa
```

Protein sequence:

NCBI Reference Sequence: NP_005338.1

LOCUS NP_005338

ACCESSION NP_005338

```
      1 mklslvaaml lllsaaraee edkkedvgtv vgidlgttys cvgvfkngrv
eiiandqgnr
       61 itpsyvaftp egerligdaa knqltsnpen tvfdakrlig rtwndpsvqq
dikflpfkvv
      121 ekktkpyiqv digggqtktf apeeisamvl tkmketaeay lgkkvthavv
tvpayfndaq
      181 rqatkdagti aglnvmriin eptaaaiayg ldkregekni lvfdlgggtf
dvslltidng
```

```
      241 vfevvatngd thlggedfdq rvmehfikly kkktgkdvrk dnravqklrr
evekakrals
      301 sqhqarieie sfyegedfse tltrakfeel nmdlfrstmk pvqkvledsd
lkksdideiv
      361 lvggstripk iqqlvkeffn gkepsrginp deavaygaav qagvlsgdqd
tgdlvlldvc
      421 pltlgietvg gvmtkliprn tvvptkksqi fstasdnqpt vtikvyeger
pltkdnhllg
      481 tfdltgippa prgvpqievt feidvngilr vtaedkgtgn knkititndq
nrltpeeier
      541 mvndaekfae edkklkerid trnelesyay slknqigdke klggklssed
ketmekavee
      601 kiewleshqd adiedfkakk keleeivqpi isklygsagp pptgeedtae kdel
```

Grp75

Official Symbol: HSPA9

Official Name: heat shock 70kDa protein 9 (mortalin)

Gene ID: 3313

Organism: Homo sapiens

Other Aliases: CSA; MOT; MOT2; GRP75; PBP74; GRP-75; HSPA9B; MTHSP75

Other Designations: 75 kDa glucose-regulated protein; heat shock 70kD protein 9B; mortalin, perinuclear; mortalin-2; p66-mortalin; peptide-binding protein 74; stress-70 protein, mitochondrial

Nucleotide sequence:

NCBI Reference Sequence: NM_004134.6

LOCUS NM_004134

ACCESSION NM_004134

```
      1 ttcctcccct ggactctttc tgagctcaga gccgccgcag ccgggacagg
agggcaggct
       61 ttctccaacc atcatgctgc ggagcatatt acctgtacgc cctggctccg
ggagcggcag
      121 tcgagtatcc tctggtcagg cggcgcgggc ggcgcctcag cggaagagcg
ggcctctggg
      181 ccgcagtgac caacccccgc ccctcacccc acgtggttgg aggtttccag
aagcgctgcc
      241 gccaccgcat cgcgcagctc tttgccgtcg gagcgcttgt ttgctgcctc
gtactcctcc
      301 atttatccgc catgataagt gccagccgag ctgcagcagc ccgtctcgtg
ggcgccgcag
      361 cctcccgggg ccctacggcc gcccgccacc aggatagctg gaatggcctt
agtcatgagg
      421 cttttagact tgtttcaagg cgggattatg catcagaagc aatcaaggga
gcagttgttg
```

481 gtattgattt gggtactacc aactcctgcg tggcagttat ggaaggtaaa caagcaaagg
541 tgctggagaa tgccgaaggt gccagaacca ccccttcagt tgtggccttt acagcagatg
601 gtgagcgact tgttggaatg ccggccaagc gacaggctgt caccaaccca aacaatacat
661 tttatgctac caagcgtctc attggccggc gatatgatga tcctgaagta cagaaagaca
721 ttaaaaatgt tccctttaaa attgtccgtg cctccaatgg tgatgcctgg gttgaggctc
781 atgggaaatt gtattctccg agtcagattg gagcatttgt gttgatgaag atgaaagaga
841 ctgcagaaaa ttacttgggg cacacagcaa aaaatgctgt gatcacagtc ccagcttatt
901 tcaatgactc gcagagacag gccactaaag atgctggcca gatatctgga ctgaatgtgc
961 ttcgggtgat taatgagccc acagctgctg ctcttgccta tggtctagac aaatcagaag
1021 acaaagtcat tgctgtatat gatttaggtg gtggaacttt tgatatttct atcctggaaa
1081 ttcagaaagg agtatttgag gtgaaatcca caaatgggga taccttctta ggtggggaag
1141 actttgacca ggccttgcta cggcacattg tgaaggagtt caagagagag acaggggttg
1201 atttgactaa agacaacatg gcacttcaga gggtacggga agctgctgaa aaggctaaat
1261 gtgaactctc ctcatctgtg cagactgaca tcaatttgcc ctatcttaca atggattctt
1321 ctggacccaa gcatttgaat atgaagttga cccgtgctca atttgaaggg attgtcactg
1381 atctaatcag aaggactatc gctccatgcc aaaaagctat gcaagatgca gaagtcagca
1441 agagtgacat aggagaagtg attcttgtgg gtggcatgac taggatgccc aaggttcagc
1501 agactgtaca ggatcttttt ggcagagccc caagtaaagc tgtcaatcct gatgaggctg
1561 tggccattgg agctgccatt cagggaggtg tgttggccgg cgatgtcacg gatgtgctgc
1621 tccttgatgt cactcccctg tctctgggta ttgaaactct aggaggtgtc tttaccaaac
1681 ttattaatag gaataccact attccaacca agaagagcca ggtattctct actgccgctg
1741 atggtcaaac gcaagtggaa attaaagtgt gtcagggtga aagagagatg gctggagaca
1801 acaaactcct tggacagttt actttgattg gaattccacc agcccctcgt ggagttcctc
1861 agattgaagt tacatttgac attgatgcca atgggatagt acatgtttct gctaaagata
1921 aaggcacagg acgtgagcag cagattgtaa tccagtcttc tggtggatta agcaaagatg
1981 atattgaaaa tatggttaaa aatgcagaga aatatgctga agaagaccgg cgaaagaagg
2041 aacgagttga agcagttaat atggctgaag gaatcattca cgacacagaa accaagatgg
2101 aagaattcaa ggaccaatta cctgctgatg agtgcaacaa gctgaaagaa gagatttcca
2161 aaatgaggga gctcctggct agaaaagaca gcgaaacagg agaaaatatt agacaggcag
2221 catcctctct tcagcaggca tcactgaagc tgttcgaaat ggcatacaaa aagatggcat

```
     2281 ctgagcgaga aggctctgga agttctggca ctggggaaca aaaggaagat
caaaaggagg
     2341 aaaaacagta ataatagcag aaattttgaa gccagaagga caacatatga
agcttaggag
     2401 tgaagagact tcctgagcag aaatgggcga acttcagtct ttttactgtg
tttttgcagt
     2461 attctatata taatttcctt aatttgtaaa tttagtgacc attagctagt
gatcatttaa
     2521 tggacagtga ttctaacagt ataaagttca caatattcta tgtccctagc
ctgtcatttt
     2581 tcagctgcat gtaaaaggag gtaggatgaa ttgatcatta taaagattta
actattttat
     2641 gctgaagtga ccatattttc aaggggtgaa accatctcgc acacagcaat
gaaggtagtc
     2701 atccatagac ttgaaatgag accacatatg gggatgagat ccttctagtt
agcctagtac
     2761 tgctgtactg gcctgtatgt acatggggtc cttcaactga ggccttgcaa
gtcaagctgg
     2821 ctgtgccatg tttgtagatg gggcagagga atctagaaca atgggaaact
tagctattta
     2881 tattaggtac agctattaaa acaaggtagg aatgaggcta gacctttaac
ttccctaagg
     2941 catacttttc tagctacctt ctgccctgtg tctggcacct acatccttga
tgattgttct
     3001 cttacccatt ctggaatttt ttttttttta aataaataca gaaagcatct
tgatctcttg
     3061 tttgtgaggg gtgatgccct gagatttagc ttcaagaata tgccatggct
catgcttccc
     3121 atatttccca aagagggaaa tacaggattt gctaacactg gttaaaaatg
caaattcaag
     3181 atttggaagg gctgttataa tgaaataatg agcagtatca gcatgtgcaa
atcttgtttg
     3241 aaggatttta ttttctcccc ttagaccttt ggtacattta gaatcttgaa
agtttctaga
     3301 tctctaacat gaaagtttct agatctctaa catgaaagtt tttagatctc
taacatgaaa
     3361 accaaggtgg ctattttcag gttgctttca gctccaagta gaaataacca
gaattggctt
     3421 acattaaaga aactgcatct agaaataagt cctaagatac tatttctatg
gctcaaaaat
     3481 aaaaggaacc cagatttctt tcccta
```

Protein sequence:

NCBI Reference Sequence: NP_004125.3

LOCUS NP_004125

ACCESSION NP_004125

```
      1 misasraaaa rlvgaaasrg ptaarhqdsw nglsheafrl vsrrdyasea
ikgavvgidl
       61 gttnscvavm egkqakvlen aegarttpsv vaftadgerl vgmpakrqav
tnpnntfyat
      121 krligrrydd pevqkdiknv pfkivrasng dawveahgkl yspsqigafv
lmkmketaen
      181 ylghtaknav itvpayfnds qrqatkdagq isglnvlrvi neptaaalay
gldksedkvi
```

```
      241 avydlgggtf disileiqkg vfevkstngd tflggedfdq allrhivkef
kretgvdltk
      301 dnmalqrvre aaekakcels ssvqtdinlp yltmdssgpk hlnmkltraq
fegivtdlir
      361 rtiapcqkam qdaevsksdi gevilvggmt rmpkvqqtvq dlfgrapska
vnpdeavaig
      421 aaiqggvlag dvtdvllldv tplslgietl ggvftklinr nttiptkksq
vfstaadgqt
      481 qveikvcqge remagdnkll gqftligipp aprgvpqiev tfdidangiv
hvsakdkgtg
      541 reqqiviqss gglskddien mvknaekyae edrrkkerve avnmaegiih
dtetkmeefk
      601 dqlpadecnk lkeeiskmre llarkdsetg enirqaassl qqaslklfem
aykkmasere
      661 gsgssgtgeq kedqkeekq
```

TIMP1

Official Symbol: TIMP1

Official Name: TIMP metallopeptidase inhibitor 1

Gene ID: Gene ID: 7076

Organism: Homo sapiens

Other Aliases: RP1-230G1.3, CLGI, EPA, EPO, HCI, TIMP

Other Designations: TIMP-1; collagenase inhibitor; erythroid potentiating activity; erythroid-potentiating activity; fibroblast collagenase inhibitor; metalloproteinase inhibitor 1; tissue inhibitor of metalloproteinases 1

Nucleotide sequence:

NCBI Reference Sequence: NM_003254.2

LOCUS NM_003254

ACCESSION NM_003254

```
      1 tttcgtcggc ccgcccttg gcttctgcac tgatggtggg tggatgagta
atgcatccag
       61 gaagcctgga ggcctgtggt ttccgcaccc gctgccaccc ccgcccctag
cgtggacatt
      121 tatcctctag cgctcaggcc ctgccgccat cgccgcagat ccagcgccca
gagagacacc
      181 agagaaccca ccatggcccc ctttgagccc ctggcttctg gcatcctgtt
gttgctgtgg
      241 ctgatagccc ccagcagggc ctgcacctgt gtcccacccc acccacagac
ggccttctgc
      301 aattccgacc tcgtcatcag ggccaagttc gtggggacac cagaagtcaa
ccagaccacc
      361 ttataccagc gttatgagat caagatgacc aagatgtata aagggttcca
agccttaggg
      421 gatgccgctg acatccggtt cgtctacacc cccgccatgg agagtgtctg
cggatacttc
      481 cacaggtccc acaaccgcag cgaggagttt ctcattgctg gaaaactgca
ggatggactc
```

```
      541 ttgcacatca ctacctgcag ttttgtggct ccctggaaca gcctgagctt
agctcagcgc
      601 cggggcttca ccaagaccta cactgttggc tgtgaggaat gcacagtgtt
tccctgttta
      661 tccatcccct gcaaactgca gagtggcact cattgcttgt ggacggacca
gctcctccaa
      721 ggctctgaaa agggcttcca gtcccgtcac cttgcctgcc tgcctcggga
gccagggctg
      781 tgcacctggc agtccctgcg gtcccagata gcctgaatcc tgcccggagt
ggaagctgaa
      841 gcctgcacag tgtccaccct gttcccactc ccatctttct tccggacaat
gaaataaaga
      901 gttaccaccc agcagaaaaa aaaaaaaaaa a
```

Protein sequence:

NCBI Reference Sequence: NP_003245.1

LOCUS NP_003245

ACCESSION NP_003245

```
      1 mapfeplasg illllwliap sractcvpph pqtafcnsdl virakfvgtp
evnqttlyqr
       61 yeikmtkmyk gfqalgdaad irfvytpame svcgyfhrsh nrseefliag
klqdgllhit
      121 tcsfvapwns lslaqrrgft ktytvgceec tvfpclsipc klqsgthclw
tdqllqgsek
      181 gfqsrhlacl prepglctwq slrsqia
```

PTX3

Official Symbol: PTX3

Official Name: pentraxin 3, long

Gene ID: 5806

Organism: Homo sapiens

Other Aliases: TNFAIP5, TSG-14

Other Designations: TNF alpha-induced protein 5; pentaxin-related gene, rapidly induced by IL-1 beta, tumor necrosis factor, alpha-induced protein 5; pentaxin-related protein **PTX3**; pentraxin-3; pentraxin-related gene, rapidly induced by IL-1 beta; pentraxin-related protein **PTX3**; tumor necrosis factor alpha-induced protein 5; tumor necrosis factor, alpha-induced protein 5; tumor necrosis factor-inducible gene 14 protein; tumor necrosis factor-inducible protein TSG-14

Nucleotide sequence:

NCBI Reference Sequence: NM_002852.3

LOCUS NM_002852

ACCESSION NM_002852

```
      1 attcatcccc attcaggctt tcctcagcat ttattaagga ctctctgctc
cagcctctca
       61 ctctcactct cctccgctca aactcagctc acttgagagt ctcctcccgc
cagctgtgga
      121 aagaactttg cgtctctcca gcaatgcatc tccttgcgat tctgttttgt
gctctctggt
      181 ctgcagtgtt ggccgagaac tcggatgatt atgatctcat gtatgtgaat
ttggacaacg
      241 aaatagacaa tggactccat cccactgagg accccacgcc gtgcgcctgc
ggtcaggagc
      301 actcggaatg ggacaagctc ttcatcatgc tggagaactc gcagatgaga
gagcgcatgc
      361 tgctgcaagc cacggacgac gtcctgcggg gcgagctgca gaggctgcgg
gaggagctgg
      421 gccggctcgc ggaaagcctg gcgaggccgt gcgcgccggg ggctcccgca
gaggccaggc
      481 tgaccagtgc tctggacgag ctgctgcagg cgacccgcga cgcgggccgc
aggctggcgc
      541 gtatggaggg cgcggaggcg cagcgcccag aggaggcggg gcgcgccctg
gccgcggtgc
      601 tagaggagct gcggcagacg cgagccgacc tgcacgcggt gcagggctgg
gctgcccgga
      661 gctggctgcc ggcaggttgt gaaacagcta ttttattccc aatgcgttcc
aagaagattt
      721 ttggaagcgt gcatccagtg agaccaatga ggcttgagtc ttttagtgcc
tgcatttggg
      781 tcaaagccac agatgtatta aacaaaacca tcctgttttc ctatggcaca
aagaggaatc
      841 catatgaaat ccagctgtat ctcagctacc aatccatagt gtttgtggtg
ggtggagagg
      901 agaacaaact ggttgctgaa gccatggttt ccctgggaag gtggacccac
ctgtgcggca
      961 cctggaattc agaggaaggg ctcacatcct tgtgggtaaa tggtgaactg
gcggctacca
     1021 ctgttgagat ggccacaggt cacattgttc ctgagggagg aatcctgcag
attggccaag
     1081 aaaagaatgg ctgctgtgtg ggtggtggct ttgatgaaac attagccttc
tctgggagac
     1141 tcacaggctt caatatctgg gatagtgttc ttagcaatga agagataaga
gagaccggag
     1201 gagcagagtc ttgtcacatc cgggggaata ttgttgggtg gggagtcaca
gagatccagc
     1261 cacatggagg agctcagtat gtttcataaa tgttgtgaaa ctccacttga
agccaaagaa
     1321 agaaactcac acttaaaaca catgccagtt gggaaggtct gaaaactcag
tgcataatag
     1381 gaacacttga gactaatgaa agagagagtt gagaccaatc tttatttgta
ctggccaaat
     1441 actgaataaa cagttgaagg aaagacattg gaaaaagctt ttgaggataa
tgttactaga
     1501 ctttatgcca tggtgctttc agtttaatgc tgtgtctctg tcagataaac
tctcaaataa
     1561 ttaaaaagga ctgtattgtt gaacagaggg acaattgttt tacttttctt
tggttaattt
     1621 tgttttggcc agagatgaat tttacattgg aagaataaca aaataagatt
tgttgtccat
     1681 tgttcattgt tattggtatg taccttatta caaaaaaaag atgaaaacat
atttatacta
     1741 caaggtgact taacaactat aaatgtagtt tatgtgttat aatcgaatgt
cacgtttttg
```

```
     1801 agaagatagt catataagtt atattgcaaa agggatttgt attaatttaa
gactattttt
     1861 gtaaagctct actgtaaata aaatatttta taaaactagc tcacgtcatt
taattataaa
     1921 tttaagagat gttttggaaa aaaaaaaaaa aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_002843.2

LOCUS NP_002843

ACCESSION NP_002843

```
      1 mhllailfca lwsavlaens ddydlmyvnl dneidnglhp tedptpcacg
qehsewdklf
       61 imlensqmre rmllqatddv lrgelqrlre elgrlaesla rpcapgapae
arltsaldel
      121 lqatrdagrr larmegaeaq rpeeagrala avleelrqtr adlhavqgwa
arswlpagce
      181 tailfpmrsk kifgsvhpvr pmrlesfsac iwvkatdvln ktilfsygtk
rnpyeiqlyl
      241 syqsivfvvg geenklvaea mvslgrwthl cgtwnseegl tslwvngela
attvematgh
      301 ivpeggilqi gqekngccvg ggfdetlafs grltgfniwd svlsneeire
tggaeschir
      361 gnivgwgvte iqphggaqyv s
```

HSP76

Official Symbol: HSPA6

Official Name: heat shock 70kDa protein 6 (HSP70B')

Gene ID: 3310

Organism: Homo sapiens

Other Aliases:

Other Designations: heat shock 70 kDa protein 6; heat shock 70 kDa protein B'; heat shock 70kD protein 6 (HSP70B')

Nucleotide sequence:

NCBI Reference Sequence: NM_002155.3

LOCUS NM_002155

ACCESSION NM_002155

```
      1 agagccagcc cggaggagct agaaccttcc ccgcatttct ttcagcagcc
tgagtcagag
       61 gcgggctggc ctggcgtagc cgcccagcct cgcggctcat gccccgatct
gcccgaacct
```

121 tctcccgggg tcagcgccgc gccgcgccac ccggctgagt cagcccgggc gggcgagagg
181 ctctcaactg ggcgggaagg tgcgggaagg tgcggaaagg ttcgcgaaag ttcgcggcgg
241 cgggggtcgg gtgaggcgca aaaggataaa aagcccgtgg aagcggagct gagcagatcc
301 gagccgggct ggctgcagag aaaccgcagg gagagcctca ctgctgagcg cccctcgacg
361 gcggagcggc agcagcctcc gtggcctcca gcatccgaca agaagcttca gccatgcagg
421 ccccacggga gctcgcggtg ggcatcgacc tgggcaccac ctactcgtgc gtgggcgtgt
481 ttcagcaggg ccgcgtggag atcctggcca acgaccaggg caaccgcacc acgcccagct
541 acgtggcctt caccgacacc gagcggctgg tcggggacgc ggccaagagc caggcggccc
601 tgaaccccca caacaccgtg ttcgatgcca agcggctgat cgggcgcaag ttcgcggaca
661 ccacggtgca gtcggacatg aagcactggc ccttccgggt ggtgagcgag ggcggcaagc
721 ccaaggtgcg cgtatgctac cgcggggagg acaagacgtt ctaccccgag gagatctcgt
781 ccatggtgct gagcaagatg aaggagacgg ccgaggcgta cctgggccag cccgtgaagc
841 acgcagtgat caccgtgccc gcctatttca atgactcgca gcgccaggcc accaaggacg
901 cgggggccat cgcggggctc aacgtgttgc ggatcatcaa tgagcccacg gcagctgcca
961 tcgcctatgg gctggaccgg cggggcgcgg gagagcgcaa cgtgctcatt tttgacctgg
1021 gtgggggcac cttcgatgtg tcggttctct ccattgacgc tggtgtcttt gaggtgaaag
1081 ccactgctgg agatacccac ctgggaggag aggacttcga caaccggctc gtgaaccact
1141 tcatggaaga attccggcgg aagcatggga aggacctgag cgggaacaag cgtgccctgc
1201 gcaggctgcg cacagcctgt gagcgcgcca agcgcaccct gtcctccagc acccaggcca
1261 ccctggagat agactccctg ttcgagggcg tggacttcta cacgtccatc actcgtgccc
1321 gctttgagga actgtgctca gacctcttcc gcagcaccct ggagccggtg gagaaggccc
1381 tgcgggatgc caagctggac aaggcccaga ttcatgacgt cgtcctggtg gggggctcca
1441 ctcgcatccc caaggtgcag aagttgctgc aggacttctt caacggcaag gagctgaaca
1501 agagcatcaa ccctgatgag gctgtggcct atggggctgc tgtgcaggcg gccgtgttga
1561 tgggggacaa atgtgagaaa gtgcaggatc tcctgctgct ggatgtggct cccctgtctc
1621 tggggctgga gacagcaggt ggggtgatga ccacgctgat ccagaggaac gccactatcc
1681 ccaccaagca gacccagact ttcaccacct actcggacaa ccagcctggg gtcttcatcc
1741 aggtgtatga gggtgagagg gccatgacca aggacaacaa cctgctgggg cgttttgaac
1801 tcagtggcat ccctcctgcc ccacgtggag tcccccagat agaggtgacc tttgacattg
1861 atgctaatgg catcctgagc gtgacagcca ctgacaggag cacaggtaag gctaacaaga

```
     1921 tcaccatcac caatgacaag ggccggctga gcaaggagga ggtggagagg
atggttcatg
     1981 aagccgagca gtacaaggct gaggatgagg cccagaggga cagagtggct
gccaaaaact
     2041 cgctggaggc ccatgtcttc catgtgaaag gttctttgca agaggaaagc
cttagggaca
     2101 agattcccga agaggacagg cgcaaaatgc aagacaagtg tcgggaagtc
cttgcctggc
     2161 tggagcacaa ccagctggca gagaaggagg agtatgagca tcagaagagg
gagctggagc
     2221 aaatctgtcg ccccatcttc tccaggctct atgggggggcc tggtgtccct
gggggcagca
     2281 gttgtggcac tcaagcccgc caggggggacc ccagcaccgg ccccatcatt
gaggaggttg
     2341 attgaatggc ccttcgtgat aagtcagctg tgactgtcag ggctatgcta
tgggccttct
     2401 agactgtctt ctatgatcct gcccttcaga gatgaacttt ccctccaaag
ctagaacttt
     2461 cttcccagga taactgaagt cttttgactt tttgggggga gggcggttca
tcctcttctg
     2521 cttcaaataa aaagtcatta atttattaaa acttgtgtgg cactttaaca
ttgctttcac
     2581 ctatattttg tgtactttgt tacttgcatg tatgaatttt gttatgtaaa
atatagttat
     2641 agacctaaat aaaaaaaaaa aaaa
```

Protein sequence:

NCBI Reference Sequence: NP_002146.2

LOCUS NP_002146

ACCESSION NP_002146

```
      1 mqaprelavg idlgttyscv gvfqqgrvei landqgnrtt psyvaftdte
rlvgdaaksq
       61 aalnphntvf dakrligrkf adttvqsdmk hwpfrvvseg gkpkvrvcyr
gedktfypee
      121 issmvlskmk etaeaylgqp vkhavitvpa yfndsqrqat kdagaiagln
vlriinepta
      181 aaiaygldrr gagernvlif dlgggtfdvs vlsidagvfe vkatagdthl
ggedfdnrlv
      241 nhfmeefrrk hgkdlsgnkr alrrlrtace rakrtlssst qatleidslf
egvdfytsit
      301 rarfeelcsd lfrstlepve kalrdakldk aqihdvvlvg gstripkvqk
llqdffngke
      361 lnksinpdea vaygaavqaa vlmgdkcekv qdllllldvap lslgletagg
vmttliqrna
      421 tiptkqtqtf ttysdnqpgv fiqvyegera mtkdnnllgr felsgippap
rgvpqievtf
      481 didangilsv tatdrstgka nkititndkg rlskeeverm vheaeqykae
deaqrdrvaa
      541 knsleahvfh vkgslqeesl rdkipeedrr kmqdkcrevl awlehnqlae
keeyehqkre
      601 leqicrpifs rlyggpgvpg gsscgtqarq gdpstgpiie evd
```

PDIA4

Official Symbol: PDIA4

Official Name: protein disulfide isomerase family A, member 4

Gene ID: 9601

Organism: Homo sapiens

Other Aliases: ERP70, ERP72, ERp-72

Other Designations: ER protein 70; ER protein 72; endoplasmic reticulum resident protein 70; endoplasmic reticulum resident protein 72; protein disulfide isomerase related protein (calcium-binding protein, intestinal-related); protein disulfide isomerase-associated 4; protein disulfide-isomerase A4

Nucleotide sequence:

NCBI Reference Sequence: NM_004911.4

LOCUS NM_004911

ACCESSION NM_004911

```
      1 gttttaaacg cgcagccgag ggccgcgcgc aggagtaggg agggcctagg
gcggcggagc
       61 cgactcgtcg cggccgaggc gcgcgcggtc cgtgccggcg tcagtctggg
attggccggc
      121 ccgcgacttc ctccgccccc tgccaatcgc cggggacgac ttccgtgggt
ttttccggct
      181 cccccgcgtc gctaaggagc gacgggctgt cggccagacc ccgagttctc
ggtgcgctca
      241 gcggccgccg acgctaggag gccgcgctcc gcccccgcta ccatgaggcc
ccggaaagcc
      301 ttcctgctcc tgctgctctt ggggctggtg cagctgctgg ccgtggcggg
tgccgagggc
      361 ccggacgagg attcttctaa cagagaaaat gccattgagg atgaagagga
ggaggaggag
      421 gaagatgatg atgaggaaga agacgacttg gaagttaagg aagaaaatgg
agtcttggtc
      481 ctaaatgatg caaactttga taattttgtg gctgacaaag acacagtgct
gctggagttt
      541 tatgctccat ggtgtggaca ttgcaagcag tttgctccgg aatatgaaaa
aattgccaac
      601 atattaaagg ataaagatcc tcccattcct gttgccaaga tcgatgcaac
ctcagcgtct
      661 gtgctggcca gcaggtttga tgtgagtggc taccccacca tcaagatcct
taagaagggg
      721 caggctgtag actacgaggg ctccagaacc caggaagaaa ttgttgccaa
ggtcagagaa
      781 gtctcccagc ccgactggac gcctccacca gaagtcacgc ttgtgttgac
caaagagaac
      841 tttgatgaag ttgtgaatga tgcagatatc attctggtgg agttttatgc
cccatggtgt
      901 ggacactgca agaaacttgc ccccgagtat gagaaggccg ccaaggagct
cagcaagcgt
      961 tctcctccaa ttcccctggc aaaggtcgac gccaccgcag aaacagacct
ggccaagagg
```

1021 tttgatgtct ctggctatcc caccctgaaa attttccgca aaggaaggcc
ttatgactac
1081 aacggcccac gagaaaaata tggaatcgtt gattacatga tcgagcagtc
cgggcctccc
1141 tccaaggaga ttctgaccct gaagcaggtc caggagttcc tgaaggatgg
agacgatgtc
1201 atcatcatcg gggtctttaa gggggagagt gacccagcct accagcaata
ccaggatgcc
1261 gctaacaacc tgagagaaga ttacaaattt caccacactt tcagcacaga
aatagcaaag
1321 ttcttgaaag tctcccaggg gcagttggtt gtaatgcagc ctgagaaatt
ccagtccaag
1381 tatgagcccc ggagccacat gatggacgtc cagggctcca cccaggactc
ggccatcaag
1441 gacttcgtgc tgaagtacgc cctgcccctg gttggccacc gcaaggtgtc
aaacgatgct
1501 aagcgctaca ccaggcgccc cctggtggtc gtctactaca gtgtggactt
cagctttgat
1561 tacagagctg caactcagtt ttggcggagc aaagtcctag aggtggccaa
ggacttccct
1621 gagtacacct ttgccattgc ggacgaagag gactatgctg gggaggtgaa
ggacctgggg
1681 ctcagcgaga gtggggagga tgtcaatgcc gccatcctgg acgagagtgg
gaagaagttc
1741 gccatggagc cagaggagtt tgactctgac accctccgcg agtttgtcac
tgctttcaaa
1801 aaaggaaaac tgaagccagt catcaaatcc cagccagtgc ccaagaacaa
caagggaccc
1861 gtcaaggtcg tggtgggaaa gacctttgac tccattgtga tggaccccaa
gaaggacgtc
1921 ctcatcgagt tctacgcgcc atggtgcggg cactgcaagc agctagagcc
cgtgtacaac
1981 agcctggcca agaagtacaa gggccaaaag ggcctggtca tcgccaagat
ggacgccact
2041 gccaacgacg tccccagcga ccgctataag gtggagggct tccccaccat
ctacttcgcc
2101 cccagtgggg acaaaaagaa cccagttaaa tttgagggtg gagacagaga
tctggagcat
2161 ttgagcaagt ttatagaaga acatgccaca aaactgagca ggaccaagga
agagctttga
2221 aggcctgagg tctgcggaag gtgggaggag gcagacgccc tgcgtggccc
atggtcgggg
2281 cgtccacgcc gaggccggca acaaacgaca gtatctcgga ttccttttt
ttttttttta
2341 atttttata ctttggtgtt tcacttcatg ctctgaatac tgaataacca
tgaatgactg
2401 aatagtttag tccagatttt tacagaggat acatctattt ttatcattat
ttggggtttg
2461 aaaaattttt ttttacacct tctaatttct ttatttctca aagcagataa
ttcttctgtg
2521 tgaaaatgtt ttctttttt aatttaaggt ttaaaattcc ttttccaaat
catgttgatt
2581 ttgctctttg ctttttcgtt gtctgagaaa ttgttggcgt agatttggct
tctggtatgt
2641 gtttctgatt gcttcctgtt gagcacaaag tgagagctgc cactgagcag
ccctgccagg
2701 ggtgctgttt caggctgggc atcgccaggc ggcctccctg caaaccaagg
gctgggggca
2761 aaggggcatg atccagggtc ccccagggtg ggctcagctc cagggagagg
ccacccacgt

```
     2821 ggcagcccca cctcttgaga gcccccagtg ccggagcaga aaggaccctg
gacccagagg
     2881 cagatactgc ggggtggtag aaaaggtaga gtaggctgtg gcaatggaat
aaaacacgat
      2941 taaaaacgtt aaaaaaaaaa aaaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_004902.1

LOCUS NP_004902

ACCESSION NP_004902

```
      1 mrprkaflll lllglvqlla vagaegpded ssnrenaied eeeeeeeddd
eeeddlevke
       61 engvlvlnda nfdnfvadkd tvllefyapw cghckqfape yekianilkd
kdppipvaki
      121 datsasvlas rfdvsgypti kilkkgqavd yegsrtqeei vakvrevsqp
dwtpppevtl
      181 vltkenfdev vndadiilve fyapwcghck klapeyekaa kelskrsppi
plakvdatae
      241 tdlakrfdvs gyptlkifrk grpydyngpr ekygivdymi eqsgppskei
ltlkqvqefl
      301 kdgddviiig vfkgesdpay qqyqdaannl redykfhhtf steiakflkv
sqgqlvvmqp
      361 ekfqskyepr shmmdvqgst qdsaikdfvl kyalplvghr kvsndakryt
rrplvvvyys
      421 vdfsfdyraa tqfwrskvle vakdfpeytf aiadeedyag evkdlglses
gedvnaaild
      481 esgkkfamep eefdsdtlre fvtafkkgkl kpviksqpvp knnkgpvkvv
vgktfdsivm
      541 dpkkdvlief yapwcghckq lepvynslak kykgqkglvi akmdatandv
psdrykvegf
      601 ptiyfapsgd kknpvkfegg drdlehlskf ieehatklsr tkeel
```

PDIA1

Official Symbol: P4HB

Official Name: prolyl 4-hydroxylase, beta polypeptide

Gene ID: 5034

Organism: Homo sapiens

Other Aliases: DSI, ERBA2L, GIT, P4Hbeta, PDI, **PDIA1**, PHDB, PO4DB, PO4HB, PROHB

Other Designations: cellular thyroid hormone-binding protein; collagen prolyl 4-hydroxylase beta; glutathione-insulin transhydrogenase; p55; procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide; prolyl 4-hydroxylase subunit beta; protein disulfide isomerase family A, member 1; protein disulfide isomerase-associated 1; protein disulfide

isomerase/oxidoreductase; protein disulfide-isomerase; protocollagen hydroxylase; thyroid hormone-binding protein p55

Nucleotide sequence:

NCBI Reference Sequence: NM_000918.3

LOCUS NM_000918

ACCESSION NM_000918

```
      1 gagcctcgaa gtccgccggc caatcgaagg cgggccccag cggcgcgtgc
gcgccgcggc
       61 cagcgcgcgc gggcgggggg gcaggcgcgc cccggaccca ggatttataa
aggcgaggcc
      121 gggaccggcg cgcgctctcg tcgcccccgc tgtcccggcg gcgccaaccg
aagcgccccg
      181 cctgatccgt gtccgacatg ctgcgccgcg ctctgctgtg cctggccgtg
gccgccctgg
      241 tgcgcgccga cgcccccgag gaggaggacc acgtcctggt gctgcggaaa
agcaacttcg
      301 cggaggcgct ggcggcccac aagtacctgc tggtggagtt ctatgccccc
tggtgtggcc
      361 actgcaaggc tctggccccc gagtatgcca aagccgctgg gaagctgaag
gcagaaggtt
      421 ccgagatcag gttggccaag gtggacgcca cggaggagtc tgacctggcc
cagcagtacg
      481 gcgtgcgcgg ctatcccacc atcaagttct tcaggaatgg agacacggct
tcccccaagg
      541 aatatacagc tggcagagag gctgatgaca tcgtgaactg gctgaagaag
cgcacgggcc
      601 cggctgccac caccctgcct gacggcgcag ctgcagagtc cttggtggag
tccagcgagg
      661 tggctgtcat cggcttcttc aaggacgtgg agtcggactc tgccaagcag
tttttgcagg
      721 cagcagaggc catcgatgac ataccatttg ggatcacttc caacagtgac
gtgttctcca
      781 aataccagct cgacaaagat ggggttgtcc tctttaagaa gtttgatgaa
ggccggaaca
      841 actttgaagg ggaggtcacc aaggagaacc tgctggactt tatcaaacac
aaccagctgc
      901 cccttgtcat cgagttcacc gagcagacag ccccgaagat ttttggaggt
gaaatcaaga
      961 ctcacatcct gctgttcttg cccaagagtg tgtctgacta tgacggcaaa
ctgagcaact
     1021 tcaaaacagc agccgagagc ttcaagggca agatcctgtt catcttcatc
gacagcgacc
     1081 acaccgacaa ccagcgcatc ctcgagttct ttggcctgaa gaaggaagag
tgcccggccg
     1141 tgcgcctcat caccctggag gaggagatga ccaagtacaa gcccgaatcg
gaggagctga
     1201 cggcagagag gatcacagag ttctgccacc gcttcctgga gggcaaaatc
aagccccacc
     1261 tgatgagcca ggagctgccg gaggactggg acaagcagcc tgtcaaggtg
cttgttggga
     1321 agaactttga agacgtggct tttgatgaga aaaaaaacgt ctttgtggag
ttctatgccc
     1381 catggtgtgg tcactgcaaa cagttggctc ccatttggga taaactggga
gagacgtaca
```

```
     1441 aggaccatga gaacatcgtc atcgccaaga tggactcgac tgccaacgag
gtggaggccg
     1501 tcaaagtgca cagcttcccc acactcaagt tctttcctgc cagtgccgac
aggacggtca
     1561 ttgattacaa cggggaacgc acgctggatg gttttaagaa attcctggag
agcggtggcc
     1621 aggatggggc aggggatgat gacgatctcg aggacctgga agaagcagag
gagccagaca
     1681 tggaggaaga cgatgatcag aaagctgtga aagatgaact gtaatacgca
aagccagacc
     1741 cgggcgctgc cgagacccct cgggggctgc acacccagca gcagcgcacg
cctccgaagc
     1801 ctgcggcctc gcttgaagga gggcgtcgcc ggaaacccag ggaacctctc
tgaagtgaca
     1861 cctcaccccc acacaccgtc cgttcacccc cgtctcttcc ttctgctttt
cggtttttgg
     1921 aaagggatcc atctccaggc agcccaccct ggtggggctt gtttcctgaa
accatgatgt
     1981 acttttttcat acatgagtct gtccagagtg cttgctaccg tgttcggagt
ctcgctgcct
     2041 ccctcccgcg ggaggtttct cctctttttg aaaattccgt ctgtgggatt
tttagacatt
     2101 tttcgacatc agggtatttg ttccaccttg gccaggcctc ctcggagaag
cttgtccccc
     2161 gtgtgggagg gacggagccg gactggacat ggtcactcag taccgcctgc
agtgtcgcca
     2221 tgactgatca tggctcttgc atttttgggt aaatggagac ttccggatcc
tgtcagggtg
     2281 tcccccatgc ctggaagagg agctggtggc tgccagccct ggggcccggc
acaggcctgg
     2341 gccttccccc tccctcaagc cagggctcct cctcctgtcg tgggctcatt
gtgaccactg
     2401 gcctctctac agcacggcct gtggcctgtt caaggcagaa ccacgaccct
tgactcccgg
     2461 gtggggaggt ggccaaggat gctggagctg aatcagacgc tgacagttct
tcaggcattt
     2521 ctatttcaca atcgaattga acacattggc caaataaagt tgaaatttta
ccacctgtaa
     2581 aaaaaaaaaa aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_000909.2

LOCUS NP_000909

ACCESSION NP_000909

```
      1 mlrrallcla vaalvradap eeedhvlvlr ksnfaealaa hkyllvefya
pwcghckala
       61 peyakaagkl kaegseirla kvdateesdl aqqygvrgyp tikffrngdt
aspkeytagr
      121 eaddivnwlk krtgpaattl pdgaaaeslv essevavigf fkdvesdsak
qflqaaeaid
      181 dipfgitsns dvfskyqldk dgvvlfkkfd egrnnfegev tkenlldfik
hnqlplvief
      241 teqtapkifg geikthillf lpksvsdydg klsnfktaae sfkgkilfif
idsdhtdnqr
      301 ileffglkke ecpavrlitl eeemtkykpe seeltaerit efchrflegk
ikphlmsqel
```

361 pedwdkqpvk vlvgknfedv afdekknvfv efyapwcghc kqlapiwdkl getykdheni
421 viakmdstan eveavkvhsf ptlkffpasa drtvidynge rtldgfkkfl esggqdgagd
481 dddledleea eepdmeeddd qkavkdel

CA2D1

Official Symbol: CACNA2D1

Official Name: calcium channel, voltage-dependent, alpha 2/delta subunit 1

Gene ID: 781

Organism: Homo sapiens

Other Aliases: H_DJ0560O14.1, CACNA2, CACNL2A, CCHL2A

Other Designations: calcium channel, L type, alpha 2 polypeptide; dihydropyridine-sensitive L-type, calcium channel alpha-2/delta subunit; voltage-dependent calcium channel subunit alpha-2/delta-1; voltage-gated calcium channel subunit alpha-2/delta-1

Nucleotide sequence:

NCBI Reference Sequence: NM_000722.2

LOCUS NM_000722

ACCESSION NM_000722

1 cggcggaggc aaggcggccg cggcgcggag cagccgacgc acgctagtgg gtccgcccgc
61 caccgcccct tcctcggcgt ccgctcccgc ccttgccgtc ccccgcgcgg ctccgcgcct
121 cgggccccgg gcgcagccag ccctccagac gcccgcggtc ccggcggcgt gtgctgctct
181 tcctccgccc gcggtttcca gcgccgctcc ttcccccgct tgggcaggga gggggcattg
241 atcttcgatc gcgaagatgg ctgctggctg cctgctggcc ttgactctga cacttttcca
301 atctttgctc atcggcccct cgtcggagga gccgttccct tcggccgtca ctatcaaatc
361 atgggtggat aagatgcaag aagaccttgt cacactggca aaaacagcaa gtggagtcaa
421 tcagcttgtt gatatttatg agaaatatca agatttgtat actgtggaac caaataatgc
481 acgccagctg gtagaaattg cagccaggga tattgagaaa cttctgagca acagatctaa
541 agccctggtg cgcctggcat tggaagcgga gaaagttcaa gcagctcacc agtggagaga
601 agattttgca agcaatgaag ttgtctacta caatgcaaag gatgatctcg atcctgagaa
661 aaatgacagt gagccaggca gccagaggat aaaacctgtt ttcattgaag atgctaattt

```
      721 tggacgacaa atatcttatc agcacgcagc agtccatatt cctactgaca
tctatgaggg
      781 ctcaacaatt gtgttaaatg aactcaactg gacaagtgcc ttagatgaag
ttttcaaaaa
      841 gaatcgcgag gaagaccctt cattattgtg gcaggttttt ggcagtgcca
ctggcctagc
      901 tcgatattat ccagcttcac catgggttga taatagtaga actccaaata
agattgacct
      961 ttatgatgta cgcagaagac catggtacat ccaaggagct gcatctccta
aagacatgct
     1021 tattctggtg gatgtgagtg gaagtgttag tggattgaca cttaaactga
tccgaacatc
     1081 tgtctccgaa atgttagaaa ccctctcaga tgatgatttc gtgaatgtag
cttcatttaa
     1141 cagcaatgct caggatgtaa gctgttttca gcaccttgtc caagcaaatg
taagaaataa
     1201 aaaagtgttg aaagacgcgg tgaataatat cacagccaaa ggaattacag
attataagaa
     1261 gggctttagt tttgcttttg aacagctgct taattataat gtttccagag
caaactgcaa
     1321 taagattatt atgctattca cggatggagg agaagagaga gcccaggaga
tatttaacaa
     1381 atacaataaa gataaaaaag tacgtgtatt cacgttttca gttggtcaac
acaattatga
     1441 cagaggacct attcagtgga tggcctgtga aaacaaaggt tattattatg
aaattccttc
     1501 cattggtgca ataagaatca atactcagga atatttggat gttttgggaa
gaccaatggt
     1561 tttagcagga gacaaagcta agcaagtcca atggacaaat gtgtacctgg
atgcattgga
     1621 actgggactt gtcattactg gaactcttcc ggtcttcaac ataaccggcc
aatttgaaaa
     1681 taagacaaac ttaaagaacc agctgattct tggtgtgatg ggagtagatg
tgtctttgga
     1741 agatattaaa agactgacac cacgttttac actgtgcccc aatgggtatt
actttgcaat
     1801 cgatcctaat ggttatgttt tattacatcc aaatcttcag ccaaagaacc
ccaaatctca
     1861 ggagccagta acattggatt tccttgatgc agagttagag aatgatatta
aagtggagat
     1921 tcgaaataag atgattgatg gggaaagtgg agaaaaaaca ttcagaactc
tggttaaatc
     1981 tcaagatgag agatatattg acaaaggaaa caggacatac acatggacac
ctgtcaatgg
     2041 cacagattac agtttggcct tggtattacc aacctacagt ttttactata
taaaagccaa
     2101 actagaagag acaataactc aggccagatc aaaaaagggc aaaatgaagg
attcggaaac
     2161 cctgaagcca gataattttg aagaatctgg ctatacattc atagcaccaa
gagattactg
     2221 caatgacctg aaaatatcgg ataataacac tgaatttctt ttaaatttca
acgagtttat
     2281 tgatagaaaa actccaaaca acccatcatg taacgcggat ttgattaata
gagtcttgct
     2341 tgatgcaggc tttacaaatg aacttgtcca aaattactgg agtaagcaga
aaaatatcaa
     2401 gggagtgaaa gcacgatttg ttgtgactga tggtgggatt accagagttt
atcccaaaga
     2461 ggctggagaa aattggcaag aaaacccaga gacatatgag gacagcttct
ataaaaggag
```

```
     2521 cctagataat gataactatg ttttcactgc tccctacttt aacaaaagtg
gacctggtgc
     2581 ctatgaatcg ggcattatgg taagcaaagc tgtagaaata tatattcaag
ggaaacttct
     2641 taaacctgca gttgttggaa ttaaaattga tgtaaattcc tggatagaga
atttcaccaa
     2701 aacctcaatc agagatccgt gtgctggtcc agtttgtgac tgcaaaagaa
acagtgacgt
     2761 aatggattgt gtgattctgg atgatggtgg gtttcttctg atggcaaatc
atgatgatta
     2821 tactaatcag attggaagat tttttggaga gattgatccc agcttgatga
gacacctggt
     2881 taatatatca gtttatgctt ttaacaaatc ttatgattat cagtcagtat
gtgagcccgg
     2941 tgctgcacca aaacaaggag caggacatcg ctcagcatat gtgccatcag
tagcagacat
     3001 attacaaatt ggctggtggg ccactgctgc tgcctggtct attctacagc
agtttctctt
     3061 gagtttgacc tttccacgac tccttgaggc agttgagatg gaggatgatg
acttcacggc
     3121 ctccctgtcc aagcagagct gcattactga acaaacccag tatttcttcg
ataacgacag
     3181 taaatcattc agtggtgtat tagactgtgg aaactgttcc agaatctttc
atggagaaaa
     3241 gcttatgaac accaacttaa tattcataat ggttgagagc aaagggacat
gtccatgtga
     3301 cacacgactg ctcatacaag cggagcagac ttctgacggt ccaaatcctt
gtgacatggt
     3361 taagcaaccc agataccgaa aagggcctga tgtctgcttt gataacaatg
tcttggagga
     3421 ttatactgac tgtggtggtg tttctggatt aaatccctcc ctgtggtata
tcattggaat
     3481 ccagtttcta ctactttggc tggtatctgg cagcacacac cgcctgttat
gaccttctaa
     3541 aaaccaaatc tgcatagtta aactccagac cctgccaaaa catgagccct
gccctcaatt
     3601 acagtaacgt agggtcagct ataaaatcag acaaacatta gctgggcctg
ttccatggca
     3661 taacactaag gcgcagactc ctaaggcacc cactggctgc atgtcagggt
gtcagatcct
     3721 taaacgtgtg tgaatgctgc atcatctatg tgtaacatca aagcaaaatc
ctatacgtgt
     3781 cctctattgg aaaatttggg agtttgttgt tgcattgttg gt
```

Protein sequence:

NCBI Reference Sequence: NP_000713.2

LOCUS NP_000713

ACCESSION NP_000713

```
      1 maagcllalt ltlfqsllig psseepfpsa vtikswvdkm qedlvtlakt
asgvnqlvdi
       61 yekyqdlytv epnnarqlve iaardiekll snrskalvrl aleaekvqaa
hqwredfasn
      121 evvyynakdd ldpekndsep gsqrikpvfi edanfgrqis yqhaavhipt
diyegstivl
```

```
      181 nelnwtsald evfkknreed psllwqvfgs atglaryypa spwvdnsrtp
nkidlydvrr
      241 rpwyiqgaas pkdmlilvdv sgsvsgltlk lirtsvseml etlsdddfvn
vasfnsnaqd
      301 vscfqhlvqa nvrnkkvlkd avnnitakgi tdykkgfsfa feqllnynvs
rancnkiiml
      361 ftdggeeraq eifnkynkdk kvrvftfsvg qhnydrgpiq wmacenkgyy
yeipsigair
      421 intqeyldvl grpmvlagdk akqvqwtnvy ldalelglvi tgtlpvfnit
gqfenktnlk
      481 nqlilgvmgv dvsledikrl tprftlcpng yyfaidpngy vllhpnlqpk
npksqepvtl
      541 dfldaelend ikveirnkmi dgesgektfr tlvksqdery idkgnrtytw
tpvngtdysl
      601 alvlptysfy yikakleeti tqarskkgkm kdsetlkpdn feesgytfia
prdycndlki
      661 sdnnteflln fnefidrktp nnpscnadli nrvlldagft nelvqnywsk
qknikgvkar
      721 fvvtdggitr vypkeagenw qenpetyeds fykrsldndn yvftapyfnk
sgpgayesgi
      781 mvskaveiyi qgkllkpavv gikidvnswi enftktsird pcagpvcdck
rnsdvmdcvi
      841 lddggfllma nhddytnqig rffgeidpsl mrhlvnisvy afnksydyqs
vcepgaapkq
      901 gaghrsayvp svadilqigw wataaawsil qqfllsltfp rlleavemed
ddftaslskq
      961 sciteqtqyf fdndsksfsg vldcgncsri fhgeklmntn lifimveskg
tcpcdtrlli
     1021 qaeqtsdgpn pcdmvkqpry rkgpdvcfdn nvledytdcg gvsglnpslw
yiigiqflll
     1081 wlvsgsthrl l
```

GPAT1

Official Symbol: GPAM

Official Name: glycerol-3-phosphate acyltransferase, mitochondrial

Gene ID: 57678

Organism: Homo sapiens

Other Aliases: RP11-426E5.2, GPAT, GPAT1

Other Designations: GPAT-1; glycerol 3-phosphate acyltransferase, mitochondrial; glycerol-3-phosphate acyltransferase 1, mitochondrial

Nucleotide sequence:

NCBI Reference Sequence: NM_001244949.

LOCUS NM_001244949

ACCESSION NM_001244949

```
      1 tgcgtcatca gggtgcgcca ctgcagctgg cattggccgg gactggaagt
gcgggcttct
       61 gcagcagccg aagctggagc tgctagggca gcagcggctc ccctgttgta
tggacattct
      121 gcacccgaaa ctgatagctg agtcctgaag ttttatgtta tgaaacagaa
gaactttcat
      181 cccagcacat gatttgggaa ttacactttg tgacatggat gaatctgcac
tgacccttgg
      241 tacaatagat gtttcttatc tgccacattc atcagaatac agtgttggtc
gatgtaagca
      301 cacaagtgag gaatggggtg agtgtggctt tagacccacc atcttcagat
ctgcaacttt
      361 aaaatggaaa gaaagcctaa tgagtcggaa aaggccattt gttggaagat
gttgttactc
      421 ctgcactccc cagagctggg acaaattttt caaccccagt atcccgtctt
tgggtttgcg
      481 gaatgttatt tatatcaatg aaactcacac aagacaccgc ggatggcttg
caagacgcct
      541 ttcttacgtt ctttttattc aagagcgaga tgtgcataag ggcatgtttg
ccaccaatgt
      601 gactgaaaat gtgctgaaca gcagtagagt acaagaggca attgcagaag
tggctgctga
      661 attaaaccct gatggttctg cccagcagca atcaaaagcc gttaacaaag
tgaaaaagaa
      721 agctaaaagg attcttcaag aaatggttgc cactgtctca ccggcaatga
tcagactgac
      781 tgggtgggtg ctgctaaaac tgttcaacag cttcttttgg aacattcaaa
ttcacaaagg
      841 tcaacttgag atggttaaag ctgcaactga gacgaatttg ccgcttctgt
ttctaccagt
      901 tcatagatcc catattgact atctgctgct cactttcatt ctcttctgcc
ataacatcaa
      961 agcaccatac attgcttcag gcaataatct caacatccca atcttcagta
ccttgatcca
     1021 taagcttggg ggcttcttca tacgacgaag gctcgatgaa acaccagatg
gacggaaaga
     1081 tgttctctat agagctttgc tccatgggca tatagttgaa ttacttcgac
agcagcaatt
     1141 cttggagatc ttcctggaag gcacacgttc taggagtgga aaaacctctt
gtgctcgggc
     1201 aggacttttg tcagttgtgg tagatactct gtctaccaat gtcatcccag
acatcttgat
     1261 aatacctgtt ggaatctcct atgatcgcat tatcgaaggt cactacaatg
gtgaacaact
     1321 gggcaaacct aagaagaatg agagcctgtg gagtgtagca agaggtgtta
ttagaatgtt
     1381 acgaaaaaac tatggttgtg tccgagtgga ttttgcacag ccattttcct
taaaggaata
     1441 tttagaaagc caaagtcaga aaccggtgtc tgctctactt tccctggagc
aagcgttgtt
     1501 accagctata cttccttcaa gacccagtga tgctgctgat gaaggtagag
acacgtccat
     1561 taatgagtcc agaaatgcaa cagatgaatc cctacgaagg aggttgattg
caaatctggc
     1621 tgagcatatt ctattcactg ctagcaagtc ctgtgccatt atgtccacac
acattgtggc
     1681 ttgcctgctc ctctacagac acaggcaggg aattgatctc tccacattgg
tcgaagactt
     1741 ctttgtgatg aaagaggaag tcctggctcg tgattttgac ctggggttct
caggaaattc
```

1801 agaagatgta gtaatgcatg ccatacagct gctgggaaat tgtgtcacaa
tcacccacac
1861 tagcaggaac gatgagtttt ttatcacccc cagcacaact gtcccatcag
tcttcgaact
1921 caacttctac agcaatgggg tacttcatgt ctttatcatg gaggccatca
tagcttgcag
1981 cctttatgca gttctgaaca agaggggact ggggggtccc actagcaccc
cacctaacct
2041 gatcagccag gagcagctgg tgcggaaggc ggccagcctg tgctaccttc
tctccaatga
2101 aggcaccatc tcactgcctt gccagacatt ttaccaagtc tgccatgaaa
cagtaggaaa
2161 gtttatccag tatggcattc ttacagtggc agagcacgat gaccaggaag
atatcagtcc
2221 tagtcttgct gagcagcagt gggacaagaa gcttccagaa cctttgtctt
ggagaagtga
2281 tgaagaagat gaagacagtg actttgggga ggaacagcga gattgctacc
tgaaggtgag
2341 ccaatccaag gagcaccagc agtttatcac cttcttacag agactccttg
ggcctttgct
2401 ggaggcctac agctctgctg ccatctttgt tcacaacttc agtggtcctg
ttccagaacc
2461 tgagtatctg caaaagttgc acaaatacct aataaccaga acagaaagaa
atgttgcagt
2521 atatgctgag agtgccacat attgtcttgt gaagaatgct gtgaaaatgt
ttaaggatat
2581 tggggttttc aaggagacca aacaaaagag agtgtctgtt ttagaactga
gcagcacttt
2641 tctacctcaa tgcaaccgac aaaaacttct agaatatatt ctgagttttg
tggtgctgta
2701 ggtaacgtgt ggcactgctg gcaaatgaag gtcatgagat gagttccttg
taggtaccag
2761 cttctggctc aagagttgaa ggtgccatcg cagggtcagg cctgccctgt
cccgaagtga
2821 tctcctggaa gacaagtgcc ttctccctcc atggatctgt gatcttccca
gctctgcatc
2881 aacacagcag cctgcagata acacttgggg ggacctcagc ctctattcgc
aactcataat
2941 ccgtagacta caagatgaaa tctcaataaa ttatttttga gtttattaaa
gattgacatt
3001 ttaagtacaa cttttaagga ctaattactg tgatggacac agaaatgtag
ctgtgttctg
3061 gaactgaatc ttacatggta tacttagtgc tgctgggtaa tttgttggta
tattatctgg
3121 ttagtggtta atgcttcctt taaaaataat tgagtcatcc attcactctt
tttcagtttt
3181 atctgtcaat agtagctaca tttttaatgg gagcaccttt tatcccaaag
tgctttataa
3241 attgagtgga ctgatatata tcacacccag gtatcactgt gctgtccttt
gctgtcagat
3301 ttagaaatgt ttttaagagc tatgtgaaaa cagacaatat tagtttaggt
cgggaactga
3361 gatattgtaa tcaaatagtt aacatcagga agttaatttg gctggcaaaa
ttctagggaa
3421 acttggccag aaaactggtg ttgaaggctt ttgctcatat aaacaagtgc
cattgagttt
3481 caaatgacca gcaaatatat ttagaaccct tcctgtttta tgtctgtacc
tcgtccaccc
3541 ctcaggtaat acctgcctct cacaggtaca gctgtttctt ggaaatcctc
caaccaaata

3601 gcagttttcc taacttgatt agcttgagct gacagactgt tagaatacag
ttctctggcc
3661 acagctgatg agggctttct gtactgcaca cagattgtgt actgcacccc
agtccaggtg
3721 actggtaccc actcgagttg tgccgtgcac aacctgtcca gtatatgcat
gtggtggccc
3781 tactgactgg taatggttag aggcatttat ggatttttag ctttgaggaa
aaaccatgac
3841 ttttaacaaa tttttatggg ttatatgcct aaaacccttat gccacatagt
ggtaaataat
3901 tatgaaaaat ggtctgttca taattggtag gtgccttttg tgagcaggga
gcataattat
3961 tggtttatta tggtaattat ggtgattttt taaatatcat gtaatgttaa
aacgttttct
4021 aacagtttac tgttgcttat ctccaagata ttatggaatt aagaattttt
ccagatgagt
4081 gttacataga ttctttgaat ttagtataaa agtactgaga attaagtttg
tacttccata
4141 agcttggatt ttaaacactg atagtatctc atgagtaatg tgtgttttgg
gagagggagg
4201 gatgctgatt gatatttcac attgtatgaa ataccatgtt tgaaactcat
agcaataatg
4261 ctatgctgtt gtgatccctc tcaagttctg catttaaaat atatttttc
tttataggaa
4321 ttgatgtata ccatgaagtc attgtcagtt gtagtagctc tgatgttgaa
tgagatatca
4381 tgttttagca ttccatttta ctgactaggg tagaagaaca cttttcttgg
ctacatttgg
4441 aggataccca gggagtcttg ggtgttcctt atctggggaa gcaaacattt
cactagtctc
4501 ttttttcat cctttaaatt gtaaattaag gattactcaa gctcaccatt
attcaagatt
4561 gggactcgct tcccagtcga cactctgccc tgcctgtcat tgctgcaaag
agctgctgct
4621 ttgccaacct aagcaaagaa aatacggctt ctcttgcatt attttccctt
ttggttggtt
4681 tgttttctag aagtacgttc agatgctttg gggaatgcaa tgtatgattt
gctagctctc
4741 tcaccactta actcactgtg aggataaata tgcatgcttt ttgtaattaa
ctggtgcttt
4801 gaaaatcttt tttaagggag aaaaatctca accaaagtta tgctcatcca
gacaagctga
4861 cctttgagtt aatttcagca caactcattc ttcagtgcct catgactgaa
aacaaaaaac
4921 aaaaaaacga aagcatcttc acaatgaagc ttccagatag caccgttttg
ctaaaagata
4981 cattctcatt gttttccaac agtgatggct tccacataag gttaaacaaa
ctaggtgctt
5041 gtaaataatt tattacagtt tactctatcg catttctgta acatgaaatg
catgcccttc
5101 ttcaggggaa gactgtggtc aagttaaaaa aaaaaaacaa tattaaacaa
catgaaactg
5161 cagtctgttt ttgaaaatga gaatgtccta agtgattcag aagagaggag
ggaagttgtg
5221 cactctgaaa atgcatgaaa aacaaaggca aaaactagtg ggaaatgtgt
agaactgtta
5281 actgagatgg cttcgagtct tccttctgga atctgttaaa tttcacaaag
tcatgagggt
5341 aaatggagaa aatatttctg ggattacaat gaatgtaagc ccaaattgtg
gaattgccag

```
     5401 taacctggat ggggaaaagc atttcccata gcactccatg taatatgagt
gctctgtgag
     5461 atgttcatca gtgttttata gaaatggtgt tgctgggaaa ccaagtttgc
acctggaaac
     5521 ttacaatgca ctttagcgca gtaagggctt ggcatccggt agtgaaaaac
tgtctaaccc
     5581 agcattgccc aaactatttt gacaccagga ccttttctc ctttgggata
cttatgaacc
     5641 tctcactaat gtcctgtgga gaacattttg ggaaacacta tgttagatag
ttctttaagg
     5701 agacaaaacg gtaatgaaca gatagcactg gggcagaata tgcatgcatt
ttgtaacgtc
     5761 cagtgtggcg ttgaatagat gtgtatttcc tcccctgcag aaaataagca
cagaaaatta
     5821 taatgtaggt gatcggagct ctttcctttg atagagagaa cagccccaat
gatcctggct
     5881 ttttcactga acgtatcaga atacatggat gaattggggt aaataaggtt
ttaattcaga
     5941 tctagaagaa agtattgtac gtttgaatgc agatttttat ccacagatag
ttgtagtgtt
     6001 tagacatgac aggacctatc gttgaggttt ctaagactta ctatgggctg
taaacctgtt
     6061 ttttaaaact attttagaaa cctgagactt gccgtctggc attttagttt
aatacaaact
     6121 aatgattgca tttgaaagag attcttgacc ttatttctaa acgtctagag
ctctgaaatg
     6181 tcttgatgga aggtattaaa ctatttgcct gttgtacaaa gaaatgttaa
gactcgtgaa
     6241 aagaattact ataaggtact gtgaaataac tgcgattttg tgagcaaaac
atacttggaa
     6301 atgctgattg atttttatgc ttgttagtgt attgcaagaa acacagaaaa
tgtagttttg
     6361 ttttaataaa ccaaaaattg aacatacaaa aaaaaaaaaa aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001231878.1

LOCUS NP_001231878

ACCESSION NP_001231878

```
      1 mdesaltlgt idvsylphss eysvgrckht seewgecgfr ptifrsatlk
wkeslmsrkr
       61 pfvgrccysc tpqswdkffn psipslglrn viyinethtr hrgwlarrls
yvlfiqerdv
      121 hkgmfatnvt envlnssrvq eaiaevaael npdgsaqqqs kavnkvkkka
krilqemvat
      181 vspamirltg wvllklfnsf fwniqihkgq lemvkaatet nlpllflpvh
rshidylllt
      241 filfchnika pyiasgnnln ipifstlihk lggffirrrl detpdgrkdv
lyrallhghi
      301 vellrqqqfl eiflegtrsr sgktscarag llsvvvdtls tnvipdilii
pvgisydrii
      361 eghyngeqlg kpkkneslws vargvirmlr knygcvrvdf aqpfslkeyl
esqsqkpvsa
      421 llsleqallp ailpsrpsda adegrdtsin esrnatdesl rrrlianlae
hilftasksc
```

```
      481 aimsthivac lllyrhrqgi dlstlvedff vmkeevlard fdlgfsgnse
dvvmhaiqll
      541 gncvtitht rndeffitps ttvpsvfeln fysngvlhvf imeaiiacsl
yavlnkrglg
      601 gptstppnli sqeqlvrkaa slcyllsneg tislpcqtfy qvchetvgkf
iqygiltvae
      661 hddqedisps laeqqwdkkl peplswrsde ededsdfgee qrdcylkvsq
skehqqfitf
      721 lqrllgplle ayssaaifvh nfsgpvpepe ylqklhkyli trternvavy
aesatyclvk
      781 navkmfkdig vfketkqkrv svlelsstfl pqcnrqklle yilsfvvl
```

TAZ

Official Symbol: TAZ

Official Name: tafazzin

Gene ID: 6901

Organism: Homo sapiens

Other Aliases: XX-FW83563B9.3, BTHS, CMD3A, EFE, EFE2, G4.5, LVNCX, Taz1

Other Designations: protein G4.5

Nucleotide sequence:

NCBI Reference Sequence: NM_000116.3

LOCUS NM_000116

ACCESSION NM_000116

```
      1 tttccggcgg ttgcaccggg ccggggtgcc agcgcccgcc ttcccgtttc
ctcccgttcc
       61 gcagcgcgcc cacggcctgt gaccccggcg accgctcccc agtgacgaga
gagcggggcc
      121 gggcgctgct ccggcctgac ctgcgaaggg acctcggtcc agtcccctgt
tgcgccgcgc
      181 cccctgtccg tccgtgcgcg ggccagtcag gggccagtgt ctcgagcggt
cgaggtcgca
      241 gacctagagg cgccccacag gccggcccgg ggcgctggga gcgccggccg
cgggccgggt
      301 ggggatgcct ctgcacgtga agtggccgtt ccccgcggtg ccgccgctca
cctggaccct
      361 ggccagcagc gtcgtcatgg gcttggtggg cacctacagc tgcttctgga
ccaagtacat
      421 gaaccacctg accgtgcaca acagggaggt gctgtacgag ctcatcgaga
agcgaggccc
      481 ggccacgccc ctcatcaccg tgtccaatca ccagtcctgc atggacgacc
ctcatctctg
      541 ggggatcctg aaactccgcc acatctggaa cctgaagttg atgcgttgga
cccctgcagc
```

```
      601 tgcagacatc tgcttcacca aggagctaca ctcccacttc ttcagcttgg
gcaagtgtgt
      661 gcctgtgtgc cgaggagcag aatttttcca agcagagaat gaggggaaag
gtgttctaga
      721 cacaggcagg cacatgccag gtgctggaaa aagaagagag aaaggagatg
gcgtctacca
      781 gaaggggatg gacttcattt tggagaagct caaccatggg gactgggtgc
atatcttccc
      841 agaagggaaa gtgaacatga gttccgaatt cctgcgtttc aagtggggaa
tcgggcgcct
      901 gattgctgag tgtcatctca accccatcat cctgcccctg tggcatgtcg
gaatgaatga
      961 cgtccttcct aacagtccgc cctacttccc ccgctttgga cagaaaatca
ctgtgctgat
     1021 cgggaagccc ttcagtgccc tgcctgtact cgagcggctc cgggcggaga
acaagtcggc
     1081 tgtggagatg cggaaagccc tgacggactt cattcaagag gaattccagc
atctgaagac
     1141 tcaggcagag cagctccaca accacctcca gcctgggaga taggccttgc
ttgctgcctt
     1201 ctggattctt ggcccgcaca gagctggggc tgagggatgg actgatgctt
ttagctcaaa
     1261 cgtggctttt agacagattt gttcatagac cctctcaagt gccctctccg
agctggtagg
     1321 cattccagct cctccgtgct tcctcagtta cacaaaggac ctcagctgct
tctcccactt
     1381 ggccaagcag ggaggaagaa gcttaggcag ggctctcttt ccttcttgcc
ttcagatgtt
     1441 ctctcccagg ggctggcttc aggagggagc atagaaggca ggtgagcaac
cagttggcta
     1501 ggggagcagg gggcccacca gagctgtgga gaggggaccc taagactcct
cggcctggct
     1561 cctacccacc gcccttgccg aaccaggagc tgctcactac ctcctcaggg
atggccgttg
     1621 gccacgtctt ccttctgcct gagcttcccc ccgaccacag gccctttcct
caggcaaggt
     1681 ctggcctcag gtgggccgca ggcgggaaaa gcagcccttg gccagaagtc
aagcccagcc
     1741 acgtggagcc tagagtgagg gcctgaggtc tggctgcttg cccccatgct
ggcgccaaca
     1801 acttctccat cctttctgcc tctcaacatc acttgaatcc tagggcctgg
gttttcatgt
     1861 ttttgaaaca gaaccataaa gcatatgtgt tggcttgttg taaaaaaaaa
aaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_000107.1

LOCUS NP_000107

ACCESSION NP_000107

```
      1 mplhvkwpfp avppltwtla ssvvmglvgt yscfwtkymn hltvhnrevl
yeliekrgpa
       61 tplitvsnhq scmddphlwg ilklrhiwnl klmrwtpaaa dicftkelhs
hffslgkcvp
      121 vcrgaeffqa enegkgvldt grhmpgagkr rekgdgvyqk gmdfilekln
hgdwvhifpe
```

```
      181 gkvnmssefl rfkwgigrli aechlnpiil plwhvgmndv lpnsppyfpr
fgqkitvlig
      241 kpfsalpvle rlraenksav emrkaltdfi qeefqhlktq aeqlhnhlqp gr
```

CO1A2

Official Symbol: COL1A2

Official Name: collagen, type I, alpha 2

Gene ID: 1278

Organism: Homo sapiens

Other Aliases: OI4

Other Designations: alpha 2(I)-collagen; alpha-2 type I collagen; collagen I, alpha-2 polypeptide; collagen alpha-2(I) chain; collagen of skin, tendon and bone, alpha-2 chain; type I procollagen

Nucleotide sequence:

NCBI Reference Sequence: NM_000089.3

LOCUS NM_000089

ACCESSION NM_000089

```
      1 gtgtcccata gtgtttccaa acttggaaag ggcgggggag ggcgggagga
tgcggagggc
       61 ggaggtatgc agacaacgag tcagagtttc cccttgaaag cctcaaaagt
gtccacgtcc
      121 tcaaaaagaa tggaaccaat ttaagaagcc agccccgtgg ccacgtccct
tcccccattc
      181 gctccctcct ctgcgccccc gcaggctcct cccagctgtg gctgcccggg
cccccagccc
      241 cagccctccc attggtggag gccctttgg aggcacccta gggccaggga
aacttttgcc
      301 gtataaatag ggcagatccg ggctttatta ttttagcacc acggcagcag
gaggtttcgg
      361 ctaagttgga ggtactggcc acgactgcat gcccgcgccc gccaggtgat
acctccgccg
      421 gtgacccagg ggctctgcga cacaaggagt ctgcatgtct aagtgctaga
catgctcagc
      481 tttgtggata cgcggacttt gttgctgctt gcagtaacct tatgcctagc
aacatgccaa
      541 tctttacaag aggaaactgt aagaaagggc ccagccggag atagaggacc
acgtggagaa
      601 aggggtccac caggcccccc aggcagagat ggtgaagatg gtcccacagg
ccctcctggt
      661 ccacctggtc ctcctggccc ccctggtctc ggtgggaact ttgctgctca
gtatgatgga
      721 aaaggagttg gacttggccc tggaccaatg ggcttaatgg gacctagagg
cccacctggt
      781 gcagctggag ccccaggccc tcaaggtttc caaggacctg ctggtgagcc
tggtgaacct
```

841 ggtcaaactg gtcctgcagg tgctcgtggt ccagctggcc ctcctggcaa
ggctggtgaa
901 gatggtcacc ctggaaaacc cggacgacct ggtgagagag gagttgttgg
accacagggt
961 gctcgtggtt tccctggaac tcctggactt cctggcttca aaggcattag
gggacacaat
1021 ggtctggatg gattgaaggg acagcccggt gctcctggtg tgaagggtga
acctggtgcc
1081 cctggtgaaa atggaactcc aggtcaaaca ggagcccgtg ggcttcctgg
tgagagagga
1141 cgtgttggtg cccctggccc agctggtgcc cgtggcagtg atggaagtgt
gggtcccgtg
1201 ggtcctgctg gtcccattgg gtctgctggc cctccaggct tcccaggtgc
ccctggcccc
1261 aagggtgaaa ttggagctgt tggtaacgct ggtcctgctg gtcccgccgg
tccccgtggt
1321 gaagtgggtc ttccaggcct ctccggcccc gttggacctc ctggtaatcc
tggagcaaac
1381 ggccttactg gtgccaaggg tgctgctggc cttcccggcg ttgctggggc
tcccggcctc
1441 cctggacccc gcggtattcc tggccctgtt ggtgctgccg gtgctactgg
tgccagagga
1501 cttgttggtg agcctggtcc agctggctcc aaaggagaga gcggtaacaa
gggtgagccc
1561 ggctctgctg ggccccaagg tcctcctggt cccagtggtg aagaaggaaa
gagaggccct
1621 aatggggaag ctggatctgc cggccctcca ggacctcctg ggctgagagg
tagtcctggt
1681 tctcgtggtc ttcctggagc tgatggcaga gctggcgtca tgggccctcc
tggtagtcgt
1741 ggtgcaagtg gccctgctgg agtccgagga cctaatggag atgctggtcg
ccctggggag
1801 cctggtctca tgggacccag aggtcttcct ggttcccctg gaaatatcgg
ccccgctgga
1861 aaagaaggtc ctgtcggcct ccctggcatc gacggcaggc ctggcccaat
tggcccagct
1921 ggagcaagag gagagcctgg caacattgga ttccctggac ccaaaggccc
cactggtgat
1981 cctggcaaaa acggtgataa aggtcatgct ggtcttgctg gtgctcgggg
tgctccaggt
2041 cctgatggaa acaatggtgc tcagggacct cctggaccac agggtgttca
aggtggaaaa
2101 ggtgaacagg gtccccctgg tcctccaggc ttccagggtc tgcctggccc
ctcaggtccc
2161 gctggtgaag ttggcaaacc aggagaaagg ggtctccatg gtgagtttgg
tctccctggt
2221 cctgctggtc caagagggga acgcggtccc ccaggtgaga gtggtgctgc
cggtcctact
2281 ggtcctattg gaagccgagg tccttctgga cccccagggc ctgatggaaa
caagggtgaa
2341 cctggtgtgg ttggtgctgt gggcactgct ggtccatctg gtcctagtgg
actcccagga
2401 gagaggggtg ctgctggcat acctggaggc aagggagaaa agggtgaacc
tggtctcaga
2461 ggtgaaattg gtaaccctgg cagagatggt gctcgtggtg ctcctggtgc
tgtaggtgcc
2521 cctggtcctg ctggagccac aggtgaccgg ggcgaagctg gggctgctgg
tcctgctggt
2581 cctgctggtc ctcggggaag ccctggtgaa cgtggtgagg tcggtcctgc
tggccccaat

2641 ggatttgctg gtcctgctgg tgctgctggt caacctggtg ctaaaggaga aagaggagcc
2701 aaagggccta agggtgaaaa cggtgttgtt ggtcccacag gccccgttgg agctgctggc
2761 ccagctggtc caaatggtcc ccccggtcct gctggaagtc gtggtgatgg aggcccccct
2821 ggtatgactg gtttccctgg tgctgctgga cggactggtc ccccaggacc ctctggtatt
2881 tctggccctc ctggtccccc tggtcctgct gggaaagaag ggcttcgtgg tcctcgtggt
2941 gaccaaggtc cagttggccg aactggagaa gtaggtgcag ttggtccccc tggcttcgct
3001 ggtgagaagg gtccctctgg agaggctggt actgctggac ctcctggcac tccaggtcct
3061 cagggtcttc ttggtgctcc tggtattctg ggtctccctg gctcgagagg tgaacgtggt
3121 ctaccaggtg ttgctggtgc tgtgggtgaa cctggtcctc ttggcattgc cggccctcct
3181 ggggcccgtg gtcctcctgg tgctgtgggt agtcctggag tcaacggtgc tcctggtgaa
3241 gctggtcgtg atggcaaccc tgggaacgat ggtcccccag gtcgcgatgg tcaacccgga
3301 cacaagggag agcgcggtta ccctggcaat attggtcccg ttggtgctgc aggtgcacct
3361 ggtcctcatg gccccgtggg tcctgctggc aaacatggaa accgtggtga aactggtcct
3421 tctggtcctg ttggtcctgc tggtgctgtt ggcccaagag gtcctagtgg cccacaaggc
3481 attcgtggcg ataagggaga gcccggtgaa aaggggccca gaggtcttcc tggcttaaag
3541 ggacacaatg gattgcaagg tctgcctggt atcgctggtc accatggtga tcaaggtgct
3601 cctggctccg tgggtcctgc tggtcctagg ggccctgctg gtccttctgg ccctgctgga
3661 aaagatggtc gcactggaca tcctggtaca gttggacctg ctggcattcg aggccctcag
3721 ggtcaccaag gccctgctgg ccccccctggt ccccctggcc ctcctggacc tccaggtgta
3781 agcggtggtg gttatgactt tggttacgat ggagacttct acagggctga ccagcctcgc
3841 tcagcacctt ctctcagacc caaggactat gaagttgatg ctactctgaa gtctctcaac
3901 aaccagattg agacccttct tactcctgaa ggctctagaa agaacccagc tcgcacatgc
3961 cgtgacttga gactcagcca cccagagtgg agcagtggtt actactggat tgaccctaac
4021 caaggatgca ctatggatgc tatcaaagta tactgtgatt tctctactgg cgaaacctgt
4081 atccgggccc aacctgaaaa catcccagcc aagaactggt ataggagctc caaggacaag
4141 aaacacgtct ggctaggaga aactatcaat gctggcagcc agtttgaata taatgtagaa
4201 ggagtgactt ccaaggaaat ggctacccaa cttgccttca tgcgcctgct ggccaactat
4261 gcctctcaga acatcaccta ccactgcaag aacagcattg catacatgga tgaggagact
4321 ggcaacctga aaaaggctgt cattctacag ggctctaatg atgttgaact tgttgctgag
4381 ggcaacagca ggttcactta cactgttctt gtagatggct gctctaaaaa gacaaatgaa

```
     4441 tggggaaaga caatcattga atacaaaaca aataagccat cacgcctgcc
cttccttgat
     4501 attgcacctt tggacatcgg tggtgctgac caggaattct ttgtggacat
tggcccagtc
     4561 tgtttcaaat aaatgaactc aatctaaatt aaaaaagaaa gaaatttgaa
aaaactttct
     4621 ctttgccatt tcttcttctt ctttttaac tgaaagctga atccttccat
ttcttctgca
     4681 catctacttg cttaaattgt gggcaaaaga gaaaaagaag gattgatcag
agcattgtgc
     4741 aatacagttt cattaactcc ttcccccgct cccccaaaaa tttgaatttt
tttttcaaca
     4801 ctcttacacc tgttatggaa aatgtcaacc tttgtaagaa aaccaaaata
aaaattgaaa
     4861 aataaaaacc ataaacattt gcaccacttg tggcttttga atatcttcca
cagagggaag
     4921 tttaaaaccc aaacttccaa aggtttaaac tacctcaaaa cactttccca
tgagtgtgat
     4981 ccacattgtt aggtgctgac ctagacagag atgaactgag gtccttgttt
tgttttgttc
     5041 ataatacaaa ggtgctaatt aatagtattt cagatacttg aagaatgttg
atggtgctag
     5101 aagaatttga gaagaaatac tcctgtattg agttgtatcg tgtggtgtat
tttttaaaaa
     5161 atttgattta gcattcatat tttccatctt attcccaatt aaaagtatgc
agattatttg
     5221 cccaaatctt cttcagattc agcatttgtt ctttgccagt ctcattttca
tcttcttcca
     5281 tggttccaca gaagctttgt ttcttgggca agcagaaaaa ttaaattgta
cctattttgt
     5341 atatgtgaga tgtttaaata aattgtgaaa aaaatgaaat aaagcatgtt
tggttttcca
     5401 aaagaacata t
```

Protein sequence:

NCBI Reference Sequence: NP_000080.2

LOCUS NP_000080

ACCESSION NP_000080

```
      1 mlsfvdtrtl lllavtlcla tcqslqeetv rkgpagdrgp rgergppgpp
grdgedgptg
       61 ppgppgppgp pglggnfaaq ydgkgvglgp gpmglmgprg ppgaagapgp
qgfqgpagep
      121 gepgqtgpag argpagppgk agedghpgkp grpgergvvg pqgargfpgt
pglpgfkgir
      181 ghngldglkg qpgapgvkge pgapgengtp gqtgarglpg ergrvgapgp
agargsdgsv
      241 gpvgpagpig sagppgfpga pgpkgeigav gnagpagpag prgevglpgl
sgpvgppgnp
      301 gangltgakg aaglpgvaga pglpgprgip gpvgaagatg arglvgepgp
agskgesgnk
      361 gepgsagpqg ppgpsgeegk rgpngeagsa gppgppglrg spgsrglpga
dgragvmgpp
      421 gsrgasgpag vrgpngdagr pgepglmgpr glpgspgnig pagkegpvgl
pgidgrpgpi
```

```
      481 gpagargepg nigfpgpkgp tgdpgkngdk ghaglagarg apgpdgnnga
qgppgpqgvq
      541 ggkgeqgppg ppgfqglpgp sgpagevgkp gerglhgefg lpgpagprge
rgppgesgaa
      601 gptgpigsrg psgppgpdgn kgepgvvgav gtagpsgpsg lpgergaagi
pggkgekgep
      661 glrgeignpg rdgargapga vgapgpagat gdrgeagaag pagpagprgs
pgergevgpa
      721 gpngfagpag aagqpgakge rgakgpkgen gvvgptgpvg aagpagpngp
pgpagsrgdg
      781 gppgmtgfpg aagrtgppgp sgisgppgpp gpagkeglrg prgdqgpvgr
tgevgavgpp
      841 gfagekgpsg eagtagppgt pgpqgllgap gilglpgsrg erglpgvaga
vgepgplgia
      901 gppgargppg avgspgvnga pgeagrdgnp gndgppgrdg qpghkgergy
pgnigpvgaa
      961 gapgphgpvg pagkhgnrge tgpsgpvgpa gavgprgpsg pqgirgdkge
pgekgprglp
     1021 glkghnglqg lpgiaghhgd qgapgsvgpa gprgpagpsg pagkdgrtgh
pgtvgpagir
     1081 gpqghqgpag ppgppgppgp pgvsgggydf gydgdfyrad qprsapslrp
kdyevdatlk
     1141 slnnqietll tpegsrknpa rtcrdlrlsh pewssgyywi dpnqgctmda
ikvycdfstg
     1201 etciraqpen ipaknwyrss kdkkhvwlge tinagsqfey nvegvtskem
atqlafmrll
     1261 anyasqnity hcknsiaymd eetgnlkkav ilqgsndvel vaegnsrfty
tvlvdgcskk
     1321 tnewgktiie yktnkpsrlp fldiapldig gadqeffvdi gpvcfk
```

LAMC1

Official Symbol: LAMC1

Official Name: laminin, gamma 1 (formerly LAMB2)

Gene ID: 3915

Organism: Homo sapiens

Other Aliases: RP11-181K3.1, LAMB2

Other Designations: S-LAM gamma; S-laminin subunit gamma; laminin B2 chain; laminin subunit gamma-1; laminin-10 subunit gamma; laminin-11 subunit gamma; laminin-2 subunit gamma; laminin-3 subunit gamma; laminin-4 subunit gamma; laminin-6 subunit gamma; laminin-7 subunit gamma; laminin-8 subunit gamma; laminin-9 subunit gamma

Nucleotide sequence:

NCBI Reference Sequence: NM_002293.3

LOCUS NM_002293

ACCESSION NM_002293

```
      1 gtgcaggctg ctcccggggt aggtgaggga agcgcggagg cggcgcgcgg
gggcagtggt
       61 cggcgagcag cgcggtcctc gctaggggcg cccacccgtc agtctctccg
gcgcgagccg
      121 ccgccaccgc ccgcgccgga gtcaggcccc tgggccccca ggctcaagca
gcgaagcggc
      181 ctccgggggga cgccgctagg cgagaggaac gcgccggtgc ccttgccttc
gccgtgaccc
      241 agcgtgcggg cggcgggatg agagggagcc atcgggccgc gccggccctg
cggccccggg
      301 ggcggctctg gcccgtgctg gccgtgctgg cggcggccgc cgcggcgggc
tgtgcccagg
      361 cagccatgga cgagtgcacg gacgagggcg ggcggccgca gcgctgcatg
cccgagttcg
      421 tcaacgccgc cttcaacgtg actgtggtgg ccaccaacac gtgtgggact
ccgcccgagg
      481 aatactgtgt gcagaccggg gtgaccgggg tcaccaagtc ctgtcacctg
tgcgacgccg
      541 ggcagcccca cctgcagcac ggggcagcct tcctgaccga ctacaacaac
caggccgaca
      601 ccacctggtg gcaaagccag accatgctgg ccggggtgca gtaccccagc
tccatcaacc
      661 tcacgctgca cctgggaaaa gcttttgaca tcacctatgt gcgtctcaag
ttccacacca
      721 gccgcccgga gagctttgcc atttacaagc gcacacggga agacgggccc
tggattcctt
      781 accagtacta cagtggttcc tgtgagaaca cctactccaa ggcaaaccgc
ggcttcatca
      841 ggacaggagg ggacgagcag caggccttgt gtactgatga attcagtgac
atttctcccc
      901 tcactggggg caacgtggcc ttttctaccc tggaaggaag gcccagcgcc
tataactttg
      961 acaatagccc tgtgctgcag gaatgggtaa ctgccactga catcagagta
actcttaatc
     1021 gcctgaacac ttttggagat gaagtgttta acgatcccaa agttctcaag
tcctattatt
     1081 atgccatctc tgattttgct gtaggtggca gatgtaaatg taatggacac
gcaagcgagt
     1141 gtatgaagaa cgaatttgat aagctggtgt gtaattgcaa acataacaca
tatggagtag
     1201 actgtgaaaa gtgtcttcct ttcttcaatg accggccgtg gaggagggca
actgcggaaa
     1261 gtgccagtga atgcctgccc tgtgattgca atggtcgatc ccaggaatgc
tacttcgacc
     1321 ctgaactcta tcgttccact ggccatgggg gccactgtac caactgccag
gataacacag
     1381 atggcgccca ctgtgagagg tgccgagaga acttcttccg ccttggcaac
aatgaagcct
     1441 gctcttcatg ccactgtagt cctgtgggct ctctaagcac acagtgtgat
agttacggca
     1501 gatgcagctg taagccagga gtgatggggg acaaatgtga ccgttgccag
cctggattcc
     1561 attctctcac tgaagcagga tgcaggccat gctcttgtga tccctctggc
agcatagatg
     1621 aatgtaatat tgaaacagga agatgtgttt gcaaagacaa tgtcgaaggc
ttcaattgtg
     1681 aaagatgcaa acctggattt tttaatctgg aatcatctaa tcctcggggt
tgcacaccct
```

1741 gcttctgctt tgggcattct tctgtctgta caaacgctgt tggctacagt
gtttattcta
1801 tctcctctac ctttcagatt gatgaggatg ggtggcgtgc ggaacagaga
gatggctctg
1861 aagcatctct cgagtggtcc tctgagaggc aagatatcgc cgtgatctca
gacagctact
1921 ttcctcggta cttcattgct cctgcaaagt tcttgggcaa gcaggtgttg
agttatggtc
1981 agaacctctc cttctccttt cgagtggaca ggcgagatac tcgcctctct
gcagaagacc
2041 ttgtgcttga gggagctggc ttaagagtat ctgtaccctt gatcgctcag
ggcaattcct
2101 atccaagtga gaccactgtg aagtatgtct tcaggctcca tgaagcaaca
gattaccctt
2161 ggaggcctgc tcttacccct tttgaatttc agaagctcct aaacaacttg
acctctatca
2221 agatacgtgg gacatacagt gagagaagtg ctggatattt ggatgatgtc
accctggcaa
2281 gtgctcgtcc tgggcctgga gtccctgcaa cttgggtgga gtcctgcacc
tgtcctgtgg
2341 gatatggagg gcagttttgt gagatgtgcc tctcaggtta cagaagagaa
actcctaatc
2401 ttggaccata cagtccatgt gtgctttgcg cctgcaatgg acacagcgag
acctgtgatc
2461 ctgagacagg tgtttgtaac tgcagagaca atacggctgg cccgcactgt
gagaagtgca
2521 gtgatgggta ctatggagat tcaactgcag gcacctcctc cgattgccaa
ccctgtccgt
2581 gtcctggagg ttcaagttgt gctgttgttc ccaagacaaa ggaggtggtg
tgcaccaact
2641 gtcctactgg caccactggt aagagatgtg agctctgtga tgatggctac
tttggagacc
2701 ccctgggtag aaacggccct gtgagacttt gccgcctgtg ccagtgcagt
gacaacatcg
2761 atcccaatgc agttggaaat tgcaatcgct tgacgggaga atgcctgaag
tgcatctata
2821 acactgctgg cttctattgt gaccggtgca aagacggatt ttttggaaat
cccctggctc
2881 ccaatccagc agacaaatgc aaagcctgca attgcaatct gtatgggacc
atgaagcagc
2941 agagcagctg taaccccgtg acggggcagt gtgaatgttt gcctcacgtg
actggccagg
3001 actgtggtgc ttgtgaccct ggattctaca atctgcagag tgggcaaggc
tgtgagaggt
3061 gtgactgcca tgccttgggc tccaccaatg ggcagtgtga catccgcacc
ggccagtgtg
3121 agtgccagcc cggcatcact ggtcagcact gtgagcgctg tgaggtcaac
cactttgggt
3181 ttggacctga aggctgcaaa ccctgtgact gtcatcctga gggatctctt
tcacttcagt
3241 gcaaagatga tggtcgctgt gaatgcagag aaggctttgt gggaaatcgc
tgtgaccagt
3301 gtgaagaaaa ctatttctac aatcggtctt ggcctggctg ccaggaatgt
ccagcttgtt
3361 accggctggt aaaggataag gttgctgatc atagagtgaa gctccaggaa
ttagagagtc
3421 tcatagcaaa ccttggaact ggggatgaga tggtgacaga tcaagccttc
gaggatagac
3481 taaaggaagc agagagggaa gttatggacc tccttcgtga ggcccaggat
gtcaaagatg

3541 ttgaccagaa tttgatggat cgcctacaga gagtgaataa cactctgtcc
agccaaatta
3601 gccgtttaca gaatatccgg aataccattg aagagactgg aaacttggct
gaacaagcgc
3661 gtgcccatgt agagaacaca gagcggttga ttgaaatcgc atccagagaa
cttgagaaag
3721 caaaagtcgc tgctgccaat gtgtcagtca ctcagccaga atctacaggg
gacccaaaca
3781 acatgactct tttggcagaa gaggctcgaa agcttgctga acgtcataaa
caggaagctg
3841 atgacattgt tcgagtggca aagacagcca atgatacgtc aactgaggca
tacaacctgc
3901 ttctgaggac actggcagga gaaaatcaaa cagcatttga gattgaagag
cttaatagga
3961 agtatgaaca agcgaagaac atctcacagg atctggaaaa acaagctgcc
cgagtacatg
4021 aggaggccaa aagggccggt gacaaagctg tggagatcta tgccagcgtg
gctcagctga
4081 gccctttgga ctctgagaca ctggagaatg aagcaaataa cataaagatg
gaagctgaga
4141 atctggaaca actgattgac cagaaattaa aagattatga ggacctcaga
gaagatatga
4201 gagggaagga acttgaagtc aagaaccttc tggagaaagg caagactgaa
cagcagaccg
4261 cagaccaact cctagcccga gctgatgctg ccaaggccct cgctgaagaa
gctgcaaaga
4321 agggacggga taccttacaa gaagctaatg acattctcaa caacctgaaa
gattttgata
4381 ggcgtgtgaa cgataacaag acggccgcag aggaggcact aaggaagatt
cctgccatca
4441 accagaccat cactgaagcc aatgaaaaga ccagagaagc ccagcaggcc
ctgggcagtg
4501 ctgcggcgga tgccacagag gccaagaaca aggcccatga ggcggagagg
atcgcgagcg
4561 ctgtccaaaa gaatgccacc agcaccaagg cagaagctga aagaactttt
gcagaagtta
4621 cagatctgga taatgaggtg aacaatatgt tgaagcaact gcaggaagca
gaaaaagagc
4681 taaagagaaa acaagatgac gctgaccagg acatgatgat ggcagggatg
gcttcacagg
4741 ctgctcaaga agccgagatc aatgccagaa aagccaaaaa ctctgttact
agcctcctca
4801 gcattattaa tgacctcttg gagcagctgg ggcagctgga tacagtggac
ctgaataagc
4861 taaacgagat tgaaggcacc ctaaacaaag ccaaagatga aatgaaggtc
agcgatcttg
4921 ataggaaagt gtctgacctg gagaatgaag ccaagaagca ggaggctgcc
atcatggact
4981 ataaccgaga tatcgaggag atcatgaagg acattcgcaa tctggaggac
atcaggaaga
5041 ccttaccatc tggctgcttc aacaccccgt ccattgaaaa gccctagtgt
ctttagggct
5101 ggaaggcagc atccctctga caggggggca gttgtgaggc cacagagtgc
cttgacacaa
5161 agattacatt tttcagaccc ccactcctct gctgctgtcc atgactgtcc
ttttgaacca
5221 ggaaaagtca cagagtttaa agagaagcaa attaaacatc ctgaatcggg
aacaaagggt
5281 tttatctaat aaagtgtctc ttccattcac gttgctacct tacccacact
ttcccttctg

5341 atttgcgtga ggacgtggca tcctacgtta ctgtacagtg gcataagcac
atcgtgtgag
5401 cccatgtatg ctggggtaga gcaagtagcc ctcccctgtc tcatcgatac
cagcagaacc
5461 tcctcagtct cagtactctt gtttctatga aggaaaagtt tggctactaa
cagtagcatt
5521 gtgatggcca gtatatccag tccatggata aagaaaatgc atctgcatct
cctacccctc
5581 ttccttctaa gcaaaaggaa ataaacatcc tgtgccaaag gtattggtca
tttagaatgt
5641 cggtagccat ccatcagtgc ttttagttat tatgagtgta ggacactgag
ccatccgtgg
5701 gtcaggatgc aattatttat aaaagtctcc aggtgaacat ggctgaagat
ttttctagta
5761 tattaataat tgactaggaa gatgaacttt ttttcagatc tttgggcagc
tgataattta
5821 aatctggatg ggcagcttgc actcaccaat agaccaaaag acatcttttg
atattcttat
5881 aaatggaact tacacagaag aaatagggat atgataacca ctaaaatttt
gttttcaaaa
5941 tcaaactaat tcttacagct tttttattag ttagtcttgg aactagtgtt
aagtatctgg
6001 cagagaacag ttaatcccta aggtcttgac aaaacagaag aaaaacaagc
ctcctcgtcc
6061 tagtcttttc tagcaaaggg ataaaactta gatggcagct tgtactgtca
gaatcccgtg
6121 tatccatttg ttcttctgtt ggagagatga gacatttgac ccttagctcc
agttttcttc
6181 tgatgtttcc atcttccaga atccctcaaa aaacattgtt tgccaaatcc
tggtggcaaa
6241 tacttgcact cagtatttca cacagctgcc aacgctatcg agttcctgca
ctttgtgatt
6301 taaatccact ctaaaccttc cctctaagtg tagagggaag acccttacgt
ggagtttcct
6361 agtgggcttc tcaacttttg atcctcagct ctgtggtttt aagaccacag
tgtgacagtt
6421 ccctgccaca cacccccttc ctcctaccaa cccacctttg agattcatat
atagccttta
6481 acactatgca actttgtact ttgcgtagca ggggcggggt ggggggaaag
aaactattat
6541 ctgacacact ggtgctatta attatttcaa atttatattt ttgtgtgaat
gttttgtgtt
6601 ttgtttatca tgattataga ataaggaatt tatgtaaata tacttagtcc
tatttctaga
6661 atgacactct gttcactttg ctcaattttt cctcttcact ggcacaatgt
atctgaatac
6721 ctccttccct cccttctaga attctttgga ttgtactcca aagaattgtg
ccttgtgttt
6781 gcagcatctc cattctctaa aattaatata attgctttcc tccacaccca
gccactgtaa
6841 agaggtaact tgggtcctct tccattgcag tcctgatgat cctaacctgc
agcacggtgg
6901 ttttacaatg ttccagagca ggaacgccag gttgacaagc tatggtagga
ttaggaaagt
6961 ttgctgaaga ggatctttga cgccacagtg ggactagcca ggaatgaggg
agaaatgccc
7021 tttctggcaa ttgttggagc tggataggta agttttataa gggagtacat
tttgactgag
7081 cacttagggc atcaggaaca gtgctactta ctgatgggta gactgggaga
ggtggtgtaa

```
     7141 cttagttctt gatgatccca cttcctgttt ccatctgctt gggatatacc
agagtttacc
     7201 acaagtgttt tgacgatata ctcctgagct ttcactctgc tgcttctccc
aggcctcttc
     7261 tactatggca ggagatgtgg cgtgctgttg caaagttttc acgtcattgt
ttcctggcta
     7321 gttcatttca ttaagtggct acatcctaac atatgcattt ggtcaaggtt
gcagaagagg
     7381 actgaagatt gactgccaag ctagtttggg tgaagttcac tccagcaagt
ctcaggccac
     7441 aatggggtgg tttggtttgg tttcctttta actttctttt tgttatttgc
ttttctcctc
     7501 cacctgtgtg gtatattttt taagcagaat tttatttttt aaaataaaag
gttctttaca
     7561 agatgatacc ttaattacac tcccgcaaca cagccattat tttattgtct
agctccagtt
     7621 atctgtattt tatgtaatgt aattgacagg atggctgctg cagaatgctg
gttgacacag
     7681 ggattattat actgctattt ttccctgaat ttttttcctt tgaattccaa
ctgtggacct
     7741 tttatatgtg ccttcacttt agctgtttgc cttaatctct acagccttgc
tctccggggt
     7801 ggttaataaa atgcaacact tggcattttt atgttttaag aaaaacagta
ttttatttat
      7861 aataaaatct gaatatttgt aacccttt
```

Protein sequence:

NCBI Reference Sequence: NP_002284.3

LOCUS NP_002284

ACCESSION NP_002284

```
      1 mrgshraapa lrprgrlwpv lavlaaaaaa gcaqaamdec tdeggrpqrc
mpefvnaafn
       61 vtvvatntcg tppeeycvqt gvtgvtksch lcdagqphlq hgaafltdyn
nqadttwwqs
      121 qtmlagvqyp ssinltlhlg kafdityvrl kfhtsrpesf aiykrtredg
pwipyqyysg
      181 scentyskan rgfirtggde qqalctdefs displtggnv afstlegrps
aynfdnspvl
      241 qewvtatdir vtlnrlntfg devfndpkvl ksyyyaisdf avggrckcng
hasecmknef
      301 dklvcnckhn tygvdcekcl pffndrpwrr ataesasecl pcdcngrsqe
cyfdpelyrs
      361 tghgghctnc qdntdgahce rcrenffrlg nneacsschc spvgslstqc
dsygrcsckp
      421 gvmgdkcdrc qpgfhsltea gcrpcscdps gsidecniet grcvckdnve
gfncerckpg
      481 ffnlessnpr gctpcfcfgh ssvctnavgy svysisstfq ididgwraeq
rdgseaslew
      541 sserqdiavi sdsyfpryfi apakflgkqv lsygqnlsfs frvdrrdtrl
saedlvlega
      601 glrvsvplia qgnsypsett vkyvfrlhea tdypwrpalt pfefqkllnn
ltsikirgty
```

```
      661 sersagyldd vtlasarpgp gvpatwvesc tcpvgyggqf cemclsgyrr
etpnlgpysp
      721 cvlcacnghs etcdpetgvc ncrdntagph cekcsdgyyg dstagtssdc
qpcpcpggss
      781 cavvpktkev vctncptgtt gkrcelcddg yfgdplgrng pvrlcrlcqc
sdnidpnavg
      841 ncnrltgecl kciyntagfy cdrckdgffg nplapnpadk ckacncnlyg
tmkqqsscnp
      901 vtgqceclph vtgqdcgacd pgfynlqsgq gcercdchal gstngqcdir
tgqcecqpgi
      961 tgqhcercev nhfgfgpegc kpcdchpegs lslqckddgr cecregfvgn
rcdqceenyf
     1021 ynrswpgcqe cpacyrlvkd kvadhrvklq eleslianlg tgdemvtdqa
fedrlkeaer
     1081 evmdllreaq dvkdvdqnlm drlqrvnntl ssqisrlqni rntieetgnl
aeqarahven
     1141 terlieiasr elekakvaaa nvsvtqpest gdpnnmtlla eearklaerh
kqeaddivrv
     1201 aktandtste aynlllrtla genqtafeie elnrkyeqak nisqdlekqa
arvheeakra
     1261 gdkaveiyas vaqlspldse tleneannik meaenleqli dqklkdyedl
redmrgkele
     1321 vknllekgkt eqqtadqlla radaakalae eaakkgrdtl qeandilnnl
kdfdrrvndn
     1381 ktaaeealrk ipainqtite anektreaqq algsaaadat eaknkaheae
riasavqkna
     1441 tstkaeaert faevtdldne vnnmlkqlqe aekelkrkqd dadqdmmmag
masqaaqeae
     1501 inarkaknsv tsllsiindl leqlgqldtv dlnklneieg tlnkakdemk
vsdldrkvsd
      1561 leneakkqea aimdynrdie eimkdirnle dirktlpsgc fntpsiekp
```

SPRC

Official Symbol: SPARC

Official Name: secreted protein, acidic, cysteine-rich (osteonectin)

Gene ID: 6678

Organism: Homo sapiens

Other Aliases: ON

Other Designations: BM-40; basement-membrane protein 40; cysteine-rich protein; osteonectin; secreted protein acidic and rich in cysteine

Nucleotide sequence:

NCBI Reference Sequence: NM_003118.3

LOCUS NM_003118

ACCESSION NM_003118

```
      1 gggagaagga ggaggccggg ggaaggagga gacaggagga ggagggacca
cggggtggag
```

```
       61 gggagataga cccagcccag agctctgagt ggtttcctgt tgcctgtctc
taaacccctc
      121 cacattcccg cggtccttca gactgcccgg agagcgcgct ctgcctgccg
cctgcctgcc
      181 tgccactgag ggttcccagc accatgaggg cctggatctt ctttctcctt
tgcctggccg
      241 ggagggcctt ggcagcccct cagcaagaag ccctgcctga tgagacagag
gtggtggaag
      301 aaactgtggc agaggtgact gaggtatctg tgggagctaa tcctgtccag
gtggaagtag
      361 gagaatttga tgatggtgca gaggaaaccg aagaggaggt ggtggcggaa
aatccctgcc
      421 agaaccacca ctgcaaacac ggcaaggtgt gcgagctgga tgagaacaac
acccccatgt
      481 gcgtgtgcca ggaccccacc agctgcccag cccccattgg cgagtttgag
aaggtgtgca
      541 gcaatgacaa caagaccttc gactcttcct gccacttctt tgccacaaag
tgcaccctgg
      601 agggcaccaa gaagggccac aagctccacc tggactacat cgggccttgc
aaatacatcc
      661 ccccttgcct ggactctgag ctgaccgaat tccccctgcg catgcgggac
tggctcaaga
      721 acgtcctggt caccctgtat gagagggatg aggacaacaa ccttctgact
gagaagcaga
      781 agctgcgggt gaagaagatc catgagaatg agaagcgcct ggaggcagga
gaccaccccg
      841 tggagctgct ggcccgggac ttcgagaaga actataacat gtacatcttc
cctgtacact
      901 ggcagttcgg ccagctggac cagcacccca ttgacgggta cctctcccac
accgagctgg
      961 ctccactgcg tgctcccctc atccccatgg agcattgcac cacccgcttt
ttcgagacct
     1021 gtgacctgga caatgacaag tacatcgccc tggatgagtg ggccggctgc
ttcggcatca
     1081 agcagaagga tatcgacaag gatcttgtga tctaaatcca ctccttccac
agtaccggat
     1141 tctctcttta accctcccct tcgtgtttcc cccaatgttt aaaatgtttg
gatggtttgt
     1201 tgttctgcct ggagacaagg tgctaacata gatttaagtg aatacattaa
cggtgctaaa
     1261 aatgaaaatt ctaacccaag acatgacatt cttagctgta acttaactat
taaggccttt
     1321 tccacacgca ttaatagtcc catttttctc ttgccatttg tagctttgcc
cattgtctta
     1381 ttggcacatg ggtggacacg gatctgctgg gctctgcctt aaacacacat
tgcagcttca
     1441 acttttctct ttagtgttct gtttgaaact aatacttacc gagtcagact
ttgtgttcat
     1501 ttcatttcag ggtcttggct gcctgtgggc ttccccaggt ggcctggagg
tgggcaaagg
     1561 gaagtaacag acacacgatg ttgtcaagga tggttttggg actagaggct
cagtggtggg
     1621 agagatccct gcagaaccca ccaaccagaa cgtggtttgc ctgaggctgt
aactgagaga
     1681 aagattctgg ggctgtgtta tgaaaatata gacattctca cataagccca
gttcatcacc
     1741 atttcctcct ttacctttca gtgcagtttc ttttcacatt aggctgttgg
ttcaaacttt
     1801 tgggagcacg gactgtcagt tctctgggaa gtggtcagcg catcctgcag
ggcttctcct
```

1861 cctctgtctt ttggagaacc agggctcttc tcaggggctc tagggactgc caggctgttt
1921 cagccaggaa ggccaaaatc aagagtgaga tgtagaaagt tgtaaaatag aaaaagtgga
1981 gttggtgaat cggttgttct ttcctcacat ttggatgatt gtcataaggt ttttagcatg
2041 ttcctccttt tcttcaccct ccccttttt cttctattaa tcaagagaaa cttcaaagtt
2101 aatgggatgg tcggatctca caggctgaga actcgttcac ctccaagcat ttcatgaaaa
2161 agctgcttct tattaatcat acaaactctc accatgatgt gaagagtttc acaaatcctt
2221 caaaataaaa agtaatgact tagaaactgc cttcctgggt gatttgcatg tgtcttagtc
2281 ttagtcacct tattatcctg acacaaaaac acatgagcat acatgtctac acatgactac
2341 acaaatgcaa acctttgcaa acacattatg cttttgcaca cacacacctg tacacacaca
2401 ccggcatgtt tatacacagg gagtgtatgg ttcctgtaag cactaagtta gctgttttca
2461 tttaatgacc tgtggtttaa ccctttgat cactaccacc attatcagca ccagactgag
2521 cagctatatc cttttattaa tcatggtcat tcattcattc attcattcac aaaatattta
2581 tgatgtattt actctgcacc aggtcccatg ccaagcactg gggacacagt tatggcaaag
2641 tagacaaagc atttgttcat ttggagctta gagtccagga ggaatacatt agataatgac
2701 acaatcaaat ataaattgca agatgtcaca ggtgtgatga agggagagta ggagagacca
2761 tgagtatgtg taacaggagg acacagcatt attctagtgc tgtactgttc cgtacggcag
2821 ccactaccca catgtaactt tttaagattt aaatttaaat tagttaacat tcaaaacgca
2881 gctccccaat cacactagca acatttcaag tgcttgagag ccatgcatga ttagtggtta
2941 ccctattgaa taggtcagaa gtagaatctt ttcatcatca cagaaagttc tattggacag
3001 tgctcttcta gatcatcata agactacaga gcacttttca aagctcatgc atgttcatca
3061 tgttagtgtc gtattttgag ctggggtttt gagactcccc ttagagatag agaaacagac
3121 ccaagaaatg tgctcaattg caatgggcca catacctaga tctccagatg tcatttcccc
3181 tctcttattt taagttatgt taagattact aaaacaataa aagctcctaa aaaatcaaac
3241 tgtattctgg tgttctcttc tacacagtgg gagggcgagc agtaggagag attggcccat
3301 ttggtgctgg ccatttgagg aatgcaagcc cagcactagt ctcataatct ctaggaatct
3361 gtagagagag gaattgaagt aaatttcagc attggctcat tcagtcattc ggcgacattc
3421 atcaggtacc tgcaatgtgt taggggatct tatgagtagg cagcgtgcgt gatccttgct
3481 cccctggagc tttctaacat tctagcaggc agaccacaca taaatttgca atactgtttc
3541 tgataaaaac gtgctgtaaa ggaaataaag cagagaacta tcatggaaaa aaaaaaaaaa
3601 aaaa

Protein sequence:

NCBI Reference Sequence: NP_003109.1

LOCUS NP_003109

ACCESSION NP_003109

```
      1 mrawiffllc lagralaapq qealpdetev veetvaevte vsvganpvqv
evgefddgae
       61 eteeevvaen pcqnhhckhg kvceldennt pmcvcqdpts cpapigefek
vcsndnktfd
      121 sschffatkc tlegtkkghk lhldyigpck yippcldsel tefplrmrdw
lknvlvtlye
      181 rdednnllte kqklrvkkih enekrleagd hpvellardf eknynmyifp
vhwqfgqldq
      241 hpidgylsht elaplrapli pmehcttrff etcdldndky ialdewagcf
gikqkdidkd
      301 lvi
```

P3H1

Official Symbol: LEPRE1

Official Name: leucine proline-enriched proteoglycan (leprecan) 1

Gene ID: 64175

Organism: Homo sapiens

Other Aliases: PSEC0109, GROS1, OI8, **P3H1**

Other Designations: growth suppressor 1; leprecan; leucine- and proline-enriched proteoglycan 1; prolyl 3-hydroxylase 1

Nucleotide sequence:

NCBI Reference Sequence: NM_001146289.1

LOCUS NM_001146289

ACCESSION NM_001146289

```
      1 atgcgccgcc cggcttggaa ggtggggctt cgcccggggg cgggccttcg
ccgggggtag
       61 gactccggcc ttggtggcgg gtggctggcg gttccgttag gtctgaggga
gcgatggcgg
      121 tacgcgcgtt gaagctgctg accacactgc tggctgtcgt ggccgctgcc
tcccaagccg
      181 aggtcgagtc cgaggcagga tggggcatgg tgacgcctga tctgctcttc
gccgagggga
      241 ccgcagccta cgcgcgcggg gactggcccg gggtggtcct gagcatggaa
cgggcgctgc
      301 gctcccgggc agccctccgc gcccttcgcc tgcgctgccg cacccagtgt
gccgccgact
```

361 tcccgtggga gctggacccc gactggtccc ccagcccggc ccaggcctcg ggcgccgccg
421 ccctgcgcga cctgagcttc ttcgggggcc ttctgcgtcg cgctgcctgc ctgcgccgct
481 gcctcgggcc gccggccgcc cactcgctca gcgaagagat ggagctggag ttccgcaagc
541 ggagccccta caactacctg caggtcgcct acttcaagat caacaagttg gagaaagctg
601 ttgctgcagc acacaccttc ttcgtgggca atcctgagca catggaaatg cagcagaacc
661 tagactatta ccaaaccatg tctggagtga aggaggccga cttcaaggat cttgagactc
721 aaccccatat gcaagaattt cgactgggag tgcgactcta ctcagaggaa cagccacagg
781 aagctgtgcc ccacctagag gcggcgctgc aagaatactt tgtggcctat gaggagtgcc
841 gtgccctctg cgaagggccc tatgactacg atggctacaa ctaccttgag tacaacgctg
901 acctcttcca ggccatcaca gatcattaca tccaggtcct caactgtaag cagaactgtg
961 tcacggagct tgcttcccac ccaagtcgag agaagccctt tgaagacttc ctcccatcgc
1021 attataatta tctgcagttt gcctactata acattgggaa ttatacacag gctgttgaat
1081 gtgccaagac ctatcttctc ttcttcccca atgacgaggt gatgaaccaa aatttggcct
1141 attatgcagc tatgcttgga gaagaacaca ccagatccat cggcccccgt gagagtgcca
1201 aggagtaccg acagcgaagc ctactggaaa aagaactgct tttcttcgct tatgatgttt
1261 ttggaattcc ctttgtggat ccggattcat ggactccaga agaagtgatt cccaagagat
1321 tgcaagagaa acagaagtca gaacgggaaa cagccgtacg catctcccag gagattggga
1381 accttatgaa ggaaatcgag acccttgtgg aagagaagac caaggagtca ctggatgtga
1441 gcagactgac ccgggaaggt ggccccctgc tgtatgaagg catcagtctc accatgaact
1501 ccaaactcct gaatggttcc cagcgggtgg tgatggacgg cgtaatctct gaccacgagt
1561 gtcaggagct gcagagactg accaatgtgg cagcaacctc aggagatggc taccggggtc
1621 agacctcccc acatactccc aatgaaaagt tctatggtgt cactgtcttc aaagccctca
1681 agctggggca agaaggcaaa gttcctctgc agagtgccca cctgtactac aacgtgacgg
1741 agaaggtgcg gcgcatcatg gagtcctact tccgcctgga tacgcccctc tacttttcct
1801 actctcatct ggtgtgccgc actgccatcg aagaggtcca ggcagagagg aaggatgata
1861 gtcatccagt ccacgtggac aactgcatcc tgaatgccga gaccctcgtg tgtgtcaaag
1921 agcccccagc ctacaccttc cgcgactaca gcgccatcct ttacctaaat ggggacttcg
1981 atggcggaaa cttttatttc actgaactgg atgccaagac cgtgacggca gaggtgcagc
2041 ctcagtgtgg aagagccgtg ggattctctt caggcactga aaacccacat ggagtgaagg
2101 ctgtcaccag ggggcagcgc tgtgccatcg ccctgtggtt caccctggac cctcgacaca

```
     2161 gcgagcgggt gagagcagct cgagcgggac agggtgcagg cagatgacct
ggtgaagatg
     2221 ctcttcagcc cagaagagat ggacctctcc caggagcagc ccctggatgc
ccagcagggc
     2281 ccccccgaac ctgcacaaga gtctctctca ggcagtgaat cgaagcccaa
ggatgagcta
     2341 tgacagcgtc caggtcagac ggatgggtga ctagacccat ggagaggaac
tcttctgcac
     2401 tctgagctgg ccagcccctc ggggctgcag agcagtgagc ctacatctgc
cactcagccg
     2461 aggggaccct gctcacagcc ttctacatgg tgctactgct cttggagtgg
acatgaccag
     2521 acaccgcacc ccctggatct ggctgagggc tcaggacaca ggcccagcca
cccccagggg
     2581 cctccacagg ccgctgcatg acagcgatac agtacttaag tgtctgtgta
gacaaccaaa
     2641 gaataaatga ttcatggttt tttttacttg gtttgttcag acaatggaaa
tttgcccatt
     2701 ctgtcaaaaa aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001139761.1

LOCUS NP_001139761

ACCESSION NP_001139761

```
      1 mavralkllt tllavvaaas qaeveseagw gmvtpdllfa egtaayargd
wpgvvlsmer
       61 alrsraalra lrlrcrtqca adfpweldpd wspspaqasg aaalrdlsff
ggllrraacl
      121 rrclgppaah slseemelef rkrspynylq vayfkinkle kavaaahtff
vgnpehmemq
      181 qnldyyqtms gvkeadfkdl etqphmqefr lgvrlyseeq pqeavphlea
alqeyfvaye
      241 ecralcegpy dydgynyley nadlfqaitd hyiqvlnckq ncvtelashp
srekpfedfl
      301 pshynylqfa yynignytqa vecaktyllf fpndevmnqn layyaamlge
ehtrsigpre
      361 sakeyrqrsl lekellffay dvfgipfvdp dswtpeevip krlqekqkse
retavrisqe
      421 ignlmkeiet lveektkesl dvsrltregg pllyegislt mnskllngsq
rvvmdgvisd
      481 hecqelqrlt nvaatsgdgy rgqtsphtpn ekfygvtvfk alklgqegkv
plqsahlyyn
      541 vtekvrrime syfrldtply fsyshlvcrt aieevqaerk ddshpvhvdn
cilnaetlvc
      601 vkeppaytfr dysailylng dfdggnfyft eldaktvtae vqpqcgravg
fssgtenphg
      661 vkavtrgqrc aialwftldp rhservraar agqgagr
```

CO6A1

Official Symbol: COL6A1

Official Name: collagen, type VI, alpha 1

Gene ID: 1291

Organism: Homo sapiens

Other Aliases: OPLL

Other Designations: alpha 1 (VI) chain (61 AA); collagen VI, alpha-1 polypeptide; collagen alpha-1(VI) chain

Nucleotide sequence:

NCBI Reference Sequence: NM_001848.2

LOCUS NM_001848

ACCESSION NM_001848

```
      1 gctctcactc tggctgggag cagaaggcag cctcggtctc tgggcggcgg
cggcggccca
       61 ctctgccctg gccgcgctgt gtggtgaccg caggccccag acatgagggc
ggcccgtgct
      121 ctgctgcccc tgctgctgca ggcctgctgg acagccgcgc aggatgagcc
ggagaccccg
      181 agggccgtgg ccttccagga ctgccccgtg gacctgttct ttgtgctgga
cacctctgag
      241 agcgtggccc tgaggctgaa gccctacggg gccctcgtgg acaaagtcaa
gtccttcacc
      301 aagcgcttca tcgacaacct gagggacagg tactaccgct gtgaccgaaa
cctggtgtgg
      361 aacgcaggcg cgctgcacta cagtgacgag gtggagatca tccaaggcct
cacgcgcatg
      421 cctggcggcc gcgacgcact caaaagcagc gtggacgcgg tcaagtactt
tgggaagggc
      481 acctacaccg actgcgctat caagaagggg ctggagcagc tcctcgtggg
gggctcccac
      541 ctgaaggaga ataagtacct gattgtggtg accgacgggc acccccctgga
gggctacaag
      601 gaaccctgtg gggggctgga ggatgctgtg aacgaggcca agcacctggg
cgtcaaagtc
      661 ttctcggtgg ccatcacacc cgaccacctg gagccgcgtc tgagcatcat
cgccacggac
      721 cacacgtacc ggcgcaactt cacggcggct gactggggcc agagccgcga
cgcagaggag
      781 gccatcagcc agaccatcga caccatcgtg gacatgatca aaaataacgt
ggagcaagtg
      841 tgctgctcct tcgaatgcca gcctgcaaga ggacctccgg ggctccgggg
cgaccccggc
      901 tttgagggag aacgaggcaa gccggggctc ccaggagaga agggagaagc
cggagatcct
      961 ggaagacccg gggacctcgg acctgttggg taccagggaa tgaagggaga
aaaagggagc
     1021 cgtggggaga agggctccag gggacccaag ggctacaagg gagagaaggg
caagcgtggc
     1081 atcgacgggg tggacggcgt gaagggggag atggggtacc caggcctgcc
aggctgcaag
     1141 ggctcgcccg ggtttgacgg cattcaagga ccccctggcc ccaagggaga
ccccggtgcc
```

1201 tttggactga aaggagaaaa gggcgagcct ggagctgacg gggaggcggg gagaccaggg
1261 agctcgggac catctggaga cgagggccag ccgggagagc ctgggccccc cggagagaaa
1321 ggagaggcgg gcgacgaggg gaacccagga cctgacggtg cccccgggga gcggggtggc
1381 cctggagaga gaggaccacg ggggacccca ggcacgcggg gaccaagagg agaccctggt
1441 gaagctggcc cgcagggtga tcagggaaga gaaggccccg ttggtgtccc tggagacccg
1501 ggcgaggctg gccctatcgg acctaaaggc taccgaggcg atgagggtcc cccagggtcc
1561 gagggtgcca gaggagcccc aggacctgcc ggacccccctg gagacccggg gctgatgggt
1621 gaaaggggag aagacggccc cgctggaaat ggcaccgagg gcttccccgg cttccccggg
1681 tatccgggca acaggggcgc tcccgggata aacggcacga agggctaccc cggcctcaag
1741 ggggacgagg gagaagccgg ggaccccgga gacgataaca acgacattgc accccgagga
1801 gtcaaaggag caaaggggta ccggggtccc gagggccccc agggaccccc aggacaccaa
1861 ggaccgcctg ggccggacga atgcgagatt ttggacatca tcatgaaaat gtgctcttgc
1921 tgtgaatgca agtgcggccc catcgacctc ctgttcgtgc tggacagctc agagagcatt
1981 ggcctgcaga acttcgagat tgccaaggac ttcgtcgtca aggtcatcga ccggctgagc
2041 cgggacgagc tggtcaagtt cgagccaggg cagtcgtacg cgggtgtggt gcagtacagc
2101 cacagccaga tgcaggagca cgtgagcctg cgcagcccca gcatccggaa cgtgcaggag
2161 ctcaaggaag ccatcaagag cctgcagtgg atggcgggcg gcaccttcac gggggaggcc
2221 ctgcagtaca cgcgggacca gctgctgccg cccagcccga acaaccgcat cgccctggtc
2281 atcactgacg ggcgctcaga cactcagagg gacaccacac cgctcaacgt gctctgcagc
2341 cccggcatcc aggtggtctc cgtgggcatc aaagacgtgt ttgacttcat cccaggctca
2401 gaccagctca atgtcatttc ttgccaaggc ctggcaccat cccagggccg gcccggcctc
2461 tcgctggtca aggagaacta tgcagagctg ctggaggatg ccttcctgaa gaatgtcacc
2521 gcccagatct gcatagacaa gaagtgtcca gattacacct gccccatcac gttctcctcc
2581 ccggctgaca tcaccatcct gctggacggc tccgccagcg tgggcagcca caactttgac
2641 accaccaagc gcttcgccaa gcgcctggcc gagcgcttcc tcacagcggg caggacggac
2701 cccgcccacg acgtgcgggt ggcggtggtg cagtacagcg gcacgggcca gcagcgccca
2761 gagcgggcgt cgctgcagtt cctgcagaac tacacggccc tggccagtgc cgtcgatgcc
2821 atggacttta tcaacgacgc caccgacgtc aacgatgccc tgggctatgt gacccgcttc
2881 taccgcgagg cctcgtccgg cgctgccaag aagaggctgc tgctcttctc agatggcaac
2941 tcgcagggcg ccacgcccgc tgccatcgag aaggccgtgc aggaagccca gcgggcaggc

```
     3001 atcgagatct tcgtggtggt cgtgggccgc caggtgaatg agccccacat
ccgcgtcctg
     3061 gtcaccggca agacggccga gtacgacgtg gcctacggcg agagccacct
gttccgtgtc
     3121 cccagctacc aggccctgct ccgcggtgtc ttccaccaga cagtctccag
gaaggtggcg
     3181 ctgggctagc ccaccctgca cgccggcacc aaaccctgtc ctcccacccc
tccccactca
     3241 tcactaaaca gagtaaaatg tgatgcgaat tttcccgacc aacctgattc
gctagatttt
     3301 ttttaaggaa aagcttggaa agccaggaca caacgctgct gcctgctttg
tgcagggtcc
     3361 tccggggctc agccctgagt tggcatcacc tgcgcagggc cctctggggc
tcagccctga
     3421 gctagtgtca cctgcacagg gccctctgag gctcagccct gagctggcgt
cacctgtgca
     3481 gggccctctg gggctcagcc ctgagctggc ctcacctggg ttccccaccc
cgggctctcc
     3541 tgccctgccc tcctgcccgc cctccctcct gcctgcgcag ctccttccct
aggcacctct
     3601 gtgctgcatc ccaccagcct gagcaagacg ccctctcggg gcctgtgccg
cactagcctc
     3661 cctctcctct gtccccatag ctggtttttc ccaccaatcc tcacctaaca
gttactttac
     3721 aattaaactc aaagcaagct cttctcctca gcttggggca gccattggcc
tctgtctcgt
     3781 tttgggaaac caaggtcagg aggccgttgc agacataaat ctcggcgact
cggccccgtc
     3841 tcctgagggt cctgctggtg accggcctgg accttggccc tacagccctg
gaggccgctg
     3901 ctgaccagca ctgaccccga cctcagagag tactcgcagg ggcgctggct
gcactcaaga
     3961 ccctcgagat taacggtgct aaccccgtct gctcctccct cccgcagaga
ctggggcctg
     4021 gactggacat gagagcccct tggtgccaca gagggctgtg tcttactaga
aacaacgcaa
     4081 acctctcctt cctcagaata gtgatgtgtt cgacgtttta tcaaaggccc
cctttctatg
     4141 ttcatgttag ttttgctcct tctgtgtttt tttctgaacc atatccatgt
tgctgacttt
     4201 tccaaataaa ggttttcact cctctaaaaa aaaaaaaaaa aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001839.2

LOCUS NP_001839

ACCESSION NP_001839

```
      1 mraarallpl llqacwtaaq depetprava fqdcpvdlff vldtsesval
rlkpygalvd
       61 kvksftkrfi dnlrdryyrc drnlvwnaga lhysdeveii qgltrmpggr
dalkssvdav
      121 kyfgkgtytd caikkgleql lvggshlken kylivvtdgh plegykepcg
gledavneak
```

```
      181 hlgvkvfsva itpdhleprl siiatdhtyr rnftaadwgq srdaeeaisq
tidtivdmik
      241 nnveqvccsf ecqpargppg lrgdpgfege rgkpglpgek geagdpgrpg
dlgpvgyqgm
      301 kgekgsrgek gsrgpkgykg ekgkrgidgv dgvkgemgyp glpgckgspg
fdgiqgppgp
      361 kgdpgafglk gekgepgadg eagrpgssgp sgdegqpgep gppgekgeag
degnpgpdga
      421 pgerggpger gprgtpgtrg prgdpgeagp qgdqgregpv gvpgdpgeag
pigpkgyrgd
      481 egppgsegar gapgpagppg dpglmgerge dgpagngteg fpgfpgypgn
rgapgingtk
      541 gypglkgdeg eagdpgddnn diaprgvkga kgyrgpegpq gppghqgppg
pdeceildii
      601 mkmcscceck cgpidllfvl dssesiglqn feiakdfvvk vidrlsrdel
vkfepgqsya
      661 gvvqyshsqm qehvslrsps irnvqelkea ikslqwmagg tftgealqyt
rdqllppspn
      721 nrialvitdg rsdtqrdttp lnvlcspgiq vvsvgikdvf dfipgsdqln
viscqglaps
      781 qgrpglslvk enyaelleda flknvtaqic idkkcpdytc pitfsspadi
tilldgsasv
      841 gshnfdttkr fakrlaerfl tagrtdpahd vrvavvqysg tgqqrperas
lqflqnytal
      901 asavdamdfi ndatdvndal gyvtrfyrea ssgaakkrll lfsdgnsqga
tpaaiekavq
      961 eaqragieif vvvvgrqvne phirvlvtgk taeydvayge shlfrvpsyq
allrgvfhqt
     1021 vsrkvalg
```

CRTAP

Official Symbol: CRTAP

Official Name: cartilage associated protein

Gene ID: 10491

Organism: Homo sapiens

Other Aliases: CASP, LEPREL3, OI7

Other Designations: cartilage-associated protein; leprecan-like 3

Nucleotide sequence:

NCBI Reference Sequence: NM_006371.4

LOCUS NM_006371

ACCESSION NM_006371

```
      1 aggctggcgt ccccgccccg aaagcactgg gcccgccgcg tcgcaccgtc
ctctttcctt
       61 tccttctccc tccccttttc ccttccttcg tcccttcctt ccttcctttc
gccgggcgcg
```

121 atggagccgg ggcgccgggg ggccgcggcg ctgctagcgc tgctgtgcgt
ggcctgcgcg
181 ctgcgcgccg ggcgcgccca atacgaacgc tacagcttcc gcagcttccc
acgggacgag
241 ctgatgccgc tcgagtcggc ctaccggcac gcgctggaca agtacagcgg
cgagcactgg
301 gccgagagcg tgggctacct ggagatcagc ctgcggctgc accgcttgct
gcgcgacagc
361 gaggccttct gccaccgcaa ctgcagcgcc gcgccgcagc ccgagcccgc
cgccggcctc
421 gccagctatc ccgagctgcg cctcttcggg ggcctgctgc gccgcgcgca
ctgcctcaag
481 cgctgcaagc agggcctgcc agccttccgc cagtcccagc ccagccgcga
ggtgctggcg
541 gacttccagc gccgcgagcc ctacaagttc ctgcagttcg cttacttcaa
ggcaaataat
601 ctccccaaag ccatcgccgc tgctcacacc tttctactga agcatcctga
tgacgaaatg
661 atgaagagga acatggcata ttataagagc ctgcctggtg ccgaggacta
cattaaagac
721 ctggaaacca agtcatatga aagcctgttc atccgagcag tgcgggcata
caacggtgag
781 aactggagaa catccatcac agacatggag ctggcccttc ccgacttctt
caaagccttt
841 tacgagtgtc tcgcagcctg cgagggttcc agggagatca aggacttcaa
ggatttctac
901 ctttccatag cagatcatta tgtagaagtt ctggaatgca aaatacagtg
tgaagagaac
961 ctcaccccag ttataggagg ctatccggtt gagaaatttg tggctaccat
gtatcattac
1021 ttgcagtttg cctattataa gttgaacgac ctgaagaatg cagccccctg
tgcagtcagc
1081 tatctgctct ttgatcagaa tgacaaggtc atgcagcaga acctggtgta
ttaccagtac
1141 cacagggaca cttggggcct ctcggatgag cacttccagc ccagacctga
agcagttcag
1201 ttctttaatg tgaccacact ccagaaggag ctgtatgact ttgctaagga
aaatataatg
1261 gatgatgatg agggagaagt tgtggaatat gtggatgacc tcttggaact
ggaggagacc
1321 agctagccca cagcaaccaa agagacttcc tcttggcgtt caggaaacac
agattctttg
1381 tccttttccc aacagcccag gctgttgata cctcagagcc ttctctttac
tctccaaagt
1441 gaaagggaag cccccgtctc tctaactgca tgtcatcagg ggtgagcctg
cctttcctat
1501 cttcacacct gccacctcat gttcacacct atctttctca cctttttttt
gagatggagt
1561 ctcgctctct tgcccaggct ggagtgcaat ggcacgttct cagctcactg
caacctccgc
1621 ctcttgggtt caagcaattc tgctgcatca gcctcccgag tacctgggat
tacaggcatg
1681 tgccaccacg cccggctaat tttgtatttt tagtagagac ggggttttgc
catgttggcc
1741 aggctggtct cgaactcttg acttcagatg atccatctgc cttggcctcc
cacagtgctg
1801 ggattacagg cgtgagccac catgcccggc ctctttctca cctttacacc
tgtcttctta
1861 tcctcacatc tgttttcaca ccttcatccc tgtcttcctc atgttcacac
ttgtcttccc

1921 catgttcata gctgcctttc ttaccatttt ggtttgaagg gcagtcttct
ctggcttgtt
1981 tttttgtttt tcccagaaaa tcagtattat tttttaaata agaaaaacat
tcctagaaga
2041 tgataattgt gaaaacctcc tttggcttat ttgcttttcc agattttagt
ctcctttctc
2101 cccatccggg aaagatggtg gaagacatag gctaaatttc tccagcctca
caatggtctt
2161 cacttggtct gacttgtacc aattctagca cccactgaaa aacaagttga
gtagagagtg
2221 tagagtgcag aaatgtggct tttgccccac tttgcatctc caaaattaca
acggttggcc
2281 gatcccattt gaggacaatg cttagttata agtctccgag ttggaaaagg
aagaaagcca
2341 gagctgtcta gtttcattca ttctttcagt aaatatttat tgagtaccta
ctgtgtgcta
2401 ggcattgacc tgggaactag agatacttca cagaataaca gggaaagttc
cctgtgctca
2461 tggagcttac attctacagg gagaaagaga tagccaatac ataggaataa
atatatacaa
2521 ggtatcatgt agtgataatt gctgtggaga aaaataaagc aggggaggga
gtaagaaatc
2581 ctggagatga ggctgcagtt ttaaatgggg cctcactggg aatgtgacgt
tgagcagaga
2641 cgttagggaa gtggatcctg gacaaggcat tccaggcaga ggaacaggat
gtgcactgcc
2701 ccaaagtgag aacttgctct acgtggtcag gaaagagcag ggagaccaag
cagagtcgtg
2761 ggcaggggta gaatggaagg agaggcggct ggggaggaca ggtggtggag
ggccttggct
2821 tctgctaagt gagatgggaa ccactggagg gtttgaacag aggagtgcct
tgattgattt
2881 atattttgca agggtcattc tagctgccat attgtgaaaa actttagtgg
acaagggcag
2941 aaggaagagg gaagacctgt taggaagcta ctgcaaggtt ccaggcttgg
gcctgggcca
3001 cagcaacagc agtggtcaaa tatctagatt tattttgaaa agagccaata
ggatttgctg
3061 agagtttgaa tgtggagtgt aagagaagga agagttaatg atgacattaa
ggtttttggc
3121 ctgaatagca ggaaagatgg agttaccagt tactgaaata gggaaggatg
ggctgggtaa
3181 gtatggaatt tggtgcaaag caggctgtct gtggttggaa tgggaggttc
tggctgcaaa
3241 tcaaagtgga gagttctctc aggtcaggtc tgcagcagag ctcgagacag
ggatctgaat
3301 gcacttggtt tattgttggg ggtgctctca gaaggaacct gtgaaagcct
ttatcagtca
3361 tttattggct gtgagaagtt ctctgggagt gtgggtacat ttgaaggcaa
gtgacttcag
3421 ttgagggcaa gtctctggaa aagaggctgt aggcatctgg cagctaccat
gcatggtagt
3481 gtgttggggg tgggggtcct gggcactggc tgtgtgaagg gatctggcag
ggcaccacag
3541 cgccccctac tgaaccatca gcatgtcagt ggcatttaaa gccatgcagc
tggaggggcc
3601 actgagattg tctctgagta ttactgagaa gcaacagaaa agagccatgg
atggagccct
3661 tgggctctct gggaaatggg aaatcagcca aaggactgag aaggagttac
cttaaggtca

3721 gagaaaacca agagagtgtg gtgttctgga agctgagctt tctttattca
acctcattcc
3781 cttctccaaa taagccactt gtgtagttgg gcccctccag ggttgaaggc
aagaggagaa
3841 aggcacagcg tttgggaaac aagacttttc ctgcaatagc ctgggaagga
ataaaaggat
3901 agagtgtttg ggtttttgtg taatggtggt taattggggt ggaacactca
cacgttgtgc
3961 tttttctggg cttcccttat cccccagaac actctaccaa cctcggggaa
ctcgggcaca
4021 tccttctgtt tctccttcag ctctatcctg ctttcctcat cccttctgac
accacgtcct
4081 cactcacctg cacaagaatc cctgcatcag gttctccttt gagggtaccc
acccaggaca
4141 gtcccctacc acttctgtct tgggctgaag ttgcccacgt ccacaaaatc
tgtactccca
4201 gcgggggtgt ttggcccgag gagtcagtgt tattactggt ggatgcaccg
tgtccacagc
4261 agcccccaat cccagcgatg cgtcagatct tacgtggctt cctgctgggg
gagatggcct
4321 tcacccacgg gatgccgggt tctcctttct ttcctcaccc caacctttac
tccaccagag
4381 aaacttcctt ttgaactcag tggggaagag ggtgatgaga caggactaga
aagtagtggg
4441 ggacccagcg agtggacgcc ctgctccggg attcctgagt ctgtaaatag
tgtgcccagc
4501 agctgtgaac tccccttata gcctcaggct gcagtgtcct tcccagctgt
gtgagaaaat
4561 gaaagccgac gtccacaggg acccaggcag ggttgggtgt tgtgactcac
tccacctctg
4621 tgccctgcag aggtactgtt gggtccttgt cttgtgagcc tggggtgagc
tctctgtaca
4681 tgttgttgtt ccacgtatgg gttgacttgg catgctgggg ggtcctcgtt
cactctctga
4741 agttggcctc ctttcactgg ggattgaaaa gcacctccac ccctacccta
gtgatgtccc
4801 ctgaggaccc gggtgatagt acagtcaata ttgtcagtac tttgctttga
ttgaaggctg
4861 tagagctgag ttaccaaaat ttctatttca aaggaaacca aaccttaaaa
aaaaaaaaca
4921 aaaactgggc tgggtcttcc aaacctacca tgaaaccctg gtgtgcaggc
tgcactcaat
4981 gacctcaacc caacacctcc ctgagtgtgc ttcttggaag agcctagaag
attcctggat
5041 ggagacccca ttggttcagc ctcaagtctg gcccgtcttc gaaaaaacaa
acacatttgt
5101 aagctttgtg ggagcttcca ggcctgctct aagatgcctt gcttgtcctt
tgacccatca
5161 gcatggagct cagtggttgc tgtttggttc tgcaggctgg tggggaggcc
gcccatcgtg
5221 gtggggcatc tgtccagccc cattgccact cagggcatcc aaacaggagg
cacccgctgg
5281 gaagggtcta aagatactcc ttgtggccac tgctactgtt cacacttgac
ttgtggagaa
5341 gcgaagggct gaggggaggt ttgtgtacac ccatgtattt aaaagtgact
gactgactga
5401 aatgagcaca taccgacata tgcaacatac taataccttc ctgattttcg
agactttcta
5461 attactacaa ctaacctgtt gtgctcacct ctggaattca gaaagagagc
cactgcgagc

```
     5521 actgaccaca agggctgcct taggaaggaa atgtgtgctt tcaggagttc
ttattagggg
     5581 aattttaaga gcacaaaaat tttattgaac ccaccttagt gacaaacaga
aaatgatttg
     5641 ttagattgtc agttggaagg ggtttattat gactgctggt gaaataaaca
gtgaccagtt
     5701 tctgccaaca tttatggaaa taacgtttct aggttttaaa tgtgagccgt
aaactgaagc
     5761 tacttcagtt aaaaaaaaaa aaatcaatac ttaaactgta gggaaaaggt
ggattggtgc
     5821 cagagaaaac attatttaat agtgtcagaa catggaactg caacacagtt
tgtagcaagg
     5881 cactgagaaa aacccacaac taatccttca tcgcagctgt ctgaactctt
gatcatctgc
     5941 catcccccaa acagtgactt ctttttttct ggtgactcca ggcctgaatg
acctagtgtg
     6001 gagagtttaa actatgtaca agggaggaaa gaaaaaaagg aaaggaacct
taagtaagcc
     6061 tcagccaaag gttttatgca ttggacttcc tgtgcctgcc cccaggggca
agactgatcc
     6121 ccatgcctgt gcccatgaca tctccctgaa agaggacacc atgacagccc
ggctttgcct
     6181 tgactgaccc actgctaccc cagaaataag aatcaagagc agctattgtt
atccttagag
     6241 tgttttccgt ctagggccgg ctcgtgaaca gccacatatc cttgcacctg
acactgtccc
     6301 cacacaaata gctggctttc gttgcttgtt gaatgaatga gtgagttggc
tctatatccc
     6361 cttggagctg gccggtaaga tattagtgcc tcattttaca agaagagaat
aggaaggaat
     6421 taagcaattt ggccaacaga tacaagatag attccagagt tttatctccc
actttagggt
     6481 ggcagccagt aggccaaact ccaaagaccg ttgctgatgt ctttttctgc
ctcccctctt
     6541 tgggttagtg tggtatgtac aagctcactc ttgttgaaaa ttagaaaata
gttgaaaaca
     6601 aaaggttttt gtttttcttt ttaaatcaca ttaaatgttt tacattgctt
aaaaaaaaaa
     6661 aaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_006362.1

LOCUS NP_006362

ACCESSION NP_006362

```
      1 mepgrrgaaa llallcvaca lragraqyer ysfrsfprde lmplesayrh
aldkysgehw
       61 aesvgyleis lrlhrllrds eafchrncsa apqpepaagl asypelrlfg
gllrrahclk
      121 rckqglpafr qsqpsrevla dfqrrepykf lqfayfkann lpkaiaaaht
fllkhpddem
      181 mkrnmayyks lpgaedyikd letksyeslf iravraynge nwrtsitdme
lalpdffkaf
      241 yeclaacegs reikdfkdfy lsiadhyvev leckiqceen ltpviggypv
ekfvatmyhy
```

```
      301 lqfayyklnd lknaapcavs yllfdqndkv mqqnlvyyqy hrdtwglsde
hfqprpeavq
      361 ffnvttlqke lydfakenim dddegevvey vddlleleet s
```

SERPH

Official Symbol: SERPINH1

Official Name: serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)

Gene ID: 871

Organism: Homo sapiens

Other Aliases: PIG14, AsTP3, CBP1, CBP2, HSP47, OI10, PPROM, RA-A47, SERPINH2, gp46

Other Designations: 47 kDa heat shock protein; arsenic-transactivated protein 3; cell proliferation-inducing gene 14 protein; colligin-1; colligin-2; rheumatoid arthritis antigen A-47; rheumatoid arthritis-related antigen RA-A47; serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1); serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 2, (collagen-binding protein 2); serpin H1

Nucleotide sequence:

NCBI Reference Sequence: NM_001207014.1

LOCUS NM_001207014

ACCESSION NM_001207014

```
      1 agtaggaccc aggggccggg aggcgccggc agagggaggg gccgggggcc
ggggaggttt
       61 tgagggaggt ctttggcttt ttttggcgga gctggggcgc cctccggaag
cgtttccaac
      121 tttccagaag tttctcggga cgggcaggag ggggtgggga ctgccatata
tagatcccgg
      181 gagcagggga gcgggctaag agtagaatcg tgtcgcggct cgagagcgag
agtcacgtcc
      241 cggcgctagc ccagcccgac ccagaatgaa aaaggcaggc attgacctcc
ctctgaggca
      301 gtttccaggc ccaccgtggt gcacgcaaac cacttcctgg ccatgcgctc
cctcctgctt
      361 ctcagcgcct tctgcctcct ggaggcggcc ctggccgccg aggtgaagaa
acctgcagcc
      421 gcagcagctc ctggcactgc ggagaagttg agccccaagg cggccacgct
tgccgagcgc
      481 agcgccggcc tggccttcag cttgtaccag gccatggcca aggaccaggc
agtggagaac
      541 atcctggtgt cacccgtggt ggtggcctcg tcgctagggc tcgtgtcgct
gggcggcaag
      601 gcgaccacgg cgtcgcaggc caaggcagtg ctgagcgccg agcagctgcg
cgacgaggag
```

661 gtgcacgccg gcctgggcga gctgctgcgc tcactcagca actccacggc
gcgcaacgtg
721 acctggaagc tgggcagccg actgtacgga cccagctcag tgagcttcgc
tgatgacttc
781 gtgcgcagca gcaagcagca ctacaactgc gagcactcca agatcaactt
ccgcgacaag
841 cgcagcgcgc tgcagtccat caacgagtgg gccgcgcaga ccaccgacgg
caagctgccc
901 gaggtcacca aggacgtgga gcgcacggac ggcgccctgc tagtcaacgc
catgttcttc
961 aagccacact gggatgagaa attccaccac aagatggtgg acaaccgtgg
cttcatggtg
1021 actcggtcct ataccgtggg tgtcatgatg atgcaccgga caggcctcta
caactactac
1081 gacgacgaga aggaaaagct gcaaatcgtg gagatgcccc tggcccacaa
gctctccagc
1141 ctcatcatcc tcatgcccca tcacgtggag cctctcgagc gccttgaaaa
gctgctaacc
1201 aaagagcagc tgaagatctg gatggggaag atgcagaaga aggctgttgc
catctccttg
1261 cccaagggtg tggtggaggt gacccatgac ctgcagaaac acctggctgg
gctgggcctg
1321 actgaggcca ttgacaagaa caaggccgac ttgtcacgca tgtcaggcaa
gaaggacctg
1381 tacctggcca gcgtgttcca cgccaccgcc tttgagttgg acacagatgg
caaccccttt
1441 gaccaggaca tctacgggcg cgaggagctg cgcagcccca agctgttcta
cgccgaccac
1501 cccttcatct tcctagtgcg ggacacccaa agcggctccc tgctattcat
tgggcgcctg
1561 gtccggccta agggtgacaa gatgcgagac gagttatagg gcctcagggt
gcacacagga
1621 tggcaggagg catccaaagg ctcctgagac acatgggtgc tattggggtt
ggggggggagg
1681 tgaggtacca gccttggata ctccatgggg tgggggtgga aaaacagacc
ggggttcccg
1741 tgtgcctgag cggaccttcc cagctagaat tcactccact tggacatggg
ccccagatac
1801 catgatgctg agcccggaaa ctccacatcc tgtgggacct gggccatagt
cattctgcct
1861 gccctgaaag tcccagatca agcctgcctc aatcagtatt catatttata
gccaggtacc
1921 ttctcacctg tgagaccaaa ttgagctagg ggggtcagcc agccctcttc
tgacactaaa
1981 acacctcagc tgcctcccca gctctatccc aacctctccc aactataaaa
ctaggtgctg
2041 cagcccctgg gaccaggcac ccccagaatg acctggccgc agtgaggcgg
attgagaagg
2101 agctcccagg aggggcttct gggcagactc tggtcaagaa gcatcgtgtc
tggcgttgtg
2161 gggatgaact ttttgttttg tttcttcctt ttttagttct tcaaagatag
ggagggaagg
2221 gggaacatga gcctttgttg ctatcaatcc aagaacttat ttgtacattt
ttttttttcaa
2281 taaaactttt ccaatgacat tttgttggag cgtggaagaa aaaaaaaaaa aaa

Protein sequence:

NCBI Reference Sequence: NP_001193943.1

LOCUS NP_001193943

ACCESSION NP_001193943

```
      1 mrsllllsaf clleaalaae vkkpaaaaap gtaeklspka atlaersagl
afslyqamak
       61 dqavenilvs pvvvasslgl vslggkatta sqakavlsae qlrdeevhag
lgellrslsn
      121 starnvtwkl gsrlygpssv sfaddfvrss kqhyncehsk infrdkrsal
qsinewaaqt
      181 tdgklpevtk dvertdgall vnamffkphw dekfhhkmvd nrgfmvtrsy
tvgvmmmhrt
      241 glynyyddek eklqivempl ahklssliil mphhvepler leklltkeql
kiwmgkmqkk
      301 avaislpkgv vevthdlqkh laglglteai dknkadlsrm sgkkdlylas
vfhatafeld
      361 tdgnpfdqdi ygreelrspk lfyadhpfif lvrdtqsgsl lfigrlvrpk
gdkmrdel
```

ITB1

Official Symbol: ITGB1

Official Name: integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)

Gene ID: 3688

Organism: Homo sapiens

Other Aliases: RP11-479G22.2, CD29, FNRB, GPIIA, MDF2, MSK12, VLA-BETA, VLAB

Other Designations: integrin VLA-4 beta subunit; integrin beta-1; very late activation protein, beta polypeptide

Nucleotide sequence:

NCBI Reference Sequence: NM_002211.3

LOCUS NM_002211

ACCESSION NM_002211

```
      1 atcagacgcg cagaggaggc ggggccgcgg ctggtttcct gccggggggc
ggctctgggc
       61 cgccgagtcc cctcctcccg cccctgagga ggaggagccg ccgccacccg
ccgcgcccga
      121 cacccgggag gccccgccag cccgcgggag aggcccagcg ggagtcgcgg
aacagcaggc
      181 ccgagcccac cgcgccgggc cccggacgcc gcgcggaaaa gatgaattta
caaccaattt
      241 tctggattgg actgatcagt tcagtttgct gtgtgtttgc tcaaacagat
gaaaatagat
      301 gtttaaaagc aaatgccaaa tcatgtggag aatgtataca agcagggcca
aattgtgggt
```

361 ggtgcacaaa ttcaacattt ttacaggaag gaatgcctac ttctgcacga
tgtgatgatt
421 tagaagcctt aaaaaagaag ggttgccctc cagatgacat agaaaatccc
agaggctcca
481 aagatataaa gaaaaataaa aatgtaacca accgtagcaa aggaacagca
gagaagctca
541 agccagagga tattactcag atccaaccac agcagttggt tttgcgatta
agatcagggg
601 agccacagac atttacatta aaattcaaga gagctgaaga ctatcccatt
gacctctact
661 accttatgga cctgtcttac tcaatgaaag acgatttgga gaatgtaaaa
agtcttggaa
721 cagatctgat gaatgaaatg aggaggatta cttcggactt cagaattgga
tttggctcat
781 ttgtggaaaa gactgtgatg ccttacatta gcacaacacc agctaagctc
aggaaccctt
841 gcacaagtga acagaactgc accagcccat ttagctacaa aaatgtgctc
agtcttacta
901 ataaaggaga agtatttaat gaacttgttg gaaaacagcg catatctgga
aatttggatt
961 ctccagaagg tggtttcgat gccatcatgc aagttgcagt ttgtggatca
ctgattggct
1021 ggaggaatgt tacacggctg ctggtgtttt ccacagatgc cgggtttcac
tttgctggag
1081 atgggaaact tggtggcatt gttttaccaa atgatggaca atgtcacctg
gaaaataata
1141 tgtacacaat gagccattat tatgattatc cttctattgc tcaccttgtc
cagaaactga
1201 gtgaaaataa tattcagaca atttttgcag ttactgaaga atttcagcct
gtttacaagg
1261 agctgaaaaa cttgatccct aagtcagcag taggaacatt atctgcaaat
tctagcaatg
1321 taattcagtt gatcattgat gcatacaatt ccctttcctc agaagtcatt
ttggaaaacg
1381 gcaaattgtc agaaggcgta acaataagtt acaaatctta ctgcaagaac
ggggtgaatg
1441 gaacagggga aaatggaaga aaatgttcca atatttccat tggagatgag
gttcaatttg
1501 aaattagcat aacttcaaat aagtgtccaa aaaaggattc tgacagcttt
aaaattaggc
1561 ctctgggctt tacggaggaa gtagaggtta ttcttcagta catctgtgaa
tgtgaatgcc
1621 aaagcgaagg catccctgaa agtcccaagt gtcatgaagg aaatgggaca
tttgagtgtg
1681 gcgcgtgcag gtgcaatgaa gggcgtgttg gtagacattg tgaatgcagc
acagatgaag
1741 ttaacagtga agacatggat gcttactgca ggaaagaaaa cagttcagaa
atctgcagta
1801 acaatggaga gtgcgtctgc ggacagtgtg tttgtaggaa gagggataat
acaaatgaaa
1861 tttattctgg caaattctgc gagtgtgata atttcaactg tgatagatcc
aatggcttaa
1921 tttgtggagg aaatggtgtt tgcaagtgtc gtgtgtgtga gtgcaacccc
aactacactg
1981 gcagtgcatg tgactgttct ttggatacta gtacttgtga agccagcaac
ggacagatct
2041 gcaatggccg gggcatctgc gagtgtggtg tctgtaagtg tacagatccg
aagtttcaag
2101 ggcaaacgtg tgagatgtgt cagacctgcc ttggtgtctg tgctgagcat
aaagaatgtg

2161 ttcagtgcag agccttcaat aaaggagaaa agaaagacac atgcacacag
gaatgttcct
2221 attttaacat taccaaggta gaaagtcggg acaaattacc ccagccggtc
caacctgatc
2281 ctgtgtccca ttgtaaggag aaggatgttg acgactgttg gttctatttt
acgtattcag
2341 tgaatgggaa caacgaggtc atggttcatg ttgtggagaa tccagagtgt
cccactggtc
2401 cagacatcat tccaattgta gctggtgtgg ttgctggaat tgttcttatt
ggccttgcat
2461 tactgctgat atggaagctt ttaatgataa ttcatgacag aagggagttt
gctaaatttg
2521 aaaaggagaa aatgaatgcc aaatgggaca cgggtgaaaa tcctatttat
aagagtgccg
2581 taacaactgt ggtcaatccg aagtatgagg gaaaatgagt actgcccgtg
caaatcccac
2641 aacactgaat gcaaagtagc aatttccata gtcacagtta ggtagcttta
gggcaatatt
2701 gccatggttt tactcatgtg caggttttga aaatgtacaa tatgtataat
ttttaaaatg
2761 ttttattatt ttgaaaataa tgttgtaatt catgccaggg actgacaaaa
gacttgagac
2821 aggatggtta ctcttgtcag ctaaggtcac attgtgcctt tttgaccttt
tcttcctgga
2881 ctattgaaat caagcttatt ggattaagtg atatttctat agcgattgaa
agggcaatag
2941 ttaaagtaat gagcatgatg agagtttctg ttaatcatgt attaaaactg
atttttagct
3001 ttacaaatat gtcagtttgc agttatgcag aatccaaagt aaatgtcctg
ctagctagtt
3061 aaggattgtt ttaaatctgt tattttgcta tttgcctgtt agacatgact
gatgacatat
3121 ctgaaagaca agtatgttga gagttgctgg tgtaaaatac gtttgaaata
gttgatctac
3181 aaaggccatg ggaaaaattc agagagttag gaaggaaaaa ccaatagctt
taaaacctgt
3241 gtgccatttt aagagttact taatgtttgg taacttttat gccttcactt
tacaaattca
3301 agccttagat aaaagaaccg agcaattttc tgctaaaaag tccttgattt
agcactattt
3361 acatacaggc catactttac aaagtatttg ctgaatgggg acctttttgag
ttgaatttat
3421 tttattattt ttattttgtt taatgtctgg tgctttctgt cacctcttct
aatcttttaa
3481 tgtatttgtt tgcaattttg gggtaagact ttttttatga gtacttttttc
tttgaagttt
3541 tagcggtcaa tttgcctttt taatgaacat gtgaagttat actgtggcta
tgcaacagct
3601 ctcacctacg cgagtcttac tttgagttag tgccataaca gaccactgta
tgtttacttc
3661 tcaccatttg agttgcccat cttgtttcac actagtcaca ttcttgtttt
aagtgccttt
3721 agttttaaca gttcactttt tacagtgcta tttactgaag ttatttatta
aatatgccta
3781 aaatacttaa atcggatgtc ttgactctga tgtattttat caggttgtgt
gcatgaaatt
3841 tttatagatt aaagaagttg aggaaaagca aaaaaaaaa

Protein sequence:

NCBI Reference Sequence: NP_002202.2

LOCUS NP_002202

ACCESSION NP_002202

```
      1 mnlqpifwig lissvccvfa qtdenrclka nakscgeciq agpncgwctn
stflqegmpt
       61 sarcddleal kkkgcppddi enprgskdik knknvtnrsk gtaeklkped
itqiqpqqlv
      121 lrlrsgepqt ftlkfkraed ypidlyylmd lsysmkddle nvkslgtdlm
nemrritsdf
      181 rigfgsfvek tvmpyisttp aklrnpctse qnctspfsyk nvlsltnkge
vfnelvgkqr
      241 isgnldspeg gfdaimqvav cgsligwrnv trllvfstda gfhfagdgkl
ggivlpndgq
      301 chlennmytm shyydypsia hlvqklsenn iqtifavtee fqpvykelkn
lipksavgtl
      361 sanssnviql iidaynslss evilengkls egvtisyksy ckngvngtge
ngrkcsnisi
      421 gdevqfeisi tsnkcpkkds dsfkirplgf teevevilqy icececqseg
ipespkcheg
      481 ngtfecgacr cnegrvgrhc ecstdevnse dmdaycrken sseicsnnge
cvcgqcvcrk
      541 rdntneiysg kfcecdnfnc drsnglicgg ngvckcrvce cnpnytgsac
dcsldtstce
      601 asngqicngr gicecgvckc tdpkfqgqtc emcqtclgvc aehkecvqcr
afnkgekkdt
      661 ctqecsyfni tkvesrdklp qpvqpdpvsh ckekdvddcw fyftysvngn
nevmvhvven
      721 pecptgpdii pivagvvagi vliglallli wkllmiihdr refakfekek
mnakwdtgen
      781 piyksavttv vnpkyegk
```

FKB10

Official Symbol: FKBP10

Official Name: FK506 binding protein 10, 65 kDa

Gene ID: 60681

Organism: Homo sapiens

Other Aliases: PSEC0056, FKBP65, OI11, OI6, PPIASE, hFKBP65

Other Designations: 65 kDa FK506-binding protein; 65 kDa FKBP; FK506-binding protein 10; FKBP-10; FKBP-65; PPIase **FKBP10**; immunophilin FKBP65; peptidyl-prolyl cis-trans isomerase **FKBP10**; rotamase

Nucleotide sequence:

NCBI Reference Sequence: NM_021939.3

LOCUS NM_021939

ACCESSION NM_021939

```
      1 cccgagcctc tctccctggc caggccccag gtctcgcagc cagggatgga
gatgggggga
       61 gggggaacct agagttcttt gtagtgcctc cctcagactc taacacactc
agcctggccc
      121 cctcctccta ttgcaacccc ctcccccgct cctcccggcc aggccagctc
agtcttccca
      181 gcccccattc cacgtggacc agccagggcg ggggtaggga aagaggacag
gaagaggggg
      241 agccagttct gggaggcggg gggaaggagg ttggtggcga ctccctcgct
cgccctcact
      301 gccggcggtc ccaactccag gcaccatgtt ccccgcgggc ccccccagcc
acagcctcct
      361 ccggctcccc ctgctgcagt tgctgctact ggtggtgcag gccgtgggga
gggggctggg
      421 ccgcgccagc ccggccgggg gccccctgga agatgtggtc atcgagaggt
accacatccc
      481 cagggcctgt ccccgggaag tgcagatggg ggattttgtg cgctaccact
acaacggcac
      541 ttttgaagat ggcaagaagt ttgattcaag ctatgatcgc aacaccttgg
tggccatcgt
      601 ggtgggtgtg gggcgcctca tcactggcat ggaccgaggc ctcatgggca
tgtgtgtcaa
      661 cgagcggcga cgcctcattg tgcctcccca cctgggctat gggagcatcg
gcctggcggg
      721 gctcattcca ccggatgcca ccctctactt cgatgtggtt ctgctggatg
tgtggaacaa
      781 ggaagacacc gtgcaggtga gcacattgct gcgcccgccc cactgccccc
gcatggtcca
      841 ggacggcgac tttgtccgct accactacaa tggcaccctg ctggacggca
cctccttcga
      901 caccagctac agtaagggcg gcacttatga cacctacgtc ggctctggtt
ggctgatcaa
      961 gggcatggac caggggctgc tgggcatgtg tcctggagag agaaggaaga
ttatcatccc
     1021 tccattcctg gcctatggcg agaaaggcta tgggacagtg atccccccac
aggcctcgct
     1081 ggtctttcac gtcctcctga ttgacgtgca caacccgaag gacgctgtcc
agctagagac
     1141 gctggagctc ccccccggct gtgtccgcag agccggggcc ggggacttca
tgcgctacca
     1201 ctacaatggc tccttgatgg acggcaccct cttcgattcc agctactccc
gcaaccacac
     1261 ctacaatacc tatatcgggc agggttacat catccccggg atggaccagg
ggctgcaggg
     1321 tgcctgcatg ggggaacgcc ggagaattac catccccccg cacctcgcct
atggggagaa
     1381 tggaactgga gacaagatcc ctggctctgc cgtgctaatc ttcaacgtcc
atgtcattga
     1441 cttccacaac cctgcggatg tggtggaaat caggacactg tcccggccat
ctgagacctg
     1501 caatgagacc accaagcttg gggactttgt tcgataccat tacaactgtt
ctttgctgga
     1561 cggcacccag ctgttcacct cgcatgacta cggggccccc caggaggcga
ctctcggggc
     1621 caacaaggtg atcgaaggcc tggacacggg cctgcagggc atgtgtgtgg
gagagaggcg
     1681 gcagctcatc gtgcccccgc acctggccca cggggagagt ggagcccggg
gagtcccagg
     1741 cagtgctgtg ctgctgtttg aggtggagct ggtgtcccgg gaggatgggc
tgcccacagg
```

```
     1801 ctacctgttt gtgtggcaca aggaccctcc tgccaacctg tttgaagaca
tggacctcaa
     1861 caaggatggc gaggtccctc cggaggagtt ctccaccttc atcaaggctc
aagtgagtga
     1921 gggcaaagga cgcctcatgc ctgggcagga ccctgagaaa accataggag
acatgttcca
     1981 gaaccaggac cgcaaccagg acggcaagat cacagtcgac gagctcaagc
tgaagtcaga
     2041 tgaggacgag gagcgggtcc acgaggagct ctgaggggca gggagcctgg
ccaggcctga
     2101 gacacagagg cccactgcga gggggacagt ggcggtggga ctgacctgct
gacagtcacc
     2161 ctccctctgc tgggatgagg tccaggagcc aactaaaaca atggcagagg
agacatctct
     2221 ggtgttccca ccaccctaga tgaaaatcca cagcacagac ctctaccgtg
tttctcttcc
     2281 atccctaaac cacttcctta aaatgtttgg atttgcaaag ccaatttggg
gcctgtggag
     2341 cctggggttg gatagggcca tggctggtcc cccaccatac ctcccctcca
catcactgac
     2401 acagctgagc ttgttatcca tctccccaaa ctttctcttt ctttgtactt
cttgtcatcc
     2461 ccactcccag ccccttttcc tctatgtgac agctccctag gacccctctg
ccttcctccc
     2521 caatcctgac tggctcctag ggaaggggaa ggctcctgga gggcagccct
acctctccca
     2581 tgccctttgc cctcctccct cgcctccagt ggaggctgag ctgaccctgg
gctgctggag
     2641 gccagactgg gctgtagtta gcttttcatc cctaaagaag gctcctttcc
ctaaggaacc
     2701 atagaagaga ggaagaaaac aaagggcatg tgtgagggaa gctgcttggg
tgggtgttag
     2761 ggctatgaaa tcttggattt ggggctgagg ggtgggaggg agggcagagc
tctgcacact
     2821 caaaggctaa actggtgtca gtccttttt cctttgttcc aaataaaaga
ttaaaccaat
     2881 ggcaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_068758.3

LOCUS NP_068758

ACCESSION NP_068758

```
      1 mfpagppshs llrlpllqll llvvqavgrg lgraspaggp ledvvieryh
ipracprevq
       61 mgdfvryhyn gtfedgkkfd ssydrntlva ivvgvgrlit gmdrglmgmc
vnerrrlivp
      121 phlgygsigl aglippdatl yfdvvlldvw nkedtvqvst llrpphcprm
vqdgdfvryh
      181 yngtlldgts fdtsyskggt ydtyvgsgwl ikgmdqgllg mcpgerrkii
ippflaygek
      241 gygtvippqa slvfhvllid vhnpkdavql etlelppgcv rragagdfmr
yhyngslmdg
      301 tlfdssysrn htyntyigqg yiipgmdqgl qgacmgerrr itipphlayg
engtgdkipg
```

```
      361 savlifnvhv idfhnpadvv eirtlsrpse tcnettklgd fvryhyncsl
ldgtqlftsh
      421 dygapqeatl gankviegld tglqgmcvge rrqlivpphl ahgesgargv
pgsavllfev
      481 elvsredglp tgylfvwhkd ppanlfedmd lnkdgevppe efstfikaqv
segkgrlmpg
      541 qdpektigdm fqnqdrnqdg kitvdelklk sdedeervhe el
```

FINC

Official Symbol: FN1

Official Name: fibronectin 1

Gene ID: 2335

Organism: Homo sapiens

Other Aliases: CIG, ED-B, **FINC**, FN, FNZ, GFND, GFND2, LETS, MSF

Other Designations: cold-insoluble globulin; fibronectin; migration-stimulating factor

Nucleotide sequence:

NCBI Reference Sequence: NM_002026.2

LOCUS NM_002026

ACCESSION NM_002026

```
      1 gcccgcgccg gctgtgctgc acagggggag gagagggaac cccaggcgcg
agcgggaaga
       61 ggggacctgc agccacaact tctctggtcc tctgcatccc ttctgtccct
ccacccgtcc
      121 ccttccccac cctctggccc ccaccttctt ggaggcgaca acccccggga
ggcattagaa
      181 gggatttttc ccgcaggttg cgaagggaag caaacttggt ggcaacttgc
ctcccggtgc
      241 gggcgtctct cccccaccgt ctcaacatgc ttaggggtcc ggggcccggg
ctgctgctgc
      301 tggccgtcca gtgcctgggg acagcggtgc cctccacggg agcctcgaag
agcaagaggc
      361 aggctcagca aatggttcag ccccagtccc cggtggctgt cagtcaaagc
aagcccggtt
      421 gttatgacaa tggaaaacac tatcagataa atcaacagtg ggagcggacc
tacctaggca
      481 atgcgttggt ttgtacttgt tatggaggaa gccgaggttt taactgcgag
agtaaacctg
      541 aagctgaaga gacttgcttt gacaagtaca ctgggaacac ttaccgagtg
ggtgacactt
      601 atgagcgtcc taaagactcc atgatctggg actgtacctg catcggggct
gggcgaggga
```

661 gaataagctg taccatcgca aaccgctgcc atgaaggggg tcagtcctac
aagattggtg
721 acacctggag gagaccacat gagactggtg gttacatgtt agagtgtgtg
tgtcttggta
781 atggaaaagg agaatggacc tgcaagccca tagctgagaa gtgttttgat
catgctgctg
841 ggacttccta tgtggtcgga gaaacgtggg agaagcccta ccaaggctgg
atgatggtag
901 attgtacttg cctgggagaa ggcagcggac gcatcacttg cacttctaga
aatagatgca
961 acgatcagga cacaaggaca tcctatagaa ttggagacac ctggagcaag
aaggataatc
1021 gaggaaacct gctccagtgc atctgcacag gcaacggccg aggagagtgg
aagtgtgaga
1081 ggcacacctc tgtgcagacc acatcgagcg gatctggccc cttcaccgat
gttcgtgcag
1141 ctgtttacca accgcagcct cacccccagc ctcctcccta tggccactgt
gtcacagaca
1201 gtggtgtggt ctactctgtg gggatgcagt ggctgaagac acaaggaaat
aagcaaatgc
1261 tttgcacgtg cctgggcaac ggagtcagct gccaagagac agctgtaacc
cagacttacg
1321 gtggcaactc aaatggagag ccatgtgtct taccattcac ctacaatggc
aggacgttct
1381 actcctgcac cacagaaggg cgacaggacg gacatctttg gtgcagcaca
acttcgaatt
1441 atgagcagga ccagaaatac tctttctgca cagaccacac tgttttggtt
cagactcgag
1501 gaggaaattc caatggtgcc ttgtgccact tccccttcct atacaacaac
cacaattaca
1561 ctgattgcac ttctgagggc agaagagaca acatgaagtg gtgtgggacc
acacagaact
1621 atgatgccga ccagaagttt gggttctgcc ccatggctgc ccacgaggaa
atctgcacaa
1681 ccaatgaagg ggtcatgtac cgcattggag atcagtggga taagcagcat
gacatgggtc
1741 acatgatgag gtgcacgtgt gttgggaatg gtcgtgggga atggacatgc
attgcctact
1801 cgcagcttcg agatcagtgc attgttgatg acatcactta caatgtgaac
gacacattcc
1861 acaagcgtca tgaagagggg cacatgctga actgtacatg cttcggtcag
ggtcggggca
1921 ggtggaagtg tgatcccgtc gaccaatgcc aggattcaga gactgggacg
ttttatcaaa
1981 ttggagattc atgggagaag tatgtgcatg gtgtcagata ccagtgctac
tgctatggcc
2041 gtggcattgg ggagtggcat tgccaacctt tacagaccta tccaagctca
agtggtcctg
2101 tcgaagtatt tatcactgag actccgagtc agcccaactc ccaccccatc
cagtggaatg
2161 caccacagcc atctcacatt tccaagtaca ttctcaggtg gagacctaaa
aattctgtag
2221 gccgttggaa ggaagctacc ataccaggcc acttaaactc ctacaccatc
aaaggcctga
2281 agcctggtgt ggtatacgag ggccagctca tcagcatcca gcagtacggc
caccaagaag
2341 tgactcgctt tgacttcacc accaccagca ccagcacacc tgtgaccagc
aacaccgtga
2401 caggagagac gactcccttt tctcctcttg tggccacttc tgaatctgtg
accgaaatca

2461 cagccagtag ctttgtggtc tcctgggtct cagcttccga caccgtgtcg
ggattccggg
2521 tggaatatga gctgagtgag gagggagatg agccacagta cctggatctt
ccaagcacag
2581 ccacttctgt gaacatccct gacctgcttc ctggccgaaa atacattgta
aatgtctatc
2641 agatatctga ggatggggag cagagtttga tcctgtctac ttcacaaaca
acagcgcctg
2701 atgcccctcc tgacccgact gtggaccaag ttgatgacac ctcaattgtt
gttcgctgga
2761 gcagacccca ggctcccatc acagggtaca gaatagtcta ttcgccatca
gtagaaggta
2821 gcagcacaga actcaacctt cctgaaactg caaactccgt caccctcagt
gacttgcaac
2881 ctggtgttca gtataacatc actatctatg ctgtggaaga aaatcaagaa
agtacacctg
2941 ttgtcattca acaagaaacc actggcaccc cacgctcaga tacagtgccc
tctcccaggg
3001 acctgcagtt tgtggaagtg acagacgtga aggtcaccat catgtggaca
ccgcctgaga
3061 gtgcagtgac cggctaccgt gtggatgtga tccccgtcaa cctgcctggc
gagcacgggc
3121 agaggctgcc catcagcagg aacacctttg cagaagtcac cgggctgtcc
cctggggtca
3181 cctattactt caaagtcttt gcagtgagcc atgggaggga gagcaagcct
ctgactgctc
3241 aacagacaac caaactggat gctcccacta acctccagtt tgtcaatgaa
actgattcta
3301 ctgtcctggt gagatggact ccacctcggg cccagataac aggataccga
ctgaccgtgg
3361 gccttacccg aagaggacag cccaggcagt acaatgtggg tccctctgtc
tccaagtacc
3421 cactgaggaa tctgcagcct gcatctgagt acaccgtatc cctcgtggcc
ataaagggca
3481 accaagagag ccccaaagcc actggagtct ttaccacact gcagcctggg
agctctattc
3541 caccttacaa caccgaggtg actgagacca ccattgtgat cacatggacg
cctgctccaa
3601 gaattggttt taagctgggt gtacgaccaa gccagggagg agaggcacca
cgagaagtga
3661 cttcagactc aggaagcatc gttgtgtccg gcttgactcc aggagtagaa
tacgtctaca
3721 ccatccaagt cctgagagat ggacaggaaa gagatgcgcc aattgtaaac
aaagtggtga
3781 caccattgtc tccaccaaca aacttgcatc tggaggcaaa ccctgacact
ggagtgctca
3841 cagtctcctg ggagaggagc accaccccag acattactgg ttatagaatt
accacaaccc
3901 ctacaaacgg ccagcaggga aattctttgg aagaagtggt ccatgctgat
cagagctcct
3961 gcacttttga taacctgagt cccggcctgg agtacaatgt cagtgtttac
actgtcaagg
4021 atgacaagga aagtgtccct atctctgata ccatcatccc agctgttcct
cctcccactg
4081 acctgcgatt caccaacatt ggtccagaca ccatgcgtgt cacctgggct
ccacccccat
4141 ccattgattt aaccaacttc ctggtgcgtt actcacctgt gaaaaatgag
gaagatgttg
4201 cagagttgtc aatttctcct tcagacaatg cagtggtctt aacaaatctc
ctgcctggta

4261 cagaatatgt agtgagtgtc tccagtgtct acgaacaaca tgagagcaca
cctcttagag
4321 gaagacagaa aacaggtctt gattccccaa ctggcattga cttttctgat
attactgcca
4381 actcttttac tgtgcactgg attgctcctc gagccaccat cactggctac
aggatccgcc
4441 atcatcccga gcacttcagt gggagacctc gagaagatcg ggtgccccac
tctcggaatt
4501 ccatcaccct caccaacctc actccaggca cagagtatgt ggtcagcatc
gttgctctta
4561 atggcagaga ggaaagtccc ttattgattg gccaacaatc aacagtttct
gatgttccga
4621 gggacctgga agttgttgct gcgaccccca ccagcctact gatcagctgg
gatgctcctg
4681 ctgtcacagt gagatattac aggatcactt acggagagac aggaggaaat
agccctgtcc
4741 aggagttcac tgtgcctggg agcaagtcta cagctaccat cagcggcctt
aaacctggag
4801 ttgattatac catcactgtg tatgctgtca ctggccgtgg agacagcccc
gcaagcagca
4861 agccaatttc cattaattac cgaacagaaa ttgacaaacc atcccagatg
caagtgaccg
4921 atgttcagga caacagcatt agtgtcaagt ggctgccttc aagttcccct
gttactggtt
4981 acagagtaac caccactccc aaaaatggac caggaccaac aaaaactaaa
actgcaggtc
5041 cagatcaaac agaaatgact attgaaggct tgcagcccac agtggagtat
gtggttagtg
5101 tctatgctca gaatccaagc ggagagagtc agcctctggt tcagactgca
gtaaccaaca
5161 ttgatcgccc taaaggactg gcattcactg atgtggatgt cgattccatc
aaaattgctt
5221 gggaaagccc acaggggcaa gtttccaggt acagggtgac ctactcgagc
cctgaggatg
5281 gaatccatga gctattccct gcacctgatg gtgaagaaga cactgcagag
ctgcaaggcc
5341 tcagaccggg ttctgagtac acagtcagtg tggttgcctt gcacgatgat
atggagagcc
5401 agcccctgat tggaacccag tccacagcta ttcctgcacc aactgacctg
aagttcactc
5461 aggtcacacc cacaagcctg agcgcccagt ggacaccacc caatgttcag
ctcactggat
5521 atcgagtgcg ggtgaccccc aaggagaaga ccggaccaat gaaagaaatc
aaccttgctc
5581 ctgacagctc atccgtggtt gtatcaggac ttatggtggc caccaaatat
gaagtgagtg
5641 tctatgctct taaggacact ttgacaagca gaccagctca gggagttgtc
accactctgg
5701 agaatgtcag cccaccaaga agggctcgtg tgacagatgc tactgagacc
accatcacca
5761 ttagctggag aaccaagact gagacgatca ctggcttcca agttgatgcc
gttccagcca
5821 atggccagac tccaatccag agaaccatca agccagatgt cagaagctac
accatcacag
5881 gtttacaacc aggcactgac tacaagatct acctgtacac cttgaatgac
aatgctcgga
5941 gctcccctgt ggtcatcgac gcctccactg ccattgatgc accatccaac
ctgcgtttcc
6001 tggccaccac acccaattcc ttgctggtat catggcagcc gccacgtgcc
aggattaccg

6061 gctacatcat caagtatgag aagcctgggt ctcctcccag agaagtggtc
cctcggcccc
6121 gccctggtgt cacagaggct actattactg gcctggaacc gggaaccgaa
tatacaattt
6181 atgtcattgc cctgaagaat aatcagaaga gcgagcccct gattggaagg
aaaaagacag
6241 acgagcttcc ccaactggta acccttccac accccaatct tcatggacca
gagatcttgg
6301 atgttccttc cacagttcaa aagacccctt tcgtcaccca ccctgggtat
gacactggaa
6361 atggtattca gcttcctggc acttctggtc agcaacccag tgttgggcaa
caaatgatct
6421 ttgaggaaca tggttttagg cggaccacac cgcccacaac ggccaccccc
ataaggcata
6481 ggccaagacc atacccgccg aatgtaggac aagaagctct ctctcagaca
accatctcat
6541 gggccccatt ccaggacact tctgagtaca tcatttcatg tcatcctgtt
ggcactgatg
6601 aagaaccctt acagttcagg gttcctggaa cttctaccag tgccactctg
acaggcctca
6661 ccagaggtgc cacctacaac atcatagtgg aggcactgaa agaccagcag
aggcataagg
6721 ttcgggaaga ggttgttacc gtgggcaact ctgtcaacga aggcttgaac
caacctacgg
6781 atgactcgtg ctttgacccc tacacagttt cccattatgc cgttggagat
gagtgggaac
6841 gaatgtctga atcaggcttt aaactgttgt gccagtgctt aggctttgga
agtggtcatt
6901 tcagatgtga ttcatctaga tggtgccatg acaatggtgt gaactacaag
attggagaga
6961 agtgggaccg tcagggagaa aatggccaga tgatgagctg cacatgtctt
gggaacggaa
7021 aaggagaatt caagtgtgac cctcatgagg caacgtgtta tgatgatggg
aagacatacc
7081 acgtaggaga acagtggcag aaggaatatc tcggtgccat ttgctcctgc
acatgctttg
7141 gaggccagcg gggctggcgc tgtgacaact gccgcagacc tgggggtgaa
cccagtcccg
7201 aaggcactac tggccagtcc tacaaccagt attctcagag ataccatcag
agaacaaaca
7261 ctaatgttaa ttgcccaatt gagtgcttca tgcctttaga tgtacaggct
gacagagaag
7321 attcccgaga gtaaatcatc tttccaatcc agaggaacaa gcatgtctct
ctgccaagat
7381 ccatctaaac tggagtgatg ttagcagacc cagcttagag ttcttctttc
tttcttaagc
7441 cctttgctct ggaggaagtt ctccagcttc agctcaactc acagcttctc
caagcatcac
7501 cctgggagtt tcctgagggt tttctcataa atgagggctg cacattgcct
gttctgcttc
7561 gaagtattca ataccgctca gtattttaaa tgaagtgatt ctaagatttg
gtttgggatc
7621 aataggaaag catatgcagc caaccaagat gcaaatgttt tgaaatgata
tgaccaaaat
7681 tttaagtagg aaagtcaccc aaacacttct gctttcactt aagtgtctgg
cccgcaatac
7741 tgtaggaaca agcatgatct tgttactgtg atattttaaa tatccacagt
actcactttt
7801 tccaaatgat cctagtaatt gcctagaaat atctttctct tacctgttat
ttatcaattt

```
     7861 ttcccagtat ttttatacgg aaaaaattgt attgaaaaca cttagtatgc
agttgataag
     7921 aggaatttgg tataattatg gtgggtgatt atttttata ctgtatgtgc
caaagcttta
     7981 ctactgtgga aagacaactg ttttaataaa agatttacat tccacaactt
gaagttcatc
     8041 tatttgatat aagacacctt cgggggaaat aattcctgtg aatattcttt
ttcaattcag
     8101 caaacatttg aaaatctatg atgtgcaagt ctaattgttg atttcagtac
aagattttct
     8161 aaatcagttg ctacaaaaac tgattggttt ttgtcacttc atctcttcac
taatggagat
     8221 agctttacac tttctgcttt aatagattta agtggacccc aatatttatt
aaaattgcta
     8281 gtttaccgtt cagaagtata atagaaataa tctttagttg ctcttttcta
accattgtaa
     8341 ttcttccctt cttccctcca cctttccttc attgaataaa cctctgttca
aagagattgc
      8401 ctgcaaggga aataaaaatg actaagatat taaaaaaaaa aaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_002017.1

LOCUS NP_002017

ACCESSION NP_002017

```
      1 mlrgpgpgll llavqclgta vpstgasksk rqaqqmvqpq spvavsqskp
gcydngkhyq
       61 inqqwertyl gnalvctcyg gsrgfncesk peaeetcfdk ytgntyrvgd
tyerpkdsmi
      121 wdctcigagr grisctianr cheggqsyki gdtwrrphet ggymlecvcl
gngkgewtck
      181 piaekcfdha agtsyvvget wekpyqgwmm vdctclgegs gritctsrnr
cndqdtrtsy
      241 rigdtwskkd nrgnllqcic tgngrgewkc erhtsvqtts sgsgpftdvr
aavyqpqphp
      301 qpppyghcvt dsgvvysvgm qwlktqgnkq mlctclgngv scqetavtqt
yggnsngepc
      361 vlpftyngrt fyscttegrq dghlwcstts nyeqdqkysf ctdhtvlvqt
rggnsngalc
      421 hfpflynnhn ytdctsegrr dnmkwcgttq nydadqkfgf cpmaaheeic
ttnegvmyri
      481 gdqwdkqhdm ghmmrctcvg ngrgewtcia ysqlrdqciv dditynvndt
fhkrheeghm
      541 lnctcfgqgr grwkcdpvdq cqdsetgtfy qigdswekyv hgvryqcycy
grgigewhcq
      601 plqtypsssg pvevfitetp sqpnshpiqw napqpshisk yilrwrpkns
vgrwkeatip
      661 ghlnsytikg lkpgvvyegq lisiqqyghq evtrfdfttt ststpvtsnt
vtgettpfsp
      721 lvatsesvte itassfvvsw vsasdtvsgf rveyelseeg depqyldlps
tatsvnipdl
      781 lpgrkyivnv yqisedgeqs lilstsqtta pdappdptvd qvddtsivvr
wsrpqapitg
      841 yrivyspsve gsstelnlpe tansvtlsdl qpgvqyniti yaveenqest
pvviqqettg
      901 tprsdtvpsp rdlqfvevtd vkvtimwtpp esavtgyrvd vipvnlpgeh
gqrlpisrnt
```

```
      961 faevtglspg vtyyfkvfav shgreskplt aqqttkldap tnlqfvnetd
stvlvrwtpp
     1021 raqitgyrlt vgltrrgqpr qynvgpsvsk yplrnlqpas eytvslvaik
gnqespkatg
     1081 vfttlqpgss ippyntevte ttivitwtpa prigfklgvr psqggeapre
vtsdsgsivv
     1141 sgltpgveyv ytiqvlrdgq erdapivnkv vtplspptnl hleanpdtgv
ltvswerstt
     1201 pditgyritt tptngqqgns leevvhadqs sctfdnlspg leynvsvytv
kddkesvpis
     1261 dtiipavppp tdlrftnigp dtmrvtwapp psidltnflv ryspvkneed
vaelsispsd
     1321 navvltnllp gteyvvsvss vyeqhestpl rgrqktglds ptgidfsdit
ansftvhwia
     1381 pratitgyri rhhpehfsgr predrvphsr nsitltnltp gteyvvsiva
lngreespll
     1441 igqqstvsdv prdlevvaat ptslliswda pavtvryyri tygetggnsp
vqeftvpgsk
     1501 statisglkp gvdytitvya vtgrgdspas skpisinyrt eidkpsqmqv
tdvqdnsisv
     1561 kwlpssspvt gyrvtttpkn gpgptktkta gpdqtemtie glqptveyvv
svyaqnpsge
     1621 sqplvqtavt nidrpkglaf tdvdvdsiki awespqgqvs ryrvtysspe
dgihelfpap
     1681 dgeedtaelq glrpgseytv svvalhddme sqpligtqst aipaptdlkf
tqvtptslsa
     1741 qwtppnvqlt gyrvrvtpke ktgpmkeinl apdsssvvvs glmvatkyev
svyalkdtlt
     1801 srpaqgvvtt lenvspprra rvtdatetti tiswrtktet itgfqvdavp
angqtpiqrt
     1861 ikpdvrsyti tglqpgtdyk iylytlndna rsspvvidas taidapsnlr
flattpnsll
     1921 vswqpprari tgyiikyekp gspprevvpr prpgvteati tglepgteyt
iyvialknnq
     1981 ksepligrkk tdelpqlvtl phpnlhgpei ldvpstvqkt pfvthpgydt
gngiqlpgts
     2041 gqqpsvgqqm ifeehgfrrt tppttatpir hrprpyppnv gqealsqtti
swapfqdtse
     2101 yiischpvgt deeplqfrvp gtstsatltg ltrgatynii vealkdqqrh
kvreevvtvg
     2161 nsvneglnqp tddscfdpyt vshyavgdew ermsesgfkl lcqclgfgsg
hfrcdssrwc
     2221 hdngvnykig ekwdrqgeng qmmsctclgn gkgefkcdph eatcyddgkt
yhvgeqwqke
     2281 ylgaicsctc fgggrgwrcd ncrrpggeps pegttgqsyn qysqryhqrt
ntnvncpiec
     2341 fmpldvqadr edsre
```

CYB5

Official Symbol: CYB5A

Official Name: cytochrome b5 type A (microsomal)

Gene ID: 1528

Organism: Homo sapiens

Other Aliases: **CYB5**, MCB5

Other Designations: cytochrome b5; type 1 cyt-b5

Note – there are three difference isoforms

Isoform 1

Nucleotide sequence:

NCBI Reference Sequence: NM_148923.3

LOCUS NM_148923

ACCESSION NM_148923

```
1 gcgccccgcc cctgagccgg ccgcccagcc cccagtgggg ttcccggcgc ggggaatgtc
       61 ccgggtggag ctggctgagt cgcgcgctct gctccacccg acggggctgt
gtgtgctggg
      121 cctggctcgc ggcgaaccga gatggcagag cagtcggacg aggccgtgaa
gtactacacc
      181 ctagaggaga ttcagaagca caaccacagc aagagcacct ggctgatcct
gcaccacaag
      241 gtgtacgatt tgaccaaatt tctggaagag catcctggtg gggaagaagt
tttaagggaa
      301 caagctggag gtgacgctac tgagaacttt gaggatgtcg ggcactctac
agatgccagg
      361 gaaatgtcca aaacattcat cattggggag ctccatccag atgacagacc
aaagttaaac
      421 aagcctccgg aaactcttat cactactatt gattctagtt ccagttggtg
gaccaactgg
      481 gtgatccctg ccatctctgc agtggccgtc gccttgatgt atcgcctata
catggcagag
      541 gactgaacac ctcctcagaa gtcagcgcag gaagagcctg ctttggacac
gggagaaaag
      601 aagccattgc taactacttc aactgacaga aaccttcact tgaaaacaat
gattttaata
      661 tatctctttc tttttcttcc gacattagaa acaaaacaaa aagaactgtc
ctttctgcgc
      721 tcaaattttt cgagtgtgcc tttttattca tctactttat tttgatgttt
ccttaatgtg
      781 taatttactt attataagca tgatctttta aaaatatatt tggcttttaa
agtatgcaaa
      841 aaaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_683725.1

LOCUS NP_683725

ACCESSION NP_683725

```
      1 maeqsdeavk yytleeiqkh nhskstwlil hhkvydltkf leehpggeev
lreqaggdat
```

61 enfedvghst daremsktfi igelhpddrp klnkppetli ttidssssww tnwvipaisa
121 vavalmyrly maed

Isoform 2

Nucleotide sequence:

NCBI Reference Sequence: NM_001914.3

LOCUS NM_001914

ACCESSION NM_001914

1 gcgccccgcc cctgagccgg ccgcccagcc cccagtgggg ttcccggcgc ggggaatgtc
61 ccgggtggag ctggctgagt cgcgcgctct gctccacccg acggggctgt gtgtgctggg
121 cctggctcgc ggcgaaccga gatggcagag cagtcggacg aggccgtgaa gtactacacc
181 ctagaggaga ttcagaagca caaccacagc aagagcacct ggctgatcct gcaccacaag
241 gtgtacgatt tgaccaaatt tctggaagag catcctggtg gggaagaagt tttaagggaa
301 caagctggag gtgacgctac tgagaacttt gaggatgtcg ggcactctac agatgccagg
361 gaaatgtcca aaacattcat cattggggag ctccatccag atgacagacc aaagttaaac
421 aagcctccgg aaccttaaag gcggtgtttc aaggaaactc ttatcactac tattgattct
481 agttccagtt ggtggaccaa ctgggtgatc cctgccatct ctgcagtggc cgtcgccttg
541 atgtatcgcc tatacatggc agaggactga acacctcctc agaagtcagc gcaggaagag
601 cctgctttgg acacgggaga aaagaagcca ttgctaacta cttcaactga cagaaacctt
661 cacttgaaaa caatgatttt aatatatctc tttctttttc ttccgacatt agaaacaaaa
721 caaaaagaac tgtcctttct gcgctcaaat ttttcgagtg tgccttttta ttcatctact
781 ttattttgat gtttccttaa tgtgtaattt acttattata agcatgatct tttaaaaata
841 tatttggctt ttaaagtatg caaaaaaaaa aaaa

Protein sequence:

NCBI Reference Sequence: NP_001905.1

LOCUS NP_001905

ACCESSION NP_001905

1 maeqsdeavk yytleeiqkh nhskstwlil hhkvydltkf leehpggeev lreqaggdat
61 enfedvghst daremsktfi igelhpddrp klnkppep

Isoform 3

Nucleotide sequence:

NCBI Reference Sequence: NM_001190807.2

LOCUS NM_001190807

ACCESSION NM_001190807

```
      1 gcgccccgcc cctgagccgg ccgcccagcc cccagtgggg ttcccggcgc
ggggaatgtc
       61 ccgggtggag ctggctgagt cgcgcgctct gctccacccg acggggctgt
gtgtgctggg
      121 cctggctcgc ggcgaaccga gatggcagag cagtcggacg aggccgtgaa
gtactacacc
      181 ctagaggaga ttcagaagca caaccacagc aagagcacct ggctgatcct
gcaccacaag
      241 gtgtacgatt tgaccaaatt tctggaagag catcctggtg gggaagaagt
tttaagggaa
      301 caagctggag gtgacgctac tgagaacttt gaggatgtcg ggcactctac
agatgccagg
      361 gaaatgtcca aaacattcat cattggggag ctccatccag aaactcttat
cactactatt
      421 gattctagtt ccagttggtg gaccaactgg gtgatccctg ccatctctgc
agtggccgtc
      481 gccttgatgt atcgcctata catggcagag gactgaacac ctcctcagaa
gtcagcgcag
      541 gaagagcctg ctttggacac gggagaaaag aagccattgc taactacttc
aactgacaga
      601 aaccttcact tgaaaacaat gattttaata tatctctttc tttttcttcc
gacattagaa
      661 acaaaacaaa aagaactgtc ctttctgcgc tcaaattttt cgagtgtgcc
tttttattca
      721 tctactttat tttgatgttt ccttaatgtg taatttactt attataagca
tgatctttta
      781 aaaatatatt tggcttttaa agtatgcaaa aaaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001177736.1

LOCUS NP_001177736

ACCESSION NP_001177736

```
      1 maeqsdeavk yytleeiqkh nhskstwlil hhkvydltkf leehpggeev
lreqaggdat
       61 enfedvghst daremsktfi igelhpetli ttidssssww tnwvipaisa
vavalmyrly
      121 maed
```

PAI1

Official Symbol: SERPINE1

Official Name: serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1

Gene ID: 5054

Organism: Homo sapiens

Other Aliases: PAI, PAI-1, PAI1, PLANH1

Other Designations: endothelial plasminogen activator inhibitor; plasminogen activator inhibitor 1; serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1; serpin E1

Nucleotide sequence (Isoform 1):

NCBI Reference Sequence: NM_000602.4

LOCUS NM_000602

ACCESSION NM_000602

```
      1 ggcccacaga ggagcacagc tgtgtttggc tgcagggcca agagcgctgt
caagaagacc
       61 cacacgcccc cctccagcag ctgaattcct gcagctcagc agccgccgcc
agagcaggac
      121 gaaccgccaa tcgcaaggca cctctgagaa cttcaggatg cagatgtctc
cagccctcac
      181 ctgcctagtc ctgggcctgg cccttgtctt tggtgaaggg tctgctgtgc
accatccccc
      241 atcctacgtg gcccacctgg cctcagactt cggggtgagg gtgtttcagc
aggtggcgca
      301 ggcctccaag gaccgcaacg tggttttctc accctatggg gtggcctcgg
tgttggccat
      361 gctccagctg acaacaggag gagaaaccca gcagcagatt caagcagcta
tgggattcaa
      421 gattgatgac aagggcatgg cccccgccct ccggcatctg tacaaggagc
tcatggggcc
      481 atggaacaag gatgagatca gcaccacaga cgcgatcttc gtccagcggg
atctgaagct
      541 ggtccagggc ttcatgcccc acttcttcag gctgttccgg agcacggtca
agcaagtgga
      601 cttttcagag gtggagagag ccagattcat catcaatgac tgggtgaaga
cacacacaaa
      661 aggtatgatc agcaacttgc ttgggaaagg agccgtggac cagctgacac
ggctggtgct
      721 ggtgaatgcc ctctacttca acggccagtg gaagactccc ttccccgact
ccagcaccca
      781 ccgccgcctc ttccacaaat cagacggcag cactgtctct gtgcccatga
tggctcagac
      841 caacaagttc aactatactg agttcaccac gcccgatggc cattactacg
acatcctgga
      901 actgccctac cacggggaca ccctcagcat gttcattgct gccccttatg
aaaaagaggt
```

961 gcctctctct gccctcacca acattctgag tgcccagctc atcagccact
ggaaaggcaa
1021 catgaccagg ctgccccgcc tcctggttct gcccaagttc tccctggaga
ctgaagtcga
1081 cctcaggaag cccctagaga acctgggaat gaccgacatg ttcagacagt
ttcaggctga
1141 cttcacgagt ctttcagacc aagagcctct ccacgtcgcg caggcgctgc
agaaagtgaa
1201 gatcgaggtg aacgagagtg gcacggtggc ctcctcatcc acagctgtca
tagtctcagc
1261 ccgcatggcc cccgaggaga tcatcatgga cagacccttc ctctttgtgg
tccggcacaa
1321 ccccacagga acagtccttt tcatgggcca agtgatggaa ccctgaccct
ggggaaagac
1381 gccttcatct gggacaaaac tggagatgca tcgggaaaga agaaactccg
aagaaaagaa
1441 ttttagtgtt aatgactctt tctgaaggaa gagaagacat ttgccttttg
ttaaaagatg
1501 gtaaaccaga tctgtctcca agaccttggc ctctccttgg aggaccttta
ggtcaaactc
1561 cctagtctcc acctgagacc ctgggagaga agtttgaagc acaactccct
taaggtctcc
1621 aaaccagacg gtgacgcctg cgggaccatc tggggcacct gcttccaccc
gtctctctgc
1681 ccactcgggt ctgcagacct ggttcccact gaggcccttt gcaggatgga
actacggggc
1741 ttacaggagc ttttgtgtgc ctggtagaaa ctatttctgt tccagtcaca
ttgccatcac
1801 tcttgtactg cctgccaccg cggaggaggc tggtgacagg ccaaaggcca
gtggaagaaa
1861 caccctttca tctcagagtc cactgtggca ctggccaccc ctccccagta
caggggtgct
1921 gcaggtggca gagtgaatgt cccccatcat gtggcccaac tctcctggcc
tggccatctc
1981 cctccccaga aacagtgtgc atgggttatt ttggagtgta ggtgacttgt
ttactcattg
2041 aagcagattt ctgcttcctt ttatttttat aggaatagag gaagaaatgt
cagatgcgtg
2101 cccagctctt caccccccaa tctcttggtg gggaggggtg tacctaaata
tttatcatat
2161 ccttgccctt gagtgcttgt tagagagaaa gagaactact aaggaaaata
atattattta
2221 aactcgctcc tagtgtttct ttgtggtctg tgtcaccgta tctcaggaag
tccagccact
2281 tgactggcac acacccctcc ggacatccag cgtgacggag cccacactgc
caccttgtgg
2341 ccgcctgaga ccctcgcgcc ccccgcgccc ctctttttcc ccttgatgga
aattgaccat
2401 acaatttcat cctccttcag gggatcaaaa ggacggagtg gggggacaga
gactcagatg
2461 aggacagagt ggtttccaat gtgttcaata gatttaggag cagaaatgca
aggggctgca
2521 tgacctacca ggacagaact ttccccaatt acagggtgac tcacagccgc
attggtgact
2581 cacttcaatg tgtcatttcc ggctgctgtg tgtgagcagt ggacacgtga
ggggggggtg
2641 ggtgagagag acaggcagct cggattcaac taccttagat aatatttctg
aaaacctacc
2701 agccagaggg tagggcacaa agatggatgt aatgcacttt gggaggccaa
ggcgggagga

```
     2761 ttgcttgagc ccaggagttc aagaccagcc tgggcaacat accaagaccc
ccgtctcttt
     2821 aaaaatatat atattttaaa tatacttaaa tatatatttc taatatcttt
aaatatatat
     2881 atatatttta aagaccaatt tatgggagaa ttgcacacag atgtgaaatg
aatgtaatct
     2941 aatagaagcc taatcagccc accatgttct ccactgaaaa atcctctttc
tttggggttt
     3001 ttctttcttt ctttttgat tttgcactgg acggtgacgt cagccatgta
caggatccac
     3061 aggggtggtg tcaaatgcta ttgaaattgt gttgaattgt atgctttttc
acttttgata
     3121 aataaacatg taaaaatgtt tcaaaaaaat aataaaataa ataaatacga
agaatatgtc
     3181 aggacagtca aaaaaaaaaa aaaaaaa
```

Protein sequence (isoform 1):

NCBI Reference Sequence: NP_000593.1

LOCUS NP_000593

ACCESSION NP_000593

```
      1 mqmspaltcl vlglalvfge gsavhhppsy vahlasdfgv rvfqqvaqas
kdrnvvfspy
       61 gvasvlamlq lttggetqqq iqaamgfkid dkgmapalrh lykelmgpwn
kdeisttdai
      121 fvqrdlklvq gfmphffrlf rstvkqvdfs everarfiin dwvkthtkgm
isnllgkgav
      181 dqltrlvlvn alyfngqwkt pfpdssthrr lfhksdgstv svpmmaqtnk
fnytefttpd
      241 ghyydilelp yhgdtlsmfi aapyekevpl saltnilsaq lishwkgnmt
rlprllvlpk
      301 fsletevdlr kplenlgmtd mfrqfqadft slsdqeplhv aqalqkvkie
vnesgtvass
      361 stavivsarm apeeiimdrp flfvvrhnpt gtvlfmgqvm ep
```

Nucleotide sequence (isoform 2):

NCBI Reference Sequence: NM_001165413.2

LOCUS NM_001165413

ACCESSION NM_001165413

```
      1 ggcccacaga ggagcacagc tgtgtttggc tgcagggcca agagcgctgt
caagaagacc
       61 cacacgcccc cctccagcag ctgaattcct gcagctcagc agccgccgcc
agagcaggac
      121 gaaccgccaa tcgcaaggca cctctgagaa cttcaggatg cagatgtctc
cagccctcac
      181 ctgcctagtc ctgggcctgg cccttgtctt tggtgaaggg tctgctgtgc
accatccccc
      241 atcctacgtg gcgcaggcct ccaaggaccg caacgtggtt ttctcaccct
atggggtggc
```

301 ctcggtgttg gccatgctcc agctgacaac aggaggagaa acccagcagc
agattcaagc
361 agctatggga ttcaagattg atgacaaggg catggccccc gccctccggc
atctgtacaa
421 ggagctcatg gggccatgga acaaggatga gatcagcacc acagacgcga
tcttcgtcca
481 gcgggatctg aagctggtcc agggcttcat gccccacttc ttcaggctgt
tccggagcac
541 ggtcaagcaa gtggactttt cagaggtgga gagagccaga ttcatcatca
atgactgggt
601 gaagacacac acaaaaggta tgatcagcaa cttgcttggg aaaggagccg
tggaccagct
661 gacacggctg gtgctggtga atgccctcta cttcaacggc cagtggaaga
ctcccttccc
721 cgactccagc acccaccgcc gcctcttcca caaatcagac ggcagcactg
tctctgtgcc
781 catgatggct cagaccaaca agttcaacta tactgagttc accacgcccg
atggccatta
841 ctacgacatc ctggaactgc cctaccacgg ggacaccctc agcatgttca
ttgctgcccc
901 ttatgaaaaa gaggtgcctc tctctgccct caccaacatt ctgagtgccc
agctcatcag
961 ccactggaaa ggcaacatga ccaggctgcc ccgcctcctg gttctgccca
agttctccct
1021 ggagactgaa gtcgacctca ggaagcccct agagaacctg ggaatgaccg
acatgttcag
1081 acagtttcag gctgacttca cgagtctttc agaccaagag cctctccacg
tcgcgcaggc
1141 gctgcagaaa gtgaagatcg aggtgaacga gagtggcacg gtggcctcct
catccacagc
1201 tgtcatagtc tcagcccgca tggcccccga ggagatcatc atggacagac
ccttcctctt
1261 tgtggtccgg cacaacccca caggaacagt ccttttcatg ggccaagtga
tggaaccctg
1321 accctgggga aagacgcctt catctgggac aaaactggag atgcatcggg
aaagaagaaa
1381 ctccgaagaa aagaatttta gtgttaatga ctctttctga aggaagagaa
gacatttgcc
1441 ttttgttaaa agatggtaaa ccagatctgt ctccaagacc ttggcctctc
cttggaggac
1501 ctttaggtca aactccctag tctccacctg agaccctggg agagaagttt
gaagcacaac
1561 tcccttaagg tctccaaacc agacggtgac gcctgcggga ccatctgggg
cacctgcttc
1621 cacccgtctc tctgcccact cgggtctgca gacctggttc ccactgaggc
cctttgcagg
1681 atggaactac ggggcttaca ggagcttttg tgtgcctggt agaaactatt
tctgttccag
1741 tcacattgcc atcactcttg tactgcctgc caccgcggag gaggctggtg
acaggccaaa
1801 ggccagtgga agaaacaccc tttcatctca gagtccactg tggcactggc
caccctccc
1861 cagtacaggg gtgctgcagg tggcagagtg aatgtccccc atcatgtggc
ccaactctcc
1921 tggcctggcc atctccctcc ccagaaacag tgtgcatggg ttattttgga
gtgtaggtga
1981 cttgtttact cattgaagca gatttctgct tccttttatt tttataggaa
tagaggaaga
2041 aatgtcagat gcgtgcccag ctcttcaccc cccaatctct tggtggggag
gggtgtacct

```
     2101 aaatatttat catatccttg cccttgagtg cttgttagag agaaagagaa
ctactaagga
     2161 aaataatatt atttaaactc gctcctagtg tttctttgtg gtctgtgtca
ccgtatctca
     2221 ggaagtccag ccacttgact ggcacacacc cctccggaca tccagcgtga
cggagcccac
     2281 actgccacct tgtggccgcc tgagaccctc gcgccccccg cgcccctctt
tttccccttg
     2341 atggaaattg accatacaat ttcatcctcc ttcaggggat caaaaggacg
gagtgggggg
     2401 acagagactc agatgaggac agagtggttt ccaatgtgtt caatagattt
aggagcagaa
     2461 atgcaagggg ctgcatgacc taccaggaca gaactttccc caattacagg
gtgactcaca
     2521 gccgcattgg tgactcactt caatgtgtca tttccggctg ctgtgtgtga
gcagtggaca
     2581 cgtgaggggg gggtgggtga gagagacagg cagctcggat tcaactacct
tagataatat
     2641 ttctgaaaac ctaccagcca gagggtaggg cacaaagatg gatgtaatgc
actttgggag
     2701 gccaaggcgg gaggattgct tgagcccagg agttcaagac cagcctgggc
aacataccaa
     2761 gacccccgtc tctttaaaaa tatatatatt ttaaatatac ttaaatatat
atttctaata
     2821 tctttaaata tatatatata ttttaaagac caatttatgg gagaattgca
cacagatgtg
     2881 aaatgaatgt aatctaatag aagcctaatc agcccaccat gttctccact
gaaaaatcct
     2941 ctttctttgg ggtttttctt tctttctttt ttgattttgc actggacggt
gacgtcagcc
     3001 atgtacagga tccacagggg tggtgtcaaa tgctattgaa attgtgttga
attgtatgct
     3061 ttttcacttt tgataaataa acatgtaaaa atgtttcaaa aaaataataa
aataaataaa
     3121 tacgaagaat atgtcaggac agtcaaaaaa aaaaaaaaaa aa
```

Protein sequence (isoform 2):

NCBI Reference Sequence: NP_001158885.1

LOCUS NP_001158885

ACCESSION NP_001158885

```
     1 mqmspaltcl vlglalvfge gsavhhppsy vaqaskdrnv vfspygvasv
lamlqlttgg
    61 etqqqiqaam gfkiddkgma palrhlykel mgpwnkdeis ttdaifvqrd
lklvqgfmph
   121 ffrlfrstvk qvdfsevera rfiindwvkt htkgmisnll gkgavdqltr
lvlvnalyfn
   181 gqwktpfpds sthrrlfhks dgstvsvpmm aqtnkfnyte fttpdghyyd
ilelpyhgdt
   241 lsmfiaapye kevplsaltn ilsaqlishw kgnmtrlprl lvlpkfslet
evdlrkplen
   301 lgmtdmfrqf qadftslsdq eplhvaqalq kvkievnesg tvassstavi
vsarmapeei
   361 imdrpflfvv rhnptgtvlf mgqvmep
```

MPR1

Official Symbol: IGF2R

Official Name: insulin-like growth factor 2 receptor

Gene ID: 3482

Organism: Homo sapiens

Other Aliases: CD222, CIMPR, M6P-R, **MPR1**, MPRI

Other Designations: 300 kDa mannose 6-phosphate receptor; CI Man-6-P receptor; CI-MPR; IGF-II receptor; M6P/IGF2 receptor; M6P/IGF2R; M6PR; MPR 300; cation-independent mannose-6 phosphate receptor; cation-independent mannose-6-phosphate receptor; insulin-like growth factor II receptor

Nucleotide sequence:

NCBI Reference Sequence: NM_000876.2

LOCUS NM_000876

ACCESSION NM_000876

```
      1 cgagcccagt cgagccgcgc tcacctcggg ctcccgctcc gtctccacct
ccgcctttgc
       61 cctggcggcg cgaccccgtc ccgggcgcgg cccccagcag tcgcgcgccg
ttagcctcgc
      121 gcccgccgcg cagtccgggc ccggcgcgat gggggccgcc gccggccgga
gcccccacct
      181 ggggcccgcg cccgcccgcc gcccgcagcg ctctctgctc ctgctgcagc
tgctgctgct
      241 cgtcgctgcc ccggggtcca cgcaggccca ggccgccccg ttccccgagc
tgtgcagtta
      301 tacatgggaa gctgttgata ccaaaaataa tgtactttat aaaatcaaca
tctgtggaag
      361 tgtggatatt gtccagtgcg ggccatcaag tgctgtttgt atgcacgact
tgaagacacg
      421 cacttatcat tcagtgggtg actctgtttt gagaagtgca accagatctc
tcctggaatt
      481 caacacaaca gtgagctgtg accagcaagg cacaaatcac agagtccaga
gcagcattgc
      541 cttcctgtgt gggaaaaccc tgggaactcc tgaatttgta actgcaacag
aatgtgtgca
      601 ctactttgag tggaggacca ctgcagcctg caagaaagac atatttaaag
caaataagga
      661 ggtgccatgc tatgtgtttg atgaagagtt gaggaagcat gatctcaatc
ctctgatcaa
      721 gcttagtggt gcctacttgg tggatgactc cgatccggac acttctctat
tcatcaatgt
      781 ttgtagagac atagacacac tacgagaccc aggttcacag ctgcgggcct
gtccccccgg
      841 cactgccgcc tgcctggtaa gaggacacca ggcgtttgat gttggccagc
cccgggacgg
      901 actgaagctg gtgcgcaagg acaggcttgt cctgagttac gtgagggaag
aggcaggaaa
```

961 gctagacttt tgtgatggtc acagccctgc ggtgactatt acatttgttt
gcccgtcgga
1021 gcggagagag ggcaccattc ccaaactcac agctaaatcc aactgccgct
atgaaattga
1081 gtggattact gagtatgcct gccacagaga ttacctggaa agtaaaactt
gttctctgag
1141 cggcgagcag caggatgtct ccatagacct cacaccactt gcccagagcg
gaggttcatc
1201 ctatatttca gatggaaaag aatatttgtt ttatttgaat gtctgtggag
aaactgaaat
1261 acagttctgt aataaaaaac aagctgcagt ttgccaagtg aaaaagagcg
atacctctca
1321 agtcaaagca gcaggaagat accacaatca gaccctccga tattcggatg
gagacctcac
1381 cttgatatat tttggaggtg atgaatgcag ctcagggttt cagcggatga
gcgtcataaa
1441 ctttgagtgc aataaaaccg caggtaacga tgggaaagga actcctgtat
tcacagggga
1501 ggttgactgc acctacttct tcacatggga cacggaatac gcctgtgtta
aggagaagga
1561 agacctcctc tgcggtgcca ccgacgggaa gaagcgctat gacctgtccg
cgctggtccg
1621 ccatgcagaa ccagagcaga attgggaagc tgtggatggc agtcagacgg
aaacagagaa
1681 gaagcatttt ttcattaata tttgtcacag agtgctgcag gaaggcaagg
cacgagggtg
1741 tcccgaggac gcggcagtgt gtgcagtgga taaaaatgga agtaaaaatc
tgggaaaatt
1801 tatttcctct cccatgaaag agaaaggaaa cattcaactc tcttattcag
atggtgatga
1861 ttgtggtcat ggcaagaaaa ttaaaactaa tatcacactt gtatgcaagc
caggtgatct
1921 ggaaagtgca ccagtgttga gaacttctgg ggaaggcggt tgcttttatg
agtttgagtg
1981 gcacacagct gcggcctgtg tgctgtctaa gacagaaggg gagaactgca
cggtctttga
2041 ctcccaggca gggttttctt ttgacttatc acctctcaca aagaaaaatg
gtgcctataa
2101 agttgagaca aagaagtatg acttttatat aaatgtgtgt ggcccggtgt
ctgtgagccc
2161 ctgtcagcca gactcaggag cctgccaggt ggcaaaaagt gatgagaaga
cttggaactt
2221 gggtctgagt aatgcgaagc tttcatatta tgatgggatg atccaactga
actacagagg
2281 cggcacaccc tataacaatg aaagacacac accgagagct acgctcatca
cctttctctg
2341 tgatcgagac gcgggagtgg gcttccctga atatcaggaa gaggataact
ccacctacaa
2401 cttccggtgg tacaccagct atgcctgccc ggaggagccc ctggaatgcg
tagtgaccga
2461 cccctccacg ctggagcagt acgacctctc cagtctggca aaatctgaag
gtggccttgg
2521 aggaaactgg tatgccatgg acaactcagg ggaacatgtc acgtggagga
aatactacat
2581 taacgtgtgt cggcctctga atccagtgcc gggctgcaac cgatatgcat
cggcttgcca
2641 gatgaagtat gaaaaagatc agggctcctt cactgaagtg gtttccatca
gtaacttggg
2701 aatggcaaag accggcccgg tggttgagga cagcggcagc ctccttctgg
aatacgtgaa

2761 tgggtcggcc tgcaccacca gcgatggcag acagaccaca tataccacga
ggatccatct
2821 cgtctgctcc aggggcaggc tgaacagcca ccccatcttt tctctcaact
gggagtgtgt
2881 ggtcagtttc ctgtggaaca cagaggctgc ctgtcccatt cagacaacga
cggatacaga
2941 ccaggcttgc tctataaggg atcccaacag tggatttgtg tttaatctta
atccgctaaa
3001 cagttcgcaa ggatataacg tctctggcat tgggaagatt tttatgttta
atgtctgcgg
3061 cacaatgcct gtctgtggga ccatcctggg aaaacctgct tctggctgtg
aggcagaaac
3121 ccaaactgaa gagctcaaga attggaagcc agcaaggcca gtcggaattg
agaaaagcct
3181 ccagctgtcc acagagggct tcatcactct gacctacaaa gggcctctct
ctgccaaagg
3241 taccgctgat gcttttatcg tccgctttgt ttgcaatgat gatgtttact
cagggcccct
3301 caaattcctg catcaagata tcgactctgg gcaagggatc cgaaacactt
actttgagtt
3361 tgaaaccgcg ttggcctgtg ttccttctcc agtggactgc caagtcaccg
acctggctgg
3421 aaatgagtac gacctgactg gcctaagcac agtcaggaaa ccttggacgg
ctgttgacac
3481 ctctgtcgat gggagaaaga ggactttcta tttgagcgtt tgcaatcctc
tcccttacat
3541 tcctggatgc cagggcagcg cagtggggtc ttgcttagtg tcagaaggca
atagctggaa
3601 tctgggtgtg gtgcagatga gtccccaagc cgcggcgaat ggatctttga
gcatcatgta
3661 tgtcaacggt gacaagtgtg ggaaccagcg cttctccacc aggatcacgt
ttgagtgtgc
3721 tcagatatcg ggctcaccag catttcagct tcaggatggt tgtgagtacg
tgtttatctg
3781 gagaactgtg gaagcctgtc ccgttgtcag agtggaaggg gacaactgtg
aggtgaaaga
3841 cccaaggcat ggcaacttgt atgacctgaa gcccctgggc ctcaacgaca
ccatcgtgag
3901 cgctggcgaa tacacttatt acttccgggt ctgtgggaag ctttcctcag
acgtctgccc
3961 cacaagtgac aagtccaagg tggtctcctc atgtcaggaa aagcgggaac
cgcagggatt
4021 tcacaaagtg gcaggtctcc tgactcagaa gctaacttat gaaaatggct
tgttaaaaat
4081 gaacttcacg gggggggaca cttgccataa ggtttatcag cgctccacag
ccatcttctt
4141 ctactgtgac cgcggcaccc agcggccagt atttctaaag gagacttcag
attgttccta
4201 cttgtttgag tggcgaacgc agtatgcctg cccacctttc gatctgactg
aatgttcatt
4261 caaagatggg gctggcaact ccttcgacct ctcgtccctg tcaaggtaca
gtgacaactg
4321 ggaagccatc actgggacgg gggacccgga gcactacctc atcaatgtct
gcaagtctct
4381 ggccccgcag gctggcactg agccgtgccc tccagaagca gccgcgtgtc
tgctgggtgg
4441 ctccaagccc gtgaacctcg gcagggtaag ggacggacct cagtggagag
atggcataat
4501 tgtcctgaaa tacgttgatg gcgacttatg tccagatggg attcggaaaa
agtcaaccac

4561 catccgattc acctgcagcg agagccaagt gaactccagg cccatgttca
tcagcgccgt
4621 ggaggactgt gagtacacct ttgcctggcc cacagccaca gcctgtccca
tgaagagcaa
4681 cgagcatgat gactgccagg tcaccaaccc aagcacagga cacctgtttg
atctgagctc
4741 cttaagtggc agggcgggat tcacagctgc ttacagcgag aaggggttgg
tttacatgag
4801 catctgtggg gagaatgaaa actgccctcc tggcgtgggg gcctgctttg
gacagaccag
4861 gattagcgtg ggcaaggcca acaagaggct gagatacgtg gaccaggtcc
tgcagctggt
4921 gtacaaggat gggtcccctt gtccctccaa atccggcctg agctataaga
gtgtgatcag
4981 tttcgtgtgc aggcctgagg ccgggccaac caataggccc atgctcatct
ccctggacaa
5041 gcagacatgc actctcttct tctcctggca cacgccgctg gcctgcgagc
aagcgaccga
5101 atgttccgtg aggaatggaa gctctattgt tgacttgtct ccccttattc
atcgcactgg
5161 tggttatgag gcttatgatg agagtgagga tgatgcctcc gataccaacc
ctgatttcta
5221 catcaatatt tgtcagccac taaatcccat gcacggagtg ccctgtcctg
ccggagccgc
5281 tgtgtgcaaa gttcctattg atggtccccc catagatatc ggccgggtag
caggaccacc
5341 aatactcaat ccaatagcaa atgagattta cttgaatttt gaaagcagta
ctccttgctt
5401 agcggacaag catttcaact acacctcgct catcgcgttt cactgtaaga
gaggtgtgag
5461 catgggaacg cctaagctgt taaggaccag cgagtgcgac tttgtgttcg
aatgggagac
5521 tcctgtcgtc tgtcctgatg aagtgaggat ggatggctgt accctgacag
atgagcagct
5581 cctctacagc ttcaacttgt ccagcctttc cacgagcacc tttaaggtga
ctcgcgactc
5641 gcgcacctac agcgttgggg tgtgcacctt tgcagtcggg ccagaacaag
gaggctgtaa
5701 ggacggagga gtctgtctgc tctcaggcac caaggggggca tcctttggac
ggctgcaatc
5761 aatgaaactg gattacaggc accaggatga agcggtcgtt ttaagttacg
tgaatggtga
5821 tcgttgccct ccagaaaccg atgacggcgt cccctgtgtc ttccccttca
tattcaatgg
5881 gaagagctac gaggagtgca tcatagagag cagggcgaag ctgtggtgta
gcacaactgc
5941 ggactacgac agagaccacg agtggggctt ctgcagacac tcaaacagct
accggacatc
6001 cagcatcata tttaagtgtg atgaagatga ggacattggg aggccacaag
tcttcagtga
6061 agtgcgtggg tgtgatgtga catttgagtg gaaaacaaaa gttgtctgcc
ctccaaagaa
6121 gttggagtgc aaattcgtcc agaaacacaa aacctacgac ctgcggctgc
tctcctctct
6181 caccgggtcc tggtccctgg tccacaacgg agtctcgtac tatataaatc
tgtgccagaa
6241 aatatataaa gggcccctgg gctgctctga aagggccagc atttgcagaa
ggaccacaac
6301 tggtgacgtc caggtcctgg gactcgttca cacgcagaag ctgggtgtca
taggtgacaa

6361 agttgttgtc acgtactcca aaggttatcc gtgtggtgga aataagaccg catcctccgt
6421 gatagaattg acctgtacaa agacggtggg cagacctgca ttcaagaggt ttgatatcga
6481 cagctgcact tactacttca gctgggactc ccgggctgcc tgcgccgtga agcctcagga
6541 ggtgcagatg gtgaatggga ccatcaccaa ccctataaat ggcaagagct tcagcctcgg
6601 agatatttat tttaagctgt tcagagcctc tggggacatg aggaccaatg gggacaacta
6661 cctgtatgag atccaacttt cctccatcac aagctccaga aacccggcgt gctctggagc
6721 caacatatgc caggtgaagc ccaacgatca gcacttcagt cggaaagttg gaacctctga
6781 caagaccaag tactaccttc aagacggcga tctcgatgtc gtgtttgcct cttcctctaa
6841 gtgcggaaag gataagacca agtctgtttc ttccaccatc ttcttccact gtgaccctct
6901 ggtggaggac gggatccccg agttcagtca cgagactgcc gactgccagt acctcttctc
6961 ttggtacacc tcagccgtgt gtcctctggg ggtgggcttt gacagcgaga atcccgggga
7021 cgacgggcag atgcacaagg ggctgtcaga acggagccag gcagtcggcg cggtgctcag
7081 cctgctgctg gtggcgctca cctgctgcct gctggccctg ttgctctaca agaaggagag
7141 gagggaaaca gtgataagta agctgaccac ttgctgtagg agaagttcca acgtgtccta
7201 caaatactca aaggtgaata aggaagaaga gacagatgag aatgaaacag agtggctgat
7261 ggaagagatc cagctgcctc ctccacggca gggaaaggaa gggcaggaga acggccatat
7321 taccaccaag tcagtgaaag ccctcagctc cctgcatggg gatgaccagg acagtgagga
7381 tgaggttctg accatcccag aggtgaaagt tcactcgggc aggggagctg gggcagagag
7441 ctcccaccca gtgagaaacg cacagagcaa tgcccttcag gagcgtgagg acgatagggt
7501 ggggctggtc aggggtgaga aggcgaggaa agggaagtcc agctctgcac agcagaagac
7561 agtgagctcc accaagctgg tgtccttcca tgacgacagc gacgaggacc tcttacacat
7621 ctgactccgc agtgcctgca ggggagcacg gagccgcggg acagccaagc acctccaacc
7681 aaataagact tccactcgat gatgcttcta taattttgcc tttaacagaa actttcaaaa
7741 gggaagagtt tttgtgatgg gggagagggt gaaggaggtc aggccccact ccttcctgat
7801 tgtttacagt cattggaata aggcatggct cagatcggcc acagggcggt accttgtgcc
7861 cagggttttg ccccaagtcc tcatttaaaa gcataaggcc ggacgcatct caaaacagag
7921 ggctgcattc gaagaaaccc ttgctgcttt agtcccgata gggtatttga ccccgatata
7981 ttttagcatt ttaattctct ccccctattt attgactttg acaattactc aggtttgaga
8041 aaaaggaaaa aaaaacagcc accgtttctt cctgccagca ggggtgtgat gtaccagttt
8101 gtccatcttg agatggtgag gctgtcagtg tatggggcag cttccggcgg gatgttgaac

```
     8161 tggtcattaa tgtgtcccct gagttggagc tcattctgtc tcttttctct
tttgctttct
     8221 gtttcttaag ggcacacaca cgtgcgtgcg agcacacaca cacatacgtg
cacagggtcc
     8281 ccgagtgcct aggttttgga gagtttgcct gttctatgcc tttagtcagg
aatggctgca
     8341 cctttttgca tgatatcttc aagcctgggc gtacagagca catttgtcag
tatttttgcc
     8401 ggctggtgaa ttcaaacaac ctgcccaaag attgatttgt gtgtttgtgt
gtgtgtgtgt
     8461 gtgtgtgtgt gtgtgtgtga gtggagttga ggtgtcagag aaaatgaatt
ttttccagat
     8521 ttggggtata ggtctcatct cttcaggttc tcatgatacc acctttactg
tgcttatttt
     8581 tttaagaaaa aagtgttgat caaccattcg acctataaga agccttaatt
tgcacagtgt
     8641 gtgacttaca gaaactgcat gaaaaatcat gggccagagc ctcggcccta
gcattgcact
     8701 tggcctcatg ctggagggag gctgggcggg tacagcgcgg aggaggaggg
aggccaggcg
     8761 ggcatggcgt ggaggaggag ggaggccggg cggtcacagc atggaggagg
agggaggcgc
     8821 tgctggtgtt cttattctgg cggcagcgcc tttcctgcca tgtttagtga
atgacttttc
     8881 tcgcattgta gaattgtata tagactctgg tgttctattg ctgagaagca
aaccgccctg
     8941 cagcatccct cagcctgtac cggtttggct ggcttgtttg atttcaacat
gagtgtattt
     9001 tttaaaattg atttttctct tcatttttt ttcaatcaac tttactgtaa
tataaagtat
      9061 tcaacaattt caataaaaga taaattatta aaa
```

Protein sequence:

NCBI Reference Sequence: NP_000867.2

LOCUS NP_000867

ACCESSION NP_000867

```
      1 mgaaagrsph lgpaparrpq rsllllqlll lvaapgstqa qaapfpelcs
ytweavdtkn
       61 nvlykinicg svdivqcgps savcmhdlkt rtyhsvgdsv lrsatrslle
fnttvscdqq
      121 gtnhrvqssi aflcgktlgt pefvtatecv hyfewrttaa ckkdifkank
evpcyvfdee
      181 lrkhdlnpli klsgaylvdd sdpdtslfin vcrdidtlrd pgsqlracpp
gtaaclvrgh
      241 qafdvgqprd glklvrkdrl vlsyvreeag kldfcdghsp avtitfvcps
erregtipkl
      301 taksncryei ewiteyachr dylesktcsl sgeqqdvsid ltplaqsggs
syisdgkeyl
      361 fylnvcgete iqfcnkkqaa vcqvkksdts qvkaagryhn qtlrysdgdl
tliyfggdec
      421 ssgfqrmsvi nfecnktagn dgkgtpvftg evdctyfftw dteyacvkek
edllcgatdg
      481 kkrydlsalv rhaepeqnwe avdgsqtete kkhffinich rvlqegkarg
cpedaavcav
      541 dkngsknlgk fisspmkekg niqlsysdgd dcghgkkikt nitlvckpgd
lesapvlrts
```

601 geggcfyefe whtaaacvls ktegenctvf dsqagfsfdl spltkkngay
kvetkkydfy
661 invcgpvsvs pcqpdsgacq vaksdektwn lglsnaklsy ydgmiqlnyr
ggtpynnerh
721 tpratlitfl cdrdagvgfp eyqeednsty nfrwytsyac peeplecvvt
dpstleqydl
781 sslakseggl ggnwyamdns gehvtwrkyy invcrplnpv pgcnryasac
qmkyekdqgs
841 ftevvsisnl gmaktgpvve dsgsllleyv ngsacttsdg rqttyttrih
lvcsrgrlns
901 hpifslnwec vvsflwntea acpiqtttdt dqacsirdpn sgfvfnlnpl
nssqgynvsg
961 igkifmfnvc gtmpvcgtil gkpasgceae tqteelknwk parpvgieks
lqlstegfit
1021 ltykgplsak gtadafivrf vcnddvysgp lkflhqdids gqgirntyfe
fetalacvps
1081 pvdcqvtdla gneydltgls tvrkpwtavd tsvdgrkrtf ylsvcnplpy
ipgcqgsavg
1141 sclvsegnsw nlgvvqmspq aaangslsim yvngdkcgnq rfstritfec
aqisgspafq
1201 lqdgceyvfi wrtveacpvv rvegdncevk dprhgnlydl kplglndtiv
sageytyyfr
1261 vcgklssdvc ptsdkskvvs scqekrepqg fhkvaglltq kltyengllk
mnftggdtch
1321 kvyqrstaif fycdrgtqrp vflketsdcs ylfewrtqya cppfdltecs
fkdgagnsfd
1381 lsslsrysdn weaitgtgdp ehylinvcks lapqagtepc ppeaaacllg
gskpvnlgrv
1441 rdgpqwrdgi ivlkyvdgdl cpdgirkkst tirftcsesq vnsrpmfisa
vedceytfaw
1501 ptatacpmks nehddcqvtn pstghlfdls slsgragfta aysekglvym
sicgenencp
1561 pgvgacfgqt risvgkankr lryvdqvlql vykdgspcps ksglsyksvi
sfvcrpeagp
1621 tnrpmlisld kqtctlffsw htplaceqat ecsvrngssi vdlsplihrt
ggyeaydese
1681 ddasdtnpdf yinicqplnp mhgvpcpaga avckvpidgp pidigrvagp
pilnpianei
1741 ylnfesstpc ladkhfnyts liafhckrgv smgtpkllrt secdfvfewe
tpvvcpdevr
1801 mdgctltdeq llysfnlssl ststfkvtrd srtysvgvct favgpeqggc
kdggvcllsg
1861 tkgasfgrlq smkldyrhqd eavvlsyvng drcppetddg vpcvfpfifn
gksyeeciie
1921 sraklwcstt adydrdhewg fcrhsnsyrt ssiifkcded edigrpqvfs
evrgcdvtfe
1981 wktkvvcppk kleckfvqkh ktydlrllss ltgswslvhn gvsyyinlcq
kiykgplgcs
2041 erasicrrtt tgdvqvlglv htqklgvigd kvvvtyskgy pcggnktass
vieltctktv
2101 grpafkrfdi dsctyyfswd sraacavkpq evqmvngtit npingksfsl
gdiyfklfra
2161 sgdmrtngdn ylyeiqlssi tssrnpacsg anicqvkpnd qhfsrkvgts
dktkyylqdg
2221 dldvvfasss kcgkdktksv sstiffhcdp lvedgipefs hetadcqylf
swytsavcpl
2281 gvgfdsenpg ddgqmhkgls ersqavgavl slllvaltcc llalllykke
rretvisklt
2341 tccrrssnvs ykyskvnkee etdenetewl meeiqlpppr qgkegqengh
ittksvkals

```
     2401 slhgddqdse devltipevk vhsgrgagae sshpvrnaqs nalqereddr
vglvrgekar
     2461 kgksssaqqk tvsstklvsf hddsdedllh i
```

1A69

Official Symbol: HLA-A

Official Name: major histocompatibility complex, class I, A

Gene ID: 3105

Organism: Homo sapiens

Other Aliases: DAQB-90C11.16-002, HLAA

Other Designations: HLA class I histocompatibility antigen, A-1 alpha chain; MHC class I antigen **HLA-A** heavy chain; antigen presenting molecule; leukocyte antigen class I-A

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_002116.7

LOCUS NM_002116

ACCESSION NM_002116

```
      1 gagaagccaa tcagtgtcgt cgcggtcgct gttctaaagc ccgcacgcac
ccaccgggac
       61 tcagattctc cccagacgcc gaggatggcc gtcatggcgc cccgaaccct
cctcctgcta
      121 ctctcggggg ccctggccct gacccagacc tgggcgggct cccactccat
gaggtatttc
      181 ttcacatccg tgtcccggcc cggccgcggg gagccccgct tcatcgccgt
gggctacgtg
      241 gacgacacgc agttcgtgcg gttcgacagc gacgccgcga gccagaggat
ggagccgcgg
      301 gcgccgtgga tagagcagga ggggccggag tattgggacc aggagacacg
gaatgtgaag
      361 gcccagtcac agactgaccg agtggacctg gggaccctgc gcggctacta
caaccagagc
      421 gaggccggtt ctcacaccat ccagataatg tatggctgcg acgtggggtc
ggacgggcgc
      481 ttcctccgcg ggtaccggca ggacgcctac gacggcaagg attacatcgc
cctgaacgag
      541 gacctgcgct cttggaccgc ggcggacatg gcggctcaga tcaccaagcg
caagtgggag
      601 gcggcccatg aggcggagca gttgagagcc tacctggatg gcacgtgcgt
ggagtggctc
      661 cgcagatacc tggagaacgg gaaggagacg ctgcagcgca cggacccccc
caagacacat
      721 atgacccacc accccatctc tgaccatgag gccaccctga ggtgctgggc
cctgggcttc
      781 taccctgcgg agatcacact gacctggcag cgggatgggg aggaccagac
ccaggacacg
```

```
      841 gagctcgtgg agaccaggcc tgcaggggat ggaaccttcc agaagtgggc
ggctgtggtg
      901 gtgccttctg gagaggagca gagatacacc tgccatgtgc agcatgaggg
tctgcccaag
      961 cccctcaccc tgagatggga gctgtcttcc cagcccacca tccccatcgt
gggcatcatt
     1021 gctggcctgg ttctccttgg agctgtgatc actggagctg tggtcgctgc
cgtgatgtgg
     1081 aggaggaaga gctcagatag aaaaggaggg agttacactc aggctgcaag
cagtgacagt
     1141 gcccagggct ctgatgtgtc cctcacagct tgtaaagtgt gagacagctg
ccttgtgtgg
     1201 gactgagagg caagagttgt tcctgccctt ccctttgtga cttgaagaac
cctgactttg
     1261 tttctgcaaa ggcacctgca tgtgtctgtg ttcgtgtagg cataatgtga
ggaggtgggg
     1321 agaccacccc acccccatgt ccaccatgac cctcttccca cgctgacctg
tgctccctcc
     1381 ccaatcatct ttcctgttcc agagaggtgg ggctgaggtg tctccatctc
tgtctcaact
     1441 tcatggtgca ctgagctgta acttcttcct tccctattaa aattagaacc
ttagtataaa
     1501 tttactttct caaattcttg ccatgagagg ttgatgagtt aattaaagga
gaagattcct
     1561 aaaatttgag agacaaaata aatggaagac atgagaacct tccagagtcc
aaaaaaaaaa
     1621 aaaaaaaaaa aaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_002107.3

LOCUS NP_002107

ACCESSION NP_002107

```
      1 mavmaprtll lllsgalalt qtwagshsmr yfftsvsrpg rgeprfiavg
yvddtqfvrf
       61 dsdaasqrme prapwieqeg peywdqetrn vkaqsqtdrv dlgtlrgyyn
qseagshtiq
      121 imygcdvgsd grflrgyrqd aydgkdyial nedlrswtaa dmaaqitkrk
weaaheaeql
      181 rayldgtcve wlrrylengk etlqrtdppk thmthhpisd heatlrcwal
gfypaeitlt
      241 wqrdgedqtq dtelvetrpa gdgtfqkwaa vvvpsgeeqr ytchvqhegl
pkpltlrwel
      301 ssqptipivg iiaglvllga vitgavvaav mwrrkssdrk ggsytqaass
dsaqgsdvsl
      361 tackv
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001242758.1

LOCUS NM_001242758

ACCESSION NM_001242758

```
      1 gagaagccaa tcagtgtcgt cgcggtcgct gttctaaagt ccgcacgcac
ccaccgggac
       61 tcagattctc cccagacgcc gaggatggcc gtcatggcgc cccgaaccct
cctcctgcta
      121 ctctcggggg ccctggccct gacccagacc tgggcgggct cccactccat
gaggtatttc
      181 ttcacatccg tgtcccggcc cggccgcggg gagccccgct tcatcgccgt
gggctacgtg
      241 gacgacacgc agttcgtgcg gttcgacagc gacgccgcga gccagaagat
ggagccgcgg
      301 gcgccgtgga tagagcagga ggggccggag tattgggacc aggagacacg
gaatatgaag
      361 gcccactcac agactgaccg agcgaacctg gggaccctgc gcggctacta
caaccagagc
      421 gaggacggtt ctcacaccat ccagataatg tatggctgcg acgtggggcc
ggacgggcgc
      481 ttcctccgcg ggtaccggca ggacgcctac gacggcaagg attacatcgc
cctgaacgag
      541 gacctgcgct cttggaccgc ggcggacatg gcagctcaga tcaccaagcg
caagtgggag
      601 gcggtccatg cggcggagca gcggagagtc tacctggagg gccggtgcgt
ggacgggctc
      661 cgcagatacc tggagaacgg gaaggagacg ctgcagcgca cggacccccc
caagacacat
      721 atgacccacc accccatctc tgaccatgag gccaccctga ggtgctgggc
cctgggcttc
      781 taccctgcgg agatcacact gacctggcag cgggatgggg aggaccagac
ccaggacacg
      841 gagctcgtgg agaccaggcc tgcaggggat ggaaccttcc agaagtgggc
ggctgtggtg
      901 gtgccttctg gagaggagca gagatacacc tgccatgtgc agcatgaggg
tctgcccaag
      961 cccctcaccc tgagatggga gctgtcttcc cagcccacca tccccatcgt
gggcatcatt
     1021 gctggcctgg ttctccttgg agctgtgatc actggagctg tggtcgctgc
cgtgatgtgg
     1081 aggaggaaga gctcagatag aaaaggaggg agttacactc aggctgcaag
cagtgacagt
     1141 gcccagggct ctgatgtgtc tctcacagct tgtaaagtgt gagacagctg
ccttgtgtgg
     1201 gactgagagg caagagttgt tcctgccctt ccctttgtga cttgaagaac
cctgactttg
     1261 tttctgcaaa ggcacctgca tgtgtctgtg ttcgtgtagg cataatgtga
ggaggtgggg
     1321 agagcacccc acccccatgt ccaccatgac cctcttccca cgctgacctg
tgctccctct
     1381 ccaatcatct ttcctgttcc agagaggtgg ggctgaggtg tctccatctc
tgtctcaact
     1441 tcatggtgca ctgagctgta acttcttcct tccctattaa aattagaacc
tgagtataaa
     1501 tttacttttct caaattcttg ccatgagagg ttgatgagtt aattaaagga
gaagattcct
      1561 aaaatttgag agacaaaatt aatggaacgc atgagaacct tccagagtcc a
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001229687.1

LOCUS NP_001229687

ACCESSION NP_001229687

```
      1 mavmaprtll lllsgalalt qtwagshsmr yfftsvsrpg rgeprfiavg
yvddtqfvrf
       61 dsdaasqkme prapwieqeg peywdqetrn mkahsqtdra nlgtlrgyyn
qsedgshtiq
      121 imygcdvgpd grflrgyrqd aydgkdyial nedlrswtaa dmaaqitkrk
weavhaaeqr
      181 rvylegrcvd glrrylengk etlqrtdppk thmthhpisd heatlrcwal
gfypaeitlt
      241 wqrdgedqtq dtelvetrpa gdgtfqkwaa vvvpsgeeqr ytchvqhegl
pkpltlrwel
      301 ssqptipivg iiaglvllga vitgavvaav mwrrkssdrk ggsytqaass
dsaqgsdvsl
      361 tackv
```

P4HA2

Official Symbol: P4HA2

Official Name: prolyl 4-hydroxylase, alpha polypeptide II

Gene ID: 8974

Organism: Homo sapiens

Other Aliases: UNQ290/PRO330

Other Designations: 4-PH alpha 2; 4-PH alpha-2; C-P4Halpha(II); collagen prolyl 4-hydroxylase alpha(II); procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide II; procollagen-proline,2-oxoglutarate-4-dioxygenase subunit alpha-2; prolyl 4-hydroxylase subunit alpha-2

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_004199.2

LOCUS NM_004199

ACCESSION NM_004199

```
      1 agcgttgttt ttccttggca gctgcggaga cccgtgataa ttcgttaact
aattcaacaa
       61 acgggaccct tctgtgtgcc agaaaccgca agcagttgct aacccagtgg
gacaggcgga
      121 ttggaagagc gggaaggtcc tggcccagag cagtgtggtg agcgctgtgc
tggaagggaa
      181 tgcgggcagt gggtacttgg tagagcactg actgcctccg gccagaggac
ttcccggagg
      241 aggtgaccca tgagctggag tggtcagagg aaggctggca aaagggcatc
gtggacagag
      301 gaacagccta tgtgagtggg agcagagacc ttggccaatg ccattcctta
tggccttgta
      361 gtggaagcaa ggtgatgggg aaggaacact gtaggggata gctgtccacg
gacgctgtct
      421 acaagaccct ggagtgagat aacgtgcctg gtactgtgcc ctgcatgtgt
aagatgccca
```

481 gttgaccttc gcagcaggag cctggatcag ggcacttcct gcctcaggta
ttgctggaca
541 gcccagacac ttccctctgt gaccatgaaa ctctgggtgt ctgcattgct
gatggcctgg
601 tttggtgtcc tgagctgtgt gcaggccgaa ttcttcacct ctattgggca
catgactgac
661 ctgatttatg cagagaaaga gctggtgcag tctctgaaag agtacatcct
tgtggaggaa
721 gccaagcttt ccaagattaa gagctgggcc aacaaaatgg aagccttgac
tagcaagtca
781 gctgctgatg ctgagggcta cctggctcac cctgtgaatg cctacaaact
ggtgaagcgg
841 ctaaacacag actggcctgc gctggaggac cttgtcctgc aggactcagc
tgcaggtttt
901 atcgccaacc tctctgtgca gcggcagttc ttccccactg atgaggacga
gataggagct
961 gccaaagccc tgatgagact tcaggacaca tacaggctgg acccaggcac
aatttccaga
1021 ggggaacttc caggaaccaa gtaccaggca atgctgagtg tggatgactg
ctttgggatg
1081 ggccgctcgg cctacaatga aggggactat tatcatacgg tgttgtggat
ggagcaggtg
1141 ctaaagcagc ttgatgccgg ggaggaggcc accacaacca agtcacaggt
gctggactac
1201 ctcagctatg ctgtcttcca gttgggtgat ctgcaccgtg ccctggagct
cacccgccgc
1261 ctgctctccc ttgacccaag ccacgaacga gctggaggga atctgcggta
ctttgagcag
1321 ttattggagg aagagagaga aaaaacgtta acaaatcaga cagaagctga
gctagcaacc
1381 ccagaaggca tctatgagag gcctgtggac tacctgcctg agagggatgt
ttacgagagc
1441 ctctgtcgtg gggagggtgt caaactgaca ccccgtagac agaagaggct
tttctgtagg
1501 taccaccatg gcaacagggc cccacagctg ctcattgccc ccttcaaaga
ggaggacgag
1561 tgggacagcc cgcacatcgt caggtactac gatgtcatgt ctgatgagga
aatcgagagg
1621 atcaaggaga tcgcaaaacc taaacttgca cgagccaccg ttcgtgatcc
caagacagga
1681 gtcctcactg tcgccagcta ccgggtttcc aaaagctcct ggctagagga
agatgatgac
1741 cctgttgtgg cccgagtaaa tcgtcggatg cagcatatca cagggttaac
agtaaagact
1801 gcagaattgt tacaggttgc aaattatgga gtgggaggac agtatgaacc
gcacttcgac
1861 ttctctagga atgatgagcg agatactttc aagcatttag ggacggggaa
tcgtgtggct
1921 actttcttaa actacatgag tgatgtagaa gctggtggtg ccaccgtctt
ccctgatctg
1981 ggggctgcaa tttggcctaa gaagggtaca gctgtgttct ggtacaacct
cttgcggagc
2041 ggggaaggtg actaccgaac aagacatgct gcctgccctg tgcttgtggg
ctgcaagtgg
2101 gtctccaata agtggttcca tgaacgagga caggagttct tgagaccttg
tggatcaaca
2161 gaagttgact gacatccttt tctgtccttc cccttcctgg tccttcagcc
catgtcaacg
2221 tgacagacac ctttgtatgt tcctttgtat gttcctatca ggctgatttt
tggagaaatg

```
     2281 aatgtttgtc tggagcagag ggagaccata ctagggcgac tcctgtgtga
ctgaagtccc
     2341 agcccttcca ttcagcctgt gccatccctg gccccaaggc taggatcaaa
gtggctgcag
     2401 cagagttagc tgtctagcgc ctagcaaggt gcctttgtac ctcaggtgtt
ttaggtgtga
     2461 gatgtttcag tgaaccaaag ttctgatacc ttgtttacat gtttgttttt
atggcatttc
     2521 tatctattgt ggctttacca aaaaataaaa tgtccctacc agaagcctta
aaaaaaaaaa
     2581 aaaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_004190.1

LOCUS NP_004190

ACCESSION NP_004190

```
      1 mklwvsallm awfgvlscvq aefftsighm tdliyaekel vqslkeyilv
eeaklskiks
       61 wankmealts ksaadaegyl ahpvnayklv krlntdwpal edlvlqdsaa
gfianlsvqr
      121 qffptdedei gaakalmrlq dtyrldpgti srgelpgtky qamlsvddcf
gmgrsayneg
      181 dyyhtvlwme qvlkqldage eatttksqvl dylsyavfql gdlhralelt
rrllsldpsh
      241 eraggnlryf eqlleeerek tltnqteael atpegiyerp vdylperdvy
eslcrgegvk
      301 ltprrqkrlf cryhhgnrap qlliapfkee dewdsphivr yydvmsdeei
erikeiakpk
      361 laratvrdpk tgvltvasyr vsksswleed ddpvvarvnr rmqhitgltv
ktaellqvan
      421 ygvgggqyeph fdfsrnderd tfkhlgtgnr vatflnymsd veaggatvfp
dlgaaiwpkk
      481 gtavfwynll rsgegdyrtr haacpvlvgc kwvsnkwfhe rgqeflrpcg stevd
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001017973.1

LOCUS NM_001017973

ACCESSION NM_001017973

```
      1 agcgttgttt ttccttggca gctgcggaga cccgtgataa ttcgttaact
aattcaacaa
       61 acgggaccct tctgtgtgcc agaaaccgca agcagttgct aacccagtgg
gacaggcgga
      121 ttggaagagc gggaaggtcc tggcccagag cagtgtggtg agcgctgtgc
tggaagggaa
      181 tgcgggcagt gggtacttgg tagagcactg actgcctccg gccagaggac
ttcccggagg
      241 aggtgaccca tgagctggag tggtcagagg aaggctggca aaagggcatc
gtggacagag
      301 gaacagccta tgtgagtggg agcagagacc ttggccaatg ccattcctta
tggccttgta
```

361 gtggaagcaa ggtgatgggg aaggaacact gtaggggata gctgtccacg
gacgctgtct
421 acaagaccct ggagtgagat aacgtgcctg gtactgtgcc ctgcatgtgt
aagatgccca
481 gttgaccttc gcagcaggag cctggatcag ggcacttcct gcctcaggta
ttgctggaca
541 gcccagacac ttccctctgt gaccatgaaa ctctgggtgt ctgcattgct
gatggcctgg
601 tttggtgtcc tgagctgtgt gcaggccgaa ttcttcacct ctattgggca
catgactgac
661 ctgatttatg cagagaaaga gctggtgcag tctctgaaag agtacatcct
tgtggaggaa
721 gccaagcttt ccaagattaa gagctgggcc aacaaaatgg aagccttgac
tagcaagtca
781 gctgctgatg ctgagggcta cctggctcac cctgtgaatg cctacaaact
ggtgaagcgg
841 ctaaacacag actggcctgc gctggaggac cttgtcctgc aggactcagc
tgcaggtttt
901 atcgccaacc tctctgtgca gcggcagttc ttccccactg atgaggacga
gataggagct
961 gccaaagccc tgatgagact tcaggacaca tacaggctgg acccaggcac
aatttccaga
1021 ggggaacttc caggaaccaa gtaccaggca atgctgagtg tggatgactg
ctttgggatg
1081 ggccgctcgg cctacaatga aggggactat tatcatacgg tgttgtggat
ggagcaggtg
1141 ctaaagcagc ttgatgccgg ggaggaggcc accacaacca agtcacaggt
gctggactac
1201 ctcagctatg ctgtcttcca gttgggtgat ctgcaccgtg ccctggagct
cacccgccgc
1261 ctgctctccc ttgacccaag ccacgaacga gctggaggga atctgcggta
ctttgagcag
1321 ttattggagg aagagagaga aaaaacgtta acaaatcaga cagaagctga
gctagcaacc
1381 ccagaaggca tctatgagag gcctgtggac tacctgcctg agagggatgt
ttacgagagc
1441 ctctgtcgtg gggagggtgt caaactgaca ccccgtagac agaagaggct
tttctgtagg
1501 taccaccatg gcaacagggc cccacagctg ctcattgccc ccttcaaaga
ggaggacgag
1561 tgggacagcc cgcacatcgt caggtactac gatgtcatgt ctgatgagga
aatcgagagg
1621 atcaaggaga tcgcaaaacc taaacttgca cgagccaccg ttcgtgatcc
caagacagga
1681 gtcctcactg tcgccagcta ccgggtttcc aaaagctcct ggctagagga
agatgatgac
1741 cctgttgtgg cccgagtaaa tcgtcggatg cagcatatca cagggttaac
agtaaagact
1801 gcagaattgt tacaggttgc aaattatgga gtgggaggac agtatgaacc
gcacttcgac
1861 ttctctaggc gacctttga cagcggcctc aaaacagagg ggaataggtt
agcgacgttt
1921 cttaactaca tgagtgatgt agaagctggt ggtgccaccg tcttccctga
tctgggggct
1981 gcaatttggc ctaagaaggg tacagctgtg ttctggtaca acctcttgcg
gagcggggaa
2041 ggtgactacc gaacaagaca tgctgcctgc cctgtgcttg tgggctgcaa
gtgggtctcc
2101 aataagtggt tccatgaacg aggacaggag ttcttgagac cttgtggatc
aacagaagtt

```
     2161 gactgacatc cttttctgtc cttccccttc ctggtccttc agcccatgtc
aacgtgacag
     2221 acacctttgt atgttccttt gtatgttcct atcaggctga tttttggaga
aatgaatgtt
     2281 tgtctggagc agagggagac catactaggg cgactcctgt gtgactgaag
tcccagccct
     2341 tccattcagc ctgtgccatc cctggcccca aggctaggat caaagtggct
gcagcagagt
     2401 tagctgtcta gcgcctagca aggtgccttt gtacctcagg tgttttaggt
gtgagatgtt
     2461 tcagtgaacc aaagttctga taccttgttt acatgtttgt ttttatggca
tttctatcta
     2521 ttgtggcttt accaaaaaat aaaatgtccc taccagaagc cttaaaaaaa
aaaaaaaaaa
     2581 aa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001017973.1

LOCUS NP_001017973

ACCESSION NP_001017973

```
      1 mklwvsallm awfgvlscvq aefftsighm tdliyaekel vqslkeyilv
eeaklskiks
       61 wankmealts ksaadaegyl ahpvnayklv krlntdwpal edlvlqdsaa
gfianlsvqr
      121 qffptdedei gaakalmrlq dtyrldpgti srgelpgtky qamlsvddcf
gmgrsayneg
      181 dyyhtvlwme qvlkqldage eatttksqvl dylsyavfql gdlhralelt
rrllsldpsh
      241 eraggnlryf eqlleeerek tltnqteael atpegiyerp vdylperdvy
eslcrgegvk
      301 ltprrqkrlf cryhhgnrap qlliapfkee dewdsphivr yydvmsdeei
erikeiakpk
      361 laratvrdpk tgvltvasyr vsksswleed ddpvvarvnr rmqhitgltv
ktaellqvan
      421 ygvgggqyeph fdfsrrpfds glktegnrla tflnymsdve aggatvfpdl
gaaiwpkkgt
      481 avfwynllrs gegdyrtrha acpvlvgckw vsnkwfherg qeflrpcgst evd
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_001017974.1

LOCUS NM_001017974

ACCESSION NM_001017974

```
      1 aagggaggag gcgccgagct gaccgggcga cgccgcggga ggttctggaa
acgccgggag
       61 ctgcgagtgt ccagacactt ccctctgtga ccatgaaact ctgggtgtct
gcattgctga
      121 tggcctggtt tggtgtcctg agctgtgtgc aggccgaatt cttcacctct
attgggcaca
      181 tgactgacct gatttatgca gagaaagagc tggtgcagtc tctgaaagag
tacatccttg
```

241 tggaggaagc caagctttcc aagattaaga gctgggccaa caaaatggaa gccttgacta
301 gcaagtcagc tgctgatgct gagggctacc tggctcaccc tgtgaatgcc tacaaactgg
361 tgaagcggct aaacacagac tggcctgcgc tggaggacct tgtcctgcag gactcagctg
421 caggttttat cgccaacctc tctgtgcagc ggcagttctt ccccactgat gaggacgaga
481 taggagctgc caaagccctg atgagacttc aggacacata caggctggac ccaggcacaa
541 tttccagagg ggaacttcca ggaaccaagt accaggcaat gctgagtgtg gatgactgct
601 ttgggatggg ccgctcggcc tacaatgaag gggactatta tcatacggtg ttgtggatgg
661 agcaggtgct aaagcagctt gatgccgggg aggaggccac cacaaccaag tcacaggtgc
721 tggactacct cagctatgct gtcttccagt tgggtgatct gcaccgtgcc ctggagctca
781 cccgccgcct gctctccctt gacccaagcc acgaacgagc tggagggaat ctgcggtact
841 ttgagcagtt attggaggaa gagagagaaa aaacgttaac aaatcagaca gaagctgagc
901 tagcaacccc agaaggcatc tatgagaggc ctgtggacta cctgcctgag agggatgttt
961 acgagagcct ctgtcgtggg gagggtgtca aactgacacc ccgtagacag aagaggcttt
1021 tctgtaggta ccaccatggc aacagggccc cacagctgct cattgccccc ttcaaagagg
1081 aggacgagtg ggacagcccg cacatcgtca ggtactacga tgtcatgtct gatgaggaaa
1141 tcgagaggat caaggagatc gcaaaaccta aacttgcacg agccaccgtt cgtgatccca
1201 agacaggagt cctcactgtc gccagctacc gggtttccaa aagctcctgg ctagaggaag
1261 atgatgaccc tgttgtggcc cgagtaaatc gtcggatgca gcatatcaca gggttaacag
1321 taaagactgc agaattgtta caggttgcaa attatggagt gggaggacag tatgaaccgc
1381 acttcgactt ctctaggcga ccttttgaca gcggcctcaa aacagagggg aataggttag
1441 cgacgtttct taactacatg agtgatgtag aagctggtgg tgccaccgtc ttccctgatc
1501 tgggggctgc aatttggcct aagaagggta cagctgtgtt ctggtacaac ctcttgcgga
1561 gcggggaagg tgactaccga acaagacatg ctgcctgccc tgtgcttgtg ggctgcaagt
1621 gggtctccaa taagtggttc catgaacgag gacaggagtt cttgagacct tgtggatcaa
1681 cagaagttga ctgacatcct tttctgtcct tccccttcct ggtccttcag cccatgtcaa
1741 cgtgacagac acctttgtat gttcctttgt atgttcctat caggctgatt tttggagaaa
1801 tgaatgtttg tctggagcag agggagacca tactagggcg actcctgtgt gactgaagtc
1861 ccagcccttc cattcagcct gtgccatccc tggccccaag gctaggatca aagtggctgc
1921 agcagagtta gctgtctagc gcctagcaag gtgcctttgt acctcaggtg ttttaggtgt
1981 gagatgtttc agtgaaccaa agttctgata ccttgtttac atgtttgttt ttatggcatt

```
     2041 tctatctatt gtggctttac caaaaaataa aatgtcccta ccagaagcct
taaaaaaaaa
     2101 aaaaaaaaaa
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_001017974.1

LOCUS NP_001017974

ACCESSION NP_001017974

```
      1 mklwvsallm awfgvlscvq aefftsighm tdliyaekel vqslkeyilv
eeaklskiks
       61 wankmealts ksaadaegyl ahpvnayklv krlntdwpal edlvlqdsaa
gfianlsvqr
      121 qffptdedei gaakalmrlq dtyrldpgti srgelpgtky qamlsvddcf
gmgrsayneg
      181 dyyhtvlwme qvlkqldage eatttksqvl dylsyavfql gdlhralelt
rrllsldpsh
      241 eraggnlryf eqlleeerek tltnqteael atpegiyerp vdylperdvy
eslcrgegvk
      301 ltprrqkrlf cryhhgnrap qlliapfkee dewdsphivr yydvmsdeei
erikeiakpk
      361 laratvrdpk tgvltvasyr vsksswleed ddpvvarvnr rmqhitgltv
ktaellqvan
      421 ygvgggqyeph fdfsrrpfds glktegnrla tflnymsdve aggatvfpdl
gaaiwpkkgt
      481 avfwynllrs gegdyrtrha acpvlvgckw vsnkwfherg qeflrpcgst evd
```

Nucleotide sequence (variant 4):

NCBI Reference Sequence: NM_001142598.1

LOCUS NM_001142598

ACCESSION NM_001142598

```
      1 aagggaggag gcgccgagct gaccgggcga cgccgcggga ggttctggaa
acgccgggag
       61 ctgcgagtgt ccagctgcgg agacccgtga taattcgtta actaattcaa
caaacgggac
      121 ccttctgtgt gccagaaacc gcaagcagtt gctaacccag tgggacaggc
ggattggaag
      181 agcgggaagg tcctggccca gagcagtgtg acacttccct ctgtgaccat
gaaactctgg
      241 gtgtctgcat tgctgatggc ctggtttggt gtcctgagct gtgtgcaggc
cgaattcttc
      301 acctctattg ggcacatgac tgacctgatt tatgcagaga aagagctggt
gcagtctctg
      361 aaagagtaca tccttgtgga ggaagccaag ctttccaaga ttaagagctg
ggccaacaaa
      421 atggaagcct tgactagcaa gtcagctgct gatgctgagg gctacctggc
tcaccctgtg
      481 aatgcctaca aactggtgaa gcggctaaac acagactggc ctgcgctgga
ggaccttgtc
      541 ctgcaggact cagctgcagg ttttatcgcc aacctctctg tgcagcggca
gttcttcccc
```

```
      601 actgatgagg acgagatagg agctgccaaa gccctgatga gacttcagga
cacatacagg
      661 ctggacccag gcacaatttc cagaggggaa cttccaggaa ccaagtacca
ggcaatgctg
      721 agtgtggatg actgctttgg gatgggccgc tcggcctaca atgaagggga
ctattatcat
      781 acggtgttgt ggatggagca ggtgctaaag cagcttgatg ccggggagga
ggccaccaca
      841 accaagtcac aggtgctgga ctacctcagc tatgctgtct tccagttggg
tgatctgcac
      901 cgtgccctgg agctcacccg ccgcctgctc tcccttgacc caagccacga
acgagctgga
      961 gggaatctgc ggtactttga gcagttattg gaggaagaga gagaaaaaac
gttaacaaat
     1021 cagacagaag ctgagctagc aaccccagaa ggcatctatg agaggcctgt
ggactacctg
     1081 cctgagaggg atgtttacga gagcctctgt cgtggggagg gtgtcaaact
gacaccccgt
     1141 agacagaaga ggcttttctg taggtaccac catggcaaca gggccccaca
gctgctcatt
     1201 gccccccttca aagaggagga cgagtgggac agcccgcaca tcgtcaggta
ctacgatgtc
     1261 atgtctgatg aggaaatcga gaggatcaag gagatcgcaa aacctaaact
tgcacgagcc
     1321 accgttcgtg atcccaagac aggagtcctc actgtcgcca gctaccgggt
ttccaaaagc
     1381 tcctggctag aggaagatga tgaccctgtt gtggcccgag taaatcgtcg
gatgcagcat
     1441 atcacagggt taacagtaaa gactgcagaa ttgttacagg ttgcaaatta
tggagtggga
     1501 ggacagtatg aaccgcactt cgacttctct aggcgacctt ttgacagcgg
cctcaaaaca
     1561 gaggggaata ggttagcgac gtttcttaac tacatgagtg atgtagaagc
tggtggtgcc
     1621 accgtcttcc ctgatctggg ggctgcaatt tggcctaaga agggtacagc
tgtgttctgg
     1681 tacaacctct tgcggagcgg ggaaggtgac taccgaacaa gacatgctgc
ctgccctgtg
     1741 cttgtgggct gcaagtgggt ctccaataag tggttccatg aacgaggaca
ggagttcttg
     1801 agaccttgtg gatcaacaga agttgactga catccttttc tgtccttccc
cttcctggtc
     1861 cttcagccca tgtcaacgtg acagacacct ttgtatgttc ctttgtatgt
tcctatcagg
     1921 ctgatttttg gagaaatgaa tgtttgtctg gagcagaggg agaccatact
agggcgactc
     1981 ctgtgtgact gaagtcccag cccttccatt cagcctgtgc catccctggc
cccaaggcta
     2041 ggatcaaagt ggctgcagca gagttagctg tctagcgcct agcaaggtgc
ctttgtacct
     2101 caggtgtttt aggtgtgaga tgtttcagtg aaccaaagtt ctgatacctt
gtttacatgt
     2161 ttgtttttat ggcatttcta tctattgtgg ctttaccaaa aaataaaatg
tccctaccag
     2221 aagccttaaa aaaaaaaaaa aaaaaa
```

Protein sequence (variant 4):

NCBI Reference Sequence: NP_001136070.1

LOCUS NP_001136070

ACCESSION NP_001136070

```
      1 mklwvsallm awfgvlscvq aefftsighm tdliyaekel vqslkeyilv
eeaklskiks
       61 wankmealts ksaadaegyl ahpvnayklv krlntdwpal edlvlqdsaa
gfianlsvqr
      121 qffptdedei gaakalmrlq dtyrldpgti srgelpgtky qamlsvddcf
gmgrsayneg
      181 dyyhtvlwme qvlkqldage eatttksqvl dylsyavfql gdlhralelt
rrllsldpsh
      241 eraggnlryf eqlleeerek tltnqteael atpegiyerp vdylperdvy
eslcrgegvk
      301 ltprrqkrlf cryhhgnrap qlliapfkee dewdsphivr yydvmsdeei
erikeiakpk
      361 laratvrdpk tgvltvasyr vsksswleed ddpvvarvnr rmqhitgltv
ktaellqvan
      421 ygvgggqyeph fdfsrrpfds glktegnrla tflnymsdve aggatvfpdl
gaaiwpkkgt
      481 avfwynllrs gegdyrtrha acpvlvgckw vsnkwfherg qeflrpcgst evd
```

Nucleotide sequence: (variant 5)

NCBI Reference Sequence: NM_001142599.1

LOCUS NM_001142599

ACCESSION NM_001142599

```
      1 aagggaggag gcgccgagct gaccgggcga cgccgcggga ggttctggaa
acgccgggag
       61 ctgcgagtgt ccagctgcgg agacccgtga taattcgtta actaattcaa
caaacgggac
      121 ccttctgtgt gccagaaacc gcaagcagtt gctaacccag tgggacaggc
ggattggaag
      181 agcgggaagg tcctggccca gagcagtgtg acacttccct ctgtgaccat
gaaactctgg
      241 gtgtctgcat tgctgatggc ctggtttggt gtcctgagct gtgtgcaggc
cgaattcttc
      301 acctctattg ggcacatgac tgacctgatt tatgcagaga aagagctggt
gcagtctctg
      361 aaagagtaca tccttgtgga ggaagccaag ctttccaaga ttaagagctg
ggccaacaaa
      421 atggaagcct tgactagcaa gtcagctgct gatgctgagg gctacctggc
tcaccctgtg
      481 aatgcctaca aactggtgaa gcggctaaac acagactggc ctgcgctgga
ggaccttgtc
      541 ctgcaggact cagctgcagg ttttatcgcc aacctctctg tgcagcggca
gttcttcccc
      601 actgatgagg acgagatagg agctgccaaa gccctgatga gacttcagga
cacatacagg
      661 ctggacccag gcacaatttc cagaggggaa cttccaggaa ccaagtacca
ggcaatgctg
      721 agtgtggatg actgctttgg gatgggccgc tcggcctaca atgaagggga
ctattatcat
      781 acggtgttgt ggatggagca ggtgctaaag cagcttgatg ccggggagga
ggccaccaca
      841 accaagtcac aggtgctgga ctacctcagc tatgctgtct tccagttggg
tgatctgcac
```

```
      901 cgtgccctgg agctcacccg ccgcctgctc tcccttgacc caagccacga
acgagctgga
      961 gggaatctgc ggtactttga gcagttattg gaggaagaga gagaaaaaac
gttaacaaat
     1021 cagacagaag ctgagctagc aaccccagaa ggcatctatg agaggcctgt
ggactacctg
     1081 cctgagaggg atgtttacga gagcctctgt cgtggggagg gtgtcaaact
gacaccccgt
     1141 agacagaaga ggcttttctg taggtaccac catggcaaca gggccccaca
gctgctcatt
     1201 gccccccttca aagaggagga cgagtgggac agcccgcaca tcgtcaggta
ctacgatgtc
     1261 atgtctgatg aggaaatcga gaggatcaag gagatcgcaa aacctaaact
tgcacgagcc
     1321 accgttcgtg atcccaagac aggagtcctc actgtcgcca gctaccgggt
ttccaaaagc
     1381 tcctggctag aggaagatga tgaccctgtt gtggcccgag taaatcgtcg
gatgcagcat
     1441 atcacagggt taacagtaaa gactgcagaa ttgttacagg ttgcaaatta
tggagtggga
     1501 ggacagtatg aaccgcactt cgacttctct aggaatgatg agcgagatac
tttcaagcat
     1561 ttagggacgg ggaatcgtgt ggctactttc ttaaactaca tgagtgatgt
agaagctggt
     1621 ggtgccaccg tcttccctga tctgggggct gcaatttggc ctaagaaggg
tacagctgtg
     1681 ttctggtaca acctcttgcg gagcggggaa ggtgactacc gaacaagaca
tgctgcctgc
     1741 cctgtgcttg tgggctgcaa gtgggtctcc aataagtggt tccatgaacg
aggacaggag
     1801 ttcttgagac cttgtggatc aacagaagtt gactgacatc cttttctgtc
cttccccttc
     1861 ctggtccttc agcccatgtc aacgtgacag acacctttgt atgttccttt
gtatgttcct
     1921 atcaggctga tttttggaga aatgaatgtt tgtctggagc agagggagac
catactaggg
     1981 cgactcctgt gtgactgaag tcccagccct tccattcagc ctgtgccatc
cctggcccca
     2041 aggctaggat caaagtggct gcagcagagt tagctgtcta gcgcctagca
aggtgccttt
     2101 gtacctcagg tgttttaggt gtgagatgtt tcagtgaacc aaagttctga
taccttgttt
     2161 acatgtttgt ttttatggca tttctatcta ttgtggcttt accaaaaaat
aaaatgtccc
     2221 taccagaagc cttaaaaaaa aaaaaaaaaa aa
```

Protein sequence:

NCBI Reference Sequence: NP_001136071.1

LOCUS NP_001136071

ACCESSION NP_001136071

```
      1 mklwvsallm awfgvlscvq aefftsighm tdliyaekel vqslkeyilv
eeaklskiks
       61 wankmealts ksaadaegyl ahpvnayklv krlntdwpal edlvlqdsaa
gfianlsvqr
```

```
      121 qffptdedei gaakalmrlq dtyrldpgti srgelpgtky qamlsvddcf
gmgrsayneg
      181 dyyhtvlwme qvlkqldage eatttksqvl dylsyavfql gdlhralelt
rrllsldpsh
      241 eraggnlryf eqlleeerek tltnqteael atpegiyerp vdylperdvy
eslcrgegvk
      301 ltprrqkrlf cryhhgnrap qlliapfkee dewdsphivr yydvmsdeei
erikeiakpk
      361 laratvrdpk tgvltvasyr vsksswleed ddpvvarvnr rmqhitgltv
ktaellqvan
      421 ygvgggqyeph fdfsrnderd tfkhlgtgnr vatflnymsd veaggatvfp
dlgaaiwpkk
      481 gtavfwynll rsgegdyrtr haacpvlvgc kwvsnkwfhe rgqeflrpcg stevd
```

HNRPG

Official Symbol: RBMX

Official Name: RNA binding motif protein, X-linked

Gene ID: 27316

Organism: Homo sapiens

Other Aliases: RP11-1114A5.1, **HNRPG**, RBMXP1, RBMXRT, RNMX, hnRNP-G

Other Designations: RNA binding motif protein, X chromosome; RNA-binding motif protein, X chromosome; glycoprotein p43; heterogeneous nuclear ribonucleoprotein G; hnRNP G

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_002139.3

LOCUS NM_002139

ACCESSION NM_002139

```
      1 ggtccttcag cctcgttccc gggcagtata aagtttgctg tctcctttgt
tcgccctcgt
       61 tgcgcagtag tgctagcggc ttcgcggttc ggtcctcgca cccggcagcc
gccactggtg
      121 ctgagctgct aggaagcccc tatcgccgag ctcgttggag cttgaaccca
ttgtcacccc
      181 tccgactcac cggcccaaaa aaaaaaaaac atggttgaag cagatcgccc
aggaaagctc
      241 ttcattggtg ggcttaatac ggaaacaaat gagaaagctc ttgaagcagt
atttggcaaa
      301 tatggacgaa tagtggaagt actcttgatg aaagaccgtg aaaccaacaa
atcaagagga
      361 tttgcttttg tcacctttga aagcccagca gacgctaagg atgcagccag
agacatgaat
      421 ggaaagtcat tagatggaaa agccatcaag gtggaacaag ccaccaaacc
atcatttgaa
```

```
      481 agtggtagac gtggaccgcc tccacctcca agaagtagag gccctccaag
aggtcttaga
      541 ggtggaagag gaggaagtgg aggaaccagg ggacctccct cacggggagg
acacatggat
      601 gacggtggat attccatgaa ttttaacatg agttcttcca ggggaccact
cccagtaaaa
      661 agaggaccac caccaagaag tgggggtcct cctcctaaga gatctgcacc
ttcaggacca
      721 gttcgcagta gcagtggaat gggaggaaga gctcctgtat cacgtggaag
agatagttat
      781 ggaggtccac ctcgaaggga accgctgccc tctcgtagag atgtttattt
gtccccaaga
      841 gatgatgggt attctactaa agacagctat tcaagcagag attacccaag
ttctcgtgat
      901 actagagatt atgcaccacc accacgagat tatacttacc gtgattatgg
tcattccagt
      961 tcacgtgatg actatccatc aagaggatat agcgatagag atggatatgg
tcgtgatcgt
     1021 gactattcag atcatccaag tggaggttcc tacagagatt catatgagag
ttatggtaac
     1081 tcacgtagtg ctccacctac acgagggccc ccgccatctt atggtggaag
cagtcgctat
     1141 gatgattaca gcagctcacg tgacggatat ggtggaagtc gagacagtta
ctcaagcagc
     1201 cgaagtgatc tctactcaag tggtcgtgat cgggttggca gacaagaaag
agggcttccc
     1261 ccttctatgg aaaggggggta ccctcctcca cgtgattcct acagcagttc
aagccgcgga
     1321 gcaccaagag gtggtggccg tggaggaagc cgatctgata gaggggggagg
cagaagcaga
     1381 tactagaaac aaacaaaact ttggaccaaa atcccagttc aaagaaacaa
aaagtggaaa
     1441 ctattctatc ataactaccc aaggactact aaaaggaaaa attgtgttac
ttttttaaa
     1501 ttccctgtta agttcccctc cataattttt atgttcttgt gaggaaaaaa
gtaaaacatg
     1561 tttaatttta tttgactttc gcattgcttt tcaacaagca aatgttaaat
gtgttaagac
     1621 ttgtactagt gttgtaactt tccaagtaaa agtatcccct aaaggccact
tcctatctga
     1681 tttttcccag caaatgaggc aggcaattct aagatcttcc acaaaacatc
tagccatcta
     1741 aaatggagag atgaatcatt ctacctatac aaacaagcta gctattagag
ggtggttggg
     1801 gtatgctact cataagattt cagggtgtct tccaactgaa atctcaatgt
tctcagtacg
     1861 aaaaacctga aatcacatgc ctatgtaagg aaagtgctat tcacccagta
aacccaaaaa
     1921 agcaaatgga taatgctggc cattttgcct ttctgacatt tccttgggaa
tctgcaagaa
     1981 cctccccttt cccttccccc aataagacca tttaagtgtg tgttaaacaa
ctacagaata
      2041 ctaaataaaa agtttggcca aaaccaacca tgaagctgca aaaaaaaaaa
aaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_002130.2

LOCUS NP_002130

ACCESSION NP_002130

```
      1 mveadrpgkl figglntetn ekaleavfgk ygrivevllm kdretnksrg
fafvtfespa
       61 dakdaardmn gksldgkaik veqatkpsfe sgrrgppppp rsrgpprglr
ggrggsggtr
      121 gppsrgghmd dggysmnfnm sssrgplpvk rgppprsggp ppkrsapsgp
vrsssgmggr
      181 apvsrgrdsy ggpprreplp srrdvylspr ddgystkdsy ssrdypssrd
trdyappprd
      241 ytyrdyghss srddypsrgy sdrdgygrdr dysdhpsggs yrdsyesygn
srsapptrgp
      301 ppsyggssry ddysssrdgy ggsrdsysss rsdlyssgrd rvgrqerglp
psmergyppp
      361 rdsysssssrg aprgggrggs rsdrgggrsr y
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001164803.1

LOCUS NM_001164803

ACCESSION NM_001164803

```
      1 ggtccttcag cctcgttccc gggcagtata aagtttgctg tctcctttgt
tcgccctcgt
       61 tgcgcagtag tgctagcggc ttcgcggttc ggtcctcgca cccggcagcc
gccactggtg
      121 ctgagctgct aggaagcccc tatcgccgag ctcgttggag cttgaaccca
ttgtcacccc
      181 tccgactcac cggcccaaaa aaaaaaaaac atggttgaag cagatcgccc
aggaaagctc
      241 ttcattggtg ggcttaatac ggaaacaaat gagaaagctc ttgaagcagt
atttggcaaa
      301 tatggacgaa tagtggaagt actcttgatg aaagaccgtg aaaccaacaa
atcaagagga
      361 tttgcttttg tcacctttga aagcccagca gacgctaagg atgcagccag
agacatgaat
      421 ggaaagctcc tgtatcacgt ggaagagata gttatggagg tccacctcga
agggaaccgc
      481 tgccctctcg tagagatgtt tatttgtccc caagagatga tgggtattct
actaaagaca
      541 gctattcaag cagagattac ccaagttctc gtgatactag agattatgca
ccaccaccac
      601 gagattatac ttaccgtgat tatggtcatt ccagttcacg tgatgactat
ccatcaagag
      661 gatatagcga tagagatgga tatggtcgtg atcgtgacta ttcagatcat
ccaagtggag
      721 gttcctacag agattcatat gagagttatg gttggtgatt ttgctcatta
tggtcgtgga
      781 gtgctgattg attcacagta gataaagctg gcagtaagaa atgctaagag
ttgttgaagc
      841 agaaggcggc tgattgtcaa taagtcacta cagttgcata agcagtgctg
tcagaattgg
      901 tttggtgcag gcaatagatt ttgccttcag gggttcctgt ggatctgagg
aaggcatcag
      961 tgttgattaa cactcataac tagggagtga ctggtagtta cttaaagcaa
gtaattgacc
```

```
     1021 aaatggaaaa ggggaagtaa ttaaggaaat tggtaagtgg aggtagtcag
gaagttcttg
     1081 tggttcttca catagatttt acagctttgg ctttcatttt gtttagctaa
agtcatgggg
     1141 acaactcttc aatttagaac ttaagttgaa ttataaaaat gatggatata
agtggtagct
     1201 gtatctagtg aagtgtctgt cagtaagtga aacatttttt ggtggtggct
tatccacaaa
     1261 cagtttagtt gtagaataaa acttatgagt gacatctgga aagtaaccat
gctaagatgg
     1321 caagcacact ggaaacaatt aggccacttg gctttctttt gctgtattgt
tttataagcc
     1381 tactttacct cccagtcttg gaaacaagtt ttagtttttt attggtttgg
agactagagc
     1441 caatagtata atgttctcaa aggaaacaga cttgagttgt tggattagag
gaactaaccc
     1501 aacttatatg attttttttt tgtttttgtc gtgtagttat ggcactgtct
tatttggaac
     1561 atttgcaact agggataata caacattttt aactctcatt tgacaaccta
ctactaatca
     1621 cagaccacaa gggtaatgac caaatttatg tggtttttgc actccatagt
tgtcttagcc
     1681 caatctttct atactcttac gattacttgg gttaacgctt ctgtgaggac
cttctggctc
     1741 ttgagatacc ctaaatattt aagatattta gatatcttga agatagtata
ggatatagag
     1801 attgtaccaa ataggaatat aaggagtatg ttaaaatgac cagatacctg
tttgatagtt
     1861 tactgaccta gcagatgtgt ggaaaaggaa tcagatcttg attcttctgg
gtttatactg
     1921 gttgtaaaac agaatgatac agaaaatgtt ttccttgttt aactggtagt
tgaacataga
     1981 acttgggtat tatagatcac ttttcacttt ttggaatgtt ttgtattgaa
acttaataaa
     2041 actttaacat ggaaaaaaaa aaaaaaaaaa a
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001158275.1

LOCUS NP_001158275

ACCESSION NP_001158275

```
      1 mveadrpgkl figglntetn ekaleavfgk ygrivevllm kdretnksrg
fafvtfespa
       61 dakdaardmn gkllyhveei vmevhlegnr cplvemficp qemmgillkt
aiqaeitqvl
      121 vileimhhhh eiiltvimvi pvhvmtihqe diaiemdmvv ivtiqiiqve
vpteihmrvm
      181 vgdfahygrg vlidsq
```

IBP7

Official Symbol: IGFBP7

Official Name: insulin-like growth factor binding protein 7

Gene ID: 3490

Organism: Homo sapiens

Other Aliases: AGM, FSTL2, IBP-7, IGFBP-7, IGFBP-7v, IGFBPRP1, MAC25, PSF, RAMSVPS, TAF

Other Designations: IGF-binding protein 7; IGFBP-rP1; PGI2-stimulating factor; angiomodulin; insulin-like growth factor-binding protein 7; prostacyclin-stimulating factor; tumor-derived adhesion factor

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001553.2

LOCUS NM_001553

ACCESSION NM_001553

```
      1 actcgcgccc ttgccgctgc caccgcaccc cgccatggag cggccgtcgc
tgcgcgccct
       61 gctcctcggc gccgctgggc tgctgctcct gctcctgccc ctctcctctt
cctcctcttc
      121 ggacacctgc ggcccctgcg agccggcctc ctgcccgccc ctgcccccgc
tgggctgcct
      181 gctgggcgag acccgcgacg cgtgcggctg ctgccctatg tgcgcccgcg
gcgagggcga
      241 gccgtgcggg ggtggcggcg ccggcagggg gtactgcgcg ccgggcatgg
agtgcgtgaa
      301 gagccgcaag aggcggaagg gtaaagccgg ggcagcagcc ggcggtccgg
gtgtaagcgg
      361 cgtgtgcgtg tgcaagagcc gctacccggt gtgcggcagc gacggcacca
cctacccgag
      421 cggctgccag ctgcgcgccg ccagccagag ggccgagagc cgcggggaga
aggccatcac
      481 ccaggtcagc aagggcacct gcgagcaagg tccttccata gtgacgcccc
ccaaggacat
      541 ctggaatgtc actggtgccc aggtgtactt gagctgtgag gtcatcggaa
tcccgacacc
      601 tgtcctcatc tggaacaagg taaaaagggg tcactatgga gttcaaagga
cagaactcct
      661 gcctggtgac cgggacaacc tggccattca gacccggggt ggcccagaaa
agcatgaagt
      721 aactggctgg gtgctggtat ctcctctaag taaggaagat gctggagaat
atgagtgcca
      781 tgcatccaat tcccaaggac aggcttcagc atcagcaaaa attacagtgg
ttgatgcctt
      841 acatgaaata ccagtgaaaa aaggtgaagg tgccgagcta taaacctcca
gaatattatt
      901 agtctgcatg gttaaaagta gtcatggata actacattac ctgttcttgc
ctaataagtt
      961 tcttttaatc caatccacta acactttagt tatattcact ggttttacac
agagaaatac
     1021 aaaataaaga tcacacatca agactatcta caaaaattta ttatatattt
acagaagaaa
     1081 agcatgcata tcattaaaca aataaaatac tttttatcac aacacagtaa
aaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001544.1

LOCUS NP_001544

ACCESSION NP_001544

```
      1 merpslrall lgaagllllll lplsssssssd tcgpcepasc pplpplgcll
getrdacgcc
       61 pmcargegep cggggagrgy capgmecvks rkrrkgkaga aaggpgvsgv
cvcksrypvc
      121 gsdgttypsg cqlraasqra esrgekaitq vskgtceqgp sivtppkdiw
nvtgaqvyls
      181 cevigiptpv liwnkvkrgh ygvqrtellp gdrdnlaiqt rggpekhevt
gwvlvsplsk
      241 edageyecha snsqgqasas akitvvdalh eipvkkgega el
```

Nucleotide sequence (variant 2)

NCBI Reference Sequence: NM_001253835.1

LOCUS NM_001253835

ACCESSION NM_001253835

```
      1 actcgcgccc ttgccgctgc caccgcaccc cgccatggag cggccgtcgc
tgcgcgccct
       61 gctcctcggc gccgctgggc tgctgctcct gctcctgccc ctctcctctt
cctcctcttc
      121 ggacacctgc ggcccctgcg agccggcctc ctgcccgccc ctgccccgc
tgggctgcct
      181 gctgggcgag acccgcgacg cgtgcggctg ctgccctatg tgcgcccgcg
gcgagggcga
      241 gccgtgcggg ggtggcggcg ccggcagggg gtactgcgcg ccgggcatgg
agtgcgtgaa
      301 gagccgcaag aggcggaagg gtaaagccgg ggcagcagcc ggcggtccgg
gtgtaagcgg
      361 cgtgtgcgtg tgcaagagcc gctacccggt gtgcggcagc gacggcacca
cctacccgag
      421 cggctgccag ctgcgcgccg ccagccagag ggccgagagc cgcggggaga
aggccatcac
      481 ccaggtcagc aagggcacct gcgagcaagg tccttccata gtgacgcccc
ccaaggacat
      541 ctggaatgtc actggtgccc aggtgtactt gagctgtgag gtcatcggaa
tcccgacacc
      601 tgtcctcatc tggaacaagg taaaaagggg tcactatgga gttcaaagga
cagaactcct
      661 gcctggtgac cgggacaacc tggccattca gacccggggt ggcccagaaa
agcatgaagt
      721 aactggctgg gtgctggtat ctcctctaag taaggaagat gctggagaat
atgagtgcca
      781 tgcatccaat tcccaaggac aggcttcagc atcagcaaaa attacagtgg
ttgatgcctt
      841 acatgaaata ccagtgaaaa aaggtacaca ataaatctca cagccattta
aaaatgacta
      901 gtacatttgc tttaaaaaga acagaactaa gtatgaaagt atcagacgta
gctattgatg
      961 aaattctgta gttagcaacc cataagggca ttaagtatgc cattaaaatg
tacagcatga
     1021 gactccaaaa gattatctgg atgggtgact g
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001240764.1

LOCUS NP_001240764

ACCESSION NP_001240764

```
      1 merpslrall lgaagllllll lplsssssssd tcgpcepasc pplpplgcll
getrdacgcc
       61 pmcargegep cggggagrgy capgmecvks rkrrkgkaga aaggpgvsgv
cvcksrypvc
      121 gsdgttypsg cqlraasqra esrgekaitq vskgtceqgp sivtppkdiw
nvtgaqvyls
      181 cevigiptpv liwnkvkrgh ygvqrtellp gdrdnlaiqt rggpekhevt
gwvlvsplsk
      241 edageyecha snsqgqasas akitvvdalh eipvkkgtq
```

1C17

Official Symbol: HLA-C

Official Name: major histocompatibility complex, class I, C

Gene ID: 3107

Organism: Homo sapiens

Other Aliases: XXbac-BCX101P6.2, D6S204, HLA-JY3, HLC-C, PSORS1

Other Designations: HLA class I histocompatibility antigen, C alpha chain; HLA class I histocompatibility antigen, Cw-1 alpha chain; MHC class I antigen heavy chain **HLA-C**; human leukocyte antigen-C alpha chain; major histocompatibility antigen **HLA-C**

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_002117.5

LOCUS NM_002117

ACCESSION NM_002117

```
      1 tccgcagtcc cggttctaaa gtccccagtc acccacccgg actcacattc
tccccagagg
       61 ccgagatgcg ggtcatggcg ccccgagccc tcctcctgct gctctcggga
ggcctggccc
      121 tgaccgagac ctgggcctgc tcccactcca tgaggtattt cgacaccgcc
gtgtcccggc
      181 ccggccgcgg agagccccgc ttcatctcag tgggctacgt ggacgacacg
cagttcgtgc
      241 ggttcgacag cgacgccgcg agtccgagag gggagccgcg ggcgccgtgg
gtggagcagg
      301 aggggccgga gtattgggac cgggagacac agaagtacaa gcgccaggca
caggctgacc
```

```
      361 gagtgagcct gcggaacctg cgcggctact acaaccagag cgaggacggg
tctcacaccc
      421 tccagaggat gtctggctgc gacctggggc ccgacgggcg cctcctccgc
gggtatgacc
      481 agtccgccta cgacggcaag gattacatcg ccctgaacga ggacctgcgc
tcctggaccg
      541 ccgcggacac cgcggctcag atcacccagc gcaagttgga ggcggcccgt
gcggcggagc
      601 agctgagagc ctacctggag ggcacgtgcg tggagtggct ccgcagatac
ctggagaacg
      661 ggaaggagac gctgcagcgc gcagaacccc caaagacaca cgtgacccac
cacccctct
      721 ctgaccatga ggccaccctg aggtgctggg ccctgggctt ctaccctgcg
gagatcacac
      781 tgacctggca gcgggatggg gaggaccaga cccaggacac cgagcttgtg
gagaccaggc
      841 cagcaggaga tggaaccttc cagaagtggg cagctgtggt ggtgccttct
ggacaagagc
      901 agagatacac gtgccatatg cagcacgagg ggctgcaaga gcccctcacc
ctgagctggg
      961 agccatcttc ccagcccacc atccccatca tgggcatcgt tgctggcctg
gctgtcctgg
     1021 ttgtcctagc tgtccttgga gctgtggtca ccgctatgat gtgtaggagg
aagagctcag
     1081 gtggaaaagg agggagctgc tctcaggctg cgtgcagcaa cagtgcccag
ggctctgatg
     1141 agtctctcat cacttgtaaa gcctgagaca gctgcctgtg tgggactgag
atgcaggatt
     1201 tcttcacacc tctcctttgt gacttcaaga gcctctggca tctctttctg
caaaggcacc
     1261 tgaatgtgtc tgcgttcctg ttagcataat gtgaggaggt ggagagacag
cccacccccg
     1321 tgtccaccgt gacccctgtc cccacactga cctgtgttcc ctccccgatc
atctttcctg
     1381 ttccagagag gtggggctgg atgtctccat ctctgtctca aattcatggt
gcactgagct
     1441 gcaacttctt acttccctaa tgaagttaag aacctgaata taaatttgtg
ttctcaaata
     1501 tttgctatga agcgttgatg gattaattaa ataagtcaat tcctagaagt
tgagagagca
      561 aataaagacc tgagaacctt ccagaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_002108.4

LOCUS NP_002108

ACCESSION NP_002108

```
      1 mrvmaprall lllsgglalt etwacshsmr yfdtavsrpg rgeprfisvg
yvddtqfvrf
       61 dsdaasprge prapwveqeg peywdretqk ykrqaqadrv slrnlrgyyn
qsedgshtlq
      121 rmsgcdlgpd grllrgydqs aydgkdyial nedlrswtaa dtaaqitqrk
leaaraaeql
      181 raylegtcve wlrrylengk etlqraeppk thvthhplsd heatlrcwal
gfypaeitlt
      241 wqrdgedqtq dtelvetrpa gdgtfqkwaa vvvpsgqeqr ytchmqhegl
qepltlswep
```

```
      301 ssqptipimg ivaglavlvv lavlgavvta mmcrrkssgg kggscsqaac
snsaqgsdes
      361 litcka
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001243042.1

LOCUS NM_001243042

ACCESSION NM_001243042

```
      1 tccgcagtcc cggttctaaa gtccccagtc acccacccgg actcacattc
tccccagagg
       61 ccgagatgcg ggtcatggcg ccccgagccc tcctcctgct gctctcggga
ggcctggccc
      121 tgaccgagac ctgggcctgc tcccactcca tgaggtattt cgacaccgcc
gtgtcccggc
      181 ccggccgcgg agagccccgc ttcatctcag tgggctacgt ggacgacacg
cagttcgtgc
      241 ggttcgacag cgacgccgcg agtccgagag gggagccgcg ggcgccgtgg
gtggagcagg
      301 aggggccgga gtattgggac cgggagacac agaactacaa gcgccaggca
caggctgacc
      361 gagtgagcct gcggaacctg cgcggctact acaaccagag cgaggacggg
tctcacaccc
      421 tccagaggat gtatggctgc gacctggggc ccgacgggcg cctcctccgc
gggtatgacc
      481 agtccgccta cgacggcaag gattacatcg ccctgaacga ggacctgcgc
tcctggaccg
      541 ccgcggacac cgcggctcag atcacccagc gcaagttgga ggcggcccgt
gcggcggagc
      601 agctgagagc ctacctggag ggcacgtgcg tggagtggct ccgcagatac
ctggagaacg
      661 ggaaggagac gctgcagcgc gcagaacccc caaagacaca cgtgacccac
cacccctct
      721 ctgaccatga ggccaccctg aggtgctggg ccctgggctt ctaccctgcg
gagatcacac
      781 tgacctggca gcgggatggg gaggaccaga cccaggacac cgagcttgtg
gagaccaggc
      841 cagcaggaga tggaaccttc cagaagtggg cagctgtggt ggtgccttct
ggacaagagc
      901 agagatacac gtgccatatg cagcacgagg ggctgcaaga gcccctcacc
ctgagctggg
      961 agccatcttc ccagcccacc atccccatca tgggcatcgt tgctggcctg
gctgtcctgg
     1021 ttgtcctagc tgtccttgga gctgtggtca ccgctatgat gtgtaggagg
aagagctcag
     1081 gtggaaaagg agggagctgc tctcaggctg cgtgcagcaa cagtgcccag
ggctctgatg
     1141 agtctctcat cacttgtaaa gcctgagaca gctgcctgtg tgggactgag
atgcaggatt
     1201 tcttcacacc tctcctttgt gacttcaaga gcctctggca tctctttctg
caaaggcgtc
     1261 tgaatgtgtc tgcgttcctg ttagcataat gtgaggaggt ggagagacag
cccacccccg
     1321 tgtccaccgt gacccctgtc cccacactga cctgtgttcc ctccccgatc
atctttcctg
     1381 ttccagagag gtggggctgg atgtctccat ctctgtctca aattcatggt
gcactgagct
```

```
     1441 gcaacttctt acttccctaa tgaagttaag aacctgaata taaatttgtg
ttctcaaata
     1501 tttgctatga agcgttgatg gattaattaa ataagtcaat tcctagaagt
tgagagagca
      1561 aataaagacc tgagaacctt ccagaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001229971.1

LOCUS NP_001229971

ACCESSION NP_001229971

```
      1 mrvmaprall lllsgglalt etwacshsmr yfdtavsrpg rgeprfisvg
yvddtqfvrf
       61 dsdaasprge prapwveqeg peywdretqn ykrqaqadrv slrnlrgyyn
qsedgshtlq
      121 rmygcdlgpd grllrgydqs aydgkdyial nedlrswtaa dtaaqitqrk
leaaraaeql
      181 raylegtcve wlrrylengk etlqraeppk thvthhplsd heatlrcwal
gfypaeitlt
      241 wqrdgedqtq dtelvetrpa gdgtfqkwaa vvvpsgqeqr ytchmqhegl
qepltlswep
      301 ssqptipimg ivaglavlvv lavlgavvta mmcrrkssgg kggscsqaac
snsaqgsdes
      361 litcka
```

RRAS2

Official Symbol: RRAS2

Official Name: related RAS viral (r-ras) oncogene homolog 2

Gene ID: 22800

Organism: Homo sapiens

Other Aliases: TC21

Other Designations: ras-like protein TC21; ras-related protein R-Ras2; teratocarcinoma oncogene

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_012250.5

LOCUS NM_012250

ACCESSION NM_012250

```
      1 cagacggcca tttgtggcgg cgctggaggc tgcgttcggc aggcgctgcg
gagacgcgta
       61 gaggagcgcg ccccccggcc gctgccgccc ctggcccgtg ccgtcacccc
gcttctccgc
```

121 gcctcgggcg gtacccagcc agtccccagc gccgcgctac cgcgctgacc
ggccctccag
181 acgcctcccg gtacccggga ccccagcccg gccgctcgcc cgcagcccgc
cggccgcaca
241 cgtccccgga gccgggccta gggcgggcgg cagcggcggc tcggcgcagt
caggctgggc
301 tctgtagcgt ccccatggcc gcggccggct ggcgggacgg ctccggccag
gagaagtacc
361 ggctcgtggt ggtcggcggg ggcggcgtgg gcaagtcggc gctcaccatc
cagttcatcc
421 agtcctattt tgtaacggat tatgatccaa ccattgaaga ttcttacaca
aagcagtgtg
481 tgatagatga cagagcagcc cggctagata ttttggatac agcaggacaa
gaagagtttg
541 gagccatgag agaacagtat atgaggactg gcgaaggctt cctgttggtc
ttttcagtca
601 cagatagagg cagttttgaa gaaatctata agtttcaaag acagattctc
agagtaaagg
661 atcgtgatga gttcccaatg attttaattg gtaataaagc agatctggat
catcaaagac
721 aggtaacaca ggaagaagga caacagttag cacggcagct taaggtaaca
tacatggagg
781 catcagcaaa gattaggatg aatgtagatc aagctttcca tgaacttgtc
cgggttatca
841 ggaaatttca agagcaggaa tgtcctcctt caccagaacc aacacggaaa
gaaaaagaca
901 agaaaggctg ccattgtgtc attttctaga atcccttcag ttttagctac
caacggccag
961 gaaaagccct catcttctct ttctctcctc agtttacatc ttgttggtac
ctttctagcc
1021 ttagacaaat gatcaccatg ttagccttag acgaagaagc tggctagtcc
tttctgtgaa
1081 gctaatacaa tggtcatttc cagacaaatt taaaggaaac actaaggctg
cttcaaagat
1141 tatctgattc ctttaaaata tatgtctata tacacagaca tgctcttttt
ttaagtgctt
1201 acattttaat agagatgaat cagttttgga atctaagctg tttgccaagc
tgaagctaca
1261 ggttgtgaaa taatttttaa cttttggaat catactgcct actgttactc
taaatagaaa
1321 tatagggttt tttttaatgt gaatttttgc ctatctttaa acatttcaat
gtcagccttt
1381 gttaacctta aatacactga attgaatcta caaaagtgaa ccatctcaga
cctttactga
1441 tactacaact tttgttttct gatggccaaa ataccaaatg cctgttgtat
ttatggatta
1501 aaaactgctt ataaaaccct gtgttactac tcctactctt ggagatgata
atattctatg
1561 tggtcaaata tttggactca tttaggactt agatatttca gtgtacttga
ttttttaatt
1621 taactctttt tcacagccac gctaagggta aaaaggaata atttccttct
gtcttccttt
1681 tcaagtattt ctgggtaagg gattcaaaaa actaaaactg tttttgtttg
taatataaaa
1741 tatggaattg atctttccag ggtcagagat gattaatgtt tttgctatat
acttttatac
1801 attattttct tatcaaacta gttaacaagt atttttatat gtttgtaagc
agatatgctt
1861 tcatagcata ccttgtgtat atgtaaagat aagtatttaa ttctcactgt
tcacttttaa

```
     1921 ctgacaaaga aaaacaagtg gaaactacag aaactgtggt agaactttta
cttgctggtc
     1981 tggtcttggt tgtacccatc tttggccagt cacataacta ctcaagaaac
cttcccaata
     2041 gagtacaaca ggatgagact ctgaaatcac tttcagtatt ccctgctaga
tattgattgt
     2101 tatttcaagt attaagtgta agcttttaat ggataattag tataactgtg
gatggcatct
     2161 gattttgttt ttaattctgt ggattgtgtt taagcaattc aatagtatgt
tcctgatttt
     2221 gagatgctaa gtggtattgc acagttgtca ctttatcaag tgtgtacaac
agtcccatga
     2281 agtttataga gcataccctt gtatagcttc aggtgctaga attaaaattg
atctgttatc
     2341 acaagaaaaa aaaaaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_036382.2

LOCUS NP_036382

ACCESSION NP_036382

```
      1 maaagwrdgs gqekyrlvvv ggggvgksal tiqfiqsyfv tdydptieds
ytkqcviddr
       61 aarldildta gqeefgamre qymrtgegfl lvfsvtdrgs feeiykfqrq
ilrvkdrdef
      121 pmilignkad ldhqrqvtqe egqqlarqlk vtymeasaki rmnvdqafhe
lvrvirkfqe
      181 qecppspept rkekdkkgch cvif
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001102669.2

LOCUS NM_001102669

ACCESSION NM_001102669

```
      1 gggagggcgg agctggaagg gtgggaagca ccgatccacc ttattgctct
ggccgaggcc
       61 agagacctcc gggagaggct gggccaccga gccgggcttt actgctccga
gggtccgggc
      121 gtggggctgg agctggagcc ccgcgcgctg cttttccagc cgcctgcggc
cgcgccttca
      181 ccgtcggggc gatagcggtg gcaacttggc cgcggctccg cgtggtctcc
gggcttcccc
      241 gcgccgcctg agccggagct gcccgcttca atcctatttt gtaacggatt
atgatccaac
      301 cattgaagat tcttacacaa agcagtgtgt gatagatgac agagcagccc
ggctagatat
      361 tttggataca gcaggacaag aagagtttgg agccatgaga gaacagtata
tgaggactgg
      421 cgaaggcttc ctgttggtct tttcagtcac agatagaggc agttttgaag
aaatctataa
```

```
      481 gtttcaaaga cagattctca gagtaaagga tcgtgatgag ttcccaatga
ttttaattgg
      541 taataaagca gatctggatc atcaaagaca ggtaacacag gaagaaggac
aacagttagc
      601 acggcagctt aaggtaacat acatggaggc atcagcaaag attaggatga
atgtagatca
      661 agctttccat gaacttgtcc gggttatcag gaaatttcaa gagcaggaat
gtcctccttc
      721 accagaacca acacggaaag aaaaagacaa gaaaggctgc cattgtgtca
ttttctagaa
      781 tcccttcagt tttagctacc aacggccagg aaaagccctc atcttctctt
tctctcctca
      841 gtttacatct tgttggtacc tttctagcct tagacaaatg atcaccatgt
tagccttaga
      901 cgaagaagct ggctagtcct ttctgtgaag ctaatacaat ggtcatttcc
agacaaattt
      961 aaaggaaaca ctaaggctgc ttcaaagatt atctgattcc tttaaaatat
atgtctatat
     1021 acacagacat gctctttttt taagtgctta cattttaata gagatgaatc
agttttggaa
     1081 tctaagctgt ttgccaagct gaagctacag gttgtgaaat aatttttaac
ttttggaatc
     1141 atactgccta ctgttactct aaatagaaat atagggtttt ttttaatgtg
aatttttgcc
     1201 tatctttaaa catttcaatg tcagcctttg ttaaccttaa atacactgaa
ttgaatctac
     1261 aaaagtgaac catctcagac ctttactgat actacaactt ttgttttctg
atggccaaaa
     1321 taccaaatgc ctgttgtatt tatggattaa aaactgctta taaaaccctg
tgttactact
     1381 cctactcttg gagatgataa tattctatgt ggtcaaatat ttggactcat
ttaggactta
     1441 gatatttcag tgtacttgat tttttaattt aactcttttt cacagccacg
ctaagggtaa
     1501 aaaggaataa tttccttctg tcttcctttt caagtatttc tgggtaaggg
attcaaaaaa
     1561 ctaaaactgt ttttgtttgt aatataaaat atggaattga tctttccagg
gtcagagatg
     1621 attaatgttt ttgctatata cttttataca ttattttctt atcaaactag
ttaacaagta
     1681 tttttatatg tttgtaagca gatatgcttt catagcatac cttgtgtata
tgtaaagata
     1741 agtatttaat tctcactgtt cacttttaac tgacaaagaa aaacaagtgg
aaactacaga
     1801 aactgtggta gaacttttac ttgctggtct ggtcttggtt gtacccatct
ttggccagtc
     1861 acataactac tcaagaaacc ttcccaatag agtacaacag gatgagactc
tgaaatcact
     1921 ttcagtattc cctgctagat attgattgtt atttcaagta ttaagtgtaa
gcttttaatg
     1981 gataattagt ataactgtgg atggcatctg attttgtttt taattctgtg
gattgtgttt
     2041 aagcaattca atagtatgtt cctgattttg agatgctaag tggtattgca
cagttgtcac
     2101 tttatcaagt gtgtacaaca gtcccatgaa gtttatagag catacccttg
tatagcttca
     2161 ggtgctagaa ttaaaattga tctgttatca caagaaaaaa aaaaaaaaa
```

Protein sequence (varaiant 2):

NCBI Reference Sequence: NP_001096139.1

LOCUS NP_001096139

ACCESSION NP_001096139

```
      1 mreqymrtge gfllvfsvtd rgsfeeiykf qrqilrvkdr defpmilign
kadldhqrqv
       61 tqeegqqlar qlkvtymeas akirmnvdqa fhelvrvirk fqeqecppsp
eptrkekdkk
      121 gchcvif
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_001177314.1

LOCUS NM_001177314

ACCESSION NM_001177314

```
      1 atctcagatg catgcagctc ctgctgggcg gtttcattct ctgccagcca
ttcattcaca
       61 ttaaaagcaa tggcccatta aggaaataag tggatgatgc tcctccatca
cccatgtcct
      121 attttgtaac ggattatgat ccaaccattg aagattctta cacaaagcag
tgtgtgatag
      181 atgacagagc agcccggcta gatattttgg atacagcagg acaagaagag
tttggagcca
      241 tgagagaaca gtatatgagg actggcgaag gcttcctgtt ggtcttttca
gtcacagata
      301 gaggcagttt tgaagaaatc tataagtttc aaagacagat tctcagagta
aaggatcgtg
      361 atgagttccc aatgatttta attggtaata aagcagatct ggatcatcaa
agacaggtaa
      421 cacaggaaga aggacaacag ttagcacggc agcttaaggt aacatacatg
gaggcatcag
      481 caaagattag gatgaatgta gatcaagctt tccatgaact tgtccgggtt
atcaggaaat
      541 ttcaagagca ggaatgtcct ccttcaccag aaccaacacg gaaagaaaaa
gacaagaaag
      601 gctgccattg tgtcattttc tagaatccct tcagttttag ctaccaacgg
ccaggaaaag
      661 ccctcatctt ctctttctct cctcagttta catcttgttg gtacctttct
agccttagac
      721 aaatgatcac catgttagcc ttagacgaag aagctggcta gtcctttctg
tgaagctaat
      781 acaatggtca tttccagaca aatttaaagg aaacactaag gctgcttcaa
agattatctg
      841 attcctttaa aatatatgtc tatatacaca gacatgctct ttttttaagt
gcttacattt
      901 taatagagat gaatcagttt tggaatctaa gctgtttgcc aagctgaagc
tacaggttgt
      961 gaaataattt ttaacttttg gaatcatact gcctactgtt actctaaata
gaaatatagg
     1021 gtttttttta atgtgaattt ttgcctatct ttaaacattt caatgtcagc
ctttgttaac
     1081 cttaaataca ctgaattgaa tctacaaaag tgaaccatct cagaccttta
ctgatactac
     1141 aacttttgtt ttctgatggc caaaatacca aatgcctgtt gtatttatgg
attaaaaact
```

```
     1201 gcttataaaa ccctgtgtta ctactcctac tcttggagat gataatattc
tatgtggtca
     1261 aatatttgga ctcatttagg acttagatat ttcagtgtac ttgatttttt
aatttaactc
     1321 tttttcacag ccacgctaag ggtaaaaagg aataatttcc ttctgtcttc
cttttcaagt
     1381 atttctgggt aagggattca aaaaactaaa actgtttttg tttgtaatat
aaaatatgga
     1441 attgatcttt ccagggtcag agatgattaa tgtttttgct atatactttt
atacattatt
     1501 ttcttatcaa actagttaac aagtattttt atatgtttgt aagcagatat
gctttcatag
     1561 cataccttgt gtatatgtaa agataagtat ttaattctca ctgttcactt
ttaactgaca
     1621 aagaaaaaca agtggaaact acagaaactg tggtagaact tttacttgct
ggtctggtct
     1681 tggttgtacc catctttggc cagtcacata actactcaag aaaccttccc
aatagagtac
     1741 aacaggatga gactctgaaa tcactttcag tattccctgc tagatattga
ttgttatttc
     1801 aagtattaag tgtaagcttt taatggataa ttagtataac tgtggatggc
atctgatttt
     1861 gtttttaatt ctgtggattg tgtttaagca attcaatagt atgttcctga
ttttgagatg
     1921 ctaagtggta ttgcacagtt gtcactttat caagtgtgta caacagtccc
atgaagttta
     1981 tagagcatac ccttgtatag cttcaggtgc tagaattaaa attgatctgt
tatcacaaga
     2041 aaaaaaaaaa aaaa
      //
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_001170785.1

LOCUS NP_001170785

ACCESSION NP_001170785

```
     1 msyfvtdydp tiedsytkqc viddraarld ildtagqeef gamreqymrt gegfllvfsv
    61 tdrgsfeeiy kfqrqilrvk drdefpmili gnkadldhqr qvtqeegqql arqlkvtyme
   121 asakirmnvd qafhelvrvi rkfqeqecpp speptrkekd kkgchcvif
```

Nucleotide sequence (variant 4):

NCBI Reference Sequence : NM_001177315.1

LOCUS NM_001177315

ACCESSION NM_001177315

```
     1 attgctctgg ccgaggccag agacctccgg gagaggctgg gccaccgagc
cgggctttac
      61 tgctccgagg gtccgggcgt ggggctggag ctggagcccc gcgcgctgct
tttccagccg
     121 cctgcggccg cgccttcacc gtcggggcga tagcggtggc aacttggccg
cggctccgcg
     181 tggtctccgg gcttccccgc gccgcctgag ccggagctgc ccgcttcaag
tactgtgtat
```

241 ttctttgttc ctattttgta acggattatg atccaaccat tgaagattct
tacacaaagc
301 agtgtgtgat agatgacaga gcagcccggc tagatatttt ggatacagca
ggacaagaag
361 agtttggagc catgagagaa cagtatatga ggactggcga aggcttcctg
ttggtctttt
421 cagtcacaga tagaggcagt tttgaagaaa tctataagtt tcaaagacag
attctcagag
481 taaaggatcg tgatgagttc ccaatgattt taattggtaa taaagcagat
ctggatcatc
541 aaagacaggt aacacaggaa gaaggacaac agttagcacg gcagcttaag
gtaacataca
601 tggaggcatc agcaaagatt aggatgaatg tagatcaagc tttccatgaa
cttgtccggg
661 ttatcaggaa atttcaagag caggaatgtc ctccttcacc agaaccaaca
cggaaagaaa
721 aagacaagaa aggctgccat tgtgtcattt tctagaatcc cttcagtttt
agctaccaac
781 ggccaggaaa agccctcatc ttctctttct ctcctcagtt tacatcttgt
tggtaccttt
841 ctagccttag acaaatgatc accatgttag ccttagacga agaagctggc
tagtcctttc
901 tgtgaagcta atacaatggt catttccaga caaatttaaa ggaaacacta
aggctgcttc
961 aaagattatc tgattccttt aaaatatatg tctatataca cagacatgct
ctttttttaa
1021 gtgcttacat tttaatagag atgaatcagt tttggaatct aagctgtttg
ccaagctgaa
1081 gctacaggtt gtgaaataat ttttaacttt tggaatcata ctgcctactg
ttactctaaa
1141 tagaaatata gggtttttt taatgtgaat ttttgcctat ctttaaacat
ttcaatgtca
1201 gcctttgtta accttaaata cactgaattg aatctacaaa agtgaaccat
ctcagacctt
1261 tactgatact acaacttttg ttttctgatg gccaaaatac caaatgcctg
ttgtatttat
1321 ggattaaaaa ctgcttataa aaccctgtgt tactactcct actcttggag
atgataatat
1381 tctatgtggt caaatatttg gactcattta ggacttagat atttcagtgt
acttgatttt
1441 ttaatttaac tctttttcac agccacgcta agggtaaaaa ggaataattt
ccttctgtct
1501 tccttttcaa gtatttctgg gtaagggatt caaaaaacta aaactgtttt
tgtttgtaat
1561 ataaaatatg gaattgatct ttccagggtc agagatgatt aatgtttttg
ctatatactt
1621 ttatacatta ttttcttatc aaactagtta acaagtattt ttatatgttt
gtaagcagat
1681 atgctttcat agcatacctt gtgtatatgt aaagataagt atttaattct
cactgttcac
1741 ttttaactga caaagaaaaa caagtggaaa ctacagaaac tgtggtagaa
cttttacttg
1801 ctggtctggt cttggttgta cccatctttg gccagtcaca taactactca
agaaaccttc
1861 ccaatagagt acaacaggat gagactctga aatcactttc agtattccct
gctagatatt
1921 gattgttatt tcaagtatta agtgtaagct tttaatggat aattagtata
actgtggatg
1981 gcatctgatt ttgtttttaa ttctgtggat tgtgtttaag caattcaata
gtatgttcct

```
     2041 gattttgaga tgctaagtgg tattgcacag ttgtcacttt atcaagtgtg
tacaacagtc
     2101 ccatgaagtt tatagagcat acccttgtat agcttcaggt gctagaatta
aaattgatct
     2161 gttatcacaa gaaaaaaaaa aaaaaa
```

Protein sequence (variant 4):

NCBI Reference Sequence: NP_001170786.1

LOCUS NP_001170786

ACCESSION NP_001170786

```
      1 mreqymrtge gfllvfsvtd rgsfeeiykf qrqilrvkdr defpmilign
kadldhqrqv
       61 tqeegqqlar qlkvtymeas akirmnvdqa fhelvrvirk fqeqecppsp
eptrkekdkk
      121 gchcvif
```

TSP1

Official Symbol: THBS1

Official Name: thrombospondin 1

Gene ID: 7057

Organism: Homo sapiens

Other Aliases: THBS, THBS-1, TSP, TSP-1, **TSP1**

Other Designations: thrombospondin-1; thrombospondin-1p180

Nucleotide sequence:

NCBI Reference Sequence: NM_003246.2

LOCUS NM_003246

ACCESSION NM_003246

```
      1 agccgctgcg cccgagctgg cctgcgagtt cagggctcct gtcgctctcc
aggagcaacc
       61 tctactccgg acgcacaggc attccccgcg cccctccagc cctcgccgcc
ctcgccaccg
      121 ctcccggccg ccgcgctccg gtacacacag gatccctgct gggcaccaac
agctccacca
      181 tggggctggc ctggggacta ggcgtcctgt tcctgatgca tgtgtgtggc
accaaccgca
      241 ttccagagtc tggcggagac aacagcgtgt ttgacatctt tgaactcacc
ggggccgccc
      301 gcaaggggtc tgggcgccga ctggtgaagg gccccgaccc ttccagccca
gctttccgca
```

361 tcgaggatgc caacctgatc ccccctgtgc ctgatgacaa gttccaagac
ctggtggatg
421 ctgtgcgggc agaaaagggt ttcctccttc tggcatccct gaggcagatg
aagaagaccc
481 ggggcacgct gctggccctg gagcggaaag accactctgg ccaggtcttc
agcgtggtgt
541 ccaatggcaa ggcgggcacc ctggacctca gcctgaccgt ccaaggaaag
cagcacgtgg
601 tgtctgtgga agaagctctc ctggcaaccg gccagtggaa gagcatcacc
ctgtttgtgc
661 aggaagacag ggcccagctg tacatcgact gtgaaaagat ggagaatgct
gagttggacg
721 tccccatcca aagcgtcttc accagagacc tggccagcat cgccagactc
cgcatcgcaa
781 agggggggcgt caatgacaat ttccaggggg tgctgcagaa tgtgaggttt
gtctttggaa
841 ccacaccaga agacatcctc aggaacaaag gctgctccag ctctaccagt
gtcctcctca
901 cccttgacaa caacgtggtg aatggttcca gccctgccat ccgcactaac
tacattggcc
961 acaagacaaa ggacttgcaa gccatctgcg gcatctcctg tgatgagctg
tccagcatgg
1021 tcctggaact caggggcctg cgcaccattg tgaccacgct gcaggacagc
atccgcaaag
1081 tgactgaaga gaacaaagag ttggccaatg agctgaggcg gcctccccta
tgctatcaca
1141 acggagttca gtacagaaat aacgaggaat ggactgttga tagctgcact
gagtgtcact
1201 gtcagaactc agttaccatc tgcaaaaagg tgtcctgccc catcatgccc
tgctccaatg
1261 ccacagttcc tgatggagaa tgctgtcctc gctgttggcc cagcgactct
gcggacgatg
1321 gctggtctcc atggtccgag tggacctcct gttctacgag ctgtggcaat
ggaattcagc
1381 agcgcggccg ctcctgcgat agcctcaaca accgatgtga gggctcctcg
gtccagacac
1441 ggacctgcca cattcaggag tgtgacaaga gatttaaaca ggatggtggc
tggagccact
1501 ggtccccgtg gtcatcttgt tctgtgacat gtggtgatgg tgtgatcaca
aggatccggc
1561 tctgcaactc tcccagcccc cagatgaacg ggaaaccctg tgaaggcgaa
gcgcgggaga
1621 ccaaagcctg caagaaagac gcctgcccca tcaatggagg ctggggtcct
tggtcaccat
1681 gggacatctg ttctgtcacc tgtggaggag gggtacagaa acgtagtcgt
ctctgcaaca
1741 accccacacc ccagtttgga ggcaaggact gcgttggtga tgtaacagaa
aaccagatct
1801 gcaacaagca ggactgtcca attgatggat gcctgtccaa tccctgcttt
gccggcgtga
1861 agtgtactag ctaccctgat ggcagctgga aatgtggtgc ttgtcccct
ggttacagtg
1921 gaaatggcat ccagtgcaca gatgttgatg agtgcaaaga agtgcctgat
gcctgcttca
1981 accacaatgg agagcaccgg tgtgagaaca cggaccccgg ctacaactgc
ctgccctgcc
2041 ccccacgctt caccggctca cagcccttcg gccagggtgt cgaacatgcc
acggccaaca
2101 aacaggtgtg caagccccgt aacccctgca cggatgggac ccacgactgc
aacaagaacg

2161 ccaagtgcaa ctacctgggc cactatagcg accccatgta ccgctgcgag
tgcaagcctg
2221 gctacgctgg caatggcatc atctgcgggg aggacacaga cctggatggc
tggcccaatg
2281 agaacctggt gtgcgtggcc aatgcgactt accactgcaa aaaggataat
tgccccaacc
2341 ttcccaactc agggcaggaa gactatgaca aggatggaat tggtgatgcc
tgtgatgatg
2401 acgatgacaa tgataaaatt ccagatgaca gggacaactg tccattccat
tacaacccag
2461 ctcagtatga ctatgacaga gatgatgtgg gagaccgctg tgacaactgt
ccctacaacc
2521 acaacccaga tcaggcagac acagacaaca atggggaagg agacgcctgt
gctgcagaca
2581 ttgatggaga cggtatcctc aatgaacggg acaactgcca gtacgtctac
aatgtggacc
2641 agagagacac tgatatggat ggggttggag atcagtgtga caattgcccc
ttggaacaca
2701 atccggatca gctggactct gactcagacc gcattggaga tacctgtgac
aacaatcagg
2761 atattgatga agatggccac cagaacaatc tggacaactg tccctatgtg
cccaatgcca
2821 accaggctga ccatgacaaa gatggcaagg gagatgcctg tgaccacgat
gatgacaacg
2881 atggcattcc tgatgacaag gacaactgca gactcgtgcc caatcccgac
cagaaggact
2941 ctgacggcga tggtcgaggt gatgcctgca aagatgattt tgaccatgac
agtgtgccag
3001 acatcgatga catctgtcct gagaatgttg acatcagtga gaccgatttc
cgccgattcc
3061 agatgattcc tctggacccc aaagggacat cccaaaatga ccctaactgg
gttgtacgcc
3121 atcagggtaa agaactcgtc cagactgtca actgtgatcc tggactcgct
gtaggttatg
3181 atgagtttaa tgctgtggac ttcagtggca ccttcttcat caacaccgaa
agggacgatg
3241 actatgctgg atttgtcttt ggctaccagt ccagcagccg cttttatgtt
gtgatgtgga
3301 agcaagtcac ccagtcctac tgggacacca accccacgag ggctcaggga
tactcgggcc
3361 tttctgtgaa agttgtaaac tccaccacag ggcctggcga gcacctgcgg
aacgccctgt
3421 ggcacacagg aaacacccct ggccaggtgc gcaccctgtg gcatgaccct
cgtcacatag
3481 gctggaaaga tttcaccgcc tacagatggc gtctcagcca caggccaaag
acgggtttca
3541 ttagagtggt gatgtatgaa gggaagaaaa tcatggctga ctcaggaccc
atctatgata
3601 aaacctatgc tggtggtaga ctagggttgt ttgtcttctc tcaagaaatg
gtgttcttct
3661 ctgacctgaa atacgaatgt agagatccct aatcatcaaa ttgttgattg
aaagactgat
3721 cataaaccaa tgctggtatt gcaccttctg gaactatggg cttgagaaaa
cccccaggat
3781 cacttctcct tggcttcctt cttttctgtg cttgcatcag tgtggactcc
tagaacgtgc
3841 gacctgcctc aagaaaatgc agttttcaaa aacagactca gcattcagcc
tccaatgaat
3901 aagacatctt ccaagcatat aaacaattgc tttggtttcc ttttgaaaaa
gcatctactt

3961 gcttcagttg ggaaggtgcc cattccactc tgcctttgtc acagagcagg
gtgctattgt
4021 gaggccatct ctgagcagtg gactcaaaag cattttcagg catgtcagag
aagggaggac
4081 tcactagaat tagcaaacaa aaccaccctg acatcctcct tcaggaacac
ggggagcaga
4141 ggccaaagca ctaaggggag ggcgcatacc cgagacgatt gtatgaagaa
aatatggagg
4201 aactgttaca tgttcggtac taagtcattt tcaggggatt gaaagactat
tgctggattt
4261 catgatgctg actggcgtta gctgattaac ccatgtaaat aggcacttaa
atagaagcag
4321 gaaagggaga caaagactgg cttctggact tcctccctga tccccaccct
tactcatcac
4381 ctgcagtggc cagaattagg gaatcagaat caaaccagtg taaggcagtg
ctggctgcca
4441 ttgcctggtc acattgaaat tggtggcttc attctagatg tagcttgtgc
agatgtagca
4501 ggaaaatagg aaaacctacc atctcagtga gcaccagctg cctcccaaag
gaggggcagc
4561 cgtgcttata tttttatggt tacaatggca caaaattatt atcaacctaa
ctaaaacatt
4621 ccttttctct tttttcctga attatcatgg agttttctaa ttctctcttt
tggaatgtag
4681 attttttta aatgctttac gatgtaaaat atttatttt tacttattct
ggaagatctg
4741 gctgaaggat tattcatgga acaggaagaa gcgtaaagac tatccatgtc
atctttgttg
4801 agagtcttcg tgactgtaag attgtaaata cagattattt attaactctg
ttctgcctgg
4861 aaatttaggc ttcatacgga aagtgtttga gagcaagtag ttgacattta
tcagcaaatc
4921 tcttgcaaga acagcacaag gaaaatcagt ctaataagct gctctgcccc
ttgtgctcag
4981 agtggatgtt atgggattct ttttttctct gttttatctt ttcaagtgga
attagttggt
5041 tatccatttg caaatgtttt aaattgcaaa gaaagccatg aggtcttcaa
tactgtttta
5101 ccccatccct tgtgcatatt tccagggaga aggaaagcat atacactttt
ttctttcatt
5161 tttccaaaag agaaaaaaat gacaaaaggt gaaacttaca tacaaatatt
acctcatttg
5221 ttgtgtgact gagtaaagaa tttttggatc aagcggaaag agtttaagtg
tctaacaaac
5281 ttaaagctac tgtagtacct aaaaagtcag tgttgtacat agcataaaaa
ctctgcagag
5341 aagtattccc aataaggaaa tagcattgaa atgttaaata caatttctga
aagttatgtt
5401 ttttttctat catctggtat accattgctt tatttttata aattattttc
tcattgccat
5461 tggaatagat atctcagatt gtgtagatat gctatttaaa taatttatca
ggaaatactg
5521 cctgtagagt tagtatttct atttttatat aatgtttgca cactgaattg
aagaattgtt
5581 ggtttttttct ttttttttgtt ttgtttttttt ttttttttttt ttttgctttt
gacctcccat
5641 ttttactatt tgccaatacc tttttctagg aatgtgcttt tttttgtaca
catttttatc
5701 cattttacat tctaaagcag tgtaagttgt atattactgt ttcttatgta
caaggaacaa

```
      5761 caataaatca tatggaaatt tatatttata aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_003237.2

LOCUS NP_003237

ACCESSION NP_003237

```
      1 mglawglgvl flmhvcgtnr ipesggdnsv fdifeltgaa rkgsgrrlvk
gpdpsspafr
       61 iedanlippv pddkfqdlvd avraekgfll laslrqmkkt rgtllalerk
dhsgqvfsvv
      121 sngkagtldl sltvqgkqhv vsveeallat gqwksitlfv qedraqlyid
cekmenaeld
      181 vpiqsvftrd lasiarlria kggvndnfqg vlqnvrfvfg ttpedilrnk
gcssstsvll
      241 tldnnvvngs spairtnyig hktkdlqaic giscdelssm vlelrglrti
vttlqdsirk
      301 vteenkelan elrrpplcyh ngvqyrnnee wtvdsctech cqnsvtickk
vscpimpcsn
      361 atvpdgeccp rcwpsdsadd gwspwsewts cstscgngiq qrgrscdsln
nrcegssvqt
      421 rtchiqecdk rfkqdggwsh wspwsscsvt cgdgvitrir lcnspspqmn
gkpcegeare
      481 tkackkdacp inggwgpwsp wdicsvtcgg gvqkrsrlcn nptpqfggkd
cvgdvtenqi
      541 cnkqdcpidg clsnpcfagv kctsypdgsw kcgacppgys gngiqctdvd
eckevpdacf
      601 nhngehrcen tdpgynclpc pprftgsqpf gqgvehatan kqvckprnpc
tdgthdcnkn
      661 akcnylghys dpmyrceckp gyagngiicg edtdldgwpn enlvcvanat
yhckkdncpn
      721 lpnsgqedyd kdgigdacdd dddndkipdd rdncpfhynp aqydydrddv
gdrcdncpyn
      781 hnpdqadtdn ngegdacaad idgdgilner dncqyvynvd qrdtdmdgvg
dqcdncpleh
      841 npdqldsdsd rigdtcdnnq didedghqnn ldncpyvpna nqadhdkdgk
gdacdhdddn
      901 dgipddkdnc rlvpnpdqkd sdgdgrgdac kddfdhdsvp diddicpenv
disetdfrrf
      961 qmipldpkgt sqndpnwvvr hqgkelvqtv ncdpglavgy defnavdfsg
tffinterdd
     1021 dyagfvfgyq sssrfyvvmw kqvtqsywdt nptraqgysg lsvkvvnstt
gpgehlrnal
     1081 whtgntpgqv rtlwhdprhi gwkdftayrw rlshrpktgf irvvmyegkk
imadsgpiyd
     1141 ktyaggrlgl fvfsqemvff sdlkyecrdp
```

EDIL3

Official Symbol: EDIL3

Official Name: EGF-like repeats and discoidin I-like domains 3

Gene ID: 10085

Organism: Homo sapiens

Other Aliases: DEL1

Other Designations: EGF-like repeat and discoidin I-like domain-containing protein 3; developmental endothelial locus-1; developmentally-regulated endothelial cell locus 1 protein; integrin-binding protein DEL1

Nucleotide sequence:

NCBI Reference Sequence: NM_005711.3

LOCUS NM_005711

ACCESSION NM_005711

```
      1 agaagccccg cagccgccgc gcggagaaca gcgacagccg agcgcccggt
ccgcctgtct
       61 gccggtgggt ctgcctgccc gcgcagcaga cccggggcgg ccgcgggagc
ccgcgccccg
      121 cccgccgcgc ctctgccggg acccacccgc agcggagggc tgagcccgcc
ggcggctccc
      181 cggagctcac ccacctccgc gcgccggagc gcaggcaaaa ggggaggaaa
ggctcctctc
      241 tttagtcacc actctcgccc tctccaagaa tttgtttaac aaagcgctga
ggaaagagaa
      301 cgtcttcttg aattctttag taggggcgga gtctgctgct gccctgcgct
gccacctcgg
      361 ctacactgcc ctccgcgacg acccctgacc agccggggtc acgtccggga
gacgggatca
      421 tgaagcgctc ggtagccgtc tggctcttgg tcgggctcag cctcggtgtc
ccccagttcg
      481 gcaaaggtga tatttgtgat cccaatccat gtgaaaatgg aggtatctgt
ttgccaggat
      541 tggctgatgg ttccttttcc tgtgagtgtc cagatggctt cacagacccc
aactgttcta
      601 gtgttgtgga ggttgcatca gatgaagaag aaccaacttc agcaggtccc
tgcactccta
      661 atccatgcca taatggagga acctgtgaaa taagtgaagc ataccgaggg
gatacattca
      721 taggctatgt ttgtaaatgt ccccgaggat ttaatgggat tcactgtcag
cacaacataa
      781 atgaatgcga agttgagcct tgcaaaaatg gtggaatatg tacagatctt
gttgctaact
      841 attcctgtga gtgcccaggc gaatttatgg gaagaaattg tcaatacaaa
tgctcaggcc
      901 cactgggaat tgaaggtgga attatatcaa accagcaaat cacagcttcc
tctactcacc
      961 gagctctttt tggactccaa aaatggtatc cctactatgc acgtcttaat
aagaaggggc
     1021 ttataaatgc gtggacagct gcagaaaatg acagatggcc gtggattcag
ataaatttgc
     1081 aaaggaaaat gagagttact ggtgtgatta cccaaggagc caagaggatt
ggaagcccag
     1141 agtatataaa atcctacaaa attgcctaca gtaatgatgg aaagacttgg
gcaatgtaca
```

1201 aagtgaaagg caccaatgaa gacatggtgt ttcgtggaaa cattgataac
aacactccat
1261 atgctaactc tttcacaccc cccataaaag ctcagtatgt aagactctat
ccccaagttt
1321 gtcgaagaca ttgcactttg cgaatggaac ttcttggctg tgaactgtcg
ggttgttctg
1381 agcctctggg tatgaaatca ggacatatac aagactatca gatcactgcc
tccagcatct
1441 tcagaacgct caacatggac atgttcactt gggaaccaag gaaagctcgg
ctggacaagc
1501 aaggcaaagt gaatgcctgg acctctggcc acaatgacca gtcacaatgg
ttacaggtgg
1561 atcttcttgt tccaaccaaa gtgactggca tcattacaca aggagctaaa
gattttggtc
1621 atgtacagtt tgttggctcc tacaaactgg cttacagcaa tgatggagaa
cactggactg
1681 tataccagga tgaaaagcaa agaaaagata aggttttcca gggaaatttt
gacaatgaca
1741 ctcacagaaa aaatgtcatc gaccctccca tctatgcacg acacataaga
atccttcctt
1801 ggtcctggta cgggaggatc acattgcggt cagagctgct gggctgcaca
gaggaggaat
1861 gaggggaggc tacatttcac aaccctcttc cctatttccc taaaagtatc
tccatggaat
1921 gaactgtgca aaatctgtag gaaactgaat ggtttttttt ttttttcat
gaaaaagtgc
1981 tcaaattatg gtaggcaact aacggtgttt ttaaggggggt ctaagcctgc
cttttcaatg
2041 atttaatttg attttatttt atccgtcaaa tctcttaagt aacaacacat
taagtgtgaa
2101 ttacttttct ctcattgttt cctgaattat tcgcattggt agaaatatat
tagggaaaga
2161 aagtagcctt ctttttatag caagagtaaa aaagtctcaa agtcatcaaa
taagagcaag
2221 agttgataga gcttttacaa tcaatactca cctaattctg ataaaaggaa
tactgcaatg
2281 ttagcaataa gtttttttct tctgtaatga ctctacgtta tcctgtttcc
ctgtgcctac
2341 caaacactgt caatgtttat tacaaaattt taaagaagaa tatgtaacat
gcagtactga
2401 tattataatt ctcattttac tttcattatt tctaataaga gattatgtga
cttctttttc
2461 ttttagttct attctacatt cttaatattg tatattacct gaataattca
attttttttct
2521 aattgaattt cctattagtt gactaaaaga agtgtcatgt ttactcatat
atgtagaaca
2581 tgactgccta tcagtagatt gatctgtatt taatattcgt taattaaatc
tgcagtttta
2641 tttttgaagg aagccataac tatttaattt ccaaataatt gcttcataaa
gaatcccata
2701 ctctcagttt gcacaaaaga acaaaaaata tatatgtctc tttaaattta
aatcttcatt
2761 tagatggtaa ttacatatcc ttatatttac tttaaaaaat cggcttattt
gtttatttta
2821 taaaaaattt agcaaagaaa tattaatata gtgctgcata gtttggccaa
gcatactcat
2881 catttctttg ttcagctcca catttcctgt gaaactaaca tcttattgag
atttgaaact
2941 ggtggtagtt tcccaggaag gcacaggtgg agtt

Protein sequence:

NCBI Reference Sequence: NP_005702.3

LOCUS NP_005702

ACCESSION NP_005702

```
      1 mkrsvavwll vglslgvpqf gkgdicdpnp cenggiclpg ladgsfscec
pdgftdpncs
       61 svvevasdee eptsagpctp npchnggtce iseayrgdtf igyvckcprg
fngihcqhni
      121 necevepckn ggictdlvan yscecpgefm grncqykcsg plgieggiis
nqqitassth
      181 ralfglqkwy pyyarlnkkg linawtaaen drwpwiqinl qrkmrvtgvi
tqgakrigsp
      241 eyiksykiay sndgktwamy kvkgtnedmv frgnidnntp yansftppik
aqyvrlypqv
      301 crrhctlrme llgcelsgcs eplgmksghi qdyqitassi frtlnmdmft
weprkarldk
      361 qgkvnawtsg hndqsqwlqv dllvptkvtg iitqgakdfg hvqfvgsykl
aysndgehwt
      421 vyqdekqrkd kvfqgnfdnd thrknvidpp iyarhirilp wswygritlr
sellgcteee
```

HMOX1

Official Symbol: HMOX1

Official Name: heme oxygenase (decycling) 1

Gene ID: 3162

Organism: Homo sapiens

Other Aliases: CTA-286B10.6, HO-1, HSP32, bK286B10

Other Designations: heat shock protein, 32-kD; heme oxygenase 1

Nucleotide sequence:

NCBI Reference Sequence: NM_002133.2

LOCUS NM_002133

ACCESSION NM_002133

```
      1 aaatgtgacc ggccgcggct ccggcagtca acgcctgcct cctctcgagc
gtcctcagcg
       61 cagccgccgc ccgcggagcc agcacgaacg agcccagcac cggccggatg
gagcgtccgc
      121 aacccgacag catgccccag gatttgtcag aggccctgaa ggaggccacc
aaggaggtgc
      181 acacccaggc agagaatgct gagttcatga ggaactttca gaagggccag
gtgacccgag
      241 acggcttcaa gctggtgatg gcctccctgt accacatcta tgtggccctg
gaggaggaga
      301 ttgagcgcaa caaggagagc ccagtcttcg cccctgtcta cttcccagaa
gagctgcacc
```

```
      361 gcaaggctgc cctggagcag gacctggcct tctggtacgg gccccgctgg
caggaggtca
      421 tcccctacac accagccatg cagcgctatg tgaagcggct ccacgaggtg
gggcgcacag
      481 agcccgagct gctggtggcc cacgcctaca cccgctacct gggtgacctg
tctgggggcc
      541 aggtgctcaa aaagattgcc cagaaagccc tggacctgcc cagctctggc
gagggcctgg
      601 ccttcttcac cttccccaac attgccagtg ccaccaagtt caagcagctc
taccgctccc
      661 gcatgaactc cctggagatg actcccgcag tcaggcagag ggtgatagaa
gaggccaaga
      721 ctgcgttcct gctcaacatc cagctctttg aggagttgca ggagctgctg
acccatgaca
      781 ccaaggacca gagcccctca cgggcaccag ggcttcgcca gcgggccagc
aacaaagtgc
      841 aagattctgc ccccgtggag actcccagag ggaagccccc actcaacacc
cgctcccagg
      901 ctccgcttct ccgatgggtc cttacactca gctttctggt ggcgacagtt
gctgtagggc
      961 tttatgccat gtgaatgcag gcatgctggc tcccagggcc atgaactttg
tccggtggaa
     1021 ggccttcttt ctagagaggg aattctcttg gctggcttcc ttaccgtggg
cactgaaggc
     1081 tttcagggcc tccagccctc tcactgtgtc cctctctctg gaaaggagga
aggagcctat
     1141 ggcatcttcc ccaacgaaaa gcacatccag gcaatggcct aaacttcaga
gggggcgaag
     1201 ggatcagccc tgcccttcag catcctcagt tcctgcagca gagcctggaa
gacaccctaa
     1261 tgtggcagct gtctcaaacc tccaaaagcc ctgagtttca agtatccttg
ttgacacggc
     1321 catgaccact ttccccgtgg gccatggcaa tttttacaca aacctgaaaa
gatgttgtgt
     1381 cttgtgtttt tgtcttattt ttgttggagc cactctgttc ctggctcagc
ctcaaatgca
     1441 gtatttttgt tgtgttctgt tgtttttata gcagggttgg ggtggttttt
gagccatgcg
     1501 tgggtgggga gggaggtgtt taacggcact gtggccttgg tctaactttt
gtgtgaaata
      1561 ataaacaaca ttgtctgata gtagcttgaa aaaaaaaaaa aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_002124.1

LOCUS NP_002124

ACCESSION NP_002124

```
      1 merpqpdsmp qdlsealkea tkevhtqaen aefmrnfqkg qvtrdgfklv
maslyhiyva
       61 leeeiernke spvfapvyfp eelhrkaale qdlafwygpr wqevipytpa
mqryvkrlhe
      121 vgrtepellv ahaytrylgd lsggqvlkki aqkaldlpss geglafftfp
niasatkfkq
      181 lyrsrmnsle mtpavrqrvi eeaktaflln iqlfeelqel lthdtkdqsp
srapglrqra
      241 snkvqdsapv etprgkppln trsqapllrw vltlsflvat vavglyam
```

NUCB1

Official Symbol: NUCB1

Official Name: nucleobindin 1

Gene ID: 4924

Organism: Homo sapiens

Other Aliases: CALNUC, NUC

Other Designations: nucleobindin-1

Nucleotide sequence:

NCBI Reference Sequence: NM_006184.5

LOCUS NM_006184

ACCESSION NM_006184

```
      1 gcggaagtta tttttccccc ggccggcagg gagttgtagt tatctttgaa
agccttctct
       61 ctcttttggc ataggcggga aatgtggcgt caagaggctg ggattccgag
aaagaagcga
      121 ggctttcacg aaaatgcgag cggcttcggc aaggggcggg gccgctggac
gtggatgaaa
      181 gctacagagc cagcgtggac caatcagacc tctttggggc ggggcctctg
tggataagtg
      241 ggcgtggtct agggggggaag tcaccaagaa cgcaccggag gtccttgccc
gccctggaaa
      301 acgccctctg cggtgaagga gagaccacac tgccatgcct ccctctgggc
cccgaggaac
      361 cctccttctg ttgccgctgc tgctgctgct cctgcttcgc gccgtgctgg
ctgtcccccct
      421 ggagcgaggg gcgcccaaca aggaggagac ccctgcgact gagagtcccg
acacaggcct
      481 gtactaccac cggtacctcc aggaggtcat cgatgtactg gagacggatg
ggcatttccg
      541 agagaagctg caggctgcca atgcggagga catcaagagc gggaagctga
gccgagagct
      601 ggactttgtc agccaccacg tccgcaccaa gctggatgag ctcaagcgac
aggaggtgtc
      661 acggctgcgg atgctgctca aggccaagat ggacgccgag caggatccca
atgtacaggt
      721 ggatcatctg aatctcctga aacagtttga acacctggac cctcagaacc
agcatacatt
      781 cgaggcccgc gacctggagc tgctgatcca gacggccacc cgggaccttg
cccagtacga
      841 cgcagcccat catgaagagt tcaagcgcta cgagatgctt aaggaacacg
agagacggcg
      901 ttatctggag tcactgggag aggagcagag aaaggaggcg gagaggaagc
tggaagagca
      961 acagcgccgg caccgcgagc accctaaagt caacgtgcct ggcagccaag
cccagttgaa
```

```
     1021 ggaggtgtgg gaggagctgg atggactgga ccccaacagg tttaacccca
agaccttctt
     1081 catactgcat gatatcaaca gtgatggtgt cctggatgag caggagctgg
aggcactctt
     1141 caccaaggag ctggagaaag tgtacgaccc aaagaatgag gaggacgaca
tgcgggagat
     1201 ggaggaggag cgactgcgca tgcgggagca tgtgatgaag aatgtggaca
ccaaccagga
     1261 ccgcctcgtg accctggagg agttcctcgc atccactcag aggaaggagt
ttggggacac
     1321 cggggagggc tgggagacag tggagatgca ccctgcctac accgaggaag
agctgaggcg
     1381 ctttgaagag gagctggctg cccgggaggc agagctgaat gccaaggccc
agcgcctcag
     1441 ccaggagaca gaggctctag ggcggtccca gggccgcctg gaggcccaga
agagagagct
     1501 gcagcaggct gtgctgcaca tggagcagcg gaagcagcag cagcagcagc
agcaaggcca
     1561 caaggccccg gctgcccacc ctgaggggca gctcaagttc cacccagaca
cagacgatgt
     1621 acctgtccca gctccagccg gtgaccagaa ggaggtggac acttcagaaa
agaaacttct
     1681 cgagcggctc cctgaggttg aggtgcccca gcatctgtga tcctccggga
ccccagccct
     1741 caggattcct gatgctccaa ggcgactgat gggcgctgga tgaagtggca
cagtcagctt
     1801 ccctgggggc tggtgtcatg ttgggctcct ggggcggggg cacggcctgg
catttcacgc
     1861 attgctgcca ccccaggtcc acctgtctcc actttcacag cctccaagtc
tgtggctctt
     1921 cccttctgtc ctccgagggg cttgccttct ctcgtgtcca gtgaggtgct
cagtgatcgg
     1981 cttaacttag agaagcccgc cccctcccct tctccgtctg tcccaagagg
gtctgctctg
     2041 agcctgcgtt cctaggtggc tcggcctcag ctgcctgggt tgtggccgcc
ctagcatcct
     2101 gtatgcccac agctactgga atccccgctg ctgctccggg ccaagcttct
ggttgattaa
     2161 tgagggcatg gggtggtccc tcaagacctt cccctacctt ttgtggaacc
agtgatgcct
     2221 caaagacagt gtcccctcca cagctgggtg ccaggggcag gggatcctca
gtatagccgg
     2281 tgaaccctga taccaggagc ctgggcctcc ctgaacccct ggcttccagc
catctcatcg
     2341 ccagcctcct cctggacctc ttggccccca gcccccttccc cacacagccc
cagaagggtc
     2401 ccagagctga ccccactcca ggacctaggc ccagcccctc agcctcatct
ggagcccctg
     2461 aagaccagtc ccacccacct ttctggcctc atctgacact gctccgcatc
ctgctgtgtg
     2521 tcctgttcca tgttccggtt ccatccaaat acactttctg gaacaaatgc
atggctccaa
      2581 aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_006175.2

LOCUS NP_006175

ACCESSION NP_006175

```
      1 mppsgprgtl lllpllllll lravlavple rgapnkeetp atespdtgly
yhrylqevid
       61 vletdghfre klqaanaedi ksgklsreld fvshhvrtkl delkrqevsr
lrmllkakmd
      121 aeqdpnvqvd hlnllkqfeh ldpqnqhtfe ardlelliqt atrdlaqyda
ahheefkrye
      181 mlkeherrry leslgeeqrk eaerkleeqq rrhrehpkvn vpgsqaqlke
vweeldgldp
      241 nrfnpktffi lhdinsdgvl deqelealft kelekvydpk needdmreme
eerlrmrehv
      301 mknvdtnqdr lvtleeflas tqrkefgdtg egwetvemhp ayteeelrrf
eeelaareae
      361 lnakaqrlsq etealgrsqg rleaqkrelq qavlhmeqrk qqqqqqqghk
apaahpegql
      421 kfhpdtddvp vpapagdqke vdtsekklle rlpevevpqh l
```

CS010

Official Symbol: C19orf10

Official Name: chromosome 19 open reading frame 10

Gene ID: 56005

Organism: Homo sapiens

Other Aliases: EUROIMAGE1875335, IL25, IL27, IL27w, R33729_1, SF20

Other Designations: UPF0556 protein C19orf10; interleukin 25; interleukin 27 working designation; interleukin-25; stromal cell-derived growth factor SF20

Nucleotide sequence:

NCBI Reference Sequence: NM_019107.3

LOCUS NM_019107

ACCESSION NM_019107

```
      1 ggcggacgct ccacgtgtcc ctcgccgcgc cccgtctacc cgcccctgcc
ctgaggaccc
       61 tagtccaaca tggcggcgcc cagcggaggg tggaacggcg tcggcgcgag
cttgtgggcc
      121 gcgctgctcc taggggccgt ggcgctgagg ccggcggagg cggtgtccga
gcccacgacg
      181 gtggcgtttg acgtgcggcc cggcggcgtc gtgcattcct tctcccataa
cgtgggcccg
      241 ggggacaaat atacgtgtat gttcacttac gcctctcaag gagggaccaa
tgagcaatgg
      301 cagatgagtc tggggaccag cgaagaccac cagcacttca cctgcaccat
ctggaggccc
      361 caggggaagt cctatctgta cttcacacag ttcaaggcag aggtgcgggg
cgctgagatt
```

```
      421 gagtacgcca tggcctactc taaagccgca tttgaaaggg aaagtgatgt
ccctctgaaa
      481 actgaggaat ttgaagtgac caaaacagca gtggctcaca ggcccggggc
attcaaagct
      541 gagctgtcca agctggtgat tgtggccaag gcatcgcgca ctgagctgtg
accagcagcc
      601 ctgttgcggg tggcaccttc tcatctccgg tgaagctgaa ggggcctgtg
tccctgaaag
      661 ggccagcaca tcactggttt tctaggaggg actcttaagt tttctacctg
ggctgacgtt
      721 gccttgtccg gaggggcttg cagggtggct gaagccctgg ggcagagaac
agagggtcca
      781 gggccctcct ggctcccaac agcttctcag ttcccacttc ctgctgagct
cttctggact
      841 caggatcgca gatccggggc acaaagaggg tggggaacat gggggctatg
ctggggaaag
      901 cagccatgct ccccccgacc tccagccgag catccttcat gagcctgcag
aactgctttc
      961 ctatgtttac ccaggggacc tcctttcaga tgaactggga agagatgaaa
tgttttttca
      1021 tatttaaata aataagaaca ttaaaaagca aaaaaaaaaa aaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_061980.1

LOCUS NP_061980

ACCESSION NP_061980

```
      1 maapsggwng vgaslwaall lgavalrpae avsepttvaf dvrpggvvhs
fshnvgpgdk
       61 ytcmftyasq ggtneqwqms lgtsedhqhf tctiwrpqgk sylyftqfka
evrgaeieya
      121 mayskaafer esdvplktee fevtktavah rpgafkaels klvivakasr tel
```

PLIN2

Official Symbol: PLIN2

Official Name: perilipin 2

Gene ID: 123

Organism: Homo sapiens

Other Aliases: RP11-151J10.1, ADFP, ADRP

Other Designations: adipophilin; adipose differentiation-related protein; perilipin-2

Nucleotide sequence:

NCBI Reference Sequence: NM_001122.3

LOCUS NM_001122

ACCESSION NM_001122

```
      1 ccgagggtga cactcgggct tgggacaggg cgtgctgccg cgggtcacgt
gctgcggagg
       61 cttggggagg ggcggcgagg cggggtttat agcccgggcg cccgcgggcc
ccacgctttg
      121 accgggtcgt ggcagccgga gtcgtcttcg ggacgcgcct gctcttcgcc
tttcgctgca
      181 gtccgtcgat ttctttctcc aggaagaaaa atggcatccg ttgcagttga
tccacaaccg
      241 agtgtggtga ctcgggtggt caacctgccc ttggtgagct ccacgtatga
cctcatgtcc
      301 tcagcctatc tcagtacaaa ggaccagtat ccctacctga agtctgtgtg
tgagatggca
      361 gagaacggtg tgaagaccat cacctccgtg gccatgacca gtgctctgcc
catcatccag
      421 aagctagagc cgcaaattgc agttgccaat acctatgcct gtaaggggct
agacaggatt
      481 gaggagagac tgcctattct gaatcagcca tcaactcaga ttgttgccaa
tgccaaaggc
      541 gctgtgactg gggcaaaaga tgctgtgacg actactgtga ctggggccaa
ggattctgtg
      601 gccagcacga tcacaggggt gatggacaag accaaagggg cagtgactgg
cagtgtggag
      661 aagaccaagt ctgtggtcag tggcagcatt aacacagtct tggggagtcg
gatgatgcag
      721 ctcgtgagca gtggcgtaga aaatgcactc accaaatcag agctgttggt
agaacagtac
      781 ctccctctca ctgaggaaga actagaaaaa gaagcaaaaa aagttgaagg
atttgatctg
      841 gttcagaagc caagttatta tgttagactg ggatccctgt ctaccaagct
tcactcccgt
      901 gcctaccagc aggctctcag cagggttaaa gaagctaagc aaaaaagcca
acagaccatt
      961 tctcagctcc attctactgt tcacctgatt gaatttgcca ggaagaatgt
gtatagtgcc
     1021 aatcagaaaa ttcaggatgc tcaggataag ctctacctct catgggtaga
gtggaaaagg
     1081 agcattggat atgatgatac tgatgagtcc cactgtgctg agcacattga
gtcacgtact
     1141 cttgcaattg cccgcaacct gactcagcag ctccagacca cgtgccacac
cctcctgtcc
     1201 aacatccaag gtgtaccaca gaacatccaa gatcaagcca agcacatggg
ggtgatggca
     1261 ggcgacatct actcagtgtt ccgcaatgct gcctccttta aagaagtgtc
tgacagcctc
     1321 ctcacttcta gcaaggggca gctgcagaaa atgaaggaat ctttagatga
cgtgatggat
     1381 tatcttgtta acaacacgcc cctcaactgg ctggtaggtc ccttttatcc
tcagctgact
     1441 gagtctcaga atgctcagga ccaaggtgca gagatggaca agagcagcca
ggagacccag
     1501 cgatctgagc ataaaactca ttaaacctgc ccctatcact agtgcatgct
gtggccagac
     1561 agatgacacc ttttgttatg ttgaaattaa cttgctaggc aaccctaaat
tgggaagcaa
     1621 gtagctagta taaaggccct caattgtagt tgtttccagc tgaattaaga
gctttaaagt
     1681 ttctggcatt agcagatgat ttctgttcac ctggtaagaa aagaatgata
ggcttgtcag
```

```
     1741 agcctatagc cagaactcag aaaaaattca aatgcactta tgttctcatt
ctatggccat
     1801 tgtgttgcct ctgttactgt ttgtattgaa taaaaacatc ttcatgtggg
ctggggtaga
     1861 aactggtgtc tgctctggtg tgatctgaaa aggcgtcttc actgctttat
ctcatgatgc
     1921 ttgcttgtaa aacttgattt tagtttttca tttctcaaat aggaatacta
cctttgaatt
     1981 caataaaatt cactgcagga tagaccagtt aaaaaaaaaa aaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001113.2

LOCUS NP_001113

ACCESSION NP_001113.2

```
      1 masvavdpqp svvtrvvnlp lvsstydlms saylstkdqy pylksvcema
engvktitsv
       61 amtsalpiiq klepqiavan tyackgldri eerlpilnqp stqivanakg
avtgakdavt
      121 ttvtgakdsv astitgvmdk tkgavtgsve ktksvvsgsi ntvlgsrmmq
lvssgvenal
      181 tksellveqy lplteeelek eakkvegfdl vqkpsyyvrl gslstklhsr
ayqqalsrvk
      241 eakqksqqti sqlhstvhli efarknvysa nqkiqdaqdk lylswvewkr
sigyddtdes
      301 hcaehiesrt laiarnltqq lqttchtlls niqgvpqniq dqakhmgvma
gdiysvfrna
      361 asfkevsdsl ltsskgqlqk mkeslddvmd ylvnntplnw lvgpfypqlt
esqnaqdqga
      421 emdkssqetq rsehkth
```

ATP5A

Official Symbol: ATP5A1

Official Name: ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle

Gene ID: 498

Organism: Homo sapiens

Other Aliases: ATP5A, ATP5AL2, ATPM, MOM2, OMR, ORM, hATP1

Other Designations: ATP synthase alpha chain, mitochondrial; ATP synthase subunit alpha, mitochondrial; ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle; ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 2, non-cardiac muscle-like 2; ATP sythase (F1-ATPase) alpha subunit; mitochondrial ATP synthetase, oligomycin-resistant

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001001937.1

LOCUS NM_001001937

ACCESSION NM_001001937

```
      1 tctggcattg caagcctcgc ttcgttgcca cttcccagct cttcccgcct
tccgcggtat
       61 aatcaacact acgagagata gagccgccta gaaccagtcc ggaggctgcg
gctgcagaag
      121 taccgcctgc ggagtaactg caaagatgct gtccgtgcgc gttgctgcgg
ccgtggtccg
      181 cgcccttcct cggcgggccg gactggtctc cagaaatgct ttgggttcat
ctttcattgc
      241 tgcaaggaac ttccatgcct ctaacactca tcttcaaaag actgggactg
ctgagatgtc
      301 ctctattctt gaagagcgta ttcttggagc tgatacctct gttgatcttg
aagaaactgg
      361 gcgtgtctta agtattggtg atggtattgc ccgcgtacat gggctgagga
atgttcaagc
      421 agaagaaatg gtagagtttt cttcaggctt aaagggtatg tccttgaact
tggaacctga
      481 caatgttggt gttgtcgtgt ttggaaatga taaactaatt aaggaaggag
atatagtgaa
      541 gaggacagga gccattgtgg acgttccagt tggtgaggag ctgttgggtc
gtgtagttga
      601 tgcccttggt aatgctattg atggaaaggg tccaattggt tccaagacgc
gtaggcgagt
      661 tggtctgaaa gcccccggta tcattcctcg aatttcagtg cgggaaccaa
tgcagactgg
      721 cattaaggct gtggatagct tggtgccaat tggtcgtggt cagcgtgaac
tgattattgg
      781 tgaccgacag actgggaaaa cctcaattgc tattgacaca atcattaacc
agaaacgttt
      841 caatgatgga tctgatgaaa agaagaagct gtactgtatt tatgttgcta
ttggtcaaaa
      901 gagatccact gttgcccagt tggtgaagag acttacagat gcagatgcca
tgaagtacac
      961 cattgtggtg tcggctacgg cctcggatgc tgccccactt cagtacctgg
ctccttactc
     1021 tggctgttcc atgggagagt attttagaga caatggcaaa catgctttga
tcatctatga
     1081 cgacttatcc aaacaggctg ttgcttaccg tcagatgtct ctgttgctcc
gccgaccccc
     1141 tggtcgtgag gcctatcctg gtgatgtgtt ctacctacac tcccggttgc
tggagagagc
     1201 agccaaaatg aacgatgctt ttggtggtgg ctccttgact gctttgccag
tcatagaaac
     1261 acaggctggt gatgtgtctg cttacattcc aacaaatgtc atttccatca
ctgacggaca
     1321 gatcttcttg gaaacagaat tgttctacaa aggtatccgc cctgcaatta
acgttggtct
     1381 gtctgtatct cgtgtcggat ccgctgccca aaccagggct atgaagcagg
tagcaggtac
     1441 catgaagctg gaattggctc agtatcgtga ggttgctgct tttgcccagt
tcggttctga
     1501 cctcgatgct gccactcaac aacttttgag tcgtggcgtg cgtctaactg
agttgctgaa
```

```
     1561 gcaaggacag tattctccca tggctattga agaacaagtg gctgttatct
atgcgggtgt
     1621 aaggggatat cttgataaac tggagcccag caagattaca aagtttgaga
atgctttctt
     1681 gtctcatgtc gtcagccagc accaagcctt gttgggcact atcagggctg
atggaaagat
     1741 ctcagaacaa tcagatgcaa agctgaaaga gattgtaaca aatttcttgg
ctggatttga
     1801 agcttaaact cctgtggatt cacatcaaat accagttcag ttttgtcatt
gttctagtaa
     1861 attagttcca tttgtaaaag ggttactctc atactcctta tgtacagaaa
tcacatgaaa
     1921 aataaaggtt ccataatgca tagttaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001001937.1

LOCUS NP_001001937

ACCESSION NP_001001937

```
      1 mlsvrvaaav vralprragl vsrnalgssf iaarnfhasn thlqktgtae
mssileeril
       61 gadtsvdlee tgrvlsigdg iarvhglrnv qaeemvefss glkgmslnle
pdnvgvvvfg
      121 ndklikegdi vkrtgaivdv pvgeellgrv vdalgnaidg kgpigsktrr
rvglkapgii
      181 prisvrepmq tgikavdslv pigrgqreli igdrqtgkts iaidtiinqk
rfndgsdekk
      241 klyciyvaig qkrstvaqlv krltdadamk ytivvsatas daaplqylap
ysgcsmgeyf
      301 rdngkhalii yddlskqava yrqmslllrr ppgreaypgd vfylhsrlle
raakmndafg
      361 ggsltalpvi etqagdvsay iptnvisitd gqifletelf ykgirpainv
glsvsrvgsa
      421 aqtramkqva gtmklelaqy revaafaqfg sdldaatqql lsrgvrltel
lkqgqyspma
      481 ieeqvaviya gvrgyldkle pskitkfena flshvvsqhq allgtiradg
kiseqsdakl
      541 keivtnflag fea
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_004046.5

LOCUS NM_004046

ACCESSION NM_004046

```
      1 ggggcagtac ttccgggtca ggtgggccgg ctgtcttgac cttctttgcg
gctcggccat
       61 tttgtcccag tcagtccgga ggctgcggct gcagaagtac cgcctgcgga
gtaactgcaa
      121 agatgctgtc cgtgcgcgtt gctgcggccg tggtccgcgc ccttcctcgg
cgggccggac
      181 tggtctccag aaatgctttg ggttcatctt tcattgctgc aaggaacttc
catgcctcta
```

241 acactcatct tcaaaagact gggactgctg agatgtcctc tattcttgaa
gagcgtattc
301 ttggagctga tacctctgtt gatcttgaag aaactgggcg tgtcttaagt
attggtgatg
361 gtattgcccg cgtacatggg ctgaggaatg ttcaagcaga agaaatggta
gagttttctt
421 caggcttaaa gggtatgtcc ttgaacttgg aacctgacaa tgttggtgtt
gtcgtgtttg
481 gaaatgataa actaattaag gaaggagata tagtgaagag gacaggagcc
attgtggacg
541 ttccagttgg tgaggagctg ttgggtcgtg tagttgatgc ccttggtaat
gctattgatg
601 gaaagggtcc aattggttcc aagacgcgta ggcgagttgg tctgaaagcc
cccggtatca
661 ttcctcgaat ttcagtgcgg gaaccaatgc agactggcat taaggctgtg
gatagcttgg
721 tgccaattgg tcgtggtcag cgtgaactga ttattggtga ccgacagact
gggaaaacct
781 caattgctat tgacacaatc attaaccaga aacgtttcaa tgatggatct
gatgaaaaga
841 agaagctgta ctgtatttat gttgctattg gtcaaaagag atccactgtt
gcccagttgg
901 tgaagagact tacagatgca gatgccatga agtacaccat tgtggtgtcg
gctacggcct
961 cggatgctgc cccacttcag tacctggctc cttactctgg ctgttccatg
ggagagtatt
1021 ttagagacaa tggcaaacat gctttgatca tctatgacga cttatccaaa
caggctgttg
1081 cttaccgtca gatgtctctg ttgctccgcc gacccccctgg tcgtgaggcc
tatcctggtg
1141 atgtgttcta cctacactcc cggttgctgg agagagcagc caaaatgaac
gatgcttttg
1201 gtggtggctc cttgactgct ttgccagtca tagaaacaca ggctggtgat
gtgtctgctt
1261 acattccaac aaatgtcatt tccatcactg acggacagat cttcttggaa
acagaattgt
1321 tctacaaagg tatccgccct gcaattaacg ttggtctgtc tgtatctcgt
gtcggatccg
1381 ctgcccaaac cagggctatg aagcaggtag caggtaccat gaagctggaa
ttggctcagt
1441 atcgtgaggt tgctgctttt gcccagttcg gttctgacct cgatgctgcc
actcaacaac
1501 ttttgagtcg tggcgtgcgt ctaactgagt tgctgaagca aggacagtat
tctcccatgg
1561 ctattgaaga acaagtggct gttatctatg cgggtgtaag gggatatctt
gataaactgg
1621 agcccagcaa gattacaaag tttgagaatg ctttcttgtc tcatgtcgtc
agccagcacc
1681 aagccttgtt gggcactatc agggctgatg gaaagatctc agaacaatca
gatgcaaagc
1741 tgaaagagat tgtaacaaat ttcttggctg gatttgaagc ttaaactcct
gtggattcac
1801 atcaaatacc agttcagttt tgtcattgtt ctagtaaatt agttccattt
gtaaaagggt
1861 tactctcata ctccttatgt acagaaatca catgaaaaat aaaggttcca
taatgcatag
1921 ttaaaaa

Protein sequence (variant 2):

NCBI Reference Sequence: NP_004037.1

LOCUS NP_004037

ACCESSION NP_004037

```
      1 mlsvrvaaav vralprragl vsrnalgssf iaarnfhasn thlqktgtae
mssileeril
       61 gadtsvdlee tgrvlsigdg iarvhglrnv qaeemvefss glkgmslnle
pdnvgvvvfg
      121 ndklikegdi vkrtgaivdv pvgeellgrv vdalgnaidg kgpigsktrr
rvglkapgii
      181 prisvrepmq tgikavdslv pigrgqreli igdrqtgkts iaidtiinqk
rfndgsdekk
      241 klyciyvaig qkrstvaqlv krltdadamk ytivvsatas daaplqylap
ysgcsmgeyf
      301 rdngkhalii yddlskqava yrqmslllrr ppgreaypgd vfylhsrlle
raakmndafg
      361 ggsltalpvi etqagdvsay iptnvisitd gqifletelf ykgirpainv
glsvsrvgsa
      421 aqtramkqva gtmklelaqy revaafaqfg sdldaatqql lsrgvrltel
lkqgqyspma
      481 ieeqvaviya gvrgyldkle pskitkfena flshvvsqhq allgtiradg
kiseqsdakl
      541 keivtnflag fea
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_001257334.1

LOCUS NM_001257334

ACCESSION NM_001257334

```
      1 ggggcagtac ttccgggtca ggtgggccgg ctgtcttgac cttctttgcg
gctcggccat
       61 tttgtcccag tcagtccgga ggctgcggct gcagaagtac cgcctgcgga
gtaactgcaa
      121 agatgctgtc cgtgcgcgtt gctgcggccg tggtccgcgc ccttcctcgg
cgggccggac
      181 tggtctccag aaatgctttg ggttcatctt tcattgctgc aaggaacttc
catgcctcta
      241 acactcatct tcaaaagact gggactgctg agatgtcctc tattcttgaa
gagcgtattc
      301 ttggagctga tacctctgtt gatcttgaag aaactgggcg tgtcttaagt
attggtgatg
      361 gtattgcccg cgtacatggg ctgaggaatg ttcaagcaga agaaatggta
gagttttctt
      421 caggcttaaa gggtatgtcc ttgaacttgg aacctgacaa tgttggtgtt
gtcgtgtttg
      481 gaaatgataa actaattaag gaaggagata tagtgaagag gacaggagcc
attgtggacg
      541 gtccaattgg ttccaagacg cgtaggcgag ttggtctgaa agcccccggt
atcattcctc
      601 gaatttcagt gcgggaacca atgcagactg gcattaaggc tgtggatagc
ttggtgccaa
      661 ttggtcgtgg tcagcgtgaa ctgattattg gtgaccgaca gactgggaaa
acctcaattg
```

```
      721 ctattgacac aatcattaac cagaaacgtt tcaatgatgg atctgatgaa
aagaagaagc
      781 tgtactgtat ttatgttgct attggtcaaa agagatccac tgttgcccag
ttggtgaaga
      841 gacttacaga tgcagatgcc atgaagtaca ccattgtggt gtcggctacg
gcctcggatg
      901 ctgccccact tcagtacctg gctccttact ctggctgttc catgggagag
tattttagag
      961 acaatggcaa acatgctttg atcatctatg acgacttatc caaacaggct
gttgcttacc
     1021 gtcagatgtc tctgttgctc cgccgacccc ctggtcgtga ggcctatcct
ggtgatgtgt
     1081 tctacctaca ctcccggttg ctggagagag cagccaaaat gaacgatgct
tttggtggtg
     1141 gctccttgac tgctttgcca gtcatagaaa cacaggctgg tgatgtgtct
gcttacattc
     1201 caacaaatgt catttccatc actgacggac agatcttctt ggaaacagaa
ttgttctaca
     1261 aaggtatccg ccctgcaatt aacgttggtc tgtctgtatc tcgtgtcgga
tccgctgccc
     1321 aaaccagggc tatgaagcag gtagcaggta ccatgaagct ggaattggct
cagtatcgtg
     1381 aggttgctgc ttttgcccag ttcggttctg acctcgatgc tgccactcaa
caacttttga
     1441 gtcgtggcgt gcgtctaact gagttgctga agcaaggaca gtattctccc
atggctattg
     1501 aagaacaagt ggctgttatc tatgcgggtg taaggggata tcttgataaa
ctggagccca
     1561 gcaagattac aaagtttgag aatgctttct tgtctcatgt cgtcagccag
caccaagcct
     1621 tgttgggcac tatcagggct gatggaaaga tctcagaaca atcagatgca
aagctgaaag
     1681 agattgtaac aaatttcttg gctggatttg aagcttaaac tcctgtggat
tcacatcaaa
     1741 taccagttca gttttgtcat tgttctagta aattagttcc atttgtaaaa
gggttactct
     1801 catactcctt atgtacagaa atcacatgaa aaataaaggt tccataatgc
atagttaaaa
     1861 a
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_001244263.1

LOCUS NP_001244263

ACCESSION NP_001244263

```
      1 mlsvrvaaav vralprragl vsrnalgssf iaarnfhasn thlqktgtae
mssileeril
       61 gadtsvdlee tgrvlsigdg iarvhglrnv qaeemvefss glkgmslnle
pdnvgvvvfg
      121 ndklikegdi vkrtgaivdg pigsktrrrv glkapgiipr isvrepmqtg
ikavdslvpi
      181 grgqreliig drqtgktsia idtiinqkrf ndgsdekkkl yciyvaigqk
rstvaqlvkr
      241 ltdadamkyt ivvsatasda aplqylapys gcsmgeyfrd ngkhaliiyd
dlskqavayr
```

301 qmslllrrpp greaypgdvf ylhsrllera akmndafggg sltalpviet
qagdvsayip
361 tnvisitdgq ifletelfyk girpainvgl svsrvgsaaq tramkqvagt
mklelaqyre
421 vaafaqfgsd ldaatqqlls rgvrltellk qgqyspmaie eqvaviyagv
rgyldkleps
481 kitkfenafl shvvsqhqal lgtiradgki seqsdaklke ivtnflagfe a

Nucleotide sequence (variant 4):

NCBI Reference Sequence: NM_001001935.2

LOCUS NM_001001935

ACCESSION NM_001001935

1 ggggcagtac ttccgggtca ggtgggccgg ctgtcttgac cttctttgcg
gctcggccat
61 tttgtcccag tcagtccgga ggctgcggct gcagaagtac cgcctgcgga
gtaactgcaa
121 agatgctgtc cgtgcgcgtt gctgcggccg tggtccgcgc ccttcctcgg
cgggccggac
181 tggtctccag aaatgctttg ggttcatctt tcattgctgc aaggaacttc
catgcctcta
241 acactcatct tcaaaagact ggtaagttat tatttctcag tctacgccgc
acttactaga
301 tgaagatata aattacatac atcgtataac tgtgggactg ctgagatgtc
ctctattctt
361 gaagagcgta ttcttggagc tgatacctct gttgatcttg aagaaactgg
gcgtgtctta
421 agtattggtg atggtattgc ccgcgtacat gggctgagga atgttcaagc
agaagaaatg
481 gtagagtttt cttcaggctt aaagggtatg tccttgaact tggaacctga
caatgttggt
541 gttgtcgtgt ttggaaatga taaactaatt aaggaaggag atatagtgaa
gaggacagga
601 gccattgtgg acgttccagt tggtgaggag ctgttgggtc gtgtagttga
tgcccttggt
661 aatgctattg atggaaaggg tccaattggt tccaagacgc gtaggcgagt
tggtctgaaa
721 gcccccggta tcattcctcg aatttcagtg cgggaaccaa tgcagactgg
cattaaggct
781 gtggatagct tggtgccaat tggtcgtggt cagcgtgaac tgattattgg
tgaccgacag
841 actgggaaaa cctcaattgc tattgacaca atcattaacc agaaacgttt
caatgatgga
901 tctgatgaaa agaagaagct gtactgtatt tatgttgcta ttggtcaaaa
gagatccact
961 gttgcccagt tggtgaagag acttacagat gcagatgcca tgaagtacac
cattgtggtg
1021 tcggctacgg cctcggatgc tgccccactt cagtacctgg ctccttactc
tggctgttcc
1081 atgggagagt attttagaga caatggcaaa catgctttga tcatctatga
cgacttatcc
1141 aaacaggctg ttgcttaccg tcagatgtct ctgttgctcc gccgaccccc
tggtcgtgag
1201 gcctatcctg gtgatgtgtt ctacctacac tcccggttgc tggagagagc
agccaaaatg

```
     1261 aacgatgctt ttggtggtgg ctccttgact gctttgccag tcatagaaac
acaggctggt
     1321 gatgtgtctg cttacattcc aacaaatgtc atttccatca ctgacggaca
gatcttcttg
     1381 gaaacagaat tgttctacaa aggtatccgc cctgcaatta acgttggtct
gtctgtatct
     1441 cgtgtcggat ccgctgccca aaccagggct atgaagcagg tagcaggtac
catgaagctg
     1501 gaattggctc agtatcgtga ggttgctgct tttgcccagt tcggttctga
cctcgatgct
     1561 gccactcaac aacttttgag tcgtggcgtg cgtctaactg agttgctgaa
gcaaggacag
     1621 tattctccca tggctattga agaacaagtg gctgttatct atgcgggtgt
aaggggatat
     1681 cttgataaac tggagcccag caagattaca aagtttgaga atgctttctt
gtctcatgtc
     1741 gtcagccagc accaagcctt gttgggcact atcagggctg atggaaagat
ctcagaacaa
     1801 tcagatgcaa agctgaaaga gattgtaaca aatttcttgg ctggatttga
agcttaaact
     1861 cctgtggatt cacatcaaat accagttcag ttttgtcatt gttctagtaa
attagttcca
     1921 tttgtaaaag ggttactctc atactcctta tgtacagaaa tcacatgaaa
aataaaggtt
      1981 ccataatgca tagttaaaaa
```

Protein sequence (variant 4):

NCBI Reference Sequence: NP_001001935.1

LOCUS NP_001001935

ACCESSION NP_001001935.1

```
      1 mssileeril gadtsvdlee tgrvlsigdg iarvhglrnv qaeemvefss
glkgmslnle
       61 pdnvgvvvfg ndklikegdi vkrtgaivdv pvgeellgrv vdalgnaidg
kgpigsktrr
      121 rvglkapgii prisvrepmq tgikavdslv pigrgqreli igdrqtgkts
iaidtiinqk
      181 rfndgsdekk klyciyvaig qkrstvaqlv krltdadamk ytivvsatas
daaplqylap
      241 ysgcsmgeyf rdngkhalii yddlskqava yrqmslllrr ppgreaypgd
vfylhsrlle
      301 raakmndafg ggsltalpvi etqagdvsay iptnvisitd gqifletelf
ykgirpainv
      361 glsvsrvgsa aqtramkqva gtmklelaqy revaafaqfg sdldaatqql
lsrgvrltel
      421 lkqgqyspma ieeqvaviya gvrgyldkle pskitkfena flshvvsqhq
allgtiradg
      481 kiseqsdakl keivtnflag fea
```

Nucleotide sequence (variant 5):

NCBI Reference Sequence: NM_001257335.1

LOCUS NM_001257335

ACCESSION NM_001257335

```
      1 ggggcagtac ttccgggtca ggtgggccgg ctgtcttgac cttctttgcg
gctcggccat
       61 tttgtcccag tcagtccgga ggctgcggct gcagaagtac cgcctgcgga
gtaactgcaa
      121 agatgctgtc cgtgcgcgtt gctgcggccg tggtccgcgc ccttcctcgg
cgggccggac
      181 tggtgagcac cgaaggccgg catgatgcag gcggccgggt ggggctgcag
ggtggtggtg
      241 cgccggctcg ggcgctctct gcaggagggc gaggggctgt ggcgaatgcc
gccatcttgc
      301 acccgtggct tctccggctg gacagagcag gcgacacagg tgccctttg
ctcgtcacct
      361 gcgcagaggc agaatggtac agggcagaca gttaactcga tggtgtccag
agacagggcc
      421 tcaagattcc tgtcttcggc tgacagcggc cctagaaggg ggatcttggg
tgaaggtcag
      481 ggcttgggcg ctagctctcc gaggcctgtt ctgaatcggt ctccagaaat
gctttgggtt
      541 catctttcat tgctgcaagg aacttccatg cctctaacac tcatcttcaa
aagactggga
      601 ctgctgagat gtcctctatt cttgaagagc gtattcttgg agctgatacc
tctgttgatc
      661 ttgaagaaac tgggcgtgtc ttaagtattg gtgatggtat tgcccgcgta
catgggctga
      721 ggaatgttca agcagaagaa atggtagagt tttcttcagg cttaaagggt
atgtccttga
      781 acttggaacc tgacaatgtt ggtgttgtcg tgtttggaaa tgataaacta
attaaggaag
      841 gagatatagt gaagaggaca ggagccattg tggacgttcc agttggtgag
gagctgttgg
      901 gtcgtgtagt tgatgccctt ggtaatgcta ttgatggaaa gggtccaatt
ggttccaaga
      961 cgcgtaggcg agttggtctg aaagcccccg gtatcattcc tcgaatttca
gtgcgggaac
     1021 caatgcagac tggcattaag gctgtggata gcttggtgcc aattggtcgt
ggtcagcgtg
     1081 aactgattat tggtgaccga cagactggga aaacctcaat tgctattgac
acaatcatta
     1141 accagaaacg tttcaatgat ggatctgatg aaaagaagaa gctgtactgt
atttatgttg
     1201 ctattggtca aaagagatcc actgttgccc agttggtgaa gagacttaca
gatgcagatg
     1261 ccatgaagta caccattgtg gtgtcggcta cggcctcgga tgctgcccca
cttcagtacc
     1321 tggctcctta ctctggctgt tccatgggag agtattttag agacaatggc
aaacatgctt
     1381 tgatcatcta tgacgactta tccaaacagg ctgttgctta ccgtcagatg
tctctgttgc
     1441 tccgccgacc ccctggtcgt gaggcctatc ctggtgatgt gttctaccta
cactcccggt
     1501 tgctggagag agcagccaaa atgaacgatg cttttggtgg tggctccttg
actgctttgc
     1561 cagtcataga aacacaggct ggtgatgtgt ctgcttacat tccaacaaat
gtcatttcca
     1621 tcactgacgg acagatcttc ttggaaacag aattgttcta caaaggtatc
cgccctgcaa
     1681 ttaacgttgg tctgtctgta tctcgtgtcg gatccgctgc ccaaaccagg
gctatgaagc
```

```
     1741 aggtagcagg taccatgaag ctggaattgg ctcagtatcg tgaggttgct
gcttttgccc
     1801 agttcggttc tgacctcgat gctgccactc aacaactttt gagtcgtggc
gtgcgtctaa
     1861 ctgagttgct gaagcaagga cagtattctc ccatggctat tgaagaacaa
gtggctgtta
     1921 tctatgcggg tgtaagggga tatcttgata aactggagcc cagcaagatt
acaaagtttg
     1981 agaatgcttt cttgtctcat gtcgtcagcc agcaccaagc cttgttgggc
actatcaggg
     2041 ctgatggaaa gatctcagaa caatcagatg caaagctgaa agagattgta
acaaatttct
     2101 tggctggatt tgaagcttaa actcctgtgg attcacatca aataccagtt
cagttttgtc
     2161 attgttctag taaattagtt ccatttgtaa aagggttact ctcatactcc
ttatgtacag
     2221 aaatcacatg aaaaataaag gttccataat gcatagttaa aaa
```

Protein sequence (variant 5):

NCBI Reference Sequence: NP_001244264.1

LOCUS NP_001244264

ACCESSION NP_001244264

```
      1 mssileeril gadtsvdlee tgrvlsigdg iarvhglrnv qaeemvefss
glkgmslnle
       61 pdnvgvvvfg ndklikegdi vkrtgaivdv pvgeellgrv vdalgnaidg
kgpigsktrr
      121 rvglkapgii prisvrepmq tgikavdslv pigrgqreli igdrqtgkts
iaidtiinqk
      181 rfndgsdekk klyciyvaig qkrstvaqlv krltdadamk ytivvsatas
daaplqylap
      241 ysgcsmgeyf rdngkhalii yddlskqava yrqmslllrr ppgreaypgd
vfylhsrlle
      301 raakmndafg ggsltalpvi etqagdvsay iptnvisitd gqifletelf
ykgirpainv
      361 glsvsrvgsa aqtramkqva gtmklelaqy revaafaqfg sdldaatqql
lsrgvrltel
      421 lkqgqyspma ieeqvaviya gvrgyldkle pskitkfena flshvvsqhq
allgtiradg
      481 kiseqsdakl keivtnflag fea
```

HSPA9

(See entry for GRP75 above)

MARS

Official Symbol: MARS

Official Name: methionyl-tRNA synthetase

Gene ID: 4141

Organism: Homo sapiens

Other Aliases: METRS, MRS, MTRNS

Other Designations: cytosolic methionyl-tRNA synthetase; methionine tRNA ligase 1, cytoplasmic; methionine--tRNA ligase, cytoplasmic

Nucleotide sequence:

NCBI Reference Sequence: NM_004990.3

LOCUS NM_004990

ACCESSION NM_004990

```
      1 aaatagtcta ctttccggta gcggtgccag ggcagtggcc taatacggaa
ctccatttcc
       61 cggcgtgcct cgcggaggcc gctgaactca gaagcgggag gccggttccg
gttgcatcag
      121 cgagggattc acggcgaaat gagactgttc gtgagtgatg gcgtcccggg
ttgcttgccg
      181 gtgctggccg ccgccgggag agcccggggc agagcagagg tgctcatcag
cactgtaggc
      241 ccggaagatt gtgtggtccc gttcctgacc cggcctaagg tccctgtctt
gcagctggat
      301 agcggcaact acctcttctc cactagtgca atctgccgat attttttttt
gttatctggc
      361 tgggagcaag atgacctcac taaccagtgg ctggaatggg aagcgacaga
gctgcagcca
      421 gctttgtctg ctgccctgta ctatttagtg gtccaaggca agaaggggga
agatgttctt
      481 ggttcagtgc ggagagccct gactcacatt gaccacagct tgagtcgtca
gaactgtcct
      541 ttcctggctg gggagacaga atctctagcc gacattgttt tgtggggagc
cctataccca
      601 ttactgcaag atcccgccta cctccctgag gagctgagtg ccctgcacag
ctggttccag
      661 acactgagta cccaggaacc atgtcagcga gctgcagaga ctgtactgaa
acagcaaggt
      721 gtcctggctc tccggcctta cctccaaaag cagccccagc ccagccccgc
tgagggaagg
      781 gctgtcacca atgagcctga ggaggaggag ctggctaccc tatctgagga
ggagattgct
      841 atggctgtta ctgcttggga gaagggccta gaaagtttgc ccccgctgcg
gccccagcag
      901 aatccagtgt tgcctgtggc tggagaaagg aatgtgctca tcaccagtgc
cctcccttac
      961 gtcaacaatg tcccccacct tgggaacatc attggttgtg tgctcagtgc
cgatgtcttt
     1021 gccaggtact ctcgcctccg ccagtggaac accctctatc tgtgtgggac
agatgagtat
     1081 ggtacagcaa cagagaccaa ggctctggag gagggactaa ccccccagga
gatctgcgac
     1141 aagtaccaca tcatccatgc tgacatctac cgctggttta acatttcgtt
tgatattttt
     1201 ggtcgcacca ccactccaca gcagaccaaa atcacccagg acattttcca
gcagttgctg
```

```
     1261 aaacgaggtt ttgtgctgca agatactgtg gagcaactgc gatgtgagca
ctgtgctcgc
     1321 ttcctggctg accgcttcgt ggagggcgtg tgtcccttct gtggctatga
ggaggctcgg
     1381 ggtgaccagt gtgacaagtg tggcaagctc atcaatgctg tcgagcttaa
gaagcctcag
     1441 tgtaaagtct gccgatcatg ccctgtggtg cagtcgagcc agcacctgtt
tctggacctg
     1501 cctaagctgg agaagcgact ggaggagtgg ttggggagga cattgcctgg
cagtgactgg
     1561 acacccaatg cccagtttat cacccgttct tggcttcggg atggcctcaa
gccacgctgc
     1621 ataacccgag acctcaaatg gggaacccct gtacccttag aaggttttga
agacaaggta
     1681 ttctatgtct ggtttgatgc cactattggc tatctgtcca tcacagccaa
ctacacagac
     1741 cagtgggaga gatggtggaa gaacccagag caagtggacc tgtatcagtt
catggccaaa
     1801 gacaatgttc ctttccatag cttagtcttt ccttgctcag ccctaggagc
tgaggataac
     1861 tataccttgg tcagccacct cattgctaca gagtacctga actatgagga
tgggaaattc
     1921 tctaagagcc gcggtgtggg agtgtttggg gacatggccc aggacacggg
gatccctgct
     1981 gacatctggc gcttctatct gctgtacatt cggcctgagg gccaggacag
tgctttctcc
     2041 tggacggacc tgctgctgaa gaataattct gagctgctta acaacctggg
caacttcatc
     2101 aacagagctg ggatgtttgt gtctaagttc tttgggggct atgtgcctga
gatggtgctc
     2161 acccctgatg atcagcgcct gctggcccat gtcaccctgg agctccagca
ctatcaccag
     2221 ctacttgaga aggttcggat ccgggatgcc ttgcgcagta tcctcaccat
atctcgacat
     2281 ggcaaccaat atattcaggt gaatgagccc tggaagcgga ttaaaggcag
tgaggctgac
     2341 aggcaacggg caggaacagt gactggcttg gcagtgaata tagctgcctt
gctctctgtc
     2401 atgcttcagc cttacatgcc cacggttagt gccacaatcc aggcccagct
gcagctccca
     2461 cctccagcct gcagtatcct gctgacaaac ttcctgtgta ccttaccagc
aggacaccag
     2521 attggcacag tcagtccctt gttccaaaaa ttggaaaatg accagattga
aagtttaagg
     2581 cagcgctttg gaggggggcca ggcaaaaacg tccccgaagc cagcagttgt
agagactgtt
     2641 acaacagcca agccacagca gatacaagcg ctgatggatg aagtgacaaa
acaaggaaac
     2701 attgtccgag aactgaaagc acaaaaggca gacaagaacg aggttgctgc
ggaggtggcg
     2761 aaactcttgg atctaaagaa acagttggct gtagctgagg ggaaaccccc
tgaagcccct
     2821 aaaggcaaga agaaaaagta aaagaccttg gctcatagaa agtcacttta
atagataggg
      2881 acagtaataa ataaatgtac aatctctata tacaaaaaaa aaaaaaaaaa aa
```

Protein sequence:

NCBI Reference Sequence: NP_004981.2

LOCUS NP_004981

ACCESSION NP_004981

```
      1 mrlfvsdgvp gclpvlaaag rargraevli stvgpedcvv pfltrpkvpv
lqldsgnylf
       61 stsaicryff llsgweqddl tnqwleweat elqpalsaal yylvvqgkkg
edvlgsvrra
      121 lthidhslsr qncpflaget esladivlwg alypllqdpa ylpeelsalh
swfqtlstqe
      181 pcqraaetvl kqqgvlalrp ylqkqpqpsp aegravtnep eeeelatlse
eeiamavtaw
      241 ekgleslppl rpqqnpvlpv agernvlits alpyvnnvph lgniigcvls
advfarysrl
      301 rqwntlylcg tdeygtatet kaleegltpq eicdkyhiih adiyrwfnis
fdifgrtttp
      361 qqtkitqdif qqllkrgfvl qdtveqlrce hcarfladrf vegvcpfcgy
eeargdqcdk
      421 cgklinavel kkpqckvcrs cpvvqssqhl fldlpklekr leewlgrtlp
gsdwtpnaqf
      481 itrswlrdgl kprcitrdlk wgtpvplegf edkvfyvwfd atigylsita
nytdqwerww
      541 knpeqvdlyq fmakdnvpfh slvfpcsalg aednytlvsh liateylnye
dgkfsksrgv
      601 gvfgdmaqdt gipadiwrfy llyirpegqd safswtdlll knnsellnnl
gnfinragmf
      661 vskffggyvp emvltpddqr llahvtlelq hyhqllekvr irdalrsilt
isrhgnqyiq
      721 vnepwkrikg seadrqragt vtglavniaa llsvmlqpym ptvsatiqaq
lqlpppacsi
      781 lltnflctlp aghqigtvsp lfqklendqi eslrqrfggg qaktspkpav
vetvttakpq
      841 qiqalmdevt kqgnivrelk aqkadkneva aevaklldlk kqlavaegkp
peapkgkkkk
```

SENP1

Official Symbol: SENP1

Official Name: SUMO1/sentrin specific peptidase 1

Gene ID: 29843

Organism: Homo sapiens

Other Aliases: SuPr-2

Other Designations: SUMO1/sentrin specific protease 1; sentrin-specific protease 1; sentrin/SUMO-specific protease SENP1

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001267594.1

LOCUS NM_001267594

ACCESSION NM_001267594

```
      1 attccgagta cgagaaagcg aaaaagccca gactgaaaag ggtactgaga
aattacgact
       61 agtcttaaat gctcccttcg cttctcgggc ctcgccacac cgcgcaggcg
ccccactggt
      121 ccttaactct gttctttgac ctcctgcccc agccccctcc tcttcagcca
cctagcgact
      181 cttccggtgc tgtgaaggcg gttccggttc gcggcggttc ccgggttttg
cgttccgcgc
      241 ccggccggaa acccccttcgc atggcagccg gttccggttc ggactttgta
tctttgctaa
      301 agtcagtgat gtgaaaagac ttgaaatgga tgatattgct gataggatga
ggatggatgc
      361 tggagaagtg actttagtga accacaactc cgtattcaaa acccacctcc
tgccacaaac
      421 aggttttcca gaggaccagc tttcgctttc tgaccagcag attttatctt
ccaggcaagg
      481 acatttggac cgatctttta catgttccac aagaagtgca gcttataatc
caagctatta
      541 ctcagataat ccttcctcag acagttttct tggctcaggc gatttaagaa
cctttggcca
      601 gagtgcaaat ggccaatgga gaaattctac cccatcgtca agctcatctt
tacaaaaatc
      661 aagaaacagc cgaagtcttt acctcgaaac ccgaaagacc tcaagtggat
tatcaaacag
      721 ttttgcggga aagtcaaacc atcactgcca tgtatctgca tatgaaaaat
cttttcctat
      781 taaacctgtt ccaagtccat cttggagtgg ttcatgtcgt cgaagtcttt
tgagccccaa
      841 gaaaactcag aggcgacatg ttagtacagc agaagagaca gttcaagaag
aagaaagaga
      901 gatttacaga cagctgctac agatggtcac agggaaacag tttactatag
ccaaacccac
      961 cacacatttt cctttacacc tgtctcgatg tcttagttcc agtaaaaata
ctttgaaaga
     1021 ctcactgttt aaaaatggaa actcttgtgc atctcagatc attggctctg
atacttcatc
     1081 atctggatct gccagcattt taactaacca ggaacagctg tcccacagtg
tatattccct
     1141 atcttcttat accccagatg ttgcatttgg atccaaagat tctggtactc
ttcatcatcc
     1201 ccatcatcac cactctgttc cacatcagcc agataactta gcagcttcaa
atacacaatc
     1261 tgaaggatca gactctgtga ttttactgaa agtgaaagat tcccagactc
caactcccag
     1321 ttctactttc ttccaggcag agctgtggat caaagaatta actagtgttt
atgattctcg
     1381 agcacgagaa agattgcgcc agattgaaga acagaaggca ttggccttac
agcttcaaaa
     1441 ccagagattg caggagcggg aacattcagt acatgattca gtagaactac
atcttcgtgt
     1501 acctcttgaa aaggagattc ctgttactgt tgtccaagaa acacaaaaaa
aaggtcataa
     1561 attaactgat agtgaagatg aatttcctga aattacagag gaaatggaga
aagaaataaa
     1621 gaatgtattt cgtaatggga atcaggatga agttctcagt gaagcatttc
gcctgaccat
     1681 tacacgcaaa gatattcaaa ctctaaacca tctgaattgg ctcaatgatg
agatcatcaa
```

1741 tttctacatg aatatgctga tggagcgaag taaagagaag ggcttgccaa
gtgtgcatgc
1801 atttaatacc tttttcttca ctaaattaaa aacggctggt tatcaggcag
tgaaacgttg
1861 gacaaagaaa gtagatgtat tttctgttga cattcttttg gtgcccattc
acctgggagt
1921 acactggtgt ctagctgttg tggactttag aaagaagaat attacctatt
acgactccat
1981 gggtgggata aacaatgaag cctgcagaat actcttgcaa tacctaaagc
aagaaagcat
2041 tgacaagaaa aggaaagagt ttgacaccaa tggctggcag cttttcagca
agaaaagcca
2101 ggagattcct cagcagatga atggaagtga ctgtgggatg tttgcctgca
aatatgctga
2161 ctgtattacc aaagacagac caatcaactt cacacagcaa cacatgccat
acttccggaa
2221 gcggatggtc tgggagatcc tccaccgaaa actcttgtga agactgtctc
acttagcaga
2281 ccttgaccat gtgggggacc agctctttgt tgtctacagc cagagacctt
ggaaacagct
2341 gctcccagcc ctctgctgtt gtaacaccct tgatcctgga ccaggccctg
gcgagatgca
2401 ttcacaagca catctgcctt tccttttgta tctcagatac tatttttgca
aagaaacttt
2461 ggtgctgtga aaggggtgag ggacatccct aagctgaaga gagagactgc
ttttcacttc
2521 ttcagttctg ccatcttgtt ttcaaagggc tccagcctca ctcagtccct
aattatggga
2581 ctgagaaaag cttggaaaga atcttggttt catataaatt cttgttgtta
ggccttacta
2641 agaagtagga aagggcatgg gcaaaaggta gggataaaaa ccaccagcat
atacatggac
2701 atacacacac acccacacac acaaacacac acacacacac aattttcacg
atgtatggtc
2761 aggaatgtga ctgtaaactg gactttgggg cccaggcata agtcccttcc
tccaggacct
2821 ttcctattta tatgtcccta tacaaaatcc atctgctttt atacgtagct
gttttatcat
2881 ctgtagcttc atcctatccg gaggcacagc acatgagccc tggacaggtc
ccaaagttcc
2941 aagcagtcct ttccgtgaaa gcaggggttt gcatgtgcta ccaacacatg
atacggggaa
3001 gacccaccca gggagcggtt tcagtggcgc aacaaagcac cacttttact
gttgcctact
3061 tctgaccaag aagaaaaagg accttagtat ttagcataaa attccagcgc
tggatgaatg
3121 cagatctagt ttggtctgtg gctagtttaa atatgtttct aaccacagag
aatttcatat
3181 atatatacat atatatatac acatacatat atatatatat atatgtatgt
ataaaatttc
3241 acagggatat gctttttttt ttaaagactg aatgtgttca ccatttagcc
tgtagattta
3301 tttccatttt ccaaattcca gcacacagag atcccagccc ctatgagtag
ggtgtttgtg
3361 gactacctaa tggaatattt ttgaggcctg gatgaacttt gccatatggg
tagaggttac
3421 agagggaggt gatattttca gctaaaaaaa aaaacgggtg gagtttggac
tgatcaactt
3481 gagatttaaa aactgctatt ccttttgttc tttctagcat ctctccccac
cctctgagag

```
     3541 ctcctcaggc ttagatagtg aagtgatcaa atgccagtgt cattttgtac
ttaagttcca
     3601 aagtaggaac attttatact tttttctgta ttgtaatagg tagttttgta
tgaaatcttt
     3661 tctcctctcc cgttgtaccg cattctttcc agcattgtgc tttttccctg
ggcttatttg
     3721 aaaattttac tgttttatac aagctcgttt agtacatttt tctatgtttt
accacaagtt
     3781 acaatttgaa aagaaaacta tttttttaa atattccatt gttaactgaa
tgttactgtt
     3841 tccactccag caactacatg tcctcccttc aactgcctgc cttttgggga
aagaccacct
     3901 tttgtgtgtt tgtttttttct ctctctttct ttccctttct ctttctatct
ctctttattt
     3961 ttctttcttt ttctttgttt ttgagttttc tataggaaat aaatagcttt
ctatatatga
     4021 gttgctgggg accttcacat tctcttttag aaagctgtgg catgcagtct
cattgcagga
     4081 ctcctggaat attgtctggt tcttggtatt tactgtatgt aagcaacaac
ttgaaaggtg
     4141 gcaatatggt gtcgatttgg actatgaatc aaaagacctt tttcaggttc
tttcactatt
     4201 gtctgggggga ctcagaacaa gattgttctc tgtatttatt gtttgtccat
ttaggtaaca
     4261 tctgtcttac cttcctcaca gactttgtac agaccaaagc aacaaatatt
tattgccatg
     4321 tatagcagaa aatgaaacat gcaacaaaag cactttgaaa aatatataag
gaattgttga
     4381 gcctgtctga atttgggccc cctttctgac taatgcagtt ttgcacaagg
tagaagttag
     4441 tgaccctgag accatcttac caccctggac ctggtccaaa tacagactta
cacagtggac
     4501 cattctttcc tgagctagcc aacaagagca ggagtagtat ctggaaactt
tccccttttgt
     4561 ttaggggtag gctttgatga ccaggaaaaa aaaaaaggta tttctgcatt
ttatggccca
     4621 aaggcatgtt attaatatct tatgtaattt actttaaact aaataagact
tttttctcct
      4681 gtgtaaaaaa aaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001254523.1

LOCUS NP_001254523

ACCESSION NP_001254523

```
      1 mddiadrmrm dagevtlvnh nsvfkthllp qtgfpedqls lsdqqilssr
qghldrsftc
       61 strsaaynps yysdnpssds flgsgdlrtf gqsangqwrn stpsssslq
ksrnsrslyl
      121 etrktssgls nsfagksnhh chvsayeksf pikpvpspsw sgscrrslls
pkktqrrhvs
      181 taeetvqeee reiyrqllqm vtgkqftiak ptthfplhls rclsskntl
kdslfkngns
      241 casqiigsdt sssgsasilt nqeqlshsvy slssytpdva fgskdsgtlh
hphhhhsvph
      301 qpdnlaasnt qsegsdsvil lkvkdsqtpt psstffqael wikeltsvyd
srarerlrqi
```

361 eeqkalalql qnqrlqereh svhdsvelhl rvplekeipv tvvqetqkkg
hkltdsedef
421 peiteemeke iknvfrngnq devlseafrl titrkdiqtl nhlnwlndei
infymnmlme
481 rskekglpsv hafntffftk lktagyqavk rwtkkvdvfs vdillvpihl
gvhwclavvd
541 frkknityyd smgginneac rillqylkqe sidkkrkefd tngwqlfskk
sqeipqqmng
601 sdcgmfacky adcitkdrpi nftqqhmpyf rkrmvweilh rkll

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001267595.1

LOCUS NM_001267595

ACCESSION NM_001267595

1 atggagggag cctggtccag gagactgtgt cggacccggt cagccggccc
gggctggact
61 gggcggaagc ggggagcact gtggggccgg cgccttttcc tctgccccgc
cccctgggac
121 cacctcccct ccccccctgct gtccggtggc cgcgctgtgg ccgccggtgg
ccgttaggct
181 acctgaggcc gttccttctg gtctctctct cctgggccgc ggagagaccg
tctccctgcc
241 gttacagcag gccccatccc agcgcccagc cgtacttggg gaaaggccgg
ttgcgattcc
301 ggggctttcc cgccagagct gggtcttctc tggggagagc tgttttcacc
gggaagctcg
361 gctttctgtg gtaccggctt catctcccgc cttccttgag acccgagtga
tatttcttga
421 ctacttctgc gtctcacgta aacatttctc caactctcct actctgtggt
atctccctga
481 gatgtgatat cgctagtgcc accatcagaa agaaacgtct ggaccctcct
gctcaggact
541 ttgtatcttt gctaaagtca gtgatgtgaa aagacttgaa atggatgata
ttgctgatag
601 gatgaggatg gatgctggag aagtgacttt agtgaaccac aactccgtat
tcaaaaccca
661 cctcctgcca caaacaggtt ttccagagga ccagctttcg ctttctgacc
agcagatttt
721 atcttccagg caaggacatt tggaccgatc ttttacatgt tccacaagaa
gtgcagctta
781 taatccaagc tattactcag ataatccttc ctcagacagt tttcttggct
caggcgattt
841 aagaaccttt ggccagagtg caaatggcca atggagaaat tctaccccat
cgtcaagctc
901 atctttacaa aaatcaagaa acagccgaag tctttacctc gaaacccgaa
agacctcaag
961 tggattatca aacagttttg cgggaaagtc aaaccatcac tgccatgtat
ctgcatatga
1021 aaaatctttt cctattaaac ctgttccaag tccatcttgg agtggttcat
gtcgtcgaag
1081 tcttttgagc cccaagaaaa ctcagaggcg acatgttagt acagcagaag
agacagttca
1141 agaagaagaa agagagattt acagacagct gctacagatg gtcacaggga
aacagtttac
1201 tatagccaaa cccaccacac attttccttt acacctgtct cgatgtctta
gttccagtaa

1261 aaatactttg aaagactcac tgtttaaaaa tggaaactct tgtgcatctc
agatcattgg
1321 ctctgatact tcatcatctg gatctgccag cattttaact aaccaggaac
agctgtccca
1381 cagtgtatat tccctatctt cttatacccc agatgttgca tttggatcca
aagattctgg
1441 tactcttcat catccccatc atcaccactc tgttccacat cagccagata
acttagcagc
1501 ttcaaataca caatctgaag gatcagactc tgtgatttta ctgaaagtga
aagattccca
1561 gactccaact cccagttcta ctttcttcca ggcagagctg tggatcaaag
aattaactag
1621 tgtttatgat tctcgagcac gagaaagatt gcgccagatt gaagaacaga
aggcattggc
1681 cttacagctt caaaaccaga gattgcagga gcgggaacat tcagtacatg
attcagtaga
1741 actacatctt cgtgtacctc ttgaaaagga gattcctgtt actgttgtcc
aagaaacaca
1801 aaaaaaaggt cataaattaa ctgatagtga agatgaattt cctgaaatta
cagaggaaat
1861 ggagaaagaa ataaagaatg tatttcgtaa tgggaatcag gatgaagttc
tcagtgaagc
1921 atttcgcctg accattacac gcaaagatat tcaaactcta aaccatctga
attggctcaa
1981 tgatgagatc atcaatttct acatgaatat gctgatggag cgaagtaaag
agaagggctt
2041 gccaagtgtg catgcattta ataccttttt cttcactaaa ttaaaaacgg
ctggttatca
2101 ggcagtgaaa cgttggacaa agaaagtaga tgtattttct gttgacattc
ttttggtgcc
2161 cattcacctg ggagtacact ggtgtctagc tgttgtggac tttagaaaga
agaatattac
2221 ctattacgac tccatgggtg ggataaacaa tgaagcctgc agaatactct
tgcaatacct
2281 aaagcaagaa agcattgaca agaaaaggaa agagtttgac accaatggct
ggcagctttt
2341 cagcaagaaa agccaggaga ttcctcagca gatgaatgga agtgactgtg
ggatgtttgc
2401 ctgcaaatat gctgactgta ttaccaaaga cagaccaatc aacttcacac
agcaacacat
2461 gccatacttc cggaagcgga tggtctggga gatcctccac cgaaaactct
tgtgaagact
2521 gtctcactta gcagaccttg accatgtggg ggaccagctc tttgttgtct
acagccagag
2581 accttggaaa cagctgctcc cagccctctg ctgttgtaac acccttgatc
ctggaccagg
2641 ccctggcgag atgcattcac aagcacatct gcctttcctt ttgtatctca
gatactattt
2701 ttgcaaagaa actttggtgc tgtgaaaggg gtgagggaca tccctaagct
gaagagagag
2761 actgcttttc acttcttcag ttctgccatc ttgttttcaa agggctccag
cctcactcag
2821 tccctaatta tgggactgag aaaagcttgg aaagaatctt ggtttcatat
aaattcttgt
2881 tgttaggcct tactaagaag taggaaaggg catgggcaaa aggtagggat
aaaaaccacc
2941 agcatataca tggacataca cacacaccca cacacacaaa cacacacaca
cacacaattt
3001 tcacgatgta tggtcaggaa tgtgactgta aactggactt tggggcccag
gcataagtcc

3061 cttcctccag gacctttcct atttatatgt ccctatacaa aatccatctg
cttttatacg
3121 tagctgtttt atcatctgta gcttcatcct atccggaggc acagcacatg
agccctggac
3181 aggtcccaaa gttccaagca gtcctttccg tgaaagcagg ggtttgcatg
tgctaccaac
3241 acatgatacg gggaagaccc acccagggag cggtttcagt ggcgcaacaa
agcaccactt
3301 ttactgttgc ctacttctga ccaagaagaa aaaggacctt agtatttagc
ataaaattcc
3361 agcgctggat gaatgcagat ctagtttggt ctgtggctag tttaaatatg
tttctaacca
3421 cagagaattt catatatata tacatatata tatacacata catatatata
tatatatatg
3481 tatgtataaa atttcacagg gatatgcttt ttttttaaa gactgaatgt
gttcaccatt
3541 tagcctgtag atttatttcc attttccaaa ttccagcaca cagagatccc
agcccctatg
3601 agtagggtgt ttgtggacta cctaatggaa tatttttgag gcctggatga
actttgccat
3661 atgggtagag gttacagagg gaggtgatat tttcagctaa aaaaaaaaac
gggtggagtt
3721 tggactgatc aacttgagat ttaaaaactg ctattccttt tgttctttct
agcatctctc
3781 cccaccctct gagagctcct caggcttaga tagtgaagtg atcaaatgcc
agtgtcattt
3841 tgtacttaag ttccaaagta ggaacatttt atacttttt ctgtattgta
ataggtagtt
3901 ttgtatgaaa tcttttctcc tctcccgttg taccgcattc tttccagcat
tgtgcttttt
3961 ccctgggctt atttgaaaat tttactgttt tatacaagct cgtttagtac
atttttctat
4021 gttttaccac aagttacaat ttgaaaagaa aactattttt tttaaatatt
ccattgttaa
4081 ctgaatgtta ctgtttccac tccagcaact acatgtcctc ccttcaactg
cctgcctttt
4141 ggggaaagac cacctttgt gtgtttgttt tttctctctc tttctttccc
tttctctttc
4201 tatctctctt tatttttctt tctttttctt tgtttttgag ttttctatag
gaaataaata
4261 gctttctata tatgagttgc tggggacctt cacattctct tttagaaagc
tgtggcatgc
4321 agtctcattg caggactcct ggaatattgt ctggttcttg gtatttactg
tatgtaagca
4381 acaacttgaa aggtggcaat atggtgtcga tttggactat gaatcaaaag
acctttttca
4441 ggttctttca ctattgtctg ggggactcag aacaagattg ttctctgtat
ttattgtttg
4501 tccatttagg taacatctgt cttaccttcc tcacagactt tgtacagacc
aaagcaacaa
4561 atatttattg ccatgtatag cagaaaatga aacatgcaac aaaagcactt
tgaaaaatat
4621 ataaggaatt gttgagcctg tctgaatttg ggcccccttt ctgactaatg
cagttttgca
4681 caaggtagaa gttagtgacc ctgagaccat cttaccaccc tggacctggt
ccaaatacag
4741 acttacacag tggaccattc tttcctgagc tagccaacaa gagcaggagt
agtatctgga
4801 aactttcccc tttgtttagg ggtaggcttt gatgaccagg aaaaaaaaaa
aggtatttct

```
     4861 gcattttatg gcccaaaggc atgttattaa tatcttatgt aatttacttt
aaactaaata
      4921 agacttttt ctcctgtgta aaaaaaaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001254524.1

LOCUS NP_001254524

ACCESSION NP_001254524

```
      1 mddiadrmrm dagevtlvnh nsvfkthllp qtgfpedqls lsdqqilssr
qghldrsftc
       61 strsaaynps yysdnpssds flgsgdlrtf gqsangqwrn stpsssssq
ksrnsrslyl
      121 etrktssgls nsfagksnhh chvsayeksf pikpvpspsw sgscrrslls
pkktqrrhvs
      181 taeetvqeee reiyrqllqm vtgkqftiak ptthfplhls rclsssknt
kdslfkngns
      241 casqiigsdt sssgsasilt nqeqlshsvy slssytpdva fgskdsgtlh
hphhhhsvph
      301 qpdnlaasnt qsegsdsvil lkvkdsqtpt psstffqael wikeltsvyd
srarerlrqi
      361 eeqkalalql qnqrlqereh svhdsvelhl rvplekeipv tvvqetqkkg
hkltdsedef
      421 peiteemeke iknvfrngnq devlseafrl titrkdiqtl nhlnwlndei
infymnmlme
      481 rskekglpsv hafntffftk lktagyqavk rwtkkvdvfs vdillvpihl
gvhwclavvd
      541 frkknityyd smgginneac rillqylkqe sidkkrkefd tngwqlfskk
sqeipqqmng
      601 sdcgmfacky adcitkdrpi nftqqhmpyf rkrmvweilh rkll
```

ATPIF1

Official Symbol: ATPIF1

Official Name: ATPase inhibitory factor 1

Gene ID: 93974

Organism: Homo sapiens

Other Aliases: RP5-1092A3.1, ATPI, ATPIP, IP

Other Designations: ATP synthase inhibitor protein; ATPase inhibitor protein; ATPase inhibitor, mitochondrial; IF(1); IF1; inhibitor of F(1)F(o)-ATPase

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_016311.4

LOCUS NM_016311

ACCESSION NM_016311

```
      1 gaccagattg ggtgcttggc cgtccctgcc attagcgcgt aacgagagac
tgcttgctgc
       61 ggcagagacg ccagaggtgc agctccagca gcaatggcag tgacggcgtt
ggcggcgcgg
      121 acgtggcttg gcgtgtgggg cgtgaggacc atgcaagccc gaggcttcgg
ctcggatcag
      181 tccgagaatg tcgaccgggg cgcgggctcc atccgggaag ccggtggggc
cttcggaaag
      241 agagagcagg ctgaagagga acgatatttc cgagcacaga gtagagaaca
actggcagct
      301 ttgaaaaaac accatgaaga agaaatcgtt catcataaga aggagattga
gcgtctgcag
      361 aaagaaattg agcgccataa gcagaagatc aaaatgctaa aacatgatga
ttaagtgcac
      421 accgtgtgcc atagaatggc acatgtcatt gcccacttct gtgtagacat
ggttctggtt
      481 taactaatat ttgtctgtgt gctactaaca gattataata aattgtcatc
agtgaactgt
      541 gaaaaaaaaa aaaaaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_057395.1

LOCUS NP_057395

ACCESSION NP_057395

```
      1 mavtalaart wlgvwgvrtm qargfgsdqs envdrgagsi reaggafgkr
eqaeeeryfr
      61 aqsreqlaal kkhheeeivh hkkeierlqk eierhkqkik mlkhdd
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_178190.2

LOCUS NM_178190

ACCESSION NM_178190

```
      1 gaccagattg ggtgcttggc cgtccctgcc attagcgcgt aacgagagac
tgcttgctgc
       61 ggcagagacg ccagaggtgc agctccagca gcaatggcag tgacggcgtt
ggcggcgcgg
      121 acgtggcttg gcgtgtgggg cgtgaggacc atgcaagccc gaggcttcgg
ctcggatcag
      181 tccgagaatg tcgaccgggg cgcgggctcc atccgggaag ccggtggggc
cttcggaaag
      241 agagagcagg ctgaagagga acgatatttc cgacattaca ggttatgctt
tgagatctct
      301 ttggggtgaa ggattgaaat taaaccctga gccaccgtgt ccttgtagag
cacagagtag
      361 agaacaactg gcagctttga aaaaacacca tgaagaagaa atcgttcatc
ataagaagga
      421 gattgagcgt ctgcagaaag aaattgagcg ccataagcag aagatcaaaa
tgctaaaaca
```

```
      481 tgatgattaa gtgcacaccg tgtgccatag aatggcacat gtcattgccc
acttctgtgt
      541 agacatggtt ctggtttaac taatatttgt ctgtgtgcta ctaacagatt
ataataaatt
      601 gtcatcagtg aactgtgaaa aaaaaaaaaa aaaaaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_835497.1

LOCUS NP_835497

ACCESSION NP_835497

```
      1 mavtalaart wlgvwgvrtm qargfgsdqs envdrgagsi reaggafgkr
eqaeeeryfr
      61 hyrlcfeisl g
```

Nucleotide sequence (variant 3)

NCBI Reference Sequence: NM_178191.2

LOCUS NM_178191

ACCESSION NM_178191

```
      1 gaccagattg ggtgcttggc cgtccctgcc attagcgcgt aacgagagac
tgcttgctgc
       61 ggcagagacg ccagaggtgc agctccagca gcaatggcag tgacggcgtt
ggcggcgcgg
      121 acgtggcttg gcgtgtgggg cgtgaggacc atgcaagccc gaggcttcgg
ctcggatcag
      181 tccgagaatg tcgaccgggg cgcgggctcc atccgggaag ccggtggggc
cttcggaaag
      241 agagagcagg ctgaagagga acgatatttc cggtgaggct caccgggtcc
caagtccagc
      301 cctggatctc ccaatggcct tccaatcctt aaactgccaa tcgccccacc
cgttcctacc
      361 tggtgccttg ggcgccccat cccccaacag aactcccggg ccccaatcca
gtataccta
      421 acccttgatg tcccgaccgt tgccacgtat agggcactcc cagttacctg
cacaacagtt
      481 tcaggccccc aaaccgtttc caccggcggg tctccaaaac aacccacggc
tcaactcctc
      541 ctttatcatt accatctccc gcgtggagtt ctcctcaggt cgtgcgaaac
acccccagat
      601 tcttcgcaca gtgtctagat ccgaccgccc aacgtttgcc tcccagcctg
actccctcgg
      661 cccttaccca cctgtcaccc cctctacgct ctccttcctc gccagcacgc
cttagctttg
      721 caagcctgca tgcattcagg cttctcaggt gtttctagac ccccgactcc
gcaagagtga
      781 ggatgatggg agctggtcat gggagctact tatggttgga caccatcttc
taaaggcttt
      841 tgccctactc agcccaacct agacctgtag atttccctct cctgcttagg
agtatggagt
```

```
      901 gggctgggcc tccctttgcc agccttgagt tatctttaac tgacttctgt
ccactctgga
      961 gagcagtgag gaattaatct tgcttttgct tgtcctttgg cctttcactt
ctgccttctg
     1021 ttgagaatta tcaccatgac acctgccata ccgtatagag agccaaggta
cagccgttag
     1081 agactatcta attgagcccc tacattttgt agttaaggaa aactgaggcc
taaatgtgac
     1141 caaaccaaca ttgtaatcca gtcccttctt ggaacctaaa ttgaactgcc
aagtactgcg
     1201 catgcaagag acccttttatt ggccttacag tgggccattc atttctatag
gcaaagaaag
     1261 ctctagacag attggaatag gaaatggata tttgcctttt agctacaccc
ctttgtctgt
     1321 cttcctcatt ttgttccttt ttttttccct aaaggggagt caagttccct
gggttgttcc
     1381 cctcataagg tattagggac ttgtgtcaca tctctctgga gttttctatt
ttaaagagga
     1441 atctgaaagc aataagctct ttggtcttct taagatggct acacctcaat
ttaagatggg
     1501 gtattctttc actagttgag gagtagaaga ggatgaccag ctagactccc
atggaattgg
     1561 aactcctatt ccttgcttag acattacagg ttatgctttg agatctcttt
ggggtgaagg
     1621 attgaaatta aaccctgagc caccgtgtcc ttgtagagca cagagtagag
aacaactggc
     1681 agctttgaaa aaacaccatg aagaagaaat cgttcatcat aagaaggaga
ttgagcgtct
     1741 gcagaaagaa attgagcgcc ataagcagaa gatcaaaatg ctaaaacatg
atgattaagt
     1801 gcacaccgtg tgccatagaa tggcacatgt cattgcccac ttctgtgtag
acatggttct
     1861 ggtttaacta atatttgtct gtgtgctact aacagattat aataaattgt
catcagtgaa
     1921 ctgtgaaaaa aaaaaaaaaa aaaa
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_835498.1

LOCUS NP_835498

ACCESSION NP_835498

1 mavtalaart wlgvwgvrtm qargfgsdqs envdrgagsi reaggafgkr eqaeeeryfr

VAMP3

Official Symbol: VAMP3

Official Name: vesicle-associated membrane protein 3 (cellubrevin)

Gene ID: 9341

Organism: Homo sapiens

Other Aliases: CEB

Other Designations: VAMP-3; cellubrevin; synaptobrevin-3; vesicle-associated membrane protein 3

Nucleotide sequence:

NCBI Reference Sequence: NM_004781.3

LOCUS NM_004781

ACCESSION NM_004781

```
      1 agtgacgtct ttgccccgcg ccgcgccgtc ccacccatct ccctggcctc
cggtcccaac
       61 ttcgcttctc tgctgaccct ctctcgtcgc cgctgccgcc gccgcagctg
ccaaaatgtc
      121 tacaggtcca actgctgcca ctggcagtaa tcgaagactt cagcagacac
aaaatcaagt
      181 agatgaggtg gtggacataa tgcgagttaa cgtggacaag gttctggaaa
gagaccagaa
      241 gctctctgag ttagacgacc gtgcagacgc actgcaggca ggcgcttctc
aatttgaaac
      301 gagcgcagcc aagttgaaga ggaaatattg gtggaagaat tgcaagatgt
gggcaatcgg
      361 gattactgtt ctggttatct tcatcatcat catcatcgtg tgggttgtct
cttcatgaag
      421 aaccagcgga actcaaaact gctgttcaag aaacctcttc aagacttttg
acttagaacc
      481 tgctatatta tcaagcttac ctactgttat ctctaaaatt ttttttgtgt
taatgtaaag
      541 ttgaatttct aggaaacgtg cctttgtttt ttaatatgca ctccaaatta
gaaggccggc
      601 cccgtccaca ttttgcacag tgcctttaca gatttacgta tgggctgatg
aagaggcctt
      661 cttaagttcc agagtgctat aatctagatg taatgttgtc actaattaat
tgccattact
      721 cccagttagt taccccttgtc atttggcatt attttcagaa ccacatttta
aacctttggg
      781 taatcagatt tccaacttat gccttccaga aaaaaacact actgcctaac
acaaatctgt
      841 gataacaaca ggctgtgcct tattttgata attttctgat tccctagaag
agaaccctct
      901 acttttttgta agcactactg actctcgctg tatttaagat gctggtgaag
agcttttgct
      961 cttgcattag atttgaagat gtttacattg ttgttattgt tatgtatcac
ttgctaaaaa
     1021 tattgtttta atcagagata acctctttaa aaaaattttt aaagaactat
ggctatgacc
     1081 aaagcttcta ttttgccaaa aagttaaata ccgataaaat ggccttaagt
gtattcctga
     1141 cagttaaatt cagaaacgtg ccaaatggaa ctcaaggtgc cccttcagaa
ttaaaatcat
     1201 taccttgtgt gtgaaccttc tacatcttca taggcctttc ttccttttga
aaggctgtag
     1261 acagtgtggc tcccccttctg attcagtatt ttgcatgggg gttagagaag
gtttgaggta
     1321 gactctgacc gtctcataaa agagttctac ccagcagttg gcagattatc
agctgtggac
     1381 tccagcatgt ttctgataat tatgcaagca acaattctgt agcctcaagt
aagaccacct
```

```
     1441 gtgaacttga tcattatctg gcccaaatat gaagataaac tataactttg
gagtttgttt
     1501 cctatttgta ttcacattct gcttcctaaa tcagttttct aaattatgcc
tgcaattagg
     1561 cattggtcag gggtgaatgg ctcttttcac agagagtagc caaccagaga
cctttgcttt
     1621 gatatcatca actgcagaga atgctgttga tgggaatgct ggaagcagaa
actttgtcat
     1681 cggaaaaact tttcttgtat gcatgagact caacatcagg atccacagct
taaagatggg
     1741 aattcaggta tgaaagaaaa caggcaagga ggcactgagg gagaaagaca
cagactttat
     1801 cgctctgtgg ctcattgtta ctggaatatt ctaaaactct tgttcacatg
ctattatgac
     1861 ttataaagca gcaacagctg aggcgcacca ggacacagct tccatttctt
taacgtctgt
     1921 tcccttaaca tcgctgaaat gatttactgt tgaagagatg ccttgcggtg
tggccagctg
     1981 tgaggagaaa gcagctggca gtgttaggac attagtccac cttcagcgca
gggtctctgg
     2041 ccgggtctga ctcagaaacc ttggtactcg ccccttggcc acagtgccca
gacccatgta
     2101 acccactggc tcctgcatta acccagaaat acctcgcttc tatctgtgca
cttagctggg
     2161 aacttaccca ctgtaatcac ctaaataaag tgtttataaa catgaaaaaa
aaaaaaaaaa
      2221 aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_004772.1

LOCUS NP_004772

ACCESSION NP_004772

```
      1 mstgptaatg snrrlqqtqn qvdevvdimr vnvdkvlerd qklselddra
dalqagasqf
      61 etsaaklkrk ywwknckmwa igitvlvifi iiiivwvvss
```

VAPA

Official Symbol: VAPA

Official Name: VAMP (vesicle-associated membrane protein)-associated protein A, 33kDa

Gene ID: 9218

Organism: Homo sapiens

Other Aliases: VAP-33, VAP-A, VAP33, hVAP-33

Other Designations: 33 kDa VAMP-associated protein; VAMP-A; VAMP-associated protein A; vesicle-associated membrane protein-associated protein A

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_003574.5

LOCUS NM_003574

ACCESSION NM_003574

```
      1 agaagctccc ggccgggggc gcgcacgtag gcacgcagag gccgtcacgt
gggtcgccga
       61 ggctcgcaag tgcgcgtggc cgtggcggct ggtgtggggt tgagtcagtt
gtgggacccg
      121 gagctgctga cccagcgggt ggcccaccga accggtgaca cagcggcagg
cgttagggct
      181 cgggagccgc gagcctggcc tcgtcctaga gctcggccga gccgtcgccg
ccgtcgtccc
      241 ccgcccccag tcagcaaacc gccgccgcgg gcgcgccccc gctctgcgct
gtctctccga
      301 tggcgtccgc ctcaggggcc atggcgaagc acgagcagat cctggtcctc
gatccgccca
      361 cagacctcaa attcaaaggc cccttcacag atgtagtcac tacaaatctt
aaattgcgaa
      421 atccatcgga tagaaaagtg tgtttcaaag tgaagactac agcacctcgc
cggtactgtg
      481 tgaggcccaa cagtggaatt attgacccag ggtcaactgt gactgtttca
gtaatgctac
      541 agccctttga ctatgatccg aatgaaaaga gtaaacacaa gtttatggta
cagacaattt
      601 ttgctccacc aaacacttca gatatggaag ctgtgtggaa agaggcaaaa
cctgatgaat
      661 taatggattc caaattgaga tgcgtatttg aaatgcccaa tgaaaatgat
aaattgggta
      721 taactccacc agggaatgct ccgactgtca cttcaatgag cagcatcaac
aacacagttg
      781 caacacctgc cagttatcac acgaaggatg accccagggg actcagtgtg
ttgaaacagg
      841 agaaacagaa gaatgatatg gaacctagca aagctgttcc actgaatgca
tctaagcaag
      901 atggacctat gccaaaacca cacagtgttt cacttaatga taccgaaaca
aggaaactaa
      961 tggaagagtg taaaagactt cagggagaaa tgatgaagct atcagaagaa
aatcggcacc
     1021 tgagagatga aggtttaagg ctcagaaagg tagcacattc ggataaacct
ggatcaacct
     1081 caactgcatc cttcagagat aatgtcacca gtcctcttcc ttcacttctt
gttgtaattg
     1141 cagccatttt cattggattc tttctaggga aattcatctt gtagagtgaa
gcatgcagag
     1201 tgctgtttct ttttttttt ttctcttgac cagaaaaaga tttgtttacc
taccatttca
     1261 ttggtagtat ggcccacggt gaccattttt ttgtgtgtac agcgtcatat
aggctttgcc
     1321 tttaatgatc tcttacggtt agaaaacaca ataaaaacaa actgttcggc
tactggacag
     1381 gttgtatatt accagatcat cactagcaga tgtcagttgc acattgagtc
ctttatgaaa
     1441 ttcataaata aagaattgtt ctttctttgt ggttttaata agagttcaag
aattgttcag
     1501 agtcttgtaa atgttatttt aataatccct ttaaatttta tctgttgctg
ttacctcttg
     1561 aaatatgatt tatttagatt gctaatccca ctcattcagg aaatgccaag
aggtattcct
```

1621 tggggaaatg gtgcctctta cagtgtaaat ttttcctcct ttacctttgc taatatcatg
1681 gcagaatttt tcttatccct tgtgaggcag ttgttgactg agtttttcat ccttacaatc
1741 ctgtcccatg gtatttaaca taaaaaaaaa taaaactgtt aacagattct tgctcgatag
1801 cttgtttgtg tctgtcgtgt tattagaggg aactccacta tatatggtca cttgaaatta
1861 tgatgcaaag gtttctcttg cattgaaacc ctcttggata ttacagtatt tttaattgaa
1921 agtcctaatt ctgttaagga aaggagttga ttaaatttta aggtaccact ggtattttgg
1981 gagattataa tcagtttgtt ttcaagataa tagaaaataa ggtccatgag aatagaagtt
2041 atgtgatttc agtgagttga tgtgtacagc atggctgtgc tccatctgat ttaccccatt
2101 cttaagttct gagagtatgt tctcaaggaa gatttaactc tctttggttt taaattactt
2161 tttaaccagc ctaataaata agtcttacta cttttcataa tatttcataa tagttaaaag
2221 taggtgtttt tttcgtgctc aatttggcac tcaaaataat gttcattatg gaagtttggt
2281 aatactgagc aagcctgtgg aattttcttt atgaaaaatg attttagcct ttgcaaatgt
2341 taaccatgtg aaacacattt tcagtataag tatgcgttac agggtttgat actttcctgc
2401 acttaggttt gtcctattct tcatttattc atactaggat agaaaatttt ggaatcagaa
2461 aatagatcca gtgtttagct acatacaatc tagtacaagt gaatttttat tcttaaacat
2521 aggtgtgttg gctctttttt taaaagatgc gctctacctg aaaaggaaat tggattttag
2581 aactggatgt ggtgcagtga agtattttag gcccaggtct gtgtacacat tttatagaag
2641 aatgaagtac tctgaagtat tttggttgcc ttttcatttc aactgtgttt tgaatttgtc
2701 agatcacaca tatattgtgt tattgggcgc tgtggtatct tttataaaac ctcttgcttg
2761 tgtgcaaaag ttcctaaaag gaaacacaag taatgcctat ccattactag catgctatgc
2821 tgcatgcttt actgccattg ctgtatgctt tactgtcttt gtaaaaatcc ccctctcccc
2881 ttttctggta actggaaaag catgctaaaa atagtcttat attttcaccc cataaatgca
2941 gaatcagtaa ttccttggct taaagctctt atataatcaa tattattggt ggtaaatacc
3001 aagtttggta tctcatagct atctttttt aaagaaatta agttcttgaa aatttagcca
3061 aatcccgttt tatgggaatg ctctttagaa ttcattttgt tcagcccctt tgttctatgg
3121 ttgagaaatc tgaggcctta cgaaggttaa gagaactttc cccgtgtctc acaggtaggt
3181 agaggcagag ctggaactag atatctggtc tgttgactct agctcagtgt cttctggtaa
3241 ctgttgaaaa ttgtcttagt ttgagagatg gctgaaataa tgaacataaa atgctattta
3301 taataacaag tatatgtgaa atttcttatt gtaagactac taccggctta ctgttgaata
3361 gtttggttat agtgtttagg ctagaaatgc ctcccacatt ggtaataaac attacaaaat

3421 acaatgtatt tttaggtagg cattttataa aatgcattat gccatggttg
cttttgagat
3481 agattgtagt ctgggtagca tctttaaaat gtatgtgggc ttaactgttg
ttcatatcag
3541 gagatgctct gattgtatag gtgagactct gtttctgtta tttttaattg
ctgtatgaaa
3601 tgtgatcaga ttattttact accaacagtt atagtttgaa agtccaactg
tattaattga
3661 ctgataatat gataatatag agattaaatt gtttgtcttc attccttata
tgtttagaag
3721 tttttgcttt gtctgcctgc ttacttgtat atgtaagcat gagggaaata
cactgttgct
3781 aatactgaaa ttacaatcaa gtaactaagg ccttgagttc atatgtgaca
ctgaatgcac
3841 tagcttcctt cgttctataa ctaatgtacc ttaacttccc ccattcttat
atttacaaga
3901 agctaagtca ttatgttctg agtgtgtggt atgttccctt aaaaaaaaat
gacacttgga
3961 agaaaaatgt atgaaattca gaaattccga tcaaagaaaa gtaattcttt
ctttttttt
4021 ttgagacaga gtcttgcttt gttgcccagg ctggagggca gtggtgtgat
ctcacctcac
4081 tgcagcttcc gcctcctggg ttcaagtgat tctcatggct cagccgcctg
agtagctggg
4141 attacaggtg tgagccaaca agcccggcta atttttgtat ttttagtaga
gacaaggttt
4201 caccatgttg gtcaggctgg gctcaaactc ctgaatccgc ctgcctcggc
ctcccaaagt
4261 gctgggatta caggtgtgag ctgccgcacc cagccaagaa aaataatact
cttaaatact
4321 tagatgttca cctaaagttg atattatttg gtatgggaat tacttttgaa
ctgtaatctt
4381 tcagattaca ccactttgaa aacaagtttt aacagtaggg taaaaatata
gtttttgagg
4441 gtattcccaa cttgtgatct tctaccactt tagagacatt caagtaatag
ttttcttaga
4501 gctttgcaca ttcctattca ctgagatttt aaaaatttca cctttattcg
agggaaggat
4561 caatgcttat taccatttgg aaaaacgaag atcagaaggt aaatgatctt
tattttctag
4621 ctttaaaggg aaattaaacc attcatgaat aaactttaaa aatgtgaagt
gtccttttcc
4681 ttttcacaat acaaaaaaaa tttcaacaga ttgtgtggtt tgtgcattta
tatcctgtta
4741 agcattaata gctaatcact gggacttgaa ttctgatggc agatagtctc
ttgcttagtg
4801 agatggagtt aactattttt tagtaggaag tgagaacagc tgattttcat
gccacgtttc
4861 atagccccac ttttggtaga ctaccaccac gcttcttcgc gtaagcagtg
gcatcttggg
4921 aatgaatgcc cagccgctcg tgggttggtg caaagaagta taaacatata
tcactaagga
4981 aaaagaaagt ttgtcttgcc cttctgacac agtgtgtgca cttcaggcaa
tttttggaaa
5041 atataaaaaa ttccaaattc tgcctttcag cagcatcaat tgctaggaac
atttcattca
5101 tttccctgta atattaatgt tctttaagca taatcactaa ttataagttg
tatcctattt
5161 ttttccagct taatttctgt ggtttattga aaaccaagta taaatgtgac
taaaagcatt

5221 ttgctttgtt tttatagtta actttcttaa ggttatggac attttataat
gtaacatttg
5281 attggcctgg cctcttgaca attcccttct agttatgcat atcctcctgt
tgccacattt
5341 cttgttttaa aactcagttt cttgttttcc agttgttgct atgtataaca
cccatcttga
5401 aagagagtat ataggaagtt attcagataa cttttgtagt agtgatattc
aactatagca
5461 gtaccttaac tcatgatgag cttaggaaca taaaagataa ttgttgcttg
aatagcaccc
5521 ccagagatac tgacctaatt ggtctggggt ggagatctgg catggtagtt
tttttcaagc
5581 tccaatcatc ggccagacag ttgctttatg taggttttta aatgccaaag
gcagatatga
5641 agtagattta attaagactt gacttcagca atacagggga acttaaaata
cttatttttc
5701 tttaaactgc aggagtcact gttaggtatt gcttaaaaaa aattgcataa
aagctttgct
5761 tgtcaagtta ggattgctgg aataccacta aagatttttg acttgtgaat
aaatgagctg
5821 tcatcgcaaa aaggcgattt gagaaatgtg ggcttcagta ttaattgcca
ttttgctgac
5881 acccagtgta cctacctacc tgagaaattt attttgtcca tcatgtattt
ctcaaagcaa
5941 aaggtggttt tcaagtataa tgtcgttttc aacatgctta ttacttagtt
ttacgtcagc
6001 tcatttcatc atcattgata acttgtgaaa tacttatctc catcctatgg
aataggggag
6061 acgggtttag acaggttcaa ttagctcaag tctacacagc tgaagtagca
gagaaagtgg
6121 gatctagatg gtctgatcct agtgatctac catatgaagg acatagtttg
tgtcctggtc
6181 caagtcaaat attgactcct cacaaacagt aagtatggca attttgtgat
gcctttgatt
6241 ccactttaca tggagtacta ttatttgtga aatgtcttta agatttttgg
tcttaaattt
6301 ttgaagactg ctttcccccc ttatctccca gaaaattgag aagaagtaaa
ctcctgccca
6361 ctaacaatct cagtccgtga acaaaaccaa catgaacatt cctaaacaag
agtgtgtgtt
6421 actctaagaa gaaggctata gaatttatgg aaatggctta tgtaacctac
aagactggag
6481 aacagaatgt gactggcctt ttctaatggt cctttaagat ttaatgatta
aagcaagagt
6541 tttttataat tgactttgtg gtctaaattc ttgatactgt ttataattct
acaaagaaca
6601 aaaattgtta tgtactatag gcacttaaga accctgagga aaaataatac
aatgtgtgtg
6661 tgtgagagag agagtgagtt actgacattg ttccaaaaaa aaaaaaaaaa
aaaaaaaaaa
6721 tgtggagggt tgaaatggta aggaattgga atcttttgta ttttcgagca
ataagaattc
6781 ctattcttgt ttcaaataga ggtttgttag gaattacagt tgtggggagc
aaactttctt
6841 ttttgtgctg ttttaattca aaatgtatat ccttaattgt atataatatg
tagataaata
6901 tatgagggta ttaagctact ttgaattaaa tttaaggata tatttcacat
gaaaacaaat
6961 acaaacgaga atcaaaataa agttttgcaa agta

Protein sequence (variant 1):

NCBI Reference Sequence: NP_003565.4

LOCUS NP_003565

ACCESSION NP_003565

```
      1 masasgamak heqilvldpp tdlkfkgpft dvvttnlklr npsdrkvcfk
vkttaprryc
       61 vrpnsgiidp gstvtvsvml qpfdydpnek skhkfmvqti fappntsdme
avwkeakpde
      121 lmdsklrcvf empnendklg itppgnaptv tsmssinntv atpasyhtkd
dprglsvlkq
      181 ekqkndmeps kavplnaskq dgpmpkphsv slndtetrkl meeckrlqge
mmklseenrh
      241 lrdeglrlrk vahsdkpgst stasfrdnvt splpsllvvi aaifigfflg kfil
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_194434.2

LOCUS NM_194434

ACCESSION NM_194434

```
      1 agaagctccc ggccgggggc gcgcacgtag gcacgcagag gccgtcacgt
gggtcgccga
       61 ggctcgcaag tgcgcgtggc cgtggcggct ggtgtggggt tgagtcagtt
gtgggacccg
      121 gagctgctga cccagcgggt ggcccaccga accggtgaca cagcggcagg
cgttagggct
      181 cgggagccgc gagcctggcc tcgtcctaga gctcggccga gccgtcgccg
ccgtcgtccc
      241 ccgcccccag tcagcaaacc gccgccgcgg gcgcgccccc gctctgcgct
gtctctccga
      301 tggcgtccgc ctcaggggcc atggcgaagc acgagcagat cctggtcctc
gatccgccca
      361 cagacctcaa attcaaaggc cccttcacag atgtagtcac tacaaatctt
aaattgcgaa
      421 atccatcgga tagaaaagtg tgtttcaaag tgaagactac agcacctcgc
cggtactgtg
      481 tgaggcccaa cagtggaatt attgacccag ggtcaactgt gactgtttca
gtaatgctac
      541 agccctttga ctatgatccg aatgaaaaga gtaaacacaa gtttatggta
cagacaattt
      601 ttgctccacc aaacacttca gatatggaag ctgtgtggaa agaggcaaaa
cctgatgaat
      661 taatggattc caaattgaga tgcgtatttg aaatgcccaa tgaaaatgat
aaattgaatg
      721 atatggaacc tagcaaagct gttccactga atgcatctaa gcaagatgga
cctatgccaa
      781 aaccacacag tgtttcactt aatgataccg aaacaaggaa actaatggaa
gagtgtaaaa
      841 gacttcaggg agaaatgatg aagctatcag aagaaaatcg gcacctgaga
gatgaaggtt
      901 taaggctcag aaaggtagca cattcggata aacctggatc aacctcaact
gcatccttca
```

961 gagataatgt caccagtcct cttccttcac ttcttgttgt aattgcagcc
attttcattg
1021 gattctttct agggaaattc atcttgtaga gtgaagcatg cagagtgctg
tttctttttt
1081 tttttttctc ttgaccagaa aaagatttgt ttacctacca tttcattggt
agtatggccc
1141 acggtgacca ttttttttgtg tgtacagcgt catataggct ttgcctttaa
tgatctctta
1201 cggttagaaa acacaataaa aacaaactgt tcggctactg gacaggttgt
atattaccag
1261 atcatcacta gcagatgtca gttgcacatt gagtccttta tgaaattcat
aaataaagaa
1321 ttgttctttc tttgtggttt taataagagt tcaagaattg ttcagagtct
tgtaaatgtt
1381 attttaataa tccctttaaa ttttatctgt tgctgttacc tcttgaaata
tgatttattt
1441 agattgctaa tcccactcat tcaggaaatg ccaagaggta ttccttgggg
aaatggtgcc
1501 tcttacagtg taaatttttc ctcctttacc tttgctaata tcatggcaga
atttttctta
1561 tcccttgtga ggcagttgtt gactgagttt ttcatcctta caatcctgtc
ccatggtatt
1621 taacataaaa aaaaataaaa ctgttaacag attcttgctc gatagcttgt
ttgtgtctgt
1681 cgtgttatta gagggaactc cactatatat ggtcacttga aattatgatg
caaaggtttc
1741 tcttgcattg aaaccctctt ggatattaca gtatttttaa ttgaaagtcc
taattctgtt
1801 aaggaaagga gttgattaaa ttttaaggta ccactggtat tttgggagat
tataatcagt
1861 ttgttttcaa gataatagaa aataaggtcc atgagaatag aagttatgtg
atttcagtga
1921 gttgatgtgt acagcatggc tgtgctccat ctgatttacc ccattcttaa
gttctgagag
1981 tatgttctca aggaagattt aactctcttt ggttttaaat tacttttaa
ccagcctaat
2041 aaataagtct tactactttt cataatattt cataatagtt aaaagtaggt
gtttttttcg
2101 tgctcaattt ggcactcaaa ataatgttca ttatggaagt ttggtaatac
tgagcaagcc
2161 tgtggaattt tctttatgaa aaatgatttt agcctttgca aatgttaacc
atgtgaaaca
2221 cattttcagt ataagtatgc gttacagggt ttgatacttt cctgcactta
ggtttgtcct
2281 attcttcatt tattcatact aggatagaaa attttggaat cagaaaatag
atccagtgtt
2341 tagctacata caatctagta caagtgaatt tttattctta aacataggtg
tgttggctct
2401 ttttttaaaa gatgcgctct acctgaaaag gaaattggat tttagaactg
gatgtggtgc
2461 agtgaagtat tttaggccca ggtctgtgta cacattttat agaagaatga
agtactctga
2521 agtattttgg ttgccttttc atttcaactg tgttttgaat ttgtcagatc
acacatatat
2581 tgtgttattg ggcgctgtgg tatcttttat aaaacctctt gcttgtgtgc
aaaagttcct
2641 aaaaggaaac acaagtaatg cctatccatt actagcatgc tatgctgcat
gctttactgc
2701 cattgctgta tgctttactg tctttgtaaa aatccccctc tccccttttc
tggtaactgg

2761 aaaagcatgc taaaaatagt cttatatttt caccccataa atgcagaatc
agtaattcct
2821 tggcttaaag ctcttatata atcaatatta ttggtggtaa ataccaagtt
tggtatctca
2881 tagctatctt tttttaaaga aattaagttc ttgaaaattt agccaaatcc
cgttttatgg
2941 gaatgctctt tagaattcat tttgttcagc ccctttgttc tatggttgag
aaatctgagg
3001 ccttacgaag gttaagagaa ctttccccgt gtctcacagg taggtagagg
cagagctgga
3061 actagatatc tggtctgttg actctagctc agtgtcttct ggtaactgtt
gaaaattgtc
3121 ttagtttgag agatggctga aataatgaac ataaaatgct atttataata
acaagtatat
3181 gtgaaatttc ttattgtaag actactaccg gcttactgtt gaatagtttg
gttatagtgt
3241 ttaggctaga aatgcctccc acattggtaa taaacattac aaaatacaat
gtatttttag
3301 gtaggcattt tataaaatgc attatgccat ggttgctttt gagatagatt
gtagtctggg
3361 tagcatcttt aaaatgtatg tgggcttaac tgttgttcat atcaggagat
gctctgattg
3421 tataggtgag actctgtttc tgttattttt aattgctgta tgaaatgtga
tcagattatt
3481 ttactaccaa cagttatagt ttgaaagtcc aactgtatta attgactgat
aatatgataa
3541 tatagagatt aaattgtttg tcttcattcc ttatatgttt agaagttttt
gctttgtctg
3601 cctgcttact tgtatatgta agcatgaggg aaatacactg ttgctaatac
tgaaattaca
3661 atcaagtaac taaggccttg agttcatatg tgacactgaa tgcactagct
tccttcgttc
3721 tataactaat gtaccttaac ttcccccatt cttatattta caagaagcta
agtcattatg
3781 ttctgagtgt gtggtatgtt cccttaaaaa aaaatgacac ttggaagaaa
aatgtatgaa
3841 attcagaaat tccgatcaaa gaaaagtaat tctttctttt ttttttgag
acagagtctt
3901 gctttgttgc ccaggctgga gggcagtggt gtgatctcac ctcactgcag
cttccgcctc
3961 ctgggttcaa gtgattctca tggctcagcc gcctgagtag ctgggattac
aggtgtgagc
4021 caacaagccc ggctaatttt tgtattttta gtagagacaa ggtttcacca
tgttggtcag
4081 gctgggctca aactcctgaa tccgcctgcc tcggcctccc aaagtgctgg
gattacaggt
4141 gtgagctgcc gcacccagcc aagaaaaata atactcttaa atacttagat
gttcacctaa
4201 agttgatatt atttggtatg ggaattactt ttgaactgta atctttcaga
ttacaccact
4261 ttgaaaacaa gttttaacag tagggtaaaa atatagtttt tgagggtatt
cccaacttgt
4321 gatcttctac cactttagag acattcaagt aatagttttc ttagagcttt
gcacattcct
4381 attcactgag attttaaaaa tttcaccttt attcgaggga aggatcaatg
cttattacca
4441 tttggaaaaa cgaagatcag aaggtaaatg atctttattt tctagcttta
aagggaaatt
4501 aaaccattca tgaataaact ttaaaaatgt gaagtgtcct tttccttttc
acaatacaaa

4561 aaaaatttca acagattgtg tggtttgtgc atttatatcc tgttaagcat
taatagctaa
4621 tcactgggac ttgaattctg atggcagata gtctcttgct tagtgagatg
gagttaacta
4681 ttttttagta ggaagtgaga acagctgatt ttcatgccac gtttcatagc
cccacttttg
4741 gtagactacc accacgcttc ttcgcgtaag cagtggcatc ttgggaatga
atgcccagcc
4801 gctcgtgggt tggtgcaaag aagtataaac atatatcact aaggaaaaag
aaagtttgtc
4861 ttgcccttct gacacagtgt gtgcacttca ggcaattttt ggaaaatata
aaaaattcca
4921 aattctgcct ttcagcagca tcaattgcta ggaacatttc attcatttcc
ctgtaatatt
4981 aatgttcttt aagcataatc actaattata agttgtatcc tatttttttc
cagcttaatt
5041 tctgtggttt attgaaaacc aagtataaat gtgactaaaa gcattttgct
ttgtttttat
5101 agttaacttt cttaaggtta tggacatttt ataatgtaac atttgattgg
cctggcctct
5161 tgacaattcc cttctagtta tgcatatcct cctgttgcca catttcttgt
tttaaaactc
5221 agtttcttgt tttccagttg ttgctatgta taacacccat cttgaaagag
agtatatagg
5281 aagttattca gataactttt gtagtagtga tattcaacta tagcagtacc
ttaactcatg
5341 atgagcttag gaacataaaa gataattgtt gcttgaatag cacccccaga
gatactgacc
5401 taattggtct ggggtggaga tctggcatgg tagttttttt caagctccaa
tcatcggcca
5461 gacagttgct ttatgtaggt ttttaaatgc caaaggcaga tatgaagtag
atttaattaa
5521 gacttgactt cagcaataca ggggaactta aaatacttat ttttctttaa
actgcaggag
5581 tcactgttag gtattgctta aaaaaaattg cataaaagct ttgcttgtca
agttaggatt
5641 gctggaatac cactaaagat ttttgacttg tgaataaatg agctgtcatc
gcaaaaaggc
5701 gatttgagaa atgtgggctt cagtattaat tgccattttg ctgacaccca
gtgtacctac
5761 ctacctgaga aatttatttt gtccatcatg tatttctcaa agcaaaaggt
ggttttcaag
5821 tataatgtcg ttttcaacat gcttattact tagttttacg tcagctcatt
tcatcatcat
5881 tgataacttg tgaaatactt atctccatcc tatggaatag gggagacggg
tttagacagg
5941 ttcaattagc tcaagtctac acagctgaag tagcagagaa agtgggatct
agatggtctg
6001 atcctagtga tctaccatat gaaggacata gtttgtgtcc tggtccaagt
caaatattga
6061 ctcctcacaa acagtaagta tggcaatttt gtgatgcctt tgattccact
ttacatggag
6121 tactattatt tgtgaaatgt ctttaagatt tttggtctta aatttttgaa
gactgctttc
6181 ccccttttatc tcccagaaaa ttgagaagaa gtaaactcct gcccactaac
aatctcagtc
6241 cgtgaacaaa accaacatga acattcctaa acaagagtgt gtgttactct
aagaagaagg
6301 ctatagaatt tatggaaatg gcttatgtaa cctacaagac tggagaacag
aatgtgactg

```
     6361 gccttttcta atggtccttt aagatttaat gattaaagca agagtttttt
ataattgact
     6421 ttgtggtcta aattcttgat actgtttata attctacaaa gaacaaaaat
tgttatgtac
     6481 tataggcact taagaaccct gaggaaaaat aatacaatgt gtgtgtgtga
gagagagagt
     6541 gagttactga cattgttcca aaaaaaaaaa aaaaaaaaaa aaaaatgtgg
agggttgaaa
     6601 tggtaaggaa ttggaatctt ttgtattttc gagcaataag aattcctatt
cttgtttcaa
     6661 atagaggttt gttaggaatt acagttgtgg ggagcaaact ttctttttg
tgctgtttta
     6721 attcaaaatg tatatcctta attgtatata atatgtagat aaatatatga
gggtattaag
     6781 ctactttgaa ttaaatttaa ggatatattt cacatgaaaa caaatacaaa
cgagaatcaa
      6841 aataaagttt tgcaaagta
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_919415.2

LOCUS NP_919415

ACCESSION NP_919415

```
      1 masasgamak heqilvldpp tdlkfkgpft dvvttnlklr npsdrkvcfk
vkttaprryc
       61 vrpnsgiidp gstvtvsvml qpfdydpnek skhkfmvqti fappntsdme
avwkeakpde
      121 lmdsklrcvf empnendkln dmepskavpl naskqdgpmp kphsvslndt
etrklmeeck
      181 rlqgemmkls eenrhlrdeg lrlrkvahsd kpgststasf rdnvtsplps
llvviaaifi
      241 gfflgkfil
```

HNRNPD

Official Symbol: HNRNPD

Official Name: heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa

Gene ID: 3184

Organism: Homo sapiens

Other Aliases: AUF1, AUF1A, HNRPD, P37, hnRNPD0

Other Designations: ARE-binding protein AUFI, type A; heterogeneous nuclear ribonucleoprotein D0; hnRNP D0

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_031370.2

LOCUS NM_031370

ACCESSION NM_031370

```
      1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
      301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
      361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
      421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac
cgcgtctgga
      481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag
taagaacgag
      541 gaggatgaag gccattcaaa ctcctcccca cgacactctg aagcagcgac
ggcacagcgg
      601 gaagaatgga aaatgtttat aggaggcctt agctgggaca ctacaaagaa
agatctgaag
      661 gactactttt ccaaatttgg tgaagttgta gactgcactc tgaagttaga
tcctatcaca
      721 gggcgatcaa ggggttttgg ctttgtgcta tttaaagaat cggagagtgt
agataaggtc
      781 atggatcaaa aagaacataa attgaatggg aaggtgattg atcctaaaag
ggccaaagcc
      841 atgaaaacaa aagagccggt taaaaaaatt tttgttggtg gcctttctcc
agatacacct
      901 gaagagaaaa taagggagta ctttggtggt tttggtgagg tggaatccat
agagctcccc
      961 atggacaaca agaccaataa gaggcgtggg ttctgcttta ttacctttaa
ggaagaagaa
     1021 ccagtgaaga agataatgga aaagaaatac cacaatgttg gtcttagtaa
atgtgaaata
     1081 aaagtagcca tgtcgaagga acaatatcag caacagcaac agtggggatc
tagaggagga
     1141 tttgcaggaa gagctcgtgg aagaggtggt ggccccagtc aaaactggaa
ccagggatat
     1201 agtaactatt ggaatcaagg ctatggcaac tatggatata acagccaagg
ttacggtggt
     1261 tatggaggat atgactacac tggttacaac aactactatg gatatggtga
ttatagcaac
     1321 cagcagagtg gttatgggaa ggtatccagg cgaggtggtc atcaaaatag
ctacaaacca
     1381 tactaaatta ttccatttgc aacttatccc caacaggtgg tgaagcagta
ttttccaatt
     1441 tgaagattca tttgaaggtg gctcctgcca cctgctaata gcagttcaaa
ctaaattttt
     1501 tgtatcaagt ccctgaatgg aagtatgacg ttgggtccct ctgaagttta
attctgagtt
     1561 ctcattaaaa gaaatttgct ttcattgttt tatttcttaa ttgctatgct
tcagaatcaa
     1621 tttgtgtttt atgccctttc ccccagtatt gtagagcaag tcttgtgtta
aaagcccagt
```

```
     1681 gtgacagtgt catgatgtag tagtgtctta ctggtttttt aataaatcct
tttgtataaa
     1741 aatgtattgg ctcttttatc atcagaatag gaaaaattgt catggattca
agttattaaa
     1801 agcataagtt tggaagacag gcttgccgaa attgaggaca tgattaaaat
tgcagtgaag
     1861 tttgaaatgt ttttagcaaa atctaatttt tgccataatg tgtcctccct
gtccaaattg
     1921 ggaatgactt aatgtcaatt tgtttgttgg ttgttttaat aatacttcct
tatgtagcca
     1981 ttaagattta tatgaatatt ttcccaaatg cccagttttt gcttaatatg
tattgtgctt
     2041 tttagaacaa atctggataa atgtgcaaaa gtaccccttt gcacagatag
ttaatgtttt
     2101 atgcttccat taaataaaaa ggacttaaaa tctgttaatt ataatagaaa
tgcggctagt
     2161 tcagagagat ttttagagct gtggtggact tcatagatga attcaagtgt
tgagggagga
      2221 ttaaagaaat atataccgtg tttatgtgtg tgtgctt
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_112738.1

LOCUS NP_112738

ACCESSION NP_112738

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedeghs nssprhseaa taqreewkmf igglswdttk
kdlkdyfskf
      121 gevvdctlkl dpitgrsrgf gfvlfkeses vdkvmdqkeh klngkvidpk
rakamktkep
      181 vkkifvggls pdtpeekire yfggfgeves ielpmdnktn krrgfcfitf
keeepvkkim
      241 ekkyhnvgls kceikvamsk eqyqqqqqwg srggfagrar grgggpsqnw
nqgysnywnq
      301 gygnygynsq gyggyggydy tgynnyygyg dysnqqsgyg kvsrrgghqn sykpy
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_031369.2

LOCUS NM_031369

ACCESSION NM_031369

```
      1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
```

301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac
cgcgtctgga
481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag
taagaacgag
541 gaggatgaag ggaaaatgtt tataggaggc cttagctggg acactacaaa
gaaagatctg
601 aaggactact tttccaaatt tggtgaagtt gtagactgca ctctgaagtt
agatcctatc
661 acagggcgat caaggggttt tggctttgtg ctatttaaag aatcggagag
tgtagataag
721 gtcatggatc aaaaagaaca taaattgaat gggaaggtga ttgatcctaa
aagggccaaa
781 gccatgaaaa caaaagagcc ggttaaaaaa atttttgttg gtggcctttc
tccagataca
841 cctgaagaga aaataaggga gtactttggt ggttttggtg aggtggaatc
catagagctc
901 cccatggaca acaagaccaa taagaggcgt gggttctgct ttattacctt
taaggaagaa
961 gaaccagtga agaagataat ggaaaagaaa taccacaatg ttggtcttag
taaatgtgaa
1021 ataaaagtag ccatgtcgaa ggaacaatat cagcaacagc aacagtgggg
atctagagga
1081 ggatttgcag gaagagctcg tggaagaggt ggtggcccca gtcaaaactg
gaaccaggga
1141 tatagtaact attggaatca aggctatggc aactatggat ataacagcca
aggttacggt
1201 ggttatggag gatatgacta cactggttac aacaactact atggatatgg
tgattatagc
1261 aaccagcaga gtggttatgg gaaggtatcc aggcgaggtg gtcatcaaaa
tagctacaaa
1321 ccatactaaa ttattccatt tgcaacttat ccccaacagg tggtgaagca
gtattttcca
1381 atttgaagat tcatttgaag gtggctcctg ccacctgcta atagcagttc
aaactaaatt
1441 ttttgtatca agtccctgaa tggaagtatg acgttgggtc cctctgaagt
ttaattctga
1501 gttctcatta aaagaaattt gctttcattg ttttatttct taattgctat
gcttcagaat
1561 caatttgtgt tttatgccct ttcccccagt attgtagagc aagtcttgtg
ttaaaagccc
1621 agtgtgacag tgtcatgatg tagtagtgtc ttactggttt tttaataaat
ccttttgtat
1681 aaaaatgtat tggctctttt atcatcagaa taggaaaaat tgtcatggat
tcaagttatt
1741 aaaagcataa gtttggaaga caggcttgcc gaaattgagg acatgattaa
aattgcagtg
1801 aagtttgaaa tgtttttagc aaaatctaat ttttgccata atgtgtcctc
cctgtccaaa
1861 ttgggaatga cttaatgtca atttgtttgt tggttgtttt aataatactt
ccttatgtag
1921 ccattaagat ttatatgaat attttcccaa atgcccagtt tttgcttaat
atgtattgtg
1981 ctttttagaa caaatctgga taaatgtgca aaagtacccc tttgcacaga
tagttaatgt
2041 tttatgcttc cattaaataa aaaggactta aaatctgtta attataatag
aaatgcggct

```
     2101 agttcagaga gatttttaga gctgtggtgg acttcataga tgaattcaag
tgttgaggga
      2161 ggattaaaga aatatatacc gtgtttatgt gtgtgtgctt
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_112737.1

LOCUS NP_112737

ACCESSION NP_112737

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedegkm figglswdtt kkdlkdyfsk fgevvdctlk
ldpitgrsrg
      121 fgfvlfkese svdkvmdqke hklngkvidp krakamktke pvkkifvggl
spdtpeekir
      181 eyfggfgeve sielpmdnkt nkrrgfcfit fkeeepvkki mekkyhnvgl
skceikvams
      241 keqyqqqqqw gsrggfagra rgrgggpsqn wnqgysnywn qgygnygyns
qgyggyggyd
      301 ytgynnyygy gdysnqqsgy gkvsrrgghq nsykpy
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_002138.3

LOCUS NM_002138

ACCESSION NM_002138

```
      1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
      301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
      361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
      421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac
cgcgtctgga
      481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag
taagaacgag
      541 gaggatgaag gccattcaaa ctcctcccca cgacactctg aagcagcgac
ggcacagcgg
      601 gaagaatgga aaatgtttat aggaggcctt agctgggaca ctacaaagaa
agatctgaag
      661 gactactttt ccaaatttgg tgaagttgta gactgcactc tgaagttaga
tcctatcaca
```

```
      721 gggcgatcaa ggggttttgg ctttgtgcta tttaaagaat cggagagtgt
agataaggtc
      781 atggatcaaa aagaacataa attgaatggg aaggtgattg atcctaaaag
ggccaaagcc
      841 atgaaaacaa aagagccggt taaaaaaatt tttgttggtg gcctttctcc
agatacacct
      901 gaagagaaaa taagggagta ctttggtggt tttggtgagg tggaatccat
agagctcccc
      961 atggacaaca agaccaataa gaggcgtggg ttctgcttta ttacctttaa
ggaagaagaa
     1021 ccagtgaaga agataatgga aaagaaatac cacaatgttg gtcttagtaa
atgtgaaata
     1081 aaagtagcca tgtcgaagga acaatatcag caacagcaac agtggggatc
tagaggagga
     1141 tttgcaggaa gagctcgtgg aagaggtggt gaccagcaga gtggttatgg
gaaggtatcc
     1201 aggcgaggtg gtcatcaaaa tagctacaaa ccatactaaa ttattccatt
tgcaacttat
     1261 ccccaacagg tggtgaagca gtattttcca atttgaagat tcatttgaag
gtggctcctg
     1321 ccacctgcta atagcagttc aaactaaatt ttttgtatca agtccctgaa
tggaagtatg
     1381 acgttgggtc cctctgaagt ttaattctga gttctcatta aaagaaattt
gctttcattg
     1441 ttttatttct taattgctat gcttcagaat caatttgtgt tttatgccct
ttcccccagt
     1501 attgtagagc aagtcttgtg ttaaaagccc agtgtgacag tgtcatgatg
tagtagtgtc
     1561 ttactggttt tttaataaat ccttttgtat aaaaatgtat tggctctttt
atcatcagaa
     1621 taggaaaaat tgtcatggat tcaagttatt aaaagcataa gtttggaaga
caggcttgcc
     1681 gaaattgagg acatgattaa aattgcagtg aagtttgaaa tgtttttagc
aaaatctaat
     1741 ttttgccata atgtgtcctc cctgtccaaa ttgggaatga cttaatgtca
atttgtttgt
     1801 tggttgtttt aataatactt ccttatgtag ccattaagat ttatatgaat
attttcccaa
     1861 atgcccagtt tttgcttaat atgtattgtg ctttttagaa caaatctgga
taaatgtgca
     1921 aaagtacccc tttgcacaga tagttaatgt tttatgcttc cattaaataa
aaaggactta
     1981 aaatctgtta attataatag aaatgcggct agttcagaga gatttttaga
gctgtggtgg
     2041 acttcataga tgaattcaag tgttgaggga ggattaaaga aatatatacc
gtgtttatgt
     2101 gtgtgtgctt
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_002129.2

LOCUS NP_002129

ACCESSION NP_002129

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
```

```
       61 aesegakida skneedeghs nssprhseaa taqreewkmf igglswdttk
kdlkdyfskf
      121 gevvdctlkl dpitgrsrgf gfvlfkeses vdkvmdqkeh klngkvidpk
rakamktkep
      181 vkkifvggls pdtpeekire yfggfgeves ielpmdnktn krrgfcfitf
keeepvkkim
      241 ekkyhnvgls kceikvamsk eqyqqqqqwg srggfagrar grggdqqsgy
gkvsrrgghq
      301 nsykpy
```

Nucleotide sequence (variant 4):

NCBI Reference Sequence: NM_001003810.1

LOCUS NM_001003810

ACCESSION NM_001003810

```
      1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
      301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
      361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
      421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac
cgcgtctgga
      481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag
taagaacgag
      541 gaggatgaag ggaaaatgtt tataggaggc cttagctggg acactacaaa
gaaagatctg
      601 aaggactact tttccaaatt tggtgaagtt gtagactgca ctctgaagtt
agatcctatc
      661 acagggcgat caaggggttt tggctttgtg ctatttaaag aatcggagag
tgtagataag
      721 gtcatggatc aaaaagaaca taaattgaat gggaaggtga ttgatcctaa
aagggccaaa
      781 gccatgaaaa caaaagagcc ggttaaaaaa atttttgttg gtggcctttc
tccagataca
      841 cctgaagaga aaataaggga gtactttggt ggttttggtg aggtggaatc
catagagctc
      901 cccatggaca acaagaccaa taagaggcgt gggttctgct ttattacctt
taaggaagaa
      961 gaaccagtga agaagataat ggaaaagaaa taccacaatg ttggtcttag
taaatgtgaa
     1021 ataaaagtag ccatgtcgaa ggaacaatat cagcaacagc aacagtgggg
atctagagga
     1081 ggatttgcag gaagagctcg tggaagaggt ggtgaccagc agagtggtta
tgggaaggta
     1141 tccaggcgag gtggtcatca aaatagctac aaaccatact aaattattcc
atttgcaact
```

```
     1201 tatccccaac aggtggtgaa gcagtatttt ccaatttgaa gattcatttg
aaggtggctc
     1261 ctgccacctg ctaatagcag ttcaaactaa atttttgta tcaagtccct
gaatggaagt
     1321 atgacgttgg gtccctctga agtttaattc tgagttctca ttaaaagaaa
tttgctttca
     1381 ttgttttatt tcttaattgc tatgcttcag aatcaatttg tgttttatgc
cctttccccc
     1441 agtattgtag agcaagtctt gtgttaaaag cccagtgtga cagtgtcatg
atgtagtagt
     1501 gtcttactgg ttttttaata aatccttttg tataaaaatg tattggctct
tttatcatca
     1561 gaataggaaa aattgtcatg gattcaagtt attaaaagca taagtttgga
agacaggctt
     1621 gccgaaattg aggacatgat taaaattgca gtgaagtttg aaatgttttt
agcaaaatct
     1681 aatttttgcc ataatgtgtc ctccctgtcc aaattgggaa tgacttaatg
tcaatttgtt
     1741 tgttggttgt tttaataata cttccttatg tagccattaa gatttatatg
aatattttcc
     1801 caaatgccca gtttttgctt aatatgtatt gtgcttttta gaacaaatct
ggataaatgt
     1861 gcaaaagtac ccctttgcac agatagttaa tgttttatgc ttccattaaa
taaaaaggac
     1921 ttaaaatctg ttaattataa tagaaatgcg gctagttcag agagattttt
agagctgtgg
     1981 tggacttcat agatgaattc aagtgttgag ggaggattaa agaaatatat
accgtgttta
     2041 tgtgtgtgtg ctt
```

Protein sequence (variant 4):

NCBI Reference Sequence: NP_001003810.1

LOCUS NP_001003810

ACCESSION NP_001003810

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedegkm figglswdtt kkdlkdyfsk fgevvdctlk
ldpitgrsrg
      121 fgfvlfkese svdkvmdqke hklngkvidp krakamktke pvkkifvggl
spdtpeekir
      181 eyfggfgeve sielpmdnkt nkrrgfcfit fkeeepvkki mekkyhnvgl
skceikvams
      241 keqyqqqqqw gsrggfagra rgrggdqqsg ygkvsrrggh qnsykpy
```

BSG

Official Symbol: BSG

Official Name: basigin (Ok blood group)

Gene ID: basigin (Ok blood group)

Organism: Homo sapiens

Other Aliases: UNQ6505/PRO21383, 5F7, CD147, EMMPRIN, M6, OK, TCSF

Other Designations: CD147 antigen; OK blood group antigen; basigin; collagenase stimulatory factor; extracellular matrix metalloproteinase inducer; leukocyte activation antigen M6; tumor cell-derived collagenase stimulatory factor

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001728.3

LOCUS NM_001728

ACCESSION NM_001728

```
      1 gtacatgcga gcgtgtgcgc gcgtgcgcag gcggggcgac cggcgtcccc
ggcgctcgcc
       61 ccgcccccga gatgacgccg tgcgtgcgcg cgcccggtcc gcgcctccgc
cgctttttat
      121 agcggccgcg ggcggcggcg gcagcggttg gaggttgtag gaccggcgag
gaataggaat
      181 catggcggct gcgctgttcg tgctgctggg attcgcgctg ctgggcaccc
acggagcctc
      241 cggggctgcc ggcttcgtcc aggcgccgct gtcccagcag aggtgggtgg
ggggcagtgt
      301 ggagctgcac tgcgaggccg tgggcagccc ggtgcccgag atccagtggt
ggtttgaagg
      361 gcagggtccc aacgacacct gctcccagct ctgggacggc gcccggctgg
accgcgtcca
      421 catccacgcc acctaccacc agcacgcggc cagcaccatc tccatcgaca
cgctcgtgga
      481 ggaggacacg ggcacttacg agtgccgggc cagcaacgac ccggatcgca
accacctgac
      541 ccgggcgccc agggtcaagt gggtccgcgc ccaggcagtc gtgctagtcc
tggaacccgg
      601 cacagtcttc actaccgtag aagaccttgg ctccaagata ctcctcacct
gctccttgaa
      661 tgacagcgcc acagaggtca cagggcaccg ctggctgaag gggggcgtgg
tgctgaagga
      721 ggacgcgctg cccggccaga aaacggagtt caaggtggac tccgacgacc
agtggggaga
      781 gtactcctgc gtcttcctcc ccgagcccat gggcacggcc aacatccagc
tccacgggcc
      841 tcccagagtg aaggctgtga agtcgtcaga acacatcaac gaggggggaga
cggccatgct
      901 ggtctgcaag tcagagtccg tgccacctgt cactgactgg gcctggtaca
agatcactga
      961 ctctgaggac aaggccctca tgaacggctc cgagagcagg ttcttcgtga
gttcctcgca
     1021 gggccggtca gagctacaca ttgagaacct gaacatggag gccgaccccg
gccagtaccg
     1081 gtgcaacggc accagctcca agggctccga ccaggccatc atcacgctcc
gcgtgcgcag
     1141 ccacctggcc gccctctggc ccttcctggg catcgtggct gaggtgctgg
tgctggtcac
```

```
     1201 catcatcttc atctacgaga agcgccggaa gcccgaggac gtcctggatg
atgacgacgc
     1261 cggctctgca cccctgaaga gcagcgggca gcaccagaat gacaaaggca
agaacgtccg
     1321 ccagaggaac tcttcctgag gcaggtggcc cgaggacgct ccctgctcca
cgtctgcgcc
     1381 gccgccggag tccactccca gtgcttgcaa gattccaagt tctcacctct
taaagaaaac
     1441 ccaccccgta gattcccatc atacacttcc ttctttttta aaaaagttgg
gttttctcca
     1501 ttcaggattc tgttccttag gttttttcc ttctgaagtg tttcacgaga
gcccgggagc
     1561 tgctgccctg cggccccgtc tgtggctttc agcctctggg tctgagtcat
ggccgggtgg
     1621 gcggcacagc cttctccact ggccggagtc agtgccaggt ccttgccctt
tgtggaaagt
     1681 cacaggtcac acgaggggcc ccgtgtcctg cctgtctgaa gccaatgctg
tctggttgcg
     1741 ccatttttgt gcttttatgt ttaattttat gagggccacg ggtctgtgtt
cgactcagcc
     1801 tcagggacga ctctgacctc ttggccacag aggactcact tgcccacacc
gagggcgacc
     1861 ccgtcacagc ctcaagtcac tcccaagccc cctccttgtc tgtgcatccg
ggggcagctc
     1921 tggagggggt ttgctgggga actggcgcca tcgccgggac tccagaaccg
cagaagcctc
     1981 cccagctcac ccctggagga cggccggctc tctatagcac cagggctcac
gtgggaaccc
     2041 ccctcccacc caccgccaca ataaagatcg cccccacctc caccctcaaa
aaaaaaaaaa
     2101 aaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001719.2

LOCUS NP_001719

ACCESSION NP_001719

```
      1 maaalfvllg fallgthgas gaagfvqapl sqqrwvggsv elhceavgsp
vpeiqwwfeg
       61 qgpndtcsql wdgarldrvh ihatyhqhaa stisidtlve edtgtyecra
sndpdrnhlt
      121 raprvkwvra qavvlvlepg tvfttvedlg skilltcsln dsatevtghr
wlkggvvlke
      181 dalpgqktef kvdsddqwge yscvflpepm gtaniqlhgp prvkavksse
hinegetaml
      241 vcksesvppv tdwawykitd sedkalmngs esrffvsssq grselhienl
nmeadpgqyr
      301 cngtsskgsd qaiitlrvrs hlaalwpflg ivaevlvlvt iifiyekrrk
pedvldddda
      361 gsaplkssgq hqndkgknvr qrnss
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_198589.2

LOCUS NM_198589

ACCESSION NM_198589

```
      1 gtacatgcga gcgtgtgcgc gcgtgcgcag gcggggcgac cggcgtcccc
ggcgctcgcc
       61 ccgcccccga gatgacgccg tgcgtgcgcg cgcccggtcc gcgcctccgc
cgctttttat
      121 agcggccgcg ggcggcggcg gcagcggttg gaggttgtag gaccggcgag
gaataggaat
      181 catggcggct gcgctgttcg tgctgctggg attcgcgctg ctgggcaccc
acggagcctc
      241 cggggctgcc ggcacagtct tcactaccgt agaagacctt ggctccaaga
tactcctcac
      301 ctgctccttg aatgacagcg ccacagaggt cacagggcac cgctggctga
agggggggcgt
      361 ggtgctgaag gaggacgcgc tgcccggcca gaaaacggag ttcaaggtgg
actccgacga
      421 ccagtgggga gagtactcct gcgtcttcct ccccgagccc atgggcacgg
ccaacatcca
      481 gctccacggg cctcccagag tgaaggctgt gaagtcgtca gaacacatca
acgaggggga
      541 gacggccatg ctggtctgca agtcagagtc cgtgccacct gtcactgact
gggcctggta
      601 caagatcact gactctgagg acaaggccct catgaacggc tccgagagca
ggttcttcgt
      661 gagttcctcg cagggccggt cagagctaca cattgagaac ctgaacatgg
aggccgaccc
      721 cggccagtac cggtgcaacg gcaccagctc caagggctcc gaccaggcca
tcatcacgct
      781 ccgcgtgcgc agccacctgg ccgccctctg gcccttcctg ggcatcgtgg
ctgaggtgct
      841 ggtgctggtc accatcatct tcatctacga gaagcgccgg aagcccgagg
acgtcctgga
      901 tgatgacgac gccggctctg cacccctgaa gagcagcggg cagcaccaga
atgacaaagg
      961 caagaacgtc cgccagagga actcttcctg aggcaggtgg cccgaggacg
ctccctgctc
     1021 cacgtctgcg ccgccgccgg agtccactcc cagtgcttgc aagattccaa
gttctcacct
     1081 cttaaagaaa acccaccccg tagattccca tcatacactt ccttcttttt
taaaaaagtt
     1141 gggttttctc cattcaggat tctgttcctt aggtttttt ccttctgaag
tgtttcacga
     1201 gagcccggga gctgctgccc tgcggccccg tctgtggctt tcagcctctg
ggtctgagtc
     1261 atggccgggt gggcggcaca gccttctcca ctggccggag tcagtgccag
gtccttgccc
     1321 tttgtggaaa gtcacaggtc acacgagggg ccccgtgtcc tgcctgtctg
aagccaatgc
     1381 tgtctggttg cgccattttt gtgcttttat gtttaatttt atgagggcca
cgggtctgtg
     1441 ttcgactcag cctcagggac gactctgacc tcttggccac agaggactca
cttgcccaca
     1501 ccgagggcga ccccgtcaca gcctcaagtc actcccaagc cccctccttg
tctgtgcatc
     1561 cgggggcagc tctggagggg gtttgctggg gaactggcgc catcgccggg
actccagaac
     1621 cgcagaagcc tccccagctc acccctggag gacggccggc tctctatagc
accagggctc
     1681 acgtgggaac ccccctccca cccaccgcca caataaagat cgcccccacc
tccaccctca
      1741 aaaaaaaaaa aaaaaaaaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_940991.1

LOCUS NP_940991

ACCESSION NP_940991

```
      1 maaalfvllg fallgthgas gaagtvfttv edlgskillt cslndsatev
tghrwlkggv
       61 vlkedalpgq ktefkvdsdd qwgeyscvfl pepmgtaniq lhgpprvkav
kssehinege
      121 tamlvckses vppvtdwawy kitdsedkal mngsesrffv sssqgrselh
ienlnmeadp
      181 gqyrcngtss kgsdqaiitl rvrshlaalw pflgivaevl vlvtiifiye
krrkpedvld
      241 dddagsaplk ssgqhqndkg knvrqrnss
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_198590.2

LOCUS NM_198590

ACCESSION NM_198590

```
      1 cccgccagtg tagccacatt cctgcccctt tccagttagc ccttcgcgtt
cggcttagtc
       61 tgcggtcctc ttgcattgcg actccgagtt taacttccaa cacacacttt
caacctccaa
      121 gagacgcccc cacctgtgtc gccccaatag cgacttttct caccgtggtc
gccgcggaac
      181 ttcaagggtc cttcctaccc gcgttgctga gagtctgggt ttacgcgtca
cctcgggcgg
      241 gacccgatcc tccgctcctg aggcccccac aatgaagcag tcggacgcgt
ctccccaaga
      301 aagccggcac agtcttcact accgtagaag accttggctc caagatactc
ctcacctgct
      361 ccttgaatga cagcgccaca gaggtcacag ggcaccgctg gctgaagggg
ggcgtggtgc
      421 tgaaggagga cgcgctgccc ggccagaaaa cggagttcaa ggtggactcc
gacgaccagt
      481 ggggagagta ctcctgcgtc ttcctccccg agcccatggg cacggccaac
atccagctcc
      541 acgggcctcc cagagtgaag gctgtgaagt cgtcagaaca catcaacgag
ggggagacgg
      601 ccatgctggt ctgcaagtca gagtccgtgc cacctgtcac tgactgggcc
tggtacaaga
      661 tcactgactc tgaggacaag gccctcatga acggctccga gagcaggttc
ttcgtgagtt
      721 cctcgcaggg ccggtcagag ctacacattg agaacctgaa catggaggcc
gaccccggcc
      781 agtaccggtg caacggcacc agctccaagg gctccgacca ggccatcatc
acgctccgcg
      841 tgcgcagcca cctggccgcc ctctggccct tcctgggcat cgtggctgag
gtgctggtgc
      901 tggtcaccat catcttcatc tacgagaagc gccggaagcc cgaggacgtc
ctggatgatg
```

```
      961 acgacgccgg ctctgcaccc ctgaagagca gcgggcagca ccagaatgac
aaaggcaaga
     1021 acgtccgcca gaggaactct tcctgaggca ggtggcccga ggacgctccc
tgctccacgt
     1081 ctgcgccgcc gccggagtcc actcccagtg cttgcaagat tccaagttct
cacctcttaa
     1141 agaaaaccca ccccgtagat tcccatcata cacttccttc ttttttaaaa
aagttgggtt
     1201 ttctccattc aggattctgt tccttaggtt tttttccttc tgaagtgttt
cacgagagcc
     1261 cgggagctgc tgccctgcgg ccccgtctgt ggctttcagc ctctgggtct
gagtcatggc
     1321 cgggtgggcg gcacagcctt ctccactggc cggagtcagt gccaggtcct
tgccctttgt
     1381 ggaaagtcac aggtcacacg aggggccccg tgtcctgcct gtctgaagcc
aatgctgtct
     1441 ggttgcgcca tttttgtgct tttatgttta attttatgag ggccacgggt
ctgtgttcga
     1501 ctcagcctca gggacgactc tgacctcttg gccacagagg actcacttgc
ccacaccgag
     1561 ggcgaccccg tcacagcctc aagtcactcc caagcccccc ccttgtctgt
gcatccgggg
     1621 gcagctctgg agggggtttg ctggggaact ggcgccatcg ccgggactcc
agaaccgcag
     1681 aagcctcccc agctcacccc tggaggacgg ccggctctct atagcaccag
ggctcacgtg
     1741 ggaacccccc tcccacccac cgccacaata aagatcgccc ccacctccac
cctcaaaaaa
      1801 aaaaaaaaaa aaaa
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_940992.1

LOCUS NP_940992

ACCESSION NP_940992

```
      1 mgtaniqlhg pprvkavkss ehinegetam lvcksesvpp vtdwawykit
dsedkalmng
       61 sesrffvsss qgrselhien lnmeadpgqy rcngtsskgs dqaiitlrvr
shlaalwpfl
      121 givaevlvlv tiifiyekrr kpedvldddd agsaplkssg qhqndkgknv rqrnss
```

Nucleotide sequence (variant 4):

NCBI Reference Sequence: NM_198591.2

LOCUS NM_198591

ACCESSION NM_198591

```
      1 cccgccagtg tagccacatt cctgcccctt tccagttagc ccttcgcgtt
cggcttagtc
       61 tgcggtcctc ttgcattgcg actccgagtt taacttccaa cacacacttt
caacctccaa
      121 gagacgcccc cacctgtgtc gccccaatag cgacttttct caccgtggtc
gccgcggaac
```

```
      181 ttcaagggtc cttcctaccc gcgttgctga gagtctgggt ttacgcgtca
cctcgggcgg
      241 gacccgatcc tccgctcctg aggcccccac aatgaagcag tcggacgcgt
ctccccaaga
      301 aagggtggac tccgacgacc agtggggaga gtactcctgc gtcttcctcc
ccgagcccat
      361 gggcacggcc aacatccagc tccacgggcc tcccagagtg aaggctgtga
agtcgtcaga
      421 acacatcaac gaggggggaga cggccatgct ggtctgcaag tcagagtccg
tgccacctgt
      481 cactgactgg gcctggtaca agatcactga ctctgaggac aaggccctca
tgaacggctc
      541 cgagagcagg ttcttcgtga gttcctcgca gggccggtca gagctacaca
ttgagaacct
      601 gaacatggag gccgaccccg gccagtaccg gtgcaacggc accagctcca
agggctccga
      661 ccaggccatc atcacgctcc gcgtgcgcag ccacctggcc gccctctggc
ccttcctggg
      721 catcgtggct gaggtgctgg tgctggtcac catcatcttc atctacgaga
agcgccggaa
      781 gcccgaggac gtcctggatg atgacgacgc cggctctgca cccctgaaga
gcagcgggca
      841 gcaccagaat gacaaaggca agaacgtccg ccagaggaac tcttcctgag
gcaggtggcc
      901 cgaggacgct ccctgctcca cgtctgcgcc gccgccggag tccactccca
gtgcttgcaa
      961 gattccaagt tctcacctct taaagaaaac ccaccccgta gattcccatc
atacacttcc
     1021 ttctttttta aaaaagttgg gttttctcca ttcaggattc tgttccttag
gttttttttcc
     1081 ttctgaagtg tttcacgaga gcccgggagc tgctgccctg cggccccgtc
tgtggctttc
     1141 agcctctggg tctgagtcat ggccgggtgg gcggcacagc cttctccact
ggccggagtc
     1201 agtgccaggt ccttgccctt tgtggaaagt cacaggtcac acgaggggcc
ccgtgtcctg
     1261 cctgtctgaa gccaatgctg tctggttgcg ccatttttgt gcttttatgt
ttaattttat
     1321 gagggccacg ggtctgtgtt cgactcagcc tcagggacga ctctgacctc
ttggccacag
     1381 aggactcact tgcccacacc gaggggcgacc ccgtcacagc ctcaagtcac
tcccaagccc
     1441 cctccttgtc tgtgcatccg ggggcagctc tggaggggggt ttgctgggga
actggcgcca
     1501 tcgccgggac tccagaaccg cagaagcctc cccagctcac ccctggagga
cggccggctc
     1561 tctatagcac cagggctcac gtgggaaccc ccctcccacc caccgccaca
ataaagatcg
     1621 cccccacctc caccctcaaa aaaaaaaaaa aaaaaaa
```

Protein sequence (variant 4):

NCBI Reference Sequence: NP_940993.1

LOCUS NP_940993

ACCESSION NP_940993

```
      1 mkqsdaspqe rvdsddqwge yscvflpepm gtaniqlhgp prvkavksse
hinegetaml
       61 vcksesvppv tdwawykitd sedkalmngs esrffvsssq grselhienl
nmeadpgqyr
      121 cngtsskgsd qaiitlrvrs hlaalwpflg ivaevlvlvt iifiyekrrk
pedvldddda
      181 gsaplkssgq hqndkgknvr qrnss
```

EIF4A3

Official Symbol: EIF4A3

Official Name: eukaryotic translation initiation factor 4A3

Gene ID: 9775

Organism: Homo sapiens

Other Aliases: DDX48, NMP265, NUK34, eIF4AIII

Other Designations: ATP-dependent RNA helicase DDX48; ATP-dependent RNA helicase eIF4A-3; DEAD (Asp-Glu-Ala-Asp) box polypeptide 48; DEAD box protein 48; NMP 265; eIF-4A-III; eIF4A-III; eukaryotic initiation factor 4A-III; eukaryotic initiation factor 4A-like NUK-34; eukaryotic translation initiation factor 4A; hNMP 265; nuclear matrix protein 265

Nucleotide sequence:

NCBI Reference Sequence: NM_014740.3

LOCUS NM_014740

ACCESSION NM_014740

```
      1 acgcacgcac gtctctcgct ttcgcatact taaggcgtct gttctcggca
gcggcacagc
       61 gaggtcggca gcggcacagc gaggtcggca gcggcacagc gaggtcggca
gcggcacagc
      121 gaggtcggca gcggcagcga ggtcggcagc ggcacagcga ggtcggcagc
ggcagcgagg
      181 tcggcagcgg cgcgcgctgt gctcttccgc ggactctgaa tcatggcgac
cacggccacg
      241 atggcgacct cgggctcggc gcgaaagcgg ctgctcaaag aggaagacat
gactaaagtg
      301 gaattcgaga ccagcgagga ggtggatgtg acccccacgt tcgacaccat
gggcctgcgg
      361 gaggacctgc tgcggggcat ctacgcttac ggttttgaaa aaccatcagc
aatccagcaa
      421 cgagcaatca agcagatcat caaagggaga gatgtcatcg cacagtctca
gtccggcaca
      481 ggaaaaacag ccaccttcag tatctcagtc ctccagtgtt tggatattca
ggttcgtgaa
      541 actcaagctt tgatcttggc tcccacaaga gagttggctg tgcagatcca
gaaggggctg
      601 cttgctctcg gtgactacat gaatgtccag tgccatgcct gcattggagg
caccaatgtt
```

```
      661 ggcgaggaca tcaggaagct ggattacgga cagcatgttg tcgcgggcac
tccagggcgt
      721 gtttttgata tgattcgtcg cagaagccta aggacacgtg ctatcaaaat
gttggttttg
      781 gatgaagctg atgaaatgtt gaataaaggt ttcaaagagc agatttacga
tgtatacagg
      841 tacctgcctc cagccacaca ggtggttctc atcagtgcca cgctgccaca
cgagattctg
      901 gagatgacca acaagttcat gaccgaccca atccgcatct tggtgaaacg
tgatgaattg
      961 actctggaag gcatcaagca atttttcgtg gcagtggaga gggaagagtg
gaaatttgac
     1021 actctgtgtg acctctacga cacactgacc atcactcagg cggtcatctt
ctgcaacacc
     1081 aaaagaaagg tggactggct gacggagaaa atgagggaag ccaacttcac
tgtatcctca
     1141 atgcatggag acatgcccca gaaagagcgg gagtccatca tgaaggagtt
ccggtcgggc
     1201 gccagccgag tgcttatttc tacagatgtc tgggccaggg ggttggatgt
ccctcaggtg
     1261 tccctcatca ttaactatga tctccctaat aacagagaat tgtacataca
cagaattggg
     1321 agatcaggtc gatacggccg gaagggtgtg gccattaact ttgtaaagaa
tgacgacatc
     1381 cgcatcctca gagatatcga gcagtactat tccactcaga ttgatgagat
gccgatgaac
     1441 gttgctgatc ttatctgaag cagcagatca gtgggatgag ggagactgtt
cacctgctgt
     1501 gtactcctgt ttggaagtat ttagatccag attctactta atggggttta
tatggacttt
     1561 cttctcataa atggcctgcc gtctcccttc ctttgaagag gatatgggga
ttctgctctc
     1621 ttttcttatt tacatgtaaa taatacattg ttctaagtct ttttcattaa
aaatttaaaa
     1681 cttttcccat aaactctata cttctaaggt gccaccacct tctctagtaa ctta
```

Protein sequence:

NCBI Reference Sequence: NP_055555.1

LOCUS NP_055555

ACCESSION NP_055555

```
      1 mattatmats gsarkrllke edmtkvefet seevdvtptf dtmglredll
rgiyaygfek
       61 psaiqqraik qiikgrdvia qsqsgtgkta tfsisvlqcl diqvretqal
ilaptrelav
      121 qiqkgllalg dymnvqchac iggtnvgedi rkldygqhvv agtpgrvfdm
irrrslrtra
      181 ikmlvldead emlnkgfkeq iydvyrylpp atqvvlisat lpheilemtn
kfmtdpiril
      241 vkrdeltleg ikqffvaver eewkfdtlcd lydtltitqa vifcntkrkv
dwltekmrea
      301 nftvssmhgd mpqkeresim kefrsgasrv listdvwarg ldvpqvslii
nydlpnnrel
      361 yihrigrsgr ygrkgvainf vknddirilr dieqyystqi dempmnvadl i
```

MTHFD1

Official Symbol: MTHFD1

Official Name: methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase

Gene ID: 4522

Organism: Homo sapiens

Other Aliases: MTHFC, MTHFD

Other Designations: 5,10-methylenetetrahydrofolate dehydrogenase, 5,10-methylenetetrahydrofolate cyclohydrolase, 10-formyltetrahydrofolate synthetase; C-1-tetrahydrofolate synthase, cytoplasmic; C1-THF synthase; cytoplasmic C-1-tetrahydrofolate synthase

Nucleotide sequence:

NCBI Reference Sequence: NM_005956.3

LOCUS NM_005956

ACCESSION NM_005956

```
      1 aattacggcc ggattccgga gtcctttcca gctccctctt cggccgggtt
tcccgccgaa
       61 tacaaaggcg cactgtgaac tggctctttc tttccgccaa tcatttccgc
cagccattca
      121 tcaccgattt tcttcatctt cccctccctc ttccgtcccg cagtccccga
cctgttagct
      181 ctcggttagt taagggactc gggtccttcc gaactgcgca tgcgccaccg
cgtctgcagg
      241 gggagaagcg ggcaggggcg caggcgcagt agtgtgatcc cctggccagt
ccctaagcac
      301 gtgggttggg ttgtcctgct tggctgcgga gggagtggaa cctcgatatt
ggtggtgtcc
      361 atcgtgggca gcggactaat aaaggccatg gcgccagcag aaatcctgaa
cgggaaggag
      421 atctccgcgc aaataagggc gagactgaaa aatcaagtca ctcagttgaa
ggagcaagta
      481 cctggtttca caccacgcct ggcaatatta caggttggca acagagatga
ttccaatctt
      541 tatataaatg tgaagctgaa ggctgctgaa gagattggga tcaaagccac
tcacattaag
      601 ttaccaagaa caaccacaga atctgaggtg atgaagtaca ttacatcttt
gaatgaagac
      661 tctactgtac atgggttctt agtgcagcta cctttagatt cagagaattc
cattaacact
      721 gaagaagtga tcaatgctat tgcacccgag aaggatgtgg atggattgac
tagcatcaat
      781 gctgggaaac ttgctagagg tgacctcaat gactgtttca ttccttgtac
gcctaaggga
      841 tgcttggaac tcatcaaaga gacaggggtg ccgattgccg gaaggcatgc
tgtggtggtt
      901 gggcgcagta aaatagttgg ggccccgatg catgacttgc ttctgtggaa
caatgccaca
```

961 gtgaccacct gccactccaa gactgcccat ctggatgagg aggtaaataa
aggtgacatc
1021 ctggtggttg caactggtca gcctgaaatg gttaaagggg agtggatcaa
acctggggca
1081 atagtcatcg actgtggaat caattatgtc ccagatgata aaaaaccaaa
tgggagaaaa
1141 gttgtgggtg atgtggcata cgacgaggcc aaagagaggg cgagcttcat
cactcctgtt
1201 cctggcggcg tagggcccat gacagttgca atgctcatgc agagcacagt
agagagtgcc
1261 aagcgtttcc tggagaaatt taagccagga aagtggatga ttcagtataa
caaccttaac
1321 ctcaagacac ctgttccaag tgacattgat atatcacgat cttgtaaacc
gaagcccatt
1381 ggtaagctgg ctcgagaaat tggtctgctg tctgaagagg tagaattata
tggtgaaaca
1441 aaggccaaag ttctgctgtc agcactagaa cgcctgaagc accggcctga
tgggaaatac
1501 gtggtggtga ctggaataac tccaacaccc ctgggagaag ggaaaagcac
aactacaatc
1561 gggctagtgc aagcccttgg tgcccatctc taccagaatg tctttgcgtg
tgtgcgacag
1621 ccttctcagg gccccacctt tggaataaaa ggtggcgctg caggaggcgg
ctactcccag
1681 gtcattccta tggaagagtt taatctccac ctcacaggtg acatccatgc
catcactgca
1741 gctaataacc tcgttgctgc ggccattgat gctcggatat ttcatgaact
gacccagaca
1801 gacaaggctc tctttaatcg tttggtgcca tcagtaaatg gagtgagaag
gttctctgac
1861 atccaaatcc gaaggttaaa gagactaggc attgaaaaga ctgaccctac
cacactgaca
1921 gatgaagaga taaacagatt tgcaagattg gacattgatc cagaaaccat
aacttggcaa
1981 agagtgttgg ataccaatga tagattcctg aggaagatca cgattggaca
ggctccaacg
2041 gagaagggtc acacacggac ggcccagttt gatatctctg tggccagtga
aattatggct
2101 gtcctggctc tcaccacttc tctagaagac atgagagaga gactgggcaa
aatggtggtg
2161 gcatccagta agaaaggaga gcccgtcagt gccgaagatc tgggggtgag
tggtgcactg
2221 acagtgctta tgaaggacgc aatcaagccc aatctcatgc agacactgga
gggcactcca
2281 gtgtttgtcc atgctggccc gtttgccaac atcgcacatg gcaattcctc
catcattgca
2341 gaccggatcg cactcaagct tgttggccca gaagggtttg tagtgacgga
agcaggattt
2401 ggagcagaca ttggaatgga aaagtttttt aacatcaaat gccggtattc
cggcctctgc
2461 ccccacgtgg tggtgcttgt tgccactgtc agggctctca agatgcacgg
gggcggcccc
2521 acggtcactg ctggactgcc tcttcccaag gcttacatac aggagaacct
ggagctggtt
2581 gaaaaaggct tcagtaactt gaagaaacaa attgaaaatg ccagaatgtt
tggaattcca
2641 gtagtagtgg ccgtgaatgc attcaagacg gatacagagt ctgagctgga
cctcatcagc
2701 cgcctttcca gagaacatgg ggcttttgat gccgtgaagt gcactcactg
ggcagaaggg

2761 ggcaagggtg ccttagccct ggctcaggcc gtccagagag cagcacaagc acccagcagc
2821 ttccagctcc tttatgacct caagctccca gttgaggata aaatcaggat cattgcacag
2881 aagatctatg gagcagatga cattgaatta cttcccgaag ctcaacacaa agctgaagtc
2941 tacacgaagc agggctttgg gaatctcccc atctgcatgg ctaaaacaca cttgtctttg
3001 tctcacaacc cagagcaaaa aggtgtccct acaggcttca ttctgcccat tcgcgacatc
3061 cgcgccagcg ttggggctgg ttttctgtac cccttagtag gaacgatgag cacaatgcct
3121 ggactcccca cccggccctg tttttatgat attgatttgg accctgaaac agaacaggtg
3181 aatggattat tctaaacaga tcaccatcca tcttcaagaa gctactttga aagtctggcc
3241 agtgtctatt caggcccact gggagttagg aagtataagt aagccaagag aagtcagccc
3301 ctgcccagaa gatctgaaac taatagtagg agtttcccca gaagtcattt tcagccttaa
3361 ttctcatcat gtataaatta acataaatca tgcatgtctg tttactttag tgacgttcca
3421 cagaataaaa ggaaacaagt ttgccatcaa aaaaaaaaaa aaaaaa

Protein sequence:

NCBI Reference Sequence: NP_005947.3

LOCUS NP_005947

ACCESSION NP_005947

1 mapaeilngk eisaqirarl knqvtqlkeq vpgftprlai lqvgnrddsn lyinvklkaa
61 eeigikathi klprtttese vmkyitslne dstvhgflvq lpldsensin teevinaiap
121 ekdvdgltsi nagklargdl ndcfipctpk gcleliketg vpiagrhavv vgrskivgap
181 mhdlllwnna tvttchskta hldeevnkgd ilvvatgqpe mvkgewikpg aividcginy
241 vpddkkpngr kvvgdvayde akerasfitp vpggvgpmtv amlmqstves akrflekfkp
301 gkwmiqynnl nlktpvpsdi disrsckpkp igklareigl lseevelyge tkakvllsal
361 erlkhrpdgk yvvvtgitpt plgegksttt iglvqalgah lyqnvfacvr qpsqgptfgi
421 kggaagggys qvipmeefnl hltgdihait aannlvaaai darifheltq tdkalfnrlv
481 psvngvrrfs diqirrlkrl giektdpttl tdeeinrfar ldidpetitw qrvldtndrf
541 lrkitigqap tekghtrtaq fdisvaseim avlalttsle dmrerlgkmv vasskkgepv
601 saedlgvsga ltvlmkdaik pnlmqtlegt pvfvhagpfa niahgnssii adrialklvg
661 pegfvvteag fgadigmekf fnikcrysgl cphvvvlvat vralkmhggg ptvtaglplp
721 kayiqenlel vekgfsnlkk qienarmfgi pvvvavnafk tdteseldli srlsrehgaf
781 davkcthwae ggkgalalaq avqraaqaps sfqllydlkl pvedkiriia qkiygaddie

```
      841 llpeaqhkae vytkqgfgnl picmakthls lshnpeqkgv ptgfilpird
irasvgagfl
      901 yplvgtmstm pglptrpcfy didldpeteq vnglf
```

ENO2

Official Symbol: ENO2

Official Name: enolase 2 (gamma, neuronal)

Gene ID: 2026

Organism: Homo sapiens

Other Aliases: NSE

Other Designations: 2-phospho-D-glycerate hydro-lyase; 2-phospho-D-glycerate hydrolyase; gamma-enolase; neural enolase; neuron specific gamma enolase; neuron-specific enolase; neurone-specific enolase

Nucleotide sequence:

NCBI Reference Sequence: NM_001975.2

LOCUS NM_001975

ACCESSION NM_001975

```
      1 acccgcgctc gtacgtgcgc ctccgccggc agctcctgac tcatcggggg
ctccgggtca
       61 catgcgcccg cgcggcccta taggcgcctc ctccgcccgc cgcccgggag
ccgcagccgc
      121 cgccgccact gccactcccg ctctctcagc gccgccgtcg ccaccgccac
cgccaccgcc
      181 actaccaccg tctgagtctg cagtcccgag atcccagcca tcatgtccat
agagaagatc
      241 tgggcccggg agatcctgga ctcccgcggg aaccccacag tggaggtgga
tctctatact
      301 gccaaaggtc ttttccgggc tgcagtgccc agtggagcct ctacgggcat
ctatgaggcc
      361 ctggagctga gggatggaga caaacagcgt tacttaggca aaggtgtcct
gaaggcagtg
      421 gaccacatca actccaccat cgcgccagcc ctcatcagct caggtctctc
tgtggtggag
      481 caagagaaac tggacaacct gatgctggag ttggatggga ctgagaacaa
atccaagttt
      541 ggggccaatg ccatcctggg tgtgtctctg gccgtgtgta aggcaggggc
agctgagcgg
      601 gaactgcccc tgtatcgcca cattgctcag ctggccggga actcagacct
catcctgcct
      661 gtgccggcct tcaacgtgat caatggtggc tctcatgctg gcaacaagct
ggccatgcag
      721 gagttcatga tcctcccagt gggagctgag agctttcggg atgccatgcg
actaggtgca
```

781 gaggtctacc atacactcaa gggagtcatc aaggacaaat acggcaagga
tgccaccaat
841 gtgggggatg aaggtggctt tgcccccaat atcctggaga acagtgaagc
cttggagctg
901 gtgaaggaag ccatcgacaa ggctggctac acggaaaaga tcgttattgg
catggatgtt
961 gctgcctcag agttttatcg tgatggcaaa tatgacttgg acttcaagtc
tcccactgat
1021 ccttcccgat acatcactgg ggaccagctg ggggcactct accaggactt
tgtcagggac
1081 tatcctgtgg tctccattga ggacccattt gaccaggatg attgggctgc
ctggtccaag
1141 ttcacagcca atgtagggat ccagattgtg ggtgatgacc tgacagtgac
caacccaaaa
1201 cgtattgagc gggcagtgga agaaaaggcc tgcaactgtc tgctgctcaa
ggtcaaccag
1261 atcggctcgg tcactgaagc catccaagcg tgcaagctgg cccaggagaa
tggctggggg
1321 gtcatggtga gtcatcgctc aggagagact gaggacacat tcattgctga
cctggtggtg
1381 gggctgtgca caggccagat caagactggt gccccgtgcc gttctgaacg
tctggctaaa
1441 tacaaccagc tcatgagaat tgaggaagag ctgggggatg aagctcgctt
tgccggacat
1501 aacttccgta atcccagtgt gctgtgattc ctctgcttgc ctggagacgt
ggaacctctg
1561 tctcatcctc ctggaacctt gctgtcctga tctgtgatag ttcacccccct
gagatcccct
1621 gagccccagg gtgcccagaa cttccctgat tgacctgctc cgctgctcct
tggcttacct
1681 gacctcttgc tgtctctgct cgccctcctt tctgtgccct actcattggg
gttccgcact
1741 ttccacttct tcctttctct ttctctcttc cctcagaaac tagaaatgtg
aatgaggatt
1801 attataaaag ggggtccgtg gaagaatgat cagcatctgt gatgggagcg
tcagggttgg
1861 tgtgctgagg tgttagagag ggaccatgtg tcacttgtgc tttgctcttg
tcccacgtgt
1921 cttccacttt gcatatgagc cgtgaactgt gcatagtgct gggatggagg
ggagtgttgg
1981 gcatgtgatc acgcctggct aataaggctt tagtgtattt atttatttat
ttatttttatt
2041 tgtttttcat tcatcccatt aatcatttcc ccataactca atggcctaaa
actggcctga
2101 cttgggggaa cgatgtgtct gtatttcatg tggctgtaga tcccaagatg
actggggtgg
2161 gaggtcttgc tagaatggga agggtcatag aaagggcctt gacatcagtt
cctttgtgtg
2221 tactcactga agcctgcgtt ggtccagagc ggaggctgtg tgcctggggg
agttttcctc
2281 tatacatctc tccccaaccc taggttccct gttcttcctc cagctgcacc
agagcaacct
2341 ctcactcccc atgccacgtt ccacagttgc caccacctct gtggcattga
aatgagcacc
2401 tccattaaag tctgaatcag tgc

Protein sequence:

NCBI Reference Sequence: NP_001966.1

LOCUS NP_001966

ACCESSION NP_001966

```
      1 msiekiware ildsrgnptv evdlytakgl fraavpsgas tgiyealelr
dgdkqrylgk
       61 gvlkavdhin stiapaliss glsvveqekl dnlmleldgt enkskfgana
ilgvslavck
      121 agaaerelpl yrhiaqlagn sdlilpvpaf nvinggshag nklamqefmi
lpvgaesfrd
      181 amrlgaevyh tlkgvikdky gkdatnvgde ggfapnilen sealelvkea
idkagyteki
      241 vigmdvaase fyrdgkydld fksptdpsry itgdqlgaly qdfvrdypvv
siedpfdqdd
      301 waawskftan vgiqivgddl tvtnpkrier aveekacncl llkvnqigsv
teaiqackla
      361 qengwgvmvs hrsgetedtf iadlvvglct gqiktgapcr serlakynql
mrieeelgde
      421 arfaghnfrn psvl
```

ATP5H

Official Symbol: ATP5H

Official Name: ATP synthase, H+ transporting, mitochondrial Fo complex, subunit d

Gene ID: 10476

Organism: Homo sapiens

Other Aliases: My032, ATPQ

Other Designations: ATP synthase D chain, mitochondrial; ATP synthase subunit d, mitochondrial; ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d; ATP synthase, H+ transporting, mitochondrial F1F0, subunit d; ATPase subunit d; My032 protein

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_006356.2

LOCUS NM_006356

ACCESSION NM_006356.

```
      1 tgacccactt ccgttacttg ctgcggagga ccgtgggcag ccagggtcgg
tgaaggatcc
       61 caaaatggct gggcgaaaac ttgctctaaa aaccattgac tgggtagctt
ttgcagagat
      121 cataccccag aaccaaaagg ccattgctag ttccctgaaa tcctggaatg
agaccctcac
      181 ctccaggttg gctgctttac ctgagaatcc accagctatc gactgggctt
actacaaggc
      241 caatgtggcc aaggctggct tggtggatga ctttgagaag aagtttaatg
cgctgaaggt
```

```
      301 tcccgtgcca gaggataaat atactgccca ggtggatgcc gaagaaaaag
aagatgtgaa
      361 atcttgtgct gagtgggtgt ctctctcaaa ggccaggatt gtagaatatg
agaaagagat
      421 ggagaagatg aagaacttaa ttccatttga tcagatgacc attgaggact
tgaatgaagc
      481 tttcccagaa accaaattag acaagaaaaa gtatccctat tggcctcacc
aaccaattga
      541 gaatttataa aattgagtcc aggaggaagc tctggccctt gtattacaca
ttctggacat
      601 taaaaataat aattatacag ttaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_006347.1

LOCUS NP_006347

ACCESSION NP_006347

```
      1 magrklalkt idwvafaeii pqnqkaiass lkswnetlts rlaalpenpp
aidwayykan
       61 vakaglvddf ekkfnalkvp vpedkytaqv daeekedvks caewvslska
riveyekeme
      121 kmknlipfdq mtiedlneaf petkldkkky pywphqpien l
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001003785.1

LOCUS NM_001003785

ACCESSION NM_001003785

```
      1 tgacccactt ccgttacttg ctgcggagga ccgtgggcag ccagggtcgg
tgaaggatcc
       61 caaaatggct gggcgaaaac ttgctctaaa aaccattgac tgggtagctt
ttgcagagat
      121 cataccccag aaccaaaagg ccattgctag ttccctgaaa tcctggaatg
agaccctcac
      181 ctccaggttg gctgctttac ctgagaatcc accagctatc gactgggctt
actacaaggc
      241 caatgtggcc aaggctggct tggtggatga ctttgagaag aaggtgaaat
cttgtgctga
      301 gtgggtgtct ctctcaaagg ccaggattgt agaatatgag aaagagatgg
agaagatgaa
      361 gaacttaatt ccatttgatc agatgaccat tgaggacttg aatgaagctt
tcccagaaac
      421 caaattagac aagaaaaagt atccctattg gcctcaccaa ccaattgaga
atttataaaa
      481 ttgagtccag gaggaagctc tggcccttgt attacacatt ctggacatta
aaaataataa
      541 ttatacagtt aaaaaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001003785.1

LOCUS NP_001003785

ACCESSION NP_001003785

```
      1 magrklalkt idwvafaeii pqnqkaiass lkswnetlts rlaalpenpp
aidwayykan
       61 vakaglvddf ekkvkscaew vslskarive yekemekmkn lipfdqmtie
dlneafpetk
      121 ldkkkypywp hqpienl
```

TRAP1

Official Symbol: TRAP1

Official Name: TNF receptor-associated protein 1

Gene ID: 10131

Organism: Homo sapiens

Other Aliases: HSP75, HSP90L

Other Designations: HSP 75; TNFR-associated protein 1; TRAP-1; heat shock protein 75 kDa, mitochondrial; tumor necrosis factor type 1 receptor associated protein; tumor necrosis factor type 1 receptor-associated protein

Nucleotide sequence:

NCBI Reference Sequence: NM_016292.2

LOCUS NM_016292

ACCESSION NM_016292

```
      1 gaggaagccc cgccccgcgc agccccgtcc cgccccttcc catcgtgtac
ggtcccgcgt
       61 ggctgcgcgc ggcgctctgg gagtacgaca tggcgcgcga gctgcgggcg
ctgctgctgt
      121 ggggccgccg cctgcggcct ttgctgcggg cgccggcgct ggcggccgtg
ccgggaggaa
      181 aaccaattct gtgtcctcgg aggaccacag cccagttggg ccccaggcga
aacccagcct
      241 ggagcttgca ggcaggacga ctgttcagca cgcagaccgc cgaggacaag
gaggaacccc
      301 tgcactcgat tatcagcagc acagagagcg tgcagggttc cacttccaaa
catgagttcc
      361 aggccgagac aaagaagctt ttggacattg ttgcccggtc cctgtactca
gaaaaagagg
      421 tgtttatacg ggagctgatc tccaatgcca gcgatgcctt ggaaaaactg
cgtcacaaac
      481 tggtgtctga cggccaagca ctgccagaaa tggagattca cttgcagacc
aatgccgaga
      541 aaggcaccat caccatccag gatactggta tcgggatgac acaggaagag
ctggtgtcca
      601 acctggggac gattgccaga tcggggtcaa aggccttcct ggatgctctg
cagaaccagg
      661 ctgaggccag cagcaagatc atcggccagt ttggagtggg tttctactca
gctttcatgg
```

```
      721 tggctgacag agtggaggtc tattcccgct cggcagcccc ggggagcctg
ggttaccagt
      781 ggctttcaga tggttctgga gtgtttgaaa tcgccgaagc ttcgggagtt
agaaccggga
      841 caaaaatcat catccacctg aaatccgact gcaaggagtt ttccagcgag
gcccgggtgc
      901 gagatgtggt aacgaagtac agcaacttcg tcagcttccc cttgtacttg
aatggaaggc
      961 ggatgaacac cttgcaggcc atctggatga tggaccccaa ggatgtccgt
gagtggcaac
     1021 atgaggagtt ctaccgctac gtcgcgcagg ctcacgacaa gccccgctac
accctgcact
     1081 ataagacgga cgcaccgctc aacatccgca gcatcttcta cgtgcccgac
atgaaaccgt
     1141 ccatgtttga tgtgagccgg gagctgggct ccagcgttgc actgtacagc
cgcaaagtcc
     1201 tcatccagac caaggccacg gacatcctgc ccaagtggct gcgcttcatc
cgaggtgtgg
     1261 tggacagtga ggacattccc ctgaacctca gccgggagct gctgcaggag
agcgcactca
     1321 tcaggaaact ccgggacgtt ttacagcaga ggctgatcaa attcttcatt
gaccagagta
     1381 aaaaagatgc tgagaagtat gcaaagtttt ttgaagatta cggcctgttc
atgcgggagg
     1441 gcattgtgac cgccaccgag caggaggtca aggaggacat agcaaagctg
ctgcgctacg
     1501 agtcctcggc gctgccctcc gggcagctaa ccagcctctc agaatacgcc
agccgcatgc
     1561 gggccggcac ccgcaacatc tactacctgt gcgcccccaa ccgtcacctg
gcagagcact
     1621 caccctacta tgaggccatg aagaagaaag acacagaggt tctcttctgc
tttgagcagt
     1681 ttgatgagct caccctgctg caccttcgtg agtttgacaa gaagaagctg
atctctgtgg
     1741 agacggacat agtcgtggat cactacaagg aggagaagtt tgaggacagg
tccccagccg
     1801 ccgagtgcct atcagagaag gagacggagg agctcatggc ctggatgaga
aatgtgctgg
     1861 ggtcgcgtgt caccaacgtg aaggtgaccc tccgactgga cacccaccct
gccatggtca
     1921 ccgtgctgga gatgggggct gcccgccact tcctgcgcat gcagcagctg
gccaagaccc
     1981 aggaggagcg cgcacagctc ctgcagccca cgctggagat caaccccagg
cacgcgctca
     2041 tcaagaagct gaatcagctg cgcgcaagcg agcctggcct ggctcagctg
ctggtggatc
     2101 agatatacga gaacgccatg attgctgctg gacttgttga cgaccctagg
gccatggtgg
     2161 gccgcttgaa tgagctgctt gtcaaggccc tggagcgaca ctgacagcca
gggggccaga
     2221 aggactgaca ccacagatga cagccccacc tccttgagct ttatttacct
aaatttaaag
      2281 gtatttctta acccgaaaaa aaaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_057376.2

LOCUS NP_057376

ACCESSION NP_057376

```
      1 marelralll wgrrlrpllr apalaavpgg kpilcprrtt aqlgprrnpa
wslqagrlfs
       61 tqtaedkeep lhsiisstes vqgstskhef qaetkklldi varslyseke
vfirelisna
      121 sdaleklrhk lvsdgqalpe meihlqtnae kgtitiqdtg igmtqeelvs
nlgtiarsgs
      181 kafldalqnq aeasskiigq fgvgfysafm vadrvevysr saapgslgyq
wlsdgsgvfe
      241 iaeasgvrtg tkiiihlksd ckefssearv rdvvtkysnf vsfplylngr
rmntlqaiwm
      301 mdpkdvrewq heefyryvaq ahdkprytlh yktdaplnir sifyvpdmkp
smfdvsrelg
      361 ssvalysrkv liqtkatdil pkwlrfirgv vdsediplnl srellqesal
irklrdvlqq
      421 rlikffidqs kkdaekyakf fedyglfmre givtateqev kediakllry
essalpsgql
      481 tslseyasrm ragtrniyyl capnrhlaeh spyyeamkkk dtevlfcfeq
fdeltllhlr
      541 efdkkklisv etdivvdhyk eekfedrspa aeclsekete elmawmrnvl
gsrvtnvkvt
      601 lrldthpamv tvlemgaarh flrmqqlakt qeeraqllqp tleinprhal
ikklnqlras
      661 epglaqllvd qiyenamiaa glvddpramv grlnellvka lerh
```

SDHA

Official Symbol: SDHA

Official Name: succinate dehydrogenase complex, subunit A, flavoprotein (Fp)

Gene ID: 6389

Organism: Homo sapiens

Other Aliases: CMD1GG, FP, PGL5, SDH1, SDH2, SDHF

Other Designations: flavoprotein subunit of complex II; succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial; succinate dehydrogenase complex flavoprotein subunit

Nucleotide sequence:

NCBI Reference Sequence: NM_004168.2

LOCUS NM_004168

ACCESSION NM_004168

```
      1 tccggcgtgg tgcgcaggcg cggtatcccc cctcccccgc cagctcgacc
ccggtgtggt
       61 gcgcaggcgc agtctgcgca gggactggcg ggactgcgcg gcggcaacag
cagacatgtc
```

121 gggggtccgg ggcctgtcgc ggctgctgag cgctcggcgc ctggcgctgg
ccaaggcgtg
181 gccaacagtg ttgcaaacag gaacccgagg ttttcacttc actgttgatg
ggaacaagag
241 ggcatctgct aaagtttcag attccatttc tgctcagtat ccagtagtgg
atcatgaatt
301 tgatgcagtg gtggtaggcg ctggaggggc aggcttgcga gctgcatttg
gcctttctga
361 ggcagggttt aatacagcat gtgttaccaa gctgtttcct accaggtcac
acactgttgc
421 agcacaggga ggaatcaatg ctgctctggg gaacatggag gaggacaact
ggaggtggca
481 tttctacgac accgtgaagg gctccgactg gctgggggac caggatgcca
tccactacat
541 gacggagcag gcccccgccg ccgtggtcga gctagaaaat tatggcatgc
cgtttagcag
601 aactgaagat gggaagattt atcagcgtgc atttggtgga cagagcctca
agtttggaaa
661 gggcgggcag gcccatcggt gctgctgtgt ggctgatcgg actggccact
cgctattgca
721 caccttatat ggaaggtctc tgcgatatga taccagctat tttgtggagt
attttgcctt
781 ggatctcctg atggagaatg gggagtgccg tggtgtcatc gcactgtgca
tagaggacgg
841 gtccatccat cgcataagag caaagaacac tgttgttgcc acaggaggct
acgggcgcac
901 ctacttcagc tgcacgtctg cccacaccag cactggcgac ggcacggcca
tgatcaccag
961 ggcaggcctt ccttgccagg acctagagtt tgttcagttc caccctacag
gcatatatgg
1021 tgctggttgt ctcattacgg aaggatgtcg tggagaggga ggcattctca
ttaacagtca
1081 aggcgaaagg tttatggagc gatacgcccc tgtcgcgaag gacctggcgt
ctagagatgt
1141 ggtgtctcgg tccatgactc tggagatccg agaaggaaga ggctgtggcc
ctgagaaaga
1201 tcacgtctac ctgcagctgc accacctacc tccagagcag ctggccacgc
gcctgcctgg
1261 catttcagag acagccatga tcttcgctgg cgtggacgtc acgaaggagc
cgatccctgt
1321 cctccccacc gtgcattata acatgggcgg cattcccacc aactacaagg
ggcaggtcct
1381 gaggcacgtg aatggccagg atcagattgt gcccggcctg tacgcctgtg
gggaggccgc
1441 ctgtgcctcg gtacatggtg ccaaccgcct cggggcaaac tcgctcttgg
acctggttgt
1501 ctttggtcgg gcatgtgccc tgagcatcga agagtcatgc aggcctggag
ataaagtccc
1561 tccaattaaa ccaaacgctg gggaagaatc tgtcatgaat cttgacaaat
tgagatttgc
1621 tgatggaagc ataagaacat cggaactgcg actcagcatg cagaagtcaa
tgcaaaatca
1681 tgctgccgtg ttccgtgtgg gaagcgtgtt gcaagaaggt tgtgggaaaa
tcagcaagct
1741 ctatggagac ctaaagcacc tgaagacgtt cgaccgggga atggtctgga
acacggacct
1801 ggtggagacc ctggagctgc agaacctgat gctgtgtgcg ctgcagacca
tctacggagc
1861 agaggcacgg aaggagtcac ggggcgcgca tgccagggaa gactacaagg
tgcggattga

```
     1921 tgagtacgat tactccaagc ccatccaggg gcaacagaag aagccctttg
aggagcactg
     1981 gaggaagcac accctgtcct atgtggacgt tggcactggg aaggtcactc
tggaatatag
     2041 acccgtgatc gacaaaactt tgaacgaggc tgactgtgcc accgtcccgc
cagccattcg
     2101 ctcctactga tgagacaaga tgtggtgatg acagaatcag cttttgtaat
tatgtataat
     2161 agctcatgca tgtgtccatg tcataactgt cttcatacgc ttctgcactc
tggggaagaa
     2221 ggagtacatt gaagggagat tggcacctag tggctgggag cttgccagga
acccagtggc
     2281 cagggagcgt ggcacttacc tttgtccctt gcttcattct tgtgagatga
taaaactggg
     2341 cacagctctt aaataaaata taaatgaaca aactttcttt tatttccaaa
aaaaaaaaaa
     2401 aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_004159.2

LOCUS NP_004159

ACCESSION NP_004159

```
      1 msgvrglsrl lsarrlalak awptvlqtgt rgfhftvdgn krasakvsds
isaqypvvdh
       61 efdavvvgag gaglraafgl seagfntacv tklfptrsht vaaqgginaa
lgnmeednwr
      121 whfydtvkgs dwlgdqdaih ymteqapaav velenygmpf srtedgkiyq
rafggqslkf
      181 gkggqahrcc cvadrtghsl lhtlygrslr ydtsyfveyf aldllmenge
crgvialcie
      241 dgsihrirak ntvvatggyg rtyfsctsah tstgdgtami traglpcqdl
efvqfhptgi
      301 ygagcliteg crgeggilin sqgerfmery apvakdlasr dvvsrsmtle
iregrgcgpe
      361 kdhvylqlhh lppeqlatrl pgisetamif agvdvtkepi pvlptvhynm
ggiptnykgq
      421 vlrhvngqdq ivpglyacge aacasvhgan rlganslldl vvfgracals
ieescrpgdk
      481 vppikpnage esvmnldklr fadgsirtse lrlsmqksmq nhaavfrvgs
vlqegcgkis
      541 klygdlkhlk tfdrgmvwnt dlvetlelqn lmlcalqtiy gaearkesrg
aharedykvr
      601 ideydyskpi qgqqkkpfee hwrkhtlsyv dvgtgkvtle yrpvidktln
eadcatvppa
      661 irsy
```

TPMA

Official Symbol: TPM4

Official Name: tropomyosin 4

Gene ID: 7171

Organism: Homo sapiens

Other Aliases:

Other Designations: TM30p1; tropomyosin alpha-4 chain; tropomyosin-4;

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001145160.1

LOCUS NM_001145160

ACCESSION NM_001145160

```
      1 ataaggccct ctcctccacc ctgccaggct cactctgccc cacagccaca
gcccctgact
       61 gccgcagccc ccacagagcc cgccgcgcac cccacgtccc ccacgccagc
gcccagccat
      121 ggaggccatc aagaagaaaa tgcagatgct gaagttggac aaggagaatg
ccatcgaccg
      181 cgcggagcag gcggaggcgg ataagaaagc cgctgaggac aagtgcaagc
aggtggagga
      241 ggagctgacg cacctccaga agaaactaaa agggacagag gacgagctgg
ataaatattc
      301 cgaggacctg aaggacgcgc aggagaagct ggagctcacg gagaagaagg
cctccgacgc
      361 tgaaggtgat gtggccgccc tcaaccgacg catccagctc gttgaggagg
agttggacag
      421 ggctcaggaa cgactggcca cggccctgca gaagctggag gaggcagaaa
aagctgcaga
      481 tgagagtgag agaggaatga aggtgataga aaaccgggcc atgaaggatg
aggagaagat
      541 ggagattcag gagatgcagc tcaaagaggc caagcacatt gcggaagagg
ctgaccgcaa
      601 atacgaggag gtagctcgta agctggtcat cctggagggt gagctggaga
gggcagagga
      661 gcgtgcggag gtgtctgaac taaaatgtgg tgacctggaa gaagaactca
agaatgttac
      721 taacaatctg aaatctctgg aggctgcatc tgaaaagtat tctgaaaagg
aggacaaata
      781 tgaagaagaa attaaacttc tgtctgacaa actgaaagag gctgagaccc
gtgctgaatt
      841 tgcagagaga acggttgcaa aactggaaaa gacaattgat gacctggaag
agaaacttgc
      901 ccaggccaaa gaagagaacg tgggcttaca tcagacactg gatcagacac
taaacgaact
      961 taactgtata taagcaaaac agaagagtct tgttccaaca gaaactctgg
agctccgtgg
     1021 gtctttctct tctcttgtaa gaagttcctt ttgttattgc catcttcgct
ttgctggaaa
     1081 tgtcaagcaa attatgaata catgaccaaa tattttgtat cggagaagct
ttgagcacca
     1141 gttaaatctc attccttccc ttttttttttc aaatggcacc agctttttca
gctctcttat
     1201 tttttcctta agtagcattt attcctaagg taggcagggt atttcctagt
aagcatactt
```

1261 tcttaagacg gaggccattt ggttcctggg agaataggca gccccacact
ttgaagaata
1321 cagaccccag tatctagtcg tggatataat taaaacgctg aagaccataa
ccttttgggt
1381 caactgttgg tcaaactata ggagagacca gggaccatca catgggtagg
gattttccat
1441 ccagagccaa taaaaggact ggtgggggcc gggggtggct attgtgggaa
gtcataaccc
1501 acagatagat caacctaaga atcctggccc ttctccactc tccaccatgc
aggacaaaca
1561 tcttctcaag cagtcaacgt agaatgcttg ggaaatagtc ataattaccc
acatatagta
1621 attaatagat ggtaattaat tgatccttga tgtgatgttc ttttgcatat
ttccttcatt
1681 ctaaagttgt tccctggccg ggagcgttgg ctttcgcctg taatcccaac
actttgggag
1741 gccaggacag atcacttgag gtcaggagtt cgagaccagc ccagccaaca
tggcgaaacc
1801 atgtctctac taaaaataca aaaattatgg tgacgcctgc ctgtagtccc
agctactcgg
1861 gaggctgagg caggaggatc gcttgaaccc aggaagtgga gactgcagtg
agccgatatc
1921 gcaccacagc gctccagcct ggtcgacaga gtgagactcc atctcaagaa
aaaataaaaa
1981 taaagttgtt ctctgaagag caaatgtctc attccagtaa tgacccactc
agcaggaata
2041 tggtggagtt cagtccaatt caggtcagcc atatccaaaa gaccacaagt
cattactaag
2101 ttgagcaaaa gagtttttat ctattagcag aaagggcctc tctggcagca
gagattaaaa
2161 actggcccaa cttcatttcc atacttcagg gaacagcaaa ttgaggattt
acttatctag
2221 gacttgaatt ccttctttgg gaccaagtta ataaaagacc aagaaactcc
tgattaaact
2281 ggataatgaa ggattctgta gacagggctg cacgtatcgg ctttgtttga
cttctctttt
2341 ctcagttaac atctcagagc tagaacattc cacattcccc agcagcgtgt
gggggctgac
2401 taaagtttac aattccaact aaaaatcacc ctgcttctgg cttatctgaa
tcccttaccc
2461 accccacccc accaccctac tcctatttat tcagcaccac actacccagg
aaatacacta
2521 gcaaattgtg caatggaata aaatccacac tttagattct tgcaactgta
tcatatgtaa
2581 tagtatcact ttttctacat tttggtcaaa taaattttta cataaactac

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001138632.1

LOCUS NP_001138632

ACCESSION NP_001138632

1 meaikkkmqm lkldkenaid raeqaeadkk aaedkckqve eelthlqkkl
kgtedeldky
61 sedlkdaqek leltekkasd aegdvaalnr riqlveeeld raqerlatal
qkleeaekaa
121 desergmkvi enramkdeek meiqemqlke akhiaeeadr kyeevarklv
ilegelerae

181 eraevselkc gdleeelknv tnnlksleaa sekysekedk yeeeikllsd klkeaetrae
241 faertvakle ktiddleekl aqakeenvgl hqtldqtlne lnci

Nucleotide sequence:

NCBI Reference Sequence (variant 2): NM_003290.2

LOCUS NM_003290

ACCESSION NM_003290

1 tttccagcag ctgtggccag cggtgccgac gtcaggccct cccccagcgg tgctgacgtc
61 ggcggtccgg ccgggtgacc tcatcgcccc gacggcagcc ggcccggggg gcggggagag
121 gcgggggcgg cccccgcgca ggcaaaggct tgggggccg gggcgcggct gtgcagctct
181 cgccggagcc gagcccagcc gagcgtccgc cgctgcccgt gcgcctctgc gcctccgcgc
241 catggccggc ctcaactccc tggaggcggt gaaacgcaag atccaggccc tgcagcagca
301 ggcggacgag gcggaagacc gcgcgcaggg cctgcagcgg gagctggacg gcgagcgcga
361 gcggcgcgag aaagctgaag gtgatgtggc cgccctcaac cgacgcatcc agctcgttga
421 ggaggagttg gacagggctc aggaacgact ggccacggcc ctgcagaagc tggaggaggc
481 agaaaaagct gcagatgaga gtgagagagg aatgaaggtg atagaaaacc gggccatgaa
541 ggatgaggag aagatggaga ttcaggagat gcagctcaaa gaggccaagc acattgcgga
601 agaggctgac cgcaaatacg aggaggtagc tcgtaagctg gtcatcctgg agggtgagct
661 ggagagggca gaggagcgtg cggaggtgtc tgaactaaaa tgtggtgacc tggaagaaga
721 actcaagaat gttactaaca atctgaaatc tctggaggct gcatctgaaa agtattctga
781 aaaggaggac aaatatgaag aagaaattaa acttctgtct gacaaactga aagaggctga
841 gacccgtgct gaatttgcag agagaacggt tgcaaaactg gaaaagacaa ttgatgacct
901 ggaagagaaa cttgcccagg ccaaagaaga gaacgtgggc ttacatcaga cactggatca
961 gacactaaac gaacttaact gtatataagc aaaacagaag agtcttgttc caacagaaac
1021 tctggagctc cgtgggtctt tctcttctct tgtaagaagt tccttttgtt attgccatct
1081 tcgctttgct ggaaatgtca agcaaattat gaatacatga ccaaatattt tgtatcggag
1141 aagctttgag caccagttaa atctcattcc ttccctttt ttttcaaatg gcaccagctt
1201 tttcagctct cttatttttt ccttaagtag catttattcc taaggtaggc agggtatttc
1261 ctagtaagca tactttctta agacggaggc catttggttc ctgggagaat aggcagcccc
1321 acactttgaa gaatacagac cccagtatct agtcgtggat ataattaaaa cgctgaagac

```
     1381 cataaccttt tgggtcaact gttggtcaaa ctataggaga gaccagggac
catcacatgg
     1441 gtagggattt tccatccaga gccaataaaa ggactggtgg gggccggggg
tggctattgt
     1501 gggaagtcat aacccacaga tagatcaacc taagaatcct ggcccttctc
cactctccac
     1561 catgcaggac aaacatcttc tcaagcagtc aacgtagaat gcttgggaaa
tagtcataat
     1621 tacccacata tagtaattaa tagatggtaa ttaattgatc cttgatgtga
tgttcttttg
     1681 catatttcct tcattctaaa gttgttccct ggccgggagc gttggctttc
gcctgtaatc
     1741 ccaacacttt gggaggccag gacagatcac ttgaggtcag gagttcgaga
ccagcccagc
     1801 caacatggcg aaaccatgtc tctactaaaa atacaaaaat tatggtgacg
cctgcctgta
     1861 gtcccagcta ctcgggaggc tgaggcagga ggatcgcttg aacccaggaa
gtggagactg
     1921 cagtgagccg atatcgcacc acagcgctcc agcctggtcg acagagtgag
actccatctc
     1981 aagaaaaaat aaaaataaag ttgttctctg aagagcaaat gtctcattcc
agtaatgacc
     2041 cactcagcag gaatatggtg gagttcagtc caattcaggt cagccatatc
caaaagacca
     2101 caagtcatta ctaagttgag caaaagagtt tttatctatt agcagaaagg
gcctctctgg
     2161 cagcagagat taaaaactgg cccaacttca tttccatact tcagggaaca
gcaaattgag
     2221 gatttactta tctaggactt gaattccttc tttgggacca agttaataaa
agaccaagaa
     2281 actcctgatt aaactggata atgaaggatt ctgtagacag ggctgcacgt
atcggctttg
     2341 tttgacttct cttttctcag ttaacatctc agagctagaa cattccacat
tccccagcag
     2401 cgtgtggggg ctgactaaag tttacaattc caactaaaaa tcaccctgct
tctggcttat
     2461 ctgaatccct tacccacccc accccaccac cctactccta tttattcagc
accacactac
     2521 ccaggaaata cactagcaaa ttgtgcaatg gaataaaatc cacactttag
attcttgcaa
     2581 ctgtatcata tgtaatagta tcacttttc tacattttgg tcaaataaat
ttttacataa
           2641 actac
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_003281.1

LOCUS NP_003281

ACCESSION NP_003281

```
      1 maglnsleav krkiqalqqq adeaedraqg lqreldgere rrekaegdva
alnrriqlve
       61 eeldraqerl atalqkleea ekaadeserg mkvienramk deekmeiqem
qlkeakhiae
      121 eadrkyeeva rklvilegel eraeeraevs elkcgdleee lknvtnnlks
leaasekyse
      181 kedkyeeeik llsdklkeae traefaertv aklektiddl eeklaqakee
nvglhqtldq
```

241 tlnelnci

ETFA

Official Symbol: ETFA

Official Name: electron-transfer-flavoprotein, alpha polypeptide

Gene ID: 2108

Organism: Homo sapiens

Other Aliases: EMA, GA2, MADD

Other Designations: alpha-ETF; electron transfer flavoprotein alpha-subunit; electron transfer flavoprotein subunit alpha, mitochondrial; electron transfer flavoprotein, alpha polypeptide; glutaric aciduria II

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_000126.3

LOCUS NM_000126

ACCESSION NM_000126

```
      1 attaggtgac tggctgaggc ggcgccagtt ggccgggcac ggggctgctg
taaggccgag
       61 gttgcggcgg aagcggagac catgttccga gcggcggctc cggggcagct
ccggcgggcg
      121 gcctcattgc tacgatttca gagtaccctg gtaatagctg agcatgcaaa
tgattcccta
      181 gcacccatta ctttaaatac cattactgca gccacacgcc ttggaggtga
agtgtcctgc
      241 ttagtagctg gaaccaaatg tgacaaggtg gcacaagatc tctgtaaagt
agcaggcata
      301 gcaaaagttc tggtggctca gcatgatgtg tacaaaggcc tacttccaga
ggaactgaca
      361 ccattgattt tggcaactca gaagcagttc aattacacac acatctgtgc
tggagcatct
      421 gccttcggaa agaacctttt gcccagagta gcagccaaac ttgaggttgc
cccgatttct
      481 gacatcattg caatcaagtc acctgacaca tttgtgagaa ctatttatgc
aggaaatgct
      541 ctatgtacag tgaagtgtga tgagaaagtg aaagtgtttt ctgtccgtgg
aacatccttt
      601 gatgctgcag caacaagtgg cggtagtgcc agttcagaaa aggcatcaag
tacttcacca
      661 gtggaaatat cagagtggct tgaccagaaa ttaacaaaaa gtgatcgacc
agagctaaca
      721 ggtgccaaag tggtggtatc tggtggtcga ggcttgaaga gtggagagaa
ctttaagttg
      781 ttatatgact tggcagatca actacatgct gcagttggtg cttcccgtgc
tgctgttgat
      841 gctggctttg ttcccaatga catgcaagtt ggacagacgg gaaaaatagt
agcaccagaa
```

```
      901 ctttatattg ctgttggaat atctggagcc atccaacatt tagctgggat
gaaagacagc
      961 aagacaattg tggcaattaa taaagaccca gaagctccaa ttttccaagt
ggcagattat
     1021 ggaatagttg cagatttatt taaggtagtt cctgaaatga ctgagatatt
gaagaaaaaa
     1081 tgaatcagga tcatgcctta aaaagaaaac ttttgttaaa gtattccact
gaaatcacag
     1141 atatttgtgg gtattataac aatcattgga aagcatggag agctacattt
cataatttga
     1201 gggaaaattt ctaacagatg ccagaatgct tgtttatggg attgctgtgt
ttccttttaa
     1261 ttatttgtgg ttccaaacaa ttattgtttg aacttttaa attctgtact
aaaatctata
     1321 ataaagcttt tccacagctt taaaactatc agaaaaaaaa aaaaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_000117.1

LOCUS NP_000117

ACCESSION NP_000117

```
      1 mfraaapgql rraasllrfq stlviaehan dslapitlnt itaatrlgge
vsclvagtkc
       61 dkvaqdlckv agiakvlvaq hdvykgllpe eltplilatq kqfnythica
gasafgknll
      121 prvaakleva pisdiiaiks pdtfvrtiya gnalctvkcd ekvkvfsvrg
tsfdaaatsg
      181 gsassekass tspveisewl dqkltksdrp eltgakvvvs ggrglksgen
fkllydladq
      241 lhaavgasra avdagfvpnd mqvgqtgkiv apelyiavgi sgaiqhlagm
kdsktivain
      301 kdpeapifqv adygivadlf kvvpemteil kkk
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001127716.1

LOCUS NM_001127716

ACCESSION NM_001127716

```
      1 attaggtgac tggctgaggc ggcgccagtt ggccgggcac ggggctgctg
taaggccgag
       61 gttgcggcgg aagcggagac catgttccga gcggcggctc cggggcagct
ccggcgggcg
      121 gtggcacaag atctctgtaa agtagcaggc atagcaaaag ttctggtggc
tcagcatgat
      181 gtgtacaaag gcctacttcc agaggaactg acaccattga ttttggcaac
tcagaagcag
      241 ttcaattaca cacacatctg tgctggagca tctgccttcg gaaagaacct
tttgcccaga
      301 gtagcagcca aacttgaggt tgccccgatt tctgacatca ttgcaatcaa
gtcacctgac
```

```
      361 acatttgtga gaactattta tgcaggaaat gctctatgta cagtgaagtg
tgatgagaaa
      421 gtgaaagtgt tttctgtccg tggaacatcc tttgatgctg cagcaacaag
tggcggtagt
      481 gccagttcag aaaaggcatc aagtacttca ccagtggaaa tatcagagtg
gcttgaccag
      541 aaattaacaa aaagtgatcg accagagcta acaggtgcca aagtggtggt
atctggtggt
      601 cgaggcttga agagtggaga gaactttaag ttgttatatg acttggcaga
tcaactacat
      661 gctgcagttg gtgcttcccg tgctgctgtt gatgctggct ttgttcccaa
tgacatgcaa
      721 gttggacaga cgggaaaaat agtagcacca gaactttata ttgctgttgg
aatatctgga
      781 gccatccaac atttagctgg gatgaaagac agcaagacaa ttgtggcaat
taataaagac
      841 ccagaagctc caattttcca agtggcagat tatggaatag ttgcagattt
atttaaggta
      901 gttcctgaaa tgactgagat attgaagaaa aaatgaatca ggatcatgcc
ttaaaaagaa
      961 aacttttgtt aaagtattcc actgaaatca cagatatttg tgggtattat
aacaatcatt
     1021 ggaaagcatg gagagctaca tttcataatt tgagggaaaa tttctaacag
atgccagaat
     1081 gcttgtttat gggattgctg tgtttccttt taattatttg tggttccaaa
caattattgt
     1141 ttgaactttt taaattctgt actaaaatct ataataaagc ttttccacag
ctttaaaact
      1201 atcagaaaaa aaaaaaaaaa aa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001121188.1

LOCUS NP_001121188

ACCESSION NP_001121188

```
      1 mfraaapgql rravaqdlck vagiakvlva qhdvykgllp eeltplilat
qkqfnythic
       61 agasafgknl lprvaaklev apisdiiaik spdtfvrtiy agnalctvkc
dekvkvfsvr
      121 gtsfdaaats ggsassekas stspveisew ldqkltksdr peltgakvvv
sggrglksge
      181 nfkllydlad qlhaavgasr aavdagfvpn dmqvgqtgki vapelyiavg
isgaiqhlag
      241 mkdsktivai nkdpeapifq vadygivadl fkvvpemtei lkkk
```

RPL8

Official Symbol: RPL8

Official Name: ribosomal protein L8

Gene ID: 6132

Organism: Homo sapiens

Other Aliases: L8

Other Designations: 60S ribosomal protein L8

Nucleotide sequence (variant 1):

NCBI Reference Sequence : NM_000973.3

LOCUS NM_000973

ACCESSION NM_000973

```
      1 agataaggcc gctcgctgac gccgtgtttc ctctttcggc cgcgctggtg
aacaggaccc
       61 gtcgccatgg gccgtgtgat ccgtggacag aggaagggcg ccgggtctgt
gttccgcgcg
      121 cacgtgaagc accgtaaagg cgctgcgcgc ctgcgcgccg tggatttcgc
tgagcggcac
      181 ggctacatca agggcatcgt caaggacatc atccacgacc cgggccgcgg
cgcgcccctc
      241 gccaaggtgg tcttccggga tccgtatcgg tttaagaagc ggacggagct
gttcattgcc
      301 gccgagggca ttcacacggg ccagtttgtg tattgcggca agaaggccca
gctcaacatt
      361 ggcaatgtgc tccctgtggg caccatgcct gagggtacaa tcgtgtgctg
cctggaggag
      421 aagcctggag accgtggcaa gctggcccgg gcatcaggga actatgccac
cgttatctcc
      481 cacaaccctg agaccaagaa gacccgtgtg aagctgccct ccggctccaa
gaaggttatc
      541 tcctcagcca acagagctgt ggttggtgtg gtggctggag gtggccgaat
tgacaaaccc
      601 atcttgaagg ctggccgggc gtaccacaaa tataaggcaa agaggaactg
ctggccacga
      661 gtacggggtg tggccatgaa tcctgtggag catccttttg gaggtggcaa
ccaccagcac
      721 atcggcaagc cctccaccat ccgcagagat gcccctgctg gccgcaaagt
gggtctcatt
      781 gctgcccgcc ggactggacg tctccgggga accaagactg tgcaggagaa
agagaactag
      841 tgctgagggc ctcaataaag tttgtgttta tgccaaaaaa aaaaaaaaaa
aaaaaaaaaa
      901 aaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_000964.1

LOCUS NP_000964

ACCESSION NP_000964

```
      1 mgrvirgqrk gagsvfrahv khrkgaarlr avdfaerhgy ikgivkdiih
dpgrgaplak
       61 vvfrdpyrfk krtelfiaae gihtgqfvyc gkkaqlnign vlpvgtmpeg
tivccleekp
      121 gdrgklaras gnyatvishn petkktrvkl psgskkviss anravvgvva
gggridkpil
```

181 kagrayhkyk akrncwprvr gvamnpvehp fgggnhqhig kpstirrdap agrkvgliaa
241 rrtgrlrgtk tvqeken

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_033301.1

LOCUS NM_033301

ACCESSION NM_033301

1 gcggcatggg cagtatccgc cgccatcctc ttccgtgagg cgcgctgaga cccggaccgg
61 ccctcctgag aggatgccgg tgcgggcgcc cgcggagagg gacccgtcgc catgggccgt
121 gtgatccgtg gacagaggaa gggcgccggg tctgtgttcc gcgcgcacgt gaagcaccgt
181 aaaggcgctg cgcgcctgcg cgccgtggat ttcgctgagc ggcacggcta catcaagggc
241 atcgtcaagg acatcatcca cgacccgggc cgcggcgcgc ccctcgccaa ggtggtcttc
301 cgggatccgt atcggtttaa gaagcggacg gagctgttca ttgccgccga gggcattcac
361 acgggccagt ttgtgtattg cggcaagaag gcccagctca acattggcaa tgtgctccct
421 gtgggcacca tgcctgaggg tacaatcgtg tgctgcctgg aggagaagcc tggagaccgt
481 ggcaagctgg cccgggcatc agggaactat gccaccgtta tctcccacaa ccctgagacc
541 aagaagaccc gtgtgaagct gccctccggc tccaagaagg ttatctcctc agccaacaga
601 gctgtggttg gtgtggtggc tggaggtggc cgaattgaca aacccatctt gaaggctggc
661 cgggcgtacc acaaatataa ggcaaagagg aactgctggc cacgagtacg gggtgtggcc
721 atgaatcctg tggagcatcc ttttggaggt ggcaaccacc agcacatcgg caagccctcc
781 accatccgca gagatgcccc tgctggccgc aaagtgggtc tcattgctgc ccgccggact
841 ggacgtctcc ggggaaccaa gactgtgcag gagaaagaga actagtgctg agggcctcaa
901 taaagtttgt gtttatgcca aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
961 aaaaaaa

Protein sequence (variant 2):

NCBI Reference Sequence: NP_150644.1

LOCUS NP_150644

ACCESSION NP_150644

1 mgrvirgqrk gagsvfrahv khrkgaarlr avdfaerhgy ikgivkdiih dpgrgaplak
61 vvfrdpyrfk krtelfiaae gihtgqfvyc gkkaqlnign vlpvgtmpeg tivccleekp

```
      121 gdrgklaras gnyatvishn petkktrvkl psgskkviss anravvgvva
gggridkpil
      181 kagrayhkyk akrncwprvr gvamnpvehp fgggnhqhig kpstirrdap
agrkvgliaa
      241 rrtgrlrgtk tvqeken
```

ARCN1

Official Symbol: ARCN1

Official Name: archain 1

Gene ID: 372

Organism: Homo sapiens

Other Aliases: COPD

Other Designations: archain vesicle transport protein 1; coatomer delta subunit; coatomer protein complex, subunit delta; coatomer protein delta-COP; coatomer subunit delta; delta-COP; delta-coat protein

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001655.4

LOCUS NM_001655

ACCESSION NM_001655

```
      1 gaagacgtgg cttggggccg ccatcttggc aagaggcgaa gcggcagcgg
ttcctgtcaa
       61 gggggcagca ggtccagagc tgctggtgct cccgttcccc agaccctacc
cctatcccca
      121 gtggagccgg agtgcgggcg cgccccacca ccgccctcac catggtgctg
ttggcagcag
      181 cggtctgcac aaaagcagga aaggctattg tttctcgaca gtttgtggaa
atgacccgaa
      241 ctcggattga gggcttatta gcagcttttc caaagctcat gaacactgga
aaacaacata
      301 cgtttgttga aacagagagt gtaagatatg tctaccagcc tatggagaaa
ctgtatatgg
      361 tactgatcac taccaaaaac agcaacattt tagaagattt ggagacccta
aggctcttct
      421 caagagtgat ccctgaatat tgccgagcct tagaagagaa tgaaatatct
gagcactgtt
      481 ttgatttgat ttttgctttt gatgaaattg tcgcactggg ataccgggag
aatgttaact
      541 tggcacagat cagaaccttc acagaaatgg attctcatga ggagaaggtg
ttcagagccg
      601 tcagagagac tcaagaacgt gaagctaagg ctgagatgcg tcgtaaagca
aaggaattac
      661 aacaggcccg aagagatgca gagagacagg gcaaaaaagc accaggattt
ggcggatttg
      721 gcagctctgc agtatctgga ggcagcacag ctgccatgat cacagagacc
atcattgaaa
```

781 ctgataaacc aaaagtggca cctgcaccag ccaggccttc aggccccagc
aaggctttaa
841 aacttggagc caaaggaaag gaagtagata actttgtgga caaattaaaa
tctgaaggtg
901 aaaccatcat gtcctctagt atgggcaagc gtacttctga agcaaccaaa
atgcatgctc
961 cacccattaa tatggaaagt gtacatatga agattgaaga aaagataaca
ttaacctgtg
1021 gacgagacgg aggattacag aatatggagt tgcatggcat gatcatgctt
aggatctcag
1081 atgacaagta tggccgaatt cgtcttcatg tggaaaatga agataagaaa
ggggtgcagc
1141 tacagaccca tccaaatgtg gataaaaaac ttttcactgc agagtctcta
attggcctga
1201 agaatccaga gaagtcattt ccagtcaaca gtgacgtagg ggtgctaaag
tggagactac
1261 aaaccacaga ggaatctttt attccactga caattaattg ctggccctcg
gagagtggaa
1321 atggctgtga tgtcaacata gaatatgagc tacaagaaga taatttagaa
ctgaatgatg
1381 tggttatcac catcccactc ccgtctggtg tcggcgcgcc tgttatcggt
gagatcgatg
1441 gggagtatcg acatgacagt cgacgaaata ccctggagtg gtgcctgcct
gtgattgatg
1501 ccaaaaataa gagtggcagc ctggagttta gcattgctgg gcagcccaat
gacttcttcc
1561 ctgttcaagt ttcctttgtc tccaagaaaa attactgtaa catacaggtt
accaaagtga
1621 cccaggtaga tggaaacagc cccgtcaggt tttccacaga gaccactttc
ctagtggata
1681 agtatgaaat tctgtaatac caagaagagg gagctgaaaa ggaaaatttt
cagattaata
1741 aagaagacgc caatgatggc tgaagagttt ttcccagatt tacaagccac
tggagacccc
1801 ttttttctga tacaatgcac gattctctgc gcgcaaggac cctcgactca
cccccatgtt
1861 tcagtgtcac agagacattc tttgataagg aaatggcaca aacataaagg
gaaaggctgc
1921 taattttctt tggcagattg tattggccag caggaaagca agctctccag
agaatgcccc
1981 cagttaaata cctcctctac ctttacctaa gttgctcctt tatttttatt
ttattattat
2041 tattattatt attattattt tttgagatgg agtctcactt tgtaacccag
gctggaatgc
2101 aatggcatga tctcagctca ctgcaacctc cgcctcctgg gttcaagcaa
gtctcctgcc
2161 tcagcctccg agtagctggg actacaggtg cacgccacca cgcctggcta
atttttttgta
2221 ttttagtaga gacggggttt caccgtgttg cccaggctgg tcgcgaactc
ctgagctcag
2281 gcaatccgcc cacctcagcc tcccaaagtg ttgggattac aggcatgagc
caccatgccc
2341 agctgctcct ttattttaat ccctaaatat aatccctaaa tatagttata
tttcatactt
2401 agtttgtttt taaaaagttt tctctgtaga aaattttaat cattcatacc
ctttaccttt
2461 aggtttttct ttctatacat tcagtcaggc actgggatca tctgtttaca
ggcattatat
2521 ttatttggca ctcctggaac aagtatatct aacccattct tgatttttgg
actattcagg

```
     2581 tgaactattt gaggggtatg gggtctagaa gttaaaagat acgcatgtct
tctgttcttt
     2641 tcccgtatca attcattcct tcatctcttt gccaagttgt tttcctttca
gggcctgtcc
     2701 ttccagttta gaacagtacc atgaatccca cttgtgtcaa tattaaagat
agctgagaag
     2761 caccttttcaa atggcacagt ccctcttcaa gatgtctaaa agaatggtta
tgtctgtcca
     2821 gttagggatt tcacatccac atgtaatcat gtctgctgct gttgctaccc
aaattttcat
     2881 ttctccacat tttgggtact taagctaaaa cgtaatggcc acagtctgta
atccattcac
     2941 attcctcagt ttcaccacct ccctcttcca gactgcactc tctgtcatca
gtcccctcct
     3001 ttctaacaga aatggggtta tgattttgaa ggctgtgggt tcagggagtc
tttgccaatc
     3061 ctgttggccc taaactatca aggaggctcc atttcaccat ttgatttttt
gcatttcagg
     3121 aggcaactga ttgtttcgat atgtacatat tactcacgta taccccattt
ccttccagtc
     3181 agcccaacat tttccaccag tctgtcccca tctctgaaat ccttccttct
ctttcccccct
     3241 aagtcttttg agtgtcatca tgtactggtg gtttctcggt tccatctcat
ccatttcctt
     3301 ttcaatggag actacagcgt cagccagctc agccttggct tttaactcaa
tattccagtc
     3361 cataggggtg gttaaaagtt gctgcaaggc tgcaggcact ggcagtggga
agaggcagac
     3421 gactagatga cttctgcact tttagctggt tgaaaagtac cactcccact
ctgaacatct
     3481 ggccgtccct gcaaagagtg tactgtgctt gaagcagagc actcacacat
aaatggctgt
     3541 gtgtggaatt gcttgccaaa gaagtttcta gcctttccct ttcccctaac
tgcatcaggg
     3601 aagaattctt atctctagct tggtttccac atgaggtttt tctgagaagg
gcttgggaca
     3661 agaagtctgt catgttagtt aagcaggcaa gaaatcctac taatccagtt
ttgtttgaaa
     3721 gttgtttgtc cgtatgattt tttaaaagtc aagtttaatt tcaaaaaacc
ttttttttct
     3781 gagattactt ttggggtaat atttaaaatg agagacattt tgtaaccctg
taaaatacat
     3841 agggaatata acattccagt gtatacaaag aaggcaaatt ctttaatcaa
ataaagcgca
     3901 ttataaaatg agatgtttat tggattattg actcactttg gtgtctgctt
gttgattcag
     3961 gatgctgtaa tgggacctaa cattaaaaat taatgacatg ttttttttaa
gagaaaaaaa
      4021 aaaaaaaaaa aa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001646.2

LOCUS NP_001646

ACCESSION NP_001646

```
     1 mvllaaavct kagkaivsrq fvemtrtrie gllaafpklm ntgkqhtfve
tesvryvyqp
```

```
       61 meklymvlit tknsniledl etlrlfsrvi peycraleen eisehcfdli
fafdeivalg
      121 yrenvnlaqi rtftemdshe ekvfravret qereakaemr rkakelqqar
rdaerqgkka
      181 pgfggfgssa vsggstaami tetiietdkp kvapaparps gpskalklga
kgkevdnfvd
      241 klksegetim sssmgkrtse atkmhappin mesvhmkiee kitltcgrdg
glqnmelhgm
      301 imlrisddky grirlhvene dkkgvqlqth pnvdkklfta esliglknpe
ksfpvnsdvg
      361 vlkwrlqtte esfipltinc wpsesgngcd vnieyelqed nlelndvvit
iplpsgvgap
      421 vigeidgeyr hdsrrntlew clpvidaknk sgslefsiag qpndffpvqv
sfvskknycn
      481 iqvtkvtqvd gnspvrfste ttflvdkyei l
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001142281.1

LOCUS NM_001142281

ACCESSION NM_001142281

```
      1 gaagacgtgg cttggggccg ccatcttggc aagaggcgaa gcggcagcgg
ttcctgtcaa
       61 gggggcagca ggtccagagc tgctggtgct cccgttcccc agaccctacc
cctatcccca
      121 gtggagccgg agtgcgggcg cgccccacca ccgccctcac catgatccct
gaatattgcc
      181 gagccttaga agagaatgaa atatctgagc actgttttga tttgattttt
gcttttgatg
      241 aaattgtcgc actgggatac cgggagaatg ttaacttggc acagatcaga
accttcacag
      301 aaatggattc tcatgaggag aaggtgttca gagccgtcag agagactcaa
gaacgtgaag
      361 ctaaggctga gatgcgtcgt aaagcaaagg aattacaaca ggcccgaaga
gatgcagaga
      421 gacagggcaa aaaagcacca ggatttggcg gatttggcag ctctgcagta
tctggaggca
      481 gcacagctgc catgatcaca gagaccatca ttgaaactga taaaccaaaa
gtggcacctg
      541 caccagccag gccttcaggc cccagcaagg ctttaaaact tggagccaaa
ggaaaggaag
      601 tagataactt tgtggacaaa ttaaaatctg aaggtgaaac catcatgtcc
tctagtatgg
      661 gcaagcgtac ttctgaagca accaaaatgc atgctccacc cattaatatg
gaaagtgtac
      721 atatgaagat tgaagaaaag ataacattaa cctgtggacg agacggagga
ttacagaata
      781 tggagttgca tggcatgatc atgcttagga tctcagatga caagtatggc
cgaattcgtc
      841 ttcatgtgga aaatgaagat aagaaagggg tgcagctaca gacccatcca
aatgtggata
      901 aaaaactttt cactgcagag tctctaattg gcctgaagaa tccagagaag
tcatttccag
      961 tcaacagtga cgtaggggtg ctaaagtgga gactacaaac cacagaggaa
tcttttattc
     1021 cactgacaat taattgctgg ccctcggaga gtggaaatgg ctgtgatgtc
aacatagaat
```

1081 atgagctaca agaagataat ttagaactga atgatgtggt tatcaccatc
ccactcccgt
1141 ctggtgtcgg cgcgcctgtt atcggtgaga tcgatgggga gtatcgacat
gacagtcgac
1201 gaaataccct ggagtggtgc ctgcctgtga ttgatgccaa aaataagagt
ggcagcctgg
1261 agtttagcat tgctgggcag cccaatgact tcttccctgt tcaagtttcc
tttgtctcca
1321 agaaaaatta ctgtaacata caggttacca aagtgaccca ggtagatgga
aacagccccg
1381 tcaggttttc cacagagacc actttcctag tggataagta tgaaattctg
taataccaag
1441 aagagggagc tgaaaaggaa aattttcaga ttaataaaga agacgccaat
gatggctgaa
1501 gagtttttcc cagatttaca agccactgga gacccctttt ttctgataca
atgcacgatt
1561 ctctgcgcgc aaggaccctc gactcacccc catgtttcag tgtcacagag
acattctttg
1621 ataaggaaat ggcacaaaca taaagggaaa ggctgctaat tttctttggc
agattgtatt
1681 ggccagcagg aaagcaagct ctccagagaa tgcccccagt taaatacctc
ctctaccttt
1741 acctaagttg ctcctttatt tttattttat tattattatt attattatta
ttatttttg
1801 agatggagtc tcactttgta acccaggctg gaatgcaatg gcatgatctc
agctcactgc
1861 aacctccgcc tcctgggttc aagcaagtct cctgcctcag cctccgagta
gctgggacta
1921 caggtgcacg ccaccacgcc tggctaattt tttgtatttt agtagagacg
gggtttcacc
1981 gtgttgccca ggctggtcgc gaactcctga gctcaggcaa tccgcccacc
tcagcctccc
2041 aaagtgttgg gattacaggc atgagccacc atgcccagct gctcctttat
tttaatccct
2101 aaatataatc cctaaatata gttatatttc atacttagtt tgtttttaaa
aagttttctc
2161 tgtagaaaat tttaatcatt catacccttt acctttaggt ttttctttct
atacattcag
2221 tcaggcactg ggatcatctg tttacaggca ttatatttat ttggcactcc
tggaacaagt
2281 atatctaacc cattcttgat ttttggacta ttcaggtgaa ctatttgagg
ggtatggggt
2341 ctagaagtta aaagatacgc atgtcttctg ttcttttccc gtatcaattc
attccttcat
2401 ctctttgcca agttgttttc ctttcagggc ctgtccttcc agtttagaac
agtaccatga
2461 atcccacttg tgtcaatatt aaagatagct gagaagcacc tttcaaatgg
cacagtccct
2521 cttcaagatg tctaaaagaa tggttatgtc tgtccagtta gggatttcac
atccacatgt
2581 aatcatgtct gctgctgttg ctacccaaat tttcatttct ccacattttg
ggtacttaag
2641 ctaaaacgta atggccacag tctgtaatcc attcacattc ctcagtttca
ccacctccct
2701 cttccagact gcactctctg tcatcagtcc cctcctttct aacagaaatg
gggttatgat
2761 tttgaaggct gtgggttcag ggagtctttg ccaatcctgt tggccctaaa
ctatcaagga
2821 ggctccattt caccatttga ttttttgcat ttcaggaggc aactgattgt
ttcgatatgt

```
     2881 acatattact cacgtatacc ccatttcctt ccagtcagcc caacattttc
caccagtctg
     2941 tccccatctc tgaaatcctt ccttctcttt ccccctaagt cttttgagtg
tcatcatgta
     3001 ctggtggttt ctcggttcca tctcatccat ttccttttca atggagacta
cagcgtcagc
     3061 cagctcagcc ttggctttta actcaatatt ccagtccata ggggtggtta
aaagttgctg
     3121 caaggctgca ggcactggca gtgggaagag gcagacgact agatgacttc
tgcactttta
     3181 gctggttgaa aagtaccact cccactctga acatctggcc gtccctgcaa
agagtgtact
     3241 gtgcttgaag cagagcactc acacataaat ggctgtgtgt ggaattgctt
gccaaagaag
     3301 tttctagcct ttccctttcc cctaactgca tcagggaaga attcttatct
ctagcttggt
     3361 ttccacatga ggtttttctg agaagggctt gggacaagaa gtctgtcatg
ttagttaagc
     3421 aggcaagaaa tcctactaat ccagttttgt ttgaaagttg tttgtccgta
tgatttttta
     3481 aaagtcaagt ttaatttcaa aaaacctttt ttttctgaga ttacttttgg
ggtaatattt
     3541 aaaatgagag acattttgta accctgtaaa atacataggg aatataacat
tccagtgtat
     3601 acaaagaagg caaattcttt aatcaaataa agcgcattat aaaatgagat
gtttattgga
     3661 ttattgactc actttggtgt ctgcttgttg attcaggatg ctgtaatggg
acctaacatt
      3721 aaaaattaat gacatgtttt ttttaagaga aaaaaaaaaa aaaaaaaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001135753.1

LOCUS NP_001135753

ACCESSION NP_001135753

```
      1 mipeycrale eneisehcfd lifafdeiva lgyrenvnla qirtftemds
heekvfravr
       61 etqereakae mrrkakelqq arrdaerqgk kapgfggfgs savsggstaa
mitetiietd
      121 kpkvapapar psgpskalkl gakgkevdnf vdklkseget imsssmgkrt
seatkmhapp
      181 inmesvhmki eekitltcgr dgglqnmelh gmimlrisdd kygrirlhve
nedkkgvqlq
      241 thpnvdkklf taesliglkn peksfpvnsd vgvlkwrlqt teesfiplti
ncwpsesgng
      301 cdvnieyelq ednlelndvv itiplpsgvg apvigeidge yrhdsrrntl
ewclpvidak
      361 nksgslefsi agqpndffpv qvsfvskkny cniqvtkvtq vdgnspvrfs
tettflvdky
      421 eil
```

DDX18

Official Symbol: DDX18

Official Name: DEAD (Asp-Glu-Ala-Asp) box polypeptide 18

Gene ID: 8886

Organism: Homo sapiens

Other Aliases: MrDb

Other Designations: ATP-dependent RNA helicase **DDX18**; DEAD box protein 18; DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated); Myc-regulated DEAD box protein

Nucleotide sequence:

NCBI Reference Sequence: NM_006773.3

LOCUS NM_006773

ACCESSION NM_006773

```
      1 gccgagctgc gcacgtgcgg ccggaaggga agtaacgtca gcctgagaac
tgagtagctg
       61 tactgtgtgg cgccttattc taggcacttg ttgggcagaa tgtcacacct
gccgatgaaa
      121 ctcctgcgta agaagatcga gaagcggaac ctcaaattgc ggcagcggaa
cctaaagttt
      181 caggggggcct caaatctgac cctatcggaa actcaaaatg gagatgtatc
tgaagaaaca
      241 atgggaagta gaaaggttaa aaaatcaaaa caaaagccca tgaatgtggg
cttatcagaa
      301 actcaaaatg gaggcatgtc tcaagaagca gtgggaaata taaaagttac
aaagtctccc
      361 cagaaatcca ctgtattaac caatggagaa gcagcaatgc agtcttccaa
ttcagaatca
      421 aaaaagaaaa agaagaaaaa gagaaaaatg gtgaatgatg ctgagcctga
tacgaaaaaa
      481 gcaaaaactg aaaacaaagg gaaatctgaa gaagaaagtg ccgagactac
taaagaaaca
      541 gaaaataatg tggagaagcc agataatgat gaagatgaga gtgaggtgcc
cagtctgccc
      601 ctgggactga caggagcttt tgaggatact tcgtttgctt ctctatgtaa
tcttgtcaat
      661 gaaaacactc tgaaggcaat aaaagaaatg ggttttacaa acatgactga
aattcagcat
      721 aaaagtatca gaccacttct ggaaggcagg gatcttctag cagctgcaaa
aacaggcagt
      781 ggtaaaaccc tggcttttct catccctgca gttgaactca ttgttaagtt
aaggttcatg
      841 cccaggaatg gaacaggagt ccttattctc tcacctacta gagaactagc
catgcaaacc
      901 tttggtgttc ttaaggagct gatgactcac cacgtgcata cctatggctt
gataatgggt
      961 ggcagtaaca gatctgctga agcacagaaa cttggtaatg ggatcaacat
cattgtggcc
     1021 acaccaggcc gtctgctgga ccatatgcag aataccccag gatttatgta
taaaaacctg
     1081 cagtgtctgg ttattgatga agctgatcgt atcttggatg tggggtttga
agaggaatta
```

1141 aagcaaatta ttaaactttt gccaacacgt agacagacta tgctcttttc
tgccacccaa
1201 actcgaaaag ttgaagacct ggcaaggatt tctctgaaaa aggagccatt
gtatgttggc
1261 gttgatgatg ataaagcgaa tgcaacagtg gatggtcttg aacagggata
tgttgtttgt
1321 ccttctgaaa agagattcct tctgctcttt acattcctta agaagaaccg
aaagaagaag
1381 cttatggtct tcttttcatc ttgtatgtct gtgaaatacc actatgagtt
gctgaactac
1441 attgatttgc ccgtcttggc cattcatgga aagcaaaagc aaaataagcg
tacaaccaca
1501 ttcttccagt tctgcaatgc agattcggga acactattgt gtacggatgt
ggcagcgaga
1561 ggactagaca ttcctgaagt cgactggatt gttcagtatg accctccgga
tgaccctaag
1621 gaatatattc atcgtgtggg tagaacagcc agaggcctaa atgggagagg
gcatgccttg
1681 ctcattttgc gcccagaaga attgggtttt cttcgctact tgaaacaatc
caaggttcca
1741 ttaagtgaat ttgacttttc ctggtctaaa atttctgaca ttcagtctca
gcttgagaaa
1801 ttgattgaaa agaattactt tcttcataag tcagcccagg aagcatataa
gtcatacata
1861 cgagcctatg attcccattc tctgaaacag atctttaatg ttaataacct
aaatttgcct
1921 caggttgctc tgtcatttgg tttcaaggtg cctcccttcg ttgatctgaa
cgtcaacagt
1981 aatgaaggca agcagaaaaa gcgaggaggt ggtggtggat ttggctacca
gaaaaccaag
2041 aaagttgaga aatccaaaat ctttaaacac attagcaaga aatcatctga
cagcaggcag
2101 ttctctcact gaacacatgc cttcctttca tcttgaataa ctttgtccta
aaatgaattt
2161 tttttcccct tgatttaaca ggatttttgt agactttaga atttggactt
acctaacaag
2221 agtataaatt gacttgggtt gcaagcactg agcactgtta cttctatcac
gtctctcttt
2281 tatttctggg atataaaaca ggctttaagt ttcttggttg cccaagggca
gagcaaggaa
2341 tatctggtgt ttcttgtgat gataatattt taattttaaa tatccctccc
tcatacaagt
2401 gtatgttacc attttaatat aattcttttt gtacctttcc ttcttgtttt
gcgaagattt
2461 ttgtggcatg gattgctgtg ctcactgctg taaaaggtga cctagtgtac
tgggcagctg
2521 gtggcggtgc agaaaagagt ctcaggttat tttttgtttt tagttatttc
ttggaccttg
2581 acagtatcta atgactcctc ctgaaaatgc tgcagtataa aagagcaaag
agctttggga
2641 aatacctaag aagcacctta agattagggt ggcattgctt ttatagattc
ttgattttaa
2701 agcaacaggc ctttctcagg tgttgcattt tttggagcaa aaactatggg
ttgtaatttg
2761 aataaagtgt cactaagcag ttataacgtt tgatggctgg ggggtaggaa
gaggatggaa
2821 ttgagatgtt tgagcctcat ttacatcaat agaggtgtaa tgtactgcat
ttcttcattt
2881 ggtaacataa caaagacttt catacaaaga acgatgatgc tcctcattaa
gatttgttta

2941 attcaaggtg gtttggattt ggtaagcctt tgcactctgt agagtactta
gaagacaagg
3001 gcaacttact tggagttaga gccaagctgt cagacggtgc ccagcacaca
ttaatgttag
3061 cttctttctg agaaaaaaat acctcttcca ggccctgaaa caaaaaatac
atttgctgtg
3121 aagattgaaa atgaacaaag ttagaaaaaa aaacagcaaa atcagtgatt
tagtcagatg
3181 agtttttcgt tgtaggagca cttgatttct agtgtgtttt gtacagtata
taactacaag
3241 atagtacatt ttgtagcagt tcaaagccaa agttgctagc atcattttgc
tgttgtgcca
3301 gttaatcata ggatcccatt aaataagtgt gctaacatcg aatatagaga
aaactggtaa
3361 agaacattcc agtaggaaaa gaaaagaaca atcttccatt tctgggcttg
gccaccatca
3421 ccctggtcgg acctgtcctg gacttccaac cttgactgct gagctcctgg
cttagcttct
3481 tgggttccta attcctggtg tttaataatt ctctccacga tcatgttttt
ctgatttttt
3541 ttttcagaaa taatgttttt taaaagacaa aaacaaaggg aagaatattt
aattactgag
3601 cagaagtaaa tactgttggc attttgtaca taatctaatt tttatatgca
tgttcatgct
3661 ttttaatttt tttatcaaaa attaagtcat ctacctacta cttgtaacca
gcttgtttca
3721 taacatgtta ttttcctgtg tcattaaata attacttcaa tgttgaaaaa
aaaaaaaaaa
3781 aaaaaaaaaa

Protein sequence:

NCBI Reference Sequence: NP_006764.3

LOCUS NP_006764

ACCESSION NP_006764

1 mshlpmkllr kkiekrnlkl rqrnlkfqga snltlsetqn gdvseetmgs
rkvkkskqkp
61 mnvglsetqn ggmsqeavgn ikvtkspqks tvltngeaam qssnseskkk
kkkkrkmvnd
121 aepdtkkakt enkgkseees aettketenn vekpdndede sevpslplgl
tgafedtsfa
181 slcnlvnent lkaikemgft nmteiqhksi rpllegrdll aaaktgsgkt
laflipavel
241 ivklrfmprn gtgvlilspt relamqtfgv lkelmthhvh tyglimggsn
rsaeaqklgn
301 giniivatpg rlldhmqntp gfmyknlqcl videadrild vgfeeelkqi
ikllptrrqt
361 mlfsatqtrk vedlarislk keplyvgvdd dkanatvdgl eqgyvvcpse
krflllftfl
421 kknrkkklmv ffsscmsvky hyellnyidl pvlaihgkqk qnkrtttffq
fcnadsgtll
481 ctdvaargld ipevdwivqy dppddpkeyi hrvgrtargl ngrghallil
rpeelgflry
541 lkqskvplse fdfswskisd iqsqleklie knyflhksaq eayksyiray
dshslkqifn
601 vnnlnlpqva lsfgfkvppf vdlnvnsneg kqkkrggggg fgyqktkkve
kskifkhisk

661 kssdsrqfsh

G3BP2

Official Symbol: G3BP2

Official Name:GTPase activating protein (SH3 domain) binding protein 2

Gene ID: GTPase activating protein (SH3 domain) binding protein 2

Organism: Homo sapiens

Other Aliases:

Other Designations: G3BP-2; GAP SH3 domain-binding protein 2; Ras-GTPase activating protein SH3 domain-binding protein 2; ras GTPase-activating protein-binding protein 2

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_203505.2

LOCUS NM_203505

ACCESSION NM_203505

1 gtgctcgggg gttccctggc cctttcggca ggggtaaaac aataagaggg ggcggtggca
61 aaggggggcgg gacgtccgtg gtccttgtcg cacgtcgcag cgcctggcgc ccgggaagag
121 gtggttgtga ggcagacgaa ctcgcggctc tccggcttcc gaggcttccg agttgtcgga
181 ggaaggggggc ggcgagcaat aagaacccgc cgcacccggt cctcagcgac tcttctgacc
241 tccgcgcgac gtacccgccg ccgccgttgg ctggagcatt tgacattgtg cagcaaagaa
301 atggttatgg agaagcccag tccgctgctt gtagggcggg agtttgtgag gcaatattat
361 actttgctga ataaagctcc ggaatattta cacaggtttt atggcaggaa ttcttcctat
421 gttcatggtg gagtagatgc tagtggaaag ccccaggaag ctgtttatgg ccaaaatgat
481 atacaccaca aagtattatc tctgaacttc agtgaatgtc atactaaaat tcgtcatgtg
541 gatgctcatg caaccttgag tgatggagta gttgtccagg tcatgggttt gctgtctaac
601 agtggacaac cagaaagaaa gtttatgcaa acctttgttc tggctcctga aggatctgtt
661 ccaaataaat tttatgttca caatgatatg tttcgttatg aagatgaagt gtttggtgat
721 tctgagcctg aacttgatga agaatcagaa gatgaagtag aagaggaaca agaagaaaga
781 caaccatctc ctgaacctgt gcaagaaaat gctaacagtg gttactatga agctcaccct

841 gtgactaatg gcatagagga gcctttggaa gaatcctctc atgaacctga
acctgagcca
901 gaatctgaaa caaagactga agagctgaaa ccacaagtgg aggagaagaa
cttagaagaa
961 ctagaggaga aatctactac tcctcctccg gcagaacctg tttctctgcc
acaagaacca
1021 ccaaaggctt tctcctgggc ttcagtgacc agtaaaaacc tgcctcctag
tggtactgtt
1081 tcttcctctg gaattccacc ccatgttaaa gcaccagtct cacagccaag
agtcgaagct
1141 aaaccagaag ttcaatctca gccacctcgt gtgcgtgaac aacgacctag
agaacgacct
1201 ggttttcctc ctagaggacc aagaccaggc agaggagata tggaacagaa
tgactctgac
1261 aaccgtagaa taattcgcta tccagatagt catcaacttt ttgttggtaa
cttgccacat
1321 gatattgatg aaaatgagct aaaggaattc ttcatgagtt ttggaaacgt
tgtggaactt
1381 cgcatcaata ccaagggtgt tgggggaaag cttccaaatt ttggttttgt
ggtttttgat
1441 gactctgaac cagttcagag aatcttaatt gcaaaaccga ttatgtttcg
aggggaagta
1501 cgtttaaatg tggaagagaa aaaaacaaga gctgcaagag agcgagaaac
cagaggtggt
1561 ggtgatgatc gcagggatat taggcgcaat gatcgaggtc ccggtggtcc
acgtggaatt
1621 gtgggtggtg gaatgatgcg tgatcgtgat ggaagaggac ctcctccaag
gggtggcatg
1681 gcacagaaac ttggctctgg aagaggaacc gggcaaatgg agggccgctt
cacaggacag
1741 cgtcgctgaa gctccactgt tggcaaagtc ttggcagtgg tacattattc
atcgtgtttg
1801 cattcttgtt aatttttttt ttggctttgg aatgtgacac agccttttg
atcatttctt
1861 tgatgtgaaa agcatctttg gttatcagtt aaattgaggt ggacattatt
tccccaattt
1921 cacaacagga ttcacattgt taatttataa atctagactt ggagaattaa
ggactgagaa
1981 atgaccatat cttaaactat ctacgacaaa gtgaacttaa aaggacatgc
ccactgaatt
2041 caggtccttt gagtaaaaaa aaaatcttct gctgcacatt ttgtttaagt
gttactgttt
2101 ctgcctgtta atgctgggaa cacaaatagt gcaatttgtg caattggaga
atcttgcctt
2161 ttttcttggc tccccccaaa aatacaaacc aacagaaact tgttatgcac
tcatcaaaat
2221 gtactaatgg gtactctgaa ctcattaaca ttgacatctg caacaggagg
caacagggaa
2281 aaaatctcat cttcttttcc agtagaaaat agtttgtgaa atgatgaggg
cattttatct
2341 gcttgctgtg accagcgtgt gtacacataa accttaacaa gactacaagt
atattccaga
2401 aggaaatcat tttagttatg aactaaataa taaaaattag aacttcaaat
gcgatggtct
2461 tgactattag accagattta gtagctccat atctaagatt tttctacctg
cccctcttca
2521 gtacagggat ggctggctgc tcaacacact cctcctcccc ttttttcctt
tctttaagct
2581 gtgtacagtg aaaattgtct ttactgtatt tttgttctct ggtaatgtaa
taagcatgat

2641 ggtgccttct attaatacat cattccagtc ttgctggtaa ttttgtacag
tatagtgtat
2701 gaattgctgt gctgcaaagc caaacagctg caaaatgttg aaaaatcatc
gaaatgtata
2761 aaaattgcag tatctttaaa atcagtaaaa tggactagca tattatttat
cttgttcttc
2821 agttaacaac tttgtgttct ctgtgggagg gagggagtcc tgtgtgtttg
tggggagagg
2881 gaaggaggaa gtcagttatt tgagtaagcc tctagttgac ttttctctta
gcctgaatgt
2941 ggacgttgaa acatatcact tcagggcttg gaaaagtcag tcaacttgac
gtacattttt
3001 agtgacattt taaaagcagt cagattctat aaatggcaag taagcctgaa
gtgaggatac
3061 tgcaattttc ggagaaaaga acagcagctc tttaagtgtt tgcattttct
atttgggggg
3121 cagggaactg tcattcattt tgcacaattc ttgaactgat gtcagcaccc
gagtggctcc
3181 tgaatttaag tctgggacga catcttttat ttttacatga atctttaaac
aattctgtga
3241 gcaaagtttg tagctgctgg attattgtct gtctttatag caagttccag
taaaccacaa
3301 gtatggcaaa gcttatccaa ttttatgctt ggagcagtca gtacatacca
gtttctgatg
3361 tttcaggcag gagtggggta aataagtgtg accacttaaa gctgctcgtt
agcatggaag
3421 acttctccat tctatctttg taaaacagac aagatatgca cttgacatag
tagcaaattg
3481 gttctgaatt atgcaactgt ttgctattta gtaaactagc aaatgatgca
tgtattttgt
3541 ttttcatgta ctgggcaata tgagtaaaat ctgtcccttt ttccccctttt
gaatgaggtc
3601 ttccatgttt gagggaaagt cttgcactat tgcatatatt ttggggacac
agattttcat
3661 agtttccatt tttgggggggc ttaaggattt tttttttttc tgtttgaaac
agttttatac
3721 tttctgatat agtacttgaa attcttacca gaaaattact ttggagtttt
gaagccttta
3781 ttaatactac ttttaaagaa gcagttgttt tattgtcaat gttttttttc
ccccaagcat
3841 attttcttgt atttctgttt ccatatatat atatatatat ataatttcca
attcaggata
3901 ttgccctgcc atccatgaaa actgttctgg caccaaaagt aatgacaaat
gttaagtgta
3961 ataatagaaa agtagagcaa agagccattc agcttcagtc tttacatacc
atgaataaaa
4021 cattaaaaca tcatatggag aagtttacat ggtgattgtt cacctgcagt
actgtggagt
4081 tttaacattt tgtcctcttt tcagtgaaac agagtaaaaa tattcatcta
ccattactgt
4141 tatttgctga ttttgtttta ttttttgatg gtaatattct atccttatga
cactattgca
4201 accaaattgg ctttaccatc ttggctttag taggtataga agacaatgga
ttaccatctt
4261 tattgctgta atgtgttaag cattatatgc tagtagaatc tagtttaatt
gtttcaggtg
4321 gaaagtattc tttgagtttc catattgaat gtgtttggac taaacaaaca
ataaactact
4381 gatgtctgca gcatttatct atgtccctaa

Protein sequence (variant 1):

NCBI Reference Sequence: NP_987101.1

LOCUS NP_987101

ACCESSION NP_987101

```
      1 mvmekpspll vgrefvrqyy tllnkapeyl hrfygrnssy vhggvdasgk
pqeavygqnd
       61 ihhkvlslnf sechtkirhv dahatlsdgv vvqvmgllsn sgqperkfmq
tfvlapegsv
      121 pnkfyvhndm fryedevfgd sepeldeese deveeeqeer qpspepvqen
ansgyyeahp
      181 vtngieeple esshepepep esetkteelk pqveeknlee leeksttppp
aepvslpqep
      241 pkafswasvt sknlppsgtv sssgipphvk apvsqprvea kpevqsqppr
vreqrprerp
      301 gfpprgprpg rgdmeqndsd nrriirypds hqlfvgnlph didenelkef
fmsfgnvvel
      361 rintkgvggk lpnfgfvvfd dsepvqrili akpimfrgev rlnveekktr
aareretrgg
      421 gddrrdirrn drgpggprgi vgggmmrdrd grgpppprggm aqklgsgrgt
gqmegrftgq
      481 rr
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_012297.4

LOCUS NM_012297

ACCESSION NM_012297

```
      1 acattccatt cgcgctccgc ggcgcgaggc aatcgtccgg tgtgtgagcc
cgggagccgg
       61 aggtgtagcg gcagagacat tgttcttgcc ggctccctac ggtgccgtgt
gtgcgtgaga
      121 gaagaccagt ctttcctcta gcatttgaca ttgtgcagca aagaaatggt
tatggagaag
      181 cccagtccgc tgcttgtagg gcgggagttt gtgaggcaat attatacttt
gctgaataaa
      241 gctccggaat atttacacag gttttatggc aggaattctt cctatgttca
tggtggagta
      301 gatgctagtg gaaagcccca ggaagctgtt tatggccaaa atgatataca
ccacaaagta
      361 ttatctctga acttcagtga atgtcatact aaaattcgtc atgtggatgc
tcatgcaacc
      421 ttgagtgatg gagtagttgt ccaggtcatg ggtttgctgt ctaacagtgg
acaaccagaa
      481 agaaagttta tgcaaacctt tgttctggct cctgaaggat ctgttccaaa
taaattttat
      541 gttcacaatg atatgtttcg ttatgaagat gaagtgtttg gtgattctga
gcctgaactt
      601 gatgaagaat cagaagatga agtagaagag gaacaagaag aaagacaacc
atctcctgaa
      661 cctgtgcaag aaaatgctaa cagtggttac tatgaagctc accctgtgac
taatggcata
      721 gaggagcctt tggaagaatc ctctcatgaa cctgaacctg agccagaatc
tgaaacaaag
```

781 actgaagagc tgaaaccaca agtggaggag aagaacttag aagaactaga
ggagaaatct
841 actactcctc ctccggcaga acctgtttct ctgccacaag aaccaccaaa
ggctttctcc
901 tgggcttcag tgaccagtaa aaacctgcct cctagtggta ctgtttcttc
ctctggaatt
961 ccaccccatg ttaaagcacc agtctcacag ccaagagtcg aagctaaacc
agaagttcaa
1021 tctcagccac ctcgtgtgcg tgaacaacga cctagagaac gacctggttt
tcctcctaga
1081 ggaccaagac caggcagagg agatatggaa cagaatgact ctgacaaccg
tagaataatt
1141 cgctatccag atagtcatca acttttgtt ggtaacttgc cacatgatat
tgatgaaaat
1201 gagctaaagg aattcttcat gagttttgga aacgttgtgg aacttcgcat
caataccaag
1261 ggtgttgggg gaaagcttcc aaattttggt tttgtggttt ttgatgactc
tgaaccagtt
1321 cagagaatct taattgcaaa accgattatg tttcgagggg aagtacgttt
aaatgtggaa
1381 gagaaaaaaa caagagctgc aagagagcga gaaaccagag gtggtggtga
tgatcgcagg
1441 gatattaggc gcaatgatcg aggtcccggt ggtccacgtg gaattgtggg
tggtggaatg
1501 atgcgtgatc gtgatggaag aggacctcct ccaaggggtg gcatggcaca
gaaacttggc
1561 tctggaagag gaaccgggca aatggagggc cgcttcacag gacagcgtcg
ctgaagctcc
1621 actgttggca aagtcttggc agtggtacat tattcatcgt gtttgcattc
ttgttaattt
1681 ttttttggc tttggaatgt gacacagcct ttttgatcat ttctttgatg
tgaaaagcat
1741 ctttggttat cagttaaatt gaggtggaca ttatttcccc aatttcacaa
caggattcac
1801 attgttaatt tataaatcta gacttggaga attaaggact gagaaatgac
catatcttaa
1861 actatctacg acaaagtgaa cttaaaagga catgcccact gaattcaggt
cctttgagta
1921 aaaaaaaaat cttctgctgc acattttgtt taagtgttac tgtttctgcc
tgttaatgct
1981 gggaacacaa atagtgcaat ttgtgcaatt ggagaatctt gccttttttc
ttggctcccc
2041 ccaaaaatac aaaccaacag aaacttgtta tgcactcatc aaaatgtact
aatgggtact
2101 ctgaactcat taacattgac atctgcaaca ggaggcaaca gggaaaaaat
ctcatcttct
2161 tttccagtag aaaatagttt gtgaaatgat gagggcattt tatctgcttg
ctgtgaccag
2221 cgtgtgtaca cataaacctt aacaagacta caagtatatt ccagaaggaa
atcattttag
2281 ttatgaacta aataataaaa attagaactt caaatgcgat ggtcttgact
attagaccag
2341 atttagtagc tccatatcta agatttttct acctgcccct cttcagtaca
gggatggctg
2401 gctgctcaac acactcctcc tccccttttt tcctttcttt aagctgtgta
cagtgaaaat
2461 tgtctttact gtatttttgt tctctggtaa tgtaataagc atgatggtgc
cttctattaa
2521 tacatcattc cagtcttgct ggtaattttg tacagtatag tgtatgaatt
gctgtgctgc

2581 aaagccaaac agctgcaaaa tgttgaaaaa tcatcgaaat gtataaaaat
tgcagtatct
2641 ttaaaatcag taaaatggac tagcatatta tttatcttgt tcttcagtta
acaactttgt
2701 gttctctgtg ggagggaggg agtcctgtgt gtttgtgggg agagggaagg
aggaagtcag
2761 ttatttgagt aagcctctag ttgacttttc tcttagcctg aatgtggacg
ttgaaacata
2821 tcacttcagg gcttggaaaa gtcagtcaac ttgacgtaca tttttagtga
cattttaaaa
2881 gcagtcagat tctataaatg gcaagtaagc ctgaagtgag gatactgcaa
ttttcggaga
2941 aaagaacagc agctctttaa gtgtttgcat tttctatttg gggggcaggg
aactgtcatt
3001 cattttgcac aattcttgaa ctgatgtcag cacccgagtg gctcctgaat
ttaagtctgg
3061 gacgacatct tttattttta catgaatctt taaacaattc tgtgagcaaa
gtttgtagct
3121 gctggattat tgtctgtctt tatagcaagt tccagtaaac cacaagtatg
gcaaagctta
3181 tccaatttta tgcttggagc agtcagtaca taccagtttc tgatgtttca
ggcaggagtg
3241 gggtaaataa gtgtgaccac ttaaagctgc tcgttagcat ggaagacttc
tccattctat
3301 ctttgtaaaa cagacaagat atgcacttga catagtagca aattggttct
gaattatgca
3361 actgtttgct atttagtaaa ctagcaaatg atgcatgtat tttgtttttc
atgtactggg
3421 caatatgagt aaaatctgtc ccttttccc cctttgaatg aggtcttcca
tgtttgaggg
3481 aaagtcttgc actattgcat atattttggg gacacagatt ttcatagttt
ccatttttgg
3541 ggggcttaag gatttttttt ttttctgttt gaaacagttt tatactttct
gatatagtac
3601 ttgaaattct taccagaaaa ttactttgga gttttgaagc ctttattaat
actacttta
3661 aagaagcagt tgttttattg tcaatgtttt ttttccccca agcatatttt
cttgtatttc
3721 tgtttccata tatatatata tatatataat ttccaattca ggatattgcc
ctgccatcca
3781 tgaaaactgt tctggcacca aaagtaatga caaatgttaa gtgtaataat
agaaaagtag
3841 agcaaagagc cattcagctt cagtctttac ataccatgaa taaaacatta
aaacatcata
3901 tggagaagtt tacatggtga ttgttcacct gcagtactgt ggagttttaa
cattttgtcc
3961 tcttttcagt gaaacagagt aaaaatattc atctaccatt actgttattt
gctgattttg
4021 ttttatttt tgatggtaat attctatcct tatgacacta ttgcaaccaa
attggcttta
4081 ccatcttggc tttagtaggt atagaagaca atggattacc atctttattg
ctgtaatgtg
4141 ttaagcatta tatgctagta gaatctagtt taattgtttc aggtggaaag
tattctttga
4201 gtttccatat tgaatgtgtt tggactaaac aaacaataaa ctactgatgt
ctgcagcatt
4261 tatctatgtc cctaa

Protein sequence (variant 2):

NCBI Reference Sequence: NP_036429.2

LOCUS NP_036429

ACCESSION NP_036429

```
      1 mvmekpspll vgrefvrqyy tllnkapeyl hrfygrnssy vhggvdasgk
pqeavygqnd
       61 ihhkvlslnf sechtkirhv dahatlsdgv vvqvmgllsn sgqperkfmq
tfvlapegsv
      121 pnkfyvhndm fryedevfgd sepeldeese deveeeqeer qpspepvqen
ansgyyeahp
      181 vtngieeple esshepepep esetkteelk pqveeknlee leeksttppp
aepvslpqep
      241 pkafswasvt sknlppsgtv sssgipphvk apvsqprvea kpevqsqppr
vreqrprerp
      301 gfpprgprpg rgdmeqndsd nrriirypds hqlfvgnlph didenelkef
fmsfgnvvel
      361 rintkgvggk lpnfgfvvfd dsepvqrili akpimfrgev rlnveekktr
aareretrgg
      421 gddrrdirrn drgpggprgi vgggmmrdrd grgpppprggm aqklgsgrgt
gqmegrftgq
      481 rr
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_203504.2

LOCUS NM_203504

ACCESSION NM_203504

```
      1 gtgctcgggg gttccctggc cctttcggca ggggtaaaac aataagaggg
ggcggtggca
       61 aaggggggcgg gacgtccgtg gtccttgtcg cacgtcgcag cgcctggcgc
ccgggaagag
      121 gtggttgtga ggcagacgaa ctcgcggctc tccggcttcc gaggcttccg
agttgtcgga
      181 ggaaggggggc ggcgagcaat aagaacccgc cgcacccggt cctcagcgac
tcttctgacc
      241 tccgcgcgac gtacccgccg ccgccgttgg ctggagcatt tgacattgtg
cagcaaagaa
      301 atggttatgg agaagcccag tccgctgctt gtagggcggg agtttgtgag
gcaatattat
      361 actttgctga ataaagctcc ggaatattta cacaggtttt atggcaggaa
ttcttcctat
      421 gttcatggtg gagtagatgc tagtggaaag ccccaggaag ctgtttatgg
ccaaaatgat
      481 atacaccaca aagtattatc tctgaacttc agtgaatgtc atactaaaat
tcgtcatgtg
      541 gatgctcatg caaccttgag tgatggagta gttgtccagg tcatgggttt
gctgtctaac
      601 agtggacaac cagaaagaaa gtttatgcaa acctttgttc tggctcctga
aggatctgtt
      661 ccaaataaat tttatgttca caatgatatg tttcgttatg aagatgaagt
gtttggtgat
      721 tctgagcctg aacttgatga agaatcagaa gatgaagtag aagaggaaca
agaagaaaga
```

781 caaccatctc ctgaacctgt gcaagaaaat gctaacagtg gttactatga
agctcaccct
841 gtgactaatg gcatagagga gcctttggaa gaatcctctc atgaacctga
acctgagcca
901 gaatctgaaa caaagactga agagctgaaa ccacaagtgg aggagaagaa
cttagaagaa
961 ctagaggaga aatctactac tcctcctccg gcagaacctg tttctctgcc
acaagaacca
1021 ccaaagccaa gagtcgaagc taaaccagaa gttcaatctc agccacctcg
tgtgcgtgaa
1081 caacgaccta gagaacgacc tggttttcct cctagaggac caagaccagg
cagaggagat
1141 atggaacaga atgactctga caaccgtaga ataattcgct atccagatag
tcatcaactt
1201 tttgttggta acttgccaca tgatattgat gaaaatgagc taaaggaatt
cttcatgagt
1261 tttggaaacg ttgtggaact tcgcatcaat accaagggtg ttgggggaaa
gcttccaaat
1321 tttggttttg tggtttttga tgactctgaa ccagttcaga gaatcttaat
tgcaaaaccg
1381 attatgtttc gaggggaagt acgtttaaat gtggaagaga aaaaaacaag
agctgcaaga
1441 gagcgagaaa ccagaggtgg tggtgatgat cgcagggata ttaggcgcaa
tgatcgaggt
1501 cccggtggtc cacgtggaat tgtgggtggt ggaatgatgc gtgatcgtga
tggaagagga
1561 cctcctccaa ggggtggcat ggcacagaaa cttggctctg gaagaggaac
cgggcaaatg
1621 gagggccgct tcacaggaca gcgtcgctga agctccactg ttggcaaagt
cttggcagtg
1681 gtacattatt catcgtgttt gcattcttgt taatttttt tttggctttg
gaatgtgaca
1741 cagccttttt gatcatttct ttgatgtgaa aagcatcttt ggttatcagt
taaattgagg
1801 tggacattat ttccccaatt tcacaacagg attcacattg ttaatttata
aatctagact
1861 tggagaatta aggactgaga aatgaccata tcttaaacta tctacgacaa
agtgaactta
1921 aaaggacatg cccactgaat tcaggtcctt tgagtaaaaa aaaaatcttc
tgctgcacat
1981 tttgtttaag tgttactgtt tctgcctgtt aatgctggga acacaaatag
tgcaatttgt
2041 gcaattggag aatcttgcct tttttcttgg ctccccccaa aaatacaaac
caacagaaac
2101 ttgttatgca ctcatcaaaa tgtactaatg ggtactctga actcattaac
attgacatct
2161 gcaacaggag gcaacaggga aaaaatctca tcttcttttc cagtagaaaa
tagtttgtga
2221 aatgatgagg gcattttatc tgcttgctgt gaccagcgtg tgtacacata
aaccttaaca
2281 agactacaag tatattccag aaggaaatca ttttagttat gaactaaata
ataaaaatta
2341 gaacttcaaa tgcgatggtc ttgactatta gaccagattt agtagctcca
tatctaagat
2401 ttttctacct gcccctcttc agtacaggga tggctggctg ctcaacacac
tcctcctccc
2461 ctttttttcct ttctttaagc tgtgtacagt gaaaattgtc tttactgtat
ttttgttctc
2521 tggtaatgta ataagcatga tggtgccttc tattaataca tcattccagt
cttgctggta

```
     2581 attttgtaca gtatagtgta tgaattgctg tgctgcaaag ccaaacagct
gcaaaatgtt
     2641 gaaaaatcat cgaaatgtat aaaaattgca gtatctttaa aatcagtaaa
atggactagc
     2701 atattattta tcttgttctt cagttaacaa ctttgtgttc tctgtgggag
ggagggagtc
     2761 ctgtgtgttt gtggggagag ggaaggagga agtcagttat ttgagtaagc
ctctagttga
     2821 cttttctctt agcctgaatg tggacgttga aacatatcac ttcagggctt
ggaaaagtca
     2881 gtcaacttga cgtacatttt tagtgacatt ttaaaagcag tcagattcta
taaatggcaa
     2941 gtaagcctga agtgaggata ctgcaatttt cggagaaaag aacagcagct
ctttaagtgt
     3001 ttgcattttc tatttggggg gcagggaact gtcattcatt ttgcacaatt
cttgaactga
     3061 tgtcagcacc cgagtggctc ctgaatttaa gtctgggacg acatctttta
tttttacatg
     3121 aatctttaaa caattctgtg agcaaagttt gtagctgctg gattattgtc
tgtctttata
     3181 gcaagttcca gtaaaccaca agtatggcaa agcttatcca attttatgct
tggagcagtc
     3241 agtacatacc agtttctgat gtttcaggca ggagtggggt aaataagtgt
gaccacttaa
     3301 agctgctcgt tagcatggaa gacttctcca ttctatcttt gtaaaacaga
caagatatgc
     3361 acttgacata gtagcaaatt ggttctgaat tatgcaactg tttgctattt
agtaaactag
     3421 caaatgatgc atgtattttg tttttcatgt actgggcaat atgagtaaaa
tctgtccctt
     3481 tttccccctt tgaatgaggt cttccatgtt tgagggaaag tcttgcacta
ttgcatatat
     3541 tttggggaca cagattttca tagtttccat ttttgggggg cttaaggatt
ttttttttt
     3601 ctgtttgaaa cagttttata ctttctgata tagtacttga aattcttacc
agaaaattac
     3661 tttggagttt tgaagccttt attaatacta cttttaaaga agcagttgtt
ttattgtcaa
     3721 tgttttttt cccccaagca tattttcttg tatttctgtt tccatatata
tatatatata
     3781 tataatttcc aattcaggat attgccctgc catccatgaa aactgttctg
gcaccaaaag
     3841 taatgacaaa tgttaagtgt aataatagaa aagtagagca aagagccatt
cagcttcagt
     3901 ctttacatac catgaataaa acattaaaac atcatatgga gaagtttaca
tggtgattgt
     3961 tcacctgcag tactgtggag ttttaacatt ttgtcctctt ttcagtgaaa
cagagtaaaa
     4021 atattcatct accattactg ttatttgctg attttgtttt atttttgat
ggtaatattc
     4081 tatccttatg acactattgc aaccaaattg gctttaccat cttggcttta
gtaggtatag
     4141 aagacaatgg attaccatct ttattgctgt aatgtgttaa gcattatatg
ctagtagaat
     4201 ctagtttaat tgtttcaggt ggaaagtatt ctttgagttt ccatattgaa
tgtgtttgga
      4261 ctaaacaaac aataaactac tgatgtctgc agcatttatc tatgtcccta a
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_987100.1

LOCUS NP_987100

ACCESSION NP_987100

```
      1 mvmekpspll vgrefvrqyy tllnkapeyl hrfygrnssy vhggvdasgk
pqeavygqnd
       61 ihhkvlslnf sechtkirhv dahatlsdgv vvqvmgllsn sgqperkfmq
tfvlapegsv
      121 pnkfyvhndm fryedevfgd sepeldeese deveeeqeer qpspepvqen
ansgyyeahp
      181 vtngieeple esshepepep esetkteelk pqveeknlee leeksttppp
aepvslpqep
      241 pkprveakpe vqsqpprvre qrprerpgfp prgprpgrgd meqndsdnrr
iirypdshql
      301 fvgnlphdid enelkeffms fgnvvelrin tkgvggklpn fgfvvfddse
pvqriliakp
      361 imfrgevrln veekktraar eretrgggdd rrdirrndrg pggprgivgg
gmmrdrdgrg
      421 ppprggmaqk lgsgrgtgqm egrftgqrr
```

UQCRH

Official Symbol: UQCRH

Official Name: ubiquinol-cytochrome c reductase hinge protein

Gene ID: 7388

Organism: Homo sapiens

Other Aliases: QCR6, UQCR8

Other Designations: complex III subunit 6; complex III subunit VIII; cytochrome b-c1 complex subunit 6, mitochondrial; cytochrome c1 non-heme 11 kDa protein; mitochondrial hinge protein; ubiquinol-cytochrome c reductase complex 11 kDa protein; ubiquinol-cytochrome c reductase, complex III subunit VIII

Nucleotide sequence:

NCBI Reference Sequence: NM_006004.2

LOCUS NM_006004

ACCESSION NM_006004

```
      1 ctgaactggg ttaggtgccg ctgttgctgc tcgtgttgaa tctagaaccg
tagccagaca
       61 tgggactgga ggacgagcaa aagatgctta ccgaatccgg agatcctgag
gaggaggaag
      121 aggaagagga ggaattagtg gatcccctaa caacagtgag agagcaatgc
gagcagttgg
      181 agaaatgtgt aaaggcccgg gagcggctag agctctgtga tgagcgtgta
tcctctcgat
```

```
      241 cacatacaga agaggattgc acggaggagc tctttgactt cttgcatgcg
agggaccatt
      301 gcgtggccca caaactcttt aacaacttga aataaatgtg tggacttaat
tcaccccagt
      361 cttcatcatc tgggcatcag aatatttcct tatggttttg gatgtaccat
ttgtttctta
      421 tttgtgtaac tgtaagttca catgaacctc atgggtttgg cttaggctgg
tagcttctat
      481 gtaattcgca atgattccat ctaaataaaa gttctatgat ctgcaaaaaa
aaaaaaaaaa
      541 aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_005995.2

LOCUS NP_005995

ACCESSION NP_005995

```
      1 mgledeqkml tesgdpeeee eeeeelvdpl ttvreqceql ekcvkarerl
elcdervssr
       61 shteedctee lfdflhardh cvahklfnnl k
```

HSPA4

Official Symbol: HSPA4

Official Name: heat shock 70kDa protein 4

Gene ID: 3308

Organism: Homo sapiens

Other Aliases: APG-2, HS24/P52, HSPH2, RY, hsp70, hsp70RY

Other Designations: heat shock 70 kDa protein 4; heat shock 70-related protein APG-2; heat shock 70kD protein 4; heat shock protein, 110 kDa; hsp70 RY

Nucleotide sequence:

NCBI Reference Sequence: NM_002154.3

LOCUS NM_002154

ACCESSION NM_002154

```
      1 gctctggtgc tgcggctccg ctctcgtcgc aacgagatct ttcgagatct
tctccgcccc
       61 cgctaccggc gcctcctctg cggccactga gccggagccg gcctgagcag
cgctctcggt
      121 tgcagtaccc actggaagga cttaggcgct cgcgtggaca ccgcaagccc
ctcagtagcc
      181 tcggcccaag aggcctgctt tccactcgct agccccgccg ggggtccgtg
tcctgtctcg
```

241 gtggccggac ccgggcccga gcccgagcag tagccggcgc catgtcggtg
gtgggcatag
301 acctgggctt ccagagctgc tacgtcgctg tggcccgcgc cggcggcatc
gagactatcg
361 ctaatgagta tagcgaccgc tgcacgccgg cttgcatttc ttttggtcct
aagaatcgtt
421 caattggagc agcagctaaa agccaggtaa tttctaatgc aaagaacaca
gtccaaggat
481 ttaaaagatt ccatggccga gcattctctg atccatttgt ggaggcagaa
aaatctaacc
541 ttgcatatga tattgtgcag ttgcctacag gattaacagg tataaaggtg
acatatatgg
601 aggaagagcg aaattttacc actgagcaag tgactgccat gcttttgtcc
aaactgaagg
661 agacagccga aagtgttctt aagaagcctg tagttgactg tgttgtttcg
gttccttgtt
721 tctatactga tgcagaaaga cgatcagtga tggatgcaac acagattgct
ggtcttaatt
781 gcttgcgatt aatgaatgaa accactgcag ttgctcttgc atatggaatc
tataagcagg
841 atcttcctgc cttagaagag aaaccaagaa atgtagtttt tgtagacatg
ggccactctg
901 cttatcaagt ttctgtatgt gcatttaata gaggaaaact gaaagttctg
gccactgcat
961 ttgacacgac attgggaggt agaaaatttg atgaagtgtt agtaaatcac
ttctgtgaag
1021 aatttgggaa gaaatacaag ctagacatta agtccaaaat ccgtgcatta
ttacgactct
1081 ctcaggagtg tgagaaactc aagaaattga tgagtgcaaa tgcttcagat
ctccctttga
1141 gcattgaatg ttttatgaat gatgttgatg tatctggaac tatgaataga
ggcaaatttc
1201 tggagatgtg caatgatctc ttagctagag tggagccacc acttcgtagt
gttttggaac
1261 aaaccaagtt aaagaaagaa gatatttatg cagtggagat agttggtggt
gctacacgaa
1321 tccctgcggt aaaagagaag atcagcaaat ttttcggtaa agaacttagt
acaacattaa
1381 atgctgatga agctgtcact cgaggctgtg cattgcagtg tgccatctta
tcgcctgctt
1441 tcaaagtcag agaattttct atcactgatg tagtaccata tccaatatct
ctgagatgga
1501 attctccagc tgaagaaggg tcaagtgact gtgaagtctt ttccaaaaat
catgctgctc
1561 ctttctctaa agttcttaca ttttatagaa aggaaccttt cactcttgag
gcctactaca
1621 gctctcctca ggatttgccc tatccagatc ctgctatagc tcagttttca
gttcagaaag
1681 tcactcctca gtctgatggc tccagttcaa aagtgaaagt caaagttcga
gtaaatgtcc
1741 atggcatttt cagtgtgtcc agtgcatctt tagtggaggt tcacaagtct
gaggaaaatg
1801 aggagccaat ggaaacagat cagaatgcaa aggaggaaga gaagatgcaa
gtggaccagg
1861 aggaaccaca tgttgaagag caacagcagc agacaccagc agaaaataag
gcagagtctg
1921 aagaaatgga gacctctcaa gctggatcca aggataaaaa gatggaccaa
ccaccccaag
1981 ccaagaaggc aaaagtgaag accagtactg tggacctgcc aatcgagaat
cagctattat

```
     2041 ggcagataga cagagagatg ctcaacttgt acattgaaaa tgagggtaag
atgatcatgc
     2101 aggataaact ggagaaggag cggaatgatg ctaagaacgc agtggaggaa
tatgtgtatg
     2161 aaatgagaga caagcttagt ggtgaatatg agaagtttgt gagtgaagat
gatcgtaaca
     2221 gttttacttt gaaactggaa gatactgaaa attggttgta tgaggatgga
gaagaccagc
     2281 caaagcaagt ttatgttgat aagttggctg aattaaaaaa tctaggtcaa
cctattaaga
     2341 tacgtttcca ggaatctgaa gaacgaccaa aattatttga agaactaggg
aaacagatcc
     2401 aacagtatat gaaaataatc agctctttca aaaacaagga ggaccagtat
gatcatttgg
     2461 atgctgctga catgacaaag gtagaaaaaa gcacaaatga agcaatggag
tggatgaata
     2521 acaagctaaa tctgcagaac aagcagagtt tgaccatgga tccagttgtc
aagtcaaaag
     2581 agattgaagc taaaattaag gagctgacaa gtacttgtag ccctataatt
tcaaagccca
     2641 aacccaaagt ggaacctcca aaagaggaac aaaaaaatgc agagcagaat
ggaccagtgg
     2701 atggacaagg agacaaccca ggcccccagg ctgctgagca gggtacagac
acagctgtgc
     2761 cttcggattc agacaagaag cttcctgaaa tggacattga ttgattccaa
cacttgtttc
     2821 tattaaaaca gactattata aagctttaag ttgtcaactt tgttctaaat
atcaactagc
     2881 gcaagtgaat actgaagatt tcttagtcag tttttagggg attttcgggg
aggggaaata
     2941 ggtaatgtat ggagcatttt cacttctaaa tagttagata cagaaattaa
gtgcattgta
     3001 tctttttcat aatggtacta tttagaagcc cagttagtct tactgagctt
atgcttcact
     3061 cctttatgtt taaccatgtg tctacaagaa taagtttgtt ttggaaagtt
gagctatagc
     3121 tacagctcta gctatccagc agacttttca ttatgactta catggcagga
gctctaatta
     3181 tgctttaaaa atctgttgtg gagattgctt taaatgctcc ctgcctggtg
tggggatggg
     3241 gtccccctct ttgtgagggc tggagcatgg cacggcatgg attaacacgg
cagaggaaca
     3301 aaggtgtgct ctgagcttct tcatatttca ccttcaccct cacctgtgtt
ctcttccctc
      3361 tctcccaata aaagggctcc catta
```

Protein sequence:

NCBI Reference Sequence: NP_002145.3

LOCUS NP_002145

ACCESSION NP_002145

```
      1 msvvgidlgf qscyvavara ggietianey sdrctpacis fgpknrsiga
aaksqvisna
       61 kntvqgfkrf hgrafsdpfv eaeksnlayd ivqlptgltg ikvtymeeer
nftteqvtam
      121 llsklketae svlkkpvvdc vvsvpcfytd aerrsvmdat qiaglnclrl
mnettavala
```

```
      181 ygiykqdlpa leekprnvvf vdmghsayqv svcafnrgkl kvlatafdtt
lggrkfdevl
      241 vnhfceefgk kykldikski rallrlsqec eklkklmsan asdlplsiec
fmndvdvsgt
      301 mnrgkflemc ndllarvepp lrsvleqtkl kkediyavei vggatripav
kekiskffgk
      361 elsttlnade avtrgcalqc ailspafkvr efsitdvvpy pislrwnspa
eegssdcevf
      421 sknhaapfsk vltfyrkepf tleayysspq dlpypdpaia qfsvqkvtpq
sdgssskvkv
      481 kvrvnvhgif svssaslvev hkseeneepm etdqnakeee kmqvdqeeph
veeqqqqtpa
      541 enkaeseeme tsqagskdkk mdqppqakka kvktstvdlp ienqllwqid
remlnlyien
      601 egkmimqdkl ekerndakna veeyvyemrd klsgeyekfv seddrnsftl
kledtenwly
      661 edgedqpkqv yvdklaelkn lgqpikirfq eseerpklfe elgkqiqqym
kiissfknke
      721 dqydhldaad mtkvekstne amewmnnkln lqnkqsltmd pvvkskeiea
kikeltstcs
      781 piiskpkpkv eppkeeqkna eqngpvdgqg dnpgpqaaeq gtdtavpsds
dkklpemdid
```

PSMA7

Official Symbol: PSMA7

Official Name: proteasome (prosome, macropain) subunit, alpha type, 7

Gene ID: 5688

Organism: Homo sapiens

Other Aliases: RP5-1005F21.4, C6, HSPC, RC6-1, XAPC7

Other Designations: proteasome subunit RC6-1; proteasome subunit XAPC7; proteasome subunit alpha 4; proteasome subunit alpha type-7

Nucleotide sequence:

NCBI Reference Sequence: NM_002792.3

LOCUS NM_002792

ACCESSION NM_002792

```
      1 gtcgccgcct gacgccgccc gtcgccggca gcgcaggaca cggcgccgag
ggtggggcgc
       61 gggcgtagtg gcgccgggag tcgcgggtgc gcgcgggccg tgagtgtgcg
cttttgagag
      121 tcgcggcgga aggagcccgg ccgccgcccg ccggcatgag ctacgaccgc
gccatcaccg
      181 tcttctcgcc cgacggccac ctcttccaag tggagtacgc gcaggaggcc
gtcaagaagg
      241 gctcgaccgc ggttggtgtt cgaggaagag acattgttgt tcttggtgtg
gagaagaagt
      301 cagtggccaa actgcaggat gaaagaacag tgcggaagat ctgtgctttg
gatgacaacg
```

```
      361 tctgcatggc ctttgcaggc ctcaccgccg atgcaaggat agtcatcaac
agggcccggg
      421 tggagtgcca gagccaccgg ctgactgtgg aggacccggt cactgtggag
tacatcaccc
      481 gctacatcgc cagtctgaag cagcgttata cgcagagcaa tgggcgcagg
ccgtttggca
      541 tctctgccct catcgtgggt ttcgactttg atggcactcc taggctctat
cagactgacc
      601 cctcgggcac ataccatgcc tggaaggcca atgccatagg tcggggtgcc
aagtcagtgc
      661 gcgagttcct ggagaagaac tatactgacg aagccattga aacagatgat
ctgaccatta
      721 agctggtgat caaggcactc ctggaagtgg ttcagtcagg tggcaaaaac
attgaacttg
      781 ctgtcatgag gcgagatcaa tccctcaaga ttttaaatcc tgaagaaatt
gagaagtatg
      841 ttgctgaaat tgaaaaagaa aaagaagaaa acgaaaagaa gaaacaaaag
aaagcatcat
      901 gatgaataaa atgtctttgc ttgtaatttt taaattcata tcaatcatgg
atgagtctcg
      961 atgtgtaggc ctttccattc catttattca cactgagtgt cctacaataa
acttccgtat
     1021 ttttaacctg ttaaaaaaaa aaaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_002783.1

LOCUS NP_002783

ACCESSION NP_002783

```
      1 msydraitvf spdghlfqve yaqeavkkgs tavgvrgrdi vvlgvekksv
aklqdertvr
       61 kicalddnvc mafagltada rivinrarve cqshrltved pvtveyitry
iaslkqrytq
      121 sngrrpfgis alivgfdfdg tprlyqtdps gtyhawkana igrgaksvre
fleknytdea
      181 ietddltikl vikallevvq sggknielav mrrdqslkil npeeiekyva
eiekekeene
      241 kkkqkkas
```

KIF5B

Official Symbol: KIF5B

Official Name: kinesin family member 5B

Gene ID: 3799

Organism: Homo sapiens

Other Aliases: KINH, KNS, KNS1, UKHC

Other Designations: conventional kinesin heavy chain; kinesin 1 (110-120kD); kinesin heavy chain; kinesin-1 heavy chain; ubiquitous kinesin heavy chain

Nucleotide sequence:

NCBI Reference Sequence: NM_004521.2

LOCUS NM_004521

ACCESSION NM_004521

```
      1 ctcctcccgc accgccctgt cgcccaacgg cggcctcagg agtgatcggg
cagcagtcgg
       61 ccggccagcg gacggcagag cgggcggacg ggtaggcccg gcctgctctt
cgcgaggagg
      121 aagaaggtgg ccactctccc ggtccccaga acctccccag cccccgcagt
ccgcccagac
      181 cgtaaagggg gacgctgagg agccgcggac gctctccccg gtgccgccgc
cgctgccgcc
      241 gccatggctg ccatgatgga tcggaagtga gcattagggt taacggctgc
cggcgccggc
      301 tcttcaagtc ccggctcccc ggccgcctcc acccggggaa gcgcagcgcg
gcgcagctga
      361 ctgctgcctc tcacggccct cgcgaccaca agccctcagg tccggcgcgt
tccctgcaag
      421 actgagcggc ggggagtggc tcccggccgc cggccccggc tgcgagaaag
atggcggacc
      481 tggccgagtg caacatcaaa gtgatgtgtc gcttcagacc tctcaacgag
tctgaagtga
      541 accgcggcga caagtacatc gccaagtttc agggagaaga cacggtcgtg
atcgcgtcca
      601 agccttatgc atttgatcgg gtgttccagt caagcacatc tcaagagcaa
gtgtataatg
      661 actgtgcaaa gaagattgtt aaagatgtac ttgaaggata taatggaaca
atatttgcat
      721 atggacaaac atcctctggg aagacacaca caatggaggg taaacttcat
gatccagaag
      781 gcatgggaat tattccaaga atagtgcaag atatttttaa ttatatttac
tccatggatg
      841 aaaatttgga atttcatatt aaggtttcat attttgaaat atatttggat
aagataaggg
      901 acctgttaga tgtttcaaag accaaccttt cagttcatga agacaaaaac
cgagttccct
      961 atgtaaaggg gtgcacagag cgttttgtat gtagtccaga tgaagttatg
gataccatag
     1021 atgaaggaaa atccaacaga catgtagcag ttacaaatat gaatgaacat
agctctagga
     1081 gtcacagtat atttcttatt aatgtcaaac aagagaacac acaaacggaa
caaaagctga
     1141 gtggaaaact ttatctggtt gatttagctg gtagtgaaaa ggttagtaaa
actggagctg
     1201 aaggtgctgt gctggatgaa gctaaaaaca tcaacaagtc actttctgct
cttggaaatg
     1261 ttatttctgc tttggctgag ggtagtacat atgttccata tcgagatagt
aaaatgacaa
     1321 gaatccttca agattcatta ggtggcaact gtagaaccac tattgtaatt
tgctgctctc
     1381 catcatcata caatgagtct gaaacaaaat ctacactctt atttggccaa
agggccaaaa
```

1441 caattaagaa cacagtttgt gtcaatgtgg agttaactgc agaacagtgg
aaaaagaagt
1501 atgaaaaaga aaaagaaaaa aataagatcc tgcggaacac tattcagtgg
cttgaaaatg
1561 agctcaacag atggcgtaat ggggagacgg tgcctattga tgaacagttt
gacaaagaga
1621 aagccaactt ggaagctttc acagtggata aagatattac tcttaccaat
gataaaccag
1681 caaccgcaat tggagttata ggaaatttta ctgatgctga aagaagaaag
tgtgaagaag
1741 aaattgctaa attatacaaa cagcttgatg acaaggatga agaaattaac
cagcaaagtc
1801 aactggtaga gaaactgaag acgcaaatgt tggatcagga ggagcttttg
gcatctacca
1861 gaagggatca agacaatatg caagctgagc tgaatcgcct tcaagcagaa
aatgatgcct
1921 ctaaagaaga agtgaaagaa gttttacagg ccctagaaga acttgctgtc
aattatgatc
1981 agaagtctca ggaagttgaa gacaaaacta aggaatatga attgcttagt
gatgaattga
2041 atcagaaatc ggcaacttta gcgagtatag atgctgagct tcagaaactt
aaggaaatga
2101 ccaaccacca gaaaaaacga gcagctgaga tgatggcatc tttactaaaa
gaccttgcag
2161 aaataggaat tgctgtggga aataatgatg taaagcagcc tgagggaact
ggcatgatag
2221 atgaagagtt cactgttgca agactctaca ttagcaaaat gaagtcagaa
gtaaaaacca
2281 tggtgaaacg ttgcaagcag ttagaaagca cacaaactga gagcaacaaa
aaaatggaag
2341 aaaatgaaaa ggagttagca gcatgtcagc ttcgtatctc tcaacatgaa
gccaaaatca
2401 agtcattgac tgaatacctt caaaatgtgg aacaaaagaa aagacagttg
gaggaatctg
2461 tcgatgccct cagtgaagaa ctagtccagc ttcgagcaca agagaaagtc
catgaaatgg
2521 aaaaggagca cttaaataag gttcagactg caaatgaagt taagcaagct
gttgaacagc
2581 agatccagag ccatagagaa actcatcaaa aacagatcag tagtttgaga
gatgaagtag
2641 aagcaaaagc aaaacttatt actgatcttc aagaccaaaa ccagaaaatg
atgttagagc
2701 aggaacgtct aagagtagaa catgagaagt tgaaagccac agatcaggaa
aagagcagaa
2761 aactacatga acttacggtt atgcaagata gacgagaaca agcaagacaa
gacttgaagg
2821 gtttggaaga gacagtggca aaagaacttc agactttaca caacctgcgc
aaactctttg
2881 ttcaggacct ggctacaaga gttaaaaaga gtgctgagat tgattctgat
gacaccggag
2941 gcagcgctgc tcagaagcaa aaaatctcct ttcttgaaaa taatcttgaa
cagctcacta
3001 aagtgcacaa acagttggta cgtgataatg cagatctccg ctgtgaactt
cctaagttgg
3061 aaaagcgact tcgagctaca gctgagagag tgaaagcttt ggaatcagca
ctgaaagaag
3121 ctaaagaaaa tgcatctcgt gatcgcaaac gctatcagca agaagtagat
cgcataaagg
3181 aagcagtcag gtcaaagaat atggccagaa gagggcattc tgcacagatt
gctaaaccta

3241 ttcgtcccgg gcaacatcca gcagcttctc caactcaccc aagtgcaatt
cgtggaggag
3301 gtgcatttgt tcagaacagc cagccagtgg cagtgcgagg tggaggaggc
aaacaagtgt
3361 aatcgtttat acatacccac aggtgttaaa aagtaatcga agtacgaaga
ggacatggta
3421 tcaagcagtc attcaatgac tataacctct actcccttgg gattgtagaa
ttataacttt
3481 taaaaaaaat gtataaatta tacctggcct gtacagctgt ttcctaccta
ctcttcttgt
3541 aaactctgct gcttcccaac acaactagag tgcaattttg gcatcttagg
agggaaaaag
3601 gacagtttac aactgtggcc ctatttatta cacagtttgt ctatcgtgtc
ttaaatttag
3661 tctttactgt gccaagctaa ctgtacctta taggactgta ctttttgtat
tttttgtgta
3721 tgtttatttt ttaatctcag tttaaattac ctagctgcta ctgcttcttg
ttttttctttt
3781 cctattaaaa cgtcttcctt tttttttctt aagagaaaat ggaacattta
ggttaaatgt
3841 ctttaaattt taccacttaa caacactaca tgcccataaa atatatccag
tcagtactgt
3901 attttaaaat cccttgaaat gatgatatca gggttaaaat tacttgtatt
gtttctgaag
3961 tttgctcctg aaaactactg tttgagcact gaaacgttac aaatgcctaa
taggcatttg
4021 agactgagca aggctacttg ttatctcatg aaatgcctgt tgccgagtta
ttttgaatag
4081 aaatatttta aagtatcaaa agcagatctt agtttaaggg agtttggaaa
aggaattata
4141 tttctctttt tcctgattct gtactcaaca agtcttgatg gaattaaaat
actctgcttt
4201 attctggtga gcctgctagc taatataagt attggacagg taataatttg
tcatctttaa
4261 tattagtaaa atgaattaag atattatagg attaaacata attttatacg
gttagtactt
4321 tattggccga cctaaattta tagcgtgtgg aaattgagaa aaatgaagaa
acaggacaga
4381 tatatgatga attaaaaata tatataggtc aattttggtc tgaaatccct
gaggtgtttt
4441 taacctgcta cactaatttg tacactaatt tatttcttta gtctagaaat
agtaaattgt
4501 ttgcaagtca ctaataatca ttagataaat tattttcttg gccatagccg
ataatttttgt
4561 aatcagtact aagtgtatac gtatttttgc cacttttttcc tcagatgatt
aaagtaagtc
4621 aacagcttat tttaggaaac tgtaaaagta atagggaaag agatttcact
atttgcttca
4681 tcagtggtag gggggcggtg actgcaactg tgttagcaga aattcacaga
gaatggggat
4741 ttaaggttag cagagaaact tggaaagttc tgtgttagga tcttgctggc
agaattaact
4801 ttttgcaaaa gttttataca cagatatttg tattaaattt ggagccatag
tcagaagact
4861 cagatcataa ttggcttatt tttctatttc cgtaactatt gtaatttcca
cttttgtaat
4921 aattttgatt taaaatataa atttatttat ttatttttttt aatagtcaaa
aatctttgct
4981 gttgtagtct gcaacctcta aaatgattgt gttgctttta ggattgatca
gaagaaacac

```
     5041 tccaaaaatt gagatgaaat gttggtgcag ccagttataa gtaatatagt
taacaagcaa
     5101 aaaaagtgct gccacctttt atgatgattt tctaaatgga gaaacatttg
gctgcatcca
     5161 catagacctt tatgttttgt tttcagttga aaacttgcct cctttggcaa
cattcgtaaa
     5221 tgaagcagaa tttttttttc tctttttcc aaatatgtta gttttgttct
tgtaagatgt
     5281 atcatgggta ttggtgctgt gtaatgaaca acgaatttta attagcatgt
ggttcagaat
     5341 atacaatgtt aggtttttaa aaagtatctt gatggttctt ttctatttat
aatttcagac
     5401 tttcataaag tgtaccaaga atttcataaa tttgttttca gtgaactgct
ttttgctatg
     5461 gtaggtcatt aaacacagca cttactctta aaaatgaaaa tttctgatca
tctaggatat
     5521 tgacacattt caatttgcag tgtctttttg actggatata ttaacgttcc
tctgaatggc
     5581 attgatagat ggttcagaag agaaactcaa tgaaataaag agaatattta
ttcatggcga
     5641 ttaattaaat tatttgccta acttaagaaa actactgtgc gtaactctca
gtttgtgctt
     5701 aactccattt gacatgaggt gacagaagag agtctgagtc tacctgtgga
atatgttggt
     5761 ttattttcag tgcttgaaga tacattcaca aatacttggt ttgggaagac
accgtttaat
     5821 tttaagttaa cttgcatgtt gtaaatgcgt tttatgttta aataaagagg
aaaatttttt
      5881 gaaatgtaaa aaaaaaaaaa aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_004512.1

LOCUS NP_004512

ACCESSION NP_004512

```
      1 madlaecnik vmcrfrplne sevnrgdkyi akfqgedtvv iaskpyafdr
vfqsstsqeq
       61 vyndcakkiv kdvlegyngt ifaygqtssg kthtmegklh dpegmgiipr
ivqdifnyiy
      121 smdenlefhi kvsyfeiyld kirdlldvsk tnlsvhedkn rvpyvkgcte
rfvcspdevm
      181 dtidegksnr hvavtnmneh ssrshsifli nvkqentqte qklsgklylv
dlagsekvsk
      241 tgaegavlde akninkslsa lgnvisalae gstyvpyrds kmtrilqdsl
ggncrttivi
      301 ccspssynes etkstllfgq raktikntvc vnveltaeqw kkkyekekek
nkilrntiqw
      361 lenelnrwrn getvpideqf dkekanleaf tvdkditltn dkpataigvi
gnftdaerrk
      421 ceeeiaklyk qlddkdeein qqsqlveklk tqmldqeell astrrdqdnm
qaelnrlqae
      481 ndaskeevke vlqaleelav nydqksqeve dktkeyells delnqksatl
asidaelqkl
      541 kemtnhqkkr aaemmasllk dlaeigiavg nndvkqpegt gmideeftva
rlyiskmkse
      601 vktmvkrckq lestqtesnk kmeenekela acqlrisqhe akikslteyl
qnveqkkrql
```

```
      661 eesvdalsee lvqlraqekv hemekehlnk vqtanevkqa veqqiqshre
thqkqisslr
      721 deveakakli tdlqdqnqkm mleqerlrve heklkatdqe ksrklheltv
mqdrreqarq
      781 dlkgleetva kelqtlhnlr klfvqdlatr vkksaeidsd dtggsaaqkq
kisflennle
      841 qltkvhkqlv rdnadlrcel pklekrlrat aervkalesa lkeakenasr
drkryqqevd
      901 rikeavrskn marrghsaqi akpirpgqhp aaspthpsai rgggafvqns
qpvavrgggg
      961 kqv
```

RPS25

Official Symbol: RPS25

Official Name: ribosomal protein S25

Gene ID: 6230

Organism: Homo sapiens

Other Aliases: S25

Other Designations: 40S ribosomal protein S25

Nucleotide sequence:

NCBI Reference Sequence: NM_001028.2

LOCUS NM_001028

ACCESSION NM_001028

```
      1 cttccttttt gtccgacatc ttgacgaggc tgcggtgtct gctgctattc
tccgagcttc
       61 gcaatgccgc ctaaggacga caagaagaag aaggacgctg gaaagtcggc
caagaaagac
      121 aaagacccag tgaacaaatc cgggggcaag gccaaaaaga agaagtggtc
caaaggcaaa
      181 gttcgggaca agctcaataa cttagtcttg tttgacaaag ctacctatga
taaactctgt
      241 aaggaagttc ccaactataa acttataacc ccagctgtgg tctctgagag
actgaagatt
      301 cgaggctccc tggccagggc agcccttcag gagctcctta gtaaaggact
tatcaaactg
      361 gtttcaaagc acagagctca agtaatttac accagaaata ccaagggtgg
agatgctcca
      421 gctgctggtg aagatgcatg aataggtcca accagctgta catttggaaa
aataaaactt
      481 tattaaatca aaaaaaaaaa aaaaaaaaaa aaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001019.1

LOCUS NP_001019

ACCESSION NP_001019

```
      1 mppkddkkkk dagksakkdk dpvnksggka kkkkwskgkv rdklnnlvlf
dkatydklck
       61 evpnyklitp avvserlkir gslaraalqe llskgliklv skhraqviyt
rntkggdapa
      121 ageda
```

HSP90AB1

Official Symbol: HSP90AB1

Official Name: heat shock protein 90kDa alpha (cytosolic), class B member 1

Gene ID: 3326

Organism: Homo sapiens

Other Aliases: RP1-302G2.1, D6S182, HSP84, HSP90-BETA, HSP90B, HSPC2, HSPCB

Other Designations: 90-kda heat shock protein beta HSP90 beta; heat shock 84 kDa; heat shock 90kD protein 1, beta; heat shock 90kDa protein 1, beta; heat shock protein HSP 90-beta; heat shock protein beta

Nucleotide sequence:

NCBI Reference Sequence: NM_007355.2

LOCUS NM_007355

ACCESSION NM_007355

```
      1 ctccggcgca gtgttgggac tgtctgggta tcggaaagca agcctacgtt
gctcactatt
       61 acgtataatc cttttctttt caagatgcct gaggaagtgc accatggaga
ggaggaggtg
      121 gagacttttg cctttcaggc agaaattgcc caactcatgt ccctcatcat
caataccttc
      181 tattccaaca aggagatttt ccttcgggag ttgatctcta atgcttctga
tgccttggac
      241 aagattcgct atgagagcct gacagaccct tcgaagttgg acagtggtaa
agagctgaaa
      301 attgacatca tccccaaccc tcaggaacgt accctgactt tggtagacac
aggcattggc
      361 atgaccaaag ctgatctcat aaataatttg ggaaccattg ccaagtctgg
tactaaagca
      421 ttcatggagg ctcttcaggc tggtgcagac atctccatga ttgggcagtt
tggtgttggc
      481 ttttattctg cctacttggt ggcagagaaa gtggttgtga tcacaaagca
caacgatgat
      541 gaacagtatg cttgggagtc ttctgctgga ggttccttca ctgtgcgtgc
tgaccatggt
```

601 gagcccattg gcaggggtac caaagtgatc ctccatctta aagaagatca
gacagagtac
661 ctagaagaga ggcgggtcaa agaagtagtg aagaagcatt ctcagttcat
aggctatccc
721 atcacccttt atttggagaa ggaacgagag aaggaaatta gtgatgatga
ggcagaggaa
781 gagaaaggtg agaaagaaga ggaagataaa gatgatgaag aaaaacccaa
gatcgaagat
841 gtgggttcag atgaggagga tgacagcggt aaggataaga agaagaaaac
taagaagatc
901 aaagagaaat acattgatca ggaagaacta aacaagacca agcctatttg
gaccagaaac
961 cctgatgaca tcacccaaga ggagtatgga gaattctaca agagcctcac
taatgactgg
1021 gaagaccact tggcagtcaa gcacttttct gtagaaggtc agttggaatt
cagggcattg
1081 ctatttattc ctcgtcgggc tccctttgac ctttttgaga acaagaagaa
aaagaacaac
1141 atcaaactct atgtccgccg tgtgttcatc atggacagct gtgatgagtt
gataccagag
1201 tatctcaatt ttatccgtgg tgtggttgac tctgaggatc tgcccctgaa
catctcccga
1261 gaaatgctcc agcagagcaa aatcttgaaa gtcattcgca aaaacattgt
taagaagtgc
1321 cttgagctct tctctgagct ggcagaagac aaggagaatt acaagaaatt
ctatgaggca
1381 ttctctaaaa atctcaagct tggaatccac gaagactcca ctaaccgccg
ccgcctgtct
1441 gagctgctgc gctatcatac ctcccagtct ggagatgaga tgacatctct
gtcagagtat
1501 gtttctcgca tgaaggagac acagaagtcc atctattaca tcactggtga
gagcaaagag
1561 caggtggcca actcagcttt tgtggagcga gtgcggaaac ggggcttcga
ggtggtatat
1621 atgaccgagc ccattgacga gtactgtgtg cagcagctca aggaatttga
tgggaagagc
1681 ctggtctcag ttaccaagga gggtctggag ctgcctgagg atgaggagga
gaagaagaag
1741 atggaagaga gcaaggcaaa gtttgagaac ctctgcaagc tcatgaaaga
aatcttagat
1801 aagaaggttg agaaggtgac aatctccaat agacttgtgt cttcaccttg
ctgcattgtg
1861 accagcacct acggctggac agccaatatg gagcggatca tgaaagccca
ggcacttcgg
1921 gacaactcca ccatgggcta tatgatggcc aaaaagcacc tggagatcaa
ccctgaccac
1981 cccattgtgg agacgctgcg gcagaaggct gaggccgaca agaatgataa
ggcagttaag
2041 gacctggtgg tgctgctgtt tgaaaccgcc ctgctatctt ctggcttttc
ccttgaggat
2101 ccccagaccc actccaaccg catctatcgc atgatcaagc taggtctagg
tattgatgaa
2161 gatgaagtgg cagcagagga acccaatgct gcagttcctg atgagatccc
ccctctcgag
2221 ggcgatgagg atgcgtctcg catggaagaa gtcgattagg ttaggagttc
atagttggaa
2281 aacttgtgcc cttgtatagt gtccccatgg gctcccactg cagcctcgag
tgcccctgtc
2341 ccacctggct ccccctgctg gtgtctagtg ttttttccc tctcctgtcc
ttgtgttgaa

```
     2401 ggcagtaaac taagggtgtc aagccccatt ccctctctac tcttgacagc
aggattggat
     2461 gttgtgtatt gtggtttatt ttattttctt cattttgttc tgaaattaaa
gtatgcaaaa
      2521 taaagaatat gccgttttaa aaaaaaaaaa aaaaaaaaaa aaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_031381.2

LOCUS NP_031381

ACCESSION NP_031381

```
      1 mpeevhhgee evetfafqae iaqlmsliin tfysnkeifl relisnasda
ldkiryeslt
       61 dpskldsgke lkidiipnpq ertltlvdtg igmtkadlin nlgtiaksgt
kafmealqag
      121 adismigqfg vgfysaylva ekvvvitkhn ddeqyawess aggsftvrad
hgepigrgtk
      181 vilhlkedqt eyleerrvke vvkkhsqfig ypitlyleke rekeisddea
eeekgekeee
      241 dkddeekpki edvgsdeedd sgkdkkkktk kikekyidqe elnktkpiwt
rnpdditqee
      301 ygefyksltn dwedhlavkh fsvegqlefr allfiprrap fdlfenkkkk
nniklyvrrv
      361 fimdscdeli peylnfirgv vdsedlplni sremlqqski lkvirknivk
kclelfsela
      421 edkenykkfy eafsknlklg ihedstnrrr lsellryhts qsgdemtsls
eyvsrmketq
      481 ksiyyitges keqvansafv ervrkrgfev vymtepidey cvqqlkefdg
kslvsvtkeg
      541 lelpedeeek kkmeeskakf enlcklmkei ldkkvekvti snrlvsspcc
ivtstygwta
      601 nmerimkaqa lrdnstmgym makkhleinp dhpivetlrq kaeadkndka
vkdlvvllfe
      661 tallssgfsl edpqthsnri yrmiklglgi dedevaaeep naavpdeipp
legdedasrm
      721 eevd
```

LMO7

Official Symbol: LMO7

Official Name: LIM domain 7

Gene ID: 4008

Organism: Homo sapiens

Other Aliases: RP11-332E3.2, FBX20, FBXO20, LOMP

Other Designations: F-box only protein 20; F-box protein Fbx20; LIM domain only 7 protein; LIM domain only protein 7; LMO-7; zinc-finger domain-containing protein

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_005358.5

LOCUS NM_005358

ACCESSION NM_005358

```
      1 ggaaagaagt ggaataatta ggaacctagg gtggggtagg gtagcaggac
atttcaaaca
       61 ttaatgagca tatgagattc caggtcttgt taaaatgcaa attctgattc
agctggtagg
      121 tgaggtctga gattgtgcat ttctaacaag cactcagata atcttaaggc
tgttggcccc
      181 agggtcacac ttatagtgat tttctagaac ccagttgggg aagtgaatct
tgggcaggag
      241 aaatacacac ctcttgcatt gagtttggag atctcatctg atataacttt
ttaagaaaga
      301 aaaataattt tccaaatatc caattgataa gctttcccac taagtggctt
tcccactaag
      361 tggctgcgtt atgaaaattg cttcactttg aaacttctgg tcttggtaat
atagaatttc
      421 tgtgttctca cagtgcttga ttgagaatat gatattgaga ttatggcata
aaatatagtg
      481 gctgtacaaa aaaaaataca ttattaggat ctctaacaat tatgtaaaag
tcattgcttc
      541 atgggtagag ctcaaacttt ggtgtgagac ctggttttat tcttggcact
tactctgagt
      601 tgtcttaggc aaattaatac cttaagcaaa aatattctca tgtacatttt
acatgagaat
      661 tataaatgaa gtacataaag tccagcagtc acaaatgtta tctattatta
ccatcgtcct
      721 aagactgcaa tcagctatag tgaaagtagt ctcaaagatt gtttcataaa
tcatcagatt
      781 cacctaattt tctaaagaat ttaaataagg agatggaatg aatagattgc
attttgtttc
      841 catgcacagg ggaactgtgc atatttcttc tgtgactcgg aaatggttta
acttttaaaa
      901 atcccaaaat agctgaagtt agcagacatg caatttacca aggatgattg
gaatttttat
      961 ctttcctgta ataatactat acccaagcac actgctcatg aggaaaacat
ttttatgtga
     1021 atcttttact cttgggggca aagaatgctg tttttctttt tgataactat
gtttatagaa
     1081 tctaaatcac cctgagcaat tatttcaaca tctaaagtta ttattaccat
tcatgtttca
     1141 tttatagcta tttgaatttt gatgaatttc aatatggtgc tacagtgata
gggcaagtgc
     1201 aaataagttc aatatatggg tacggtctaa agctatttta atttttttat
tacaactgct
     1261 atgaagaaaa ttaggatatg ccatattttc acgttttaca gttggatgtc
ctatgatgtt
     1321 ctcttccaga gaacagagct cggagctctg gaaatttgga ggcaactgat
atgtgctcat
     1381 gtctgcatct gtgtgggttg gctgtatctc agggacagag tctgcagcaa
aaaagatata
     1441 attttgagga ctgaacaaaa ttcaggaagg actattctca ttaaggcagt
aacagagaag
```

1501 aattttgaaa caaaagattt tcgagcctct ctagaaaatg gtgttctgct
gtgtgatttg
1561 attaataagc ttaaacctgg cgtcattaag aagatcaata gactgtctac
accaatagca
1621 ggattggata atataaacgt tttcttgaaa gcttgtgaac agattggatt
gaaagaagcc
1681 cagcttttcc atcctggaga tctacaggat ttatcaaatc gagtcactgt
caagcaagaa
1741 gagactgaca ggagagtgaa aaatgttttg ataacattgt actggctggg
aagaaaagca
1801 caaagcaacc cgtactataa tggtccccat cttaatttga aagcgtttga
gaatctttta
1861 ggacaagcac tgacgaaggc actcgaagac tccagcttcc tgaaaagaag
tggcagggac
1921 agtggctacg gtgacatctg gtgtcctgaa cgtggagaat ttcttgctcc
tccaaggcac
1981 cataagagag aagattcctt tgaaagcttg gactctttgg gctcgaggtc
attgacaagc
2041 tgctcctctg atatcacgtt gagaggggggg cgtgaaggtt ttgaaagtga
cacagattcg
2101 gaatttacat ttaagatgca ggattataat aaagatgata tgtcgtatcg
aaggatttcg
2161 gctgttgagc caaagactgc gttacccttc aatcgttttt tacccaacaa
aagtagacag
2221 ccatcctatg taccagcacc tctgagaaag aaaaagccag acaaacatga
ggataacaga
2281 agaagttggg caagcccggt ttatacagaa gcagatggaa cattttcaag
actctttcaa
2341 aagatttatg gtgagaatgg gagtaagtcc atgagtgatg tcagcgcaga
agatgttcaa
2401 aacttgcgtc agctgcgtta cgaggagatg cagaaaataa aatcacaatt
aaaagaacaa
2461 gatcagaaat ggcaggatga ccttgcaaaa tggaaagatc gtcgaaaaag
ttacacttca
2521 gatctgcaga agaaaaaaga agagagagaa gaaattgaaa agcaggcact
tgagaagtct
2581 aagagaagct ctaagacgtt taaggaaatg ctgcaggaca gggaatccca
aaatcaaaag
2641 tctacagttc cgtcaagaag gagaatgtat tcttttgatg atgtgctgga
ggaaggaaag
2701 cgacccccta caatgactgt gtcagaagca agttaccaga gtgagagagt
agaagagaag
2761 ggagcaactt atccttcaga aattcccaaa gaagattcta ccacttttgc
aaaaagagag
2821 gaccgtgtaa caactgaaat tcagcttcct tctcaaagtc ctgtggaaga
acaaagccca
2881 gcctctttgt cttctctgcg ttcacggagc acacaaatgg aatcaactcg
tgtttcagct
2941 tctctcccca gaagttaccg gaaaactgat acagtcaggt taacatctgt
ggtcacacca
3001 agacccttg gctctcagac aaggggaatc tcatcactcc ccagatctta
cacgatggat
3061 gatgcttgga agtataatgg agatgttgaa gacattaaga gaactccaaa
caatgtggtc
3121 agcacccctg caccaagccc ggacgcaagc caactggctt caagcttatc
tagccagaaa
3181 gaggtagcag caacagaaga agatgtgaca aggctgccct ctcctacatc
ccccttctca
3241 tctctttccc aagaccaggc tgccacttct aaagccacat tgtcttccac
atctggtctt

3301 gatttaatgt ctgaatctgg agaaggggaa atctccccac aaagagaagt
ctcaagatcc
3361 caggatcagt tcagtgatat gagaatcagc ataaaccaga cgcctgggaa
gagtcttgac
3421 tttgggttta caataaaatg ggatattcct gggatcttcg tagcatcagt
tgaagcaggt
3481 agcccagcag aattttctca gctacaagta gatgatgaaa ttattgctat
taacaacacc
3541 aagttttcat ataacgattc aaaagagtgg gaggaagcca tggctaaggc
tcaagaaact
3601 ggacacctag tgatggatgt gaggcgctat ggaaaggctg gttcacctga
aacaaagtgg
3661 attgatgcaa cttctggaat ttacaactca gaaaaatctt caaatctatc
tgtaacaact
3721 gatttctccg aaagccttca gagttctaat attgaatcca aagaaatcaa
tggaattcat
3781 gatgaaagca atgcttttga atcaaaagca tctgaatcca tttctttgaa
aaacttaaaa
3841 aggcgatcac aatttttga acaaggaagc tctgattcgg tggttcctga
tcttccagtt
3901 ccaaccatca gtgccccgag tcgctgggtg tgggatcaag aggaggagcg
gaagcggcag
3961 gagaggtggc agaaggagca ggaccgccta ctgcaggaaa aatatcaacg
tgagcaggag
4021 aaactgaggg aagagtggca aagggccaaa caggaggcag agagagagaa
ttccaagtac
4081 ttggatgagg aactgatggt cctaagctca aacagcatgt ctctgaccac
acgggagccc
4141 tctcttgcca cctgggaagc tacctggagt gaagggtcca agtcttcaga
cagagaagga
4201 acccgagcag gagaagagga gaggagacag ccacaagagg aagttgttca
tgaggaccaa
4261 ggaaagaagc cgcaggatca gcttgttatt gagagagaga ggaaatggga
gcaacagctt
4321 caggaagagc aagagcaaaa gcggcttcag gctgaggctg aggagcagaa
gcgtcctgcg
4381 gaggagcaga agcgccaggc agagatagag cgggaaacat cagtcagaat
ataccagtac
4441 aggaggcctg ttgattccta tgatatacca aagacagaag aagcatcttc
aggttttctt
4501 cctggtgaca ggaataaatc cagatctact actgaactgg atgattactc
cacaaataaa
4561 aatggaaaca ataaatattt agaccaaatt gggaacatga cctcttcaca
gaggagatcc
4621 aagaaagaac aagtaccatc aggagcagaa ttggagaggc aacaaatcct
tcaggaaatg
4681 aggaagagaa cacccttca caatgacaac agctggatcc gacagcgcag
tgccagtgtc
4741 aacaaagagc ctgttagtct tcctgggatc atgagaagag gcgaatcttt
agataacctg
4801 gactcccccc gatccaattc ttggagacag cctccttggc tcaatcagcc
cacaggattc
4861 tatgcttctt cctctgtgca agactttagt cgcccaccac ctcagctggt
gtccacatca
4921 aaccgtgcct acatgcggaa cccctcctcc agcgtgcccc caccttcagc
tggctccgtg
4981 aagacctcca ccacaggtgt ggccaccaca cagtccccca ccccgagaag
ccattcccct
5041 tcagcttcac agtcaggctc tcagctgcgt aacaggtcag tcagtgggaa
gcgcatatgc

5101 tcctactgca ataacattct gggcaaagga gccgccatga tcatcgagtc
cctgggtctt
5161 tgttatcatt tgcattgttt taagtgtgtt gcctgtgagt gtgacctcgg
aggctcttcc
5221 tcaggagctg aagtcaggat cagaaaccac caactgtact gcaacgactg
ctatctcaga
5281 ttcaaatctg gacggccaac cgccatgtga tgtaagcctc catacgaaag
cactgttgca
5341 gatagaagaa gaggtggttg ctgctcatgt agatctataa atatgtgttg
tatgtctttt
5401 ttgctttttt tttaaaaaaa agaataactt tttttgcctc tttagattac
atagaagcat
5461 tgtagtcttg gtagaaccag tatttttgtt gtttatttat aaggtaattg
tgtgtgggga
5521 aaagtgcagt atttacctgt tgaattcagc atcttgagag cacaagggaa
aaaataagaa
5581 cctacgaata tttttgaggc agataatgat ctagtttgac tttctagtta
gtggtgtttt
5641 gaagagggta ttttattgtt ttttaaaaaa aggttcttaa acattatttg
aaatagttaa
5701 tataaataca taattgcatt tgctctgttt attgtaatgt attctaaatt
aatgcagaac
5761 catatggaaa atttcattaa aatctatccc caaatgtgct ttctgtatcc
ttccttctac
5821 ctattattct gatttttaaa aatgcagtta atgtaccatt tatttgcttg
atgaagggag
5881 ctctattttc tttaccagaa atgttgctaa gtaattccca atagaaagct
gcttattttc
5941 attaatgaaa aataaccatg gtttgtatac tagaagtctt cttcagaaac
tggtgagcct
6001 ttctgttcaa ttgcatttgt aaataaactt gctgatgcat ttaacgagtg
ggtcgtcttt
6061 ttcttaggtg tatgtgtctg acctcaggcc ttttagccat atttcagtat
gtggcctttt
6121 ttgatgttat gttttatcca gtagctttac taaggtataa ttgatgtaat
aaactgcata
6181 tatttaaagt gtatactttg acaaattttg acatggtgta taccttcgaa
actatgccac
6241 agtctggatg tgtttactga aacattttaa taaggaagtt tatttttgat
aaagttatgt
6301 ttttggatac aatatatttg tatggtgaga gtgatgaatt gttggatcat
ttgaataaaa
6361 tcttttacta accccatgat aaaaggagaa gacaacagtg agcttagaat
atctataaag
6421 caaaaaatgt agtctcttgt ttaaaaaatc tggagcggga atgcaaggat
acaaaacttt
6481 agcatgcttt gagcaaaaat ttaaacttac tggaatcttt tataataatg
taagtggaat
6541 ggaggattct aggaactgag aactgtattg gaataggttc aaaatatgta
agaaatgcta
6601 atgtgggaga taaaaatttt atttagtact tattctgatt attattaaag
taataatgtg
6661 ttccttgagg ataacttgtc aaatgcccca aagcataaag aatataattc
tgaatcccaa
6721 attccaaaga caagaactct gtgtttgaat tcattctgca tataattatt
tataagtata
6781 gattgtgaat ttttccatgt tcttaaaatt atttttatct tttttcatgg
ttgcatagtg
6841 ctccattgtt tggccttggt aatatttagt tgataattcc attactgtgt
atttttcact

```
     6901 tgtttctaag atcaaacatt ttaatatgtg catgttatat ataaatatgt
aaattctgtg
     6961 atactctatg atcatctctt tctttatatt attttcatag acatgaaata
gttgctcaga
     7021 gattatgcat tttaagacac tcatagtata tattgccaaa gtggtttcca
gaaaggcact
     7081 gctggcttcg actcctataa gcagcacgtg ggcttgttca tctcactgca
tgtttatgaa
     7141 gatacagttc ttttgccttg ttctctgcct gatgtgtatg cagaggcagc
cctcaatatg
      7201 cagtggttga ataaatgaat gaagaaacca ctatcaaaaa aaaaaaaaaa aaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_005349.3

LOCUS NP_005349

ACCESSION NP_005349

```
      1 mkkirichif tfyswmsydv lfqrtelgal eiwrqlicah vcicvgwlyl
rdrvcskkdi
       61 ilrteqnsgr tilikavtek nfetkdfras lengvllcdl inklkpgvik
kinrlstpia
      121 gldninvflk aceqiglkea qlfhpgdlqd lsnrvtvkqe etdrrvknvl
itlywlgrka
      181 qsnpyyngph lnlkafenll gqaltkaled ssflkrsgrd sgygdiwcpe
rgeflapprh
      241 hkredsfesl dslgsrslts cssditlrgg regfesdtds eftfkmqdyn
kddmsyrris
      301 avepktalpf nrflpnksrq psyvpaplrk kkpdkhednr rswaspvyte
adgtfsrlfq
      361 kiygengsks msdvsaedvq nlrqlryeem qkiksqlkeq dqkwqddlak
wkdrrksyts
      421 dlqkkkeere eiekqaleks krssktfkem lqdresqnqk stvpsrrrmy
sfddvleegk
      481 rpptmtvsea syqserveek gatypseipk edsttfakre drvtteiqlp
sqspveeqsp
      541 aslsslrsrs tqmestrvsa slprsyrktd tvrltsvvtp rpfgsqtrgi
sslprsytmd
      601 dawkyngdve dikrtpnnvv stpapspdas qlasslssqk evaateedvt
rlpsptspfs
      661 slsqdqaats katlsstsgl dlmsesgege ispqrevsrs qdqfsdmris
inqtpgksld
      721 fgftikwdip gifvasveag spaefsqlqv ddeiiainnt kfsyndskew
eeamakaqet
      781 ghlvmdvrry gkagspetkw idatsgiyns ekssnlsvtt dfseslqssn
ieskeingih
      841 desnafeska sesislknlk rrsqffeqgs sdsvvpdlpv ptisapsrwv
wdqeeerkrq
      901 erwqkeqdrl lqekyqreqe klreewqrak qeaerensky ldeelmvlss
nsmslttrep
      961 slatweatws egskssdreg trageeerrq pqeevvhedq gkkpqdqlvi
ererkweqql
     1021 qeeqeqkrlq aeaeeqkrpa eeqkrqaeie retsvriyqy rrpvdsydip
kteeassgfl
     1081 pgdrnksrst telddystnk ngnnkyldqi gnmtssqrrs kkeqvpsgae
lerqqilqem
```

```
     1141 rkrtplhndn swirqrsasv nkepvslpgi mrrgesldnl dsprsnswrq
ppwlnqptgf
     1201 yasssvqdfs rpppqlvsts nraymrnpss svpppsagsv ktsttgvatt
qsptprshsp
     1261 sasqsgsqlr nrsvsgkric sycnnilgkg aamiieslgl cyhlhcfkcv
acecdlggss
     1321 sgaevrirnh qlycndcylr fksgrptam
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_015842.2

LOCUS NM_015842

ACCESSION NM_015842

```
      1 aacaggtaat gtttaacgtg ccagtcacaa agatcacaga aacagtgtat
gcccgggcat
       61 aagatagcac gactgtgtat gctctggagg actgaaaggc tgtacaagcc
ctatgtattt
      121 tttttcaaat atacatatgc atgggtcttg ctgctgcctc ttttgctgac
tgtaattgga
      181 ctttgaagct tcgaagttat atcataaaaa tttgtaacct ttgtctgaga
gagagctcag
      241 ctaagcaatc actttccact tcttttcaca ggataatata aacgttttct
tgaaagcttg
      301 tgaacagatt ggattgaaag aagcccagct tttccatcct ggagatctac
aggatttatc
      361 aaatcgagtc actgtcaagc aagaagagac tgacaggaga gtgaaaaatg
ttttgataac
      421 attgtactgg ctgggaagaa aagcacaaag caacccgtac tataatggtc
cccatcttaa
      481 tttgaaagcg tttgagaatc ttttaggaca agcactgacg aaggcactcg
aagactccag
      541 cttcctgaaa agaagtggca gggacagtgg ctacggtgac atctggtgtc
ctgaacgtgg
      601 agaatttctt gctcctccaa ggcaccataa gagagaagat tcctttgaaa
gcttggactc
      661 tttgggctcg aggtcattga caagctgctc ctctgatatc acgttgagag
gggggcgtga
      721 aggttttgaa agtgacacag attcggaatt tacatttaag atgcaggatt
ataataaaga
      781 tgatatgtcg tatcgaagga tttcggctgt tgagccaaag actgcgttac
ccttcaatcg
      841 ttttttaccc aacaaaagta gacagccatc ctatgtacca gcacctctga
gaaagaaaaa
      901 gccagacaaa catgaggata acagaagaag ttgggcaagc ccggtttata
cagaagcaga
      961 tggaacattt tcaagtaatc agaggaggat ttggggcacc aatgtggaga
actggccaac
     1021 tgtacaagga acttcaaagt cctcttgtta tttggaagag gaaaaagcaa
agacaagaag
     1081 catacccaac attgtaaagg atgatcttta tgtgcgcaag ctcagtccag
tcatgccaaa
     1141 cccagggaat gcttttgatc agtttcttcc caaatgttgg accccagaag
atgtgaactg
     1201 gaaaagaata aaaagggaaa cttataagcc atggtataaa gaatttcagg
gattcagtca
```

1261 gtttttactg cttcaggccc tccaaacata ctctgatgac atcttgtctt
ctgaaacaca
1321 taccaaaatt gatcccactt ctggcccaag gctcataacc cgcaggaaga
atctctctta
1381 tgcaccaggc tatagaagag atgacctcga gatggcagcc ctggatcctg
acttagagaa
1441 tgatgatttc tttgtcagaa agactggggc tttccatgca aatccatatg
ttctccgagc
1501 ttttgaagac tttagaaagt tctctgagca agatgattct gtagagcgag
atataatttt
1561 acagtgtaga gaaggtgaac ttgtacttcc ggatttggaa aaagatgata
tgattgttcg
1621 ccgaattcca gcacagaaga aagaagtgcc gctgtctggg gccccagata
gataccaccc
1681 agtccctttt cccgaaccct ggactcttcc tccagaaatt caagcaaaat
ttctctgtgt
1741 acttgaaagg acatgcccat ccaaagaaaa aagtaatagc tgtagaatat
tagttccttc
1801 atatcggcag aagaaagatg acatgctgac acgtaagatt cagtcctgga
aactgggaac
1861 taccgtgcct cccatcagtt tcacccctgg cccctgcagt gaggctgact
tgaagagatg
1921 ggaggccatc cgggaggcca gcagacttag gcacaagaaa aggctgatgg
tggagagact
1981 ctttcaaaag atttatggtg agaatgggag taagtccatg agtgatgtca
gcgcagaaga
2041 tgttcaaaac ttgcgtcagc tgcgttacga ggagatgcag aaaataaaat
cacaattaaa
2101 agaacaagat cagaaatggc aggatgacct tgcaaaatgg aaagatcgtc
gaaaaagtta
2161 cacttcagat ctgcagaaga aaaaagaaga gagagaagaa attgaaaagc
aggcacttga
2221 gaagtctaag agaagctcta agacgtttaa ggaaatgctg caggacaggg
aatcccaaaa
2281 tcaaaagtct acagttccgt caagaaggag aatgtattct tttgatgatg
tgctggagga
2341 aggaaagcga ccccctacaa tgactgtgtc agaagcaagt taccagagtg
agagagtaga
2401 agagaaggga gcaacttatc cttcagaaat tcccaaagaa gattctacca
cttttgcaaa
2461 aagagaggac cgtgtaacaa ctgaaattca gcttccttct caaagtcctg
tggaagaaca
2521 aagcccagcc tctttgtctt ctctgcgttc acggagcaca caaatggaat
caactcgtgt
2581 ttcagcttct ctccccagaa gttaccggaa aactgataca gtcaggttaa
catctgtggt
2641 cacaccaaga ccctttggct ctcagacaag gggaatctca tcactcccca
gatcttacac
2701 gatggatgat gcttggaagt ataatggaga tgttgaagac attaagagaa
ctccaaacaa
2761 tgtggtcagc acccctgcac caagcccgga cgcaagccaa ctggcttcaa
gcttatctag
2821 ccagaaagag gtagcagcaa cagaagaaga tgtgacaagg ctgccctctc
ctacatcccc
2881 cttctcatct ctttcccaag accaggctgc cacttctaaa gccacattgt
cttccacatc
2941 tggtcttgat ttaatgtctg aatctggaga aggggaaatc tccccacaaa
gagaagtctc
3001 aagatcccag gatcagttca gtgatatgag aatcagcata aaccagacgc
ctgggaagag

3061 tcttgacttt gggtttacaa taaaatggga tattcctggg atcttcgtag
catcagttga
3121 agcaggtagc ccagcagaat tttctcagct acaagtagat gatgaaatta
ttgctattaa
3181 caacaccaag ttttcatata acgattcaaa agagtgggag gaagccatgg
ctaaggctca
3241 agaaactgga cacctagtga tggatgtgag gcgctatgga aaggctggtt
cacctgaaac
3301 aaagtggatt gatgcaactt ctggaattta caactcagaa aaatcttcaa
atctatctgt
3361 aacaactgat ttctccgaaa gccttcagag ttctaatatt gaatccaaag
aaatcaatgg
3421 aattcatgat gaaagcaatg cttttgaatc aaaagcatct gaatccattt
ctttgaaaaa
3481 cttaaaaagg cgatcacaat tttttgaaca aggaagctct gattcggtgg
ttcctgatct
3541 tccagttcca accatcagtg ccccgagtcg ctgggtgtgg gatcaagagg
aggagcggaa
3601 gcggcaggag aggtggcaga aggagcagga ccgcctactg caggaaaaat
atcaacgtga
3661 gcaggagaaa ctgagggaag agtggcaaag ggccaaacag gaggcagaga
gagagaattc
3721 caagtacttg gatgaggaac tgatggtcct aagctcaaac agcatgtctc
tgaccacacg
3781 ggagccctct cttgccacct gggaagctac ctggagtgaa gggtccaagt
cttcagacag
3841 agaaggaacc cgagcaggag aagaggagag gagacagcca caagaggaag
ttgttcatga
3901 ggaccaagga aagaagccgc aggatcagct tgttattgag agagagagga
aatgggagca
3961 acagcttcag gaagagcaag agcaaaagcg gcttcaggct gaggctgagg
agcagaagcg
4021 tcctgcggag gagcagaagc gccaggcaga gatagagcgg gaaacatcag
tcagaatata
4081 ccagtacagg aggcctgttg attcctatga tataccaaag acagaagaag
catcttcagg
4141 ttttcttcct ggtgacagga ataaatccag atctactact gaactggatg
attactccac
4201 aaataaaaat ggaaacaata aatatttaga ccaaattggg aacatgacct
cttcacagag
4261 gagatccaag aaagaacaag taccatcagg agcagaattg gagaggcaac
aaatccttca
4321 ggaaatgagg aagagaacac cccttcacaa tgacaacagc tggatccgac
agcgcagtgc
4381 cagtgtcaac aaagagcctg ttagtcttcc tgggatcatg agaagaggcg
aatctttaga
4441 taacctggac tccccccgat ccaattcttg gagacagcct ccttggctca
atcagcccac
4501 aggattctat gcttcttcct ctgtgcaaga ctttagtcgc ccaccacctc
agctggtgtc
4561 cacatcaaac cgtgcctaca tgcggaaccc ctcctccagc gtgcccccac
cttcagctgg
4621 ctccgtgaag acctccacca caggtgtggc caccacacag tcccccaccc
cgagaagcca
4681 ttccccttca gcttcacagt caggctctca gctgcgtaac agtgtgttgc
ctgtgagtgt
4741 gacctcggag gctcttcctc aggagctgaa gtcaggatca gaaaccacca
actgtactgc
4801 aacgactgct atctcagatt caaatctgga cggccaaccg ccatgtgatg
taagcctcca

4861 tacgaaagca ctgttgcaga tagaagaaga ggtggttgct gctcatgtag
atctataaat
4921 atgtgttgta tgtctttttt gctttttttt taaaaaaaag aataactttt
tttgcctctt
4981 tagattacat agaagcattg tagtcttggt agaaccagta tttttgttgt
ttatttataa
5041 ggtaattgtg tgtggggaaa agtgcagtat ttacctgttg aattcagcat
cttgagagca
5101 caagggaaaa aataagaacc tacgaatatt tttgaggcag ataatgatct
agtttgactt
5161 tctagttagt ggtgttttga agagggtatt ttattgtttt ttaaaaaaag
gttcttaaac
5221 attatttgaa atagttaata taaatacata attgcatttg ctctgtttat
tgtaatgtat
5281 tctaaattaa tgcagaacca tatggaaaat ttcattaaaa tctatcccca
aatgtgcttt
5341 ctgtatcctt ccttctacct attattctga tttttaaaaa tgcagttaat
gtaccattta
5401 tttgcttgat gaagggagct ctattttctt taccagaaat gttgctaagt
aattcccaat
5461 agaaagctgc ttattttcat taatgaaaaa taaccatggt ttgtatacta
gaagtcttct
5521 tcagaaactg gtgagccttt ctgttcaatt gcatttgtaa ataaacttgc
tgatgcattt
5581 aacgagtggg tcgtcttttt cttaggtgta tgtgtctgac ctcaggcctt
ttagccatat
5641 ttcagtatgt ggccttttt gatgttatgt tttatccagt agctttacta
aggtataatt
5701 gatgtaataa actgcatata tttaaagtgt atactttgac aaattttgac
atggtgtata
5761 ccttcgaaac tatgccacag tctggatgtg tttactgaaa cattttaata
aggaagttta
5821 tttttgataa agttatgttt ttggatacaa tatatttgta tggtgagagt
gatgaattgt
5881 tggatcattt gaataaaatc ttttactaac cccatgataa aaggagaaga
caacagtgag
5941 cttagaatat ctataaagca aaaaatgtag tctcttgttt aaaaaatctg
gagcgggaat
6001 gcaaggatac aaaactttag catgctttga gcaaaaattt aaacttactg
gaatctttta
6061 taataatgta agtggaatgg aggattctag gaactgagaa ctgtattgga
ataggttcaa
6121 aatatgtaag aaatgctaat gtgggagata aaaattttat ttagtactta
ttctgattat
6181 tattaaagta ataatgtgtt ccttgaggat aacttgtcaa atgccccaaa
gcataaagaa
6241 tataattctg aatcccaaat tccaaagaca agaactctgt gtttgaattc
attctgcata
6301 taattattta taagtataga ttgtgaattt ttccatgttc ttaaaattat
ttttatcttt
6361 tttcatggtt gcatagtgct ccattgtttg gccttggtaa tatttagttg
ataattccat
6421 tactgtgtat ttttcacttg tttctaagat caaacatttt aatatgtgca
tgttatatat
6481 aaatatgtaa attctgtgat actctatgat catctctttc tttatattat
tttcatagac
6541 atgaaatagt tgctcagaga ttatgcattt taagacactc atagtatata
ttgccaaagt
6601 ggtttccaga aaggcactgc tggcttcgac tcctataagc agcacgtggg
cttgttcatc

```
     6661 tcactgcatg tttatgaaga tacagttctt ttgccttgtt ctctgcctga
tgtgtatgca
     6721 gaggcagccc tcaatatgca gtggttgaat aaatgaatga agaaaccact
atcaaaaaaa
      6781 aaaaaaaaaa a
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_056667.2

LOCUS NP_056667

ACCESSION NP_056667

```
      1 mqdynkddms yrrisavepk talpfnrflp nksrqpsyvp aplrkkkpdk
hednrrswas
       61 pvyteadgtf ssnqrriwgt nvenwptvqg tsksscylee ekaktrsipn
ivkddlyvrk
      121 lspvmpnpgn afdqflpkcw tpedvnwkri kretykpwyk efqgfsqfll
lqalqtysdd
      181 ilssethtki dptsgprlit rrknlsyapg yrrddlemaa ldpdlenddf
fvrktgafha
      241 npyvlrafed frkfseqdds verdiilqcr egelvlpdle kddmivrrip
aqkkevplsg
      301 apdryhpvpf pepwtlppei qakflcvler tcpskeksns crilvpsyrq
kkddmltrki
      361 qswklgttvp pisftpgpcs eadlkrweai reasrlrhkk rlmverlfqk
iygengsksm
      421 sdvsaedvqn lrqlryeemq kiksqlkeqd qkwqddlakw kdrrksytsd
lqkkkeeree
      481 iekqaleksk rssktfkeml qdresqnqks tvpsrrrmys fddvleegkr
pptmtvseas
      541 yqserveekg atypseipke dsttfakred rvtteiqlps qspveeqspa
slsslrsrst
      601 qmestrvsas lprsyrktdt vrltsvvtpr pfgsqtrgis slprsytmdd
awkyngdved
      661 ikrtpnnvvs tpapspdasq lasslssqke vaateedvtr lpsptspfss
lsqdqaatsk
      721 atlsstsgld lmsesgegei spqrevsrsq dqfsdmrisi nqtpgksldf
gftikwdipg
      781 ifvasveags paefsqlqvd deiiainntk fsyndskewe eamakaqetg
hlvmdvrryg
      841 kagspetkwi datsgiynse kssnlsvttd fseslqssni eskeingihd
esnafeskas
      901 esislknlkr rsqffeqgss dsvvpdlpvp tisapsrwvw dqeeerkrqe
rwqkeqdrll
      961 qekyqreqek lreewqrakq eaerenskyl deelmvlssn smslttreps
latweatwse
     1021 gskssdregt rageeerrqp qeevvhedqg kkpqdqlvie rerkweqqlq
eeqeqkrlqa
     1081 eaeeqkrpae eqkrqaeier etsvriyqyr rpvdsydipk teeassgflp
gdrnksrstt
     1141 elddystnkn gnnkyldqig nmtssqrrsk keqvpsgael erqqilqemr
krtplhndns
     1201 wirqrsasvn kepvslpgim rrgesldnld sprsnswrqp pwlnqptgfy
asssvqdfsr
     1261 pppqlvstsn raymrnpsss vpppsagsvk tsttgvattq sptprshsps
asqsgsqlrn
     1321 svlpvsvtse alpqelksgs ettnctatta isdsnldgqp pcdvslhtka
llqieeevva
```

1381 ahvdl

CARS

Official Symbol: CARS

Official Name: cysteinyl-tRNA synthetase

Gene ID: 833

Organism: Homo sapiens

Other Aliases: CARS1, CYSRS, MGC:11246

Other Designations: cysteine tRNA ligase 1, cytoplasmic; cysteine translase; cysteine--tRNA ligase, cytoplasmic

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_139273.3

LOCUS NM_139273

ACCESSION NM_139273

1 gtggggcgcg acttccgggg cggcggttgc atcagattct aggaagtgtc
tgtagccgca
61 gctgcgggtc cgggattccc agccatggca gattcctccg ggcagcaggg
caaaggccgg
121 cgtgtgcagc cccagtggtc ccctcctgct gggacccagc catgcagact
ccacctttac
181 aacagcctca ccaggaacaa ggaagtgttc atacctcaag atgggaaaaa
ggtgacgtgg
241 tattgctgtg ggccaaccgt ctatgacgca tctcacatgg ggcacgccag
gtcctacatc
301 tcttttgata tcttgagaag agtgttgaag gattacttca aatttgatgt
cttttattgc
361 atgaacatta cggatattga tgacaagatc atcaagaggg cccggcagaa
ccacctgttc
421 gagcagtatc gggagaagag gcctgaagcg gcacagctct tggaggatgt
tcaggccgcc
481 ctgaagccat tttcagtaaa attaaatgag accacggatc ccgataaaaa
gcagatgctc
541 gaacggattc agcacgcagt gcagcttgcc acagagccac ttgagaaagc
tgtgcagtcc
601 agactcacgg gagaggaagt caacagctgt gtggaggtgt tgctggaaga
agccaaggat
661 ttgctctctg actggctgga ttctacactt ggctgtgatg tcactgacaa
ttccatcttc
721 tccaagctgc ccaagttctg ggaggggggac ttccacagag acatggaagc
tctgaatgtt
781 ctccctccag atgtcttaac ccgggttagt gagtatgtgc cagaaattgt
gaactttgtc

```
      841 cagaagattg tggacaacgg ttacggctat gtctccaatg ggtctgtcta
ctttgataca
      901 gcgaagtttg cttctagcga gaagcactcc tatgggaagc tggtgcctga
ggccgttgga
      961 gatcagaaag cccttcaaga aggggaaggt gacctgagca tctctgcaga
ccgcctgagt
     1021 gagaagcgct ctcccaacga ctttgcctta tggaaggcct ctaagcccgg
agaaccgtcc
     1081 tggccgtgcc cttggggaaa gggtcgtccg ggctggcata tcgagtgctc
ggccatggca
     1141 ggcaccctcc taggggcttc gatggacatt cacggaggtg ggttcgacct
ccggttcccc
     1201 caccatgaca atgagctggc acagtcggag gcctactttg aaaacgactg
ctgggtcagg
     1261 tacttcctgc acacaggcca cctgaccatt gcaggctgca aaatgtcaaa
gtcactaaaa
     1321 aacttcatca ccattaaaga tgccttgaaa aagcactcag cacggcagtt
gcggctggcc
     1381 ttcctcatgc actcgtggaa ggacaccctg gactactcca gcaacaccat
ggagtcagcg
     1441 cttcaatatg agaagttctt gaatgagttt ttcttaaatg tgaaagatat
ccttcgcgct
     1501 cctgttgaca tcactggtca gtttgagaag tggggagaag aagaagcaga
actgaataag
     1561 aacttttatg acaagaagac agcaattcac aaagccctct gtgacaatgt
tgacacccgc
     1621 accgtcatgg aagagatgcg ggccttggtc agtcagtgca acctctatat
ggcagcccgg
     1681 aaagccgtga ggaagaggcc caaccaggct ctgctggaga acatcgccct
gtacctcacc
     1741 catatgctga agatctttgg ggccgtagaa gaggacagct ccctgggatt
cccggtcgga
     1801 gggcctggaa ccagcctcag tctcgaggcc acagtcatgc cctaccttca
ggtgttatca
     1861 gaattccgag aaggagtgcg gaagattgcc cgagagcaaa aagtccctga
gattctgcag
     1921 ctcagcgatg ccctgcggga caacatcctg cccgagcttg gggtgcggtt
tgaagaccac
     1981 gaaggactgc ccacagtggt gaaactggta gacagaaaca ccttattaaa
agagagagaa
     2041 gaaaagagac gggttgaaga ggagaagagg aagaagaaag aggaggcggc
ccggaggaaa
     2101 caggaacaag aagcagcaaa gctggccaag atgaagattc cccccagtga
gatgttcttg
     2161 tcagaaaccg acaaatactc caagtttgat gaaaatgtaa gcatggtctg
cccacacatg
     2221 acatggaggg caaagagctc agcaaagggc aagccaagaa gctgaagaag
ctcttcgagg
     2281 ctcaggagaa gctctacaag gaatatctgc agatggccca gaatggaagc
ttccagtgag
     2341 ggggcacagg actgactttt taaaccattg tggactagtg gctgctgtct
gcctcagtga
     2401 caatgtccca gcgctcctat catgtttaca gtcacccttg ggtcctaaat
taagagttgt
     2461 gttcatgtag gttcgtgtcg tcgttggctc tgagacattg ataataaatt
tttctcaaca
      2521 gtgagaccct caaaaaaaaa aaaaaaaaaa aaaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_644802.1

LOCUS NP_644802

ACCESSION NP_644802

```
      1 madssgqqgk grrvqpqwsp pagtqpcrlh lynsltrnke vfipqdgkkv
twyccgptvy
       61 dashmghars yisfdilrrv lkdyfkfdvf ycmnitdidd kiikrarqnh
lfeqyrekrp
      121 eaaqlledvq aalkpfsvkl nettdpdkkq mleriqhavq lateplekav
qsrltgeevn
      181 scvevlleea kdllsdwlds tlgcdvtdns ifsklpkfwe gdfhrdmeal
nvlppdvltr
      241 vseyvpeivn fvqkivdngy gyvsngsvyf dtakfassek hsygklvpea
vgdqkalqeg
      301 egdlsisadr lsekrspndf alwkaskpge pswpcpwgkg rpgwhiecsa
magtllgasm
      361 dihgggfdlr fphhdnelaq seayfendcw vryflhtghl tiagckmsks
lknfitikda
      421 lkkhsarqlr laflmhswkd tldyssntme salqyekfln efflnvkdil
rapvditgqf
      481 ekwgeeeael nknfydkkta ihkalcdnvd trtvmeemra lvsqcnlyma
arkavrkrpn
      541 qallenialy lthmlkifga veedsslgfp vggpgtslsl eatvmpylqv
lsefregvrk
      601 iareqkvpei lqlsdalrdn ilpelgvrfe dheglptvvk lvdrntllke
reekrrveee
      661 krkkkeeaar rkqeqeaakl akmkippsem flsetdkysk fdenvsmvcp
hmtwrakssa
      721 kgkprs
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001751.5

LOCUS NM_001751

ACCESSION NM_001751

```
      1 gtggggcgcg acttccgggg cggcggttgc atcagattct aggaagtgtc
tgtagccgca
       61 gctgcgggtc cgggattccc agccatggca gattcctccg ggcagcaggg
caaaggccgg
      121 cgtgtgcagc cccagtggtc ccctcctgct gggacccagc catgcagact
ccacctttac
      181 aacagcctca ccaggaacaa ggaagtgttc atacctcaag atgggaaaaa
ggtgacgtgg
      241 tattgctgtg ggccaaccgt ctatgacgca tctcacatgg ggcacgccag
gtcctacatc
      301 tcttttgata tcttgagaag agtgttgaag gattacttca aatttgatgt
cttttattgc
      361 atgaacatta cggatattga tgacaagatc atcaagaggg cccggcagaa
ccacctgttc
      421 gagcagtatc gggagaagag gcctgaagcg gcacagctct tggaggatgt
tcaggccgcc
      481 ctgaagccat tttcagtaaa attaaatgag accacggatc ccgataaaaa
gcagatgctc
      541 gaacggattc agcacgcagt gcagcttgcc acagagccac ttgagaaagc
tgtgcagtcc
```

601 agactcacgg gagaggaagt caacagctgt gtggaggtgt tgctggaaga
agccaaggat
661 ttgctctctg actggctgga ttctacactt ggctgtgatg tcactgacaa
ttccatcttc
721 tccaagctgc ccaagttctg ggaggggggac ttccacagag acatggaagc
tctgaatgtt
781 ctccctccag atgtcttaac ccgggttagt gagtatgtgc cagaaattgt
gaactttgtc
841 cagaagattg tggacaacgg ttacggctat gtctccaatg ggtctgtcta
ctttgataca
901 gcgaagtttg cttctagcga gaagcactcc tatgggaagc tggtgcctga
ggccgttgga
961 gatcagaaag cccttcaaga aggggaaggt gacctgagca tctctgcaga
ccgcctgagt
1021 gagaagcgct ctcccaacga ctttgcctta tggaaggcct ctaagcccgg
agaaccgtcc
1081 tggccgtgcc cttggggaaa gggtcgtccg ggctggcata tcgagtgctc
ggccatggca
1141 ggcaccctcc taggggcttc gatggacatt cacggaggtg ggttcgacct
ccggttcccc
1201 caccatgaca atgagctggc acagtcggag gcctactttg aaaacgactg
ctgggtcagg
1261 tacttcctgc acacaggcca cctgaccatt gcaggctgca aaatgtcaaa
gtcactaaaa
1321 aacttcatca ccattaaaga tgccttgaaa aagcactcag cacggcagtt
gcggctggcc
1381 ttcctcatgc actcgtggaa ggacaccctg gactactcca gcaacaccat
ggagtcagcg
1441 cttcaatatg agaagttctt gaatgagttt ttcttaaatg tgaaagatat
ccttcgcgct
1501 cctgttgaca tcactggtca gtttgagaag tggggagaag aagaagcaga
actgaataag
1561 aacttttatg acaagaagac agcaattcac aaagccctct gtgacaatgt
tgacacccgc
1621 accgtcatgg aagagatgcg ggccttggtc agtcagtgca acctctatat
ggcagcccgg
1681 aaagccgtga ggaagaggcc caaccaggct ctgctggaga acatcgccct
gtacctcacc
1741 catatgctga agatctttgg ggccgtagaa gaggacagct ccctgggatt
cccggtcgga
1801 gggcctggaa ccagcctcag tctcgaggcc acagtcatgc cctaccttca
ggtgttatca
1861 gaattccgag aaggagtgcg gaagattgcc cgagagcaaa aagtccctga
gattctgcag
1921 ctcagcgatg ccctgcggga caacatcctg cccgagcttg gggtgcggtt
tgaagaccac
1981 gaaggactgc ccacagtggt gaaactggta gacagaaaca ccttattaaa
agagagagaa
2041 gaaaagagac gggttgaaga ggagaagagg aagaagaaag aggaggcggc
ccggaggaaa
2101 caggaacaag aagcagcaaa gctggccaag atgaagattc cccccagtga
gatgttcttg
2161 tcagaaaccg acaaatactc caagtttgat gaaaatggtc tgcccacaca
tgacatggag
2221 ggcaaagagc tcagcaaagg gcaagccaag aagctgaaga agctcttcga
ggctcaggag
2281 aagctctaca aggaatatct gcagatggcc cagaatggaa gcttccagtg
aggggcaca
2341 ggactgactt tttaaaccat tgtggactag tggctgctgt ctgcctcagt
gacaatgtcc

```
     2401 cagcgctcct atcatgttta cagtcaccct tgggtcctaa attaagagtt
gtgttcatgt
     2461 aggttcgtgt cgtcgttggc tctgagacat tgataataaa tttttctcaa
cagtgagacc
      2521 ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001742.1

LOCUS NP_001742

ACCESSION NP_001742

```
      1 madssgqqgk grrvqpqwsp pagtqpcrlh lynsltrnke vfipqdgkkv
twyccgptvy
       61 dashmghars yisfdilrrv lkdyfkfdvf ycmnitdidd kiikrarqnh
lfeqyrekrp
      121 eaaqlledvq aalkpfsvkl nettdpdkkq mleriqhavq lateplekav
qsrltgeevn
      181 scvevlleea kdllsdwlds tlgcdvtdns ifsklpkfwe gdfhrdmeal
nvlppdvltr
      241 vseyvpeivn fvqkivdngy gyvsngsvyf dtakfassek hsygklvpea
vgdqkalqeg
      301 egdlsisadr lsekrspndf alwkaskpge pswpcpwgkg rpgwhiecsa
magtllgasm
      361 dihgggfdlr fphhdnelaq seayfendcw vryflhtghl tiagckmsks
lknfitikda
      421 lkkhsarqlr laflmhswkd tldyssntme salqyekfln efflnvkdil
rapvditgqf
      481 ekwgeeeael nknfydkkta ihkalcdnvd trtvmeemra lvsqcnlyma
arkavrkrpn
      541 qallenialy lthmlkifga veedsslgfp vggpgtslsl eatvmpylqv
lsefregvrk
      601 iareqkvpei lqlsdalrdn ilpelgvrfe dheglptvvk lvdrntllke
reekrrveee
      661 krkkkeeaar rkqeqeaakl akmkippsem flsetdkysk fdenglpthd
megkelskgq
      721 akklkklfea qeklykeylq maqngsfq
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_001014437.2

LOCUS NM_001014437

ACCESSION NM_001014437

```
      1 gtggggcgcg acttccgggg cggcggttgc atcagattct aggaagtgtc
tgtagccgca
       61 gctgcgggtc cgggattccc agccatggca gattcctccg ggcagcaggc
tcctgactac
      121 aggtccattc tgagcattag tgacgaggca gccagggcac aagccctgaa
cgagcacctc
      181 agcacgcgta gctatgtcca ggggtactca ctgtcccagg cagacgtgga
cgcgttcagg
      241 cagctctcgg ccccgcccgc tgacccccag ctcttccacg tggctcggtg
gttcaggcac
```

301 atagaagcgc tcctgggtag cccctgtggc aaaggccagc cctgcaggct
ccaagcaagc
361 aaaggccggc gtgtgcagcc ccagtggtcc cctcctgctg ggacccagcc
atgcagactc
421 caccttaca acagcctcac caggaacaag gaagtgttca tacctcaaga
tgggaaaaag
481 gtgacgtggt attgctgtgg gccaaccgtc tatgacgcat ctcacatggg
gcacgccagg
541 tcctacatct cttttgatat cttgagaaga gtgttgaagg attacttcaa
atttgatgtc
601 ttttattgca tgaacattac ggatattgat gacaagatca tcaagagggc
ccggcagaac
661 cacctgttcg agcagtatcg ggagaagagg cctgaagcgg cacagctctt
ggaggatgtt
721 caggccgccc tgaagccatt ttcagtaaaa ttaaatgaga ccacggatcc
cgataaaaag
781 cagatgctcg aacggattca gcacgcagtg cagcttgcca cagagccact
tgagaaagct
841 gtgcagtcca gactcacggg agaggaagtc aacagctgtg tggaggtgtt
gctggaagaa
901 gccaaggatt tgctctctga ctggctggat tctacacttg gctgtgatgt
cactgacaat
961 tccatcttct ccaagctgcc caagttctgg gaggggact tccacagaga
catggaagct
1021 ctgaatgttc tccctccaga tgtcttaacc cgggttagtg agtatgtgcc
agaaattgtg
1081 aactttgtcc agaagattgt ggacaacggt tacggctatg tctccaatgg
gtctgtctac
1141 tttgatacag cgaagtttgc ttctagcgag aagcactcct atgggaagct
ggtgcctgag
1201 gccgttggag atcagaaagc ccttcaagaa ggggaaggtg acctgagcat
ctctgcagac
1261 cgcctgagtg agaagcgctc tcccaacgac tttgccttat ggaaggcctc
taagcccgga
1321 gaaccgtcct ggccgtgccc ttggggaaag ggtcgtccgg gctggcatat
cgagtgctcg
1381 gccatggcag gcaccctcct aggggcttcg atggacattc acggaggtgg
gttcgacctc
1441 cggttccccc accatgacaa tgagctggca cagtcggagg cctactttga
aaacgactgc
1501 tgggtcaggt acttcctgca cacaggccac ctgaccattg caggctgcaa
aatgtcaaag
1561 tcactaaaaa acttcatcac cattaaagat gccttgaaaa agcactcagc
acggcagttg
1621 cggctggcct tcctcatgca ctcgtggaag gacaccctgg actactccag
caacaccatg
1681 gagtcagcgc ttcaatatga gaagttcttg aatgagtttt tcttaaatgt
gaaagatatc
1741 cttcgcgctc ctgttgacat cactggtcag tttgagaagt ggggagaaga
agaagcagaa
1801 ctgaataaga acttttatga caagaagaca gcaattcaca aagccctctg
tgacaatgtt
1861 gacacccgca ccgtcatgga agagatgcgg gccttggtca gtcagtgcaa
cctctatatg
1921 gcagcccgga aagccgtgag gaagaggccc aaccaggctc tgctggagaa
catcgccctg
1981 tacctcaccc atatgctgaa gatctttggg gccgtagaag aggacagctc
cctgggattc
2041 ccggtcggag ggcctggaac cagcctcagt ctcgaggcca cagtcatgcc
ctaccttcag

```
     2101 gtgttatcag aattccgaga aggagtgcgg aagattgccc gagagcaaaa
agtccctgag
     2161 attctgcagc tcagcgatgc cctgcgggac aacatcctgc ccgagcttgg
ggtgcggttt
     2221 gaagaccacg aaggactgcc cacagtggtg aaactggtag acagaaacac
cttattaaaa
     2281 gagagagaag aaaagagacg ggttgaagag gagaagagga agaagaaaga
ggaggcggcc
     2341 cggaggaaac aggaacaaga agcagcaaag ctggccaaga tgaagattcc
ccccagtgag
     2401 atgttcttgt cagaaaccga caaatactcc aagtttgatg aaaatggtct
gcccacacat
     2461 gacatggagg gcaaagagct cagcaaaggg caagccaaga agctgaagaa
gctcttcgag
     2521 gctcaggaga agctctacaa ggaatatctg cagatggccc agaatggaag
cttccagtga
     2581 gggggcacag gactgacttt ttaaaccatt gtggactagt ggctgctgtc
tgcctcagtg
     2641 acaatgtccc agcgctccta tcatgtttac agtcaccctt gggtcctaaa
ttaagagttg
     2701 tgttcatgta ggttcgtgtc gtcgttggct ctgagacatt gataataaat
ttttctcaac
     2761 agtgagaccc tcaaaaaaaa aaaaaaaaaa aaaaaaaaa
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_001014437.1

LOCUS NP_001014437

ACCESSION NP_001014437

```
      1 madssgqqap dyrsilsisd eaaraqalne hlstrsyvqg yslsqadvda
frqlsappad
       61 pqlfhvarwf rhieallgsp cgkgqpcrlq askgrrvqpq wsppagtqpc
rlhlynsltr
      121 nkevfipqdg kkvtwyccgp tvydashmgh arsyisfdil rrvlkdyfkf
dvfycmnitd
      181 iddkiikrar qnhlfeqyre krpeaaqlle dvqaalkpfs vklnettdpd
kkqmleriqh
      241 avqlateple kavqsrltge evnscvevll eeakdllsdw ldstlgcdvt
dnsifsklpk
      301 fwegdfhrdm ealnvlppdv ltrvseyvpe ivnfvqkivd ngygyvsngs
vyfdtakfas
      361 sekhsygklv peavgdqkal qegegdlsis adrlsekrsp ndfalwkask
pgepswpcpw
      421 gkgrpgwhie csamagtllg asmdihgggf dlrfphhdne laqseayfen
dcwvryflht
      481 ghltiagckm skslknfiti kdalkkhsar qlrlaflmhs wkdtldyssn
tmesalqyek
      541 flnefflnvk dilrapvdit gqfekwgeee aelnknfydk ktaihkalcd
nvdtrtvmee
      601 mralvsqcnl ymaarkavrk rpnqalleni alylthmlki fgaveedssl
gfpvggpgts
      661 lsleatvmpy lqvlsefreg vrkiareqkv peilqlsdal rdnilpelgv
rfedheglpt
      721 vvklvdrntl lkereekrrv eeekrkkkee aarrkqeqea aklakmkipp
semflsetdk
      781 yskfdenglp thdmegkels kgqakklkkl feaqeklyke ylqmaqngsf q
```

Nucleotide sequence (variant 5):

NCBI Reference Sequence: NM_001194997.1

LOCUS NM_001194997

ACCESSION NM_001194997

```
      1 gtggggcgcg acttccgggg cggcggttgc atcagattct aggaagtgtc
tgtagccgca
       61 gctgcgggtc cgggattccc agccatggca gattcctccg ggcagcaggc
tcctgactac
      121 aggtccattc tgagcattag tgacgaggca gccagggcac aagccctgaa
cgagcacctc
      181 agcacgcgta gctatgtcca ggggtactca ctgtcccagg cagacgtgga
cgcgttcagg
      241 cagctctcgg ccccgcccgc tgacccccag ctcttccacg tggctcggtg
gttcaggcac
      301 atagaagcgc tcctgggtag cccctgtggc aaaggccagc cctgcaggct
ccaagcaagc
      361 aaaggccggc gtgtgcagcc ccagtggtcc cctcctgctg ggacccagcc
atgcagactc
      421 caccttaca acagcctcac caggaacaag gaagtgttca tacctcaaga
tgggaaaaag
      481 gtgacgtggt attgctgtgg gccaaccgtc tatgacgcat ctcacatggg
gcacgccagg
      541 tcctacatct cttttgatat cttgagaaga gtgttgaagg attacttcaa
atttgatgtc
      601 ttttattgca tgaacattac ggatattgat gacaagatca tcaagagggc
ccggcagaac
      661 cacctgttcg agcagtatcg ggagaagagg cctgaagcgg cacagctctt
ggaggatgtt
      721 caggccgccc tgaagccatt ttcagtaaaa ttaaatgaga ccacggatcc
cgataaaaag
      781 cagatgctcg aacggattca gcacgcagtg cagcttgcca cagagccact
tgagaaagct
      841 gtgcagtcca gactcacggg agaggaagtc aacagctgtg tggaggtgtt
gctggaagaa
      901 gccaaggatt tgctctctga ctggctggat tctacacttg gctgtgatgt
cactgacaat
      961 tccatcttct ccaagctgcc caagttctgg gaggggggact tccacagaga
catggaagct
     1021 ctgaatgttc tccctccaga tgtcttaacc cgggttagtg agtatgtgcc
agaaattgtg
     1081 aactttgtcc agaagattgt ggacaacggt tacggctatg tctccaatgg
gtctgtctac
     1141 tttgatacag cgaagtttgc ttctagcgag aagcactcct atgggaagct
ggtgcctgag
     1201 gccgttggag atcagaaagc ccttcaagaa ggggaaggtg acctgagcat
ctctgcagac
     1261 cgcctgagtg agaagcgctc tcccaacgac tttgccttat ggaaggcctc
taagcccgga
     1321 gaaccgtcct ggccgtgccc ttggggaaag ggtcgtccgg gctggcatat
cgagtgctcg
     1381 gccatggcag gcaccctcct aggggcttcg atggacattc acggaggtgg
gttcgacctc
     1441 cggttccccc accatgacaa tgagctggca cagtcggagg cctactttga
aaacgactgc
     1501 tgggtcaggt acttcctgca cacaggccac ctgaccattg caggctgcaa
aatgtcaaag
```

```
     1561 tcactaaaaa acttcatcac cattaaagat gccttgaaaa agcactcagc
acggcagttg
     1621 cggctggcct tcctcatgca ctcgtggaag gacaccctgg actactccag
caacaccatg
     1681 gagtcagcgc ttcaatatga gaagttcttg aatgagtttt tcttaaatgt
gaaagatatc
     1741 cttcgcgctc ctgttgacat cactggtcag tttgagaagt ggggagaaga
agaagcagaa
     1801 ctgaataaga acttttatga caagaagaca gcaattcaca aagccctctg
tgacaatgtt
     1861 gacacccgca ccgtcatgga agagatgcgg gccttggtca gtcagtgcaa
cctctatatg
     1921 gcagcccgga aagccgtgag gaagaggccc aaccaggctc tgctggagaa
catcgccctg
     1981 tacctcaccc atatgctgaa gatctttggg gccgtagaag aggacagctc
cctgggattc
     2041 ccggtcggag ggcctggaac cagcctcagt ctcgaggcca cagtcatgcc
ctaccttcag
     2101 gtgttatcag aattccgaga aggagtgcgg aagattgccc gagagcaaaa
agtccctgag
     2161 attctgcagc tcagcgatgc cctgcgggac aacatcctgc ccgagcttgg
ggtgcggttt
     2221 gaagaccacg aaggactgcc cacagtggtg aaactggtag acagaaacac
cttattaaaa
     2281 gagagagaag aaaagagacg ggttgaagag gagaagagga agaagaaaga
ggaggcggcc
     2341 cggaggaaac aggaacaaga agcagcaaag ctggccaaga tgaagattcc
ccccagtgag
     2401 atgttcttgt cagaaaccga caaatactcc aagtttgatg aaaatgtaag
catggtctgc
     2461 ccacacatga catggagggc aaagagctca gcaaagggca agccaagaag
ctgaagaagc
     2521 tcttcgaggc tcaggagaag ctctacaagg aatatctgca gatggcccag
aatggaagct
     2581 tccagtgagg gggcacagga ctgacttttt aaaccattgt ggactagtgg
ctgctgtctg
     2641 cctcagtgac aatgtcccag cgctcctatc atgtttacag tcacccttgg
gtcctaaatt
     2701 aagagttgtg ttcatgtagg ttcgtgtcgt cgttggctct gagacattga
taataaattt
      2761 ttctcaacag tgagaccctc aaaaaaaaaa aaaaaaaaaa aaaaaaa
```

Protein sequence (variant 5):

NCBI Reference Sequence: NP_001181926.1

LOCUS NP_001181926

ACCESSION NP_001181926

```
      1 madssgqqap dyrsilsisd eaaraqalne hlstrsyvqg yslsqadvda
frqlsappad
       61 pqlfhvarwf rhieallgsp cgkgqpcrlq askgrrvqpq wsppagtqpc
rlhlynsltr
      121 nkevfipqdg kkvtwyccgp tvydashmgh arsyisfdil rrvlkdyfkf
dvfycmnitd
      181 iddkiikrar qnhlfeqyre krpeaaqlle dvqaalkpfs vklnettdpd
kkqmleriqh
      241 avqlateple kavqsrltge evnscvevll eeakdllsdw ldstlgcdvt
dnsifsklpk
```

```
      301 fwegdfhrdm ealnvlppdv ltrvseyvpe ivnfvqkivd ngygyvsngs
vyfdtakfas
      361 sekhsygklv peavgdqkal qegegdlsis adrlsekrsp ndfalwkask
pgepswpcpw
      421 gkgrpgwhie csamagtllg asmdihgggf dlrfphhdne laqseayfen
dcwvryflht
      481 ghltiagckm skslknfiti kdalkkhsar qlrlaflmhs wkdtldyssn
tmesalqyek
      541 flnefflnvk dilrapvdit gqfekwgeee aelnknfydk ktaihkalcd
nvdtrtvmee
      601 mralvsqcnl ymaarkavrk rpnqalleni alylthmlki fgaveedssl
gfpvggpgts
      661 lsleatvmpy lqvlsefreg vrkiareqkv peilqlsdal rdnilpelgv
rfedheglpt
      721 vvklvdrntl lkereekrrv eeekrkkkee aarrkqeqea aklakmkipp
semflsetdk
      781 yskfdenvsm vcphmtwrak ssakgkprs
```

DDX1

Official Symbol: DDX1

Official Name: DEAD (Asp-Glu-Ala-Asp) box helicase 1

Gene ID: 1653

Organism: Homo sapiens

Other Aliases: DBP-RB, UKVH5d

Other Designations: ATP-dependent RNA helicase **DDX1**; DEAD (Asp-Glu-Ala-Asp) box polypeptide 1; DEAD box polypeptide 1; DEAD box protein 1; DEAD box protein retinoblastoma; DEAD box-1; DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1

Nucleotide sequence:

NCBI Reference Sequence: NM_004939.2

LOCUS NM_004939

ACCESSION NM_004939

```
      1 ctaatcacca aacatctgct tccttctctg tagctgtgac cctgataccg
cgtggtgtgc
       61 tccgaacaca tggtgcccag aacgaaggcg gcgtccagaa gccctaggtc
ccagaggtcc
      121 gctcagcggc aggcgcataa ggcggggccg gcgcgggcct ttccttccat
cggaaccgtt
      181 ctcccggggc tgagtccctg cccggactcc gaacgccgaa gaccaggggc
cggaagcgcg
      241 cgccgccact gccacgccgt gtcagtcggg agggagggag cgagcaggcg
aagccgcgga
      301 ggacggggtg aagatggcgg ccttctccga gatgggtgta atgcctgaga
ttgcacaagc
```

361 tgtggaagag atggattggc tcctcccaac tgatatccag gctgaatcta
tcccattgat
421 cttaggagga ggtgatgtac ttatggctgc agaaacagga agtggcaaaa
ctggtgcttt
481 tagtattcca gttatccaga tagtttatga aactctgaaa gaccaacagg
aaggcaaaaa
541 aggaaaaaca acaattaaaa ctggtgcttc agtgctgaac aaatggcaga
tgaacccata
601 tgacagagga tctgcttttg caattgggtc agatggtctt tgttgtcaaa
gcagagaagt
661 aaaggaatgg catgggtgta gagctactaa aggattaatg aaagggaaac
actactatga
721 agtatcctgt catgaccaag ggttatgcag ggtcgggtgg tctaccatgc
aggcctcttt
781 ggacctaggt actgacaagt ttggatttgg ctttggtgga acaggaaaga
aatcccataa
841 caaacaattt gataattatg gagaggaatt cactatgcat gataccattg
gatgttacct
901 ggatatagat aagggacatg tcaagttctc caaaaatgga aaagatcttg
gtctggcatt
961 tgaaatacca ccacatatga aaaaccaagc cctctttcct gcctgtgttt
tgaagaatgc
1021 tgaactgaaa tttaacttcg gtgaagagga atttaagttt ccaccaaaag
atggctttgt
1081 tgctctttcc aaggcaccgg atggttacat tgtcaaatca cagcactcag
gtaatgcaca
1141 ggtgacacaa acaaagtttc tccccaatgc tccgaaagct ctcattgttg
aaccttcccg
1201 ggagttagct gaacaaactt tgaacaacat caagcagttt aagaaataca
ttgataatcc
1261 taaattaagg gagcttctga taattggagg tgttgcagcc cgggatcagc
tctctgtttt
1321 ggaaaatgga gtagatatag ttgtaggtac tccgggaaga ctagatgact
tggtgtcaac
1381 tggaaagctg aacttatctc aagttagatt cctggtcctg gatgaagctg
atgggcttct
1441 ttctcaaggt tattctgatt ttataaatag gatgcacaat cagattcctc
aggttacctc
1501 tgatggaaaa agacttcagg tgattgtttg ctctgccact ttgcattctt
tcgatgtaaa
1561 gaaactgtcc gagaagataa tgcattttcc tacatgggtt gacttaaaag
gagaagactc
1621 tgttccagat actgtacacc atgttgttgt cccagtaaat cccaaaactg
acagactctg
1681 ggaaaggctt ggaaagagcc acattagaac tgatgatgta catgcaaaag
ataacacaag
1741 acctggtgct aatagtccag agatgtggtc tgaagctatt aaaatcctga
aaggggagta
1801 tgctgtccgg gcaatcaagg aacataagat ggatcaagca attatcttct
gtagaaccaa
1861 aattgactgt gataacttgg agcagtactt tatacaacaa ggaggaggac
ctgataaaaa
1921 aggacaccag ttctcatgtg tttgtcttca tggtgacaga aagcctcatg
agagaaagca
1981 aaacttggaa agatttaaga aaggagatgt aagattcttg atttgcacag
atgtagctgc
2041 tagaggaatt gatatccacg gtgttcctta tgttataaat gtcactctgc
ccgatgaaaa
2101 gcaaaactac gtacatcgaa ttggcagagt aggaagagct gaaaggatgg
gtctggcaat

```
     2161 ttccctggtg gcaacagaaa aagaaaaggt ttggtaccat gtatgtagca
gccgtggaaa
     2221 agggtgttat aacacaagac tcaaggaaga tggaggctgt accatatggt
acaacgagat
     2281 gcagttacta tctgagatag aagaacacct gaactgtacc atttctcagg
ttgagccgga
     2341 tataaaggta ccagtggatg aatttgatgg gaaagttacc tacggtcaga
aaagggctgc
     2401 tggtggtgga agctataaag gccatgtgga tattttggca cctactgttc
aagagttggc
     2461 tgcccttgaa aaggaggcgc agacatcttt cctgcatctt ggctaccttc
ctaaccagct
     2521 gttcagaacc ttctgatttt tacatttact gaataagatt tgagtaatga
aagtctgtag
     2581 tcttaaaact ctaaaacagt tgtactgctt ccaagcagca gtatttatag
taacgtaagc
     2641 tattaatgct aactcttgca tgtcaagaaa cattagtctt aggaattctt
caaaaaatgg
      2701 catcccaatg aaaataaatt tgatgactat attttcatga aaaaaaaaaa aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_004930.1

LOCUS NP_004930

ACCESSION NP_004930

```
      1 maafsemgvm peiaqaveem dwllptdiqa esiplilggg dvlmaaetgs
gktgafsipv
       61 iqivyetlkd qqegkkgktt iktgasvlnk wqmnpydrgs afaigsdglc
cqsrevkewh
      121 gcratkglmk gkhyyevsch dqglcrvgws tmqasldlgt dkfgfgfggt
gkkshnkqfd
      181 nygeeftmhd tigcyldidk ghvkfskngk dlglafeipp hmknqalfpa
cvlknaelkf
      241 nfgeeefkfp pkdgfvalsk apdgyivksq hsgnaqvtqt kflpnapkal
ivepsrelae
      301 qtlnnikqfk kyidnpklre lliiggvaar dqlsvlengv divvgtpgrl
ddlvstgkln
      361 lsqvrflvld eadgllsqgy sdfinrmhnq ipqvtsdgkr lqvivcsatl
hsfdvkklse
      421 kimhfptwvd lkgedsvpdt vhhvvvpvnp ktdrlwerlg kshirtddvh
akdntrpgan
      481 spemwseaik ilkgeyavra ikehkmdqai ifcrtkidcd nleqyfiqqg
ggpdkkghqf
      541 scvclhgdrk pherkqnler fkkgdvrfli ctdvaargid ihgvpyvinv
tlpdekqnyv
      601 hrigrvgrae rmglaislva tekekvwyhv cssrgkgcyn trlkedggct
iwynemqlls
      661 eieehlncti sqvepdikvp vdefdgkvty gqkraagggs ykghvdilap
tvqelaalek
      721 eaqtsflhlg ylpnqlfrtf
```

CCDC22

Official Symbol: CCDC22

Official Name: coiled-coil domain containing 22

Gene ID: 28952

Organism: Homo sapiens

Other Aliases: JM1, CXorf37

Other Designations: coiled-coil domain-containing protein 22

Nucleotide sequence:

NCBI Reference Sequence: NM_014008.3

LOCUS NM_014008

ACCESSION NM_014008

```
      1 ctcacatccg gcatgcgccg tgctcgctca cagaactaca ctttccaact
ctccccacac
       61 gacccgtgac actctgtgga ccgcgagcac ggagcagggt ttctacagct
gctccccact
      121 ttctcggacc cggtcctgga cccagccccc gactccgaca cggctccacc
atggaggagg
      181 cggaccgaat cctcatccat tcgctgcgcc aggccggcac ggcagttcct
ccagatgtgc
      241 agaccttgcg cgccttcacc actgagctgg ttgtagaggc tgtggtccgc
tgcctgcgtg
      301 tgatcaaccc tgcggtgggc tctggcctca gccctctgct gcctcttgcc
atgtctgccc
      361 ggttccgcct ggccatgagc ctggctcagg cctgcatgga cctgggctat
cccttggagc
      421 ttggctatca gaacttcctc taccccagtg agcctgacct ccgagacctg
cttctcttct
      481 tggctgagcg tctgcccacc gatgcctctg aggatgcaga ccagcctgca
ggtgactcag
      541 ctattctcct ccgggccatt gggagccaaa ttcgggacca gctggcactg
ccttgggtcc
      601 cgccccacct tcgcactccc aagctgcagc acctccaggg ctcggccctc
cagaagcctt
      661 tccatgccag caggctggtc gtgccagaat tgagttccag aggtgagcca
cgggagttcc
      721 aggcgagtcc cctgctgctt ccagtcccta cccaggtgcc tcagcctgtt
ggaagggtgg
      781 cctcgctcct cgaacaccat gccctgcagc tctgccagca gacgggccgg
gaccggccag
      841 gggatgagga ctgggtccac cggacatccc gcctcccacc ccaggaggac
acacgggctc
      901 agcggcagcg gctgcagaag caactgactg agcatctgcg ccaaagctgg
ggcctgcttg
      961 gggcccccat acaagcccgg gacctgggag aactgctgca ggcctggggt
gctggggcca
     1021 agactggtgc tcctaagggc tcccgcttca cgcactcaga gaagttcacc
ttccatctgg
     1081 agccccaggc ccaggccact caggtgtcag atgtgccagc cacctcccgg
cggcctgaac
     1141 aggtcacgtg ggcagctcag gaacaggagc tcgagtccct tcgggagcag
ctggaaggag
```

```
     1201 tgaaccgcag cattgaggag gttgaggccg acatgaagac cctgggcgtc
agctttgtgc
     1261 aggcagagtc tgagtgccgg cacagcaagc tcagtacagc agagcgtgag
caggccctgc
     1321 gcctgaagag ccgcgcggtg gagctgctgc ccgatgggac tgccaacctt
gccaagctgc
     1381 agcttgtggt ggagaatagt gcccagcggg tcatccactt ggcgggtcag
tgggagaagc
     1441 accgggtccc actcctcgct gagtaccgcc acctccgaaa gctgcaggat
tgcagagagc
     1501 tggaatcttc tcgacggctg gcagagatcc aagaactgca ccagagtgtc
cgggcggctg
     1561 ctgaagaggc ccgcaggaag gaggaggtct ataagcagct gatgtcagag
ctggagactc
     1621 tgcccagaga tgtgtcccgg ctggcctaca cccagcgcat cctggagatc
gtgggcaaca
     1681 tccggaagca gaaggaagag atcaccaaga tcttgtctga tacgaaggag
cttcagaagg
     1741 aaatcaactc cctatctggg aagctggacc ggacgtttgc ggtgactgat
gagcttgtgt
     1801 tcaaggatgc caagaaggac gatgctgttc ggaaggccta taagtatcta
gctgctctgc
     1861 acgagaactg cagccagctc atccagacca tcgaggacac aggcaccatc
atgcgggagg
     1921 ttcgagacct cgaggagcag atcgagacag agctgggcaa gaagaccctc
agcaacctgg
     1981 agaagatccg ggaggactac cgagccctcc gccaggagaa cgctggcctc
ctaggccggg
     2041 tccgggaggc ctgaggagcc gccggcagag gtctctcccc agcctcaggc
agggatttgg
     2101 ggtgctggag gcagtggcca agcacatgcc ctagctactt cctccgctgt
ccagttcctc
     2161 ctgctgcggc cttggaccca gacccctgcc cactgaccgc aacccttata
tggggtgata
     2221 gtccagcatg tggggagctc ggctgcagtt tattggggac ggtactgtgg
gttgggggcc
      2281 ttggatccca aataaatgag tagttcctct gcagtctaaa aaaaaaaaaa aaa
```

Protein sequence:

NCBI Reference Sequence: NP_054727.1

LOCUS NP_054727

ACCESSION NP_054727

```
      1 meeadrilih slrqagtavp pdvqtlraft telvveavvr clrvinpavg
sglspllpla
       61 msarfrlams laqacmdlgy plelgyqnfl ypsepdlrdl llflaerlpt
dasedadqpa
      121 gdsaillrai gsqirdqlal pwvpphlrtp klqhlqgsal qkpfhasrlv
vpelssrgep
      181 refqasplll pvptqvpqpv grvasllehh alqlcqqtgr drpgdedwvh
rtsrlppqed
      241 traqrqrlqk qltehlrqsw gllgapiqar dlgellqawg agaktgapkg
srfthsekft
      301 fhlepqaqat qvsdvpatsr rpeqvtwaaq eqeleslreq legvnrsiee
veadmktlgv
      361 sfvqaesecr hsklstaere qalrlksrav ellpdgtanl aklqlvvens
aqrvihlagq
```

421 wekhrvplla eyrhlrklqd crelessrrl aeiqelhqsv raaaeearrk
eevykqlmse
481 letlprdvsr laytqrilei vgnirkqkee itkilsdtke lqkeinslsg
kldrtfavtd
541 elvfkdakkd davrkaykyl aalhencsql iqtiedtgti mrevrdleeq
ietelgkktl
601 snlekiredy ralrqenagl lgrvrea

CLIC4

Official Symbol: CLIC4

Official Name: chloride intracellular channel 4

Gene ID: 25932

Organism: Homo sapiens

Other Aliases: CLIC4L, H1, MTCLIC, huH1, p64H1

Other Designations: chloride intracellular channel 4 like; chloride intracellular channel protein 4; intracellular chloride ion channel protein p64H1

Nucleotide sequence:

NCBI Reference Sequence: NM_013943.2

LOCUS NM_013943

ACCESSION NM_013943

1 ttattttccc cggagagtcc cgaggcgccg cgccttggcc ctgcctacag
cccgaggccc
61 cgccccggc gcccctccca gccgtttgaa gcggctcggg ctgcggctgg
ctcagagtgg
121 cgcgggggc gtggggcggt gctgaggagc tgaagccgtg gccagctcga
cgccggacag
181 tccagcgagc agcacggcgg gaaccggcag ccggagcagt cccggagcag
aagcagcagc
241 agcagcagca gccctcgccg ttcgcggagc gcagccgagc cggccatggc
gttgtcgatg
301 ccgctgaatg ggctgaagga ggaggacaaa gagcccctca tcgagctctt
cgtcaaggct
361 ggcagtgatg gtgaaagcat aggaaactgc ccctttccc agaggctctt
catgattctt
421 tggctcaaag gagttgtatt tagtgtgacg actgttgacc tgaaaaggaa
gccagcagac
481 ctgcagaact tggctcccgg gacccaccca ccatttataa ctttcaacag
tgaagtcaaa
541 acggatgtaa ataagattga ggaatttctt gaagaagtct tatgccctcc
caagtactta
601 aagctttcac caaaacaccc agaatcaaat actgctggaa tggacatctt
tgccaaattc
661 tctgcatata tcaagaattc aaggccagag gctaatgaag cactggagag
gggtctcctg
721 aaaaccctgc agaaactgga tgaatatctg aattctcctc tccctgatga
aattgatgaa

781 aatagtatgg aggacataaa gttttctaca cgtaaatttc tggatggcaa
tgaaatgaca
841 ttagctgatt gcaacctgct gcccaaactg catattgtca aggtggtggc
caaaaaatat
901 cgcaactttg atattccaaa agaaatgact ggcatctgga gatacctaac
taatgcatac
961 agtagggacg agttcaccaa tacctgtccc agtgataagg aggttgaaat
agcatatagt
1021 gatgtagcca aaagactcac caagtaaaat cgcgtttgta aaagagatgt
cttcatgtct
1081 tcccctaaga atacgctttt cctaacaggc tactccttcc tgtagagcag
aaattgtatt
1141 ttgcacgaac atgcagttat tgaagattag gatcaaggat agacaaggta
tagtagttat
1201 cttaaaatat acactcctaa gcagtattat tttaaaatcc tttaccctgg
ctacctcccc
1261 tacccgggtt cccctctctt taatttggag acactccacc acaaactttt
cactttagag
1321 gtagcttgcc atctctcagg agccctcacc attgtgtcca ttcactgtgt
atagatggca
1381 gaacttttga ggtgcaatgt ttaattgtta aaaatagtag ccacgacttt
atcaggcagc
1441 cccaaactgg tgcataatgc atggtacaag aaatatttat gtatttttg
gaattttgta
1501 atatttagta agagtatatg aaaggattgc tactgtatca gaaatattgt
ttcaatttag
1561 tctatcctgg atatgtacta acgaatatta ccaccagaga agagagcttt
ctacaaaagt
1621 cactacagat tttgctatat tgctttgtag atagattttt acttttgcct
aaaagcattt
1681 atccttcata ccaattgtaa catctgacac catgtagaag ctaaaagttt
agagggagtg
1741 agggttttct caagaccttc ctcaagcatt ttatctttag aagagaaact
gatgggcacc
1801 tgatactctg tctaaatacg tttgttatat gtgttttgcc ctgtgccatt
catttggaac
1861 tttattgcat tctttatttt aaaaagcttg tttttacgta atcatagagc
ttgctatttg
1921 tacatctgtt gagcaacact acataactga tttttagttg acttagctat
agcagtacaa
1981 tgattagtaa tgtaaaaatt aacacagaaa ttaacctaag gaatgaaggg
tgggtttgtc
2041 aaaatatcaa gtaaattttt gtttctaaag tacatttaat gtagatgacc
taaagaatgc
2101 gttatccatc ctatataaaa gaaagataaa acacaggtca ccaattttct
catttcaccc
2161 catttacctt gtatagagga ttgttcattc ctttgggact aagttatagt
tatggtgagt
2221 gtgtatttac tgtagttttg cctgatctca ctcattgcac ttcctggagt
taaattttcc
2281 aacagccatg ttgaggaata gcactctgca tgtttttgtt ttgtttttcg
gggttttttt
2341 taattgaagc cctaaaccag gaattatttg tgttctaaca ggaggatgaa
cttgctgaaa
2401 ataaaacttt gctatgtatt tactcttttt taaaagacaa aagcaaaacc
agactttcta
2461 cgtactactc caaagactgt gattgtgact ataatacatt tttggtaatt
tttttatacc
2521 taatttgtat aggaagtgct atttctcata ggctgtttct tgaaatttta
agtttattgc

2581 tttaaaatgg cagtgtttct cccactttga tatgctaaca tttagtaagc
actggcttta
2641 tgaaagcggc tttttataag tatactgcat tttttgagcc tatcattaat
tagcttagta
2701 tgaaagataa gaaaatctcc atgttgtatc catttggctc aggaagattc
tttgccttac
2761 ctttcttaga actctttatt gcttatcaaa agtttgagta cccgcttggt
tttttttgg
2821 taattaaata ttgtatgatt tatctggttc aaggaagatg cactattcag
ttatctattg
2881 agaaattatt ttgcagtggt tttagtgggt gaaaatgtcc catctgcacc
agtacacagg
2941 caggcattat cattcttcac ctacttttta aatagtggca acttgggatt
cattctggtg
3001 attctgaacc ttgcctcata gcttaaagta taaaaaagat tcaagagcag
tgaggtttgt
3061 tctttccagt gaatggtgga ctgagtggtg cgaggtggag ggctaacaag
aggaaagaac
3121 tacattcttc agaatacagt gatgaaaatt cattttgaaa ctcaaatatt
ttcattttgg
3181 atattctcct gtttttatta aaccagtgat tacacctggc catccctcta
aatgttctag
3241 gaaggcatgt ctattgtgat tttgatgaag acagaattat ttttctctgt
agaaacacag
3301 ataccacttt atcagggaag ttagtcaaat gaaatggaaa ttggtaaatg
gacaaaagct
3361 agctagtaaa aaggacgacc cagcaacatg ctttaacccc attgtatgtt
tgtggaaaga
3421 gcatagttta acatcttgag aaatttggga cataaagttt tcatggtaga
cagttcatgc
3481 agtatatgaa ttgccataat ggaaataatc tgattttatt tttacaacta
acatccattc
3541 cccttcattt aaacaccttt tgtgttttac ttcagtgagg agattggagt
ctgaatggat
3601 ctgttttcca agagattctg agaaattttt gtattcagca gttggaaagc
tctctattct
3661 agttgataaa acttcccttt tttgatgtag atgcagatat tctatacagt
tctgttgtct
3721 tttactagga ctgtaaactt ttgtgataaa attcaaataa gattttattt
cttggtaatt
3781 ttggctttca caatttatct ttaaatcctt gagcaatctg tatacaatta
agagatttct
3841 gacatttatt cttacactaa atggatcaac tctaggattt aggcatgtta
acttctgttg
3901 tgttttgaat ctctccagag ttgcatgtag atagcattta tttctgtgcc
cttaaaccca
3961 tttagaaaat aactacaaag taaaaatgta gaggaaatag aaatgtattt
tttcatgaac
4021 attttgatac aaatttcatc atttaatgat tcaccaattt cttgcattaa
tttgaattta
4081 agcatttaat tcaaagagag gggagcatcc attattgata catgtgggct
tttaaaaact
4141 ccatccttta taaatagtca aggtttgggc cacacaaagt atatttttat
catggaaaaa
4201 tttcaactcc tcaagccgta atgttgaaca gaattggagt attttcttta
taatttcttg
4261 aacaggcaaa tgaaagctta ttatagaatg catgtatttt cttttatctt
tggaacatca
4321 gcaccagtat attgctggca gctattgtat taaaaaataa agtatatttt
cactatcata

```
     4381 aaggattctt ttttcccccc tcatgaaaat aaacaacaac ttggggtaaa
agtgaaaaaa
      4441 aaaaaaaaaa aa
```

Protein sequence:

NCBI Reference Sequence: NP_039234.1

LOCUS NP_039234

ACCESSION NP_039234

```
      1 malsmplngl keedkeplie lfvkagsdge signcpfsqr lfmilwlkgv
vfsvttvdlk
       61 rkpadlqnla pgthppfitf nsevktdvnk ieefleevlc ppkylklspk
hpesntagmd
      121 ifakfsayik nsrpeaneal ergllktlqk ldeylnsplp deidensmed
ikfstrkfld
      181 gnemtladcn llpklhivkv vakkyrnfdi pkemtgiwry ltnaysrdef
tntcpsdkev
      241 eiaysdvakr ltk
```

DLD

Official Symbol: DLD

Official Name: dihydrolipoamide dehydrogenase

Gene ID: 1738

Organism: Homo sapiens

Other Aliases: tcag7.39, DLDH, E3, GCSL, LAD, PHE3

Other Designations: E3 component of pyruvate dehydrogenase complex, 2-oxo-glutarate complex, branched chain keto acid dehydrogenase complex; diaphorase; dihydrolipoyl dehydrogenase, mitochondrial; glycine cleavage system L protein; glycine cleavage system protein L; lipoamide dehydrogenase; lipoamide reductase; lipoyl dehydrogenase

Nucleotide sequence:

NCBI Reference Sequence: NM_000108.3

LOCUS NM_000108

ACCESSION NM_000108

```
      1 gatgacgtag gctgcgcctg tgcatgcgca gggaggggag accttggcgg
agcggcggag
       61 gcgcccagcg gaggtgaaag tattggcgga aaggaaaata cagcggaaaa
atgcagagct
      121 ggagtcgtgt gtactgctcc ttggccaaga gaggccattt caatcgaata
tctcatggcc
```

181 tacagggact ttctgcagtg cctctgagaa cttacgcaga tcagccgatt
gatgctgatg
241 taacagttat aggttctggt cctggaggat atgttgctgc tattaaagct
gcccagttag
301 gcttcaagac agtctgcatt gagaaaaatg aaacacttgg tggaacatgc
ttgaatgttg
361 gttgtattcc ttctaaggct ttattgaaca actctcatta ttaccatatg
gcccatggaa
421 aagattttgc atctagagga attgaaatgt ccgaagttcg cttgaattta
gacaagatga
481 tggagcagaa gagtactgca gtaaaagctt taacaggtgg aattgcccac
ttattcaaac
541 agaataaggt tgttcatgtc aatggatatg gaaagataac tggcaaaaat
caagtcactg
601 ctacgaaagc tgatggcggc actcaggtta ttgatacaaa gaacattctt
atagccacgg
661 gttcagaagt tactcctttt cctggaatca cgatagatga agatacaata
gtgtcatcta
721 caggtgcttt atctttaaaa aaagttccag aaaagatggt tgttattggt
gcaggagtaa
781 taggtgtaga attgggttca gtttggcaaa gacttggtgc agatgtgaca
gcagttgaat
841 ttttaggtca tgtaggtgga gttggaattg atatggagat atctaaaaac
tttcaacgca
901 tccttcaaaa acaggggttt aaatttaaat tgaatacaaa ggttactggt
gctaccaaga
961 agtcagatgg aaaaattgat gtttctattg aagctgcttc tggtggtaaa
gctgaagtta
1021 tcacttgtga tgtactcttg gtttgcattg gccgacgacc ctttactaag
aatttgggac
1081 tagaagagct gggaattgaa ctagatccca gaggtagaat tccagtcaat
accagatttc
1141 aaactaaaat tccaaatatc tatgccattg gtgatgtagt tgctggtcca
atgctggctc
1201 acaaagcaga ggatgaaggc attatctgtg ttgaaggaat ggctggtggt
gctgtgcaca
1261 ttgactacaa ttgtgtgcca tcagtgattt acacacaccc tgaagttgct
tgggttggca
1321 aatcagaaga gcagttgaaa gaagagggta ttgagtacaa agttgggaaa
ttcccatttg
1381 ctgctaacag cagagctaag acaaatgctg acacagatgg catggtgaag
atccttgggc
1441 agaaatcgac agacagagta ctgggagcac atattcttgg accaggtgct
ggagaaatgg
1501 taaatgaagc tgctcttgct ttggaatatg gagcatcctg tgaagatata
gctagagtct
1561 gtcatgcaca tccgacctta tcagaagctt ttagagaagc aaatcttgct
gcgtcatttg
1621 gcaaatcaat caacttttga attagaagat tatatatatt tttttctgaa
atttcctggg
1681 agcttttgta gaagtcacat tcctgaacag gatattctca cagctccaag
aatttctagg
1741 actgaattat gaaacttttg gaaggtattt aataggtttg gacaaaatgg
aatactctta
1801 tatctatatt ttacataaat ttagtatttt gtttcagtgc actaatgtgt
aagacaaaaa
1861 gctacttatt gtagcatcct ggaatatctc cgtcaactca tattttcatg
ctgttcatga
1921 aagattcaat gcccctgaat ttaaatagct tttttctctg atacagaaaa
gttgaatttt

```
     1981 acatggctgg agctagaatt tgatatgtga acagttgtgt ttgaagcaca
gtgatcaagt
     2041 tatttttaat ttggttttca cattggaaac aagtcagtca ttcagatatg
attcaaatgt
     2101 ctataaaccg aactgatgta agtaaacggt ctctcacttg ttttatttaa
cctctaaatt
     2161 ctttcatttt aggggtagca tttgtgttga agaggtttta aagcttccat
tgttgtctgc
     2221 aactctgaag ggtaattata tagttaccca aattaagaga gtctatttac
ggaactcaaa
     2281 tacgtgggca ttcaaatgta ttacagtggg gaatgaagat actgaaataa
acgtcttaaa
     2341 tattcattta ctggttatca tgagtacgtg ttgagatggt catagttttt
tttatgacta
     2401 cttctagtgt atattctaat ttcttttcta ggcctgaatg tatctttatt
ttcatgttat
     2461 aggacaatat taaggcattt taaaggtcat catcctttca tctattttag
atacacctac
     2521 taaatgttta atatatactt ttggagaagt acaacataag ggagtcttta
atctgtgttt
     2581 tccttggctg ggttaatgac tgtttattta aagagtgttg taaaattgga
tgtgtggtgt
     2641 ttaaaatggc catgtcctga ggaaacttaa gtaacaaagt actaaatgct
aagtaggctt
     2701 ttgcatattg taactaaatt taagaataat tcagattaag tagttctgaa
atttggtata
     2761 gatagcatag attgtctcat gctcatgagt gacataatga ccctggattc
tgttacatac
     2821 ttctaaagaa aattgattgt tgtcttagga ggcagttaac ttggctgaac
accaactcca
     2881 cactctgtct tgtttgtagg tggcagcagc tgaaatctct tctcagttgt
tttagcttta
     2941 gctatgctgc tggaagtctt tcccatgcaa gtgtgtagtt caggggtcaa
ccagagtttg
     3001 ggcagaagga agtctgcccc ttctgtgcct cctgtttttt gggggtttcc
cctttatgtt
     3061 ccagctgttg tggttgcccc atattctgcc ttctgatcct taaccaataa
aacttggctt
     3121 ttgtttcccc ctcaagtgag aacccgttaa aaatgagaca ttgagccagt
gctgttcact
     3181 ttttaagtgc caacttccct ctactttcca cttgtttata gttgtttcca
gtgcctttag
     3241 ttttttctaa aatatatttg ttcagagttt gcagttgcta tcagcaggag
ggttggtctg
     3301 atatctgtgt gctactttgc cattattgga agtgaactct gcatcttttt
aaaaatttga
     3361 aatcccggta tcatgtgaag tgctgtttat gtaaatctca acatatccct
tactcaggga
     3421 aaaaaaagtt tttagttagg gaatagtgaa atataattta atatggaatt
ctagctgtag
     3481 agttaaatcc atctttaagt gtttacattc agtatgagaa tgcaaattta
tctgtatggg
      3541 gaataaagtc ctaggaataa aacaagtttt aagtgttca
```

Protein sequence:

NCBI Reference Sequence: NP_000099.2

LOCUS NP_000099

ACCESSION NP_000099

```
      1 mqswsrvycs lakrghfnri shglqglsav plrtyadqpi dadvtvigsg
pggyvaaika
       61 aqlgfktvci eknetlggtc lnvgcipska llnnshyyhm ahgkdfasrg
iemsevrlnl
      121 dkmmeqksta vkaltggiah lfkqnkvvhv ngygkitgkn qvtatkadgg
tqvidtknil
      181 iatgsevtpf pgitidedti vsstgalslk kvpekmvvig agvigvelgs
vwqrlgadvt
      241 aveflghvgg vgidmeiskn fqrilqkqgf kfklntkvtg atkksdgkid
vsieaasggk
      301 aevitcdvll vcigrrpftk nlgleelgie ldprgripvn trfqtkipni
yaigdvvagp
      361 mlahkaedeg iicvegmagg avhidyncvp sviythpeva wvgkseeqlk
eegieykvgk
      421 fpfaansrak tnadtdgmvk ilgqkstdrv lgahilgpga gemvneaala
leygascedi
      481 arvchahptl seafreanla asfgksinf
```

ATAD3A

Official Symbol: ATAD3A

Official Name: ATPase family, AAA domain containing 3A

Gene ID: 55210

Organism: Homo sapiens

Other Aliases: RP5-832C2.1

Other Designations: ATPase family AAA domain-containing protein 3A

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_018188.3

LOCUS NM_018188

ACCESSION NM_018188

```
      1 gtgtgtgtgg cgcctgcgca gtggcggtga ccaccggctc gcggcgcgtg
gaggctgctc
       61 ccagccgcgc gcgagtcaga ctcgggtggg ggtcccggcg gcggtagcgg
cggcggcggt
      121 gcgagcatgt cgtggctctt cggcattaac aagggcccca agggtgaagg
cgcggggccg
      181 ccgccgcctt tgccgcccgc gcagcccggg gccgagggcg gcggggaccg
cgggttggga
      241 gaccggccgg cgcccaagga caaatggagc aacttcgacc ccaccggcct
ggagcgcgcc
      301 gccaaggcgg cgcgcgagct ggagcactcg cgttatgcca aggacgccct
gaatctggca
```

361 cagatgcagg agcagacgct gcagttggag caacagtcca agctcaaaat
gcggctggaa
421 gccctgagcc tgctgcacac actagtctgg gcatggagtc tctgccgtgc
cggagccgtg
481 cagacacagg agcggctgtc aggcagtgcc agccctgagc aagtgccagc
tggtgagtgc
541 tgtgctctgc aggagtatga ggccgccgtg gagcagctca agagcgagca
gatccgggcg
601 caggctgagg agaggaggaa gaccctgagc gaggagaccc ggcagcacca
ggccagggcc
661 cagtatcaag acaagctggc ccggcagcgc tacgaggacc aactgaagca
gcagcaactt
721 ctcaatgagg agaatttacg gaagcaggag gagtccgtgc agaagcagga
agccatgcgg
781 cgagccaccg tggagcggga gatggagctg cggcacaaga atgagatgct
gcgagtggag
841 gccgaggccc gggcgcgcgc caaggccgag cgggagaatg cagacatcat
ccgcgagcag
901 atccgcctga aggcggccga gcaccgtcag accgtcttgg agtccatcag
gacggctggc
961 accttgtttg gggaaggatt ccgtgccttt gtgacagact gggacaaagt
gacagccacg
1021 gtggctgggc tgacgctgct ggctgttggg gtctactcag ccaagaatgc
cacgcttgtc
1081 gccggccgct tcatcgaggc tcggctgggg aagccgtccc tagtgaggga
gacgtcccgc
1141 atcacggtgc ttgaggcgct gcggcacccc atccaggtca gccggcggct
cctcagtcga
1201 ccccaggacg cgctggaggg tgttgtgctc agtcccagcc tggaagcacg
ggtgcgcgac
1261 atcgccatag caacaaggaa caccaagaag aaccgcagcc tgtacaggaa
catcctgatg
1321 tacgggccac caggcaccgg gaagacgctg tttgccaaga aactcgccct
gcactcaggc
1381 atggactacg ccatcatgac aggcggggac gtggcccccа tggggcggga
aggcgtgacc
1441 gccatgcaca agctctttga ctgggccaat accagccggc gcggcctcct
gctctttgtg
1501 gatgaagcgg acgccttcct tcggaagcga gccaccgaga agataagcga
ggacctcagg
1561 gccacactga acgccttcct gtaccgcacg ggccagcaca gcaacaagtt
catgctggtc
1621 ctggccagca accaaccaga gcagttcgac tgggccatca atgaccgcat
caatgagatg
1681 gtccacttcg acctgccagg gcaggaggaa cgggagcgcc tggtgagaat
gtattttgac
1741 aagtatgttc ttaagccggc cacagaagga aagcagcgcc tgaagctggc
ccagtttgac
1801 tacgggagga agtgctcgga ggtcgctcgg ctgacggagg gcatgtcggg
ccgggagatc
1861 gctcagctgg ccgtgtcctg gcaggccacg gcgtatgcct ccgaggacgg
ggtcctgacc
1921 gaggccatga tggacacccg cgtgcaagat gctgtccagc agcaccagca
gaagatgtgc
1981 tggctgaagg cggaagggcc tgggcgtggg gacgagccct ccccatcctg
agtccacagg
2041 gagatccaca gctcacggag cctggccgcg gacccctccc acccctgcct
tgccggcccc
2101 tgcacattta ggatatgctc ctgggtgggg actgggctgt gcccagggcc
tctgtccccc

```
     2161 aggatgtctt gtggtgcggg tcggccgttc tgccccccag ggcacccccct
gttgtaggca
     2221 ctggctaggg aggggcaggc ctccttcctg cccctcgaga cactcttggg
agatgcattt
     2281 tccgtctggc tcacaggggg agggtgaggc tttgcacccc agcccctgcc
caggccactg
     2341 tgagggtggg tgctggctga gccccccgggg cagcaggagc caggcaggtg
atgtctttgt
     2401 tctcggctcc cacagcagag ccaggtgagg gggcgcctgc cagggccaga
cccaggtggg
     2461 gcagcctgaa ccctgcttcc ccctgtggcc ggcatgcccc gatctttcac
acactggtga
     2521 ccctgagaga ggagggagga gggaacctgg cgggggtgtc tgaggccgca
ctgtcagctg
     2581 gccggtccaa gcctgtggct ggagctgggg tctgtttacc taataaagtc
ccacaggtgc
      2641 ctcattaaaa aaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_060658.3

LOCUS NP_060658

ACCESSION NP_060658

```
      1 mswlfginkg pkgegagppp plppaqpgae gggdrglgdr papkdkwsnf
dptgleraak
       61 aarelehsry akdalnlaqm qeqtlqleqq sklkmrleal sllhtlvwaw
slcragavqt
      121 qerlsgsasp eqvpagecca lqeyeaaveq lkseqiraqa eerrktlsee
trqhqaraqy
      181 qdklarqrye dqlkqqqlln eenlrkqees vqkqeamrra tveremelrh
knemlrveae
      241 ararakaere nadiireqir lkaaehrqtv lesirtagtl fgegfrafvt
dwdkvtatva
      301 gltllavgvy saknatlvag rfiearlgkp slvretsrit vlealrhpiq
vsrrllsrpq
      361 dalegvvlsp slearvrdia iatrntkknr slyrnilmyg ppgtgktlfa
kklalhsgmd
      421 yaimtggdva pmgregvtam hklfdwants rrglllfvde adaflrkrat
ekisedlrat
      481 lnaflyrtgq hsnkfmlvla snqpeqfdwa indrinemvh fdlpgqeere
rlvrmyfdky
      541 vlkpategkq rlklaqfdyg rkcsevarlt egmsgreiaq lavswqatay
asedgvltea
      601 mmdtrvqdav qqhqqkmcwl kaegpgrgde psps
```

Nucleotide sequence (variant 2):

NCBI Reference Sequence: NM_001170535.1

LOCUS NM_001170535

ACCESSION NM_001170535

```
      1 gtgtgtgtgg cgcctgcgca gtggcggtga ccaccggctc gcggcgcgtg
gaggctgctc
```

```
       61 ccagccgcgc gcgagtcaga ctcgggtggg ggtcccggcg gcggtagcgg
cggcggcggt
      121 gcgagcatgt cgtggctctt cggcattaac aagggcccca agggtgaagg
cgcggggccg
      181 ccgccgcctt tgccgcccgc gcagcccggg gccgagggcg gcggggaccg
cgggttggga
      241 gaccggccgg cgcccaagga caaatggagc aacttcgacc ccaccggcct
ggagcgcgcc
      301 gccaaggcgg cgcgcgagct ggagcactcg cgttatgcca aggacgccct
gaatctggca
      361 cagatgcagg agcagacgct gcagttggag caacagtcca agctcaaaga
gtatgaggcc
      421 gccgtggagc agctcaagag cgagcagatc cgggcgcagg ctgaggagag
gaggaagacc
      481 ctgagcgagg agacccggca gcaccaggcc agggcccagt atcaagacaa
gctggcccgg
      541 cagcgctacg aggaccaact gaagcagcag caacttctca atgaggagaa
tttacggaag
      601 caggaggagt ccgtgcagaa gcaggaagcc atgcggcgag ccaccgtgga
gcgggagatg
      661 gagctgcggc acaagaatga gatgctgcga gtggaggccg aggcccgggc
gcgcgccaag
      721 gccgagcggg agaatgcaga catcatccgc gagcagatcc gcctgaaggc
ggccgagcac
      781 cgtcagaccg tcttggagtc catcaggacg gctggcacct tgtttgggga
aggattccgt
      841 gcctttgtga cagactggga caaagtgaca gccacggtgg ctgggctgac
gctgctggct
      901 gttggggtct actcagccaa gaatgccacg cttgtcgccg gccgcttcat
cgaggctcgg
      961 ctggggaagc cgtccctagt gagggagacg tcccgcatca cggtgcttga
ggcgctgcgg
     1021 caccccatcc aggtcagccg gcggctcctc agtcgacccc aggacgcgct
ggagggtgtt
     1081 gtgctcagtc ccagcctgga agcacgggtg cgcgacatcg ccatagcaac
aaggaacacc
     1141 aagaagaacc gcagcctgta caggaacatc ctgatgtacg ggccaccagg
caccgggaag
     1201 acgctgtttg ccaagaaact cgccctgcac tcaggcatgg actacgccat
catgacaggc
     1261 ggggacgtgg cccccatggg gcgggaaggc gtgaccgcca tgcacaagct
ctttgactgg
     1321 gccaatacca gccggcgcgg cctcctgctc tttgtggatg aagcggacgc
cttccttcgg
     1381 aagcgagcca ccgagaagat aagcgaggac ctcagggcca cactgaacgc
cttcctgtac
     1441 cgcacgggcc agcacagcaa caagttcatg ctggtcctgg ccagcaacca
accagagcag
     1501 ttcgactggg ccatcaatga ccgcatcaat gagatggtcc acttcgacct
gccagggcag
     1561 gaggaacggg agcgcctggt gagaatgtat tttgacaagt atgttcttaa
gccggccaca
     1621 gaaggaaagc agcgcctgaa gctggcccag tttgactacg ggaggaagtg
ctcggaggtc
     1681 gctcggctga cggagggcat gtcgggccgg gagatcgctc agctggccgt
gtcctggcag
     1741 gccacggcgt atgcctccga ggacggggtc ctgaccgagg ccatgatgga
cacccgcgtg
     1801 caagatgctg tccagcagca ccagcagaag atgtgctggc tgaaggcgga
agggcctggg
```

```
     1861 cgtggggacg agccctcccc atcctgagtc cacagggaga tccacagctc
acggagcctg
     1921 gccgcggacc cctcccaccc ctgccttgcc ggcccctgca catttaggat
atgctcctgg
     1981 gtggggactg ggctgtgccc agggcctctg tcccccagga tgtcttgtgg
tgcgggtcgg
     2041 ccgttctgcc ccccagggca ccccctgttg taggcactgg ctagggaggg
gcaggcctcc
     2101 ttcctgcccc tcgagacact cttgggagat gcattttccg tctggctcac
agggggaggg
     2161 tgaggctttg caccccagcc cctgcccagg ccactgtgag ggtgggtgct
ggctgagccc
     2221 ccggggcagc aggagccagg caggtgatgt ctttgttctc ggctcccaca
gcagagccag
     2281 gtgaggggggc gcctgccagg gccagaccca ggtggggcag cctgaaccct
gcttcccct
     2341 gtggccggca tgccccgatc tttcacacac tggtgaccct gagagaggag
ggaggaggga
     2401 acctggcggg ggtgtctgag gccgcactgt cagctggccg gtccaagcct
gtggctggag
     2461 ctggggtctg tttacctaat aaagtcccac aggtgcctca ttaaaaaaaa aa
```

Protein sequence (variant 2):

NCBI Reference Sequence: NP_001164006.1

LOCUS NP_001164006

ACCESSION NP_001164006

```
      1 mswlfginkg pkgegagppp plppaqpgae gggdrglgdr papkdkwsnf
dptgleraak
       61 aarelehsry akdalnlaqm qeqtlqleqq sklkeyeaav eqlkseqira
qaeerrktls
      121 eetrqhqara qyqdklarqr yedqlkqqql lneenlrkqe esvqkqeamr
ratveremel
      181 rhknemlrve aeararakae renadiireq irlkaaehrq tvlesirtag
tlfgegfraf
      241 vtdwdkvtat vagltllavg vysaknatlv agrfiearlg kpslvretsr
itvlealrhp
      301 iqvsrrllsr pqdalegvvl spslearvrd iaiatrntkk nrslyrnilm
ygppgtgktl
      361 fakklalhsg mdyaimtggd vapmgregvt amhklfdwan tsrrglllfv
deadaflrkr
      421 atekisedlr atlnaflyrt gqhsnkfmlv lasnqpeqfd waindrinem
vhfdlpgqee
      481 rerlvrmyfd kyvlkpateg kqrlklaqfd ygrkcsevar ltegmsgrei
aqlavswqat
      541 ayasedgvlt eammdtrvqd avqqhqqkmc wlkaegpgrg depsps
```

Nucleotide sequence (variant 3):

NCBI Reference Sequence: NM_001170536.1

LOCUS NM_001170536

ACCESSION NM_001170536

```
      1 gggagccctg gcccttgccg ctcctcgccg ctgtcggcag ccacttcccg
ggcgagactg
       61 cgcccccgga gcacccccgg ccggagccgt gtcgcgtgcc gggaggatcg
gactctttcc
      121 gtcacccgtt tgcacctctg cagctgtcag gagcgggtca ggttatgcca
aggacgccct
      181 gaatctggca cagatgcagg agcagacgct gcagttggag caacagtcca
agctcaaaga
      241 gtatgaggcc gccgtggagc agctcaagag cgagcagatc cgggcgcagg
ctgaggagag
      301 gaggaagacc ctgagcgagg agacccggca gcaccaggcc agggcccagt
atcaagacaa
      361 gctggcccgg cagcgctacg aggaccaact gaagcagcag caacttctca
atgaggagaa
      421 tttacggaag caggaggagt ccgtgcagaa gcaggaagcc atgcggcgag
ccaccgtgga
      481 gcgggagatg gagctgcggc acaagaatga gatgctgcga gtggaggccg
aggcccgggc
      541 gcgcgccaag gccgagcggg agaatgcaga catcatccgc gagcagatcc
gcctgaaggc
      601 ggccgagcac cgtcagaccg tcttggagtc catcaggacg gctggcacct
tgtttgggga
      661 aggattccgt gcctttgtga cagactggga caaagtgaca gccacggtgg
ctgggctgac
      721 gctgctggct gttggggtct actcagccaa gaatgccacg cttgtcgccg
gccgcttcat
      781 cgaggctcgg ctggggaagc cgtccctagt gagggagacg tcccgcatca
cggtgcttga
      841 ggcgctgcgg caccccatcc aggtcagccg gcggctcctc agtcgacccc
aggacgcgct
      901 ggagggtgtt gtgctcagtc ccagcctgga agcacgggtg cgcgacatcg
ccatagcaac
      961 aaggaacacc aagaagaacc gcagcctgta caggaacatc ctgatgtacg
ggccaccagg
     1021 caccgggaag acgctgtttg ccaagaaact cgccctgcac tcaggcatgg
actacgccat
     1081 catgacaggc ggggacgtgg cccccatggg gcgggaaggc gtgaccgcca
tgcacaagct
     1141 ctttgactgg gccaatacca gccggcgcgg cctcctgctc tttgtggatg
aagcggacgc
     1201 cttccttcgg aagcgagcca ccgagaagat aagcgaggac ctcagggcca
cactgaacgc
     1261 cttcctgtac cgcacgggcc agcacagcaa caagttcatg ctggtcctgg
ccagcaacca
     1321 accagagcag ttcgactggg ccatcaatga ccgcatcaat gagatggtcc
acttcgacct
     1381 gccagggcag gaggaacggg agcgcctggt gagaatgtat tttgacaagt
atgttcttaa
     1441 gccggccaca gaaggaaagc agcgcctgaa gctggcccag tttgactacg
ggaggaagtg
     1501 ctcggaggtc gctcggctga cggagggcat gtcgggccgg gagatcgctc
agctggccgt
     1561 gtcctggcag gccacggcgt atgcctccga ggacggggtc ctgaccgagg
ccatgatgga
     1621 cacccgcgtg caagatgctg tccagcagca ccagcagaag atgtgctggc
tgaaggcgga
     1681 agggcctggg cgtggggacg agccctcccc atcctgagtc cacagggaga
tccacagctc
     1741 acggagcctg gccgcggacc cctcccaccc ctgccttgcc ggcccctgca
catttaggat
```

```
     1801 atgctcctgg gtggggactg ggctgtgccc agggcctctg tcccccagga
tgtcttgtgg
     1861 tgcgggtcgg ccgttctgcc ccccagggca ccccctgttg taggcactgg
ctagggaggg
     1921 gcaggcctcc ttcctgcccc tcgagacact cttgggagat gcattttccg
tctggctcac
     1981 agggggaggg tgaggctttg caccccagcc cctgcccagg ccactgtgag
ggtgggtgct
     2041 ggctgagccc ccggggcagc aggagccagg caggtgatgt ctttgttctc
ggctcccaca
     2101 gcagagccag gtgaggggggc gcctgccagg gccagaccca ggtggggcag
cctgaaccct
     2161 gcttcccccc gtggccggca tgccccgatc tttcacacac tggtgaccct
gagagaggag
     2221 ggaggaggga acctggcggg ggtgtctgag gccgcactgt cagctggccg
gtccaagcct
     2281 gtggctggag ctggggtctg tttacctaat aaagtcccac aggtgcctca
ttaaaaaaaa
      2341 aa
```

Protein sequence (variant 3):

NCBI Reference Sequence: NP_001164007.1

LOCUS NP_001164007

ACCESSION NP_001164007

```
      1 mqeqtlqleq qsklkeyeaa veqlkseqir aqaeerrktl seetrqhqar
aqyqdklarq
       61 ryedqlkqqq llneenlrkq eesvqkqeam rratvereme lrhknemlrv
eaeararaka
      121 erenadiire qirlkaaehr qtvlesirta gtlfgegfra fvtdwdkvta
tvagltllav
      181 gvysaknatl vagrfiearl gkpslvrets ritvlealrh piqvsrrlls
rpqdalegvv
      241 lspslearvr diaiatrntk knrslyrnil mygppgtgkt lfakklalhs
gmdyaimtgg
      301 dvapmgregv tamhklfdwa ntsrrglllf vdeadaflrk ratekisedl
ratlnaflyr
      361 tgqhsnkfml vlasnqpeqf dwaindrine mvhfdlpgqe ererlvrmyf
dkyvlkpate
      421 gkqrlklaqf dygrkcseva rltegmsgre iaqlavswqa tayasedgvl
teammdtrvq
      481 davqqhqqkm cwlkaegpgr gdepsps
```

PCBP2

Official Symbol: PCBP2

Official Name: poly(rC) binding protein 2

Gene ID: 5094

Organism: Homo sapiens

Other Aliases: HNRPE2, hnRNP-E2

Other Designations: alpha-CP2; heterogeneous nuclear ribonucleoprotein E2; heterogenous nuclear ribonucleoprotein E2; hnRNP E2; poly(rC)-binding protein 2

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_005016.5

LOCUS NM_005016.

ACCESSION NM_005016

```
      1 cccagaccag cagaggcagc agccggagca gccgcagcct gcgccctctc
ccgcccgccc
       61 gccctccgcc cgcccgcccg ccctccgccg ccctccaccc gccccggggt
ctctttcccc
      121 cttcctcctc ctcctcctcc acccccccтt cctcctccgc ccgcccgcgg
ggcccccctc
      181 gccttcccgc ccgcccctat tgttccgccc ccggcctccc gcccttcccc
ttcccgcccg
      241 ctccccтttt cccctcagtc gcctcgcgcc tgcagttttt ggctttcacc
cccaaccagt
      301 gaccaaagac ttgaccactc aaagtccagc tccccagaac actgctcgac
atggacaccg
      361 gtgtgattga aggtggatta aatgtcactc tcaccatccg gctacttatg
catggaaagg
      421 aagttggcag tatcatcgga aagaaaggag aatcagttaa gaagatgcgc
gaggagagtg
      481 gtgcacgtat caacatctca gaagggaatt gtcctgagag aattatcact
ttggctggac
      541 ccactaatgc catcttcaaa gcctttgcta tgatcattga caaactggaa
gaggacataa
      601 gcagctctat gaccaatagc acagctgcca gtagaccccc ggtcaccctg
aggctggtgg
      661 tccctgctag tcagtgtggc tctctcattg gaaaaggtgg atgcaagatc
aaggaaatac
      721 gagagagtac aggggctcag gtccaggtgg caggggatat gctacccaac
tcaactgagc
      781 gggccatcac tattgctggc attccacaat ccatcattga gtgtgtcaaa
cagatctgcg
      841 tggtcatgtt ggagactctc tcccagtccc ccccgaaggg cgtgaccatc
ccgtaccggc
      901 ccaagccgtc cagctctccg gtcatctttg caggtggtca ggacaggtac
agcacaggca
      961 gcgacagtgc gagctttccc cacaccaccc cgtccatgtg cctcaaccct
gacctggagg
     1021 gaccacctct agaggcctat accattcaag gacagtatgc cattccacag
ccagatttga
     1081 ccaagctgca ccagttggca atgcaacagt ctcattttcc catgacgcat
ggcaacaccg
     1141 gattcagtgg cattgaatcc agctctccag aggtgaaagg ctattgggca
ggtttggatg
     1201 catctgctca gactacttct catgaactca ccattccaaa cgatttgatt
ggctgcataa
     1261 tcgggcgtca aggcgccaaa atcaatgaga tccgtcagat gtctggggcg
cagatcaaaa
```

1321 ttgcgaaccc agtggaagga tctactgata ggcaggttac catcactgga
tctgctgcca
1381 gcattagcct ggctcaatat ctaatcaatg tcaggctttc ctcggagacg
ggtggcatgg
1441 ggagcagcta gaacaatgca gattcatcca taatcccttt ctgctgttca
ccaccaccca
1501 tgatccatct gtgtagtttc tgaacagtca gcgattccag gttttaaata
gtttgtaaat
1561 tttcagtttc tacacacttt atcatccact cgtgattttt taattaaagc
gttttaattc
1621 ctttctctgt tcagctgttg atgctgagat ccatatttag ttttataagc
ttctccctgg
1681 ttttttttt ttggctcatg aatttttctg tttgtcatgg aaatgtaaga
gtggaatatt
1741 aatacatttc agtttagttc tgtaatgtca ggaatttttc aaaaaaatta
aaagatggac
1801 tggagctttt tctttgtgaa tagaaactgg atgccacagt gattcatgtg
ggttttattc
1861 ctcttgtctt gctgttattt ttgtaccttt tatccctcaa aggacccttc
ttgggttttg
1921 aatggaagcc tttattccgg ttaagatgtt ttcttctatt ttaccacttc
catcttttt
1981 tgtggccctc gatcctattt ttccctgact ccatgcttgg ttggccctta
taaaacttgt
2041 gcccaaaaga ttgaggatta gactttccga ggacttacct gtcctagggg
agtaggcaag
2101 cacttccact agggaggggg tgggggaaag gaatgacaca tgacatacat
ggcatacaca
2161 ttaagcagtt gatcatatgt ctgactgggt tccagtttct tgggaatgtt
ggtccccttg
2221 ttcaggcttg catattttaa actaaaaatt tcagtctatt gtttttagta
acttcattta
2281 tagtcctcca taacaagtta gaaggatgta tctgctacca tttattccta
taattttaga
2341 aagttggggc ttgacattat actcatttag tgagagtaga tgcaaaaaag
tggaggggca
2401 ggagaacttc tccagacacc tcagataaag tccggagccc aaggctttat
cttaaccatg
2461 tatggtaccc cattcattca tcaagaaaac cctcaacagc tgggcctgca
tggagtgtta
2521 tatttcaagg tttttcacag gggttacagt aggacagtcc ccaccccaat
caggcaccag
2581 gataaaagca gggacttaaa cagcaccccg gttcttcagc ctgagccatc
acatgctatc
2641 agtctcctaa cctccccctg ggccttaaga cagggcttgg gcagagaaga
taaatggtgg
2701 gacaaaaaaa tgagttacat tgccacctga gaaacctcag aggggaggac
ccagccttag
2761 cctccctcct cccaagtgca aaatgtgtaa acagagtaaa cggaacagaa
aagtgcagtc
2821 taagtggttt tctctcctgc ccctcccacc gcccctcccc ccacccccta
ttatttgggg
2881 ataaagaata taaagacaac cctggctttt ctattgcctt gttgcttgct
gaatataagg
2941 aatggggtgg ggcaggaagg ggcttgccct tagccacagc tctacggctg
tgcctcattc
3001 atttccacag ctgccagtgt ccctagagtt tatcaggtga attggtcagg
ggatcagtct
3061 ccctcgagcc tgacttacgg ctgggacagc cccatctttc tgttgattat
gtggcgcata

```
     3121 tatatatata tatgtatata tataaattt atataaatat ttctctatgt
aaaaaaaaaa
     3181 aaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_005007.2

LOCUS NP_005007

ACCESSION NP_005007

```
      1 mdtgvieggl nvtltirllm hgkevgsiig kkgesvkkmr eesgarinis
egncperiit
       61 lagptnaifk afamiidkle edisssmtns taasrppvtl rlvvpasqcg
sligkggcki
      121 keirestgaq vqvagdmlpn steraitiag ipqsiiecvk qicvvmletl
sqsppkgvti
      181 pyrpkpsssp vifaggqdry stgsdsasfp httpsmclnp dlegppleay
tiqgqyaipq
      241 pdltklhqla mqqshfpmth gntgfsgies sspevkgywa gldasaqtts
heltipndli
      301 gciigrqgak ineirqmsga qikianpveg stdrqvtitg saasislaqy
linvrlsset
      361 ggmgss
```

PDLIM7

Official Symbol: PDLIM7

Official Name: PDZ and LIM domain 7

Gene ID: 9260

Organism: Homo sapiens

Other Aliases: LMP1, LMP3

Other Designations: 1110003B01Rik; LIM domain protein; LMP; Lim mineralization protein 3; PDZ and LIM domain protein 7; protein enigma

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_005451.3

LOCUS NM_005451

ACCESSION NM_005451

```
      1 agaacactgg cggccgatcc caacgaggct ccctggagcc cgacgcagag
cagcgccctg
       61 gccgggccaa gcaggagccg gcatcatgga ttccttcaaa gtagtgctgg
aggggccagc
      121 accttggggc ttccggctgc aaggggggcaa ggacttcaat gtgcccctct
ccatttcccg
```

```
      181 gctcactcct gggggcaaag cggcgcaggc cggagtggcc gtgggtgact
gggtgctgag
      241 catcgatggc gagaatgcgg gtagcctcac acacatcgaa gctcagaaca
agatccgggc
      301 ctgcggggag cgcctcagcc tgggcctcag cagggcccag ccggttcaga
gcaaaccgca
      361 gaaggcctcc gcccccgccg cggaccctcc gcggtacacc tttgcaccca
gcgtctccct
      421 caacaagacg gcccggccct ttggggcgcc cccgcccgct gacagcgccc
cgcagcagaa
      481 tggacagccg ctccgaccgc tggtcccaga tgccagcaag cagcggctga
tggagaacac
      541 agaggactgg cggccgcggc cggggacagg ccagtcgcgt tccttccgca
tccttgccca
      601 cctcacaggc accgagttca tgcaagaccc ggatgaggag cacctgaaga
aatcaagcca
      661 ggtgcccagg acagaagccc cagccccagc ctcatctaca ccccaggagc
cctggcctgg
      721 ccctaccgcc cccagcccta ccagccgccc gccctgggct gtggaccctg
cgtttgccga
      781 gcgctatgcc ccggacaaaa cgagcacagt gctgacccgg cacagccagc
cggccacgcc
      841 cacgccgctg cagagccgca cctccattgt gcaggcagct gccggagggg
tgccaggagg
      901 gggcagcaac aacggcaaga ctcccgtgtg tcaccagtgc cacaaggtca
tccggggccg
      961 ctacctggtg gcgctgggcc acgcgtacca cccggaggag tttgtgtgta
gccagtgtgg
     1021 gaaggtcctg gaagagggtg gcttctttga ggagaagggc gccatcttct
gcccaccatg
     1081 ctatgacgtg cgctatgcac ccagctgtgc caagtgcaag aagaagatta
caggcgagat
     1141 catgcacgcc ctgaagatga cctggcacgt gcactgcttt acctgtgctg
cctgcaagac
     1201 gcccatccgg aacagggcct tctacatgga ggagggcgtg ccctattgcg
agcgagacta
     1261 tgagaagatg tttggcacga aatgccatgg ctgtgacttc aagatcgacg
ctggggaccg
     1321 cttcctggag gccctgggct tcagctggca tgacacctgc ttcgtctgtg
cgatatgtca
     1381 gatcaacctg gaaggaaaga ccttctactc caagaaggac aggcctctct
gcaagagcca
     1441 tgccttctct catgtgtgag ccccttctgc ccacagctgc cgcggtggcc
cctagcctga
     1501 ggggcctgga gtcgtggccc tgcatttctg ggtagggctg gcaatggttg
ccttaaccct
     1561 ggctcctggc ccgagcctgg ggctccctgg gccctgcccc acccacctta
tcctcccacc
     1621 ccactccctc caccaccaca gcacaccggt gctggccaca ccagcccctt
ttcacctcca
     1681 gtgccacaat aaacctgtac ccagctgtg
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_005442.2

LOCUS NP_005442

ACCESSION NP_005442

```
      1 mdsfkvvleg papwgfrlqg gkdfnvplsi srltpggkaa qagvavgdwv
lsidgenags
       61 lthieaqnki racgerlslg lsraqpvqsk pqkasapaad pprytfapsv
slnktarpfg
      121 apppadsapq qngqplrplv pdaskqrlme ntedwrprpg tgqsrsfril
ahltgtefmq
      181 dpdeehlkks sqvprteapa passtpqepw pgptapspts rppwavdpaf
aeryapdkts
      241 tvltrhsqpa tptplqsrts ivqaaaggvp gggsnngktp vchqchkvir
grylvalgha
      301 yhpeefvcsq cgkvleeggf feekgaifcp pcydvryaps cakckkkitg
eimhalkmtw
      361 hvhcftcaac ktpirnrafy meegvpycer dyekmfgtkc hgcdfkidag
drflealgfs
      421 whdtcfvcai cqinlegktf yskkdrplck shafshv
```

PDCD6

Official Symbol: PDCD6

Official Name: programmed cell death 6

Gene ID: 10016

Organism: Homo sapiens

Other Aliases: ALG-2, PEF1B

Other Designations: apoptosis-linked gene 2 protein; probable calcium-binding protein ALG-2; programmed cell death protein 6

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_013232.3

LOCUS NM_013232

ACCESSION NM_013232

```
      1 gataatgcca ggccctgccc ccggcagagg cggaagcgga gtcggcctga
gaggtctctc
       61 gtcgctgcag gcgcctcagc ccagccgcgt gccttggccc atggccgcct
actcttaccg
      121 ccccggccct ggggccggcc ctgggcctgc tgcaggcgcg gcgctgccgg
accagagctt
      181 cctgtggaac gttttccaga gggtcgataa agacaggagt ggagtgatat
cagacaccga
      241 gcttcagcaa gctctctcca acggcacgtg gactcccttt aatccagtga
ctgtcaggtc
      301 gatcatatcc atgtttgacc gtgagaacaa ggccggcgtg aacttcagcg
agttcacggg
      361 tgtgtggaag tacatcacgg actggcagaa cgtcttccgc acgtacgacc
gggacaactc
      421 cgggatgatc gataagaacg agctgaagca ggccctctca ggtttcggct
accggctctc
      481 tgaccagttc cacgacatcc tcattcgaaa gtttgacagg cagggacggg
ggcagattgc
```

```
      541 cttcgacgac ttcatccagg gctgcatcgt cctgcagagg ttgacggata
tattcagacg
      601 ttacgacacg gatcaggacg gctggattca ggtgtcgtac gaacagtacc
tgtccatggt
      661 cttcagtatc gtatgaccct ggcctctcgt gaagagcagc acaacatgga
aagagccaaa
      721 atgtcacagt tcctatctgt gagggaatgg agcacaggtg cagttagatg
ctgttcttcc
      781 tttagatttt gtcacgtggg gacccagctg tacatatgtg gataagctga
ttaatggttt
      841 tgcaactgta atagtagctg tatcgttcta atgcagacat tggatttggt
gactgtctca
      901 ttgtgccatg aggtaaatgt aatgtttcag gcattctgct tgcaaaaaaa
tctatcatgt
      961 gcttttctag atgtctctgg ttctatagtg caaatgcttt tattagccaa
taggaatttt
     1021 aaaataacat ggaacttaca caaaaggctt ttcatgtgcc ttacttttt
aaaaaggagt
     1081 ttattgtatt cattggaata tgtgacgtaa gcaataaagg gaatgttaga
cgtgtaaaaa
      1141 aaaaaaaaaa a
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_037364.1

LOCUS NP_037364

ACCESSION NP_037364

```
      1 maaysyrpgp gagpgpaaga alpdqsflwn vfqrvdkdrs gvisdtelqq
alsngtwtpf
       61 npvtvrsiis mfdrenkagv nfseftgvwk yitdwqnvfr tydrdnsgmi
dknelkqals
      121 gfgyrlsdqf hdilirkfdr qgrgqiafdd fiqgcivlqr ltdifrrydt
dqdgwiqvsy
      181 eqylsmvfsi v
```

ACTR2

Official Symbol: ACTR2

Official Name: ARP2 actin-related protein 2 homolog (yeast)

Gene ID: 10097

Organism: Homo sapiens

Other Aliases: ARP2

Other Designations: actin-like protein 2; actin-related protein 2

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001005386.2

LOCUS NM_001005386

ACCESSION NM_001005386

```
      1 gagctcaccg ctgccagtcg cgctgcctgc ccgtcccacc cttttcgtgc
aggcattcag
       61 ctaaatgacg ggcggagccc ggcggcggct tccggtcggg ggaaaaaagt
tgggccgaag
      121 gaggggccgg gaagacgcaa gaggaagaag agaaaacggc cgggcggcgg
tggctgtagg
      181 ttgtgcggct gcagcggctc ttccctgggc ggacgatgga cagccagggc
aggaaggtgg
      241 tggtgtgcga caacggcacc gggtttgtga agtgtggata tgcaggctct
aactttccag
      301 aacacatctt cccagctttg gttggaagac ctattatcag atcaaccacc
aaagtgggaa
      361 acattgaaat caagaataac aaaaagatgg atcttatggt tggtgatgag
gcaagtgaat
      421 tacgatcaat gttagaagtt aactacccta tggaaaatgg catagtacga
aattgggatg
      481 acatgaaaca cctgtgggac tacacatttg gaccagagaa acttaatata
gataccagaa
      541 attgtaaaat cttactcaca gaacctccta tgaacccaac caaaaacaga
gagaagattg
      601 tagaggtaat gtttgaaact taccagtttt ccggtgtata tgtagccatc
caggcagttc
      661 tgactttgta cgctcaaggt ttattgactg gtgtagtggt agactctgga
gatggtgtga
      721 ctcacatttg cccagtatat gaaggctttt ctctccctca tcttaccagg
agactggata
      781 ttgctgggag ggatataact agatatctta tcaagctact tctgttgcga
ggatacgcct
      841 tcaaccactc tgctgatttt gaaacggttc gcatgattaa agaaaaactg
tgttacgtgg
      901 gatataatat tgagcaagag cagaaactgg ccttagaaac cacagtatta
gttgaatctt
      961 atacactccc agatggacgt atcatcaaag ttgggggaga gagatttgaa
gcaccagaag
     1021 ctttatttca gcctcacttg atcaatgttg aaggagttgg tgttgctgaa
ttgcttttta
     1081 acacaattca ggcagctgac attgatacca gatctgaatt ctacaaacac
attgtgcttt
     1141 ctggagggtc tactatgtat cctggcctgc catcacggtt ggaacgagaa
cttaaacagc
     1201 tttacttaga acgagttttg aagggtgatg tggaaaaact ttctaaattt
aagatccgca
     1261 ttgaagaccc accccgcaga aagcacatgg tattcctggg tggtgcagtt
ctagcggata
     1321 tcatgaaaga caaagacaac ttttggatga cccgacaaga gtaccaagaa
aagggtgtcc
     1381 gtgtgctaga gaaacttggt gtgactgttc gataaactcc aaagcttgtt
cccgtcatac
     1441 ccgtaatgct ttctttttc ctttattgcc aatctttgaa ctcattcaac
tccaggacat
     1501 ggaagaggcc tctctctgcc ctttgactgg aaaggtcaag ttttattctg
gtgtcttggg
     1561 gaagctttgt taaatttttg ttaatgtggg taaatctgag tttaattcaa
ctgcttccct
     1621 acatagacta gagggctaag gattctgtct gctgctttgt ttcttctaag
taggcattta
```

1681 gatcattcct ataggcttcc tattttcact ttactgctct aatgctgcta
gtcgtagtct
1741 ttagcacact aggtggtatg cctttattag cataaaacaa aaaaaacttt
aacaggagct
1801 tttacatatt actgggatgg ggggtggttc gggatgggtg ggcagctgct
gaacccttta
1861 gggcatttcc tctgtaatgt ggcgctttca actgtactgc tgcagcttta
agtaccttaa
1921 agcttctcct gtgaacttct tagggaaatg ttaggttcag aactaaagtg
ttttgggtgg
1981 gttttgttgc gggggggagg gtaacaatgg gtggtcttct gatttttatt
tttgaggttt
2041 tgtcaactgg agtacgtaga ggaactttat ttacagtact ttgatttggc
aggttttctt
2101 ctacttgtgc tctgcctgga gctgtttcca tatgatataa aaagcaagtg
tagtattcca
2161 ttactatgtg gcttagggat ttatttgttt tttaaaatca accatgttag
ctgggattag
2221 actccctaca gtccttcaat ggaaaagtaa catttaaaaa tcctttgggt
aattcgaatt
2281 acagatttaa aagagcttaa gatctggtgt tttgttaatg cttctgttta
ttccagaagc
2341 attaaggtaa cccattgcca agtatcattc ttgcaaatta ttcttttata
taactgacca
2401 gtgcttaata aaacaagcag gtacttacaa ataattactg gcagtaggtt
ataattggtg
2461 gtttaaaaat aacattggaa tacaggactt gttgccaatt gggtaatttt
cattagttgt
2521 tttgtttgtt ttgatttgaa acctggaaat acagtaaaat ttgactgttt
aaaatgttgg
2581 ccaaaaaaat caagatttaa tttttttatt tgtactgaaa aactaatcat
aactgttaat
2641 tctcagccat ctttgaagct tgaaagaaga gtctttggta ttttgtaaac
gttagcagac
2701 tttcctgcca gtgtcagaaa atcctattta tgaatcctgt cggtattcct
tggtatctga
2761 aaaaaatacc aaatagtacc atacatgagt tatttctaag tttgaaaaat
aaaaagaaat
2821 tgcatcacac taattacaaa atacaagttc tggaaaaaat atttttcttc
attttaaaac
2881 tttttttaa ctaataatag ctttgaaaga agaggcttaa tttgggggtg
gtaactaaaa
2941 tcaaaagaaa tgattgactt gagggtctct gtttggtaag aatacatcat
tagcttaaat
3001 aagcagcaga aggttagttt taattatgta gcttctgtta atattaagtg
ttttttgtct
3061 gttttacctc aatttgaaca gataagtttg cctgcatgct ggacatgcct
cagaaccctg
3121 aatagcccgt actagatctt gggaacatgg atcttagagt cactttggaa
taagttctta
3181 tataaatacc cccagccttt tgagaacggg gcttgttaaa ggacgcgtat
gtagggcccg
3241 tacctactgg cagttgggtt cagggaaatg ggattgactt ggccttcagg
ctcctttggt
3301 cataatttta aaatatggga gtagaaaaca acaaagaatg gaatggactc
ttaaaacaat
3361 gaaagagcat ttatcgtttg tcccttgaat gtagaatttg tttttgattt
cataattctg
3421 ctggtaaatg tgacagttaa aatggtgcat tatgtatata tattataatt
tagaaatacc

```
     3481 attttataat tttactattc cagggtgaca taatgcattt aaatttggga
tttgggtgga
     3541 gtattatgtt taactggagt tgtcaagtat gagtccctca ggaaaaaaaa
aaaattctgt
     3601 tttaaaaagc aatctgattc ttagctcttg aaactattgc tacttaaatt
tccaataatt
     3661 aaaaatttaa aatttttaaa ttagaattgc caatacttct acatttgaga
agggtttttt
     3721 tagaaataca tttagtaaag tccccaagac attagtctta catttaaact
ttttttctta
     3781 aaacatggtt ttggtggtta acttttacac agttctgagt actgttaata
tctggaaagt
     3841 atcttgagat atcagtggaa agctaaacag tctaaattaa catgaaatac
ttcattttga
      3901 ttgagaaaat aaaatcagat tttttcaaag tcaaaaaaaa aaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001005386.1

LOCUS NP_001005386

ACCESSION NP_001005386

```
      1 mdsqgrkvvv cdngtgfvkc gyagsnfpeh ifpalvgrpi irsttkvgni
eiknnkkmdl
       61 mvgdeaselr smlevnypme ngivrnwddm khlwdytfgp eklnidtrnc
killteppmn
      121 ptknrekive vmfetyqfsg vyvaiqavlt lyaqglltgv vvdsgdgvth
icpvyegfsl
      181 phltrrldia grditrylik llllrgyafn hsadfetvrm ikeklcyvgy
nieqeqklal
      241 ettvlvesyt lpdgriikvg gerfeapeal fqphlinveg vgvaellfnt
iqaadidtrs
      301 efykhivlsg gstmypglps rlerelkqly lervlkgdve klskfkirie
dpprrkhmvf
      361 lggavladim kdkdnfwmtr qeyqekgvrv leklgvtvr
```

TXNDC12

Official Symbol: TXNDC12

Official Name: thioredoxin domain containing 12 (endoplasmic reticulum)

Gene ID: 51060

Organism: Homo sapiens

Other Aliases: UNQ713/PRO1376, AG1, AGR1, ERP16, ERP18, ERP19, PDIA16, TLP19, hAG-1, hTLP19

Other Designations: ER protein 18; ER protein 19; anterior gradient homolog 1; endoplasmic reticulum protein ERp19; endoplasmic reticulum resident protein 18; endoplasmic reticulum resident protein 19; endoplasmic reticulum thioredoxin superfamily member, 18 kDa; protein disulfide isomerase family A,

member 16; thioredoxin domain-containing protein 12; thioredoxin-like protein p19

Nucleotide sequence:

NCBI Reference Sequence: NM_015913.3

LOCUS NM_015913

ACCESSION NM_015913

```
      1 agtgctagtg gcggcaggtg caggtggccg cgcggcatcc tggggcttgc
agtctcccga
       61 gcgttctgtt gtgtccctgc ctacaatttt agggaaccta ataaaagggt
ggtcggtatg
      121 tttttatttg ggtgtgtact tttgttaggt cgctttttcg ctatgcatta
agtacggact
      181 ttaggactca actagtacca ggaagaaaaa gaccggacat tttcacctgt
ttgttataca
      241 gcgaagggga aaaattggga agaaatcctc aagctacaag aaaaataacc
agaagcttta
      301 cactttagcc tgcagtgact tatatcctgg tgtcctaagt ccacctaagt
cagttttgca
      361 ataagaggtc ccaagtttgg ttttcttgga gcttgcatca gttggttgca
tcatccctga
      421 gtagaagatt tgcggttgca aggaaaaata aggtacagag cttctccagc
gggaaagtgc
      481 atgtctgcac ggcacgagcc cacgcaccgc agaacaggct tgccaggtct
cctcagagac
      541 cctcgcagga acctaacaat gaaatccagt tgtccagtct tgatttgtgg
aggggtaagg
      601 agaatccgag gccagtgggc aatccgccca ctgttgggag cgactgacct
cacgaatcaa
      661 taatttgctt ttgactagga agtgcagcgg ttcttggggg gaggggctgg
actgggtggc
      721 ggacgcgagg agcaacggtt ctcccgaacc tctcccccgc ccctactatc
ttggcctaca
      781 ttttcccgct ccgtcccggg acctggacac ccagaatcca cgaaaagcaa
ctcgcgctcg
      841 agaacagctc tcgtaccctt ctacgtgatc tgcaccttta agctcactcc
atcccaaacc
      901 ggaccccgga ggcaccaccc acatccgtct aacatcactt ccttcagagt
ttgaaaaaaa
      961 aaaatctggg aagtagaggt gttgtgctga gcggcgctcg gcgaactgtg
tggaccgtct
     1021 gctgggactc cggccctgcg tccgctcagc cccgtggccc cgcgcaccta
ctgccatgga
     1081 gacgcggcct cgtctcgggg ccacctgttt gctgggcttc agtttcctgc
tcctcgtcat
     1141 ctcttctgat ggacataatg ggcttggaaa gggttttgga gatcatattc
attggaggac
     1201 actggaagat gggaagaaag aagcagctgc cagtggactg cccctgatgg
tgattattca
     1261 taaatcctgg tgtggagctt gcaaagctct aaagcccaaa tttgcagaat
ctacggaaat
     1321 ttcagaactc tcccataatt ttgttatggt aaatcttgag gatgaagagg
aacccaaaga
     1381 tgaagatttc agccctgacg ggggttatat tccacgaatc ctttttctgg
atcccagtgg
```

```
     1441 caaggtgcat cctgaaatca tcaatgagaa tggaaacccc agctacaagt
atttttatgt
     1501 cagtgccgag caagttgttc aggggatgaa ggaagctcag gaaaggctga
cgggtgatgc
     1561 cttcagaaag aaacatcttg aagatgaatt gtaacatgaa tgtgcccctt
ctttcatcag
     1621 agttagtgtt ctggaaggaa agcagcaggg aagggaatat tgaggaatca
tctagaacaa
     1681 ttaagccgac caggaaacct cattcctacc tacactggaa ggagcgctct
cactgtggaa
     1741 gagttctgct aacagaagct ggtctgcatg tttgtggatc cagcggagag
tggcagactt
     1801 tcttctcctt ttccctctca cctaaatgtc aacttgtcat tgaatgtaaa
gaatgaaacc
     1861 ttctgacaca aaacttgagc cacttggatg tttactcctc gcacttaagt
atttgagtct
     1921 tttcccattt cctcccactt tactcacctt agtggtgaaa ggagactagt
agcatctttt
     1981 ctacaacgtt aaaattgcag aagtagctta tcattaaaaa acaacaacaa
caacaataac
     2041 aataaatcct aagtgtaaat cagttattct acccccctacc aaggatatca
gcctgttttt
     2101 tccctttttt ctcctgggaa taattgtggg cttcttccca aatttctaca
gcctctttcc
     2161 tcttctcatg cttgagcttc cctgtttgca cgcatgcgtg tgcaggactg
gctgtgtgct
     2221 tggactcggc tccaggtgga agcatgcttt cccttgttac tgttggagaa
actcaaacct
     2281 tcaagcccta ggtgtagcca ttttgtcaag tcatcaactg tatttttgta
ctggcattaa
     2341 caaaaaaaga gataaaatat tgtaccatta aactttaata aaactttaaa
aggaaaaaaa
      2401 aaaaaaaaaa aa
```

Protein sequence:

NCBI Reference Sequence: NP_056997.1

LOCUS NP_056997

ACCESSION NP_056997

```
      1 metrprlgat cllgfsflll vissdghngl gkgfgdhihw rtledgkkea
aasglplmvi
       61 ihkswcgack alkpkfaest eiselshnfv mvnledeeep kdedfspdgg
yiprilfldp
      121 sgkvhpeiin engnpsykyf yvsaeqvvqg mkeaqerltg dafrkkhled el
```

ANXA7

Official Symbol: ANXA7

Official Name: annexin A7

Gene ID: 310

Organism: Homo sapiens

Other Aliases: RP11-537A6.8, ANX7, SNX, SYNEXIN

Other Designations: annexin VII; annexin-7

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001156.3

LOCUS NM_001156

ACCESSION NM_001156

```
      1 ccaccctggg cccgcccccg gctccatctt gcgggagacc gggttgggct
gtgacgctgc
       61 tgctggggtc agaatgtcat acccaggcta tcccccaaca ggctacccac
ctttccctgg
      121 atatcctcct gcaggtcagg agtcatcttt tccccccttct ggtcagtatc
cttatcctag
      181 tggctttcct ccaatgggag gaggtgccta cccacaagtg ccaagtagtg
gctacccagg
      241 agctggaggc taccctgcgc ctggaggtta tccagcccct ggaggctatc
ctggtgcccc
      301 acagccaggg ggagctccat cctatcccgg agttcctcca ggccaaggat
ttggagtccc
      361 accaggtgga gcaggctttt ctgggtatcc acagccacct tcacagtctt
atggaggtgg
      421 tccagcacag gttccactac ctggtggctt tcctggagga cagatgcctt
ctcagtatcc
      481 tggaggacaa cctacttacc ctagtcagcc tgccacagtg actcaggtca
ctcaaggaac
      541 tatccgacca gctgccaact tcgatgctat aagagatgca gaaattcttc
gtaaggcaat
      601 gaagggtttt gggacagatg agcaggcaat tgtggatgtg gtggccaacc
gttccaatga
      661 tcagaggcaa aaaattaaag cagcatttaa gacctcctat ggcaaggatt
taatcaaaga
      721 tctcaaatca gagttaagtg gaaatatgga agaactgatc ctggccctct
tcatgcctcc
      781 tacgtattac gatgcctgga gcttacggaa agcaatgcag ggagcaggaa
ctcaggaacg
      841 tgtattgatt gagattttgt gcacaagaac aaatcaggaa atccgagaaa
ttgtcagatg
      901 ttatcagtca gaatttggac gagaccttga aaaggacatt aggtcagata
catcaggaca
      961 ttttgaacgt ttacttgtgt ccatgtgcca gggaaatcgt gatgagaacc
agagtataaa
     1021 ccaccaaatg gctcaggaag atgctcagcg tctctatcaa gctggtgagg
ggagactagg
     1081 gaccgatgaa tcttgcttta acatgatcct tgccacaaga agctttcctc
agctgagagc
     1141 taccatggag gcttattcta ggatggctaa tcgagacttg ttaagcagtg
tgagccgtga
     1201 gttttccgga tatgtagaaa gtggtttgaa gaccatcttg cagtgtgccc
tgaaccgccc
     1261 tgccttcttt gctgagaggc tctactatgc tatgaaaggt gctggcacag
atgactccac
     1321 cctggtccgg attgtggtca ctcgaagtga gattgacctt gtacaaataa
aacagatgtt
     1381 cgctcagatg tatcagaaga ctctgggcac aatgattgca ggtgacacga
gtggagatta
```

```
     1441 ccgaagactt cttctggcta ttgtgggcca gtaggaggga ttttttttt
tttaatgaaa
     1501 aaaaatttct attcatagct tatccttcag agcaatgacc tgcatgcagc
aatatcaaac
     1561 atcagctaac cgaaagagct ttctgtcaag gaccgtatca gggtaatgtg
cttggtttgc
     1621 acatgttgtt attgccttaa ttctaatttt attttgttct ctacatacaa
tcaatgtaaa
     1681 gccatatcac aatgatacag taatattgca atgtttgtaa accttcattc
ttactagttt
     1741 cattctaatc aagatgtcaa attgaataaa aatcacagca atctctgatt
ctgtgtaata
     1801 atattgaata attttttaga aggttactga aagctctgcc ttccggaatc
cctctaagtc
     1861 tgcttgatag agtggatagt gtgttaaaac tgtgtacttt aaaaaaaaat
tcaaccttta
     1921 catctagaat aatttgcatc tcattttgcc taaattggtt ctgtattcat
aaacactttc
     1981 cacatagaaa atagattagt attacctgtg gcacctttta agaaagggtc
aaatgtttat
     2041 atgcttaaga tacatagcct actttttttt cgcagttgtt ttctttttt
aaattgagtt
      2101 atgacaaata aaaaattgca tatatttaag gtgtacaaaa aaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001147.1

LOCUS NP_001147

ACCESSION NP_001147

```
      1 msypgypptg yppfpgyppa gqessfppsg qypypsgfpp mgggaypqvp
ssgypgaggy
       61 papggypapg gypgapqpgg apsypgvppg qgfgvppgga gfsgypqpps
qsygggpaqv
      121 plpggfpggq mpsqypgggp typsqpatvt qvtqgtirpa anfdairdae
ilrkamkgfg
      181 tdeqaivdvv anrsndqrqk ikaafktsyg kdlikdlkse lsgnmeelil
alfmpptyyd
      241 awslrkamqg agtqervlie ilctrtnqei reivrcyqse fgrdlekdir
sdtsghferl
      301 lvsmcqgnrd enqsinhqma qedaqrlyqa gegrlgtdes cfnmilatrs
fpqlratmea
      361 ysrmanrdll ssvsrefsgy vesglktilq calnrpaffa erlyyamkga
gtddstlvri
      421 vvtrseidlv qikqmfaqmy qktlgtmiag dtsgdyrrll laivgq
```

PFKM

Official Symbol: PFKM

Official Name: phosphofructokinase, muscle

Gene ID: 5213

Organism: Homo sapiens

Other Aliases: GSD7, PFK-1, PFK1, PFKA, PFKX

Other Designations: 6-phosphofructo-1-kinase; 6-phosphofructokinase, muscle type; PFK-A; phosphofructo-1-kinase isozyme A; phosphofructokinase 1; phosphofructokinase, polypeptide X; phosphofructokinase-M; phosphohexokinase

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001166686.1

LOCUS NM_001166686

ACCESSION NM_001166686

```
      1 gtcccagggg gcggggcaga ggaaaaggcg ccggccccac agtgctcccc
gcttccgccc
       61 agtccagccc gggccggctg accgggtccg acacagtctc ctggaccagg
ctccctccat
      121 cctcaccccc cccccagctt cccgccgcca ctcaccgaac cggaaccggc
tgccatgcga
      181 aggggtttcc ggccgggcgc ggaacgcaaa acccgggaac cgccgcgaac
cggaaccgcc
      241 ttcacagcac cggaagagtc gctaggaggc agccatgcat aaagacgagt
ttcatctgaa
      301 atttttcatg tgtgtgattc agtctcgcca gttagtcagg actcctcaga
gaacagctgg
      361 ggaagcttct acttccagca tgctcatacc aaagccacca ccaaagacag
acatcttgaa
      421 gagtctagat actatggatg atccagacac cgtgggaagc atacctgttt
tcaaaactga
      481 gtggatcatg acccatgaag agcaccatgc agccaaaacc ctggggattg
gcaaagccat
      541 tgctgtctta acctctggtg gagatgccca aggtatgaat gctgctgtca
gggctgtggt
      601 tcgagttggt atcttcaccg gtgcccgtgt cttctttgtc catgagggtt
atcaaggcct
      661 ggtggatggt ggagatcaca tcaaggaagc cacctgggag agcgtttcga
tgatgcttca
      721 gctgggaggc acggtgattg gaagtgcccg gtgcaaggac tttcgggaac
gagaaggacg
      781 actccgagct gcctacaacc tggtgaagcg tgggatcacc aatctctgtg
tcattggggg
      841 tgatggcagc ctcactgggg ctgacacctt ccgttctgag tggagtgact
tgttgagtga
      901 cctccagaaa gcaggtaaga tcacagatga ggaggctacg aagtccagct
acctgaacat
      961 tgtgggcctg gttgggtcaa ttgacaatga cttctgtggc accgatatga
ccattggcac
     1021 tgactctgcc ctgcatcgga tcatggaaat tgtagatgcc atcactacca
ctgcccagag
     1081 ccaccagagg acatttgtgt tagaagtaat gggccgccac tgtggatacc
tggcccttgt
     1141 cacctctctg tcctgtgggg ccgactgggt ttttattcct gaatgtccac
cagatgacga
     1201 ctgggaggaa caccttttgtc gccgactcag cgagacaagg acccgtggtt
ctcgtctcaa
```

1261 catcatcatt gtggctgagg gtgcaattga caagaatgga aaaccaatca
cctcagaaga
1321 catcaagaat ctggtggtta agcgtctggg atatgacacc cgggttactg
tcttggggca
1381 tgtgcagagg ggtgggacgc catcagcctt tgacagaatt ctgggcagca
ggatgggtgt
1441 ggaagcagtg atggcacttt tggaggggac cccagatacc ccagcctgtg
tagtgagcct
1501 ctctggtaac caggctgtgc gcctgcccct catggaatgt gtccaggtga
ccaaagatgt
1561 gaccaaggcc atggatgaga agaaatttga cgaagccctg aagctgagag
gccggagctt
1621 catgaacaac tgggaggtgt acaagcttct agctcatgtc agacccccgg
tatctaagag
1681 tggttcgcac acagtggctg tgatgaacgt gggggctccg gctgcaggca
tgaatgctgc
1741 tgttcgctcc actgtgagga ttggccttat ccagggcaac cgagtgctcg
ttgtccatga
1801 tggtttcgag ggcctggcca aggggcagat agaggaagct ggctggagct
atgttggggg
1861 ctggactggc caaggtggct ctaaacttgg gactaaaagg actctaccca
agaagagctt
1921 tgaacagatc agtgccaata taactaagtt taacattcag ggccttgtca
tcattggggg
1981 ctttgaggct tacacagggg gcctggaact gatggagggc aggaagcagt
ttgatgagct
2041 ctgcatccca tttgtggtca ttcctgctac agtctccaac aatgtccctg
gctcagactt
2101 cagcgttggg gctgacacag cactcaatac tatctgcaca acctgtgacc
gcatcaagca
2161 gtcagcagct ggcaccaagc gtcgggtgtt tatcattgag actatgggtg
gctactgtgg
2221 ctacctggct accatggctg gactggcagc tggggccgat gctgcctaca
tttttgagga
2281 gcccttcacc attcgagacc tgcaggcaaa tgttgaacat ctggtgcaaa
agatgaaaac
2341 aactgtgaaa aggggcttgg tgttaaggaa tgaaaagtgc aatgagaact
ataccactga
2401 cttcattttc aacctgtact ctgaggaggg gaagggcatc ttcgacagca
ggaagaatgt
2461 gcttggtcac atgcagcagg gtgggagccc aaccccattt gataggaatt
ttgccactaa
2521 gatgggcgcc aaggctatga actggatgtc tgggaaaatc aaagagagtt
accgtaatgg
2581 gcggatcttt gccaatactc cagattcggg ctgtgttctg gggatgcgta
agagggctct
2641 ggtcttccaa ccagtggctg agctgaagga ccagacagat tttgagcatc
gaatccccaa
2701 ggaacagtgg tggctgaaac tgaggcccat cctcaaaatc ctagccaagt
acgagattga
2761 cttggacact tcagaccatg cccacctgga gcacatcacc cggaagcggt
ccggggaagc
2821 tgccgtctaa acctctctgg agtgagggga atagattacc tgatcatggt
cagctcacac
2881 cctaataagt ccacatcttc tcagtgtttt agctgttttt ttcattaggt
ttccttttat
2941 tctgtacctt gcagccatga ccagttctgg ccaggagctg gaggagcagg
cagtgggtgg
3001 gagctccttt taggtagaat ttaacatgac ttctgcccca gctttatctg
tcacacaagg

```
     3061 ctgggcacct ctagtgctac tgctagatat cacttactca gttagaattt
tcctaaaaat
     3121 aagctttatt tatttctttg tgataacaaa gagtcttggt tcctctacta
cttttactac
     3181 agtgacaaat tgtaactaca ctaataaatg ccaactggtc actgtgcttt
tgcttctcct
     3241 gttatcatct tcctaagtgg aatgtaatac tgtcagcccc atgtatcaga
cacttgtctg
     3301 atgaagcagt aaagacgtta agggtatcac agggggtgga ggaagggatt
atctctagta
     3361 cactacttgc tggctgtctg aaaaattgtc actgccaaac tctaaaaaca
gttctaaata
     3421 gtgactgaga aggtttgttg ctggagtcag ggaataaggc agccaaatac
tctttgcaca
      3481 gttctttagt gggaagagaa attaacaata aatatcaagc actgtgaaaa
aaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001160158.1

LOCUS NP_001160158

ACCESSION NP_001160158

```
      1 mhkdefhlkf fmcviqsrql vrtpqrtage astssmlipk pppktdilks
ldtmddpdtv
       61 gsipvfktew imtheehhaa ktlgigkaia vltsggdaqg mnaavravvr
vgiftgarvf
      121 fvhegyqglv dggdhikeat wesvsmmlql ggtvigsarc kdfreregrl
raaynlvkrg
      181 itnlcviggd gsltgadtfr sewsdllsdl qkagkitdee atkssylniv
glvgsidndf
      241 cgtdmtigtd salhrimeiv daitttaqsh qrtfvlevmg rhcgylalvt
slscgadwvf
      301 ipecppdddw eehlcrrlse trtrgsrlni iivaegaidk ngkpitsedi
knlvvkrlgy
      361 dtrvtvlghv qrggtpsafd rilgsrmgve avmallegtp dtpacvvsls
gnqavrlplm
      421 ecvqvtkdvt kamdekkfde alklrgrsfm nnwevyklla hvrppvsksg
shtvavmnvg
      481 apaagmnaav rstvrigliq gnrvlvvhdg feglakgqie eagwsyvggw
tgqggsklgt
      541 krtlpkksfe qisanitkfn iqglviiggf eaytgglelm egrkqfdelc
ipfvvipatv
      601 snnvpgsdfs vgadtalnti cttcdrikqs aagtkrrvfi ietmggycgy
latmaglaag
      661 adaayifeep ftirdlqanv ehlvqkmktt vkrglvlrne kcnenyttdf
ifnlyseegk
      721 gifdsrknvl ghmqqggspt pfdrnfatkm gakamnwmsg kikesyrngr
ifantpdsgc
      781 vlgmrkralv fqpvaelkdq tdfehripke qwwlklrpil kilakyeidl
dtsdhahleh
      841 itrkrsgeaa v
```

SUB1

Official Symbol: SUB1

Official Name: SUB1 homolog (S. cerevisiae)

Gene ID: 10923

Organism: Homo sapiens

Other Aliases: P15, PC4, p14

Other Designations: activated RNA polymerase II transcription cofactor 4; activated RNA polymerase II transcriptional coactivator p15; positive cofactor 4

Nucleotide sequence:

NCBI Reference Sequence: NM_006713.3

LOCUS NM_006713

ACCESSION NM_006713

```
      1 gccccatcac gtgaccgcag ccccagcgcg gcggggccgg cgtctcctgg
ctgccgtcac
       61 ttccggttct ctgtcagtcg cgagcgaacg accaagaggg tgttcgactg
ctagagccga
      121 gcgaagcgat gcctaaatca aaggaacttg tttcttcaag ctcttctggc
agtgattctg
      181 acagtgaggt tgacaaaaag ttaaagagga aaaagcaagt tgctccagaa
aaacctgtaa
      241 agaaacaaaa gacaggtgag acttcgagag ccctgtcatc ttctaaacag
agcagcagca
      301 gcagagatga taacatgttt cagattggga aaatgaggta cgttagtgtt
cgcgatttta
      361 aaggcaaagt gctaattgat attagagaat attggatgga tcctgaaggt
gaaatgaaac
      421 caggaagaaa aggtatttct ttaaatccag aacaatggag ccagctgaag
gaacagattt
      481 ctgacattga tgatgcagta agaaaactgt aaaattcgag ccatataaat
aaaacctgta
      541 ctgttctagt tgttttaatc tgtcttttta cattggcttt tgttttctaa
atgttctcca
      601 agctattgta tgtttggatt gcagaagaat ttgtaagatg aatacttttt
tttaatgtgc
      661 attattaaaa atattgagtg aagctaattg tcaactttat taaggattac
tttgtctgcc
      721 caccacctag tgtaaaataa aatcaagtaa tacaatctta actgttgtgg
ccttttttga
      781 tcataagagt tggtactgtt taaggccaaa agtaacagtt tttatagatc
ttttagtttc
      841 aactcagctt ttacaataaa aaggatttgt attgcattga gtttataaac
ttttggtttg
      901 tgaacttcat atttgatctt ttctcttcca atcaaatgtc taggcttgtt
tgacttccac
      961 ccccaatggt ttttcactct ttttatttac ttcattttcc tttaataact
taatctcttc
     1021 atgttcagtt tttacttcac tctttattct tttctttgat tatggtatgc
ttatttggaa
     1081 agtcagtgaa actgtcaaaa tgttatctca ataagatact tatatgagaa
ctacaatcac
     1141 cgaatctact gtattcaata ttagcagatc taatttgata aacaacatgg
cttgtgtgaa
```

1201 aactgagcag gtgtttgttt acccatagtg ttctgtgtag ttattgctta gtctgcagaa
1261 aataatgact tagatgagat gtctgacttg ctttcactta ttaaacatgt tcaccatggg
1321 atgatgtctg taacatcaga tattgttcaa ctagactagg atttaataaa aattgtgaaa
1381 gcttactggc ctaacatttt attttataat attgggtatg aattatatgt agccagagat
1441 gtcattaagc tttactgtta tagtaggtaa tatggttagt ttgtagggaa aagagcatat
1501 gagcacatgc ttgtgtattt tggcctttgc cccagtagaa cagaccaatg gcattctaga
1561 cttgatgata ctaagtttta gcagacacta gtaagtggtt tgtatttaac catactgatg
1621 aagcagacag attgaggcac agattttagt ggctttgtgg caataaatag ggcatggtgt
1681 gccttaggaa aagaatgttt ataaagggaa ttataactga aattaaagga ggcggcagtg
1741 aagaggaaat aattctcttc tatctaaatg atatacatat gatattttga gatttttata
1801 acagcagtgg aacacaattc taggtagagt agaaaaagga aagttttaaa gacatataaa
1861 agattcttgt tgacaaatta tttttggtag caaatctcaa atggttacct gctattaagg
1921 tctgccatat tagagttttg cactattttg ctaccaagtt tgattcatac atctaaaaca
1981 ttttgtagtt acttgtcaag gacttaattt gaaaatcatt tgccaggcca catagttatc
2041 aatttttttt tctatcagct attctgttgt atttctaaaa catttttag atgacttttt
2101 aaagtatatt tagcagtaac cttatgaggt tcaaattggt aaatctcttg taatttagcc
2161 ttcatcgaat aataggtacc agtgtattaa aaatgtgtat tttttgcagc cccttgaacc
2221 agagtaggtt cagagaaact cccaaagttt gtactttaga cacatcatgc ttgattggta
2281 acttccctcc ttttttgggg aacatgtttg tgtcctatta acttaattgg atagattttt
2341 aaatatttct tatttttggc acacggaaag ggtagttcga gtacagaact ttgatttttg
2401 gtgtagatgc agagggaatg atgggtaaat ttcctaggtt tatgtgaatt taggggggtgt
2461 atgcattttg aaacaatcta ctaacagatg gtgctgaaat ctattaccta catgttttct
2521 agttgttcag cattatgtta atgaagcctc catataagga gtgtttctct ggcacagttg
2581 gtaagttgac tgctaacttc atttaaatgt gttactggat atgcagtata ctgaaattat
2641 taatcagttt gtgtatagga aaagagaact gggttaaaag caaattaact tgttctgaaa
2701 agaaagtata gattaatttt gttttctgtt taaattttat ctccttggta aagatttttt
2761 tttcctgggc agaaaacttg gcatttttag gcgtagatac cttaccttac aatgccaaaa
2821 tgaatttaat tccagtactc aggtttttcc ctttaacaga ctctatgtgt atcagggctt
2881 tctaatgggt ttttcctctt cgtttttaaa atgtgagtag catttgacca atttccagtg
2941 ctcttagcat tttacttaaa gaacaaccac tacaaaagaa aatctttgta atttgattgt

```
     3001 cttttgcttt gcttcattaa tgcctaagaa cttaagaata ctcctacctc
attagctact
     3061 caagatgctg tgacgatcaa atctattcta cataatgcgt ttagaaacaa
agacttgggt
     3121 gaaaaatgaa ataagtatat tctgacttgg ctattgaggg gaaaattcag
tattaagtgt
     3181 tcctcacagg agatatgtta gcagaatact ataaaagttt gaaattttta
aaaagtaaaa
     3241 gtacttaaat ttaggtatct ctcctgaaat tctttgcagt tcatttttta
tggcagttaa
     3301 tccagtgaaa cactcaaaag ttttttttt tttaaaagtg tttttccaga
taaactgtag
     3361 ggtgaacatt cacataatca caaatatgta attctgtaat tgtggaatgc
ttgtatgctt
     3421 tgttttcgta catcttccat ggagatgtct gaatataata ctccatctgt
gaatatttta
      3481 aatgttgaaa taaaagtaag aaatgtgaaa aaaaaaaaaa aa
```

Protein sequence:

NCBI Reference Sequence: NP_006704.3

LOCUS NP_006704.3

ACCESSION NP_006704.3

```
      1 mpkskelvss sssgsdsdse vdkklkrkkq vapekpvkkq ktgetsrals
sskqssssrd
       61 dnmfqigkmr yvsvrdfkgk vlidireywm dpegemkpgr kgislnpeqw
sqlkeqisdi
      121 ddavrkl
```

ACDB3

Official Symbol: ACBD3

Official Name: acyl-CoA binding domain containing 3

Gene ID: 64746

Organism: Homo sapiens

Other Aliases: GCP60, GOCAP1, GOLPH1, PAP7

Other Designations: Golgi resident protein GCP60; PBR- and PKA-associated protein 7; PKA (RIalpha)-associated protein; acyl-Coenzyme A binding domain containing 3; golgi complex associated protein 1, 60kDa; golgi phosphoprotein 1; peripheral benzodiazepine receptor-associated protein PAP7

Nucleotide sequence:

NCBI Reference Sequence: NM_022735.3

LOCUS NM_022735

ACCESSION NM_022735

```
      1 atacgtggct gccgtctgtc cccgctgagg aggtgcagca gccggagatg
gcggcggtgc
       61 tgaacgcaga gcgactcgag gtgtccgtcg acggcctcac gctcagcccg
gacccggagg
      121 agcggcctgg ggcggagggc gccccgctgc tgccgccacc gctgccaccg
ccctcgccac
      181 ctggatccgg tcgcggcccg ggcgcctcag gggagcagcc cgagcccggg
gaggcggcgg
      241 ctgggggcgc ggcggaggag gcgcggcggc tggagcagcg ctggggtttc
ggcctggagg
      301 agttgtacgg cctggcactg cgcttcttca aagaaaaaga tggcaaagca
tttcatccaa
      361 cttatgaaga aaaattgaag cttgtggcac tgcataagca agttcttatg
ggcccatata
      421 atccagacac ttgtcctgag gttggattct ttgatgtgtt ggggaatgac
aggaggagag
      481 aatgggcagc cctgggaaac atgtctaaag aggatgccat ggtggagttt
gtcaagctct
      541 taaataggtg ttgccatctc ttttcaacat atgttgcgtc ccacaaaata
gagaaggaag
      601 agcaagaaaa aaaaaggaag gaggaagagg agcgaaggcg gcgtgaagag
gaagaaagag
      661 aacgtctgca aaaggaggaa gagaaacgta ggagagaaga agaggaaagg
cttcgacggg
      721 aggaagagga aaggagacgg atagaagaag aaaggcttcg gttggagcag
caaaagcagc
      781 agataatggc agctttaaac tcccagactg ccgtgcagtt ccagcagtat
gcagcccaac
      841 agtatccagg gaactacgaa cagcagcaaa ttctcatccg ccagttgcag
gagcaacact
      901 atcagcagta catgcagcag ttgtatcaag tccagcttgc acagcaacag
gcagcattac
      961 agaaacaaca ggaagtagta gtggctgggt cttccttgcc tacatcatca
aaagtgaatg
     1021 caactgtacc aagtaatatg atgtcagtta atggacaggc caaaacacac
actgacagct
     1081 ccgaaaaaga actggaacca gaagctgcag aagaagccct ggagaatgga
ccaaaagaat
     1141 ctcttccagt aatagcagct ccatccatgt ggacacgacc tcagatcaaa
gacttcaaag
     1201 agaagattca gcaggatgca gattccgtga ttacagtggg ccgaggagaa
gtggtcactg
     1261 ttcgagtacc cacccatgaa gaaggatcat atctcttttg ggaatttgcc
acagacaatt
     1321 atgacattgg gtttggggtg tattttgaat ggacagactc tccaaacact
gctgtcagcg
     1381 tgcatgtcag tgagtccagc gatgacgacg aggaggaaga agaaaacatc
ggttgtgaag
     1441 agaaagccaa aaagaatgcc aacaagcctt tgctggatga gattgtgcct
gtgtaccgac
     1501 gggactgtca tgaggaggtg tatgctggca gccatcaata tccagggaga
ggagtctatc
     1561 tcctcaagtt tgacaactcc tactctttgt ggcggtcaaa atcagtctac
tacagagtct
     1621 attatactag ataaaaatgt tgttacaaag tctggagtct agggttgggc
agaagatgac
     1681 atttaatttg gaaatttctt tttacttttg tggagcatta gagtcacagt
ttaccttatt
     1741 gatattggtc tgatggtttg tgaactcttg ctgggaatca aaatttcctt
gagactcttt
```

1801 agcattcata ctttggggtt aaaggagatt cctcagactc atccagccct
tgggtgctga
1861 ccagcagagt cactagtgga tgctgaagtt acatgagcta catgttaaat
atttaaagtc
1921 tccaaaataa aacaccccaa cgttgacctt acccggctga tggttagccc
cttgctgcct
1981 gctccatgtg tcttatgaga gcccgtagtt acagtgtcct ctaatttgaa
atccataagt
2041 taacaagtct atatcaggtg cagctggctt tgattaaagg ccatttttaa
aacttaaaaa
2101 ctcaacacct cacagattat aatagaaaaa gaaatggcct cagtttgatc
tcgttcagaa
2161 tgacccagat tgtttctgct ttgggtgcag ctgtttagtt cagagttata
ttacagagaa
2221 ttattttctg agataatctt aaactagaat gttcaaaact aattgataat
tgaagtatca
2281 agatacgtag aacacctcag agatttttct tcaggaactt ccacaaactt
tgaatccttg
2341 tatctttatt tggtattcat actactagta gcaaaataca ggtttttttgt
tttgttttgt
2401 tttgttttgg cttcatagag tatctcaaat tgaaactttt ctgcacaaag
aataaaatta
2461 aggattttat aaactcaaat tggcacctac tgaattaaaa tacataaaat
catttaaata
2521 taattcagca tatgggaagt aacattgcac taatatggaa atcactgcca
gagacagtct
2581 attttctttt aatttgttac tacttagtca caaaccccac attattccag
tttggaatta
2641 cttattaagg agaattggaa atacatatgc ccatgcttaa attttatagc
tttaatttgt
2701 gttatttctt tattgacggg aagaggtaca tctttttttc cttactgaaa
acaaatatgg
2761 attaattgcc tcaaatttgt ataagtgatt ggctagtgat tcttgttttc
agaagggaga
2821 gtggtataga tagaaaatga caaagatggc aatatacact taatgttgtt
attgtatgtt
2881 gttactgaag tacttagatt tttaaaattt caaatcctaa atcacttctt
gtaggagggt
2941 tttcattaac tgcagtatat acagttcact acatatgggt tgtttgagtt
ttttgtgtgc
3001 tgtatttctt tctgtttttt aatacctggt tttgtacata tctaactctg
ttctcttttg
3061 gttgttcaga aactggattt tttttttctt aagcagtgct taatttgtgt
tttttaattt
3121 tgattcagaa gtagtcccag ctcataggtg ttcatactgt tacatccaga
acatttgtca
3181 ggctctctgt cagctttcat gtacatatgg tatagaaacc atggagttag
gcacttcctg
3241 gatttttttt ttatgagaaa aatactgtat ttaaaatgta aaataaactt
ttaaaaagca
3301 ggcactaata tatatttctt ccagcctttg attacaaatt tgtccttgca
catgttaaga
3361 tgaattatct cctaaaaata tcattgttct tgggagcagt gtatgttact
ttacatagca
3421 gcggttcctg tcatgtgttc atgtcagaat atttttggtt ttaaactttc
ttattgcctt
3481 tggctgttga ttagtacagt acaagtgcga tttcaaaaag atcttgaaag
taatatattt
3541 aatcaattaa aatgtttatc tgtaaaaaaa aaaaaaaaaa a

Protein sequence:

NCBI Reference Sequence: NP_073572.2

LOCUS NP_073572

ACCESSION NP_073572

```
      1 maavlnaerl evsvdgltls pdpeerpgae gapllppplp ppsppgsgrg
pgasgeqpep
       61 geaaaggaae earrleqrwg fgleelygla lrffkekdgk afhptyeekl
klvalhkqvl
      121 mgpynpdtcp evgffdvlgn drrrewaalg nmskedamve fvkllnrcch
lfstyvashk
      181 iekeeqekkr keeeerrrre eeererlqke eekrrreeee rlrreeeerr
rieeerlrle
      241 qqkqqimaal nsqtavqfqq yaaqqypgny eqqqilirql qeqhyqqymq
qlyqvqlaqq
      301 qaalqkqqev vvagsslpts skvnatvpsn mmsvngqakt htdssekele
peaaeealen
      361 gpkeslpvia apsmwtrpqi kdfkekiqqd adsvitvgrg evvtvrvpth
eegsylfwef
      421 atdnydigfg vyfewtdspn tavsvhvses sdddeeeeen igceekakkn
ankplldeiv
      481 pvyrrdchee vyagshqypg rgvyllkfdn syslwrsksv yyrvyytr
```

ASNA1

Official Symbol: ASNA1

Official Name: arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)

Gene ID: 439

Organism: Homo sapiens

Other Aliases: ARSA-I, ARSA1, ASNA-I, GET3, TRC40, hASNA-I

Other Designations: ATPase **ASNA1**; arsenical pump-driving ATPase; arsenite-stimulated ATPase; golgi to ER traffic 3 homolog; transmembrane domain recognition complex 40 kDa ATPase subunit; transmembrane domain recognition complex, 40kDa

Nucleotide sequence:

NCBI Reference Sequence: NM_004317.2

LOCUS NM_004317

ACCESSION NM_004317

```
      1 gagccagttc caaaatggcg gcaggggtgg ccgggtgggg ggttgaggca
gaggagttcg
       61 aagatgctcc tgatgtggag ccgctggagc ctacacttag caacatcatc
gagcagcgca
      121 gcctgaagtg gatcttcgtc gggggcaagg gtggtgtggg caagaccacc
tgcagctgca
      181 gcctggcagt ccagctctcc aaggggcgtg agagtgttct gatcatctcc
acagacccag
```

```
      241 cacacaacat ctcagatgct tttgaccaga agttctcaaa ggtgcctacc
aaggtcaaag
      301 gctatgacaa cctctttgct atggagattg accccagcct gggcgtggcg
gagctgcctg
      361 acgagttctt cgaggaggac aacatgctga gcatgggcaa gaagatgatg
caggaggcca
      421 tgagcgcatt tcccggcatc gatgaggcca tgagctatgc cgaggtcatg
aggctggtga
      481 agggcatgaa cttctcggtg gtggtatttg acacggcacc cacgggccac
accctgaggc
      541 tgctcaactt ccccaccatc gtggagcggg gcctgggccg gcttatgcag
atcaagaacc
      601 agatcagccc tttcatctca cagatgtgca acatgctggg cctgggggac
atgaacgcag
      661 accagctggc ctccaagctg gaggagacgc tgcccgtcat ccgctcagtc
agcgaacagt
      721 tcaaggaccc tgagcagaca actttcatct gcgtatgcat tgctgagttc
ctgtccctgt
      781 atgagacaga gaggctgatc caggagctgg ccaagtgcaa gattgacaca
cacaatataa
      841 ttgtcaacca gctcgtcttc cccgaccccg agaagccctg caagatgtgt
gaggcccgtc
      901 acaagatcca ggccaagtat ctggaccaga tggaggacct gtatgaagac
ttccacatcg
      961 tgaagctgcc gctgttaccc catgaggtgc ggggggcaga caaggtcaac
accttctcgg
     1021 ccctcctcct ggagccctac aagcccccca gtgcccagta gcacagctgc
cagccccaac
     1081 cgctgccatt tcacactcac cctccaccct ccccaccccc tcggggcaga
gtttgcacaa
     1141 agtccccccc ataatacagg gggagccact tgggcaggag gcagggaggg
gtccattccc
     1201 cctggtgggg ctggtgggga gctgtagttg ccccctacct ctcccacctc
ttgctcttca
     1261 ataaaatgat cttaaactgc aaaaaaaaaa aaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_004308.2

LOCUS NP_004308

ACCESSION NP_004308

```
      1 maagvagwgv eaeefedapd vepleptlsn iieqrslkwi fvggkggvgk
ttcscslavq
       61 lskgresvli istdpahnis dafdqkfskv ptkvkgydnl fameidpslg
vaelpdeffe
      121 ednmlsmgkk mmqeamsafp gideamsyae vmrlvkgmnf svvvfdtapt
ghtlrllnfp
      181 tiverglgrl mqiknqispf isqmcnmlgl gdmnadqlas kleetlpvir
svseqfkdpe
      241 qttficvcia eflslyeter liqelakcki dthniivnql vfpdpekpck
mcearhkiqa
      301 kyldqmedly edfhivklpl lphevrgadk vntfsallle pykppsaq
```

PSMD3

Official Symbol: PSMD3

Official Name: proteasome (prosome, macropain) 26S subunit, non-ATPase, 3

Gene ID: 5709

Organism: Homo sapiens

Other Aliases: P58, RPN3, S3, TSTA2

Other Designations: 26S proteasome non-ATPase regulatory subunit 3; 26S proteasome regulatory subunit RPN3; 26S proteasome regulatory subunit S3; proteasome subunit p58; tissue specific transplantation antigen 2

Nucleotide sequence:

NCBI Reference Sequence: NM_002809.3

LOCUS NM_002809

ACCESSION NM_002809

```
      1 gttgactcgg ccatcggcct gccgggcctg gcgtttccca gaaggcccag
cgccgggaag
       61 gggtttgcag ctgctccgtc atcgtgcggc ccgacgctat ctcgcgctcg
tgtgcaggcc
      121 cggctcggct cctggtcccc ggtgcgaggg ttaacgcgag gccccggcct
cggtccccgg
      181 actaggccgt gaccccgggt gccatgaagc aggagggctc ggcgcggcgc
cgcggcgcgg
      241 acaaggcgaa accgccgccc ggcggaggag aacaagaacc cccaccgccg
ccggcccccc
      301 aggatgtgga gatgaaagag gaggcagcga cgggtggcgg gtcgacgggg
gaggcagacg
      361 gcaagacggc ggcggcagcg gctgagcact cccagcgaga gctggacaca
gtcaccttgg
      421 aggacatcaa ggagcacgtg aaacagctag agaaagcggt ttcaggcaag
gagccgagat
      481 tcgtgctgcg ggccctgcgg atgctgcctt ccacatcacg ccgcctcaac
cactatgttc
      541 tgtataaggc tgtgcagggc ttcttcactt caaataatgc cactcgagac
tttttgctcc
      601 ccttcctgga agagcccatg gacacagagg ctgatttaca gttccgtccc
cgcacgggaa
      661 aagctgcgtc gacacccctc ctgcctgaag tggaagccta tctccaactc
ctcgtggtca
      721 tcttcatgat gaacagcaag cgctacaaag aggcacagaa gatctctgat
gatctgatgc
      781 agaagatcag tactcagaac cgccgggccc tagaccttgt agccgcaaag
tgttactatt
      841 atcacgcccg ggtctatgag ttcctggaca agctggatgt ggtgcgcagc
ttcttgcatg
      901 ctcggctccg gacagctacg cttcggcatg acgcagacgg gcaggccacc
ctgttgaacc
      961 tcctgctgcg gaattaccta cactacagct tgtacgacca ggctgagaag
ctggtgtcca
     1021 agtctgtgtt cccagagcag gccaacaaca atgagtgggc caggtacctc
tactacacag
```

```
     1081 ggcgaatcaa agccatccag ctggagtact cagaggcccg gagaacgatg
accaacgccc
     1141 ttcgcaaggc ccctcagcac acagctgtcg gcttcaaaca gacggtgcac
aagcttctca
     1201 tcgtggtgga gctgttgctg ggggagatcc ctgaccggct gcagttccgc
cagccctccc
     1261 tcaagcgctc actcatgccc tatttccttc tgactcaagc tgtcaggaca
ggaaacctag
     1321 ccaagttcaa ccaggtcctg gatcagtttg gggagaagtt tcaagcagat
gggacctaca
     1381 ccctaattat ccggctgcgg cacaacgtga ttaagacagg tgtacgcatg
atcagcctct
     1441 cctattcccg aatctccttg gctgacatcg cccagaagct gcagttggat
agccccgaag
     1501 atgcagagtt cattgttgcc aaggccatcc gggatggtgt cattgaggcc
agcatcaacc
     1561 acgagaaggg ctatgtccaa tccaaggaga tgattgacat ctattccacc
cgagagcccc
     1621 agctagcctt ccaccagcgc atctccttct gcctagatat ccacaacatg
tctgtcaagg
     1681 ccatgaggtt tcctcccaaa tcgtacaaca aggacttgga gtctgcagag
gaacggcgtg
     1741 agcgagaaca gcaggacttg gagtttgcca aggagatggc agaagatgat
gatgacagct
     1801 tcccttgagc tgggggggctg gggaggggta gggggaatgg ggacaggctc
tttccccctt
     1861 gggggtcccc tgcccagggc actgtcccca ttttcccaca cacagctcat
atgctgcatt
     1921 cgtgcagggg gtgggggtgc tgggagccag ccaccctgac ctccccccagg
gctcctcccc
     1981 agccggtgac ttactgtaca gcaggcagga gggtgggcag gcaacctccc
cgggcagggt
     2041 cctggccagc agtgtgggag caggagggga aggatagttc tgtgtactcc
tttagggagt
     2101 gggggactag aactgggatg tcttggcttg tatgtttttt gaagcttcga
ttatgatttt
      2161 taaacaataa aaagttctcc acagtgc
```

Protein sequence:

NCBI Reference Sequence: NP_002800.2

LOCUS NP_002800

ACCESSION NP_002800

```
      1 mkqegsarrr gadkakpppg ggeqeppppp apqdvemkee aatgggstge
adgktaaaaa
       61 ehsqreldtv tledikehvk qlekavsgke prfvlralrm lpstsrrlnh
yvlykavqgf
      121 ftsnnatrdf llpfleepmd teadlqfrpr tgkaastpll peveaylqll
vvifmmnskr
      181 ykeaqkisdd lmqkistqnr raldlvaakc yyyharvyef ldkldvvrsf
lharlrtatl
      241 rhdadgqatl lnlllrnylh yslydqaekl vsksvfpeqa nnnewaryly
ytgrikaiql
      301 eysearrtmt nalrkapqht avgfkqtvhk llivvelllg eipdrlqfrq
pslkrslmpy
      361 flltqavrtg nlakfnqvld qfgekfqadg tytliirlrh nviktgvrmi
slsysrisla
```

421 diaqklqlds pedaefivak airdgvieas inhekgyvqs kemidiystr epqlafhqri
481 sfcldihnms vkamrfppks ynkdlesaee rrereqqdle fakemaeddd dsfp

IDH1

Official Symbol: IDH1

Official Name: isocitrate dehydrogenase 1 (NADP+), soluble

Gene ID: 3417

Organism: Homo sapiens

Other Aliases: IDCD, IDH, IDP, IDPC, PICD

Other Designations: NADP(+)-specific ICDH; NADP-dependent isocitrate dehydrogenase, cytosolic; NADP-dependent isocitrate dehydrogenase, peroxisomal; isocitrate dehydrogenase [NADP] cytoplasmic; oxalosuccinate decarboxylase

Nucleotide sequence:

NCBI Reference Sequence: NM_005896.2

LOCUS NM_005896

ACCESSION NM_005896

1 cctgtggtcc cgggtttctg cagagtctac ttcagaagcg gaggcactgg gagtccggtt
61 tgggattgcc aggctgtggt tgtgagtctg agcttgtgag cggctgtggc gccccaactc
121 ttcgccagca tatcatcccg gcaggcgata aactacattc agttgagtct gcaagactgg
181 gaggaactgg ggtgataaga aatctattca ctgtcaaggt ttattgaagt caaaatgtcc
241 aaaaaaatca gtggcggttc tgtggtagag atgcaaggag atgaaatgac acgaatcatt
301 tgggaattga ttaaagagaa actcattttt ccctacgtgg aattggatct acatagctat
361 gatttaggca tagagaatcg tgatgccacc aacgaccaag tcaccaagga tgctgcagaa
421 gctataaaga agcataatgt tggcgtcaaa tgtgccacta tcactcctga tgagaagagg
481 gttgaggagt tcaagttgaa acaaatgtgg aaatcaccaa atggcaccat acgaaatatt
541 ctgggtggca cggtcttcag agaagccatt atctgcaaaa atatcccccg gcttgtgagt
601 ggatgggtaa aacctatcat cataggtcgt catgcttatg gggatcaata cagagcaact
661 gattttgttg ttcctgggcc tggaaaagta gagataacct acacaccaag tgacggaacc
721 caaaaggtga catacctggt acataacttt gaagaaggtg gtggtgttgc catggggatg
781 tataatcaag ataagtcaat tgaagatttt gcacacagtt ccttccaaat ggctctgtct

```
      841 aagggttggc ctttgtatct gagcaccaaa aacactattc tgaagaaata
tgatgggcgt
      901 tttaaagaca tctttcagga gatatatgac aagcagtaca agtcccagtt
tgaagctcaa
      961 aagatctggt atgagcatag gctcatcgac gacatggtgg cccaagctat
gaaatcagag
     1021 ggaggcttca tctgggcctg taaaaactat gatggtgacg tgcagtcgga
ctctgtggcc
     1081 caagggtatg gctctctcgg catgatgacc agcgtgctgg tttgtccaga
tggcaagaca
     1141 gtagaagcag aggctgccca cgggactgta acccgtcact accgcatgta
ccagaaagga
     1201 caggagacgt ccaccaatcc cattgcttcc atttttgcct ggaccagagg
gttagcccac
     1261 agagcaaagc ttgataacaa taaagagctt gccttctttg caaatgcttt
ggaagaagtc
     1321 tctattgaga caattgaggc tggcttcatg accaaggact tggctgcttg
cattaaaggt
     1381 ttacccaatg tgcaacgttc tgactacttg aatacatttg agttcatgga
taaacttgga
     1441 gaaaacttga agatcaaact agctcaggcc aaactttaag ttcatacctg
agctaagaag
     1501 gataattgtc ttttggtaac taggtctaca ggtttacatt tttctgtgtt
acactcaagg
     1561 ataaaggcaa aatcaatttt gtaatttgtt tagaagccag agtttatctt
ttctataagt
     1621 ttacagcctt tttcttatat atacagttat tgccaccttt gtgaacatgg
caagggactt
     1681 ttttacaatt tttatttttat tttctagtac cagcctagga attcggttag
tactcatttg
     1741 tattcactgt cacttttttct catgttctaa ttataaatga ccaaaatcaa
gattgctcaa
     1801 aagggtaaat gatagccaca gtattgctcc ctaaaatatg cataaagtag
aaattcactg
     1861 ccttcccctc ctgtccatga ccttgggcac agggaagttc tggtgtcata
gatatcccgt
     1921 tttgtgaggt agagctgtgc attaaacttg cacatgactg gaacgaagta
tgagtgcaac
     1981 tcaaatgtgt tgaagatact gcagtcattt ttgtaaagac cttgctgaat
gtttccaata
     2041 gactaaatac tgtttaggcc gcaggagagt ttggaatccg gaataaatac
tacctggagg
     2101 tttgtcctct ccatttttct ctttctcctc ctggcctggc ctgaatatta
tactactcta
     2161 aatagcatat ttcatccaag tgcaataatg taagctgaat ctttttttgga
cttctgctgg
     2221 cctgttttat ttcttttata taaatgtgat ttctcagaaa ttgatattaa
acactatctt
      2281 atcttctcct gaactgttga ttttaattaa aattaagtgc taattaccaa
aaaaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_005887.2

LOCUS NP_005887

ACCESSION NP_005887

```
      1 mskkisggsv vemqgdemtr iiwelikekl ifpyveldlh sydlgienrd
atndqvtkda
       61 aeaikkhnvg vkcatitpde krveefklkq mwkspngtir nilggtvfre
aiickniprl
      121 vsgwvkpiii grhaygdqyr atdfvvpgpg kveitytpsd gtqkvtylvh
nfeegggvam
      181 gmynqdksie dfahssfqma lskgwplyls tkntilkkyd grfkdifqei
ydkqyksqfe
      241 aqkiwyehrl iddmvaqamk seggfiwack nydgdvqsds vaqgygslgm
mtsvlvcpdg
      301 ktveaeaahg tvtrhyrmyq kgqetstnpi asifawtrgl ahrakldnnk
elaffanale
      361 evsietieag fmtkdlaaci kglpnvqrsd ylntfefmdk lgenlkikla qakl
```

KPNB1

Official Symbol: KPNB1

Official Name: karyopherin (importin) beta 1

Gene ID: 3837

Organism: Homo sapiens

Other Aliases: IMB1, IPO1, IPOB, Impnb, NTF97

Other Designations: PTAC97; importin 1; importin 90; importin beta-1 subunit; importin subunit beta-1; importin-90; karyopherin subunit beta-1; nuclear factor p97; pore targeting complex 97 kDa subunit

Nucleotide sequence:

NCBI Reference Sequence: NM_002265.4

LOCUS NM_002265

ACCESSION NM_002265

```
      1 ctccctcgct ccctccctgc gcgccgcctc tcactcacag cctcccttcc
ttctttctcc
       61 ctccgcctcc cgagcaccag cgcgctctga gctgccccca gggtccctcc
cccgccgcca
      121 gcagcccatt tggagggagg aagtaaggga agaggagagg aaggggagcc
ggaccgacta
      181 cccagacaga gccggtgaat gggtttgtgg tgacccccgc cccccacccc
accctccctt
      241 cccacccgac ccccaacccc catccccagt tcgagccgcc gcccgaaagg
ccgggccgtc
      301 gtcttaggag gagtcgccgc cgccgccacc tccgccatgg agctgatcac
cattctcgag
      361 aagaccgtgt ctcccgatcg gctggagctg gaagcggcgc agaagttcct
ggagcgtgcg
      421 gccgtggaga acctgcccac tttccttgtg gaactgtcca gagtgctggc
aaatccagga
      481 aacagtcagg ttgccagagt tgcagctggt ctacaaatca agaactcttt
gacatctaaa
```

541 gatccagata tcaaggcaca atatcagcag aggtggcttg ctattgatgc
taatgctcga
601 cgagaagtca agaactatgt tttgcagaca ttgggtacag aaacttaccg
gcctagttct
661 gcctcacagt gtgtggctgg tattgcttgt gcagagatcc cagtaaacca
gtggccagaa
721 ctcattcctc agctggtggc caatgtcaca aaccccaaca gcacagagca
catgaaggag
781 tcgacattgg aagccatcgg ttatatttgc caagatatag acccagagca
gctacaagat
841 aaatccaatg agattctgac tgccataatc caggggatga ggaaagaaga
gcctagtaat
901 aatgtgaagc tagctgctac gaatgcactc ctgaactcat tggagttcac
caaagcaaac
961 tttgataaag agtctgaaag gcactttatt atgcaggtgg tctgtgaagc
cacacagtgt
1021 ccagatacga gggtacgagt ggctgcttta cagaatctgg tgaagataat
gtccttatat
1081 tatcagtaca tggagacata tatgggtcct gctctttttg caatcacaat
cgaagcaatg
1141 aaaagtgaca ttgatgaggt ggctttacaa gggatagaat tctggtccaa
tgtctgtgat
1201 gaggaaatgg atttggccat tgaagcttca gaggcagcag aacaaggacg
gccccctgag
1261 cacaccagca agttttatgc gaagggagca ctacagtatc tggttccaat
cctcacacag
1321 acactaacta aacaggacga aaatgatgat gacgatgact ggaacccctg
caaagcagca
1381 ggggtgtgcc tcatgcttct ggccacctgc tgtgaagatg acattgtccc
acatgtcctc
1441 cccttcatta aagaacacat caagaaccca gattggcggt accgggatgc
agcagtgatg
1501 gcttttggtt gtatcttgga aggaccagag cccagtcagc tcaaaccact
agttatacag
1561 gctatgccca ccctaataga attaatgaaa gaccccagtg tagttgttcg
agatacagct
1621 gcatggactg taggcagaat ttgtgagctg cttcctgaag ctgccatcaa
tgatgtctac
1681 ttggctcccc tgctacagtg tctgattgag ggtctcagtg ctgaacccag
agtggcttca
1741 aatgtgtgct gggctttctc cagtctggct gaagctgctt atgaagctgc
agacgttgct
1801 gatgatcagg aagaaccagc tacttactgc ttatcttctt catttgaact
catagttcag
1861 aagctcctag agactacaga cagacctgat ggacaccaga acaacctgag
gagttctgca
1921 tatgaatctc tgatggaaat tgtgaaaaac agtgccaagg attgttatcc
tgctgtccag
1981 aaaacgactt tggtcatcat ggaacgactg caacaggttc ttcagatgga
gtcacatatc
2041 cagagcacat ccgatagaat ccagttcaat gaccttcagt ctttactctg
tgcaactctt
2101 cagaatgttc ttcggaaagt gcaacatcaa gatgctttgc agatctctga
tgtggttatg
2161 gcctccctgt taaggatgtt ccaaagcaca gctgggtctg ggggagtaca
agaggatgcc
2221 ctgatggcag ttagcacact ggtggaagtg ttgggtggtg aattcctcaa
gtacatggag
2281 gcctttaaac ccttcctggg cattggatta aaaaattatg ctgaatacca
ggtttgtttg

2341 gcagctgtgg gcttagtggg agacttgtgc cgtgccctgc aatccaacat
cataccttte
2401 tgtgacgagg tgatgcagct gcttctggaa aatttgggga atgagaacgt
ccacaggtct
2461 gtgaagccgc agattctgtc agtgtttggt gatattgccc ttgctattgg
aggagagttt
2521 aaaaaatact tagaggttgt attgaatact cttcagcagg cctcccaagc
ccaggtggac
2581 aagtcagact atgacatggt ggattatctg aatgagctaa gggaaagctg
cttggaagcc
2641 tatactggaa tcgtccaggg attaaagggg gatcaggaga acgtacaccc
ggatgtgatg
2701 ctggtacaac ccagagtaga atttattctg tctttcattg accacattgc
tggagatgag
2761 gatcacacag atggagtagt agcttgtgct gctggactaa taggggactt
atgtacagca
2821 tttgggaagg atgtactgaa attagtagaa gctaggccaa tgatccatga
attgttaact
2881 gaagggcgga gatcgaagac taacaaagca aaaacccttg ctacatgggc
aacaaaagaa
2941 ctgaggaaac tgaagaacca agcttgatct gttaccattg ggatgataac
ctgaggaccc
3001 ccactggaaa tctcccatct tttgaaaaac ctggaagtga ggagtgtgca
cggatgctga
3061 atgtttggga atgagaggat gagtgagtga ggcttgaaaa cacaccacat
tgaaaatcct
3121 gccacagcag cagccgcagc cgccaacagc agcgctgtta gtgagctaag
taagcactga
3181 cttcgtagaa aaccataaca tcggccatct tggaaaagag aaaaacaatg
gagttactta
3241 tttaaaaaaa aagaaagaaa gttatctctt cccaggagag gctagaagta
gcttttctgt
3301 cttttggcca gtgccgagtg gaatgcctgg tttgggggag gaggagggac
tgggttcagc
3361 tgtggtgctt tgttgtaaaa ggcagcctgg cctttgctac tgaggagaaa
gatggagcct
3421 gggtctcaag cccaccttcg ctgtaccttt gccacatggt actgtatgct
tgccagctag
3481 aaggagggtc agggattttt tacagtctga gaatgagtgt gtgtgagtga
ggcggtatcc
3541 acattctcaa cttcaagtca ttgcagtttc tttttcccag aaaacaaggg
gttagatgtt
3601 gcatttcata aaactaaccg aagttctgtc tactgatgca gcacaagaga
tgtaaaaaaa
3661 aaaaaaaaaa aaaaaaaaaa aacacacaca cagaggaaag acgctcttta
ggttttgttt
3721 tgttttttt ttttggtttt gttttttgtt ttttttactc tagggaaaac
actgacgaat
3781 ggtcagagct cctatcctga tcttttcatc aaggcgcctt tcctaataat
atggttcaac
3841 tgtgaatgta gaagtggggg ggaggggggga gaaaaagaaa actctggcgt
tagaggatat
3901 agaaaaatat aagtacaatt gttacaaata acgcagactt caaaaacaaa
aaaatcacaa
3961 cccaaacaaa ccaaaattta aatgatcaga attggcagca caaagaaaac
gccctctcct
4021 gacttgtatt gtggcagtct gaacgccccc agaaaattgt gccaaagagt
ttagaaaaat
4081 aaatatacaa taaaagtaaa cacatacaca caaaacagca aacttcaggt
aactattttg

```
     4141 gattgcaaac aggataaatt aaatgttcaa acaatctgat aaaataacca
tttggaaact
      4201 gaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_002256.2

LOCUS NP_002256

ACCESSION NP_002256

```
      1 melitilekt vspdrlelea aqkfleraav enlptflvel srvlanpgns
qvarvaaglq
       61 iknsltskdp dikaqyqqrw laidanarre vknyvlqtlg tetyrpssas
qcvagiacae
      121 ipvnqwpeli pqlvanvtnp nstehmkest leaigyicqd idpeqlqdks
neiltaiiqg
      181 mrkeepsnnv klaatnalln sleftkanfd keserhfimq vvceatqcpd
trvrvaalqn
      241 lvkimslyyq ymetymgpal faitieamks didevalqgi efwsnvcdee
mdlaieasea
      301 aeqgrppeht skfyakgalq ylvpiltqtl tkqdendddd dwnpckaagv
clmllatcce
      361 ddivphvlpf ikehiknpdw ryrdaavmaf gcilegpeps qlkplviqam
ptlielmkdp
      421 svvvrdtaaw tvgricellp eaaindvyla pllqcliegl saeprvasnv
cwafsslaea
      481 ayeaadvadd qeepatycls ssfelivqkl lettdrpdgh qnnlrssaye
slmeivknsa
      541 kdcypavqkt tlvimerlqq vlqmeshiqs tsdriqfndl qsllcatlqn
vlrkvqhqda
      601 lqisdvvmas llrmfqstag sggvqedalm avstlvevlg geflkymeaf
kpflgiglkn
      661 yaeyqvclaa vglvgdlcra lqsniipfcd evmqlllenl gnenvhrsvk
pqilsvfgdi
      721 alaiggefkk ylevvlntlq qasqaqvdks dydmvdylne lrescleayt
givqglkgdq
      781 envhpdvmlv qprvefilsf idhiagdedh tdgvvacaag ligdlctafg
kdvlklvear
      841 pmihellteg rrsktnkakt latwatkelr klknqa
```

DDX17

Official Symbol: DDX17

Official Name: DEAD (Asp-Glu-Ala-Asp) box helicase 17

Gene ID: 10521

Organism: Homo sapiens

Other Aliases: RP3-434P1.1, P72, RH70

Other Designations: DEAD (Asp-Glu-Ala-Asp) box polypeptide 17; DEAD box protein p72; DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 17 (72kD); RNA-dependent helicase p72; probable ATP-dependent RNA helicase **DDX17**

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_006386.4

LOCUS NM_006386

ACCESSION NM_006386

```
      1 gttaagttgg agccgactca gcggcggccg ccattttgtg cagtcgctgg
gaaggaagga
       61 gacgcctaaa ccgcggcact gcccggtttg agcgtagcca aacctgccca
ccggctttgt
      121 agccccgatt ctctgtgttt tgctcccgtc tccgacgaga gaggcggcga
cggtggcgtc
      181 tgcgacggga gacagcgcgt cggagcgaga gagcgctgcg cctgccgccg
ccccaacagc
      241 ggaggcgccg ccgccatcgg tcgtcaccag accggagccg caggccctcc
cgagcccggc
      301 catccgtgcc ccgctcccag atctctatcc ttttgggacc atgcgcggag
gaggctttgg
      361 ggaccgggac cgggatcgtg accgtggagg atttggagca agaggtggtg
gtggccttcc
      421 cccgaagaaa tttggtaatc ctggggagcg tttgcgtaaa aaaaagtggg
atttgagtga
      481 gctccccaag tttgagaaaa atttttatgt ggaacatccg gaagtagcaa
ggctgacacc
      541 atatgaggtt gatgagctac gccgaaagaa ggagattaca gtgaggggggg
gagatgtttg
      601 tcctaaaccc gtgtttgcct tccatcatgc taacttccca caatatgtaa
tggatgtgtt
      661 gatggatcag cactttacag aaccaactcc aattcagtgc cagggatttc
cgttggctct
      721 tagtggccgg gatatggtgg gcattgctca gactggctct gggaagacgt
tggcgtatct
      781 cctgcctgca attgttcata ttaaccacca gccatacttg gaaaggggag
atggcccaat
      841 ctgtctagtt ctggctccta ccagagagct tgcccagcaa gtacagcagg
tggccgatga
      901 ctatggcaaa tgttctagat tgaagagtac ttgtatttat ggaggtgctc
ctaaaggtcc
      961 ccagattcga gacttggaaa gaggtgttga gatctgcata gccactcctg
gacgtctgat
     1021 agatttcctg gagtcaggaa agacaaatct tcgccgatgt acttaccttg
tattggacga
     1081 agctgacaga atgcttgata tggggtttga accccagatc cgtaaaattg
ttgaccaaat
     1141 caggcctgat aggcagacac tgatgtggag tgcaacctgg ccaaaagaag
taagacagct
     1201 tgcagaggat ttccttcgtg attacaccca gatcaacgta ggcaatctgg
agttgagtgc
     1261 caaccacaac atcctccaga tagtggatgt ctgcatggaa agtgaaaaag
accacaagtt
     1321 gatccaacta atggaagaaa taatggctga aaaggaaaac aaaacaataa
tatttgtgga
     1381 gacaaagaga cgctgtgatg atctgactcg aaggatgcgc agagatggtt
ggccagctat
```

1441 gtgtatccat ggagacaaga gtcaaccaga aagagattgg gtacttaatg
agttccgttc
1501 tggaaaggca cccatcctta ttgctacaga tgtagcctcc cgtgggctag
atgtggaaga
1561 tgtcaagttt gtgatcaact atgactatcc aaacagctca gaggattatg
tgcaccgtat
1621 tggccgaaca gcccgtagca ccaacaaggg taccgcctat accttcttca
ccccagggaa
1681 cctaaaacag gccagagagc ttatcaaagt gctggaagag gccaatcagg
ctatcaatcc
1741 aaaactgatg cagcttgtgg accacagagg aggcggcgga ggcgggggtg
gtcgttctcg
1801 ttaccggacc acttcttcag ccaacaatcc caatctgatg tatcaggatg
agtgtgaccg
1861 aaggcttcga ggagtcaagg atggtggccg gagagactct gcaagctatc
gggatcgtag
1921 tgaaaccgat agagctggtt atgctaatgg cagtggctat ggaagtccaa
attctgcctt
1981 tggagcacaa gcaggccaat acacctatgg tcaaggcacc tatggggcag
ctgcttatgg
2041 caccagtagc tatacagctc aagaatatgg tgctggcact tatggagcta
gtagcaccac
2101 ctcaactggg agaagttcac agagctctag ccagcagttt agtgggatag
gccggtctgg
2161 gcagcagcca cagccactga tgtcacaaca gtttgcacag cctccgggag
ctaccaatat
2221 gataggttac atggggcaga ctgcctacca ataccctcct cctcctcccc
ctcctcctcc
2281 ttcacgtaaa tgaaaccact caagtggtag tgactccagc agacttaatt
acattttaag
2341 gaacactgtc tttccttttt ttttcctctt cgccttttct ttttttttcc
ttttttcttt
2401 ttttttttt aatttttccc cccaaccatc gtgatttgtc ttttcatgca
gattagttag
2461 aattcactgc caggtttctt ctgcccacca aaatgatcca gtctggaata
acattttgta
2521 aaaaaaaaaa aaatatatat atatatatat agctgactgg aagagattaa
tttcttcccc
2581 caacttcttg catgttgaag atatttgagc tatttttcat ctaaaagagt
aaggtattag
2641 gccctttgt gggagcccca tgttttgttt ttctgagttg gtggggaggg
agggaggggg
2701 agggctgaat tgttttgcag aggaagatgg catctgtgct ttaaatttct
cattactggg
2761 ttagaaaaca aagagggatt gccctgcaca ttttcttttg tgcttttaaa
tgtttcttaa
2821 gttggaacag gtttcctcgg gcctgttttg actgattgct ggagtgcatt
tgatagttaa
2881 aaattactaa ttggttttat ttcccttcac actctgcctc cccacttctc
ccccccgttac
2941 tgaaaaataa ccattttagt gtcaggctag aaattgaatt gctgagtttt
gtgtatcctt
3001 taaattaaaa accacaagtg tttattgtag tggttaaact gtagcatctc
agcatctggg
3061 tggaagctgc ctatatttct tcccagttta actggggacc atctgtgaaa
ttaattttcc
3121 atccagacag ctgctgtgag caaatgaaca taaatgctcg ctggaaattt
actaaccagt
3181 ttttatattg acctgcagtg taaaaagcac atttaattat aaacaatata
ttcaaaatgg

```
     3241 gcaaatttta ttttcaaatg cagtgtagag ctagattaaa agcaactctt
tgccacctac
     3301 tctgcccttt tggcaaagtt accttgaaca aagaatctta agggtttatt
aagaactctt
     3361 tattttcttc ataccctgtt ctctgcagtg ctttctaaca gcttctgggt
gcagattttc
     3421 ttcggcatcc ttttgcactc agcttattac aggtaggtag tgcttaagaa
aagtcatgga
     3481 ggactaaagc ctaagtcctt ttcacttttc ctccatctga aggtaggtga
gttcatcctc
     3541 ttcatggtaa tgctgtttta ccaagacttt atagcagatg gacccagaaa
gaatttttctg
     3601 ctattgtgtt cactacaaca ggatagggac atcagacagc cccagaaacc
ccttccagat
     3661 ctgatatggg actattaatt tttatgctgt taattggtat tcattcacaa
tgcagttgaa
     3721 gggggaaggc tccactgcat tctttggcta aggcctgaat gcttgctcat
ctgtaagatc
     3781 tatactcgag gttttgtttt ccttttaaaa ttctttaggg agagagggat
ggtttctgag
     3841 gggttctgaa agtatgattc aatgtgcaac atacaggtag gtcttcagca
taagctgaaa
     3901 tatatgcatg taaaaacttt gacatctttt tttttaattt tccactttct
tcttaacttt
     3961 acttctcttt ttgtcccccc cccatcttac agaagttgag gccaagggag
aatggtaggc
     4021 acagaagaaa catggcaaac tgctctgtgc tttcaaacca aagtgttccc
cccaacccca
     4081 aatttgtcta agcactggcc agtctgttgt gggcattgtt ttctacaacc
aaattctggg
     4141 tttttttctt ctttctttaa acatagaggt accaccacaa gggatgccct
actctctcgc
     4201 agctcttgaa agcatctgtt tgagggaaag gtctctgggc aagcaagtgg
ttatttggat
     4261 tgcttgcttc ccttttcca cctgggacat tgtaatcata aaataacagt
aaattccaaa
     4321 cctcaaaaac tattatggcc tgagcacagc tgaaatctag cagagtttaa
ctcttctgcc
     4381 tccatgtctg tcacttataa ttcaggttct gctgttggct tcagaacatg
agcagaagaa
     4441 tcgttttatg ctagttattg cattcatggt tgaaactcaa cttagggaaa
gggttccaat
     4501 gtattaagca atgggctgct tctccccaat cctccctaac aattcgttgt
gtggacttct
     4561 catctaaaag gttagtggct tttgcttggg atcagtgctc tctattgatg
ttcttgctgg
     4621 tctccagaca cattcctgtt gcattaagac ttgaaagact tgtagatgtg
tgatgttcag
     4681 gcacaggatg ctgaaagcta tgttactatt cttagtttgt aaattgtcct
tttgatacca
     4741 tcatcttgtt ttctttttgt aggtataaat aaaaacactg ttgacaataa
aaaaaaaaaa
      4801 aaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_006377.2

LOCUS NP_006377

ACCESSION NP_006377

```
      1 mptgfvapil cvllpsptre aatvasatgd saseresaap aaaptaeapp
psvvtrpepq
       61 alpspairap lpdlypfgtm rgggfgdrdr drdrggfgar gggglppkkf
gnpgerlrkk
      121 kwdlselpkf eknfyvehpe varltpyevd elrrkkeitv rggdvcpkpv
fafhhanfpq
      181 yvmdvlmdqh fteptpiqcq gfplalsgrd mvgiaqtgsg ktlayllpai
vhinhqpyle
      241 rgdgpiclvl aptrelaqqv qqvaddygkc srlkstciyg gapkgpqird
lergveicia
      301 tpgrlidfle sgktnlrrct ylvldeadrm ldmgfepqir kivdqirpdr
qtlmwsatwp
      361 kevrqlaedf lrdytqinvg nlelsanhni lqivdvcmes ekdhkliqlm
eeimaekenk
      421 tiifvetkrr cddltrrmrr dgwpamcihg dksqperdwv lnefrsgkap
iliatdvasr
      481 gldvedvkfv inydypnsse dyvhrigrta rstnkgtayt fftpgnlkqa
relikvleea
      541 nqainpklmq lvdhrggggg gggrsryrtt ssannpnlmy qdecdrrlrg
vkdggrrdsa
      601 syrdrsetdr agyangsgyg spnsafgaqa gqytygqgty gaaaygtssy
taqeygagty
      661 gassttstgr ssqsssqqfs gigrsgqqpq plmsqqfaqp pgatnmigym
gqtayqyppp
      721 pppppppsrk
```

M6PRBP1

Official Symbol: PLIN3

Official Name: perilipin 3

Gene ID: 10226

Organism: Homo sapiens

Other Aliases: **M6PRBP1**, PP17, TIP47

Other Designations: 47 kDa MPR-binding protein; cargo selection protein TIP47; mannose-6-phosphate receptor-binding protein 1; perilipin-3; placental protein 17; tail-interacting protein, 47 kD

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_005817.4

LOCUS NM_005817

ACCESSION NM_005817

```
      1 tggcgcgggc aatccctcaa cctgattggt cccctcgccc gtcactccag
tgcgccccca
```

```
       61 acctaccacg cagtaaaagc cacccccgcc tcggcccgga cggtttccaa
gctggttttg
      121 aagtcgcggc agctgttcct gggacgtccg gttgaccgcg cgtctgctgc
agagaccatg
      181 tctgccgacg gggcagaggc tgatggcagc acccaggtga cagtggaaga
accggtacag
      241 cagcccagtg tggtggaccg tgtggccagc atgcctctga tcagctccac
ctgcgacatg
      301 gtgtccgcag cctatgcctc caccaaggag agctacccgc acatcaagac
tgtctgcgac
      361 gcagcagaga agggagtgag gaccctcacg gcggctgctg tcagcggggc
tcagccgatc
      421 ctctccaagc tggagcccca gattgcatca gccagcgaat acgcccacag
ggggctggac
      481 aagttggagg agaacctccc catcctgcag cagcccacgg agaaggtcct
ggcggacacc
      541 aaggagcttg tgtcgtctaa ggtgtcgggg gcccaagaga tggtgtctag
cgccaaggac
      601 acggtggcca cccaattgtc ggaggcggtg gacgcgaccc gcggtgctgt
gcagagcggc
      661 gtggacaaga caaagtccgt agtgaccggc ggcgtccaat cggtcatggg
ctcccgcttg
      721 ggccagatgg tgttgagtgg ggtcgacacg gtgctgggga agtcggagga
gtgggcggac
      781 aaccacctgc cccttacgga tgccgaactg gcccgcatcg ccacatccct
ggatggcttt
      841 gacgtcgcgt ccgtgcagca gcagcggcag gaacagagct acttcgtacg
tctgggctcc
      901 ctgtcggaga ggctgcggca gcacgcctat gagcactcgc tgggcaagct
tcgagccacc
      961 aagcagaggg cacaggaggc tctgctgcag ctgtcgcagg tcctaagcct
gatggaaact
     1021 gtcaagcaag gcgttgatca gaagctggtg gaaggccagg agaagctgca
ccagatgtgg
     1081 ctcagctgga accagaagca gctccagggc cccgagaagg agccgcccaa
gccagagcag
     1141 gtcgagtccc gggcgctcac catgttccgg gacattgccc agcaactgca
ggccacctgt
     1201 acctccctgg ggtccagcat tcagggcctc cccaccaatg tgaaggacca
ggtgcagcag
     1261 gcccgccgcc aggtggagga cctccaggcc acgttttcca gcatccactc
cttccaggac
     1321 ctgtccagca gcattctggc ccagagccgt gagcgtgtcg ccagcgcccg
cgaggccctg
     1381 gaccacatgg tggaatatgt ggcccagaac acacctgtca cgtggctcgt
gggacccttt
     1441 gcccctggaa tcactgagaa agccccggag gagaagaagt agggggagag
gagaggactc
     1501 agcgggcccc gtctctataa tgcagctgtg ctctggagtc ctcaacccgg
ggctcatttc
     1561 aaacttattt tctagccact cctcccagct cttctgtgct gtccacttgg
gaagctaagg
     1621 ctctcaaaac gggcatcacc cagttgaccc atctctcagc ctctctgagc
ttggaagaag
     1681 cctgttctga gcctcaccct atcagtcagt agagagagat gtccagaaaa
aatatctttc
     1741 aggaaagttc tcccctgcag aatttttttt ccttgttaaa tatcaggaat
ataggccggg
     1801 tgcggtggct cacacctgta atcccagcac tttgggaggc tgaggcgggc
ggaacacctg
```

```
     1861 aggtcaggtg ttcgagacca gccaggccaa catggtgaaa ccccgtctct
actaaaaata
     1921 caaaaaaaaa tgagccgggc atggtagcag gtgtctgtta tcccagttag
gaggctgagg
     1981 caagagaatc tcttgaacct gagaggcgga ggttgcagtg agccaagatc
gcgccattgc
     2041 actccagcct gggggacaag agtgagactt agtctcaaaa aaaaaaaaaa
agaaaaaaaa
     2101 atcagggata tagttcatat cccacttctt tgtttacacc gatgtccctg
aatatcagcc
     2161 tgtagctaat ggacttggga tttctggtct aagtgggcct cctggggatg
gggtggtaca
     2221 ctgagcttct gagcctcatt gtagagtaga aaggtactgg ggcctgtgtg
gtaagccttg
     2281 ttgaaatgct ctggtattca gtattgcctt aataaacttc acccacaact
gcatacaggc
     2341 aaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_005808.3

LOCUS NP_005808

ACCESSION NP_005808

```
      1 msadgaeadg stqvtveepv qqpsvvdrva smplisstcd mvsaayastk
esyphiktvc
       61 daaekgvrtl taaavsgaqp ilsklepqia saseyahrgl dkleenlpil
qqptekvlad
      121 tkelvsskvs gaqemvssak dtvatqlsea vdatrgavqs gvdktksvvt
ggvqsvmgsr
      181 lgqmvlsgvd tvlgkseewa dnhlpltdae lariatsldg fdvasvqqqr
qeqsyfvrlg
      241 slserlrqha yehslgklra tkqraqeall qlsqvlslme tvkqgvdqkl
vegqeklhqm
      301 wlswnqkqlq gpekeppkpe qvesraltmf rdiaqqlqat ctslgssiqg
lptnvkdqvq
      361 qarrqvedlq atfssihsfq dlsssilaqs rervasarea ldhmveyvaq
ntpvtwlvgp
      421 fapgitekap eekk
```

EIF4A3

Official Symbol: EIF4A3

Official Name: eukaryotic translation initiation factor 4A3

Gene ID: 9775

Organism: Homo sapiens

Other Aliases: DDX48, NMP265, NUK34, eIF4AIII

Other Designations: ATP-dependent RNA helicase DDX48; ATP-dependent RNA helicase eIF4A-3; DEAD (Asp-Glu-Ala-Asp) box polypeptide 48; DEAD box

protein 48; NMP 265; eIF-4A-III; eIF4A-III; eukaryotic initiation factor 4A-III; eukaryotic initiation factor 4A-like NUK-34; eukaryotic translation initiation factor 4A; hNMP 265; nuclear matrix protein 265

Nucleotide sequence:

NCBI Reference Sequence: NM_014740.3

LOCUS NM_014740

ACCESSION NM_014740

```
      1 acgcacgcac gtctctcgct ttcgcatact taaggcgtct gttctcggca
gcggcacagc
       61 gaggtcggca gcggcacagc gaggtcggca gcggcacagc gaggtcggca
gcggcacagc
      121 gaggtcggca gcggcagcga ggtcggcagc ggcacagcga ggtcggcagc
ggcagcgagg
      181 tcggcagcgg cgcgcgctgt gctcttccgc ggactctgaa tcatggcgac
cacggccacg
      241 atggcgacct cgggctcggc gcgaaagcgg ctgctcaaag aggaagacat
gactaaagtg
      301 gaattcgaga ccagcgagga ggtggatgtg acccccacgt tcgacaccat
gggcctgcgg
      361 gaggacctgc tgcggggcat ctacgcttac ggttttgaaa aaccatcagc
aatccagcaa
      421 cgagcaatca agcagatcat caaagggaga gatgtcatcg cacagtctca
gtccggcaca
      481 ggaaaaacag ccaccttcag tatctcagtc ctccagtgtt tggatattca
ggttcgtgaa
      541 actcaagctt tgatcttggc tcccacaaga gagttggctg tgcagatcca
gaaggggctg
      601 cttgctctcg gtgactacat gaatgtccag tgccatgcct gcattggagg
caccaatgtt
      661 ggcgaggaca tcaggaagct ggattacgga cagcatgttg tcgcgggcac
tccagggcgt
      721 gtttttgata tgattcgtcg cagaagccta aggacacgtg ctatcaaaat
gttggttttg
      781 gatgaagctg atgaaatgtt gaataaaggt ttcaaagagc agatttacga
tgtatacagg
      841 tacctgcctc cagccacaca ggtggttctc atcagtgcca cgctgccaca
cgagattctg
      901 gagatgacca acaagttcat gaccgaccca atccgcatct tggtgaaacg
tgatgaattg
      961 actctggaag gcatcaagca atttttcgtg gcagtggaga gggaagagtg
gaaatttgac
     1021 actctgtgtg acctctacga cacactgacc atcactcagg cggtcatctt
ctgcaacacc
     1081 aaaagaaagg tggactggct gacggagaaa atgagggaag ccaacttcac
tgtatcctca
     1141 atgcatggag acatgcccca gaaagagcgg gagtccatca tgaaggagtt
ccggtcgggc
     1201 gccagccgag tgcttatttc tacagatgtc tgggccaggg ggttggatgt
ccctcaggtg
     1261 tccctcatca ttaactatga tctccctaat aacagagaat tgtacataca
cagaattggg
     1321 agatcaggtc gatacggccg gaagggtgtg gccattaact ttgtaaagaa
tgacgacatc
     1381 cgcatcctca gagatatcga gcagtactat tccactcaga ttgatgagat
gccgatgaac
```

```
     1441 gttgctgatc ttatctgaag cagcagatca gtgggatgag ggagactgtt
cacctgctgt
     1501 gtactcctgt ttggaagtat ttagatccag attctactta atggggttta
tatggacttt
     1561 cttctcataa atggcctgcc gtctcccttc ctttgaagag gatatgggga
ttctgctctc
     1621 ttttcttatt tacatgtaaa taatacattg ttctaagtct ttttcattaa
aaatttaaaa
      1681 cttttcccat aaactctata cttctaaggt gccaccacct tctctagtaa ctta
```

Protein sequence:

NCBI Reference Sequence: NP_055555.1

LOCUS NP_055555

ACCESSION NP_055555

```
      1 mattatmats gsarkrllke edmtkvefet seevdvtptf dtmglredll
rgiyaygfek
       61 psaiqqraik qiikgrdvia qsqsgtgkta tfsisvlqcl diqvretqal
ilaptrelav
      121 qiqkgllalg dymnvqchac iggtnvgedi rkldygqhvv agtpgrvfdm
irrrslrtra
      181 ikmlvldead emlnkgfkeq iydvyrylpp atqvvlisat lpheilemtn
kfmtdpiril
      241 vkrdeltleg ikqffvaver eewkfdtlcd lydtltitqa vifcntkrkv
dwltekmrea
      301 nftvssmhgd mpqkeresim kefrsgasrv listdvwarg ldvpqvslii
nydlpnnrel
      361 yihrigrsgr ygrkgvainf vknddirilr dieqyystqi dempmnvadl i
```

IQGAP1

Official Symbol: IQGAP1

Official Name: IQ motif containing GTPase activating protein 1

Gene ID: 8826

Organism: Homo sapiens

Other Aliases: HUMORFA01, SAR1, p195

Other Designations: RasGAP-like with IQ motifs; ras GTPase-activating-like protein IQGAP1

Nucleotide sequence:

NCBI Reference Sequence: NM_003870.3

LOCUS NM_003870

ACCESSION NM_003870

```
      1 ggaccccggc aagcccgcgc acttggcagg agctgtagct accgccgtcc
gcgcctccaa
```

```
       61 ggtttcacgg cttcctcagc agagactcgg gctcgtccgc catgtccgcc
gcagacgagg
      121 ttgacgggct gggcgtggcc cggccgcact atggctctgt cctggataat
gaaagactta
      181 ctgcagagga gatggatgaa aggagacgtc agaacgtggc ttatgagtac
ctttgtcatt
      241 tggaagaagc gaagaggtgg atggaagcat gcctagggga agatctgcct
cccaccacag
      301 aactggagga ggggcttagg aatggggtct accttgccaa actggggaac
ttcttctctc
      361 ccaaagtagt gtccctgaaa aaaatctatg atcgagaaca gaccagatac
aaggcgactg
      421 gcctccactt tagacacact gataatgtga ttcagtggtt gaatgccatg
gatgagattg
      481 gattgcctaa gatttttac ccagaaacta cagatatcta tgatcgaaag
aacatgccaa
      541 gatgtatcta ctgtatccat gcactcagtt tgtacctgtt caagctaggc
ctggcccctc
      601 agattcaaga cctatatgga aaggttgact tcacagaaga agaaatcaac
aacatgaaga
      661 ctgagttgga gaagtatggc atccagatgc ctgcctttag caagattggg
ggcatcttgg
      721 ctaatgaact gtcagtggat gaagccgcat tacatgctgc tgttattgct
attaatgaag
      781 ctattgaccg tagaattcca gccgacacat ttgcagcttt gaaaaatccg
aatgccatgc
      841 ttgtaaatct tgaagagccc ttggcatcca cttaccagga tatactttac
caggctaagc
      901 aggacaaaat gacaaatgct aaaaacagga cagaaaactc agagagagaa
agagatgttt
      961 atgaggagct gctcacgcaa gctgaaattc aaggcaatat aaacaaagtc
aatacatttt
     1021 ctgcattagc aaatatcgac ctggctttag aacaaggaga tgcactggcc
ttgttcaggg
     1081 ctctgcagtc accagccctg gggcttcgag gactgcagca acagaatagc
gactggtact
     1141 tgaagcagct cctgagtgat aaacagcaga agagacagag tggtcagact
gacccctgc
     1201 agaaggagga gctgcagtct ggagtggatg ctgcaaacag tgctgcccag
caatatcaga
     1261 gaagattggc agcagtagca ctgattaatg ctgcaatcca gaagggtgtt
gctgagaaga
     1321 ctgttttgga actgatgaat cccgaagccc agctgcccca ggtgtatcca
tttgccgccg
     1381 atctctatca gaaggagctg gctaccctgc agcgacaaag tcctgaacat
aatctcaccc
     1441 acccagagct ctctgtcgca gtggagatgt tgtcatcggt ggccctgatc
aacagggcat
     1501 tggaatcagg agatgtgaat acagtgtgga agcaattgag cagttcagtt
actggtctta
     1561 ccaatattga ggaagaaaac tgtcagaggt atctcgatga gttgatgaaa
ctgaaggctc
     1621 aggcacatgc agagaataat gaattcatta catggaatga tatccaagct
tgcgtggacc
     1681 atgtgaacct ggtggtgcaa gaggaacatg agaggatttt agccattggt
ttaattaatg
     1741 aagccctgga tgaaggtgat gcccaaaaga ctctgcaggc cctacagatt
cctgcagcta
     1801 aacttgaggg agtccttgca gaagtggccc agcattacca agacacgctg
attagagcga
```

1861 agagagagaa agcccaggaa atccaggatg agtcagctgt gttatggttg
gatgaaattc
1921 aaggtggaat ctggcagtcc aacaaagaca cccaagaagc acagaagttt
gccttaggaa
1981 tctttgccat taatgaggca gtagaaagtg gtgatgttgg caaaacactg
agtgcccttc
2041 gctcccctga tgttggcttg tatggagtca tccctgagtg tggtgaaact
taccacagtg
2101 atcttgctga agccaagaag aaaaaactgg cagtaggaga taataacagc
aagtgggtga
2161 agcactgggt aaaaggtgga tattattatt accacaatct ggagacccag
gaaggaggat
2221 gggatgaacc tccaaatttt gtgcaaaatt ctatgcagct ttctcgggag
gagatccaga
2281 gttctatctc tggggtgact gccgcatata accgagaaca gctgtggctg
gccaatgaag
2341 gcctgatcac caggctgcag gctcgctgcc gtggatactt agttcgacag
gaattccgat
2401 ccaggatgaa tttcctgaag aaacaaatcc ctgccatcac ctgcattcag
tcacagtgga
2461 gaggatacaa gcagaagaag gcatatcaag atcggttagc ttacctgcgc
tcccacaaag
2521 atgaagttgt aaagattcag tccctggcaa ggatgcacca agctcgaaag
cgctatcgag
2581 atcgcctgca gtacttccgg gaccatataa atgacattat caaaatccag
gcttttattc
2641 gggcaaacaa agctcgggat gactacaaga ctctcatcaa tgctgaggat
cctcctatgg
2701 ttgtggtccg aaaatttgtc cacctgctgg accaaagtga ccaggatttt
caggaggagc
2761 ttgaccttat gaagatgcgg gaagaggtta tcaccctcat tcgttctaac
cagcagctgg
2821 agaatgacct caatctcatg gatatcaaaa ttggactgct agtgaaaaat
aagattacgt
2881 tgcaggatgt ggtttcccac agtaaaaaac ttaccaaaaa aaataaggaa
cagttgtctg
2941 atatgatgat gataaataaa cagaagggag gtctcaaggc tttgagcaag
gagaagagag
3001 agaagttgga agcttaccag cacctgtttt atttattgca aaccaatccc
acctatctgg
3061 ccaagctcat ttttcagatg ccccagaaca agtccaccaa gttcatggac
tctgtaatct
3121 tcacactcta caactacgcg tccaaccagc gagaggagta cctgctcctg
cggctcttta
3181 agacagcact ccaagaggaa atcaagtcga aggtagatca gattcaagag
attgtgacag
3241 gaaatcctac ggttattaaa atggttgtaa gtttcaaccg tggtgcccgt
ggccagaatg
3301 ccctgagaca gatcttggcc ccagtcgtga aggaaattat ggatgacaaa
tctctcaaca
3361 tcaaaactga ccctgtggat atttacaaat cttgggttaa tcagatggag
tctcagacag
3421 gagaggcaag caaactgccc tatgatgtga cccctgagca ggcgctagct
catgaagaag
3481 tgaagacacg gctagacagc tccatcagga acatgcgggc tgtgacagac
aagtttctct
3541 cagccattgt cagctctgtg gacaaaatcc cttatgggat gcgcttcatt
gccaaagtgc
3601 tgaaggactc gttgcatgag aagttccctg atgctggtga ggatgagctg
ctgaagatta

3661 ttggtaactt gctttattat cgatacatga atccagccat tgttgctcct
gatgcctttg
3721 acatcattga cctgtcagca ggaggccagc ttaccacaga ccaacgccga
aatctgggct
3781 ccattgcaaa aatgcttcag catgctgctt ccaataagat gtttctggga
gataatgccc
3841 acttaagcat cattaatgaa tatctttccc agtcctacca gaaattcaga
cggtttttcc
3901 aaactgcttg tgatgtccca gagcttcagg ataaatttaa tgtggatgag
tactctgatt
3961 tagtaaccct caccaaacca gtaatctaca tttccattgg tgaaatcatc
aacacccaca
4021 ctctcctgtt ggatcaccag gatgccattg ctccggagca caatgatcca
atccacgaac
4081 tgctggacga cctcggcgag gtgcccacca tcgagtccct gataggggaa
agctctggca
4141 atttaaatga cccaaataag gaggcactgg ctaagacgga agtgtctctc
accctgacca
4201 acaagttcga cgtgcctgga gatgagaatg cagaaatgga tgctcgaacc
atcttactga
4261 atacaaaacg tttaattgtg gatgtcatcc ggttccagcc aggagagacc
ttgactgaaa
4321 tcctagaaac accagccacc agtgaacagg aagcagaaca tcagagagcc
atgcagagac
4381 gtgctatccg tgatgccaaa acacctgaca agatgaaaaa gtcaaaatct
gtaaaggaag
4441 acagcaacct cactcttcaa gagaagaaag agaagatcca gacaggttta
aagaagctaa
4501 cagagcttgg aaccgtggac ccaaagaaca aataccagga actgatcaac
gacattgcca
4561 gggatattcg gaatcagcgg aggtaccgac agaggagaaa ggccgaacta
gtgaaactgc
4621 aacagacata cgctgctctg aactctaagg ccacctttta tggggagcag
gtggattact
4681 ataaaagcta tatcaaaacc tgcttggata acttagccag caagggcaaa
gtctccaaaa
4741 agcctaggga aatgaaagga aagaaaagca aaaagatttc tctgaaatat
acagcagcaa
4801 gactacatga aaaaggagtt cttctggaaa ttgaggacct gcaagtgaat
cagtttaaaa
4861 atgttatatt tgaaatcagt ccaacagaag aagttggaga cttcgaagtg
aaagccaaat
4921 tcatgggagt tcaaatggag acttttatgt tacattatca ggacctgctg
cagctacagt
4981 atgaaggagt tgcagtcatg aaattatttg atagagctaa agtaaatgtc
aacctcctga
5041 tcttccttct caacaaaaag ttctacggga agtaattgat cgtttgctgc
cagcccagaa
5101 ggatgaagga aagaagcacc tcacagctcc tttctaggtc cttctttcct
cattggaagc
5161 aaagacctag ccaacaacag cacctcaatc tgatacactc ccgatgccac
atttttaact
5221 cctctcgctc tgatgggaca tttgttaccc ttttttcata gtgaaattgt
gtttcaggct
5281 tagtctgacc tttctggttt cttcattttc ttccattact taggaaagag
tggaaactcc
5341 actaaaattt ctctgtgttg ttacagtctt agaggttgca gtactatatt
gtaagctttg
5401 gtgtttgttt aattagcaat agggatggta ggattcaaat gtgtgtcatt
tagaagtgga

5461 agctattagc accaatgaca taaatacata caagacacac aactaaaatg
tcatgttatt
5521 aacagttatt aggttgtcat ttaaaaataa agttccttta tatttctgtc
ccatcaggaa
5581 aactgaagga tatggggaat cattggttat cttccattgt gtttttcttt
atggacagga
5641 gctaatggaa gtgacagtca tgttcaaagg aagcatttct agaaaaaagg
agataatgtt
5701 tttaaatttc attatcaaac ttgggcaatt ctgtttgtgt aactccccga
ctagtggatg
5761 ggagagtccc attgctaaaa ttcagctact cagataaatt cagaatgggt
caaggcacct
5821 gcctgttttt gttggtgcac agagattgac ttgattcaga gagacaattc
actccatccc
5881 tatggcagag gaatgggtta gccctaatgt agaatgtcat tgtttttaaa
actgttttat
5941 atcttaagag tgccttatta aagtatagat gtatgtctta aaatgtgggt
gataggaatt
6001 ttaaagattt atataatgca tcaaaagcct tagaataaga aaagcttttt
ttaaattgct
6061 ttatctgtat atctgaactc ttgaaactta tagctaaaac actaggattt
atctgcagtg
6121 ttcagggaga taattctgcc tttaattgtc taaaacaaaa acaaaaccag
ccaacctatg
6181 ttacacgtga gattaaaacc aatttttcc ccatttttc tccttttttc
tcttgctgcc
6241 cacattgtgc ctttatttta tgagccccag ttttctgggc ttagtttaaa
aaaaaaatca
6301 agtctaaaca ttgcatttag aaagcttttg ttcttggata aaaagtcata
cactttaaaa
6361 aaaaaaaaaa ctttttccag gaaaatatat tgaaatcatg ctgctgagcc
tctatttttct
6421 ttctttgatg ttttgattca gtattctttt atcataaatt tttagcattt
aaaaattcac
6481 tgatgtacat taagccaata aactgcttta atgaataaca aactatgtag
tgtgtcccta
6541 ttataaatgc attggagaag tatttttatg agactcttta ctcaggtgca
tggttacagc
6601 ccacagggag gcatggagtg ccatggaagg attcgccact acccagacct
tgtttttttgt
6661 tgtattttgg aagacaggtt ttttaaagaa acattttcct cagattaaaa
gatgatgcta
6721 ttacaactag cattgcctca aaaactggga ccaaccaaag tgtgtcaacc
ctgtttcctt
6781 aaaagaggct atgaatccca aaggccacat ccaagacagg caataatgag
cagagtttac
6841 agctccttta ataaaatgtg tcagtaattt taaggtttat agttccctca
acacaattgc
6901 taatgcagaa tagtgtaaaa tgcgcttcaa gaatgttgat gatgatgata
tagaattgtg
6961 gctttagtag cacagaggat gccccaacaa actcatggcg ttgaaaccac
acagttctca
7021 ttactgttat ttattagctg tagcattctc tgtctcctct ctctcctcct
ttgaccttct
7081 cctcgaccag ccatcatgac atttaccatg aatttacttc ctcccaagag
tttggactgc
7141 ccgtcagatt gttgctgcac atagttgcct ttgtatctct gtatgaaata
aaaggtcatt
7201 tgttcatgtt aaaaaaaaa

Protein sequence:

NCBI Reference Sequence: NP_003861.1

LOCUS NP_003861

ACCESSION NP_003861

```
      1 msaadevdgl gvarphygsv ldnerltaee mderrrqnva yeylchleea
krwmeaclge
       61 dlpptteleе glrngvylak lgnffspkvv slkkiydreq trykatglhf
rhtdnviqwl
      121 namdeiglpk ifypettdiy drknmprciy cihalslylf klglapqiqd
lygkvdftee
      181 einnmktele kygiqmpafs kiggilanel svdeaalhaa viaineaidr
ripadtfaal
      241 knpnamlvnl eeplastyqd ilyqakqdkm tnaknrtens ererdvyeel
ltqaeiqgni
      301 nkvntfsala nidlaleqgd alalfralqs palglrglqq qnsdwylkql
lsdkqqkrqs
      361 gqtdplqkee lqsgvdaans aaqqyqrrla avalinaaiq kgvaektvle
lmnpeaqlpq
      421 vypfaadlyq kelatlqrqs pehnlthpel svavemlssv alinralesg
dvntvwkqls
      481 ssvtgltnie eencqrylde lmklkaqaha ennefitwnd iqacvdhvnl
vvqeeheril
      541 aiglineald egdaqktlqa lqipaakleg vlaevaqhyq dtlirakrek
aqeiqdesav
      601 lwldeiqggi wqsnkdtqea qkfalgifai neavesgdvg ktlsalrspd
vglygvipec
      661 getyhsdlae akkkklavgd nnskwvkhwv kggyyyyhnl etqeggwdep
pnfvqnsmql
      721 sreeiqssis gvtaaynreq lwlaneglit rlqarcrgyl vrqefrsrmn
flkkqipait
      781 ciqsqwrgyk qkkayqdrla ylrshkdevv kiqslarmhq arkryrdrlq
yfrdhindii
      841 kiqafirank arddyktlin aedppmvvvr kfvhlldqsd qdfqeeldlm
kmreevitli
      901 rsnqqlendl nlmdikigll vknkitlqdv vshskkltkk nkeqlsdmmm
inkqkgglka
      961 lskekrekle ayqhlfyllq tnptylakli fqmpqnkstk fmdsviftly
nyasnqreey
     1021 lllrlfktal qeeikskvdq iqeivtgnpt vikmvvsfnr gargqnalrq
ilapvvkeim
     1081 ddkslniktd pvdiykswvn qmesqtgeas klpydvtpeq alaheevktr
ldssirnmra
     1141 vtdkflsaiv ssvdkipygm rfiakvlkds lhekfpdage dellkiignl
lyyrymnpai
     1201 vapdafdiid lsaggqlttd qrrnlgsiak mlqhaasnkm flgdnahlsi
ineylsqsyq
     1261 kfrrffqtac dvpelqdkfn vdeysdlvtl tkpviyisig eiinthtlll
dhqdaiapeh
     1321 ndpihelldd lgevptiesl igessgnlnd pnkealakte vsltltnkfd
vpgdenaemd
     1381 artillntkr livdvirfqp getlteilet patseqeaeh qramqrrair
daktpdkmkk
     1441 sksvkedsnl tlqekkekiq tglkkltelg tvdpknkyqe lindiardir
nqrryrqrrk
     1501 aelvklqqty aalnskatfy geqvdyyksy iktcldnlas kgkvskkpre
mkgkkskkis
     1561 lkytaarlhe kgvlleiedl qvnqfknvif eispteevgd fevkakfmgv
qmetfmlhyq
```

1621 dllqlqyegv avmklfdrak vnvnllifll nkkfygk

SFRS2

Official Symbol:SRSF2

Official Name: serine/arginine-rich splicing factor 2

Gene ID: 6427

Organism: Homo sapiens

Other Aliases: PR264, SC-35, SC35, **SFRS2**, SFRS2A, SRp30b

Other Designations: SR splicing factor 2; splicing component, 35 kDa; splicing factor SC35; splicing factor, arginine/serine-rich 2

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_003016.4

LOCUS NM_003016

ACCESSION NM_003016

1 agaaggtttc atttccgggt ggcgcgggcg ccattttgtg aggagcgata taaacgggcg
61 cagaggccgg ctgcccgccc agttgttact caggtgcgct agcctgcgga gcccgtccgt
121 gctgttctgc ggcaaggcct ttcccagtgt ccccacgcgg aaggcaactg cctgagaggc
181 gcggcgtcgc accgcccaga gctgaggaag ccggcgccag ttcgcggggc tccgggccgc
241 cactcagagc tatgagctac ggccgccccc ctcccgatgt ggagggtatg acctccctca
301 aggtggacaa cctgacctac cgcacctcgc ccgacacgct gaggcgcgtc ttcgagaagt
361 acgggcgcgt cggcgacgtg tacatcccgc gggaccgcta caccaaggag tcccgcggct
421 tcgccttcgt tcgctttcac gacaagcgcg acgctgagga cgctatggat gccatggacg
481 gggccgtgct ggacggccgc gagctgcggg tgcaaatggc gcgctacggc cgccccccgg
541 actcacacca cagccgccgg ggaccgccac cccgcaggta cgggggcggt ggctacggac
601 gccggagccg cagccctagg cggcgtcgcc gcagccgatc ccggagtcgg agccgttcca
661 ggtctcgcag ccgatctcgc tacagccgct cgaagtctcg gtcccgcact cgttctcgat
721 ctcggtcgac ctccaagtcc agatccgcac gaaggtccaa gtccaagtcc tcgtcggtct
781 ccagatctcg ttcgcggtcc aggtcccggt ctcggtccag gagtcctccc ccagtgtcca
841 agagggaatc caaatccagg tcgcgatcga agagtccccc caagtctcct gaagaggaag
901 gagcggtgtc ctcttaagaa aatggtaatg tctgggaatc cgagacacat aaccctaatt

961 cataaatggg atttggggta ggtctttttg agtcgtgtta atgtaagaat gactcctatc
1021 attaggagtg ctgctcggag gttactcacc tttgggagta atactgaaga gaggggtctg
1081 cagaaaggat gtgtatgaag cttagataat aatggctgtt tcgtaaactg tttgagacct
1141 attaatgaaa atgactattt cttgctgttt ttatccaacg tctgcatttt ccccctttaa
1201 agctgcggtc tcctgtttga taaaagaata ttggccagta ttgcagattt taactgattt
1261 ggctgatcct ccagggacca gtttctgtgg gcgtgtattg gagcaggttt gtctttaaat
1321 gttaaagatg cactatcctc ttagagaaac aatcagttca actattgttg tactgactgg
1381 gacttcatat tctaatggat gtggcaaaag aattgcaata agaagcagtg aacatttgga
1441 accccaaaag aaagttacag gtattgcact gggtggggaa aggatagtgt gtctttaact
1501 cttaaattgt ttggtcctat tttttaaaaa ggaaagggcc ctaagtagct cagatattaa
1561 agtagtattc tcaattacca aatgtttcat ttgaaacaat ttatcttaat gaaatataga
1621 ccaattctct gatctcgagt tgtttttgtt tggatacagc ccttttttttt ttctttttttt
1681 ttcttcccct tacctttctt caccttggtt atttggccag gaatacgtaa attcaaactt
1741 gtacatgctg atggtagcct ttgtgaaatt ttcctaattg ggccttttaa aaacatggct
1801 gggtggaaca tttctgtacc ctactggttt gaccagagcc ttagtaagta cgtgcctgaa
1861 actgaaacca tgtgcacttt aatggaaggt aagctgaact tctttctttt caaacctaga
1921 tgtatcggca agcagtgtaa acggaggact tggggaaaaa ggaccacata gtccatcgaa
1981 gaagagtcct tggaacaagc aactggctat tgaaaaggtt attttgtaac atttgtctaa
2041 ctttttactt gtttaagctt tgcctcagtt ggcaaacttc attttatgtg ccatttttgtt
2101 gctgttattc aaatttcttg taatttagtg aggtgaacga cttcagattt cattattgga
2161 tttggatatt tgaggtaaaa tttcattttg ttatatagtg ctgacttttt ttgtttgaaa
2221 ttaaacagat tggtaaccta atttgtggcc tcctgacttt taaggaaaac gtgtgcagcc
2281 attacacaca gcctaaagct gtcaagagat tgactcggca ttgccttcat tccttaaaat
2341 taaaaaccta caaaagttgg tgtaaatttg tatatgttat ttaccttcag atctaaatgg
2401 taatctgaac ccaaatttgt ataaagactt ttcaggtgaa aagacttgat tttttgaaag
2461 gattgtttat caaacacaat tctaatctct tctcttatgt atttttgtgc actaggcgca
2521 gttgtgtagc agttgagtaa tgctggttag ctgttaaggt ggcgtgttgc agtgcagagt
2581 gcttggctgt ttcctgtttt ctcccgattg ctcctgtgta aagatgcctt gtcgtgcaga
2641 aacaaatggc tgtccagttt attaaaatgc ctgacaactg cacttccagt cacccgggcc
2701 ttgcatataa ataacggagc atacagtgag cacatctagc tgatgataaa tacacctttt

```
     2761 tttccctctt ccccctaaaa atggtaaatc tgatcatatc tacatgtatg
aacttaacat
     2821 ggaaaatgtt aaggaagcaa atggttgtaa ctttgtaagt acttataaca
tggtgtatct
     2881 ttttgcttat gaatattctg tattataacc attgtttctg tagtttaatt
aaaacatttt
     2941 cttggtgtta gcttttctca gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
      3001 aaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_003007.2

LOCUS NP_003007

ACCESSION NP_003007

```
      1 msygrpppdv egmtslkvdn ltyrtspdtl rrvfekygrv gdvyiprdry
tkesrgfafv
       61 rfhdkrdaed amdamdgavl dgrelrvqma rygrppdshh srrgppprry
ggggygrrsr
      121 sprrrrrsrs rsrsrsrsrs rsrysrsksr srtrsrsrst sksrsarrsk
sksssvsrsr
      181 srsrsrsrsr spppvskres ksrsrskspp kspeeegavs s
```

GOLGA3

Official Symbol: GOLGA3

Official Name: golgin A3

Gene ID: 2802

Organism: Homo sapiens

Other Aliases: GCP170, MEA-2

Other Designations: Golgi membrane associated protein; Golgi peripheral membrane protein; Golgin subfamily A member 3; SY2/SY10 protein; golgi autoantigen, golgin subfamily a, 3; golgi complex-associated protein of 170 kDa; golgin-160; golgin-165; male enhanced antigen-2

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_005895.3

LOCUS NM_005895

ACCESSION NM_005895

```
      1 ggcctgggcg cgtccctgca gcgtggcggg acggccccgt tccagtcacc
cccgcctcgc
```

```
       61 tgcggtggcc tcgggcctgg gcgcccgcct tcagctgcgg cggagctggc
tctgtaaatg
      121 ccggtgcccg cgagccctcc tgaatgcttg tctgcgcccg acgagcgcgg
cctgtcccga
      181 agctgtccac tgccaccact cgggcagtgc ttgctttagc ctggcccttt
gcagctgaaa
      241 ggcgtgacat ggtgtcgggt ggttcgtggg aagtcggggt ttcaggagtc
cgtgtacttc
      301 cttgtttgtc tttgtcgctg gccgacttgt cttcattcca ggtggccaga
gcgagtgggg
      361 ccgggcgttg tcacgggtat catgatatta gctggtttga catcaagtca
tttgtgagtc
      421 atcagatctt ctcctgaaaa tgggagacac agtagggccc ctcccaggag
ctcttggctg
      481 ttgctgatgg cagaagccaa gcttgtccaa ggttcacttg tagcccctca
gcgtcagctc
      541 agctggtgtc gtcctgacca tggacggcgc gtcggccgag caagatggcc
tccaggagga
      601 cagatcccac agtggcccct cgtctctccc cgaggcccca ctgaagcccc
cgggcccact
      661 ggtgccacct gaccagcagg acaaagtcca gtgtgccgag gtaaacagag
catccacgga
      721 aggggaaagc ccggatggac ctggccaggg aggcctctgt cagaacgggc
caacgccacc
      781 cttcccagac cctccgtcgt ctctcgatcc caccacaagc ccagtgggcc
ctgatgcctc
      841 tccaggtgtg gctggtttcc atgacaacct aaggaagtct cagggaacta
gtgctgaggg
      901 cagtgttaga aaagaagctt tgcagtctct cagactcagt cttcctatgc
aagaaacgca
      961 actgtgctct acagattctc ccctgcccct ggagaaggag gagcaggtcc
gacttcaggc
     1021 tcggaagtgg ctggaagagc agctcaaaca gtacagggtg aagcgccagc
aggagaggtc
     1081 cagtcaacct gcaaccaaaa cgagactttt tagcacgctt gatcctgagc
tcatgttaaa
     1141 cccagaaaac ttaccaaggg ccagtaccct ggctatgaca aaagaatatt
ccttcctgcg
     1201 caccagtgtc cctcgggggc ctaaggtggg cagcctgggg cttccggcac
atcctaggga
     1261 gaaaaaaact tccaaatcaa gcaaaatccg gtctctggcc gattacagaa
ctgaagattc
     1321 aaatgcgggg aattctgggg gaaatgtccc ggctcccgat tctaccaagg
gttccctgaa
     1381 gcagaacaga agcagtgcgg cgtccgttgt gtctgagatc agcctgtccc
ccgacactga
     1441 cgaccgtctg gagaacacct ccctggctgg agacagcgtg tctgaggtgg
atggaaatga
     1501 cagcgacagc tcatcgtaca gcagcgcctc cacccgaggg acctatggca
ttctgtcgaa
     1561 gacagtgggc acgcaggaca ccccctatat ggtcaacggc caggagattc
ctgcggatac
     1621 cctgggccag ttcccctcca ttaaggacgt cctccaggcc gcagccgctg
agcaccaaga
     1681 ccaggggcag gaggtcaacg gggaggtgcg gagtcggaga gacagcatct
gcagcagcgt
     1741 gtccttggag agctctgcag cagaaacaca ggaggagatg ctgcaggtgc
tcaaagagaa
     1801 aatgcgactc gaaggacagc tggaagcctt gtcactggag gcgagtcagg
cacttaaaga
```

1861 gaaggctgag ctgcaggccc agctggccgc cctcagcacg aagctgcagg
cgcaggtgga
1921 gtgcagccac agcagccagc agcggcagga ttcgctgagc tcggaggtgg
acaccctgaa
1981 gcagtcgtgc tgggacctgg agcgagccat gactgacctg cagaacatgc
tggaggcaaa
2041 aaatgccagc ctggcgtcgt ccaacaacga cttgcaggtg gccgaggagc
agtaccagag
2101 gcttatggcc aaggtagagg acatgcagag gagcatgctc agcaaggaca
acacagtgca
2161 cgacctgcga cagcagatga cagccttgca gagccagctt cagcaggtgc
agctggagcg
2221 gacgacgctg accagcaagc tgaaggcgtc gcaggcggag atctcgtccc
tgcagagtgt
2281 ccggcagtgg taccagcagc agctcgccct ggcacaggag gcccgcgtca
ggctgcaggg
2341 tgagatggcc cacatccagg ttggacagat gacccaggca ggtctcctgg
agcacctgaa
2401 actcgagaat gtgtccctgt cccagcagct gacggaaact cagcacaggt
ccatgaagga
2461 gaaggggcgc atcgcggcac agctgcaggg cattgaggct gacatgttgg
atcaggaagc
2521 agccttcatg cagattcagg aggcaaagac gatggtggag gaggaccttc
agaggaggct
2581 ggaagagttt gaaggtgaga gggagcggct gcagaggatg gcggactcgg
cggcatccct
2641 ggagcagcag ctggagcagg tgaagttgac tttactccag cgagaccagc
agcttgaggc
2701 tttgcagcag gagcacctgg acctgatgaa acagctcacc ttgactcagg
aggctctgca
2761 gagcagggag cagtccctcg atgccctgca gacacactac gatgagctgc
aggccaggct
2821 gggggagctg cagggcgagg ccgcctccag ggaggacacg atctgcctcc
tgcagaacga
2881 gaagatcatc ttggaggcgg ctttgcaggc ggccaagagt ggcaaggagg
agcttgacag
2941 aggagcaaga cgcttggaag aaggtaccga ggaaacgtcg gaaactttag
agaagttaag
3001 agaagaatta gctatcaaat ccggccaggt ggaacacctg cagcaggaga
ctgctgctct
3061 gaaaaagcaa atgcaaaaaa taaaggaaca gtttctccaa caaaaggtga
tggtggaggc
3121 ctaccggcgc gacgccacct ccaaagacca gctcatcagt gagctgaaag
ccaccaggaa
3181 gaggctggac tcggagctga aggagctgcg gcaggagctg atgcaagtgc
acggggagaa
3241 gcggactgcc gaggcggagc tctcgcgcct gcacagagag gtggcccagg
tccgtcagca
3301 catggcggac cttgaagggc atctccagtc ggcgcagaag gagcgagacg
agatggaaac
3361 acacttgcag tcgttgcagt tcgataagga gcagatggtc gcggtcacag
aggccaatga
3421 ggcgctgaag aaacaaatcg aagagttgca gcaagaggcc cggaaggcca
tcacggaaca
3481 gaagcagaag atgaggcggc tgggctcaga cttgaccagc gcccagaagg
agatgaagac
3541 caaacataag gcctacgaga acgccgtggg catcctcagc cgccgcctgc
aggaggccct
3601 cgcggccaag gaggctgcgg acgcggagct gggccagctc cgagcccagg
gtggcagcag

3661 tgacagcagc ctggctctac atgaaaggat ccaggccctg gaggcggagc
tgcaggctgt
3721 cagtcatagc aagacgctgc tggaaaagga actgcaggag gtcatagcgc
tgaccagcca
3781 ggagctggag gagtcccggg agaaggtgct ggagctggag gacgagcttc
aagaatccag
3841 aggctttagg aagaagataa aacgccttga ggagtcaaac aagaagttgg
ctcttgaatt
3901 agagcacgag aaagggaagc ttacgggcct cggtcagtcc aacgcagctc
tgcgggaaca
3961 caacagcatc ctagaaacag ctttggccaa gagggaggca gacctagtcc
agttgaacct
4021 tcaggtgcag gcagttttgc agcgcaaaga agaggaggat cgccagatga
agcatcttgt
4081 ccaggccctg caggcctcac tagagaagga gaaggagaag gtgaacagcc
tcaaggagca
4141 ggtggctgct gccaaggtgg aagccgggca taaccgccgc cacttcaagg
cggcctcctt
4201 ggagctgagt gaggtgaaga aggagctgca ggccaaggaa cacctggtgc
agaagctgca
4261 ggccgaggcc gacgaccttc agattcggga ggggaaacat tcccaggaga
tagcacagtt
4321 ccaagcagag ctggccgagg cccgggcaca gctccagctc ctgcagaagc
agctggacga
4381 gcagctcagc aaacagcccg tgggaaacca agagatggaa aatctcaaat
gggaggtgga
4441 tcagaaagaa agagaaatcc agtccttgaa gcagcagctg gacttgacgg
agcagcaggg
4501 caggaaggaa ctggaagggc tacagcagct gctgcagaac gtcaagtctg
agttggagat
4561 ggcccaggaa gacctgtcca tgacccagaa ggataaattt atgctccagg
caaaagtgtc
4621 ggagctgaag aacaacatga agaccctgct ccagcagaac cagcagctca
agctggacct
4681 acgccgcggc gcggccaaga cgagaaagga gccgaaaggc gaggccagct
cttccaaccc
4741 tgccacgccc atcaagatcc cggactgccc agttcccgcc tcgctgctgg
aggagctgct
4801 gagaccaccg cccgccgtga gcaaggagcc cctcaagaac ctgaacagct
gcctccagca
4861 gctcaagcag gagatggaca gcctgcagcg ccagatggag gagcacgccc
tgacggtgca
4921 cgagtctctg tcctcgtgga cgccgctgga gccagccact gccagccctg
tgcccccggg
4981 gggtcacgcc ggcccacgcg gcgacccaca gagacacagt cagagcaggg
cttccaaaga
5041 agggccggga gagtgactgc tgtggactcg cctccgtgcg ccgctgcccc
agaaggctct
5101 tatcaatgtt atttatttga ttgtgtggtc gatgtttttc taagacatga
aatttaagtt
5161 ttgttttgcc tttaacaaga agtaaaatat atagcagaat gagagccaag
gactagaaaa
5221 acattcgaag atcacaatta gcttttcaca tggaatgacc aactcttaaa
agcctgatag
5281 gctctcggcg aggagctttg aacgtgtctg aagggttact tgtaggtcgt
ggcttctgag
5341 cggccaccga tgctgctctc tgcgggtgac agggagaggc tgcgtaactg
ggagcagctg
5401 tgtgacaggg tctgcggcac cgcgcctggc caggccggct gcagtttctc
acttccctgt

5461 tccattcagt aagagcttta cttttccgca gaaatgaaat tttatctgta
cctttggctt
5521 tttacttgtt tttttggata gccatcccac cataggatgt gtacatagat
actgaatatc
5581 ataatccaat ctttgttttt ttttttttt ttttttgag acagagtctc
gctttgttgc
5641 ccaggctgga gtgcagtggc acactctccg ctcactgcaa gctccgcctc
ccaggttcat
5701 gcgattctcc tgcctcagcc tctcgagtag ctgggattac aggcgtgcgc
cactatgcca
5761 ggctaatgtt tgtatttta gtagcaatgg ggtttcacca tgttggccag
gatggtctcg
5821 atctcctgac ctcaagtgat ctgcccatct cagcctccca aagtgctggg
attacaggcg
5881 tgagcccctg cgcccggcct gtcacccagt ctttaagaag catatgctca
tgttattgaa
5941 gaagaaccta cttattattg attgcctttt gaaaatttgt tgggaataat
ttacctgcag
6001 gatttaggga tagtcagaaa attctaagaa atataattat tttatttacc
ttctaaagcc
6061 aaatattctt acacagaaag gtcctctgtt gttctggttt tactttgttg
ctgaggatct
6121 ttccttcctg ctggtctctt cctctcaggc cactggccct gtgtgattcc
accgtggctg
6181 gccactggga aggggcagct tggacccttg gtcaggcctg acggccatca
ggaggcacaa
6241 ggacactgag gccccatatc tgatctgacc tttgggggggg cacagggaga
ggccggtgga
6301 ggaggaggag gagagcagac caggggctcc ctgcagcgac tcccgcggtt
tcccctggag
6361 tcagccaggt gtaggtcgca ggcggtaaca aacctcacac tcctgttccc
caagtgaaaa
6421 tctttaccat tgtctgtggg agcgcctgta ctcgtgtgta ggagcacctg
tacttctgca
6481 gtcatcgaga agtcctggat cttttgtggt tacaccagca tcatgtggca
agcagaggcg
6541 acttccggaa gagacaggca ggcaccgtga ggaaggtggc tgtgctctcc
caggtgtctc
6601 agagacagat gccttattta aaatcagcac gacatgtgtg agatcttctg
tttcctaccc
6661 caaatcctga aaccctgcag acactggctg actgggagag gtggggtctg
taagttgtcc
6721 cctagtttgc taagaaaatc taaaataata tttattatat gagttaggag
agagagaatg
6781 ggtccgcgtg gcctcctctg cagatgtact ggtctgaaat gaggttctga
gtcactggcc
6841 aggccagatg tgctcatgtc ggtgtctggt gtctgttttg tggagaaaac
agtatggtgt
6901 gttttaagct atttgtgttc tgttgtaata tacttttaga aggttaattg
gtaaggttaa
6961 ggtagcatta accacaaaga tgtttggtat ttaaaaaata ttctctagca
aatattggaa
7021 tttccaaaat atatcatttg tacagggtta attttgaaat aatacttgaa
aatttcatta
7081 taaatatatc ctacttttta tcttaagttg aagatgttat ttactaaatt
gttcttgtac
7141 cattagaaaa aaaaatacgg caatttacgt tcttatttat tttggctgta
ctacccctttt
7201 gttttaattt taaaatcaag aaatcgggcc gggcgcggtg gctcatgcct
gtaatcccag

7261 cactttggga ggccgaggcg ggtggatcac ctgaggtcaa gaggtccaga ccatcctggc
7321 caacatggca aaaccccgtc tttactaaac atacaaaaat tagccgggtg tcgtagtgcg
7381 cacctataat cccagctact tgggaggctg aggcaggaga atcacttgaa cccaggaggc
7441 ggagcttgca gtgagccgag atcgcgccac tgccctccag cctgggcaac agagcgagac
7501 tccgtctcaa aaataaataa atgattttaa aaaatctaaa atcgagaaat cacacattca
7561 gtggggagcg acttctcctt gcttatggga agtcctcaag tgagtgatgt tcaccatgta
7621 ttttttttttc tcttaggaca gactaattct gaaaataccg aaggaaaagt agctctatgt
7681 tctcaccccg gttttcctgc gtgtgtgccc ttgggtgcga tgcctccccc agcgctctgt
7741 ggtcgccggt gccagggccc cctctggttt ggcagggcct ggctgccttt gctccctgca
7801 gtgagtcttt tggtgttttc atgcacggct tgtgcttctg gatctgaggc ctctcgtgtt
7861 cacgcggaca cttccttcct taagaagacg cctaaaagag gaagttggaa ttttttttttt
7921 tttttttga gacagagtct cgctctgtcg cccaggctgg agtgcagtgg cgtgatctct
7981 gctcactgca agctccgcca tctgggttca agcgattctc ctgcctcatt ctccccagta
8041 gctgggatta caggtgcccg ccaccacacc agcctaattt ttgtattttt agaggggtgg
8101 agttccacca tgttggccag gctggtcttg aactcttgac ctcaggtgat cctgagcctc
8161 agcctcccaa agtgctggaa ttataggcgt gaaccaccgc ccccggctgc agttggattt
8221 ttaaattgct tttttttatt gttgaggttt ttttatctcc aagggactct cccggcactt
8281 ctaccttcca gagttacttc agtgcataaa gtttgaatta ttttgttctt gtgggcagaa
8341 gtgggaatga tggaatatcc tcacggaaaa ggcagtgaag ttgggagtac tgcttacaaa
8401 acagggtcac cagtgcatta tgtggcgtgt tcatccccac gccgtgtgtc acgggctagg
8461 gcggcgtgtt catccccaca ccgtgtgtca caacaggcta gggcacttca cgatgtcact
8521 acttgttttt ctgatgttcc aaaaacaacg taacttggtt ttcatgtgtt tttccgtggt
8581 atatgtgaga ttgatgctac gggtcttacg gactcacacc cgttcccact ctctgcaata
8641 tggatcaggc agtgtttctg ataggatgtg aaatggactc tcctcgggtg ggtccagcag
8701 gggccctgcc caccagaaca cagtccgtgc tgtgctgcgc taaggagctg gccctcaact
8761 ctccttggtg cagggttccc acaaccgagt tctagttccc tgaggtcttt aaaaacaaaa
8821 acagaatgtt gtacgtgaag attctaggag gggagggacc agcaaatctg agagaaccgt
8881 cctggggcct cccttcgagg agccctctga tgtgaggagg gacttgagtt gagtgacgct
8941 gtggtgtgag gtgttctgag ctcactgacc ggaaggtcca ggtgaatctc gtcataagtg
9001 atctcaggct ctcacaggat ccggagggaa atgtgttaga gggtctggaa aattcagtgc

```
     9061 ttttgagtta cttgttttta ttaaaaattt cctcacaaaa gagagtcctc
aagttgtggc
     9121 tgttcttggg aaaggggtca ccgtgtctga caaagtgtaa ctttaaaaag
cacgttgatt
     9181 ttttacaaat gtaagtgtgc ttgggaattc cttaaatttt gtgcaataaa
ctatttttg
      9241 gtaaagattt tc
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_005886.2

LOCUS NP_005886

ACCESSION NP_005886

```
      1 mdgasaeqdg lqedrshsgp sslpeaplkp pgplvppdqq dkvqcaevnr
astegespdg
       61 pgqgglcqng ptppfpdpps sldpttspvg pdaspgvagf hdnlrksqgt
saegsvrkea
      121 lqslrlslpm qetqlcstds plplekeeqv rlqarkwlee qlkqyrvkrq
qerssqpatk
      181 trlfstldpe lmlnpenlpr astlamtkey sflrtsvprg pkvgslglpa
hprekktsks
      241 skirsladyr tedsnagnsg gnvpapdstk gslkqnrssa asvvseisls
pdtddrlent
      301 slagdsvsev dgndsdsssy ssastrgtyg ilsktvgtqd tpymvngqei
padtlgqfps
      361 ikdvlqaaaa ehqdqgqevn gevrsrrdsi cssvslessa aetqeemlqv
lkekmrlegq
      421 lealsleasq alkekaelqa qlaalstklq aqvecshssq qrqdslssev
dtlkqscwdl
      481 eramtdlqnm leaknaslas snndlqvaee qyqrlmakve dmqrsmlskd
ntvhdlrqqm
      541 talqsqlqqv qlerttltsk lkasqaeiss lqsvrqwyqq qlalaqearv
rlqqgemahiq
      601 vgqmtqagll ehlklenvsl sqqltetqhr smkekgriaa qlqgieadml
dqeaafmqiq
      661 eaktmveedl qrrleefege rerlqrmads aasleqqleq vkltllqrdq
qlealqqehl
      721 dlmkqltltq ealqsreqsl dalqthydel qarlgelqge aasredticl
lqnekiilea
      781 alqaaksgke eldrgarrle egteetsetl eklreelaik sgqvehlqqe
taalkkqmqk
      841 ikeqflqqkv mveayrrdat skdqliselk atrkrldsel kelrqelmqv
hgekrtaeae
      901 lsrlhrevaq vrqhmadleg hlqsaqkerd emethlqslq fdkeqmvavt
eanealkkqi
      961 eelqqearka iteqkqkmrr lgsdltsaqk emktkhkaye navgilsrrl
qealaakeaa
     1021 daelgqlraq ggssdsslal heriqaleae lqavshsktl lekelqevia
ltsqeleesr
     1081 ekvleledel qesrgfrkki krleesnkkl alelehekgk ltglgqsnaa
lrehnsilet
     1141 alakreadlv qlnlqvqavl qrkeeedrqm khlvqalqas lekekekvns
lkeqvaaakv
     1201 eaghnrrhfk aaslelsevk kelqakehlv qklqaeaddl qiregkhsqe
iaqfqaelae
     1261 araqlqllqk qldeqlskqp vgnqemenlk wevdqkerei qslkqqldlt
eqqgrkeleg
```

```
     1321 lqqllqnvks elemaqedls mtqkdkfmlq akvselknnm ktllqqnqql
kldlrrgaak
     1381 trkepkgeas ssnpatpiki pdcpvpasll eellrpppav skeplknlns
clqqlkqemd
     1441 slqrqmeeha ltvheslssw tplepatasp vppgghagpr gdpqrhsqsr
askegpge
```

PH4B

Official Symbol: P4HB

Official Name: prolyl 4-hydroxylase, beta polypeptide

Gene ID: 5034

Organism: Homo sapiens

Other Aliases: DSI, ERBA2L, GIT, P4Hbeta, PDI, PDIA1, PHDB, PO4DB, PO4HB, PROHB

Other Designations: cellular thyroid hormone-binding protein; collagen prolyl 4-hydroxylase beta; glutathione-insulin transhydrogenase; p55; procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide; prolyl 4-hydroxylase subunit beta; protein disulfide isomerase family A, member 1; protein disulfide isomerase-associated 1; protein disulfide isomerase/oxidoreductase; protein disulfide-isomerase; protocollagen hydroxylase; thyroid hormone-binding protein p55

Nucleotide sequence:

NCBI Reference Sequence: NM_000918.3

LOCUS NM_000918

ACCESSION NM_000918

```
1 gagcctcgaa gtccgccggc caatcgaagg cgggccccag cggcgcgtgc gcgccgcggc
       61 cagcgcgcgc gggcgggggg gcaggcgcgc cccggaccca ggatttataa
aggcgaggcc
      121 gggaccggcg cgcgctctcg tcgcccccgc tgtcccggcg gcgccaaccg
aagcgccccg
      181 cctgatccgt gtccgacatg ctgcgccgcg ctctgctgtg cctggccgtg
gccgccctgg
      241 tgcgcgccga cgcccccgag gaggaggacc acgtcctggt gctgcggaaa
agcaacttcg
      301 cggaggcgct ggcggcccac aagtacctgc tggtggagtt ctatgcccct
tggtgtggcc
      361 actgcaaggc tctggccccc gagtatgcca aagccgctgg gaagctgaag
gcagaaggtt
      421 ccgagatcag gttggccaag gtggacgcca cggaggagtc tgacctggcc
cagcagtacg
      481 gcgtgcgcgg ctatcccacc atcaagttct tcaggaatgg agacacggct
tcccccaagg
      541 aatatacagc tggcagagag gctgatgaca tcgtgaactg gctgaagaag
cgcacgggcc
```

601 cggctgccac caccctgcct gacggcgcag ctgcagagtc cttggtggag
tccagcgagg
661 tggctgtcat cggcttcttc aaggacgtgg agtcggactc tgccaagcag
tttttgcagg
721 cagcagaggc catcgatgac ataccatttg ggatcacttc caacagtgac
gtgttctcca
781 aataccagct cgacaaagat ggggttgtcc tctttaagaa gtttgatgaa
ggccggaaca
841 actttgaagg ggaggtcacc aaggagaacc tgctggactt tatcaaacac
aaccagctgc
901 cccttgtcat cgagttcacc gagcagacag ccccgaagat ttttggaggt
gaaatcaaga
961 ctcacatcct gctgttcttg cccaagagtg tgtctgacta tgacggcaaa
ctgagcaact
1021 tcaaaacagc agccgagagc ttcaagggca agatcctgtt catcttcatc
gacagcgacc
1081 acaccgacaa ccagcgcatc ctcgagttct ttggcctgaa gaaggaagag
tgcccggccg
1141 tgcgcctcat caccctggag gaggagatga ccaagtacaa gcccgaatcg
gaggagctga
1201 cggcagagag gatcacagag ttctgccacc gcttcctgga gggcaaaatc
aagccccacc
1261 tgatgagcca ggagctgccg gaggactggg acaagcagcc tgtcaaggtg
cttgttggga
1321 agaactttga agacgtggct tttgatgaga aaaaaaacgt ctttgtggag
ttctatgccc
1381 catggtgtgg tcactgcaaa cagttggctc ccatttggga taaactggga
gagacgtaca
1441 aggaccatga gaacatcgtc atcgccaaga tggactcgac tgccaacgag
gtggaggccg
1501 tcaaagtgca cagcttcccc acactcaagt tctttcctgc cagtgccgac
aggacggtca
1561 ttgattacaa cggggaacgc acgctggatg gttttaagaa attcctggag
agcggtggcc
1621 aggatggggc aggggatgat gacgatctcg aggacctgga agaagcagag
gagccagaca
1681 tggaggaaga cgatgatcag aaagctgtga aagatgaact gtaatacgca
aagccagacc
1741 cgggcgctgc cgagacccct cgggggctgc acacccagca gcagcgcacg
cctccgaagc
1801 ctgcggcctc gcttgaagga gggcgtcgcc ggaaacccag ggaacctctc
tgaagtgaca
1861 cctcaccccct acacaccgtc cgttcacccc cgtctcttcc ttctgctttt
cggtttttgg
1921 aaagggatcc atctccaggc agcccaccct ggtggggctt gtttcctgaa
accatgatgt
1981 acttttttcat acatgagtct gtccagagtg cttgctaccg tgttcggagt
ctcgctgcct
2041 ccctcccgcg ggaggtttct cctcttttg aaaattccgt ctgtgggatt
tttagacatt
2101 tttcgacatc agggtatttg ttccaccttg gccaggcctc ctcggagaag
cttgtccccc
2161 gtgtgggagg gacggagccg gactggacat ggtcactcag taccgcctgc
agtgtcgcca
2221 tgactgatca tggctcttgc atttttgggt aaatggagac ttccggatcc
tgtcagggtg
2281 tcccccatgc ctggaagagg agctggtggc tgccagccct ggggcccggc
acaggcctgg
2341 gccttcccct tccctcaagc cagggctcct cctcctgtcg tgggctcatt
gtgaccactg

```
     2401 gcctctctac agcacggcct gtggcctgtt caaggcagaa ccacgaccct
tgactcccgg
     2461 gtggggaggt ggccaaggat gctggagctg aatcagacgc tgacagttct
tcaggcattt
     2521 ctatttcaca atcgaattga acacattggc caaataaagt tgaaatttta
ccacctgtaa
      2581 aaaaaaaaaa aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_000909.2

LOCUS NP_000909

ACCESSION NP_000909

```
      1 mlrrallcla vaalvradap eeedhvlvlr ksnfaealaa hkyllvefya
pwcghckala
       61 peyakaagkl kaegseirla kvdateesdl aqqygvrgyp tikffrngdt
aspkeytagr
      121 eaddivnwlk krtgpaattl pdgaaaeslv essevavigf fkdvesdsak
qflqaaeaid
      181 dipfgitsns dvfskyqldk dgvvlfkkfd egrnnfegev tkenlldfik
hnqlplvief
      241 teqtapkifg geikthillf lpksvsdydg klsnfktaae sfkgkilfif
idsdhtdnqr
      301 ileffglkke ecpavrlitl eeemtkykpe seeltaerit efchrflegk
ikphlmsqel
      361 pedwdkqpvk vlvgknfedv afdekknvfv efyapwcghc kqlapiwdkl
getykdheni
      421 viakmdstan eveavkvhsf ptlkffpasa drtvidynge rtldgfkkfl
esggqdgagd
      481 dddledleea eepdmeeddd qkavkdel
```

HSPA1A

Official Symbol: **HSPA1A**

Official Name: heat shock 70kDa protein 1A

Gene ID: 3303

Organism: Homo sapiens

Other Aliases: DAQB-147D11.1, HSP70-1, HSP70-1A, HSP70I, HSP72, HSPA1

Other Designations: HSP70-1/HSP70-2; HSP70.1/HSP70.2; dnaK-type molecular chaperone HSP70-1; heat shock 70 kDa protein 1/2; heat shock 70 kDa protein 1A/1B; heat shock 70kD protein 1A; heat shock-induced protein

Nucleotide sequence:

NCBI Reference Sequence: NM_005345.5

LOCUS NM_005345

ACCESSION NM_005345

```
      1 ataaaagccc aggggcaagc ggtccggata acggctagcc tgaggagctg
ctgcgacagt
       61 ccactacctt tttcgagagt gactcccgtt gtcccaaggc ttcccagagc
gaacctgtgc
      121 ggctgcaggc accggcgcgt cgagtttccg gcgtccggaa ggaccgagct
cttctcgcgg
      181 atccagtgtt ccgtttccag cccccaatct cagagcggag ccgacagaga
gcagggaacc
      241 ggcatggcca aagccgcggc gatcggcatc gacctgggca ccacctactc
ctgcgtgggg
      301 gtgttccaac acggcaaggt ggagatcatc gccaacgacc agggcaaccg
caccaccccc
      361 agctacgtgg ccttcacgga caccgagcgg ctcatcgggg atgcggccaa
gaaccaggtg
      421 gcgctgaacc cgcagaacac cgtgtttgac gcgaagcggc tgattggccg
caagttcggc
      481 gacccggtgg tgcagtcgga catgaagcac tggcctttcc aggtgatcaa
cgacggagac
      541 aagcccaagg tgcaggtgag ctacaagggg gagaccaagg cattctaccc
cgaggagatc
      601 tcgtccatgg tgctgaccaa gatgaaggag atcgccgagg cgtacctggg
ctacccggtg
      661 accaacgcgg tgatcaccgt gccggcctac ttcaacgact cgcagcgcca
ggccaccaag
      721 gatgcgggtg tgatcgcggg gctcaacgtg ctgcggatca tcaacgagcc
cacggccgcc
      781 gccatcgcct acggcctgga cagaacgggc aaggggggagc gcaacgtgct
catctttgac
      841 ctgggcgggg gcaccttcga cgtgtccatc ctgacgatcg acgacggcat
cttcgaggtg
      901 aaggccacgg ccggggacac ccacctgggt ggggaggact ttgacaacag
gctggtgaac
      961 cacttcgtgg aggagttcaa gagaaaacac aagaaggaca tcagccagaa
caagcgagcc
     1021 gtgaggcggc tgcgcaccgc ctgcgagagg gccaagagga ccctgtcgtc
cagcacccag
     1081 gccagcctgg agatcgactc cctgtttgag ggcatcgact tctacacgtc
catcaccagg
     1141 gcgaggttcg aggagctgtg ctccgacctg ttccgaagca ccctggagcc
cgtggagaag
     1201 gctctgcgcg acgccaagct ggacaaggcc cagattcacg acctggtcct
ggtcgggggc
     1261 tccacccgca tccccaaggt gcagaagctg ctgcaggact tcttcaacgg
gcgcgacctg
     1321 aacaagagca tcaaccccga cgaggctgtg gcctacgggg cggcggtgca
ggcggccatc
     1381 ctgatggggg acaagtccga gaacgtgcag gacctgctgc tgctggacgt
ggctcccctg
```

```
     1441 tcgctggggc tggagacggc cggaggcgtg atgactgccc tgatcaagcg
caactccacc
     1501 atccccacca agcagacgca gatcttcacc acctactccg acaaccaacc
cggggtgctg
     1561 atccaggtgt acgagggcga gagggccatg acgaaagaca acaatctgtt
ggggcgcttc
     1621 gagctgagcg gcatccctcc ggcccccagg ggcgtgcccc agatcgaggt
gaccttcgac
     1681 atcgatgcca acggcatcct gaacgtcacg gccacggaca agagcaccgg
caaggccaac
     1741 aagatcacca tcaccaacga caagggccgc ctgagcaagg aggagatcga
gcgcatggtg
     1801 caggaggcgg agaagtacaa agcggaggac gaggtgcagc gcgagagggt
gtcagccaag
     1861 aacgccctgg agtcctacgc cttcaacatg aagagcgccg tggaggatga
ggggctcaag
     1921 ggcaagatca gcgaggcgga caagaagaag gtgctggaca agtgtcaaga
ggtcatctcg
     1981 tggctggacg ccaacacctt ggccgagaag gacgagtttg agcacaagag
gaaggagctg
     2041 gagcaggtgt gtaaccccat catcagcgga ctgtaccagg gtgccggtgg
tcccgggcct
     2101 gggggcttcg gggctcaggg tcccaaggga gggtctgggt caggccccac
cattgaggag
     2161 gtagattagg ggcctttcca agattgctgt ttttgttttg gagcttcaag
actttgcatt
     2221 tcctagtatt tctgtttgtc agttctcaat ttcctgtgtt tgcaatgttg
aaatttttg
     2281 gtgaagtact gaacttgctt tttttccggt ttctacatgc agagatgaat
ttatactgcc
     2341 atcttacgac tatttcttct ttttaataca cttaactcag gccatttttt
aagttggtta
     2401 cttcaaagta aataaacttt aaaattcaaa aaaaaaaaaa aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_005336.3

LOCUS NP_005336

ACCESSION NP_005336

```
        1 makaaaigid lgttyscvgv fqhgkveiia ndqgnrttps yvaftdterl
igdaaknqva
       61 lnpqntvfda krligrkfgd pvvqsdmkhw pfqvindgdk pkvqvsykge
tkafypeeis
      121 smvltkmkei aeaylgypvt navitvpayf ndsqrqatkd agviaglnvl
riineptaaa
      181 iaygldrtgk gernvlifdl gggtfdvsil tiddgifevk atagdthlgg
edfdnrlvnh
      241 fveefkrkhk kdisqnkrav rrlrtacera krtlssstqa sleidslfeg
idfytsitra
      301 rfeelcsdlf rstlepveka lrdakldkaq ihdlvlvggs tripkvqkll
qdffngrdln
      361 ksinpdeava ygaavqaail mgdksenvqd llllvapls lgletaggvm
talikrnsti
      421 ptkqtqiftt ysdnqpgvli qvyegeramt kdnnllgrfe lsgippaprg
vpqievtfdi
```

```
      481 dangilnvta tdkstgkank ititndkgrl skeeiermvq eaekykaede
vqrervsakn
      541 alesyafnmk savedeglkg kiseadkkkv ldkcqevisw ldantlaekd
efehkrkele
      601 qvcnpiisgl yqgaggpgpg gfgaqgpkgg sgsgptieev d
```

Gene

Official Symbol: **HNRNPD**

Official Name: heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)

Gene ID: 3184

Organism: Homo sapiens

Other Aliases: AUF1, AUF1A, HNRPD, P37, hnRNPD0

Other Designations: ARE-binding protein AUFI, type A; heterogeneous nuclear ribonucleoprotein D0; hnRNP D0

Nucleotide sequence: **ISOFORM D**

NCBI Reference Sequence: NM_001003810.1

LOCUS NM_001003810

ACCESSION NM_001003810

```
        1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
      301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
      361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
      421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac
cgcgtctgga
      481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag
taagaacgag
      541 gaggatgaag ggaaaatgtt tataggaggc cttagctggg acactacaaa
gaaagatctg
      601 aaggactact tttccaaatt tggtgaagtt gtagactgca ctctgaagtt
agatcctatc
      661 acagggcgat caaggggttt tggctttgtg ctatttaaag aatcggagag
tgtagataag
```

```
      721 gtcatggatc aaaaagaaca taaattgaat gggaaggtga ttgatcctaa
aagggccaaa
      781 gccatgaaaa caaaagagcc ggttaaaaaa atttttgttg gtggcctttc
tccagataca
      841 cctgaagaga aaataaggga gtactttggt ggttttggtg aggtggaatc
catagagctc
      901 cccatggaca acaagaccaa taagaggcgt gggttctgct ttattacctt
taaggaagaa
      961 gaaccagtga agaagataat ggaaaagaaa taccacaatg ttggtcttag
taaatgtgaa
     1021 ataaaagtag ccatgtcgaa ggaacaatat cagcaacagc aacagtgggg
atctagagga
     1081 ggatttgcag gaagagctcg tggaagaggt ggtgaccagc agagtggtta
tgggaaggta
     1141 tccaggcgag gtggtcatca aaatagctac aaaccatact aaattattcc
atttgcaact
     1201 tatccccaac aggtggtgaa gcagtatttt ccaatttgaa gattcatttg
aaggtggctc
     1261 ctgccacctg ctaatagcag ttcaaactaa atttttgta tcaagtccct
gaatggaagt
     1321 atgacgttgg gtccctctga agtttaattc tgagttctca ttaaaagaaa
tttgctttca
     1381 ttgttttatt tcttaattgc tatgcttcag aatcaatttg tgttttatgc
cctttccccc
     1441 agtattgtag agcaagtctt gtgttaaaag cccagtgtga cagtgtcatg
atgtagtagt
     1501 gtcttactgg ttttttaata aatccttttg tataaaaatg tattggctct
tttatcatca
     1561 gaataggaaa aattgtcatg gattcaagtt attaaaagca taagtttgga
agacaggctt
     1621 gccgaaattg aggacatgat taaaattgca gtgaagtttg aaatgttttt
agcaaaatct
     1681 aatttttgcc ataatgtgtc ctccctgtcc aaattgggaa tgacttaatg
tcaatttgtt
     1741 tgttggttgt tttaataata cttccttatg tagccattaa gatttatatg
aatattttcc
     1801 caaatgccca gtttttgctt aatatgtatt gtgcttttta gaacaaatct
ggataaatgt
     1861 gcaaaagtac ccctttgcac agatagttaa tgttttatgc ttccattaaa
taaaaaggac
     1921 ttaaaatctg ttaattataa tagaaatgcg gctagttcag agagattttt
agagctgtgg
     1981 tggacttcat agatgaattc aagtgttgag ggaggattaa agaaatatat
accgtgttta
     2041 tgtgtgtgtg ctt
```

Protein sequence: **ISOFORM D**

NCBI Reference Sequence: NP_001003810.1

LOCUS NP_001003810

ACCESSION NP_001003810

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedegkm figglswdtt kkdlkdyfsk fgevvdctlk
ldpitgrsrg
```

121 fgfvlfkese svdkvmdqke hklngkvidp krakamktke pvkkifvggl spdtpeekir
181 eyfggfgeve sielpmdnkt nkrrgfcfit fkeeepvkki mekkyhnvgl skceikvams
241 keqyqqqqqw gsrggfagra rgrggdqqsg ygkvsrrggh qnsykpy

Nucleotide sequence: **ISOFORM C**

NCBI Reference Sequence: NM_002138.3

LOCUS NM_002138

ACCESSION NM_002138

1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga ggaggcgaga
61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg agtgtgcgcc
121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc tttgcagcca
181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc ggcagcggcg
241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct cggcggcagc
301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc ggcggcagcg
361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt ggcggcgaca
421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac cgcgtctgga
481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag taagaacgag
541 gaggatgaag gccattcaaa ctcctcccca cgacactctg aagcagcgac ggcacagcgg
601 gaagaatgga aaatgtttat aggaggcctt agctgggaca ctacaaagaa agatctgaag
661 gactactttt ccaaatttgg tgaagttgta gactgcactc tgaagttaga tcctatcaca
721 gggcgatcaa ggggttttgg ctttgtgcta tttaaagaat cggagagtgt agataaggtc
781 atggatcaaa aagaacataa attgaatggg aaggtgattg atcctaaaag ggccaaagcc
841 atgaaaacaa aagagccggt taaaaaaatt tttgttggtg gcctttctcc agatacacct
901 gaagagaaaa taagggagta ctttggtggt tttggtgagg tggaatccat agagctcccc
961 atggacaaca agaccaataa gaggcgtggg ttctgcttta ttacctttaa ggaagaagaa
1021 ccagtgaaga agataatgga aaagaaatac cacaatgttg gtcttagtaa atgtgaaata
1081 aaagtagcca tgtcgaagga acaatatcag caacagcaac agtggggatc tagaggagga
1141 tttgcaggaa gagctcgtgg aagaggtggt gaccagcaga gtggttatgg gaaggtatcc
1201 aggcgaggtg gtcatcaaaa tagctacaaa ccatactaaa ttattccatt tgcaacttat
1261 ccccaacagg tggtgaagca gtattttcca atttgaagat tcatttgaag gtggctcctg

```
     1321 ccacctgcta atagcagttc aaactaaatt ttttgtatca agtccctgaa
tggaagtatg
     1381 acgttgggtc cctctgaagt ttaattctga gttctcatta aaagaaattt
gctttcattg
     1441 ttttatttct taattgctat gcttcagaat caatttgtgt tttatgccct
ttcccccagt
     1501 attgtagagc aagtcttgtg ttaaaagccc agtgtgacag tgtcatgatg
tagtagtgtc
     1561 ttactggttt tttaataaat ccttttgtat aaaaatgtat tggctctttt
atcatcagaa
     1621 taggaaaaat tgtcatggat tcaagttatt aaaagcataa gtttggaaga
caggcttgcc
     1681 gaaattgagg acatgattaa aattgcagtg aagtttgaaa tgtttttagc
aaaatctaat
     1741 ttttgccata atgtgtcctc cctgtccaaa ttgggaatga cttaatgtca
atttgtttgt
     1801 tggttgtttt aataatactt ccttatgtag ccattaagat ttatatgaat
attttcccaa
     1861 atgcccagtt tttgcttaat atgtattgtg ctttttagaa caaatctgga
taaatgtgca
     1921 aaagtacccc tttgcacaga tagttaatgt tttatgcttc cattaaataa
aaaggactta
     1981 aaatctgtta attataatag aaatgcggct agttcagaga gatttttaga
gctgtggtgg
     2041 acttcataga tgaattcaag tgttgaggga ggattaaaga aatatatacc
gtgtttatgt
     2101 gtgtgtgctt
```

Protein sequence: **ISOFORM C**

NCBI Reference Sequence: NP_002129.2

LOCUS NP_002129

ACCESSION NP_002129

```
        1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedeghs nssprhseaa taqreewkmf igglswdttk
kdlkdyfskf
      121 gevvdctlkl dpitgrsrgf gfvlfkeses vdkvmdqkeh klngkvidpk
rakamktkep
      181 vkkifvggls pdtpeekire yfggfgeves ielpmdnktn krrgfcfitf
keeepvkkim
      241 ekkyhnvgls kceikvamsk eqyqqqqqwg srggfagrar grggdqqsgy
gkvsrrgghq
      301 nsykpy
```

Nucleotide sequence: **ISOFORM B**

NCBI Reference Sequence: NM_031369.2

LOCUS NM_031369

ACCESSION NM_031369

```
      1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
      301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
      361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
      421 caggggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac
cgcgtctgga
      481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag
taagaacgag
      541 gaggatgaag ggaaaatgtt tataggaggc cttagctggg acactacaaa
gaaagatctg
      601 aaggactact tttccaaatt tggtgaagtt gtagactgca ctctgaagtt
agatcctatc
      661 acagggcgat caaggggttt tggctttgtg ctatttaaag aatcggagag
tgtagataag
      721 gtcatggatc aaaaagaaca taaattgaat gggaaggtga ttgatcctaa
aagggccaaa
      781 gccatgaaaa caaaagagcc ggttaaaaaa atttttgttg gtggcctttc
tccagataca
      841 cctgaagaga aaataaggga gtactttggt ggttttggtg aggtggaatc
catagagctc
      901 cccatggaca acaagaccaa taagaggcgt gggttctgct ttattacctt
taaggaagaa
      961 gaaccagtga agaagataat ggaaaagaaa taccacaatg ttggtcttag
taaatgtgaa
     1021 ataaaagtag ccatgtcgaa ggaacaatat cagcaacagc aacagtgggg
atctagagga
     1081 ggatttgcag gaagagctcg tggaagaggt ggtggcccca gtcaaaactg
gaaccaggga
     1141 tatagtaact attggaatca aggctatggc aactatggat ataacagcca
aggttacggt
     1201 ggttatggag gatatgacta cactggttac aacaactact atggatatgg
tgattatagc
     1261 aaccagcaga gtggttatgg gaaggtatcc aggcgaggtg gtcatcaaaa
tagctacaaa
     1321 ccatactaaa ttattccatt tgcaacttat ccccaacagg tggtgaagca
gtattttcca
     1381 atttgaagat tcatttgaag gtggctcctg ccacctgcta atagcagttc
aaactaaatt
     1441 ttttgtatca agtccctgaa tggaagtatg acgttgggtc cctctgaagt
ttaattctga
     1501 gttctcatta aaagaaattt gctttcattg ttttatttct taattgctat
gcttcagaat
     1561 caatttgtgt tttatgccct ttcccccagt attgtagagc aagtcttgtg
ttaaaagccc
     1621 agtgtgacag tgtcatgatg tagtagtgtc ttactggttt tttaataaat
ccttttgtat
     1681 aaaaatgtat tggctctttt atcatcagaa taggaaaaat tgtcatggat
tcaagttatt
     1741 aaaagcataa gtttggaaga caggcttgcc gaaattgagg acatgattaa
aattgcagtg
```

```
     1801 aagtttgaaa tgtttttagc aaaatctaat ttttgccata atgtgtcctc
cctgtccaaa
     1861 ttgggaatga cttaatgtca atttgtttgt tggttgtttt aataatactt
ccttatgtag
     1921 ccattaagat ttatatgaat attttcccaa atgcccagtt tttgcttaat
atgtattgtg
     1981 ctttttagaa caaatctgga taaatgtgca aaagtacccc tttgcacaga
tagttaatgt
     2041 tttatgcttc cattaaataa aaaggactta aaatctgtta attataatag
aaatgcggct
     2101 agttcagaga gatttttaga gctgtggtgg acttcataga tgaattcaag
tgttgaggga
     2161 ggattaaaga aatatatacc gtgtttatgt gtgtgtgctt
```

Protein sequence: **ISOFORM B**

NCBI Reference Sequence: NP_112737.1

LOCUS NP_112737

ACCESSION NP_112737

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedegkm figglswdtt kkdlkdyfsk fgevvdctlk
ldpitgrsrg
      121 fgfvlfkese svdkvmdqke hklngkvidp krakamktke pvkkifvggl
spdtpeekir
      181 eyfggfgeve sielpmdnkt nkrrgfcfit fkeeepvkki mekkyhnvgl
skceikvams
      241 keqyqqqqqw gsrggfagra rgrgggpsqn wnqgysnywn qgygnygyns
qgyggyggyd
      301 ytgynnyygy gdysnqqsgy gkvsrrgghq nsykpy
```

Nucleotide sequence: **ISOFORM A**

NCBI Reference Sequence: NM_031370.2

LOCUS NM_031370

ACCESSION NM_031370

```
      1 cttccgtcgg ccattttagg tggtccgcgg cggcgccatt aaagcgagga
ggaggcgaga
       61 gcggccgccg ctggtgctta ttctttttta gtgcagcggg agagagcggg
agtgtgcgcc
      121 gcgcgagagt gggaggcgaa gggggcaggc cagggagagg cgcaggagcc
tttgcagcca
      181 cgcgcgcgcc ttccctgtct tgtgtgcttc gcgaggtaga gcgggcgcgc
ggcagcggcg
      241 gggattactt tgctgctagt ttcggttcgc ggcagcggcg ggtgtagtct
cggcggcagc
      301 ggcggagaca ctagcactat gtcggaggag cagttcggcg gggacggggc
ggcggcagcg
      361 gcaacggcgg cggtaggcgg ctcggcgggc gagcaggagg gagccatggt
ggcggcgaca
```

421 caggggcag cggcggcggc gggaagcgga gccgggaccg ggggcggaac cgcgtctgga
481 ggcaccgaag ggggcagcgc cgagtcggag ggggcgaaga ttgacgccag taagaacgag
541 gaggatgaag gccattcaaa ctcctcccca cgacactctg aagcagcgac ggcacagcgg
601 gaagaatgga aaatgtttat aggaggcctt agctgggaca ctacaaagaa agatctgaag
661 gactactttt ccaaatttgg tgaagttgta gactgcactc tgaagttaga tcctatcaca
721 gggcgatcaa ggggttttgg ctttgtgcta tttaaagaat cggagagtgt agataaggtc
781 atggatcaaa aagaacataa attgaatggg aaggtgattg atcctaaaag ggccaaagcc
841 atgaaaacaa aagagccggt taaaaaaatt tttgttggtg gcctttctcc agatacacct
901 gaagagaaaa taagggagta ctttggtggt tttggtgagg tggaatccat agagctcccc
961 atggacaaca agaccaataa gaggcgtggg ttctgcttta ttacctttaa ggaagaagaa
1021 ccagtgaaga agataatgga aaagaaatac cacaatgttg gtcttagtaa atgtgaaata
1081 aaagtagcca tgtcgaagga acaatatcag caacagcaac agtggggatc tagaggagga
1141 tttgcaggaa gagctcgtgg aagaggtggt ggccccagtc aaaactggaa ccagggatat
1201 agtaactatt ggaatcaagg ctatggcaac tatggatata acagccaagg ttacggtggt
1261 tatggaggat atgactacac tggttacaac aactactatg gatatggtga ttatagcaac
1321 cagcagagtg gttatgggaa ggtatccagg cgaggtggtc atcaaaatag ctacaaacca
1381 tactaaatta ttccatttgc aacttatccc caacaggtgg tgaagcagta ttttccaatt
1441 tgaagattca tttgaaggtg gctcctgcca cctgctaata gcagttcaaa ctaaattttt
1501 tgtatcaagt ccctgaatgg aagtatgacg ttgggtccct ctgaagttta attctgagtt
1561 ctcattaaaa gaaatttgct ttcattgttt tatttcttaa ttgctatgct tcagaatcaa
1621 tttgtgtttt atgccctttc ccccagtatt gtagagcaag tcttgtgtta aaagcccagt
1681 gtgacagtgt catgatgtag tagtgtctta ctggtttttt aataaatcct tttgtataaa
1741 aatgtattgg ctcttttatc atcagaatag gaaaaattgt catggattca agttattaaa
1801 agcataagtt tggaagacag gcttgccgaa attgaggaca tgattaaaat tgcagtgaag
1861 tttgaaatgt ttttagcaaa atctaatttt tgccataatg tgtcctccct gtccaaattg
1921 ggaatgactt aatgtcaatt tgtttgttgg ttgttttaat aatacttcct tatgtagcca
1981 ttaagattta tatgaatatt ttcccaaatg cccagttttt gcttaatatg tattgtgctt
2041 tttagaacaa atctggataa atgtgcaaaa gtaccccttt gcacagatag ttaatgtttt
2101 atgcttccat taaataaaaa ggacttaaaa tctgttaatt ataatagaaa tgcggctagt
2161 tcagagagat ttttagagct gtggtggact tcatagatga attcaagtgt tgagggagga
2221 ttaaagaaat atataccgtg tttatgtgtg tgtgctt

Protein sequence: **ISOFORM A**

NCBI Reference Sequence: NP_112738.1

LOCUS NP_112738

ACCESSION NP_112738

```
      1 mseeqfggdg aaaaataavg gsageqegam vaatqgaaaa agsgagtggg
tasggteggs
       61 aesegakida skneedeghs nssprhseaa taqreewkmf igglswdttk
kdlkdyfskf
      121 gevvdctlkl dpitgrsrgf gfvlfkeses vdkvmdqkeh klngkvidpk
rakamktkep
      181 vkkifvggls pdtpeekire yfggfgeves ielpmdnktn krrgfcfitf
keeepvkkim
      241 ekkyhnvgls kceikvamsk eqyqqqqqwg srggfagrar grgggpsqnw
nqgysnywnq
      301 gygnygynsq gyggyggydy tgynnyygyg dysnqqsgyg kvsrrgghqn sykpy
```

RPL32

Official Symbol: **RPL32**

Official Name: ribosomal protein L32

Gene ID: 6161

Organism: Homo sapiens

Other Aliases: AU020185, **rpL32**-3A

Other Designations: 60S ribosomal protein L32; snoRNA MBI-141

Nucleotide sequence: **Transcript Variant 1**

NCBI Reference Sequence: NM_000994.3

LOCUS NM_000994

ACCESSION NM_000994

```
      1 aggggttacg acccatcagc ccttgcgcgc caccgtccct tctctcttcc
tcggcgctgc
       61 ctacggaggt ggcagccatc tccttctcgg catcatggcc gccctcagac
cccttgtgaa
      121 gcccaagatc gtcaaaaaga gaaccaagaa gttcatccgg caccagtcag
accgatatgt
      181 caaaattaag cgtaactggc ggaaacccag aggcattgac aacagggttc
gtagaagatt
```

241 caagggccag atcttgatgc ccaacattgg ttatggaagc aacaaaaaaa caaagcacat
301 gctgcccagt ggcttccgga agttcctggt ccacaacgtc aaggagctgg aagtgctgct
361 gatgtgcaac aaatcttact gtgccgagat cgctcacaat gtttcctcca agaaccgcaa
421 agccatcgtg gaaagagctg cccaactggc catcagagtc accaacccca atgccaggct
481 gcgcagtgaa gaaaatgagt aggcagctca tgtgcacgtt ttctgtttaa ataaatgtaa
541 aaactgccat ctggcatctt ccttccttga ttttaagtct tcagcttctt ggccaactta
601 gtttgccaca gagattgttc ttttgcttaa gccccttgg aatctcccat ttggagggga
661 tttgtaaagg acactcagtc cttgaacagg ggaatgtggc ctcaagtgca cagactagcc
721 ttagtcatct ccagttgagg ctgggtatga ggggtacaga cttggccctc acaccaggta
781 ggttctgaga cacttgaaga agcttgtggc tcccaagcca caagtagtca ttcttagcct
841 tgcttttgta aagttaggtg acaagttatt ccatgtgatg cttgtgagaa ttgagaaaat
901 atgcatggaa atatccagat gaatttctta cacagattct tacgggatgc ctaaattgca
961 tcctgtaact tctgtccaaa aagaacagga tgatgtacaa attgctcttc caggtaatcc
1021 accacggtta actggaaaag cactttcagt ctcctataac cctcccacca gctgctgctt
1081 caggtataat gttacagcag tttgccaagg cggggaccta actggtgaca attgagcctc
1141 ttgactggta ctcagaattt agtgacacgt ggtcctgatt ttttttggag acggggtctt
1201 gctctcaccc aggctgggag tgcagtggca cactgactac agccttgacc tccccaggct
1261 caggtgatct tcccacctca gccttccaag tagctgggac tacagatgca cacctccaaa
1321 cctgggtagt ttttgaagtt tttttgtaga ggtggtctag ccatgttgcc taggctcccg
1381 aactcctgag ctcaagcaat cctgcttcag cctcccaaag tactgggatt acaggcatct
1441 tctgtagtat ataggtcatg agggatatgg gatgtggtac ttatgagaca gaaatgctta
1501 caggatgttt ttctgtaacc atcctggtca acttagcaga aatgctgcgc tgggtataat
1561 aaagcttttc tacttctagt ctagacagga atcttacaga ttgtctcctg ttcaaaacct
1621 agtcataaat atttataatg caaactggtc aaaaaaaaaa aaaaaaaa

Protein sequence: **Transcript Variant 1.**

NCBI Reference Sequence: NP_000985.1

LOCUS NP_000985

ACCESSION NP_000985

1 maalrplvkp kivkkrtkkf irhqsdryvk ikrnwrkprg idnrvrrrfk gqilmpnigy

```
       61 gsnkktkhml psgfrkflvh nvkelevllm cnksycaeia hnvssknrka
iveraaqlai
      121 rvtnpnarlr seene
```

Nucleotide sequence: **Transcript Variant 2.**

NCBI Reference Sequence: NM_001007073.1

LOCUS NM_001007073

ACCESSION NM_001007073

```
      1 aggggttacg acccatcagc ccttgcgcgc caccgtccct tctctcttcc
tcggcgctgc
       61 ctacggaggt ggcagccatc tccttctcgc tggcgattgg aagacactct
gcgacagtgt
      121 tcagtccctg ggcaggaaag cctccttcca ggattcttcc tcacctgggg
ccgcttcttc
      181 cccaaaaggc atcatggccg ccctcagacc ccttgtgaag cccaagatcg
tcaaaaagag
      241 aaccaagaag ttcatccggc accagtcaga ccgatatgtc aaaattaagc
gtaactggcg
      301 gaaacccaga ggcattgaca acagggttcg tagaagattc aagggccaga
tcttgatgcc
      361 caacattggt tatggaagca acaaaaaaac aaagcacatg ctgcccagtg
gcttccggaa
      421 gttcctggtc cacaacgtca aggagctgga agtgctgctg atgtgcaaca
aatcttactg
      481 tgccgagatc gctcacaatg tttcctccaa gaaccgcaaa gccatcgtgg
aaagagctgc
      541 ccaactggcc atcagagtca ccaaccccaa tgccaggctg cgcagtgaag
aaaatgagta
      601 ggcagctcat gtgcacgttt tctgtttaaa taaatgtaaa aactgccatc
tggcatcttc
      661 cttccttgat tttaagtctt cagcttcttg gccaacttag tttgccacag
agattgttct
      721 tttgcttaag ccccttttgga atctcccatt tggaggggat ttgtaaagga
cactcagtcc
      781 ttgaacaggg gaatgtggcc tcaagtgcac agactagcct tagtcatctc
cagttgaggc
      841 tgggtatgag gggtacagac ttggccctca caccaggtag gttctgagac
acttgaagaa
      901 gcttgtggct cccaagccac aagtagtcat tcttagcctt gcttttgtaa
agttaggtga
      961 caagttattc catgtgatgc ttgtgagaat tgagaaaata tgcatggaaa
tatccagatg
     1021 aatttcttac acagattctt acgggatgcc taaattgcat cctgtaactt
ctgtccaaaa
     1081 agaacaggat gatgtacaaa ttgctcttcc aggtaatcca ccacggttaa
ctggaaaagc
     1141 actttcagtc tcctataacc ctcccaccag ctgctgcttc aggtataatg
ttacagcagt
     1201 ttgccaaggc ggggacctaa ctggtgacaa ttgagcctct tgactggtac
tcagaattta
     1261 gtgacacgtg gtcctgattt tttttggaga cggggtcttg ctctcaccca
ggctgggagt
     1321 gcagtggcac actgactaca gccttgacct ccccaggctc aggtgatctt
cccacctcag
```

```
     1381 ccttccaagt agctgggact acagatgcac acctccaaac ctgggtagtt
tttgaagttt
     1441 ttttgtagag gtggtctagc catgttgcct aggctcccga actcctgagc
tcaagcaatc
     1501 ctgcttcagc ctcccaaagt actgggatta caggcatctt ctgtagtata
taggtcatga
     1561 gggatatggg atgtggtact tatgagacag aaatgcttac aggatgtttt
tctgtaacca
     1621 tcctggtcaa cttagcagaa atgctgcgct gggtataata aagcttttct
acttctagtc
     1681 tagacaggaa tcttacagat tgtctcctgt tcaaaaccta gtcataaata
tttataatgc
     1741 aaactggtca aaaaaaaaaa aaaaaaa
```

Protein sequence: **Transcript Variant 2.**

NCBI Reference Sequence: NP_001007074.1

LOCUS NP_001007074

ACCESSION NP_001007074

```
      1 maalrplvkp kivkkrtkkf irhqsdryvk ikrnwrkprg idnrvrrrfk
gqilmpnigy
      61 gsnkktkhml psgfrkflvh nvkelevllm cnksycaeia hnvssknrka
iveraaqlai
     121 rvtnpnarlr seene
```

Nucleotide sequence: **Transcript Variant 3.**

NCBI Reference Sequence: NM_001007074.1

LOCUS NM_001007074

ACCESSION NM_001007074

```
      1 gacctcctgg gatcgcatct ggagagtgcc tagtattctg ccagcttcgg
aaagggaggg
      61 aaagcaagcc tggcagaggc acccattcca ttcccagctt gctccgtagc
tggcgattgg
     121 aagacactct gcgacagtgt tcagtccctg ggcaggaaag cctccttcca
ggattcttcc
     181 tcacctgggg ccgcttcttc cccaaaaggc atcatggccg ccctcagacc
ccttgtgaag
     241 cccaagatcg tcaaaaagag aaccaagaag ttcatccggc accagtcaga
ccgatatgtc
     301 aaaattaagc gtaactggcg gaaacccaga ggcattgaca acagggttcg
tagaagattc
     361 aagggccaga tcttgatgcc caacattggt tatggaagca acaaaaaaac
aaagcacatg
     421 ctgcccagtg gcttccggaa gttcctggtc cacaacgtca aggagctgga
agtgctgctg
     481 atgtgcaaca aatcttactg tgccgagatc gctcacaatg tttcctccaa
gaaccgcaaa
     541 gccatcgtgg aaagagctgc ccaactggcc atcagagtca ccaaccccaa
tgccaggctg
```

```
      601 cgcagtgaag aaaatgagta ggcagctcat gtgcacgttt tctgtttaaa
taaatgtaaa
      661 aactgccatc tggcatcttc cttccttgat tttaagtctt cagcttcttg
gccaacttag
      721 tttgccacag agattgttct tttgcttaag ccccctttgga atctcccatt
tggaggggat
      781 ttgtaaagga cactcagtcc ttgaacaggg gaatgtggcc tcaagtgcac
agactagcct
      841 tagtcatctc cagttgaggc tgggtatgag gggtacagac ttggccctca
caccaggtag
      901 gttctgagac acttgaagaa gcttgtggct cccaagccac aagtagtcat
tcttagcctt
      961 gcttttgtaa agttaggtga caagttattc catgtgatgc ttgtgagaat
tgagaaaata
     1021 tgcatggaaa tatccagatg aatttcttac acagattctt acgggatgcc
taaattgcat
     1081 cctgtaactt ctgtccaaaa agaacaggat gatgtacaaa ttgctcttcc
aggtaatcca
     1141 ccacggttaa ctggaaaagc actttcagtc tcctataacc ctcccaccag
ctgctgcttc
     1201 aggtataatg ttacagcagt ttgccaaggc ggggacctaa ctggtgacaa
ttgagcctct
     1261 tgactggtac tcagaattta gtgacacgtg gtcctgattt tttttggaga
cggggtcttg
     1321 ctctcaccca ggctgggagt gcagtggcac actgactaca gccttgacct
ccccaggctc
     1381 aggtgatctt cccacctcag ccttccaagt agctgggact acagatgcac
acctccaaac
     1441 ctgggtagtt tttgaagttt ttttgtagag gtggtctagc catgttgcct
aggctcccga
     1501 actcctgagc tcaagcaatc ctgcttcagc ctcccaaagt actgggatta
caggcatctt
     1561 ctgtagtata taggtcatga gggatatggg atgtggtact tatgagacag
aaatgcttac
     1621 aggatgtttt tctgtaacca tcctggtcaa cttagcagaa atgctgcgct
gggtataata
     1681 aagcttttct acttctagtc tagacaggaa tcttacagat tgtctcctgt
tcaaaaccta
     1741 gtcataaata tttataatgc aaactggtca aaaaaaaaaa aaaaaaa
```

Protein sequence: **Transcript Variant 3.**

NCBI Reference Sequence: NP_001007075.1

LOCUS NP_001007075

ACCESSION NP_001007075

```
      1 maalrplvkp kivkkrtkkf irhqsdryvk ikrnwrkprg idnrvrrrfk
gqilmpnigy
       61 gsnkktkhml psgfrkflvh nvkelevllm cnksycaeia hnvssknrka
iveraaqlai
      121 rvtnpnarlr seene
```

Gene

Official Symbol: **ATP5H**

Official Name: ATP synthase, H+ transporting, mitochondrial Fo complex, subunit d

Gene ID: 10476

Organism: Homo sapiens

Other Aliases: My032, ATPQ

Other Designations: ATP synthase D chain, mitochondrial; ATP synthase subunit d, mitochondrial; ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d; ATP synthase, H+ transporting, mitochondrial F1F0, subunit d; ATPase subunit d; My032 protein

Nucleotide sequence: **ISOFORM B**

NCBI Reference Sequence: NM_001003785.1

LOCUS NM_001003785

ACCESSION NM_001003785

```
      1 tgacccactt ccgttacttg ctgcggagga ccgtgggcag ccagggtcgg
tgaaggatcc
       61 caaaatggct gggcgaaaac ttgctctaaa aaccattgac tgggtagctt
ttgcagagat
      121 catacccag aaccaaaagg ccattgctag ttccctgaaa tcctggaatg
agaccctcac
      181 ctccaggttg gctgctttac ctgagaatcc accagctatc gactgggctt
actacaaggc
      241 caatgtggcc aaggctggct tggtggatga ctttgagaag aaggtgaaat
cttgtgctga
      301 gtgggtgtct ctctcaaagg ccaggattgt agaatatgag aaagagatgg
agaagatgaa
      361 gaacttaatt ccatttgatc agatgaccat tgaggacttg aatgaagctt
tcccagaaac
      421 caaattagac aagaaaaagt atccctattg gcctcaccaa ccaattgaga
atttataaaa
      481 ttgagtccag gaggaagctc tggcccttgt attacacatt ctggacatta
aaaataataa
      541 ttatacagtt aaaaaa
```

Protein sequence: **ISOFORM B**

NCBI Reference Sequence: NP_001003785.1

LOCUS NP_001003785

ACCESSION NP_001003785

```
      1 magrklalkt idwvafaeii pqnqkaiass lkswnetlts rlaalpenpp
aidwayykan
```

```
       61 vakaglvddf ekkvkscaew vslskarive yekemekmkn lipfdqmtie
dlneafpetk
      121 ldkkkypywp hqpienl
```

Nucleotide sequence: **ISOFORM A**

NCBI Reference Sequence: NM_006356.2

LOCUS NM_006356

ACCESSION NM_006356

```
      1 tgacccactt ccgttacttg ctgcggagga ccgtgggcag ccagggtcgg
tgaaggatcc
       61 caaaatggct gggcgaaaac ttgctctaaa aaccattgac tgggtagctt
ttgcagagat
      121 cataccccag aaccaaaagg ccattgctag ttccctgaaa tcctggaatg
agaccctcac
      181 ctccaggttg gctgctttac ctgagaatcc accagctatc gactgggctt
actacaaggc
      241 caatgtggcc aaggctggct tggtggatga ctttgagaag aagtttaatg
cgctgaaggt
      301 tcccgtgcca gaggataaat atactgccca ggtggatgcc gaagaaaaag
aagatgtgaa
      361 atcttgtgct gagtgggtgt ctctctcaaa ggccaggatt gtagaatatg
agaaagagat
      421 ggagaagatg aagaacttaa ttccatttga tcagatgacc attgaggact
tgaatgaagc
      481 tttcccagaa accaaattag acaagaaaaa gtatccctat tggcctcacc
aaccaattga
      541 gaatttataa aattgagtcc aggaggaagc tctggccctt gtattacaca
ttctggacat
      601 taaaaataat aattatacag ttaaaaaa
```

Protein sequence: **ISOFORM A**

NCBI Reference Sequence: NP_006347.1

LOCUS NP_006347

ACCESSION NP_006347

```
      1 magrklalkt idwvafaeii pqnqkaiass lkswnetlts rlaalpenpp
aidwayykan
       61 vakaglvddf ekkfnalkvp vpedkytaqv daeekedvks caewvslska
riveyekeme
      121 kmknlipfdq mtiedlneaf petkldkkky pywphqpien l
```

PSMA1

Official Symbol: **PSMA1**

Official Name: proteasome (prosome, macropain) subunit, alpha type, 1

Gene ID: 5682

Organism: Homo sapiens

Other Aliases: HC2, NU, PROS30

Other Designations: 30 kDa prosomal protein; PROS-30; macropain subunit C2; macropain subunit nu; multicatalytic endopeptidase complex subunit C2; proteasome component C2; proteasome nu chain; proteasome subunit alpha type-1; proteasome subunit nu; proteasome subunit, alpha-type, 1; protein P30-33K

Nucleotide sequence: **ISOFORM 3**

NCBI Reference Sequence: NM_001143937.1

LOCUS NM_001143937

ACCESSION NM_001143937

```
      1 gatatctctg gaatagactg cgctaccctg cgccgccgcc gtcaaactcc
cgcagacttc
       61 tctgtagatc gctgagcgat actttcggca gcacctcctt gattctcagt
tttgctggag
      121 gccgcaacca ggcccgcgcc gccaccatgt ttcgaaatca gtatgacaat
gatgtcactg
      181 tttggagccc ccagggcagg attcatcaaa ttgaatatgc aatggaagct
gttaaacaag
      241 gttcagccac agttggtctg aaatcaaaaa ctcatgcagt tttggttgca
ttgaaaaggg
      301 cgcaatcaga gcttgcagct catcagaaaa aaattctcca tgttgacaac
catattggta
      361 tctcaattgc ggggcttact gctgatgcta gactgttatg taattttatg
cgtcaggagt
      421 gtttggattc cagatttgta ttcgatagac cactgcctgt gtctcgtctt
gtatctctaa
      481 ttggaagcag tatccttttt atgttagcat ttatggatat gaactttgaa
gggttttgat
      541 acttgtgtta attattagga atataataat aatatgacat aggtaagatt
gtgaaaactt
      601 taaaacaaca aattggattg ctctttcatt agcctttata agcaatttat
atttgctaga
      661 cacaaataag cccaacttca ggaaaatcat ctaagcatct ttttagaggg
gatttaaagt
      721 ttcttaatgg ttctagatgt cccaagaaat cctagacccc ttgtatccaa
aacaaatcag
      781 gttttagatg ggaagaaatt attttgctgg cactctttct taggttcggt
agaaagtcaa
      841 acattttata ttaggccaaa gaaatagtgt cctattgcat tatttctctg
gtggattatg
      901 caacaattaa agaataagcc agagac
```

Protein sequence: **ISOFORM 3**

NCBI Reference Sequence: NP_001137409.1

LOCUS NP_001137409

ACCESSION NP_001137409

```
      1 mfrnqydndv tvwspqgrih qieyameavk qgsatvglks kthavlvalk
raqselaahq
       61 kkilhvdnhi gisiagltad arllcnfmrq ecldsrfvfd rplpvsrlvs
ligssilfml
      121 afmdmnfegf
```

Nucleotide sequence: **ISOFORM 2**

NCBI Reference Sequence: NM_002786.3

LOCUS NM_002786

ACCESSION NM_002786

```
       1 gatatctctg gaatagactg cgctaccctg cgccgccgcc gtcaaactcc
cgcagacttc
       61 tctgtagatc gctgagcgat actttcggca gcacctcctt gattctcagt
tttgctggag
      121 gccgcaacca ggcccgcgcc gccaccatgt ttcgaaatca gtatgacaat
gatgtcactg
      181 tttggagccc ccagggcagg attcatcaaa ttgaatatgc aatggaagct
gttaaacaag
      241 gttcagccac agttggtctg aaatcaaaaa ctcatgcagt tttggttgca
ttgaaaaggg
      301 cgcaatcaga gcttgcagct catcagaaaa aaattctcca tgttgacaac
catattggta
      361 tctcaattgc ggggcttact gctgatgcta gactgttatg taattttatg
cgtcaggagt
      421 gtttggattc cagatttgta ttcgatagac cactgcctgt gtctcgtctt
gtatctctaa
      481 ttggaagcaa gacccagata ccaacacaac gatatggccg gagaccatat
ggtgttggtc
      541 tccttattgc tggttatgat gatatgggcc ctcacatttt ccaaacctgt
ccatctgcta
      601 actattttga ctgcagagcc atgtccattg gagcccgttc ccaatcagct
cgtacttact
      661 tggagagaca tatgtctgaa tttatggagt gtaatttaaa tgaactagtt
aaacatggtc
      721 tgcgtgcctt aagagagacg cttcctgcag aacaggacct gactacaaag
aatgtttcca
      781 ttggaattgt tggtaaagac ttggagttta caatctatga tgatgatgat
gtgtctccat
      841 tcctggaagg tcttgaagaa agaccacaga gaaaggcaca gcctgctcaa
cctgctgatg
      901 aacctgcaga aaaggctgat gaaccaatgg aacattaagt gataagccag
tctatatatg
      961 tattatcaaa tatgtaagaa tacaggcacc acatactgat gacaataatc
tatactttga
     1021 accaaaagtt gcagagtggt ggaatgctat gttttaggaa tcagtccaga
tgtgagtttt
     1081 ttccaagcaa cctcactgaa acctatataa tggaatacat ttttctttga
aagggtctgt
```

```
     1141 ataatcattt tctagaaagt atgggtatct atactaatgt ttttatatga
agaacatagg
     1201 tgtctttgtg gttttaaaga caactgtgaa ataaaattgt ttcaccgcct
ggtaaaaaaa
     1261 aaaaaaaaaa aaaaaaaaaa a
```

Protein sequence: **ISOFORM 2**

NCBI Reference Sequence: NP_002777.1

LOCUS NP_002777

ACCESSION NP_002777

```
      1 mfrnqydndv tvwspqgrih qieyameavk qgsatvglks kthavlvalk
raqselaahq
       61 kkilhvdnhi gisiagltad arllcnfmrq ecldsrfvfd rplpvsrlvs
ligsktqipt
      121 qrygrrpygv glliagyddm gphifqtcps anyfdcrams igarsqsart
ylerhmsefm
      181 ecnlnelvkh glralretlp aeqdlttknv sigivgkdle ftiyddddvs
pflegleerp
      241 qrkaqpaqpa depaekadep meh
```

Nucleotide sequence: **ISOFORM 1**

NCBI Reference Sequence: NM_148976.2

LOCUS NM_148976

ACCESSION NM_148976

```
      1 cggccgccca acagggacgc gagccgggac cacgccgacc cagcgtgccc
aggccgagga
       61 aagcgcggcg gcggcagtcc gaagacccac cgggactgaa agagaaggac
gaggtcatct
      121 tcggacggga ggggcaagcc agccatcctg ggaccccagg cgtgcaggtt
ctctttgagg
      181 gtattccacc ctgcaaaaag catgtattca tggtcagctc tcagcaaggc
cagtagcaga
      241 gtggtaaagg ccttggccct ccaaggctgg gaaaagacaa tgacaagtca
aatccagacc
      301 tatgttgtat gttggtctac taggtgactg tctcctggaa atgttatgca
gctcagcaag
      361 gtgaagtttc gaaatcagta tgacaatgat gtcactgttt ggagccccca
gggcaggatt
      421 catcaaattg aatatgcaat ggaagctgtt aaacaaggtt cagccacagt
tggtctgaaa
      481 tcaaaaactc atgcagtttt ggttgcattg aaaagggcgc aatcagagct
tgcagctcat
      541 cagaaaaaaa ttctccatgt tgacaaccat attggtatct caattgcggg
gcttactgct
      601 gatgctagac tgttatgtaa ttttatgcgt caggagtgtt tggattccag
atttgtattc
      661 gatagaccac tgcctgtgtc tcgtcttgta tctctaattg gaagcaagac
ccagatacca
```

```
      721 acacaacgat atggccggag accatatggt gttggtctcc ttattgctgg
ttatgatgat
      781 atgggccctc acattttcca aacctgtcca tctgctaact attttgactg
cagagccatg
      841 tccattggag cccgttccca atcagctcgt acttacttgg agagacatat
gtctgaattt
      901 atggagtgta atttaaatga actagttaaa catggtctgc gtgccttaag
agagacgctt
      961 cctgcagaac aggacctgac tacaaagaat gtttccattg gaattgttgg
taaagacttg
     1021 gagtttacaa tctatgatga tgatgatgtg tctccattcc tggaaggtct
tgaagaaaga
     1081 ccacagagaa aggcacagcc tgctcaacct gctgatgaac ctgcagaaaa
ggctgatgaa
     1141 ccaatggaac attaagtgat aagccagtct atatatgtat tatcaaatat
gtaagaatac
     1201 aggcaccaca tactgatgac aataatctat actttgaacc aaaagttgca
gagtggtgga
     1261 atgctatgtt ttaggaatca gtccagatgt gagtttttc caagcaacct
cactgaaacc
     1321 tatataatgg aatacatttt tctttgaaag ggtctgtata atcattttct
agaaagtatg
     1381 ggtatctata ctaatgtttt tatatgaaga acataggtgt ctttgtggtt
ttaaagacaa
     1441 ctgtgaaata aaattgtttc accgcctggt aaaaaaaaaa aaaaaaaaaa
aaaaaaaa
```

Protein sequence: **ISOFORM 1**

NCBI Reference Sequence: NP_683877.1

LOCUS NP_683877

ACCESSION NP_683877

```
      1 mqlskvkfrn qydndvtvws pqgrihqiey ameavkqgsa tvglksktha
vlvalkraqs
       61 elaahqkkil hvdnhigisi agltadarll cnfmrqecld srfvfdrplp
vsrlvsligs
      121 ktqiptqryg rrpygvglli agyddmgphi fqtcpsanyf dcramsigar
sqsartyler
      181 hmsefmecnl nelvkhglra lretlpaeqd lttknvsigi vgkdleftiy
ddddvspfle
      241 gleerpqrka qpaqpadepa ekadepmeh
```

PTBP1

Official Symbol: **PTBP1**

Official Name: polypyrimidine tract binding protein 1

Gene ID: 5725

Organism: Homo sapiens

Other Aliases: HNRNP-I, HNRNPI, HNRPI, PTB, PTB-1, PTB-T, PTB2, PTB3, PTB4, pPTB

Other Designations: 57 kDa RNA-binding protein PPTB-1; RNA-binding protein; heterogeneous nuclear ribonucleoprotein I; heterogeneous nuclear ribonucleoprotein polypeptide I; hnRNP I; polypyrimidine tract binding protein (heterogeneous nuclear ribonucleoprotein I); polypyrimidine tract-binding protein 1

Nucleotide sequence: **ISOFORM A**

NCBI Reference Sequence: NM_002819.4

LOCUS NM_002819

ACCESSION NM_002819

```
      1 tgcgggcgtc tccgccattt tgtgagtcta taactcggag ccgttgggtc
ggttcctgct
       61 attccggcgc ctccactccg tcccccgcgg gtctgctctg tgtgccatgg
acggcattgt
      121 cccagatata gccgttggta caaagcgggg atctgacgag cttttctcta
cttgtgtcac
      181 taacggaccg tttatcatga gcagcaactc ggcttctgca gcaaacggaa
atgacagcaa
      241 gaagttcaaa ggtgacagcc gaagtgcagg cgtcccctct agagtgatcc
acatccggaa
      301 gctccccatc gacgtcacgg agggggaagt catctccctg gggctgccct
ttgggaaggt
      361 caccaacctc ctgatgctga aggggaaaaa ccaggccttc atcgagatga
acacggagga
      421 ggctgccaac accatggtga actactacac ctcggtgacc cctgtgctgc
gcggccagcc
      481 catctacatc cagttctcca accacaagga gctgaagacc gacagctctc
ccaaccaggc
      541 gcgggcccag gcggccctgc aggcggtgaa ctcggtccag tcggggaacc
tggccttggc
      601 tgcctcggcg gcggccgtgg acgcagggat ggcgatggcc gggcagagcc
ccgtgctcag
      661 gatcatcgtg gagaacctct tctaccctgt gaccctggat gtgctgcacc
agattttctc
      721 caagttcggc acagtgttga agatcatcac cttcaccaag aacaaccagt
tccaggccct
      781 gctgcagtat gcggaccccg tgagcgccca gcacgccaag ctgtcgctgg
acgggcagaa
      841 catctacaac gcctgctgca cgctgcgcat cgacttttcc aagctcacca
gcctcaacgt
      901 caagtacaac aatgacaaga gccgtgacta cacacgccca gacctgcctt
ccggggacag
      961 ccagccctcg ctggaccaga ccatggccgc ggccttcggt gcacctggta
taatctcagc
     1021 ctctccgtat gcaggagctg gtttccctcc caccttgcc attcctcaag
ctgcaggcct
```

1081 ttccgttccg aacgtccacg gcgccctggc ccccctggcc atcccctcgg cggcggcggc
1141 agctgcggcg gcaggtcgga tcgccatccc gggcctggcg ggggcaggaa attctgtatt
1201 gctggtcagc aacctcaacc cagagagagt cacaccccaa agcctcttta ttcttttcgg
1261 cgtctacggt gacgtgcagc gcgtgaagat cctgttcaat aagaaggaga acgccctagt
1321 gcagatggcg gacggcaacc aggcccagct ggccatgagc cacctgaacg ggcacaagct
1381 gcacgggaag cccatccgca tcacgctctc gaagcaccag aacgtgcagc tgccccgcga
1441 gggccaggag gaccagggcc tgaccaagga ctacggcaac tcacccctgc accgcttcaa
1501 gaagccgggc tccaagaact tccagaacat attcccgccc tcggccacgc tgcacctctc
1561 caacatcccg ccctcagtct ccgaggagga tctcaaggtc ctgttttcca gcaatggggg
1621 cgtcgtcaaa ggattcaagt tcttccagaa ggaccgcaag atggcactga tccagatggg
1681 ctccgtggag gaggcggtcc aggccctcat tgacctgcac aaccacgacc tcggggagaa
1741 ccaccacctg cgggtctcct tctccaagtc caccatctag gggcacaggc ccccacggcc
1801 gggccccctg gcgacaactt ccatcattcc agagaaaagc cactttaaaa acagctgaag
1861 tgaccttagc agaccagaga ttttatttt ttaaagagaa atcagtttac ctgtttttaa
1921 aaaaattaaa tctagttcac cttgctcacc ctgcggtgac agggacagct caggctcttg
1981 gtgactgtgg cagcgggagt tcccggccct ccacacccgg ggccagaccc tcggggccat
2041 gccttggtgg ggcctgtgtc gggcgtgggg cctgcaggtg ggcgccccga ccacgacttg
2101 gcttccttgt gccttaaaaa acctgccttc ctgcagccac acacccaccc ggggtgtcct
2161 ggggacccaa ggggtggggg ggtcacacca gagagaggca gggggcctgg ccggctcctg
2221 caggatcatg cagctggggc gcggcggccg cggctgcgac accccaaccc cagccctcta
2281 atcaagtcac gtgattctcc cttcaccccg cccccagggc cttcccttct gcccccaggc
2341 gggctccccg ctgctccagc tgcggagctg gtcgacataa tctctgtatt atatactttg
2401 cagttgcaga cgtctgtgcc tagcaatatt tccagttgac caaatattct aatctttttt
2461 catttatatg caaaagaaat agttttaagt aacttttttat agcaagatga tacaatggta
2521 tgagtgtaat ctaaacttcc ttgtggtatt accttgtatg ctgttacttt tattttattc
2581 cttgtaatta agtcacaggc aggacccagt ttccagagag caggcggggc cgcccagtgg
2641 gtcaggcaca gggagccccg gtcctatctt agagcccctg agcttcaggg aaggggcggg
2701 cgtgtcgccg cctctggcat cgcctccggt tgccttacac cacgccttca cctgcagtcg
2761 cctagaaaac ttgctctcaa acttcagggt tttttcttcc ttcaaatttt ggaccaaagt
2821 ctcatttctg tgttttgcct gcctctgatg ctgggacccg gaaggcgggc gctcctcctg

```
     2881 tcttctctgt gctctttcta ccgcccccgc gtcctgtccc gggggctctc
ctaggatccc
     2941 ctttccgtaa aagcgtgtaa caagggtgta aatatttata atttttata
cctgttgtga
     3001 gacccgaggg gcggcggcgc ggtttttat ggtgacacaa atgtatattt
tgctaacagc
     3061 aattccaggc tcagtattgt gaccgcggag ccacagggga ccccacgcac
attccgttgc
     3121 cttacccgat ggcttgtgac gcggagagaa ccgattaaaa ccgtttgaga
aactcctccc
     3181 ttgtctagcc ctgtgttcgc tgtggacgct gtagaggcag gttggccagt
ctgtacctgg
     3241 acttcgaata aatcttctgt atcctcgctc cgttccgcct taaaaaaaaa
aaaaaaaaaa
     3301 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
```

Protein sequence: **ISOFORM A**

NCBI Reference Sequence: NP_002810.1

LOCUS NP_002810

ACCESSION NP_002810

```
      1 mdgivpdiav gtkrgsdelf stcvtngpfi mssnsasaan gndskkfkgd
srsagvpsrv
       61 ihirklpidv tegevislgl pfgkvtnllm lkgknqafie mnteeaantm
vnyytsvtpv
      121 lrgqpiyiqf snhkelktds spnqaraqaa lqavnsvqsg nlalaasaaa
vdagmamagq
      181 spvlriiven lfypvtldvl hqifskfgtv lkiitftknn qfqallqyad
pvsaqhakls
      241 ldgqniynac ctlridfskl tslnvkynnd ksrdytrpdl psgdsqpsld
qtmaaafgap
      301 giisaspyag agfpptfaip qaaglsvpnv hgalaplaip saaaaaaaag
riaipglaga
      361 gnsvllvsnl npervtpqsl filfgvygdv qrvkilfnkk enalvqmadg
nqaqlamshl
      421 nghklhgkpi ritlskhqnv qlpregqedq gltkdygnsp lhrfkkpgsk
nfqnifppsa
      481 tlhlsnipps vseedlkvlf ssnggvvkgf kffqkdrkma liqmgsveea
vqalidlhnh
      541 dlgenhhlrv sfsksti
```

Nucleotide sequence: **ISOFORM B**

NCBI Reference Sequence: NM_031990.3

LOCUS NM_031990

ACCESSION NM_031990

```
      1 tgcgggcgtc tccgccattt tgtgagtcta taactcggag ccgttgggtc
ggttcctgct
       61 attccggcgc ctccactccg tcccccgcgg gtctgctctg tgtgccatgg
acggcattgt
```

121 cccagatata gccgttggta caaagcgggg atctgacgag cttttctcta
cttgtgtcac
181 taacggaccg tttatcatga gcagcaactc ggcttctgca gcaaacggaa
atgacagcaa
241 gaagttcaaa ggtgacagcc gaagtgcagg cgtcccctct agagtgatcc
acatccggaa
301 gctccccatc gacgtcacgg agggggaagt catctccctg gggctgccct
ttgggaaggt
361 caccaacctc ctgatgctga aggggaaaaa ccaggccttc atcgagatga
acacggagga
421 ggctgccaac accatggtga actactacac ctcggtgacc cctgtgctgc
gcggccagcc
481 catctacatc cagttctcca accacaagga gctgaagacc gacagctctc
ccaaccaggc
541 gcgggcccag gcggccctgc aggcggtgaa ctcggtccag tcggggaacc
tggccttggc
601 tgcctcggcg gcggccgtgg acgcagggat ggcgatggcc gggcagagcc
ccgtgctcag
661 gatcatcgtg gagaacctct tctaccctgt gaccctggat gtgctgcacc
agattttctc
721 caagttcggc acagtgttga agatcatcac cttcaccaag aacaaccagt
tccaggccct
781 gctgcagtat gcggaccccg tgagcgccca gcacgccaag ctgtcgctgg
acgggcagaa
841 catctacaac gcctgctgca cgctgcgcat cgacttttcc aagctcacca
gcctcaacgt
901 caagtacaac aatgacaaga gccgtgacta cacacgccca gacctgcctt
ccggggacag
961 ccagccctcg ctggaccaga ccatggccgc ggccttcgcc tctccgtatg
caggagctgg
1021 tttccctccc acctttgcca ttcctcaagc tgcaggcctt tccgttccga
acgtccacgg
1081 cgccctggcc cccctggcca tcccctcggc ggcggcggca gctgcggcgg
caggtcggat
1141 cgccatcccg ggcctggcgg gggcaggaaa ttctgtattg ctggtcagca
acctcaaccc
1201 agagagagtc acaccccaaa gcctctttat tcttttcggc gtctacggtg
acgtgcagcg
1261 cgtgaagatc ctgttcaata agaaggagaa cgccctagtg cagatggcgg
acggcaacca
1321 ggcccagctg gccatgagcc acctgaacgg gcacaagctg cacgggaagc
ccatccgcat
1381 cacgctctcg aagcaccaga acgtgcagct gccccgcgag ggccaggagg
accagggcct
1441 gaccaaggac tacggcaact cacccctgca ccgcttcaag aagccgggct
ccaagaactt
1501 ccagaacata ttcccgccct cggccacgct gcacctctcc aacatcccgc
cctcagtctc
1561 cgaggaggat ctcaaggtcc tgttttccag caatgggggc gtcgtcaaag
gattcaagtt
1621 cttccagaag gaccgcaaga tggcactgat ccagatgggc tccgtggagg
aggcggtcca
1681 ggccctcatt gacctgcaca accacgacct cggggagaac caccacctgc
gggtctcctt
1741 ctccaagtcc accatctagg ggcacaggcc cccacggccg ggccccctgg
cgacaacttc
1801 catcattcca gagaaaagcc actttaaaaa cagctgaagt gaccttagca
gaccagagat
1861 tttatttttt taaagagaaa tcagtttacc tgtttttaaa aaaattaaat
ctagttcacc

```
     1921 ttgctcaccc tgcggtgaca gggacagctc aggctcttgg tgactgtggc
agcgggagtt
     1981 cccggccctc cacacccggg gccagaccct cggggccatg ccttggtggg
gcctgtgtcg
     2041 ggcgtggggc ctgcaggtgg gcgccccgac cacgacttgg cttccttgtg
ccttaaaaaa
     2101 cctgccttcc tgcagccaca cacccacccg gggtgtcctg gggacccaag
gggtgggggg
     2161 gtcacaccag agagaggcag ggggcctggc cggctcctgc aggatcatgc
agctggggcg
     2221 cggcggccgc ggctgcgaca ccccaacccc agccctctaa tcaagtcacg
tgattctccc
     2281 ttcaccccgc ccccagggcc ttcccttctg cccccaggcg ggctccccgc
tgctccagct
     2341 gcggagctgg tcgacataat ctctgtatta tatactttgc agttgcagac
gtctgtgcct
     2401 agcaatattt ccagttgacc aaatattcta atctttttc atttatatgc
aaaagaaata
     2461 gttttaagta acttttata gcaagatgat acaatggtat gagtgtaatc
taaacttcct
     2521 tgtggtatta ccttgtatgc tgttactttt attttattcc ttgtaattaa
gtcacaggca
     2581 ggacccagtt tccagagagc aggcggggcc gcccagtggg tcaggcacag
ggagccccgg
     2641 tcctatctta gagcccctga gcttcaggga aggggcgggc gtgtcgccgc
ctctggcatc
     2701 gcctccggtt gccttacacc acgccttcac ctgcagtcgc ctagaaaact
tgctctcaaa
     2761 cttcagggtt ttttcttcct tcaaattttg gaccaaagtc tcatttctgt
gttttgcctg
     2821 cctctgatgc tgggacccgg aaggcgggcg ctcctcctgt cttctctgtg
ctctttctac
     2881 cgccccccgcg tcctgtcccg ggggctctcc taggatcccc tttccgtaaa
agcgtgtaac
     2941 aagggtgtaa atatttataa ttttttatac ctgttgtgag acccgagggg
cggcggcgcg
     3001 gtttttttatg gtgacacaaa tgtatatttt gctaacagca attccaggct
cagtattgtg
     3061 accgcggagc cacaggggac cccacgcaca ttccgttgcc ttacccgatg
gcttgtgacg
     3121 cggagagaac cgattaaaac cgtttgagaa actcctccct tgtctagccc
tgtgttcgct
     3181 gtggacgctg tagaggcagg ttggccagtc tgtacctgga cttcgaataa
atcttctgta
     3241 tcctcgctcc gttccgcctt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
     3301 aaaaaaaaaa aaaaaaaaa
```

Protein sequence: **ISOFORM B**

NCBI Reference Sequence: NP_114367.1

LOCUS NP_114367

ACCESSION NP_114367

```
      1 mdgivpdiav gtkrgsdelf stcvtngpfi mssnsasaan gndskkfkgd
srsagvpsrv
       61 ihirklpidv tegevislgl pfgkvtnllm lkgknqafie mnteeaantm
vnyytsvtpv
      121 lrgqpiyiqf snhkelktds spnqaraqaa lqavnsvqsg nlalaasaaa
vdagmamagq
      181 spvlriiven lfypvtldvl hqifskfgtv lkiitftknn qfqallqyad
pvsaqhakls
      241 ldgqniynac ctlridfskl tslnvkynnd ksrdytrpdl psgdsqpsld
qtmaaafasp
      301 yagagfpptf aipqaaglsv pnvhgalapl aipsaaaaaa aagriaipgl
agagnsvllv
      361 snlnpervtp qslfilfgvy gdvqrvkilf nkkenalvqm adgnqaqlam
shlnghklhg
      421 kpiritlskh qnvqlpregq edqgltkdyg nsplhrfkkp gsknfqnifp
psatlhlsni
      481 ppsvseedlk vlfssnggvv kgfkffqkdr kmaliqmgsv eeavqalidl
hnhdlgenhh
      541 lrvsfsksti
```

Nucleotide sequence: **ISOFORM C**

NCBI Reference Sequence: NM_031991.3

LOCUS NM_031991

ACCESSION NM_031991

```
      1 tgcgggcgtc tccgccattt tgtgagtcta taactcggag ccgttgggtc
ggttcctgct
       61 attccggcgc ctccactccg tcccccgcgg gtctgctctg tgtgccatgg
acggcattgt
      121 cccagatata gccgttggta caaagcgggg atctgacgag cttttctcta
cttgtgtcac
      181 taacggaccg tttatcatga gcagcaactc ggcttctgca gcaaacggaa
atgacagcaa
      241 gaagttcaaa ggtgacagcc gaagtgcagg cgtcccctct agagtgatcc
acatccggaa
      301 gctccccatc gacgtcacgg agggggaagt catctccctg gggctgccct
ttgggaaggt
      361 caccaacctc ctgatgctga aggggaaaaa ccaggccttc atcgagatga
acacggagga
      421 ggctgccaac accatggtga actactacac ctcggtgacc cctgtgctgc
gcggccagcc
      481 catctacatc cagttctcca accacaagga gctgaagacc gacagctctc
ccaaccaggc
      541 gcgggcccag gcggccctgc aggcggtgaa ctcggtccag tcggggaacc
tggccttggc
      601 tgcctcggcg gcggccgtgg acgcagggat ggcgatggcc gggcagagcc
ccgtgctcag
      661 gatcatcgtg gagaacctct tctaccctgt gaccctggat gtgctgcacc
agattttctc
      721 caagttcggc acagtgttga agatcatcac cttcaccaag aacaaccagt
tccaggccct
      781 gctgcagtat gcggaccccg tgagcgccca gcacgccaag ctgtcgctgg
acgggcagaa
      841 catctacaac gcctgctgca cgctgcgcat cgacttttcc aagctcacca
gcctcaacgt
```

```
      901 caagtacaac aatgacaaga gccgtgacta cacacgccca gacctgcctt
ccggggacag
      961 ccagccctcg ctggaccaga ccatggccgc ggccttcggc ctttccgttc
cgaacgtcca
     1021 cggcgccctg gccccctgg ccatcccctc ggcggcggcg gcagctgcgg
cggcaggtcg
     1081 gatcgccatc ccgggcctgg cgggggcagg aaattctgta ttgctggtca
gcaacctcaa
     1141 cccagagaga gtcacacccc aaagcctctt tattcttttc ggcgtctacg
gtgacgtgca
     1201 gcgcgtgaag atcctgttca ataagaagga gaacgcccta gtgcagatgg
cggacggcaa
     1261 ccaggcccag ctggccatga gccacctgaa cgggcacaag ctgcacggga
agcccatccg
     1321 catcacgctc tcgaagcacc agaacgtgca gctgccccgc gagggccagg
aggaccaggg
     1381 cctgaccaag gactacggca actcacccct gcaccgcttc aagaagccgg
gctccaagaa
     1441 cttccagaac atattcccgc cctcggccac gctgcacctc tccaacatcc
cgccctcagt
     1501 ctccgaggag gatctcaagg tcctgttttc cagcaatggg ggcgtcgtca
aaggattcaa
     1561 gttcttccag aaggaccgca agatggcact gatccagatg ggctccgtgg
aggaggcggt
     1621 ccaggccctc attgacctgc acaaccacga cctcggggag aaccaccacc
tgcgggtctc
     1681 cttctccaag tccaccatct aggggcacag gcccccacgg ccgggccccc
tggcgacaac
     1741 ttccatcatt ccagagaaaa gccactttaa aaacagctga agtgacctta
gcagaccaga
     1801 gattttattt ttttaaagag aaatcagttt acctgttttt aaaaaaatta
aatctagttc
     1861 accttgctca ccctgcggtg acagggacag ctcaggctct tggtgactgt
ggcagcggga
     1921 gttcccggcc ctccacaccc ggggccagac cctcggggcc atgccttggt
ggggcctgtg
     1981 tcgggcgtgg ggcctgcagg tgggcgcccc gaccacgact tggcttcctt
gtgccttaaa
     2041 aaacctgcct tcctgcagcc acacacccac ccggggtgtc ctggggaccc
aaggggtggg
     2101 ggggtcacac cagagagagg caggggggcct ggccggctcc tgcaggatca
tgcagctggg
     2161 gcgcggcggc cgcggctgcg acaccccaac cccagccctc taatcaagtc
acgtgattct
     2221 cccttcaccc cgccccccagg gccttccctt ctgcccccag gcgggctccc
cgctgctcca
     2281 gctgcggagc tggtcgacat aatctctgta ttatatactt tgcagttgca
gacgtctgtg
     2341 cctagcaata tttccagttg accaaatatt ctaatctttt ttcatttata
tgcaaaagaa
     2401 atagttttaa gtaacttttt atagcaagat gatacaatgg tatgagtgta
atctaaactt
     2461 ccttgtggta ttaccttgta tgctgttact tttatttttat tccttgtaat
taagtcacag
     2521 gcaggaccca gtttccagag agcaggcggg gccgcccagt gggtcaggca
cagggagccc
     2581 cggtcctatc ttagagcccc tgagcttcag ggaaggggcg ggcgtgtcgc
cgcctctggc
     2641 atcgcctccg gttgccttac accacgcctt cacctgcagt cgcctagaaa
acttgctctc
```

```
     2701 aaacttcagg gtttttttctt ccttcaaatt ttggaccaaa gtctcatttc
tgtgttttgc
     2761 ctgcctctga tgctgggacc cggaaggcgg gcgctcctcc tgtcttctct
gtgctctttc
     2821 taccgccccc gcgtcctgtc ccgggggctc tcctaggatc ccctttccgt
aaaagcgtgt
     2881 aacaagggtg taaatattta taatttttta tacctgttgt gagacccgag
gggcggcggc
     2941 gcggtttttt atggtgacac aaatgtatat tttgctaaca gcaattccag
gctcagtatt
     3001 gtgaccgcgg agccacaggg gaccccacgc acattccgtt gccttacccg
atggcttgtg
     3061 acgcggagag aaccgattaa aaccgtttga gaaactcctc ccttgtctag
ccctgtgttc
     3121 gctgtggacg ctgtagaggc aggttggcca gtctgtacct ggacttcgaa
taaatcttct
     3181 gtatcctcgc tccgttccgc cttaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
     3241 aaaaaaaaaa aaaaaaaaaa aa
```

Protein sequence: **ISOFORM C**

NCBI Reference Sequence: NP_114368.1

LOCUS NP_114368

ACCESSION NP_114368

```
      1 mdgivpdiav gtkrgsdelf stcvtngpfi mssnsasaan gndskkfkgd
srsagvpsrv
       61 ihirklpidv tegevislgl pfgkvtnllm lkgknqafie mnteeaantm
vnyytsvtpv
      121 lrgqpiyiqf snhkelktds spnqaraqaa lqavnsvqsg nlalaasaaa
vdagmamagq
      181 spvlriiven lfypvtldvl hqifskfgtv lkiitftknn qfqallqyad
pvsaqhakls
      241 ldgqniynac ctlridfskl tslnvkynnd ksrdytrpdl psgdsqpsld
qtmaaafgls
      301 vpnvhgalap laipsaaaaa aaagriaipg lagagnsvll vsnlnpervt
pqslfilfgv
      361 ygdvqrvkil fnkkenalvq madgnqaqla mshlnghklh gkpiritlsk
hqnvqlpreg
      421 qedqgltkdy gnsplhrfkk pgsknfqnif ppsatlhlsn ippsvseedl
kvlfssnggv
      481 vkgfkffqkd rkmaliqmgs veeavqalid lhnhdlgenh hlrvsfskst i
```

AP2A1

Official Symbol: **AP2A1**

Official Name: adaptor-related protein complex 2, alpha 1 subunit

Gene ID: 160

Organism: Homo sapiens

Other Aliases: ADTAA, AP2-ALPHA, CLAPA1

Other Designations: 100 kDa coated vesicle protein A; AP-2 complex subunit alpha-1; adapter-related protein complex 2 alpha-1 subunit; adaptin, alpha A; adaptor protein complex AP-2 subunit alpha-1; alpha-adaptin A; alpha1-adaptin; clathrin assembly protein complex 2 alpha-A large chain; clathrin-associated/assembly/adaptor protein, large, alpha 1; plasma membrane adaptor HA2/AP2 adaptin alpha A subunit

Nucleotide sequence: **ISOFORM 1**

NCBI Reference Sequence: NM_014203.2

LOCUS NM_014203

ACCESSION NM_014203

```
      1 cggctcagag ctccggaccg cgggcggagg ggaggggcag ggggcggtgc
cacggcctgc
       61 cagcccgccc gcccgcccgc cagccagccc tccccgcggc cggctcggct
ccttggcgct
      121 gcctggggtc ctttccgccc ggtccccgct tgccagcccc cgctgctctg
tgccctgtcc
      181 ggccaggcct ggagccgaca ccaccgccat catgccggcc gtgtccaagg
gcgatgggat
      241 gcgggggctc gcggtgttca tctccgacat ccggaactgt aagagcaaag
aggcggaaat
      301 taagagaatc aacaaggaac tggccaacat ccgctccaag ttcaaaggag
acaaagcctt
      361 ggatggctac agtaagaaaa aatatgtgtg taaactgctt ttcatcttcc
tgcttggcca
      421 tgacattgac tttgggcaca tggaggctgt gaatctgttg agttccaata
aatacacaga
      481 gaagcaaata ggttacctgt tcatttctgt gctggtgaac tcgaactcgg
agctgatccg
      541 cctcatcaac aacgccatca agaatgacct ggccagccgc aaccccacct
tcatgtgcct
      601 ggccctgcac tgcatcgcca acgtgggcag ccgggagatg ggcgaggcct
ttgccgctga
      661 catcccccgc atcctggtgg ccggggacag catggacagt gtcaagcaga
gtgcggccct
      721 gtgcctcctt cgactgtaca aggcctcgcc tgacctggtg cccatgggcg
agtggacggc
      781 gcgtgtggta cacctgctca atgaccagca catgggtgtg gtcacggccg
ccgtcagcct
      841 catcacctgt ctctgcaaga agaacccaga tgacttcaag acgtgcgtct
ctctggctgt
      901 gtcgcgcctg agccggatcg tctcctctgc ctccaccgac ctccaggact
acacctacta
      961 cttcgtccca gcaccctggc tctcggtgaa gctcctgcgg ctgctgcagt
gctacccgcc
     1021 tccagaggat gcggctgtga aggggcggct ggtggaatgt ctggagactg
tgctcaacaa
     1081 ggcccaggag ccccccaaat ccaagaaggt gcagcattcc aacgccaaga
acgccatcct
     1141 cttcgagacc atcagcctca tcatccacta tgacagtgag cccaacctcc
tggttcgggc
```

1201 ctgcaaccag ctgggccagt tcctgcagca ccgggagacc aacctgcgct
acctggccct
1261 ggagagcatg tgcacgctgg ccagctccga gttctcccat gaagccgtca
agacgcacat
1321 tgacaccgtc atcaatgccc tcaagacgga gcgggacgtc agcgtgcggc
agcgggcggc
1381 tgacctcctc tacgccatgt gtgaccggag caatgccaag cagatcgtgt
cggagatgct
1441 gcggtacctg gagacggcag actacgccat ccgcgaggag atcgtcctga
aggtggccat
1501 cctggccgag aagtacgccg tggactacag ctggtacgtg gacaccatcc
tcaacctcat
1561 ccgcattgcg ggcgactacg tgagtgagga ggtgtggtac cgtgtgctac
agatcgtcac
1621 caaccgtgat gacgtccagg gctatgccgc caagaccgtc tttgaggcgc
tccaggcccc
1681 tgcctgtcac gagaacatgg tgaaggttgg cggctacatc cttggggagt
ttgggaacct
1741 gattgctggg gacccccgct ccagcccccc agtgcagttc tccctgctcc
actccaagtt
1801 ccatctgtgc agcgtggcca cgcgggcgct gctgctgtcc acctacatca
agttcatcaa
1861 cctcttcccc gagaccaagg ccaccatcca gggcgtcctg cgggccggct
cccagctgcg
1921 caatgctgac gtggagctgc agcagcgagc cgtggagtac ctcaccctca
gctcagtggc
1981 cagcaccgac gtcctggcca cggtgctgga ggagatgccg cccttccccg
agcgcgagtc
2041 gtccatcctg gccaagctga aacgcaagaa ggggccaggg gccggcagcg
ccctggacga
2101 tggccggagg gaccccagca gcaacgacat caacgggggc atggagccca
cccccagcac
2161 tgtgtcgacg ccctcgccct ccgccgacct cctggggctg cgggcagccc
ctccccggc
2221 agcacccccg gcttctgcag gagcagggaa ccttctggtg gacgtcttcg
atggcccggc
2281 cgcccagccc agcctggggc ccacccccga ggaggccttc ctcagcgagc
tggagccgcc
2341 tgcccccgag agccccatgg ctttgctggc tgacccagct ccagctgctg
acccaggtcc
2401 tgaggacatc ggccctccca ttccggaagc cgatgagttg ctgaataagt
ttgtgtgtaa
2461 gaacaacggg gtcctgttcg agaaccagct gctgcagatc ggagtcaagt
cagagttccg
2521 acagaacctg ggccgcatgt atctcttcta tggcaacaag acctcggtgc
agttccagaa
2581 tttctcaccc actgtggttc acccgggaga cctccagact cagctggctg
tgcagaccaa
2641 gcgcgtggcg gcgcaggtgg acggcggcgc gcaggtgcag caggtgctca
atatcgagtg
2701 cctgcgggac ttcctgacgc ccccgctgct gtccgtgcgc ttccggtacg
gtggcgcccc
2761 ccaggccctc accctgaagc tcccagtgac catcaacaag ttcttccagc
ccaccgagat
2821 ggcggcccag gatttcttcc agcgctggaa gcagctgagc ctccctcaac
aggaggcgca
2881 gaaaatcttc aaagccaacc accccatgga cgcagaagtt actaaggcca
agcttctggg
2941 gtttggctct gctctcctgg acaatgtgga ccccaaccct gagaacttcg
tgggggcggg

```
     3001 gatcatccag actaaagccc tgcaggtggg ctgtctgctt cggctggagc
ccaatgccca
     3061 ggcccagatg taccggctga ccctgcgcac cagcaaggag cccgtctccc
gtcacctgtg
     3121 tgagctgctg gcacagcagt tctgagccct ggactctgcc ccgggggatg
tggccggcac
     3181 tgggcagccc cttggactga ggcagttttg gtggatgggg gacctccact
ggtgacagag
     3241 aagacaccag ggtttggggg atgcctggga ctttcctccg gccttttgta
tttttatttt
     3301 tgttcatctg ctgctgttta cattctgggg ggttaggggg agtccccctc
cctccctttc
     3361 ccccccaagc acagagggga gaggggccag ggaagtggat gtctcctccc
ctcccacccc
     3421 accctgttgt agcccctcct accccctccc catccagggg ctgtgtatta
ttgtgagcga
     3481 ataaacagag agacgctaa
```

Protein sequence: **ISOFORM 1**

NCBI Reference Sequence: NP_055018.2

LOCUS NP_055018

ACCESSION NP_055018

```
      1 mpavskgdgm rglavfisdi rnckskeaei krinkelani rskfkgdkal
dgyskkkyvc
       61 kllfifllgh didfghmeav nllssnkyte kqigylfisv lvnsnselir
linnaikndl
      121 asrnptfmcl alhcianvgs remgeafaad iprilvagds mdsvkqsaal
cllrlykasp
      181 dlvpmgewta rvvhllndqh mgvvtaavsl itclckknpd dfktcvslav
srlsrivssa
      241 stdlqdytyy fvpapwlsvk llrllqcypp pedaavkgrl vecletvlnk
aqeppkskkv
      301 qhsnaknail fetisliihy dsepnllvra cnqlgqflqh retnlrylal
esmctlasse
      361 fsheavkthi dtvinalkte rdvsvrqraa dllyamcdrs nakqivseml
ryletadyai
      421 reeivlkvai laekyavdys wyvdtilnli riagdyvsee vwyrvlqivt
nrddvqgyaa
      481 ktvfealqap achenmvkvg gyilgefgnl iagdprsspp vqfsllhskf
hlcsvatral
      541 llstyikfin lfpetkatiq gvlragsqlr nadvelqqra veyltlssva
stdvlatvle
      601 emppfperes silaklkrkk gpgagsaldd grrdpssndi nggmeptpst
vstpspsadl
      661 lglraapppa appasagagn llvdvfdgpa aqpslgptpe eaflselepp
apespmalla
      721 dpapaadpgp edigppipea dellnkfvck nngvlfenql lqigvksefr
qnlgrmylfy
      781 gnktsvqfqn fsptvvhpgd lqtqlavqtk rvaaqvdgga qvqqvlniec
lrdfltppll
      841 svrfryggap qaltlklpvt inkffqptem aaqdffqrwk qlslpqqeaq
kifkanhpmd
      901 aevtkakllg fgsalldnvd pnpenfvgag iiqtkalqvg cllrlepnaq
aqmyrltlrt
      961 skepvsrhlc ellaqqf
```

Nucleotide sequence: **ISOFORM 2**

NCBI Reference Sequence: NM_130787.2

LOCUS NM_130787

ACCESSION NM_130787

```
      1 cggctcagag ctccggaccg cgggcggagg ggaggggcag ggggcggtgc
cacggcctgc
       61 cagcccgccc gcccgcccgc cagccagccc tccccgcggc cggctcggct
ccttggcgct
      121 gcctggggtc ctttccgccc ggtccccgct tgccagcccc cgctgctctg
tgccctgtcc
      181 ggccaggcct ggagccgaca ccaccgccat catgccggcc gtgtccaagg
gcgatgggat
      241 gcgggggctc gcggtgttca tctccgacat ccggaactgt aagagcaaag
aggcggaaat
      301 taagagaatc aacaaggaac tggccaacat ccgctccaag ttcaaaggag
acaaagcctt
      361 ggatggctac agtaagaaaa aatatgtgtg taaactgctt ttcatcttcc
tgcttggcca
      421 tgacattgac tttgggcaca tggaggctgt gaatctgttg agttccaata
aatacacaga
      481 gaagcaaata ggttacctgt tcatttctgt gctggtgaac tcgaactcgg
agctgatccg
      541 cctcatcaac aacgccatca agaatgacct ggccagccgc aaccccacct
tcatgtgcct
      601 ggccctgcac tgcatcgcca acgtgggcag ccgggagatg ggcgaggcct
ttgccgctga
      661 catccccgc atcctggtgg ccggggacag catggacagt gtcaagcaga
gtgcggccct
      721 gtgcctcctt cgactgtaca aggcctcgcc tgacctggtg cccatgggcg
agtggacggc
      781 gcgtgtggta cacctgctca atgaccagca catgggtgtg gtcacggccg
ccgtcagcct
      841 catcacctgt ctctgcaaga agaacccaga tgacttcaag acgtgcgtct
ctctggctgt
      901 gtcgcgcctg agccggatcg tctcctctgc ctccaccgac ctccaggact
acacctacta
      961 cttcgtccca gcaccctggc tctcggtgaa gctcctgcgg ctgctgcagt
gctacccgcc
     1021 tccagaggat gcggctgtga aggggcggct ggtggaatgt ctggagactg
tgctcaacaa
     1081 ggcccaggag ccccccaaat ccaagaaggt gcagcattcc aacgccaaga
acgccatcct
     1141 cttcgagacc atcagcctca tcatccacta tgacagtgag cccaacctcc
tggttcgggc
     1201 ctgcaaccag ctgggccagt tcctgcagca ccgggagacc aacctgcgct
acctggccct
     1261 ggagagcatg tgcacgctgg ccagctccga gttctcccat gaagccgtca
agacgcacat
     1321 tgacaccgtc atcaatgccc tcaagacgga gcgggacgtc agcgtgcggc
agcgggcggc
     1381 tgacctcctc tacgccatgt gtgaccggag caatgccaag cagatcgtgt
cggagatgct
     1441 gcggtacctg gagacggcag actacgccat ccgcgaggag atcgtcctga
aggtggccat
```

1501 cctggccgag aagtacgccg tggactacag ctggtacgtg gacaccatcc
tcaacctcat
1561 ccgcattgcg ggcgactacg tgagtgagga ggtgtggtac cgtgtgctac
agatcgtcac
1621 caaccgtgat gacgtccagg gctatgccgc caagaccgtc tttgaggcgc
tccaggcccc
1681 tgcctgtcac gagaacatgg tgaaggttgg cggctacatc cttggggagt
ttgggaacct
1741 gattgctggg gacccccgct ccagcccccc agtgcagttc tccctgctcc
actccaagtt
1801 ccatctgtgc agcgtggcca cgcgggcgct gctgctgtcc acctacatca
agttcatcaa
1861 cctcttcccc gagaccaagg ccaccatcca gggcgtcctg cgggccggct
cccagctgcg
1921 caatgctgac gtggagctgc agcagcgagc cgtggagtac ctcaccctca
gctcagtggc
1981 cagcaccgac gtcctggcca cggtgctgga ggagatgccg cccttccccg
agcgcgagtc
2041 gtccatcctg gccaagctga aacgcaagaa ggggccaggg gccggcagcg
ccctggacga
2101 tggccggagg gaccccagca gcaacgacat caacgggggc atggagccca
cccccagcac
2161 tgtgtcgacg ccctcgccct ccgccgacct cctggggctg cgggcagccc
ctccccccggc
2221 agcacccccg gcttctgcag gagcagggaa ccttctggtg gacgtcttcg
atggcccggc
2281 cgcccagccc agcctggggc ccacccccga ggaggccttc ctcagcccag
gtcctgagga
2341 catcggccct cccattccgg aagccgatga gttgctgaat aagtttgtgt
gtaagaacaa
2401 cggggtcctg ttcgagaacc agctgctgca gatcggagtc aagtcagagt
tccgacagaa
2461 cctgggccgc atgtatctct tctatggcaa caagacctcg gtgcagttcc
agaatttctc
2521 acccactgtg gttcacccgg gagacctcca gactcagctg gctgtgcaga
ccaagcgcgt
2581 ggcggcgcag gtggacggcg gcgcgcaggt gcagcaggtg ctcaatatcg
agtgcctgcg
2641 ggacttcctg acgccccccgc tgctgtccgt gcgcttccgg tacggtggcg
ccccccaggc
2701 cctcaccctg aagctcccag tgaccatcaa caagttcttc cagcccaccg
agatggcggc
2761 ccaggatttc ttccagcgct ggaagcagct gagcctccct caacaggagg
cgcagaaaat
2821 cttcaaagcc aaccacccca tggacgcaga agttactaag gccaagcttc
tggggtttgg
2881 ctctgctctc ctggacaatg tggaccccaa ccctgagaac ttcgtggggg
cggggatcat
2941 ccagactaaa gccctgcagg tgggctgtct gcttcggctg gagcccaatg
cccaggccca
3001 gatgtaccgg ctgaccctgc gcaccagcaa ggagcccgtc tcccgtcacc
tgtgtgagct
3061 gctggcacag cagttctgag ccctggactc tgccccgggg gatgtggccg
gcactgggca
3121 gccccttgga ctgaggcagt tttggtggat gggggacctc cactggtgac
agagaagaca
3181 ccagggtttg ggggatgcct gggactttcc tccggccttt tgtattttta
tttttgttca
3241 tctgctgctg tttacattct gggggggttag ggggagtccc cctccctccc
tttccccccc

```
     3301 aagcacagag gggagagggg ccagggaagt ggatgtctcc tcccctccca
ccccaccctg
     3361 ttgtagcccc tcctaccccc tccccatcca ggggctgtgt attattgtga
gcgaataaac
     3421 agagagacgc taa
```

Protein sequence: **ISOFORM 2**

NCBI Reference Sequence: NP_570603.2

LOCUS NP_570603

ACCESSION NP_570603

```
      1 mpavskgdgm rglavfisdi rnckskeaei krinkelani rskfkgdkal
dgyskkkyvc
       61 kllfifllgh didfghmeav nllssnkyte kqigylfisv lvnsnselir
linnaikndl
      121 asrnptfmcl alhcianvgs remgeafaad iprilvagds mdsvkqsaal
cllrlykasp
      181 dlvpmgewta rvvhllndqh mgvvtaavsl itclckknpd dfktcvslav
srlsrivssa
      241 stdlqdytyy fvpapwlsvk llrllqcypp pedaavkgrl vecletvlnk
aqeppkskkv
      301 qhsnaknail fetisliihy dsepnllvra cnqlgqflqh retnlrylal
esmctlasse
      361 fsheavkthi dtvinalkte rdvsvrqraa dllyamcdrs nakqivseml
ryletadyai
      421 reeivlkvai laekyavdys wyvdtilnli riagdyvsee vwyrvlqivt
nrddvqgyaa
      481 ktvfealqap achenmvkvg gyilgefgnl iagdprsspp vqfsllhskf
hlcsvatral
      541 llstyikfin lfpetkatiq gvlragsqlr nadvelqqra veyltlssva
stdvlatvle
      601 emppfperes silaklkrkk gpgagsaldd grrdpssndi nggmeptpst
vstpspsadl
      661 lglraapppa appasagagn llvdvfdgpa aqpslgptpe eaflspgped
igppipeade
      721 llnkfvcknn gvlfenqllq igvksefrqn lgrmylfygn ktsvqfqnfs
ptvvhpgdlq
      781 tqlavqtkrv aaqvdggaqv qqvlnieclr dfltppllsv rfryggapqa
ltlklpvtin
      841 kffqptemaa qdffqrwkql slpqqeaqki fkanhpmdae vtkakllgfg
salldnvdpn
      901 penfvgagii qtkalqvgcl lrlepnaqaq myrltlrtsk epvsrhlcel laqqf
```

TTLL12

Official Symbol: **TTLL12**

Official Name: tubulin tyrosine ligase-like family, member 12

Gene ID: 23170

Organism: Homo sapiens

Other Aliases: dJ526I14.2

Other Designations: tubulin--tyrosine ligase-like protein 12

Nucleotide sequence:

NCBI Reference Sequence: NM_015140.3

LOCUS NM_015140

ACCESSION NM_015140

```
       1 gccgacggac ggcgggcggc ggcggcggtg gcggcgctgg agtcggcgcg
ggtgctggcg
      61 ccatggaggc cgagcggggt cccgagcgcc ggcctgcgga gcgtagcagc
ccgggccaga
     121 cgccggagga gggcgcgcag gccttggccg agttcgcggc gctgcacggc
ccggcgctgc
     181 gcgcttcggg ggtccccgaa cgttactggg gccgcctcct gcacaagctg
gagcacgagg
     241 ttttcgacgc tggggaagtg tttgggatca tgcaagtgga ggaggtagaa
gaggaggagg
     301 acgaggcagc ccgggaggtg cggaagcagc agcccaaccc ggggaacgag
ctgtgctaca
     361 aggtcatcgt gaccagggag agcgggctcc aggcagccca ccccaacagc
atcttcctca
     421 tcgaccacgc ctggacgtgc cgtgtggagc acgcgcgcca gcagctgcag
caggtgcccg
     481 ggctgctgca ccgcatggcc aacctgatgg gcattgagtt ccacggtgag
ctgcccagta
     541 cagaggctgt ggccctggtg ctggaggaga tgtggaagtt caaccagacc
taccagctgg
     601 cccatgggac agctgaggag aagatgccgg tgtggtatat catggacgag
ttcggttcgc
     661 ggatccagca cgcggacgtg cccagcttcg ccacggcacc cttcttctac
atgccgcagc
     721 aggtggccta cacgctgctg tggcccctga gggacctgga cactggcgag
gaggtgaccc
     781 gagactttgc ctacggagag acggacccc tgatccggaa gtgcatgctg
ctgccctggg
     841 cccccaccga catgctggac ctcagctctt gcacacccga gccgcccgcc
gagcactacc
     901 aggccattct ggaggaaaac aaggagaagc tgccacttga catcaacccc
gtggtgcacc
     961 cccacggcca catcttcaag gtctacacgg acgtgcagca ggtggccagc
agcctcaccc
    1021 acccgcgctt caccctcacc cagagtgagg cggacgccga catcctcttc
aacttctcac
    1081 acttcaagga ctacaggaaa ctcagccagg agaggccagg cgtgctgctg
aaccagttcc
    1141 cctgcgagaa cctgctgact gtcaaggact gcctggcctc catcgcgcgc
cgggcaggtg
    1201 gccccgaggg cccaccctgg ctgccccgaa ccttcaacct gcgcactgag
ctgccccagt
    1261 ttgtcagcta cttccagcag cgggaaaggt ggggcgagga caaccactgg
atctgcaagc
```

1321 cctggaacct ggcgcgcagc ctggacaccc acgtcaccaa gagcctgcac
agcatcatcc
1381 ggcaccgaga gagcaccccc aaggttgtgt ccaagtacat cgaaagtccc
gtgttgttcc
1441 ttcgagaaga cgtgggaaag gtcaagttcg acatccgcta catcgtgctg
ctgcggtcag
1501 tgaggcccct acggttgttc gtgtatgatg tgttctggct gcggttctcc
aaccgggcct
1561 ttgcactcaa cgacctggat gactacgaga agcacttcac ggtcatgaac
tatgacccgg
1621 atgtggtgct gaagcaggtg cactgtgaag agttcatccc cgagtttgag
aagcaatacc
1681 cagaatttcc ctggacggac gtccaggctg agatcttccg ggccttcacg
gagctgttcc
1741 aggtggcctg tgccaagcca ccacccctgg gcctctgcga ctacccctca
tcccgggcca
1801 tgtatgccgt cgacctcatg ctgaagtggg acaacggccc agatggaagg
cgggtgatgc
1861 agccgcagat cctggaggtg aacttcaacc ccgactgtga gcgagcctgc
aggtaccacc
1921 ccaccttctt caacgacgtc ttcagcacct tgtttctgga ccagcccggt
ggctgccacg
1981 ttacctgcct tgtctaggca ctcgctgtcc ccaaaacctg tgcttggggc
aggattccaa
2041 cctcagttct ctgagctgct tctgcaaagg cccccatgtc cctccccaca
ccggccctgg
2101 gcatagcctc agccccaggc ctctgtcctg ccgagccatc ctcccggcgc
cacactccgg
2161 gagcacagca tcctcctctc acctgtgggt cagagcagga cagtgatggt
gtccccaggg
2221 ctgagcacca ccccacgccc tgccctcacc cctcaccacc atctgtgcac
tgatgagtct
2281 ccagtttagc caagggcttt gttcctggca tggagaattt gttcctggct
gctgtgtttc
2341 caggggggtgc tgggggaagg gttccgtgga gcgagacaag gtgtcctcgg
gagcagggtt
2401 ccaccgggaa gcgtttggga gccctgtatc acacggggca ggcgggtttc
tcttccgggg
2461 tctctgctct tatgcatcag gacgaccccg ggacggctgt ggggccccac
actgcaccca
2521 cagggctcta tgcgacaggg gcccaggaac agcctgaggc caccacccag
caagcccgcc
2581 ttatcaccca ttccagctca cccagaacct tcaccagcaa acctcctgct
gaggtcctgg
2641 caggaggcca ccgtcttgtt accgtttcct tttcgtttgc tgagggtcac
agaccccaac
2701 agggaaatca gtatctgtct tcccagtggt tgccctgctc gccgggcact
ccacggggtc
2761 ccgcccttgt gtgagatggg ccaggatcct tcggcaaggg gcgcctgggg
ctggggctga
2821 ttgtgggcgg tggagcgcca gacagaaaag gattccaatg agaacttcag
gttaaagtca
2881 gatgccacct accagggtct acagtcaaaa tgttggcttt ttcttatttt
ttaatgtatg
2941 ggagaaaaat gtaaaattcc agttcttttc taattgtgtt tctgaaatta
ggagtcagct
3001 gccagcgttt ttgtgtggct gcagtgtgcc tgggcccagc tcacgggcag
tgggtggacc
3061 taactgccca ggcaggcgag agctacttcc agagccttcc agtgcatggg
agggcagggc

```
     3121 taggtgtagc ggtgtctcct ctttgaaatt aagaactatc tttcttgtag
caaagctgca
     3181 cctgatgatg ctgcctctcc tctctgtgtt gtctgggccc ttgtttacaa
gcacgcgtta
     3241 cccttcctga ggggagccat gctctagccc ctggagggcc tgttgcaggg
gcagggcggg
     3301 cccgtcgcct ttggcagctc ctggagagct gtggacatgc agtccccctc
agttcgtgct
     3361 gcaataaagg ccatcttctc ttaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
     3421 a
      //
```

Protein sequence:

NCBI Reference Sequence: NP_055955.1

LOCUS NP_055955

ACCESSION NP_055955

```
      1 meaergperr paersspgqt peegaqalae faalhgpalr asgvperywg
rllhklehev
       61 fdagevfgim qveeveeeed eaarevrkqq pnpgnelcyk vivtresglq
aahpnsifli
      121 dhawtcrveh arqqlqqvpg llhrmanlmg iefhgelpst eavalvleem
wkfnqtyqla
      181 hgtaeekmpv wyimdefgsr iqhadvpsfa tapffympqq vaytllwplr
dldtgeevtr
      241 dfaygetdpl irkcmllpwa ptdmldlssc tpeppaehyq aileenkekl
pldinpvvhp
      301 hghifkvytd vqqvasslth prftltqsea dadilfnfsh fkdyrklsqe
rpgvllnqfp
      361 cenlltvkdc lasiarragg pegppwlprt fnlrtelpqf vsyfqqrerw
gednhwickp
      421 wnlarsldth vtkslhsiir hrestpkvvs kyiespvlfl redvgkvkfd
iryivllrsv
      481 rplrlfvydv fwlrfsnraf alndlddyek hftvmnydpd vvlkqvhcee
fipefekqyp
      541 efpwtdvqae ifraftelfq vacakppplg lcdypssram yavdlmlkwd
ngpdgrrvmq
      601 pqilevnfnp dceracryhp tffndvfstl fldqpggchv tclv
//
```

FERMT2

Official Symbol: **FERMT2**

Official Name: fermitin family member 2

Gene ID: 10979

Organism: Homo sapiens

Other Aliases: KIND2, MIG2, PLEKHC1, UNC112, UNC112B, mig-2

Other Designations: PH domain-containing family C member 1; fermitin family homolog 2; kindlin 2; kindlin-2; mitogen inducible gene 2 protein; mitogen-inducible gene 2 protein; pleckstrin homology domain containing, family C (with FERM domain) member 1; pleckstrin homology domain containing, family C member 1; pleckstrin homology domain-containing family C member

Nucleotide sequence:

NCBI Reference Sequence: NM_001134999.1

LOCUS NM_001134999

ACCESSION NM_001134999

```
      1 gggtggagcg cggggagcca ggcgaggggc cgcgacgacg ggactccatt
agccgctccg
       61 gccacaggca gcgcttcgcc agccgaggaa ccggacgcgg acaccgccgc
cccgcgagcc
      121 tccagcccct cgcctgttgc cgcgcgagtc ccgggcccgg agcgctagga
gcgcgcggaa
      181 ggagccatgg ctctggacgg gataaggatg ccagatggct gctacgcgga
cgggacgtgg
      241 gaactgagtg tccatgtgac ggacctgaac cgcgatgtca ccctgagagt
gaccggcgag
      301 gtgcacattg gaggcgtgat gcttaagctg gtggagaaac tcgatgtaaa
aaaagattgg
      361 tctgaccatg ctctctggtg ggaaaagaag agaacttggc ttctgaagac
acattggacc
      421 ttagataagt atggtattca ggcagatgct aagcttcagt tcacccctca
gcacaaactg
      481 ctccgcctgc agcttcccaa catgaagtat gtgaaggtga aagtgaattt
ctctgataga
      541 gtcttcaaag ctgtttctga catctgtaag acttttaata tcagacaccc
cgaagaactt
      601 tctctcttaa agaaacccag agatccaaca aagaaaaaaa agaagaagct
agatgaccag
      661 tctgaagatg aggcacttga attagagggg cctcttatca ctcctggatc
aggaagtata
      721 tattcaagcc caggactgta tagtaaaaca atgaccccca cttatgatgc
tcatgatgga
      781 agccccttgt caccaacttc tgcttggttt ggtgacagtg ctttgtcaga
aggcaatcct
      841 ggtatacttg ctgtcagtca accaatcacg tcaccagaaa tcttggcaaa
aatgttcaag
      901 cctcaagctc ttcttgataa agcaaaaatc aaccaaggat ggcttgattc
ctcaagatct
      961 ctcatggaac aagatgtgaa ggaaaatgag gccttgctgc tccgattcaa
gtattacagc
     1021 ttttttgatt tgaatccaaa gtatgatgca atcagaatca atcagcttta
cgagcaggcc
     1081 aaatgggcca ttctcctgga agagattgaa tgcacagaag aagaaatgat
gatgtttgca
     1141 gccctgcagt atcatatcaa taagctgtca atcatgacat cagagaatca
tttgaacaac
```

1201 agtgacaaag aagttgatga agttgatgct gccctttcag acctggagat
tactctggaa
1261 gggggtaaaa cgtcaacaat tttgggtgac attacttcca ttcctgaact
tgctgactac
1321 attaaagttt tcaagccaaa aaagctgact ctgaaaggtt acaaacaata
ttggtgcacc
1381 ttcaaagaca catccatttc ttgttataag agcaaagaag aatccagtgg
cacaccagct
1441 catcagatga acctcagggg atgtgaagtt accccagatg taaacatttc
aggccaaaaa
1501 tttaacatta aactcctgat tccagttgca gaaggcatga atgaaatctg
gcttcgttgt
1561 gacaatgaaa aacagtatgc acactggatg gcagcctgca gattagcctc
caaaggcaag
1621 accatggcgg acagttctta caacttagaa gttcagaata ttctttcctt
tctgaagatg
1681 cagcatttaa acccagatcc tcagttaata ccagagcaga tcacgactga
tataactcct
1741 gaatgtttgg tgtctccccg ctatctaaaa aagtataaga acaagcagcc
aggctatata
1801 agagatttga taacagcgag aatcttggag gcccatcaga atgtagctca
gatgagtcta
1861 attgaagcca agatgagatt tattcaagct tggcagtcac tacctgaatt
tggcatcact
1921 cacttcattg caaggttcca agggggcaaa aaagaagaac ttattggaat
tgcatacaac
1981 agactgattc ggatggatgc cagcactgga gatgcaatta aaacatggcg
tttcagcaac
2041 atgaaacagt ggaatgtcaa ctgggaaatc aaaatggtca ccgtagagtt
tgcagatgaa
2101 gtacgattgt ccttcatttg tactgaagta gattgcaaag tggttcatga
attcattggt
2161 ggctacatat ttctctcaac acgtgcaaaa gaccaaaacg agagtttaga
tgaagagatg
2221 ttctacaaac ttaccagtgg ttgggtgtga ataggaatac tgtttaatga
aactccacgg
2281 ccataacaat atttaacttt aaaagctgtt tgttatatgc tgcttaataa
agtaagcttg
2341 aaatttatca ttttatcatg aaaacttctt tgccttacca gaccagttaa
tatgtgcact
2401 aaacaagcac gactattaat ctatcatgtt atgatataat aaacttgaat
ttgtcacaca
2461 ttccttaggg ccatgaattg aaaactgaaa tagtgggcaa atcaggaaca
aaccatcact
2521 gatttactga tttaagctag ccaaactgta agaaacaagc catctatttt
aaagctatcc
2581 agggcttaac ctatatgaac tctatttatc atgtctaatg catgtgattt
aatgtatgtt
2641 taatttgata tcatgtttta aaatatccta cttctggtag ccatttaatt
cctcccccta
2701 cccccaaata aatcaggcat gcaggaggcc tgatatttag taatgtcatt
gtgtttgacc
2761 ttgaaggaaa atgctattag tccgtcgtgc ttgatttgtt tttgtccttg
aataagcatg
2821 ttatgtatat tgtctcgtgt ttttatttt acaccatatt gtattacact
tttagtattc
2881 accagcataa tcactgtctg cctaaaatat gcaactcttt gcattacaat
atgaagtaaa
2941 gttctatgaa gtatgcattt tgtgtaacta atgtaaaaac acaaatttta
taaaattgta

```
     3001 cagtttttta aaaactactc acaactagca gatggcttaa atgtagcaat
ctctgcgtta
     3061 attaaatgcc tttaagagat ataattaacg tgcagtttta atatctacta
aattaagaat
     3121 gacttcatta tgatcatgat ttgccacaat gtccttaact ctaatgcctg
gactggccat
     3181 gttctagtct gttgcgctgt tacaatctgt attggtgcta gtcagaaaat
tcctagctca
     3241 catagcccaa aagggtgcga gggagaggtg gattaccagt attgttcaat
aatccatggt
     3301 tcaaagactg tataaatgca ttttatttta aataaaagca aaacttttat
ttaataaaaa
     3361 aaaaaaaaaa aa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001128471.1

LOCUS NP_001128471

ACCESSION NP_001128471

```
      1 maldgirmpd gcyadgtwel svhvtdlnrd vtlrvtgevh iggvmlklve
kldvkkdwsd
       61 halwwekkrt wllkthwtld kygiqadakl qftpqhkllr lqlpnmkyvk
vkvnfsdrvf
      121 kavsdicktf nirhpeelsl lkkprdptkk kkkklddqse dealelegpl
itpgsgsiys
      181 spglysktmt ptydahdgsp lsptsawfgd salsegnpgi lavsqpitsp
eilakmfkpq
      241 alldkakinq gwldssrslm eqdvkeneal llrfkyysff dlnpkydair
inqlyeqakw
      301 ailleeiect eeemmmfaal qyhinklsim tsenhlnnsd kevdevdaal
sdleitlegg
      361 ktstilgdit sipeladyik vfkpkkltlk gykqywctfk dtsiscyksk
eessgtpahq
      421 mnlrgcevtp dvnisgqkfn ikllipvaeg mneiwlrcdn ekqyahwmaa
crlaskgktm
      481 adssynlevq nilsflkmqh lnpdpqlipe qittditpec lvsprylkky
knkqpgyird
      541 litarileah qnvaqmslie akmrfiqawq slpefgithf iarfqggkke
eligiaynrl
      601 irmdastgda iktwrfsnmk qwnvnweikm vtvefadevr lsfictevdc
kvvhefiggy
      661 iflstrakdq nesldeemfy kltsgwv
```

ANXA6

Official Symbol: **ANXA6**

Official Name: annexin A6

Gene ID: 309

Organism: Homo sapiens

Other Aliases: ANX6, CBP68

Other Designations: 67 kDa calelectrin; CPB-II; annexin VI (p68); annexin-6; calcium-binding protein p68; calelectrin; calphobindin II; calphobindin-II; chromobindin-20; lipocortin VI; p68; p70

Nucleotide sequence:

NCBI Reference Sequence: NM_001155.4

LOCUS NM_001155

ACCESSION NM_001155

```
      1 agaggggtgg ggtggaggag ggaggcgggc gcgccggatt ggcctctgcg
cgccacgtgt
       61 ccggctcgga gcccacggct gtcctcccgg tccgccccgc gctgcggttg
ctgctgggct
      121 aacgggctcc gatccagcga gcgctgcgtc ctcgagtccc tgcgcccgtg
cgtccgtctg
      181 cgacccgagg cctccgctgc gcgtggattc tgctgcgaac cggagaccat
ggccaaacca
      241 gcacagggtg ccaagtaccg gggctccatc catgacttcc caggctttga
ccccaaccag
      301 gatgccgagg ctctgtacac tgccatgaag ggctttggca gtgacaagga
ggccatactg
      361 gacataatca cctcacggag caacaggcag aggcaggagg tctgccagag
ctacaagtcc
      421 ctctacggca aggacctcat tgctgattta aagtatgaat tgacgggcaa
gtttgaacgg
      481 ttgattgtgg gcctgatgag gccacctgcc tattgtgatg ccaaagaaat
taaagatgcc
      541 atctcgggca ttggcactga tgagaagtgc ctcattgaga tcttggcttc
ccggaccaat
      601 gagcagatgc accagctggt ggcagcatac aaagatgcct acgagcggga
cctggaggct
      661 gacatcatcg gcgacacctc tggccacttc cagaagatgc ttgtggtcct
gctccaggga
      721 accagggagg aggatgacgt agtgagcgag gacctggtac aacaggatgt
ccaggaccta
      781 tacgaggcag gggaactgaa atggggaaca gatgaagccc agttcattta
catcttggga
      841 aatcgcagca agcagcatct tcggttggtg ttcgatgagt atctgaagac
cacagggaag
      901 ccgattgaag ccagcatccg aggggagctg tctggggact ttgagaagct
aatgctggcc
      961 gtagtgaagt gtatccggag cacccggaa tattttgctg aaaggctctt
caaggctatg
     1021 aagggcctgg ggactcggga caacaccctg atccgcatca tggtctcccg
tagtgagttg
```

1081 gacatgctcg acattcggga gatcttccgg accaagtatg agaagtccct
ctacagcatg
1141 atcaagaatg acacctctgg cgagtacaag aagactctgc tgaagctgtc
tgggggagat
1201 gatgatgctg ctggccagtt cttcccggag gcagcgcagg tggcctatca
gatgtgggaa
1261 cttagtgcag tggcccgagt agagctgaag ggaactgtgc gcccagccaa
tgacttcaac
1321 cctgacgcag atgccaaagc gctgcggaaa gccatgaagg gactcgggac
tgacgaagac
1381 acaatcatcg atatcatcac gcaccgcagc aatgtccagc ggcagcagat
ccggcagacc
1441 ttcaagtctc actttggccg ggacttaatg actgacctga agtctgagat
ctctggagac
1501 ctggcaaggc tgattctggg gctcatgatg ccaccggccc attacgatgc
caagcagttg
1561 aagaaggcca tggagggagc cggcacagat gaaaaggctc ttattgaaat
cctggccact
1621 cggaccaatg ctgaaatccg ggccatcaat gaggcctata aggaggacta
tcacaagtcc
1681 ctggaggatg ctctgagctc agacacatct ggccacttca ggaggatcct
catttctctg
1741 gccacggggc atcgtgagga gggaggagaa aacctggacc aggcacggga
agatgcccag
1801 gtggctgctg agatcttgga aatagcagac acacctagtg gagacaaaac
ttccttggag
1861 acacgtttca tgacgatcct gtgtacccgg agctatccgc acctccggag
agtcttccag
1921 gagttcatca agatgaccaa ctatgacgtg gagcacacca tcaagaagga
gatgtctggg
1981 gatgtcaggg atgcatttgt ggccattgtt caaagtgtca agaacaagcc
tctcttcttt
2041 gccgacaaac tttacaaatc catgaagggt gctggcacag atgagaagac
tctgaccagg
2101 atcatggtat cccgcagtga gattgacctg ctcaacatcc ggagggaatt
cattgagaaa
2161 tatgacaagt ctctccacca agccattgag ggtgacacct ccggagactt
cctgaaggcc
2221 ttgctggctc tctgtggtgg tgaggactag ggccacagct ttggcgggca
cttctgccaa
2281 gaaatggtta tcagcaccag ccgccatggc caagcctgat tgttccagct
ccagagacta
2341 aggaaggggc aggggtgggg ggaggggttg ggttgggctc ttatcttcag
tggagcttag
2401 gaaacgctcc cactcccacg ggccatcgag ggcccagcac ggctgagcgg
ctgaaaaacc
2461 gtagccatag atcctgtcca cctccactcc cctctgaccc tcaggctttc
ccagcttcct
2521 ccccttgcta cagcctctgc cctggtttgg gctatgtcag atccaaaaac
atcctgaacc
2581 tctgtctgta aaatgagtag tgtctgtact ttgaatgagg gggttggtgg
caggggccag
2641 ttgaatgtgc tgggcggggt ggtgggaagg atagtaaatg tgctggggca
aactgacaaa
2701 tcttcccatc catttcacca cccatctcca tccaggccgc gctagagtac
tggaccagga
2761 atttggatgc ctgggttcaa atctgcatct gccatgcact tgtttctgac
cttaggccag
2821 ccccttccc tccctgagtc tctattttct tatctacaat gagacagttg
gacaaaaaaa

```
     2881 tcttggcttc ccttctaaca ttaacttcct aaagtatgcc tccgattcat
tcccttgaca
     2941 ctttttattt ctaaggaaga aataaaaaga gatacacaaa cacataaaca
caaaaaaaaa
     3001 aa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001146.2

LOCUS NP_001146

ACCESSION NP_001146

```
      1 makpaqgaky rgsihdfpgf dpnqdaealy tamkgfgsdk eaildiitsr
snrqrqevcq
       61 sykslygkdl iadlkyeltg kferlivglm rppaycdake ikdaisgigt
dekclieila
      121 srtneqmhql vaaykdayer dleadiigdt sghfqkmlvv llqgtreedd
vvsedlvqqd
      181 vqdlyeagel kwgtdeaqfi yilgnrskqh lrlvfdeylk ttgkpieasi
rgelsgdfek
      241 lmlavvkcir stpeyfaerl fkamkglgtr dntlirimvs rseldmldir
eifrtkyeks
      301 lysmikndts geykktllkl sggdddaagq ffpeaaqvay qmwelsavar
velkgtvrpa
      361 ndfnpdadak alrkamkglg tdedtiidii thrsnvqrqq irqtfkshfg
rdlmtdlkse
      421 isgdlarlil glmmppahyd akqlkkameg agtdekalie ilatrtnaei
raineayked
      481 yhksledals sdtsghfrri lislatghre eggenldqar edaqvaaeil
eiadtpsgdk
      541 tsletrfmti lctrsyphlr rvfqefikmt nydvehtikk emsgdvrdaf
vaivqsvknk
      601 plffadklyk smkgagtdek tltrimvsrs eidllnirre fiekydkslh
qaiegdtsgd
      661 flkallalcg ged
//
```

PSMD4

Official Symbol: **PSMD4**

Official Name: proteasome (prosome, macropain) 26S subunit, non-ATPase, 4

Gene ID: 5710

Organism: Homo sapiens

Other Aliases: RP11-126K1.1, AF, AF-1, ASF, MCB1, Rpn10, S5A, pUB-R5

Other Designations: 26S proteasome non-ATPase regulatory subunit 4; 26S proteasome regulatory subunit S5A; RPN10 homolog; S5a/antisecretory factor protein; angiocidin; antisecretory factor 1; multiubiquitin chain-binding protein

Nucleotide sequence:

NCBI Reference Sequence: NM_002810.2

LOCUS NM_002810

ACCESSION NM_002810

```
      1 aattggagga gttgttgtta ggccgtcccg gagacccggt cgggagggag
gaaggtggca
       61 agatggtgtt ggaaagcact atggtgtgtg tggacaacag tgagtatatg
cggaatggag
      121 acttcttacc caccaggctg caggcccagc aggatgctgt caacatagtt
tgtcattcaa
      181 agacccgcag caaccctgag aacaacgtgg gccttatcac actggctaat
gactgtgaag
      241 tgctgaccac actcacccca gacactggcc gtatcctgtc caagctacat
actgtccaac
      301 ccaagggcaa gatcaccttc tgcacgggca tccgcgtggc ccatctggct
ctgaagcacc
      361 gacaaggcaa gaatcacaag atgcgcatca ttgcctttgt gggaagccca
gtggaggaca
      421 atgagaagga tctggtgaaa ctggctaaac gcctcaagaa ggagaaagta
aatgttgaca
      481 ttatcaattt tggggaagag gaggtgaaca cagaaaagct gacagccttt
gtaaacacgt
      541 tgaatggcaa agatggaacc ggttctcatc tggtgacagt gcctcctggg
cccagtttgg
      601 ctgatgctct catcagttct ccgattttgg ctggtgaagg tggtgccatg
ctgggtcttg
      661 gtgccagtga ctttgaattt ggagtagatc ccagtgctga tcctgagctg
gccttggccc
      721 ttcgtgtatc tatggaagag cagcggcagc ggcaggagga ggaggcccgg
cgggcagctg
      781 cagcttctgc tgctgaggcc gggattgcta cgactgggac tgaagactca
gacgatgccc
      841 tgctgaagat gaccatcagc cagcaagagt ttggccgcac tgggcttcct
gacctaagca
      901 gtatgactga ggaagagcag attgcttatg ccatgcagat gtccctgcag
ggagcagagt
      961 ttggccaggc ggaatcagca gacattgatg ccagctcagc tatggacaca
tctgagccag
     1021 ccaaggagga ggatgattac gacgtgatgc aggaccccga gttccttcag
agtgtcctag
     1081 agaacctccc aggtgtggat cccaacaatg aagccattcg aaatgctatg
ggctccctgg
     1141 cctcccaggc caccaaggac ggcaagaagg acaagaagga ggaagacaag
aagtgagact
     1201 ggagggaaag ggtagctgag tctgcttagg ggactgcatg ggaagcacgg
aatatagggt
     1261 tagatgtgtg ttatctgtaa ccattacagc ctaaataaag cttggcaact
ttttttcctt
     1321 ttttgcttca aa
```

//

Protein sequence:

NCBI Reference Sequence: NP_002801.1

LOCUS NP_002801

ACCESSION NP_002801

```
      1 mvlestmvcv dnseymrngd flptrlqaqq davnivchsk trsnpennvg
litlandcev
       61 lttltpdtgr ilsklhtvqp kgkitfctgi rvahlalkhr qgknhkmrii
afvgspvedn
      121 ekdlvklakr lkkekvnvdi infgeeevnt ekltafvntl ngkdgtgshl
vtvppgpsla
      181 dalisspila geggamlglg asdfefgvdp sadpelalal rvsmeeqrqr
qeeearraaa
      241 asaaeagiat tgtedsddal lkmtisqqef grtglpdlss mteeeqiaya
mqmslqgaef
      301 gqaesadida ssamdtsepa keeddydvmq dpeflqsvle nlpgvdpnne
airnamgsla
      361 sqatkdgkkd kkeedkk
//
```

COTL1

Official Symbol: **COTL1**

Official Name: coactosin-like 1 (Dictyostelium)

Gene ID: 23406

Organism: Homo sapiens

Other Aliases: CLP

Other Designations: coactosin-like protein

Nucleotide sequence:

NCBI Reference Sequence: NM_021149.2

LOCUS NM_021149

ACCESSION NM_021149

```
      1 cgcgctcgca gctcgcaggc gccgcgtagc cgtcgccacc gccgccagcc
cgtgcgccct
       61 cggcgcgtac ccgccgcgct cccatccccg ccgccggcca ggggcgcgct
cggccgcccc
```

```
      121 ggacagtgtc ccgctgcggc tccgcggcga tggccaccaa gatcgacaaa
gaggcttgcc
      181 gggcggcgta caacctggtg cgcgacgacg gctcggccgt catctgggtg
acttttaaat
      241 atgacggctc caccatcgtc cccggcgagc agggagcgga gtaccagcac
ttcatccagc
      301 agtgcacaga tgacgtccgg ttgtttgcct tcgtgcgctt caccaccggg
gatgccatga
      361 gcaagaggtc caagtttgcc ctcatcacgt ggatcggtga gaacgtcagc
gggctgcagc
      421 gcgccaaaac cgggacggac aagaccctgg tgaaggaggt cgtacagaat
ttcgctaagg
      481 agtttgtgat cagtgatcgg aaggagctgg aggaagattt catcaagagc
gagctgaaga
      541 aggcgggggg agccaattac gacgcccaga cggagtaacc ccagcccccg
ccacaccacc
      601 ccttgccaaa gtcatctgcc tgctccccgg gggagaggac cgccggcctc
agctactagc
      661 ccaccagccc accagggaga aaagaagcca tgagaggcag cgcccgccac
cctgtgtcca
      721 cagcccccac cttcccgctt cccttagaac cctgccgtgt cctatctcat
gacgctcatg
      781 gaacctcttt ctttgatctt ctttttcttt tctccccctc ttttttgttc
taaagaaaag
      841 tcattttgat gcaaggtcct gcctgccatc agatccgagg tgcctcctgc
agtgacccct
      901 tttcctggca tttctcttcc acgcgacgag gtctgcctag tgagatctgc
atgacctcac
      961 gttgctttcc agagcccggg cctattttgc catctcagtt ttcctggacc
ctgcttcctg
     1021 tgtaccactg aggggcagct gggccaggag ctgtgcccgg tgcctgcagc
cttcataagc
     1081 acacacgtcc attccctact aaggcccaga cctcctggta tctgccccgg
gctccctcat
     1141 cccacctcca tccggagttg cctaagatgc atgtccagca taggcaggat
tgctcggtgg
     1201 tgagaaggtt aggtccggct cagactgaat aagaagagat aaaatttgcc
ttaaaactta
     1261 cctggcagtg gctttgctgc acggtctgaa accacctgtt cccaccctct
tgaccgaaat
     1321 ttccttgtga cacagagaag ggcaaaggtc tgagcccaga gttgacggag
ggagtatttc
     1381 agggttcact tcaggggctc ccaaagcgac aagatcgtta gggagagagg
cccagggtgg
     1441 ggactgggaa tttaaggaga gctgggaacg gatcccttag gttcaggaag
cttctgtgta
     1501 agctgcgagg atggcttggg ccgaagggtt gctctgcccg ccgcgctagc
tgtgagctga
     1561 gcaaagccct gggctcacag caccccaaaa gcctgtggct tcagtcctgc
gtctgcacca
     1621 cacattcaaa aggatcgttt tgttttgttt ttaaagaaag gtgagattgg
cttggttctt
     1681 catgagcaca tttgatatag ctctttttct gtttttcctt gctcatttcg
ttttggggaa
     1741 gaaatctgta ctgtattggg attgtaaaga acatctctgc actcagacag
tttacagaaa
     1801 taaatgtttt ttttgttttt cagaaaaaaa aaaaaaaaaa aaaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_066972.1

LOCUS NP_066972

ACCESSION NP_066972

```
      1 matkidkeac raaynlvrdd gsaviwvtfk ydgstivpge qgaeyqhfiq
qctddvrlfa
       61 fvrfttgdam skrskfalit wigenvsglq raktgtdktl vkevvqnfak
efvisdrkel
      121 eedfikselk kagganydaq te
            //
```

ST13

Official Symbol: **ST13**

Official Name: suppression of tumorigenicity 13

Gene ID: 6767

Organism: Homo sapiens

Other Aliases: AAG2, FAM10A1, FAM10A4, HIP, HOP, HSPABP, HSPABP1, P48, PRO0786, SNC6

Other Designations: Hsp70-interacting protein; aging-associated protein 2; heat shock 70kD protein binding protein; hsc70-interacting protein; progesterone receptor-associated p48 protein; putative tumor suppressor **ST13**; renal carcinoma antigen NY-REN-33; suppression of tumorigenicity 13 protein

Nucleotide sequence:

NCBI Reference Sequence: NM_003932.3

LOCUS NM_003932

ACCESSION NM_003932

```
      1 gggtgggagg agccagcggc cggggaggtt ctagtctgtt ctgtcttgcg
gcagccgccc
       61 ccttctgcgc ggtcacgccg agccagcgcc tgggcctgga accgggccgt
agccccccca
      121 gtttcgccca ccacctccct accatggacc cccgcaaagt gaacgagctt
cgggcctttg
      181 tgaaaatgtg taagcaggat ccgagcgttc tgcacaccga ggaaatgcgc
ttcctgaggg
      241 agtgggtgga gagcatgggt ggtaaagtac cacctgctac tcagaaagct
aaatcagaag
```

301 aaaataccaa ggaagaaaaa cctgatagta agaaggtgga ggaagactta
aaggcagacg
361 aaccatcaag tgaggaaagt gatctagaaa ttgataaaga aggtgtgatt
gaaccagaca
421 ctgatgctcc tcaagaaatg ggagatgaaa atgcggagat aacggaggag
atgatggatc
481 aggcaaatga taaaaaagtg gctgctattg aagccctaaa tgatggtgaa
ctccagaaag
541 ccattgactt attcacagat gccatcaagc tgaatcctcg cttggccatt
ttgtatgcca
601 agagggccag tgtcttcgtc aaattacaga agccaaatgc tgccatccga
gactgtgaca
661 gagccattga aataaatcct gattcagctc agccttacaa gtggcggggg
aaagcacaca
721 gacttctagg ccactgggaa gaagcagccc atgatcttgc ccttgcctgt
aaattggatt
781 atgatgaaga tgctagtgca atgctgaaag aagttcaacc tagggcacag
aaaattgcag
841 aacatcggag aaagtatgag cgaaaacgtg aagagcgaga gatcaaagaa
agaatagaac
901 gagttaagaa ggctcgagaa gagcatgaga gagcccagag ggaggaagaa
gccagacgac
961 agtcaggagc tcagtatggc tcttttccag gtggctttcc tgggggaatg
cctggtaatt
1021 ttcccggagg aatgcctgga atgggagggg gcatgcctgg aatggctgga
atgcctggac
1081 tcaatgaaat tcttagtgat ccagaggttc ttgcagccat gcaggatcca
gaagttatgg
1141 tggctttcca ggatgtggct cagaacccag caaatatgtc aaaataccag
agcaacccaa
1201 aggttatgaa tctcatcagt aaattgtcag ccaaatttgg aggtcaagcg
taatgtcctt
1261 ctgataaata aagcccttgc tgaaggaaaa gcaacctaga tcaccttatg
gatgtcgcaa
1321 taatacaaac cagtgtacct ctgaccttct catcaagaga gctggggtgc
tttgaagata
1381 atccctaccc ctctccccca aatgcagctg aagcatttta cagtggtttg
ccattagggt
1441 attcattcag ataatgtttt cctactagga attacaaact ttaaacactt
tttaaatctt
1501 caaaatattt aaaacaaatt taaagggcct gttaattctt atatttttct
ttactaatca
1561 ttttggattt ttttctttga attattggca gggaatatac ttatgtatgg
aagattactg
1621 ctctgagtga aataaaagtt attagtgcga ggcaaacata actcatttga
ggataaagtt
1681 tgtgttggat atgtggttcc tgatgcattt tgacttgtct ttttaaatgc
tttatctttt
1741 tctttaaaga tttatttcaa taaaactaat tgggaccacc cgtatttcag
taggacctgg
1801 gtagggattg gaagtacttg gcagggcagc agcaatcttg ctgtgtttga
tataacatgc
1861 atccttgggc aggttgccct taaatcttac actgtggtga agggatgttt
tttttgtaat
1921 gctgcagtag agttggagta cttagttctc ttgttgtcca gtatatctaa
taagtgtttt
1981 tcatattatt tccacgtaag ggaaataagg tagtactttt ctttttatat
ttctatgctt
2041 aaaattctct ttcctagtca aaaattgccc aaatctgtgt ttgctttctg
cttgctacat

```
     2101 ttgtctccct tacttttctt gagctaaaga caggcttttt ccaccggcat
catcactgct
     2161 atcatcatta acagcgtaat tatacaagca tatttaatgc tgagtttaat
ttaatatgta
     2221 atacatatgg taattgtagg gtaataccca caacaactgt agtttcttac
ttggccaaga
     2281 gaatgcttat ttaagtgtta gacttccatt ctggcaaaat cttgccttat
cagaagacat
     2341 tggaaagagg gattcccttt ggtgtttggt cttctactta gaaaaaccta
ttgcagttag
     2401 tttatcttgt agtattcatc tttgtattct gaagataagg tttgaattaa
attgatacac
     2461 acagagggga accgattttt tttatccaat gtgaattata aatgagataa
tccacagtta
     2521 ttcattgtgg agttgttgag actatgaaag actcattgtc tttgtattca
gctcttaaat
     2581 agtgtaacta tatccccacc tctgcttgct ttctttccct cccctccaat
gataaagaaa
     2641 atgataaatt ttctgttgtg cattcaattc ttattttaaa taagactaag
tataggcatt
     2701 gtacctgaca ttgctacgtt tctaccagtg tttcaattta aagtgctagt
gtttaaaaac
     2761 attttcaagg gataaggcct tctgtacttt gcttatttga agaatcagtg
gtaggagcag
     2821 tgaagtaaat tctatggagt acatttctaa aataccacat ttctgaaatc
ataaataagt
     2881 ttattcaggt tctaaccctt tgctgtacac aagcagacag aaatgcatct
gttacataaa
     2941 tgagaaaaag ctattatgct gatggagcat gctttttaaa tcctttaaaa
acactcacca
     3001 tataaacttg catttgagct tgtgtgttct tttgttaatg tgtagagttc
tcctttctcg
     3061 aaattgccag tgtgtacttg gcttaactca agaacagttt cttctggatt
ccttatttga
     3121 tttatttaac ctaattatat tctaatattg caaatattac cataagtggg
taaaagtaaa
     3181 attcctcttc tgaaaaaaaa aaaaaaaaaa aaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_003923.2

LOCUS NP_003923

ACCESSION NP_003923

```
       1 mdprkvnelr afvkmckqdp svlhteemrf lrewvesmgg kvppatqkak
seentkeekp
       61 dskkveedlk adepsseesd leidkegvie pdtdapqemg denaeiteem
mdqandkkva
      121 aiealndgel qkaidlftda iklnprlail yakrasvfvk lqkpnaaird
cdraieinpd
      181 saqpykwrgk ahrllghwee aahdlalack ldydedasam lkevqpraqk
iaehrrkyer
      241 kreereiker iervkkaree heraqreeea rrqsgaqygs fpggfpggmp
gnfpggmpgm
```

```
      301 gggmpgmagm pglneilsdp evlaamqdpe vmvafqdvaq npanmskyqs
npkvmnlisk
      361 lsakfggqa
//
```

Gene

Official Symbol: **SRSF2** (also known as **SFRS2**)

Official Name: serine/arginine-rich splicing factor 2

Gene ID: 6427

Organism: Homo sapiens

Other Aliases: PR264, SC-35, SC35, **SFRS2**, SFRS2A, SRp30b

Other Designations: SR splicing factor 2; splicing component, 35 kDa; splicing factor SC35; splicing factor, arginine/serine-rich 2

Nucleotide sequence:

NCBI Reference Sequence: NM_001195427.1

LOCUS NM_001195427

ACCESSION NM_001195427

```
      1 agaaggtttc atttccgggt ggcgcgggcg ccattttgtg aggagcgata
taaacgggcg
       61 cagaggccgg ctgcccgccc agttgttact caggtgcgct agcctgcgga
gcccgtccgt
      121 gctgttctgc ggcaaggcct ttcccagtgt ccccacgcgg aaggcaactg
cctgagaggc
      181 gcggcgtcgc accgcccaga gctgaggaag ccggcgccag ttcgcggggc
tccgggccgc
      241 cactcagagc tatgagctac ggccgccccc ctcccgatgt ggagggtatg
acctccctca
      301 aggtggacaa cctgacctac cgcacctcgc ccgacacgct gaggcgcgtc
ttcgagaagt
      361 acgggcgcgt cggcgacgtg tacatcccgc gggaccgcta caccaaggag
tcccgcggct
      421 tcgccttcgt tcgctttcac gacaagcgcg acgctgagga cgctatggat
gccatggacg
      481 gggccgtgct ggacggccgc gagctgcggg tgcaaatggc gcgctacggc
cgcccccccgg
      541 actcacacca cagccgccgg ggaccgccac cccgcaggta cgggggcggt
ggctacggac
      601 gccggagccg cagccctagg cggcgtcgcc gcagccgatc ccggagtcgg
agccgttcca
      661 ggtctcgcag ccgatctcgc tacagccgct cgaagtctcg gtcccgcact
cgttctcgat
```

```
      721 ctcggtcgac ctccaagtcc agatccgcac gaaggtccaa gtccaagtcc
tcgtcggtct
      781 ccagatctcg ttcgcggtcc aggtcccggt ctcggtccag gagtcctccc
ccagtgtcca
      841 agagggaatc caaatccagg tcgcgatcga agagtccccc caagtctcct
gaagaggaag
      901 gagcggtgtc ctcttaagaa aatgatgtat cggcaagcag tgtaaacgga
ggacttgggg
      961 aaaaaggacc acatagtcca tcgaagaaga gtccttggaa caagcaactg
gctattgaaa
     1021 aggttatttt gtaacatttg tctaactttt tacttgttta agctttgcct
cagttggcaa
     1081 acttcatttt atgtgccatt ttgttgctgt tattcaaatt tcttgtaatt
tagtgaggtg
     1141 aacgacttca gatttcatta ttggatttgg atatttgagg taaaatttca
ttttgttata
     1201 tagtgctgac tttttttgtt tgaaattaaa cagattggta acctaatttg
tggcctcctg
     1261 acttttaagg aaaacgtgtg cagccattac acacagccta aagctgtcaa
gagattgact
     1321 cggcattgcc ttcattcctt aaaattaaaa acctacaaaa gttggtgtaa
atttgtatat
     1381 gttatttacc ttcagatcta aatggtaatc tgaacccaaa tttgtataaa
gacttttcag
     1441 gtgaaaagac ttgatttttt gaaaggattg tttatcaaac acaattctaa
tctcttctct
     1501 tatgtatttt tgtgcactag gcgcagttgt gtagcagttg agtaatgctg
gttagctgtt
     1561 aaggtggcgt gttgcagtgc agagtgcttg gctgtttcct gttttctccc
gattgctcct
     1621 gtgtaaagat gccttgtcgt gcagaaacaa atggctgtcc agtttattaa
aatgcctgac
     1681 aactgcactt ccagtcaccc gggccttgca tataaataac ggagcataca
gtgagcacat
     1741 ctagctgatg ataaatacac ctttttttcc ctcttccccc taaaaatggt
aaatctgatc
     1801 atatctacat gtatgaactt aacatggaaa atgttaagga agcaaatggt
tgtaactttg
     1861 taagtactta taacatggtg tatctttttg cttatgaata ttctgtatta
taaccattgt
     1921 ttctgtagtt taattaaaac attttcttgg tgttagcttt tctcagaaaa
aaaaaaaaaa
     1981 aaaaaaaaaa aaaaaaaaaa aaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001182356.1

LOCUS NP_001182356

ACCESSION NP_001182356

```
      1 msygrpppdv egmtslkvdn ltyrtspdtl rrvfekygrv gdvyiprdry
tkesrgfafv
       61 rfhdkrdaed amdamdgavl dgrelrvqma rygrppdshh srrgppprry
ggggygrrsr
```

```
      121 sprrrrrsrs rsrsrsrsrs rsrysrsksr srtrsrsrst sksrsarrsk
sksssvsrsr
      181 srsrsrsrsr spppvskres ksrsrskspp kspeeegavs s
//
```

HNRNPH1

Official Symbol: **HNRNPH1**

Official Name: heterogeneous nuclear ribonucleoprotein H1 (H)

Gene ID: 3187

Organism: Homo sapiens

Other Aliases: HNRPH, HNRPH1, hnRNPH

Other Designations: heterogeneous nuclear ribonucleoprotein H

Nucleotide sequence:

NCBI Reference Sequence: NM_001257293.1

LOCUS NM_001257293

ACCESSION NM_001257293

```
      1 acaagggacc ttatttaggt tgcgcaggcg cccgctggcc atttcgtctt
agccacgcag
       61 aagtcgcgtg tctaggtgag tcgcggtggg tcctcgcttg cagttcagcg
accacgtttg
      121 tttcgacgcc ggaccgcgta agagacgatg atgttgggca cggaaggtgg
agagggattc
      181 gtggtgaagg tccggggctt gccctggtct tgctcggccg atgaagtgca
gaggtttttt
      241 tctgactgca aaattcaaaa tggggctcaa ggtattcgtt tcatctacac
cagagaaggc
      301 agaccaagtg gcgaggcttt tgttgaactt gaatcagaag atgaagtcaa
attggccctg
      361 aaaaaagaca gagaaactat gggacacaga tatgttgaag tattcaagtc
aaacaacgtt
      421 gaaatggatt gggtgttgaa gcatactggt ccaaatagtc ctgacacggc
caatgatggc
      481 tttgtacggc ttagaggact tccctttgga tgtagcaagg aagaaattgt
tcagttcttc
      541 tcagggttgg aaatcgtgcc aaatgggata acattgccgg tggacttcca
ggggaggagt
      601 acgggggagg ccttcgtgca gtttgcttca caggaaatag ctgaaaaggc
tctaaagaaa
      661 cacaaggaaa gaatagggca caggtatatt gaaatcttta agagcagtag
agctgaagtt
```

```
      721 agaactcatt atgatccacc acgaaagctt atggccatgc agcggccagg
tccttatgac
      781 agacctgggg ctggtagagg gtataacagc attggcagag gagctggctt
tgagaggatg
      841 aggcgtggtg cttatggtgg aggctatgga ggctatgatg attacaatgg
ctataatgat
      901 ggctatggat ttgggtcaga tagatttgga agagacctca attactgttt
ttcaggaatg
      961 tctgatcaca gatacgggga tggtggctct actttccaga gcacaacagg
acactgtgta
     1021 cacatgcggg gattacctta cagagctact gagaatgaca tttataattt
tttttcaccg
     1081 ctcaaccctg tgagagtaca cattgaaatt ggtcctgatg gcagagtaac
tggtgaagca
     1141 gatgtcgagt tcgcaactca tgaagatgct gtggcagcta tgtcaaaaga
caaagcaaat
     1201 atgcaacaca gatatgtaga actcttcttg aattctacag caggagcaag
cggtggtgct
     1261 tacgaacaca gatatgtaga actcttcttg aattctacag caggagcaag
cggtggtgct
     1321 tatggtagcc aaatgatggg aggcatgggc ttgtcaaacc agtccagcta
cgggggccca
     1381 gccagccagc agctgagtgg gggttacgga ggcggctacg gtggccagag
cagcatgagt
     1441 ggatacgacc aagttttaca ggaaaactcc agtgattttc aatcaaacat
tgcataggta
     1501 accaaggagc agtgaacagc agctactaca gtagtggaag ccgtgcatct
atgggcgtga
     1561 acggaatggg agggttgtct agcatgtcca gtatgagtgg tggatgggga
atgtaattga
     1621 tcgatcctga tcactgactc ttggtcaacc ttttttttttt ttttttttttt
ttctttaaga
     1681 aaacttcagt ttaacagttt ctgcaataca agcttgtgat ttatgcttac
tctaagtgga
     1741 aatcaggatt gttatgaaga cttaaggccc agtatttttg aatacaatac
tcatctagga
     1801 tgtaacagtg aagctgagta aactataact gttaaactta agttccagct
tttctcaagt
     1861 tagttatagg atgtacttaa gcagtaagcg tatttaggta aaagcagttg
aattatgtta
     1921 aatgttgccc tttgccacgt taaattgaac actgttttgg atgcatgttg
aaagacatgc
     1981 ttttatttttt ttgtaaaaca atataggagc tgtgtctact attaaaagtg
aaacattttg
     2041 gcatgtttgt taattctagt ttcatttaat aacctgtaag gcacgtaagt
ttaagctttt
     2101 ttttttttta agttaatggg aaaaatttga gacgcaatac caatacttag
gattttggtc
     2161 ttggtgtttg tatgaaattc tgaggccttg atttaaatct ttcattgtat
tgtgatttcc
     2221 ttttaggtat attgcgctaa gtgaaacttg tcaaataaat cctcctttta
aaaactgcaa
     2281 aaaaaaaaaa aaaaaaaaaa aaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001244222.1

LOCUS NP_001244222

ACCESSION NP_001244222

```
      1 mmlgteggeg fvvkvrglpw scsadevqrf fsdckiqnga qgirfiytre
grpsgeafve
       61 lesedevkla lkkdretmgh ryvevfksnn vemdwvlkht gpnspdtand
gfvrlrglpf
      121 gcskeeivqf fsgleivpng itlpvdfqgr stgeafvqfa sqeiaekalk
khkerighry
      181 ieifkssrae vrthydpprk lmamqrpgpy drpgagrgyn sigrgagfer
mrrgaygggy
      241 ggyddyngyn dgygfgsdrf grdlnycfsg msdhrygdgg stfqsttghc
vhmrglpyra
      301 tendiynffs plnpvrvhie igpdgrvtge advefathed avaamskdka
nmqhryvelf
      361 lnstagasgg ayehryvelf lnstagasgg aygsqmmggm glsnqssygg
pasqqlsggy
      421 gggyggqssm sgydqvlqen ssdfqsnia
//
```

Gene

Official Symbol: **IQGAP1**

Official Name: IQ motif containing GTPase activating protein 1

Gene ID: 8826

Organism: Homo sapiens

Other Aliases: HUMORFA01, SAR1, p195

Other Designations: RasGAP-like with IQ motifs; ras GTPase-activating-like protein IQGAP1

Nucleotide sequence:

NCBI Reference Sequence: NM_003870.3

LOCUS NM_003870

ACCESSION NM_003870

```
      1 ggaccccggc aagcccgcgc acttggcagg agctgtagct accgccgtcc
gcgcctccaa
       61 ggtttcacgg cttcctcagc agagactcgg gctcgtccgc catgtccgcc
gcagacgagg
      121 ttgacgggct gggcgtggcc cggccgcact atggctctgt cctggataat
gaaagactta
      181 ctgcagagga gatggatgaa aggagacgtc agaacgtggc ttatgagtac
ctttgtcatt
      241 tggaagaagc gaagaggtgg atggaagcat gcctagggga agatctgcct
cccaccacag
```

301 aactggagga ggggcttagg aatggggtct accttgccaa actggggaac
ttcttctctc
361 ccaaagtagt gtccctgaaa aaaatctatg atcgagaaca gaccagatac
aaggcgactg
421 gcctccactt tagacacact gataatgtga ttcagtggtt gaatgccatg
gatgagattg
481 gattgcctaa gatttttac ccagaaacta cagatatcta tgatcgaaag
aacatgccaa
541 gatgtatcta ctgtatccat gcactcagtt tgtacctgtt caagctaggc
ctggcccctc
601 agattcaaga cctatatgga aaggttgact tcacagaaga agaaatcaac
aacatgaaga
661 ctgagttgga gaagtatggc atccagatgc ctgcctttag caagattggg
ggcatcttgg
721 ctaatgaact gtcagtggat gaagccgcat tacatgctgc tgttattgct
attaatgaag
781 ctattgaccg tagaattcca gccgacacat ttgcagcttt gaaaaatccg
aatgccatgc
841 ttgtaaatct tgaagagccc ttggcatcca cttaccagga tatactttac
caggctaagc
901 aggacaaaat gacaaatgct aaaaacagga cagaaaactc agagagagaa
agagatgttt
961 atgaggagct gctcacgcaa gctgaaattc aaggcaatat aaacaaagtc
aatacatttt
1021 ctgcattagc aaatatcgac ctggctttag aacaaggaga tgcactggcc
ttgttcaggg
1081 ctctgcagtc accagccctg gggcttcgag gactgcagca acagaatagc
gactggtact
1141 tgaagcagct cctgagtgat aaacagcaga agagacagag tggtcagact
gacccctgc
1201 agaaggagga gctgcagtct ggagtggatg ctgcaaacag tgctgcccag
caatatcaga
1261 gaagattggc agcagtagca ctgattaatg ctgcaatcca gaagggtgtt
gctgagaaga
1321 ctgttttgga actgatgaat cccgaagccc agctgcccca ggtgtatcca
tttgccgccg
1381 atctctatca gaaggagctg gctaccctgc agcgacaaag tcctgaacat
aatctcaccc
1441 acccagagct ctctgtcgca gtggagatgt tgtcatcggt ggccctgatc
aacagggcat
1501 tggaatcagg agatgtgaat acagtgtgga agcaattgag cagttcagtt
actggtctta
1561 ccaatattga ggaagaaaac tgtcagaggt atctcgatga gttgatgaaa
ctgaaggctc
1621 aggcacatgc agagaataat gaattcatta catggaatga tatccaagct
tgcgtggacc
1681 atgtgaacct ggtggtgcaa gaggaacatg agaggatttt agccattggt
ttaattaatg
1741 aagccctgga tgaaggtgat gcccaaaaga ctctgcaggc cctacagatt
cctgcagcta
1801 aacttgaggg agtccttgca gaagtggccc agcattacca agacacgctg
attagagcga
1861 agagagagaa agcccaggaa atccaggatg agtcagctgt gttatggttg
gatgaaattc
1921 aaggtggaat ctggcagtcc aacaaagaca cccaagaagc acagaagttt
gccttaggaa
1981 tctttgccat taatgaggca gtagaaagtg gtgatgttgg caaaacactg
agtgcccttc
2041 gctcccctga tgttggcttg tatggagtca tccctgagtg tggtgaaact
taccacagtg

2101 atcttgctga agccaagaag aaaaaactgg cagtaggaga taataacagc
aagtgggtga
2161 agcactgggt aaaaggtgga tattattatt accacaatct ggagacccag
gaaggaggat
2221 gggatgaacc tccaaatttt gtgcaaaatt ctatgcagct ttctcgggag
gagatccaga
2281 gttctatctc tggggtgact gccgcatata accgagaaca gctgtggctg
gccaatgaag
2341 gcctgatcac caggctgcag gctcgctgcc gtggatactt agttcgacag
gaattccgat
2401 ccaggatgaa tttcctgaag aaacaaatcc ctgccatcac ctgcattcag
tcacagtgga
2461 gaggatacaa gcagaagaag gcatatcaag atcggttagc ttacctgcgc
tcccacaaag
2521 atgaagttgt aaagattcag tccctggcaa ggatgcacca agctcgaaag
cgctatcgag
2581 atcgcctgca gtacttccgg gaccatataa atgacattat caaaatccag
gcttttattc
2641 gggcaaacaa agctcgggat gactacaaga ctctcatcaa tgctgaggat
cctcctatgg
2701 ttgtggtccg aaaatttgtc cacctgctgg accaaagtga ccaggatttt
caggaggagc
2761 ttgaccttat gaagatgcgg gaagaggtta tcaccctcat tcgttctaac
cagcagctgg
2821 agaatgacct caatctcatg gatatcaaaa ttggactgct agtgaaaaat
aagattacgt
2881 tgcaggatgt ggtttcccac agtaaaaaac ttaccaaaaa aaataaggaa
cagttgtctg
2941 atatgatgat gataaataaa cagaagggag gtctcaaggc tttgagcaag
gagaagagag
3001 agaagttgga agcttaccag cacctgtttt atttattgca aaccaatccc
acctatctgg
3061 ccaagctcat ttttcagatg ccccagaaca agtccaccaa gttcatggac
tctgtaatct
3121 tcacactcta caactacgcg tccaaccagc gagaggagta cctgctcctg
cggctcttta
3181 agacagcact ccaagaggaa atcaagtcga aggtagatca gattcaagag
attgtgacag
3241 gaaatcctac ggttattaaa atggttgtaa gtttcaaccg tggtgcccgt
ggccagaatg
3301 ccctgagaca gatcttggcc ccagtcgtga aggaaattat ggatgacaaa
tctctcaaca
3361 tcaaaactga ccctgtggat atttacaaat cttgggttaa tcagatggag
tctcagacag
3421 gagaggcaag caaactgccc tatgatgtga cccctgagca ggcgctagct
catgaagaag
3481 tgaagacacg gctagacagc tccatcagga acatgcgggc tgtgacagac
aagtttctct
3541 cagccattgt cagctctgtg gacaaaatcc cttatgggat gcgcttcatt
gccaaagtgc
3601 tgaaggactc gttgcatgag aagttccctg atgctggtga ggatgagctg
ctgaagatta
3661 ttggtaactt gctttattat cgatacatga atccagccat tgttgctcct
gatgcctttg
3721 acatcattga cctgtcagca ggaggccagc ttaccacaga ccaacgccga
aatctgggct
3781 ccattgcaaa aatgcttcag catgctgctt ccaataagat gtttctggga
gataatgccc
3841 acttaagcat cattaatgaa tatctttccc agtcctacca gaaattcaga
cggtttttcc

3901 aaactgcttg tgatgtccca gagcttcagg ataaatttaa tgtggatgag
tactctgatt
3961 tagtaaccct caccaaacca gtaatctaca tttccattgg tgaaatcatc
aacacccaca
4021 ctctcctgtt ggatcaccag gatgccattg ctccggagca caatgatcca
atccacgaac
4081 tgctggacga cctcggcgag gtgcccacca tcgagtccct gataggggaa
agctctggca
4141 atttaaatga cccaaataag gaggcactgg ctaagacgga agtgtctctc
accctgacca
4201 acaagttcga cgtgcctgga gatgagaatg cagaaatgga tgctcgaacc
atcttactga
4261 atacaaaacg tttaattgtg gatgtcatcc ggttccagcc aggagagacc
ttgactgaaa
4321 tcctagaaac accagccacc agtgaacagg aagcagaaca tcagagagcc
atgcagagac
4381 gtgctatccg tgatgccaaa acacctgaca agatgaaaaa gtcaaaatct
gtaaaggaag
4441 acagcaacct cactcttcaa gagaagaaag agaagatcca gacaggttta
aagaagctaa
4501 cagagcttgg aaccgtggac ccaaagaaca aataccagga actgatcaac
gacattgcca
4561 gggatattcg gaatcagcgg aggtaccgac agaggagaaa ggccgaacta
gtgaaactgc
4621 aacagacata cgctgctctg aactctaagg ccacctttta tggggagcag
gtggattact
4681 ataaaagcta tatcaaaacc tgcttggata acttagccag caagggcaaa
gtctccaaaa
4741 agcctaggga aatgaaagga aagaaaagca aaaagatttc tctgaaatat
acagcagcaa
4801 gactacatga aaaaggagtt cttctggaaa ttgaggacct gcaagtgaat
cagtttaaaa
4861 atgttatatt tgaaatcagt ccaacagaag aagttggaga cttcgaagtg
aaagccaaat
4921 tcatgggagt tcaaatggag acttttatgt tacattatca ggacctgctg
cagctacagt
4981 atgaaggagt tgcagtcatg aaattatttg atagagctaa agtaaatgtc
aacctcctga
5041 tcttccttct caacaaaaag ttctacggga agtaattgat cgtttgctgc
cagcccagaa
5101 ggatgaagga aagaagcacc tcacagctcc tttctaggtc cttctttcct
cattggaagc
5161 aaagacctag ccaacaacag cacctcaatc tgatacactc ccgatgccac
atttttaact
5221 cctctcgctc tgatgggaca tttgttaccc ttttttcata gtgaaattgt
gtttcaggct
5281 tagtctgacc tttctggttt cttcattttc ttccattact taggaaagag
tggaaactcc
5341 actaaaattt ctctgtgttg ttacagtctt agaggttgca gtactatatt
gtaagctttg
5401 gtgtttgttt aattagcaat agggatggta ggattcaaat gtgtgtcatt
tagaagtgga
5461 agctattagc accaatgaca taaatacata caagacacac aactaaaatg
tcatgttatt
5521 aacagttatt aggttgtcat ttaaaaataa agttccttta tatttctgtc
ccatcaggaa
5581 aactgaagga tatggggaat cattggttat cttccattgt gtttttcttt
atggacagga
5641 gctaatggaa gtgacagtca tgttcaaagg aagcatttct agaaaaaagg
agataatgtt

```
     5701 tttaaatttc attatcaaac ttgggcaatt ctgtttgtgt aactccccga
ctagtggatg
     5761 ggagagtccc attgctaaaa ttcagctact cagataaatt cagaatgggt
caaggcacct
     5821 gcctgttttt gttggtgcac agagattgac ttgattcaga gagacaattc
actccatccc
     5881 tatggcagag gaatgggtta gccctaatgt agaatgtcat tgtttttaaa
actgttttat
     5941 atcttaagag tgccttatta aagtatagat gtatgtctta aaatgtgggt
gataggaatt
     6001 ttaaagattt atataatgca tcaaaagcct tagaataaga aaagcttttt
ttaaattgct
     6061 ttatctgtat atctgaactc ttgaaactta tagctaaaac actaggattt
atctgcagtg
     6121 ttcagggaga taattctgcc tttaattgtc taaaacaaaa acaaaaccag
ccaacctatg
     6181 ttacacgtga gattaaaacc aatttttcc ccatttttc tccttttttc
tcttgctgcc
     6241 cacattgtgc ctttatttta tgagccccag ttttctgggc ttagtttaaa
aaaaaaatca
     6301 agtctaaaca ttgcatttag aaagcttttg ttcttggata aaaagtcata
cactttaaaa
     6361 aaaaaaaaaa ctttttccag gaaaatatat tgaaatcatg ctgctgagcc
tctattttct
     6421 ttctttgatg ttttgattca gtattctttt atcataaatt tttagcattt
aaaaattcac
     6481 tgatgtacat taagccaata aactgcttta atgaataaca aactatgtag
tgtgtcccta
     6541 ttataaatgc attggagaag tatttttatg agactcttta ctcaggtgca
tggttacagc
     6601 ccacagggag gcatggagtg ccatggaagg attcgccact acccagacct
tgtttttttgt
     6661 tgtattttgg aagacaggtt ttttaaagaa acattttcct cagattaaaa
gatgatgcta
     6721 ttacaactag cattgcctca aaaactggga ccaaccaaag tgtgtcaacc
ctgtttcctt
     6781 aaaagaggct atgaatccca aaggccacat ccaagacagg caataatgag
cagagtttac
     6841 agctccttta ataaaatgtg tcagtaattt taaggtttat agttccctca
acacaattgc
     6901 taatgcagaa tagtgtaaaa tgcgcttcaa gaatgttgat gatgatgata
tagaattgtg
     6961 gctttagtag cacagaggat gccccaacaa actcatggcg ttgaaaccac
acagttctca
     7021 ttactgttat ttattagctg tagcattctc tgtctcctct ctctcctcct
ttgaccttct
     7081 cctcgaccag ccatcatgac atttaccatg aatttacttc ctcccaagag
tttggactgc
     7141 ccgtcagatt gttgctgcac atagttgcct ttgtatctct gtatgaaata
aaaggtcatt
     7201 tgttcatgtt aaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_003861.1

LOCUS NP_003861

ACCESSION NP_003861

```
      1 msaadevdgl gvarphygsv ldnerltaee mderrrqnva yeylchleea
krwmeaclge
       61 dlppttelee glrngvylak lgnffspkvv slkkiydreq trykatglhf
rhtdnviqwl
      121 namdeiglpk ifypettdiy drknmprciy cihalslylf klglapqiqd
lygkvdftee
      181 einnmktele kygiqmpafs kiggilanel svdeaalhaa viaineaidr
ripadtfaal
      241 knpnamlvnl eeplastyqd ilyqakqdkm tnaknrtens ererdvyeel
ltqaeiqgni
      301 nkvntfsala nidlaleqgd alalfralqs palglrglqq qnsdwylkql
lsdkqqkrqs
      361 gqtdplqkee lqsgvdaans aaqqyqrrla avalinaaiq kgvaektvle
lmnpeaqlpq
      421 vypfaadlyq kelatlqrqs pehnlthpel svavemlssv alinralesg
dvntvwkqls
      481 ssvtgltnie eencqrylde lmklkaqaha ennefitwnd iqacvdhvnl
vvqeeheril
      541 aiglineald egdaqktlqa lqipaakleg vlaevaqhyq dtlirakrek
aqeiqdesav
      601 lwldeiqggi wqsnkdtqea qkfalgifai neavesgdvg ktlsalrspd
vglygvipec
      661 getyhsdlae akkkklavgd nnskwvkhwv kggyyyyhnl etqeggwdep
pnfvqnsmql
      721 sreeiqssis gvtaaynreq lwlaneglit rlqarcrgyl vrqefrsrmn
flkkqipait
      781 ciqsqwrgyk qkkayqdrla ylrshkdevv kiqslarmhq arkryrdrlq
yfrdhindii
      841 kiqafirank arddyktlin aedppmvvvr kfvhlldqsd qdfqeeldlm
kmreevitli
      901 rsnqqlendl nlmdikigll vknkitlqdv vshskkltkk nkeqlsdmmm
inkqkgglka
      961 lskekrekle ayqhlfyllq tnptylakli fqmpqnkstk fmdsviftly
nyasnqreey
     1021 lllrlfktal qeeikskvdq iqeivtgnpt vikmvvsfnr gargqnalrq
ilapvvkeim
     1081 ddkslniktd pvdiykswvn qmesqtgeas klpydvtpeq alaheevktr
ldssirnmra
     1141 vtdkflsaiv ssvdkipygm rfiakvlkds lhekfpdage dellkiignl
lyyrymnpai
     1201 vapdafdiid lsaggqlttd qrrnlgsiak mlqhaasnkm flgdnahlsi
ineylsqsyq
     1261 kfrrffqtac dvpelqdkfn vdeysdlvtl tkpviyisig eiinthtlll
dhqdaiapeh
     1321 ndpihelldd lgevptiesl igessgnlnd pnkealakte vsltltnkfd
vpgdenaemd
     1381 artillntkr livdvirfqp getlteilet patseqeaeh qramqrrair
daktpdkmkk
     1441 sksvkedsnl tlqekkekiq tglkkltelg tvdpknkyqe lindiardir
nqrryrqrrk
     1501 aelvklqqty aalnskatfy geqvdyyksy iktcldnlas kgkvskkpre
mkgkkskkis
     1561 lkytaarlhe kgvlleiedl qvnqfknvif eispteevgd fevkakfmgv
qmetfmlhyq
     1621 dllqlqyegv avmklfdrak vnvnllifll nkkfygk
//
```

GPSN2

Official Symbol: **TECR** (also known as **GPSN2**)

Official Name: trans-2,3-enoyl-CoA reductase

Gene ID: 9524

Organism: Homo sapiens

Other Aliases: **GPSN2**, MRT14, SC2, TER

Other Designations: glycoprotein, synaptic 2; synaptic glycoprotein SC2

Nucleotide sequence:

NCBI Reference Sequence: NM_138501.5

LOCUS NM_138501

ACCESSION NM_138501 XM_001132190 XM_001132196

```
      1 ggaggggcgg ggcggacgca gagccgcgtt tagtctatcg ctgcggttgc
gagcgctgta
       61 gggagcctgt gctgtgccgc gcagttaggc agcagcagcc gcggagcagt
agccgccgtg
      121 ggagggagcc atgaagcatt acgaggtgga gattctggac gcaaagacaa
gggagaagct
      181 gtgtttcttg gacaaggtgg agccccacgc caccattgcg gagatcaaga
acctcttcac
      241 taagacccat ccgcagtggt accccgcccg ccagtccctc cgcctggacc
ccaagggcaa
      301 gtccctgaag gatgaggatg ttctgcagaa gctgcccgtg ggcaccacgg
ccacactgta
      361 cttccgggac ctgggggccc agatcagctg ggtgacggtc ttcctaacag
agtacgcggg
      421 gccccttttc atctacctgc tcttctactt ccgagtgccc ttcatctatg
gccacaaata
      481 tgactttacg tccagtcggc atacagtggt gcacctcgcc tgcatctgtc
actcattcca
      541 ctacatcaag cgcctgctgg agacgctctt cgtgcaccgc ttctcccatg
gcactatgcc
      601 tttgcgcaac atcttcaaga actgcaccta ctactggggc ttcgccgcgt
ggatggccta
      661 ttacatcaat caccctctct acactccccc tacctacgga gctcagcagg
tgaaactggc
      721 gctcgccatc tttgtgatct gccagctcgg caacttctcc atccacatgg
ccctgcggga
      781 cctgcggccc gctgggtcca agacgcggaa gatcccatac cccaccaaga
acccccttcac
      841 gtggctcttc ctgctggtgt cctgccccaa ctacacctac gaggtggggt
cctggatcgg
      901 tttcgccatc atgacgcagt gtctcccagt ggccctgttc tccctggtgg
gcttcaccca
      961 gatgaccatc tgggccaagg gcaagcaccg cagctacctg aaggagttcc
gggactaccc
     1021 gcccctgcgc atgcccatca tccccttcct gctctgagcg ctcacccctg
ctgaggctca
```

```
     1081 gcccctcaac ccggtggcat tctgggggag gagtggggcc cacagctctc
cagcacccgg
     1141 aataaagccc gcctgcccca gtcggaaaaa aaaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_612510.1

LOCUS NP_612510

ACCESSION NP_612510 XP_001132190 XP_001132196

```
      1 mkhyeveild aktreklcfl dkvephatia eiknlftkth pqwyparqsl
rldpkgkslk
       61 dedvlqklpv gttatlyfrd lgaqiswvtv flteyagplf iyllfyfrvp
fiyghkydft
      121 ssrhtvvhla cichsfhyik rlletlfvhr fshgtmplrn ifknctyywg
faawmayyin
      181 hplytpptyg aqqvklalai fvicqlgnfs ihmalrdlrp agsktrkipy
ptknpftwlf
      241 llvscpnyty evgswigfai mtqclpvalf slvgftqmti wakgkhrsyl
kefrdypplr
      301 mpiipfll
//
```

EHD2

Official Symbol: **EHD2**

Official Name: EH-domain containing 2

Gene ID: 30846

Organism: Homo sapiens

Other Aliases: PAST2

Other Designations: EH domain containing 2; EH domain-containing protein 2; PAST homolog 2

Nucleotide sequence:

NCBI Reference Sequence: NM_014601.3

LOCUS NM_014601

ACCESSION NM_014601

```
      1 ttttgagggg ggcgcctcgt cccgcccctc cctcctgtcc tccctcccgt
cctccccgct
```

```
       61 ccgggcccca cccggctcag acggctccgg acgggaccgc gagcacaggc
cgctccgcgg
      121 gcgcttcgga tcctcgcggg accccaccct ctcccagcct gcccagcccg
ctgcagccgc
      181 cagcgcgccc cgtcggcagc tctccatctg cacgtctctc cgtgaacccc
gtgagcggtg
      241 tgcagccacc atgttcagct ggctgaagcg gggcggggca cggggccagc
agcccgaggc
      301 catccgcacg gtgacctcgg ccctcaagga gctgtaccgc acgaagctgc
tgccgctgga
      361 ggagcactac cgctttgggg ccttccactc gccggccctg gaggacgcag
acttcgacgg
      421 caagcccatg gtgctggtgg ccggccagta cagcacgggc aagaccagct
tcatccagta
      481 cctgctggag caggaggtgc ccggctcccg cgtggggcct gagcccacca
ccgactgctt
      541 tgtggccgtc atgcacgggg acactgaggg caccgtgccc ggcaacgccc
tcgtcgtgga
      601 cccggacaag cccttccgca aactcaaccc tttcggaaac accttcctca
acaggttcat
      661 gtgtgcccag ctccctaatc aggtcctgga gagcatcagc atcatcgaca
ccccgggtat
      721 cctgtcgggt gccaagcaga gagtgagccg cggctacgac ttcccggccg
tgctgcgctg
      781 gttcgcggag cgcgtggacc tcatcatcct gctctttgat gcgcacaagc
tggagatctc
      841 ggacgagttc tcagaggcca tcggcgcgtt gcggggccat gaggacaaga
tccgcgtggt
      901 gctcaacaag gccgacatgg tggagacgca gcagctgatg cgcgtctacg
gcgcgctcat
      961 gtgggcgctg ggcaaggtgg tgggcacgcc cgaggtgctg cgcgtctaca
tcggctcctt
     1021 ctggtcccag cccctcctcg tgcccgacaa ccggcgcctc ttcgagctgg
aggagcagga
     1081 cctcttccgc gacatccagg gcctgccccg gcacgcagcc ttgcgcaagc
tcaacgacct
     1141 ggtgaagagg gcccggctgg tgcgagttca cgcttacatc atcagctacc
tgaagaagga
     1201 gatgccctct gtgtttggga aggagaacaa gaagaagcag ctgatcctca
aactgcccgt
     1261 catctttgcg aagattcagc tggaacatca catctcccct ggggactttc
ctgattgcca
     1321 gaaaatgcag gagctgctga tggcgcacga cttcaccaag tttcactcgc
tgaagccgaa
     1381 gctgctagag gcactggacg agatgctgac gcacgacatc gccaagctca
tgcccctgct
     1441 gcggcaggag gagctggaga gcaccgaggt gggcgtgcag gggggcgctt
ttgagggcac
     1501 ccacatgggc ccgtttgtgg agcggggacc tgacgaggcc atggaggacg
gcgaggaggg
     1561 ctcggacgac gaggccgagt gggtggtgac caaggacaag tccaaatacg
acgagatctt
     1621 ctacaacctg gcgcctgccg acggcaagct gagcggctcc aaggccaaga
cctggatggt
     1681 ggggaccaag ctccccaact cagtgctggg gcgcatctgg aagctcagcg
atgtggaccg
     1741 cgacggcatg ctggatgatg aggagttcgc gctggccagc cacctcatcg
aggccaagct
     1801 ggaaggccac gggctgcccg ccaacctgcc ccgtcgcctg gtgccaccct
ccaagcgacg
```

1861 ccacaagggc tccgccgagt gagccggccc ccctcccatg gccctgctgt
ggctccccag
1921 ctccagtcgg ctgcacgcac acccctgctc cggctcacac acgccctgcc
tgccctccct
1981 gcccagctgt aaggaccggg ggtctccctc ctcactaccg ccagacaccc
cggtggaagc
2041 atttagaggg gaccacggga gggacaaggc ttctctgtcc gcccttcaca
cctccagcct
2101 cacgttcact taggcacatc acacacacac tggcacacgc aggcatccat
ccatccgtca
2161 ttcattcaaa tatttattga gcacctacta tgtgcccagc cctgttctag
gcactgggca
2221 ttaccataga gaacaaaata gacaaataca tctgccctca tggaaggtga
cgttcccagg
2281 agagggcacc tacacagtca cgcaaacaca cactaattcc tggcagggcc
cccagcccct
2341 cccctggctg agcagccctg tggctgaaat gactagcaga taaacagacc
cccttctgct
2401 ccgcttcctc ctgcccagcc aggcaacacc ctcaaccggc tccatcacat
cctcaggtct
2461 cgggaccatg gggggctcag aggggagaca cacctactgc ttcctcagat
gggcccctcc
2521 gcagcccctt cccttgctcg gggaaagccc ccaattctgc ccacacccat
ttatttcctt
2581 ccttccttcc ttcttttctt tccttccttc cttctttttt gtttttgccc
ccaattctgc
2641 ccatacccat ttctttcttt ccttccttcc ttctttttttg tttttgcccc
cagttctgtc
2701 cacacccctt ccctttcctg tcctgtcctt tctttctttt ttgatagaat
cttgctctgt
2761 cgcccaggct ggagtgcagt ggtgagatct cagctcactg caacctccac
ctcctgggtt
2821 gaagtgattc tcgtgcctca gcctcctgag tagctgggac tgcaggcacg
cgccaccacg
2881 cccagctaat ttttgtattt gagtagagac ggggtttcac catgttggcc
aggctggtct
2941 cgaactccgc atctcaggtg atctgctcgc cttggcctcc caaagtgatg
ggattacagg
3001 catgagccac cgtgcccggc ttcacaccca tttctttaaa aaggatcccg
tagcaggcag
3061 aaaagcccct tccatcctgc tcctctgata ctgtgccccc ttggagatat
ttccgtcctc
3121 cacccacgtg tctgtggctg gaactgccca gcctgctcct ggccccctgg
aagcctcccc
3181 acagctggta atctggactt aaggattgct gggccaccgc ctctctgcct
accaccattc
3241 catatttaag tggagcccct acgtagaaag gccccggggc tttattttag
tctccttttc
3301 agggatgtcg tgggcggggg agggggttct tggtgctaca gccctctccc
cacccctaaa
3361 gggacgccga cgctgtttgc tgccttcacc acatattagt gcttgaccct
ggcaggggac
3421 cccatggaaa agatggggaa gagcaaaata catggagacg acgcaccctc
caggatgctc
3481 gctgggattc ccacgcccac cactgtcccc caccccatgg ctgggagggg
cctctgaacg
3541 gaacagtgtc cccacagagc gaataaagcc aaggcttctt cccaaaaaaa
aaaaaaaaaa
3601 a
//

Protein sequence:

NCBI Reference Sequence: NP_055416.2

LOCUS NP_055416

ACCESSION NP_055416

```
      1 mfswlkrgga rgqqpeairt vtsalkelyr tkllpleehy rfgafhspal
edadfdgkpm
       61 vlvagqystg ktsfiqylle qevpgsrvgp epttdcfvav mhgdtegtvp
gnalvvdpdk
      121 pfrklnpfgn tflnrfmcaq lpnqvlesis iidtpgilsg akqrvsrgyd
fpavlrwfae
      181 rvdliillfd ahkleisdef seaigalrgh edkirvvlnk admvetqqlm
rvygalmwal
      241 gkvvgtpevl rvyigsfwsq pllvpdnrrl feleeqdlfr diqglprhaa
lrklndlvkr
      301 arlvrvhayi isylkkemps vfgkenkkkq lilklpvifa kiqlehhisp
gdfpdcqkmq
      361 ellmahdftk fhslkpklle aldemlthdi aklmpllrqe elestevgvq
ggafegthmg
      421 pfvergpdea medgeegsdd eaewvvtkdk skydeifynl apadgklsgs
kaktwmvgtk
      481 lpnsvlgriw klsdvdrdgm lddeefalas hlieaklegh glpanlprrl
vppskrrhkg
      541 sae
//
```

UGP2

Official Symbol: **UGP2**

Official Name: UDP-glucose pyrophosphorylase 2

Gene ID: 7360

Organism: Homo sapiens

Other Aliases: UDPG, UDPGP2, UGP1, UGPP1, UGPP2, pHC379

Other Designations: UDP-glucose diphosphorylase; UDP-glucose pyrophosphorylase 1; UDPGP; UGPase 2; UTP--glucose-1-phosphate uridylyltransferase; UTP--glucose-1-phosphate uridylyltransferase 2; UTP-glucose-1-phosphate uridyltransferase; uridyl diphosphate glucose pyrophosphorylase 2

Nucleotide sequence:

NCBI Reference Sequence: NM_001001521.1

LOCUS NM_001001521

ACCESSION NM_001001521

```
      1 tttccgcatt gaaggggctg ctccgaatgg agggggaggg gaggtgttta
ggagaaagta
       61 ggggctgtgg gtgtcgggag ccggctgacg ggtggacaag ggggggttag
cagctgggct
      121 gcgaccgtta gggaggggct caaggtgtgc atgtgtgagg gaagagagag
agagagaagg
      181 gcgcctcaga ggtgactttc agcctgcgag ccttcttccc ggggcgccat
aaacgccccc
      241 aatttcccag ctgctaaagg aagaggaaga tcttagcaaa gcaatgtctc
aagatggtgc
      301 ttctcagttc caagaagtca ttcggcaaga gctagaatta tctgtgaaga
aggaactaga
      361 aaaaatactc accacagcat catcacatga atttgagcac accaaaaaag
acctggatgg
      421 atttcggaag ctatttcata gatttttgca agaaaagggg ccttctgtgg
attggggaaa
      481 aatccagaga ccccctgaag attcgattca accctatgaa aagataaagg
ccaggggctt
      541 gcctgataat atatcttccg tgttgaacaa actagtggtg gtgaaactca
atggtggttt
      601 gggaaccagc atgggctgca aaggccctaa aagtctgatt ggtgtgagga
atgagaatac
      661 ctttctggat ctgactgttc agcaaattga acatttgaat aaaacctaca
atacagatgt
      721 tcctcttgtt ttaatgaact cttttaacac ggatgaagat accaaaaaaa
tactacagaa
      781 gtacaatcat tgtcgtgtga aaatctacac tttcaatcaa agcaggtacc
cgaggattaa
      841 taaagaatct ttacttcctg tagcaaagga cgtgtcttac tcaggggaaa
atacagaagc
      901 ttggtaccct ccaggtcatg gtgatattta cgccagtttc tacaactctg
gattgcttga
      961 tacctttata ggagaaggca aagagtatat ttttgtgtct aacatagata
atctgggtgc
     1021 cacagtggat ctgtatattc ttaatcatct aatgaaccca cccaatggaa
aacgctgtga
     1081 atttgtcatg gaagtcacaa ataaaacacg tgcagatgta aagggcggga
cactcactca
     1141 atatgaaggc aaactgagac tggtggaaat tgctcaagtg ccaaaagcac
atgtagacga
     1201 gttcaagtct gtatcaaagt tcaaaatatt taatacaaac aacctatgga
tttctcttgc
     1261 agcagttaaa agactgcagg agcaaaatgc cattgacatg gaaatcattg
tgaatgcaaa
     1321 gactttggat ggaggcctga atgtcattca attagaaact gcagtagggg
ctgccatcaa
     1381 aagttttgag aattctctag gtattaatgt gccaaggagc cgttttctgc
ctgtcaaaac
     1441 cacatcagat ctcttgctgg tgatgtcaaa cctctatagt cttaatgcag
gatctctgac
     1501 aatgagtgaa aagcgggaat ttcctacagt gcccttggtt aaattaggca
gttcttttac
     1561 gaaggttcaa gattatctaa gaagatttga aagtatacca gatatgcttg
aattggatca
     1621 cctcacagtt tcaggagatg tgacatttgg aaaaaatgtt tcattaaagg
gaacggttat
     1681 catcattgca aatcatggtg acagaattga tatcccacct ggagcagtat
tagagaacaa
     1741 gattgtgtct ggaaaccttc gcatcttgga ccactgaaat gaaaaatact
gtggacactt
```

```
     1801 aaataatggg ctagtttctt acaatgaaat gttctctagg attctaaaat
aggcaggtac
     1861 tttactatgt tactgtaccc tgcagtgttg atttttaaaa tagagttttc
tgcagtatgc
     1921 ttttagtcta agaaaagcac agatggagca atactttcct tctttgaaga
gaatcccaaa
     1981 agttagttca tcttaaagtg caatattgtt taatcttaaa actgggcaac
tttggaagaa
     2041 cttttaacag aagcctcaat gatgatcact ttgaattgct tgtgatttca
aaaataaagc
     2101 agtgaagcaa taaaaaaaaa aaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001001521.1

LOCUS NP_001001521

ACCESSION NP_001001521

```
      1 msqdgasqfq evirqelels vkkelekilt tasshefeht kkdldgfrkl
fhrflqekgp
       61 svdwgkiqrp pedsiqpyek ikarglpdni ssvlnklvvv klngglgtsm
gckgpkslig
      121 vrnentfldl tvqqiehlnk tyntdvplvl mnsfntdedt kkilqkynhc
rvkiytfnqs
      181 ryprinkesl lpvakdvsys genteawypp ghgdiyasfy nsglldtfig
egkeyifvsn
      241 idnlgatvdl yilnhlmnpp ngkrcefvme vtnktradvk ggtltqyegk
lrlveiaqvp
      301 kahvdefksv skfkifntnn lwislaavkr lqeqnaidme iivnaktldg
glnviqleta
      361 vgaaiksfen slginvprsr flpvkttsdl llvmsnlysl nagsltmsek
refptvplvk
      421 lgssftkvqd ylrrfesipd mleldhltvs gdvtfgknvs lkgtviiian
hgdridippg
      481 avlenkivsg nlrildh
            //
```

UGDH

Official Symbol: **UGDH**

Official Name: UDP-glucose 6-dehydrogenase

Gene ID: 7358

Organism: Homo sapiens

Other Aliases: GDH, UDP-GlcDH, UDPGDH, UGD

Other Designations: UDP-Glc dehydrogenase; UDP-glucose dehydrogenase; uridine diphospho-glucose dehydrogenase

Nucleotide sequence:

NCBI Reference Sequence: NM_001184700.1

LOCUS: NM_001184700

ACCESSION : NM_001184700

```
      1 gtgaaggaaa tagggacctg gccctgggcc ttgtgtagcg ggaggggggag
ctaggaagca
       61 gctgagggca gaatccagga gggcctggct gcgggggaat gaagcctccg
ccttcgcagg
      121 caaaagcctt taaatacggg ctcaggcccg ggactcagag tgtaacgcgt
ggcagcctga
      181 gggaggggcg tgcgccgaga gggagctcag atcgagcggg gcgcgggtgg
agaagctgcg
      241 gcggcgcggc ccgtaggaag gtgctgtccg aacgatcggg ataggagcgg
tccctgcgct
      301 tgctgctggg aagtggtaca atcatgtttg aaattaagaa gatctgttgc
atcggtgcag
      361 gctatgttgg aggacccaca tgtagtgtca ttgctcatat gtgtcctgaa
atcagggtaa
      421 cggttgttga tgtcaatgaa tcaagaatca atgcgtggaa ttctcctaca
cttcctattt
      481 atgagccagg actaaaagaa gtggtagaat cctgtcgagg aaaaaatctt
tttttttcta
      541 ccaatattga tgatgccatc aaagaagctg atcttgtatt tatttctgtg
ctgtccaacc
      601 ctgagtttct ggcagaggga acagccatca aggacctaaa gaacccagac
agagtactga
      661 ttggagggga tgaaactcca gagggccaga gagctgtgca ggccctgtgt
gctgtatatg
      721 agcactgggt tcccagagaa aagatcctca ccactaatac ttggtcttca
gagctttcca
      781 aactggcagc aaatgctttt cttgcccaga gaataagcag cattaactcc
ataagtgctc
      841 tgtgtgaagc aacaggagct gatgtagaag aggtagcaac agcgattgga
atggaccaga
      901 gaattggaaa caagtttcta aaagccagtg ttgggtttgg tgggagctgt
ttccaaaagg
      961 atgttctgaa tttggtttat ctctgtgagg ctctgaattt gccagaagta
gctcgttatt
     1021 ggcagcaggt catagacatg aatgactacc agaggaggag gtttgcttcc
cggatcatag
     1081 atagtctgtt taatacagta actgataaga agatagctat tttgggattt
gcattcaaaa
     1141 aggacactgg tgatacaaga gaatcttcta gtatatatat tagcaaatat
ttgatggatg
     1201 aaggtgcaca tctacatata tatgatccaa aagtacctag ggaacaaata
gttgtggatc
     1261 tttctcatcc aggtgtttca gaggatgacc aagtgtcccg gctcgtgacc
atttccaagg
     1321 atccatatga agcatgtgat ggtgcccatg ctgttgttat ttgcactgag
tgggacatgt
```

```
     1381 ttaaggaatt ggattatgaa cgcattcata aaaaaatgct aaagccagcc
tttatcttcg
     1441 atggacggcg tgtcctggat gggctccaca atgaactaca aaccattggc
ttccagattg
     1501 aaacaattgg caaaaaggtg tcttcaaaga gaattccata tgctccttct
ggtgaaattc
     1561 cgaagtttag tcttcaagat ccacctaaca agaaacctaa agtgtagaga
ttgccatttt
     1621 tatttgtgat ttttttttt ttttttggt acttcaggat agcaaatatc
tatctgctat
     1681 taaatggtaa atgaaccaag tgtttttttt tgtttttttt ttgagacaga
gtctcactgt
     1741 tgcccaggct ggagtgcagt ggtgcaatct cggctcactg caagctctgc
ttcccaggtt
     1801 cacgccattc tcctggctca gcctcccaag tagctgggac tacaggcacc
cgccacagtg
     1861 cctggctaat tttttgtatt tttagtagag acagggtttc accatgtgag
ccaggatggt
     1921 ctcaatctcc tgaccttgtg aaccacccgt ctcggcctcc caaagtgctg
ggattacagg
     1981 tgtgagccac cacgcctggc ccatgaacca agtgttttta aggaaacaaa
actatttttt
     2041 taatcatcag atttatacta gctatatgga tattagcata tctggtaatt
atgaatctag
     2101 aatttttta catattttta taatactgtt agctcagtta ttggatgagt
gaaagataat
     2161 catgttggtt ttaatagtgt caatttttgt aaaataaaaa ttaaacttca
aactctttac
     2221 tttataaatt gtccataggc cacactttaa tatcacatta taaagggaag
gacagtcttc
     2281 attcctcctg gttattggtt tgtttgtcat taaagatata ttttgaatcc
atgaaattgc
     2341 tatgctaaac agcctttaca tgtatggtct ggttaaagtt cctttgttcc
ttttgtttta
     2401 ataaaatgtg tcactgattt tttagctcaa aatcatcact gttaatttcc
agtcacccca
     2461 aatatggtta aaagattttt tttttaatca tgaagagaaa attagtagca
tttctttctc
     2521 tccccattat ttattggttt tcctcactaa tctttttttt tttagtccaa
aagccaaaaa
     2581 tatttatctt ggttttacat tttaatttcc attcttaatt gtaatttttt
tctttaaata
     2641 aggaaaccaa tataatctca tgtataaaaa cttaaatatt ttacaagtta
catatagcat
     2701 cattctaaaa taagaatttt ttttgttttc tgtctgcttt tttcttatgt
ctcttgttga
     2761 gttttatatt ttcagtggtt atttttgctt gtgttagatc attattaaaa
tatatccaat
     2821 gtccctttga tacttgtgct ctgctgagaa tgtacagttt gcattaaaca
tcccaggtct
     2881 catccttcag gaattttgca gttcaatgag aagagggaga caaatataaa
gatgaggaca
     2941 gaagcatctc tacagatgaa aattacataa ataaaacatt ctccatcaac
aactaaaaaa
     3001 aaaaaaaaaa aaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001171629.1

LOCUS: NP_001171629

ACCESSION: NP_001171629

```
      1 mfeikkicci gagyvggptc sviahmcpei rvtvvdvnes rinawnsptl
piyepglkev
       61 vescrgknlf fstniddaik eadlvfisvl snpeflaegt aikdlknpdr
vliggdetpe
      121 gqravqalca vyehwvprek ilttntwsse lsklaanafl aqrissinsi
salceatgad
      181 veevataigm dqrignkflk asvgfggscf qkdvlnlvyl cealnlpeva
rywqqvidmn
      241 dyqrrrfasr iidslfntvt dkkiailgfa fkkdtgdtre sssiyiskyl
mdegahlhiy
      301 dpkvpreqiv vdlshpgvse ddqvsrlvti skdpyeacdg ahavvictew
dmfkeldyer
      361 ihkkmlkpaf ifdgrrvldg lhnelqtigf qietigkkvs skripyapsg
eipkfslqdp
      421 pnkkpkv
      //
```

Gene

Official Symbol: **PLIN3** (also known as **M6PRBP1**)

Official Name: perilipin 3

Gene ID: 10226

Organism: Homo sapiens

Other Aliases: **M6PRBP1**, PP17, TIP47

Other Designations: 47 kDa MPR-binding protein; cargo selection protein TIP47; mannose-6-phosphate receptor-binding protein 1; perilipin-3; placental protein 17; tail-interacting protein, 47 kD

Nucleotide sequence:

NCBI Reference Sequence: NM_001164189.1

LOCUS: NM_001164189

ACCESSION : NM_001164189

```
      1 tggcgcgggc aatccctcaa cctgattggt cccctcgccc gtcactccag
tgcgccccca
       61 acctaccacg cagtaaaagc caccccgcc tcggcccgga cggtttccaa
gctggttttg
      121 aagtcgcggc agctgttcct gggacgtccg gttgaccgcg cgtctgctgc
agagaccatg
```

181 tctgccgacg gggcagaggc tgatggcagc acccaggtga cagtggaaga
accggtacag
241 cagcccagtg tggtggaccg tgtggccagc atgcctctga tcagctccac
ctgcgacatg
301 gtgtccgcag cctatgcctc caccaaggag agctacccgc acatcaagac
tgtctgcgac
361 gcagcagaga agggagtgag gaccctcacg gcggctgctg tcagcggggc
tcagccgatc
421 ctctccaagc tggagcccca gattgcatca gccagcgaat acgcccacag
ggggctggac
481 aagttggagg agaacctccc catcctgcag cagcccacgg agaaggtcct
ggcggacacc
541 aaggagcttg tgtcgtctaa ggtgtcgggg gcccaagaga tggtgtctag
cgccaaggac
601 acggtggcca cccaattgtc ggaggcggtg gacgcgaccc gcggtgctgt
gcagagcggc
661 gtggacaaga caaagtccgt agtgaccggc ggcgtccaat cggtcatggg
ctcccgcttg
721 ggccagatgg tgttgagtgg ggtcgacacg gtgctgggga agtcggagga
gtgggcggac
781 aaccacctgc cccttacgga tgccgaactg gcccgcatcg ccacatccct
ggatggcttt
841 gacgtcgcgt ccgtgcagca gcagcggcag gaacagagct acttcgtacg
tctgggctcc
901 ctgtcggaga ggctgcggca gcacgcctat gagcactcgc tgggcaagct
tcgagccacc
961 aagcagaggg cacaggaggc tctgctgcag ctgtcgcagg tcctaagcct
gatggaaact
1021 gtcaagcaag gcgttgatca gaagctggtg gaaggccagg agaagctgca
ccagatgtgg
1081 ctcagctgga accagaagca gctccagggc cccgagaagg agccgcccaa
gccagaggtc
1141 gagtcccggg cgctcaccat gttccgggac attgcccagc aactgcaggc
cacctgtacc
1201 tccctggggt ccagcattca gggcctcccc accaatgtga aggaccaggt
gcagcaggcc
1261 cgccgccagg tggaggacct ccaggccacg ttttccagca tccactcctt
ccaggacctg
1321 tccagcagca ttctggccca gagccgtgag cgtgtcgcca gcgcccgcga
ggccctggac
1381 cacatggtgg aatatgtggc ccagaacaca cctgtcacgt ggctcgtggg
acccttgcc
1441 cctggaatca ctgagaaagc cccggaggag aagaagtagg gggagaggag
aggactcagc
1501 gggccccgtc tctataatgc agctgtgctc tggagtcctc aacccggggc
tcatttcaaa
1561 cttatttct agccactcct cccagctctt ctgtgctgtc cacttgggaa
gctaaggctc
1621 tcaaaacggg catcacccag ttgacccatc tctcagcctc tctgagcttg
gaagaagcct
1681 gttctgagcc tcaccctatc agtcagtaga gagagatgtc cagaaaaaat
atctttcagg
1741 aaagttctcc cctgcagaat ttttttcct tgttaaatat caggaatata
ggccgggtgc
1801 ggtggctcac acctgtaatc ccagcacttt gggaggctga ggcgggcgga
acacctgagg
1861 tcaggtgttc gagaccagcc aggccaacat ggtgaaaccc cgtctctact
aaaaatacaa
1921 aaaaaaatga gccgggcatg gtagcaggtg tctgttatcc cagttaggag
gctgaggcaa

```
     1981 gagaatctct tgaacctgag aggcggaggt tgcagtgagc caagatcgcg
ccattgcact
     2041 ccagcctggg ggacaagagt gagacttagt ctcaaaaaaa aaaaaaaaga
aaaaaaaatc
     2101 agggatatag ttcatatccc acttctttgt ttacaccgat gtccctgaat
atcagcctgt
     2161 agctaatgga cttgggattt ctggtctaag tgggcctcct ggggatgggg
tggtacactg
     2221 agcttctgag cctcattgta gagtagaaag gtactggggc ctgtgtggta
agccttgttg
     2281 aaatgctctg gtattcagta ttgccttaat aaacttcacc cacaactgca
tacaggcaaa
     2341 aa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001157661.1

LOCUS: NP_001157661

ACCESSION: NP_001157661

```
      1 msadgaeadg stqvtveepv qqpsvvdrva smplisstcd mvsaayastk
esyphiktvc
       61 daaekgvrtl taaavsgaqp ilsklepqia saseyahrgl dkleenlpil
qqptekvlad
      121 tkelvsskvs gaqemvssak dtvatqlsea vdatrgavqs gvdktksvvt
ggvqsvmgsr
      181 lgqmvlsgvd tvlgkseewa dnhlpltdae lariatsldg fdvasvqqqr
qeqsyfvrlg
      241 slserlrqha yehslgklra tkqraqeall qlsqvlslme tvkqgvdqkl
vegqeklhqm
      301 wlswnqkqlq gpekeppkpe vesraltmfr diaqqlqatc tslgssiqgl
ptnvkdqvqq
      361 arrqvedlqa tfssihsfqd lsssilaqsr ervasareal dhmveyvaqn
tpvtwlvgpf
      421 apgitekape ekk
//
```

C14orf166

Official Symbol: **C14orf166**

Official Name: chromosome 14 open reading frame 166

Gene ID: 51637

Organism: Homo sapiens

Other Aliases: CGI-99, CGI99, CLE, CLE7, LCRP369, RLLM1

Other Designations: CLE7 homolog; RLL motif containing 1; UPF0568 protein C14orf166

Nucleotide sequence:

NCBI Reference Sequence: NM_016039.2

LOCUS: NM_016039

ACCESSION : NM_016039

```
      1 cgccgtcatt tcggagcgac tcagcgcctg cccgccctct cgccgcgtcg
ccggtgcctg
       61 cgcctcccgc tccacctcgc ttcttctctc ccggccgagg cccgggggac
cagagcgaga
      121 agcggggacc atgttccgac gcaagttgac ggctctcgac taccacaacc
ccgccggctt
      181 caactgcaaa gatgaaacag aatttagaaa cttcatcgtt tggcttgaag
accagaaaat
      241 caggcactac aagattgaag acagagggaa tttaagaaac atccacagca
gcgactggcc
      301 caagttcttt gaaaagtatc tcagagatgt taactgtcct ttcaagattc
aagatcgaca
      361 agaagctatt gactggcttc ttggtttagc tgttagactt gaatatggag
ataatgctga
      421 aaaatacaag gatttagtac ctgataattc aaaaactgct gacaatgcaa
ctaaaaatgc
      481 agaaccattg atcaatttgg atgtaaataa tcctgatttt aaggctggtg
tgatggcttt
      541 ggctaacctg cttcagattc agcgtcatga tgattacctg gtaatgctta
aggcaattcg
      601 gattttggtt caggagcgcc tgacacagga tgcagttgct aaggcaaatc
aaacaaaaga
      661 gggcttacct gttgctttag acaaacatat tcttggtttt gacacaggag
atgcagttct
      721 taatgaagct gctcaaattc tgcgattgct gcacatagag gagctcagag
agctacagac
      781 aaaaatcaac gaagccatag tagctgttca ggcaattatt gctgatccaa
agacagacca
      841 cagactggga aaagttggaa gatgaacact tgaggacttc agcttctcac
ctacttagta
      901 cagttgggaa ccatacactt ctggcatgtt tggaaatcaa aatgtcacat
tctcgggggа
      961 ggaagcccag aaaattgggt atgttctaga gatttaccac cattgcttat
tgctttttc
     1021 tttaataaag tttaggaaag tagaattttt aaaaaaaaaa aaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_057123.1

LOCUS: NP_057123

ACCESSION: NP_057123

```
      1 mfrrkltald yhnpagfnck detefrnfiv wledqkirhy kiedrgnlrn
ihssdwpkff
       61 ekylrdvncp fkiqdrqeai dwllglavrl eygdnaekyk dlvpdnskta
dnatknaepl
```

```
      121 inldvnnpdf kagvmalanl lqiqrhddyl vmlkairilv qerltqdava
kanqtkeglp
      181 valdkhilgf dtgdavlnea aqilrllhie elrelqtkin eaivavqaii
adpktdhrlg
      241 kvgr
//
```

SNRNP70

Official Symbol: SNRNP70

Official Name: small nuclear ribonucleoprotein 70kDa (U1)

Gene ID: 6625

Organism: Homo sapiens

Other Aliases: RNPU1Z, RPU1, SNRP70, Snp1, U1-70K, U170K, U1AP, U1RNP

Other Designations: U1 small nuclear ribonucleoprotein 70 kDa; U1 snRNP 70 kDa

Nucleotide sequence:

NCBI Reference Sequence: NM_003089.4

LOCUS: NM_003089

ACCESSION : NM_003089

```
      1 gcggttcggc gcggaaagcg ggaggtggag gggcggcttg gggcaagcgc
gcgcgcgcag
       61 tgcagaagcc agccccccgc ggctgaggta ctcaaggtgc ccaaaggcgg
ggtagtgacc
      121 tcgcgcgtgc gctgtgcccg cggcagcgcc gggtcctagt gtgtgggttg
ttgttggcac
      181 cgcacggcgc gtgcgcagtg aggacggcgg agggatttgc ggccgggacc
caccccctgc
      241 tccagtcgct atcggaggcc gcgcgggtgg ctgagcagcg gcctggtgcg
ctcgcttagc
      301 gggcgacgga atcagacgga cgtggacgcc cccggagtgg aagccgaagc
aggagttgtt
      361 gttgctgagg ggctgccgca gccgccgcga gcctccggac agacgccaga
gcgaggaggg
      421 cgctacgcga cttggcaaga tgacccagtt cctgccgccc aaccttctgg
ccctctttgc
      481 cccccgtgac cctattccat acctgccacc cctggagaaa ctgccacatg
aaaaacacca
      541 caatcaacct tattgtggca ttgcgccgta cattcgagag tttgaggacc
ctcgagatgc
      601 ccctcctcca actcgtgctg aaacccgaga ggagcgcatg gagaggaaaa
gacgggaaaa
      661 gattgagcgg cgacagcaag aagtggagac agagcttaaa atgtgggacc
ctcacaatga
```

```
      721 tcccaatgct caggggatg ccttcaagac tctcttcgtg gcgagagtga
attatgacac
      781 aacagaatcc aagctccgga gagagtttga ggtgtacgga cctatcaaaa
gaatacacat
      841 ggtctacagt aagcggtcag gaaagccccg tggctatgcc ttcatcgagt
acgaacacga
      901 gcgagacatg cactccgctt acaaacacgc agatggcaag aagattgatg
gcaggagggt
      961 ccttgtggac gtggagaggg gccgaaccgt gaagggctgg aggccccggc
ggctaggagg
     1021 aggcctcggt ggtaccagaa gaggaggggc tgatgtgaac atccggcatt
caggccgcga
     1081 tgacacctcc cgctacgatg agaggcccgg cccctccccg cttccgcaca
gggaccggga
     1141 ccgggaccgt gagcgggagc gcagagagcg gagccgggag cgagacaagg
agcgagaacg
     1201 gcgacgctcc cgctcccggg accggcggag gcgctcacgg agtcgcgaca
aggaggagcg
     1261 gaggcgctcc agggagcgga gcaaggacaa ggaccgggac cggaagcggc
gaagcagccg
     1321 gagtcgggag cgggcccggc gggagcggga gcgcaaggag gagctgcgtg
gcggcggtgg
     1381 cgacatggcg gagccctccg aggcgggtga cgcgcccccc gatgatgggc
ctccagggga
     1441 gctcgggcct gacggccctg acggtccaga ggaaaagggc cgggatcgtg
accgggagcg
     1501 acggcggagc caccggagcg agcgcgagcg gcgccgggac cgggatcgtg
accgtgaccg
     1561 tgaccgcgag cacaaacggg gggagcgggg cagtgagcgg ggcagggatg
aggcccgagg
     1621 tgggggcggt ggccaggaca acgggctgga gggtctgggc aacgacagcc
gagacatgta
     1681 catggagtct gagggcggcg acggctacct ggctccggag aatgggtatt
tgatggaggc
     1741 tgcgccggag tgaagaggtc gtcctctcca tctgctgtgt ttggacgcgt
tcctgcccag
     1801 ccccttgctg tcatcccctc ccccaacctt ggccacttga gtttgtcctc
caagggtagg
     1861 tgtctcattt gttctggccc cttggattta aaaataaaat taatttcctg
ttgatagtgg
     1921 gcaaaaaaaa aaaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_003080.2

LOCUS: NP_003080

ACCESSION: NP_003080

```
      1 mtqflppnll alfaprdpip ylppleklph ekhhnqpycg iapyirefed
prdappptra
       61 etreermerk rrekierrqq evetelkmwd phndpnaqgd afktlfvarv
nydttesklr
      121 refevygpik rihmvyskrs gkprgyafie yeherdmhsa ykhadgkkid
grrvlvdver
```

```
      181 grtvkgwrpr rlggglggtr rggadvnirh sgrddtsryd erpgpsplph
rdrdrdrere
      241 rrersrerdk ererrrsrsr drrrrsrsrd keerrrsrer skdkdrdrkr
rssrsrerar
      301 rererkeelr ggggdmaeps eagdappddg ppgelgpdgp dgpeekgrdr
drerrrshrs
      361 ererrrdrdr drdrdrehkr gergsergrd eargggggqd ngleglgnds
rdmymesegg
      421 dgylapengy lmeaape
            //
```

CNN2

Official Symbol: **CNN2**

Official Name: calponin 2

Gene ID: 1265

Organism: Homo sapiens

Other Aliases: none

Other Designations: calponin H2, smooth muscle; calponin-2; neutral calponin

Nucleotide sequence:

NCBI Reference Sequence: NM_004368.2

LOCUS: NM_004368

ACCESSION : NM_004368

```
      1 gaaagagtga gagccgccca cgagctctga gcagagagcc cgcaggagtg
ccacgtcccg
       61 gcggcctcgg cccctccctg cctcagtttc ccgtggcatc aaaggggcg
aggggccct
      121 ccaggcctct ggtgacgggg gtgctgtgcc caggcggggg tccgggggcg
accgaggggg
      181 ctcaggaagt ccgcggccgc aggaattcgg cgcctccagg ccttataagg
acatttgcgc
      241 tccgggccaa tcagcggcgg gggcgtggcg cgcggagccc ggcgcgtccc
aaccccgcgc
      301 cagcccggcg gtcccgtccc gtcccgtcct gtgcggcccc gtcccgccgc
ccgcccgcca
      361 gccatgagct ccacgcagtt caacaagggc ccctcgtacg ggctgtcggc
cgaggtcaag
      421 aaccggctcc tgtccaaata tgacccccag aaggaggcag agctccgcac
ctggatcgag
      481 ggactcaccg gcctctccat cggccccgac ttccagaagg gcctgaagga
tggaactatc
```

541 ttatgcacac tcatgaacaa gctacagccg ggctccgtcc ccaagatcaa
ccgctccatg
601 cagaactggc accagctaga aaacctgtcc aacttcatca aggccatggt
cagctacggc
661 atgaaccctg tggacctgtt cgaggccaac gacctgtttg agagtgggaa
catgacgcag
721 gtgcaggtgt ctcttctcgc cctggcgggg aaggccaaga ctaaggggct
gcagagcggg
781 gtggacattg gcgtcaagta ctcggagaag caggagcgga atttcgacga
tgccaccatg
841 aaggctggcc agtgcgtcat cgggctgcag atgggcacca acaaatgcgc
cagccagtcg
901 ggcatgactg cctacggcac gagaaggcat ctctatgacc ccaagaacca
tatcctgccc
961 cccatggacc actcgaccat cagcctccag atgggcacga acaagtgtgc
cagccaggtg
1021 ggcatgacgg ctcccgggac ccggcggcac atctatgata ccaagctggg
aaccgacaag
1081 tgtgacaact cctccatgtc cctgcagatg ggctacacgc agggcgccaa
ccagagcggc
1141 caggtcttcg gcctgggccg gcagatatat gaccccaagt actgcccgca
aggcacagtg
1201 gccgatgggg ctccctcggg caccggcgac tgcccggacc cgggggaggt
ccctgaatat
1261 ccccttact accaggagga ggccggctac tgaggctccc agcacgctct
ctccccacat
1321 cgtctgccca tctgggtttt tgggtttttc tgtgttttca tctttttttt
ttttttctta
1381 acccgttcag tgctgccagt caaccaaggg tctgtgagtg tcagcgtggg
atcaggcagc
1441 agagcttttt tccccttgc cttgatcctt cgcaaggctg agccactggg
ctgtgggga
1501 aggggtcaag gccatatccc aatacgtgta gggcgagggt ccctgctggc
acattcaggc
1561 tgtgctggga agaagagacc tgggcttgga aggaaccggt ccccgacggt
ttctgcttgc
1621 ctcgcctctt ccccctttg tcagctgagc agtttgtggt ttctatgccc
gcaagtttca
1681 ggaagtattc acaaaagaaa aatacatttt ttcccccagg ggtggggcaa
ggacagtgga
1741 gagagtgcta ggaaatgagt cccctgggaa aggggaccgg gccgtgatgt
taaatatctc
1801 cggctcccaa gtgactggat ttgcctagga ccttcagatc aacagacttc
agaccctcag
1861 acctgccccg gggccaggtg gagaaagtga gggccgtaca aggaagtgaa
attctgagtt
1921 gttggggcta agcctgaccc cctctccatg ctccccgccc caactcactc
tggcctcagt
1981 agatttttt ttcagttgtg gttgttgccc aggctggagt gcagtggcgc
catcttggct
2041 cactgcacct ccaccttccg ggctcaagcg attctccagc ctcagcctcc
tgagtagcta
2101 ggactgcagg tgctccacca cgcccggcta atttttgtat ttttagtaga
gatggggttt
2161 ccccatgttg gccaggctgg tctcgaactc ctggcctcag gtgtgatccg
cccgcctccg
2221 cctccccaag cgctgagatt acaggtgtga gccaccgtgc ccaggccctc
agtaggtttt
2281 aaggagtccc cagccctcct cccttctggg cccgaccagc ttatactgct
ccatcttccc

```
     2341 cggccacatg ccccgccaag tactgcacag ggacccccca cccaggggcc
ctgctccgtg
     2401 agataatgtg aaatacgact gtggaccaaa cgcaataaaa cctctgtttg
tacgaagaaa
     2461 aaaaaaaaaa aaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_004359.1

LOCUS: NP_004359

ACCESSION: NP_004359

```
      1 msstqfnkgp syglsaevkn rllskydpqk eaelrtwieg ltglsigpdf
qkglkdgtil
       61 ctlmnklqpg svpkinrsmq nwhqlenlsn fikamvsygm npvdlfeand
lfesgnmtqv
      121 qvsllalagk aktkglqsgv digvkysekq ernfddatmk agqcviglqm
gtnkcasqsg
      181 mtaygtrrhl ydpknhilpp mdhstislqm gtnkcasqvg mtapgtrrhi
ydtklgtdkc
      241 dnssmslqmg ytqganqsgq vfglgrqiyd pkycpqgtva dgapsgtgdc
pdpgevpeyp
      301 pyyqeeagy
//
```

PEBP1

Official Symbol: **PEBP1**

Official Name: phosphatidylethanolamine binding protein 1

Gene ID: 5037

Organism: Homo sapiens

Other Aliases: HCNP, HCNPpp, PBP, PEBP, PEBP-1, RKIP

Other Designations: Raf kinase inhibitory protein; hippocampal cholinergic neurostimulating peptide; neuropolypeptide h3; phosphatidylethanolamine-binding protein 1; prostatic binding protein; prostatic-binding protein; raf kinase inhibitor protein

Nucleotide sequence:

NCBI Reference Sequence: NM_002567.2

LOCUS: NM_002567

ACCESSION : NM_002567 XR_109136 XR_109137 XR_111344 XR_114620

```
      1 tgggcggcgg ctgaggcgcg tgctctcgcg tggtcgctgg gtctgcgtct
tcccgagcca
       61 gtgtgctgag ctctccgcgt cgcctctgtc gcccgcgcct ggcctaccgc
ggcactcccg
      121 gctgcacgct ctgcttggcc tcgccatgcc ggtggacctc agcaagtggt
ccgggccctt
      181 gagcctgcaa gaagtggacg agcagccgca gcacccgctg catgtcacct
acgccggggc
      241 ggcggtggac gagctgggca aagtgctgac gcccacccag gttaagaata
gacccaccag
      301 catttcgtgg gatggtcttg attcagggaa gctctacacc ttggtcctga
cagacccgga
      361 tgctcccagc aggaaggatc ccaaatacag agaatggcat catttcctgg
tggtcaacat
      421 gaagggcaat gacatcagca gtggcacagt cctctccgat tatgtgggct
cggggcctcc
      481 caagggcaca ggcctccacc gctatgtctg gctggtttac gagcaggaca
ggccgctaaa
      541 gtgtgacgag cccatcctca gcaaccgatc tggagaccac cgtggcaaat
tcaaggtggc
      601 gtccttccgt aaaaagtatg agctcagggc cccggtggct ggcacgtgtt
accaggccga
      661 gtgggatgac tatgtgccca aactgtacga gcagctgtct gggaagtagg
gggttagctt
      721 ggggacctga actgtcctgg aggccccaag ccatgttccc cagttcagtg
ttgcatgtat
      781 aatagatttc tcctcttcct gccccccttg gcatgggtga gacctgacca
gtcagatggt
      841 agttgagggt gacttttcct gctgcctggc ctttataatt ttactcactc
actctgattt
      901 atgttttgat caaatttgaa cttcattttg gggggtattt tggtactgtg
atggggtcat
      961 caaattatta atctgaaaat agcaacccag aatgtaaaaa agaaaaaact
ggggggaaaa
     1021 agaccaggtc tacagtgata gagcaaagca tcaaagaatc tttaagggag
gtttaaaaaa
     1081 aaaaaaaaaa aaaaagattg gttgcctctg cctttgtgat cctgagtcca
gaatggtaca
     1141 caatgtgatt ttatggtgat gtcactcacc tagacaacca gaggctggca
ttgaggctaa
     1201 cctccaacac agtgcatctc agatgcctca gtaggcatca gtatgtcact
ctggtccctt
     1261 taaagagcaa tcctggaaga agcaggaggg agggtggctt tgctgttgtt
gggacatggc
     1321 aatctagacc ggtagcagcg ctcgctgaca gcttgggagg aaacctgaga
tctgtgtttt
     1381 ttaaattgat cgttcttcat gggggtaaga aaagctggtc tggagttgct
gaatgttgca
     1441 ttaattgtgc tgtttgcttg tagttgaata aaaatagaaa cctgaatgaa
gaaaaaaaaa
     1501 aaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_002558.1

LOCUS: NP_002558

ACCESSION: NP_002558

```
      1 mpvdlskwsg plslqevdeq pqhplhvtya gaavdelgkv ltptqvknrp
tsiswdglds
       61 gklytlvltd pdapsrkdpk yrewhhflvv nmkgndissg tvlsdyvgsg
ppkgtglhry
      121 vwlvyeqdrp lkcdepilsn rsgdhrgkfk vasfrkkyel rapvagtcyq
aewddyvpkl
      181 yeqlsgk
//
```

ACLY

Official Symbol: **ACLY**

Official Name: ATP citrate lyase

Gene ID: 47

Organism: Homo sapiens

Other Aliases: ACL, ATPCL, CLATP

Other Designations: ATP citrate synthase; ATP-citrate (pro-S-)-lyase; ATP-citrate synthase; citrate cleavage enzyme

Nucleotide sequence:

NCBI Reference Sequence: NM_001096.2

LOCUS: NM_001096

ACCESSION : NM_001096

```
      1 agccgatggg ggcggggaaa agtccggctg ggccgggaca aaagccggat
cccgggaagc
       61 taccggctgc tggggtgctc cggattttgc ggggttcgtc gggcctgtgg
aagaagcgcc
      121 gcgcacggac ttcggcagag gtagagcagg tctctctgca gccatgtcgg
ccaaggcaat
      181 ttcagagcag acgggcaaag aactccttta caagttcatc tgtaccacct
cagccatcca
      241 gaatcggttc aagtatgctc gggtcactcc tgacacagac tgggcccgct
tgctgcagga
      301 ccacccctgg ctgctcagcc agaacttggt agtcaagcca gaccagctga
tcaaacgtcg
      361 tggaaaactt ggtctcgttg gggtcaacct cactctggat ggggtcaagt
cctggctgaa
      421 gccacggctg ggacaggaag ccacagttgg caaggccaca ggcttcctca
agaactttct
      481 gatcgagccc ttcgtccccc acagtcaggc tgaggagttc tatgtctgca
tctatgccac
```

541 ccgagaaggg gactacgtcc tgttccacca cgaggggggt gtggacgtgg
gtgatgtgga
601 cgccaaggcc cagaagctgc ttgttggcgt ggatgagaaa ctgaatcctg
aggacatcaa
661 aaaacacctg ttggtccacg cccctgaaga caagaaagaa attctggcca
gttttatctc
721 cggcctcttc aatttctacg aggacttgta cttcacctac ctcgagatca
atccccttgt
781 agtgaccaaa gatggagtct atgtccttga cttggcggcc aaggtggacg
ccactgccga
841 ctacatctgc aaagtgaagt ggggtgacat cgagttccct ccccccttcg
ggcgggaggc
901 atatccagag gaagcctaca ttgcagacct cgatgccaaa agtggggcaa
gcctgaagct
961 gaccttgctg aaccccaaag ggaggatctg gaccatggtg gccgggggtg
gcgcctctgt
1021 cgtgtacagc gataccatct gtgatctagg gggtgtcaac gagctggcaa
actatgggga
1081 gtactcaggc gcccccagcg agcagcagac ctatgactat gccaagacta
tcctctccct
1141 catgacccga gagaagcacc cagatggcaa gatcctcatc attggaggca
gcatcgcaaa
1201 cttcaccaac gtggctgcca cgttcaaggg catcgtgaga gcaattcgag
attaccaggg
1261 ccccctgaag gagcacgaag tcacaatctt tgtccgaaga ggtggcccca
actatcagga
1321 gggcttacgg gtgatgggag aagtcgggaa gaccactggg atccccatcc
atgtctttgg
1381 cacagagact cacatgacgg ccattgtggg catggccctg ggccaccggc
ccatccccaa
1441 ccagccaccc acagcggccc acactgcaaa cttcctcctc aacgccagcg
ggagcacatc
1501 gacgccagcc cccagcagga cagcatcttt ttctgagtcc agggccgatg
aggtggcgcc
1561 tgcaaagaag gccaagcctg ccatgccaca agattcagtc ccaagtccaa
gatccctgca
1621 aggaaagagc accaccctct tcagccgcca caccaaggcc attgtgtggg
gcatgcagac
1681 ccgggccgtg caaggcatgc tggactttga ctatgtctgc tcccgagacg
agccctcagt
1741 ggctgccatg gtctaccctt tcactgggga ccacaagcag aagttttact
gggggcacaa
1801 agagatcctg atccctgtct tcaagaacat ggctgatgcc atgaggaagc
atccggaggt
1861 agatgtgctc atcaactttg cctctctccg ctctgcctat gacagcacca
tggagaccat
1921 gaactatgcc cagatccgga ccatcgccat catagctgaa ggcatccctg
aggccctcac
1981 gagaaagctg atcaagaagg cggaccagaa gggagtgacc atcatcggac
ctgccactgt
2041 tggaggcatc aagcctgggt gctttaagat tggcaacaca ggtgggatgc
tggacaacat
2101 cctggcctcc aaactgtacc gcccaggcag cgtggcctat gtctcacgtt
ccggaggcat
2161 gtccaacgag ctcaacaata tcatctctcg gaccacggat ggcgtctatg
agggcgtggc
2221 cattggtggg gacaggtacc cgggctccac attcatggat catgtgttac
gctatcagga
2281 cactccagga gtcaaaatga ttgtggttct tggagagatt gggggcactg
aggaatataa

2341 gatttgccgg ggcatcaagg agggccgcct cactaagccc atcgtctgct
ggtgcatcgg
2401 gacgtgtgcc accatgttct cctctgaggt ccagtttggc catgctggag
cttgtgccaa
2461 ccaggcttct gaaactgcag tagccaagaa ccaggctttg aaggaagcag
gagtgtttgt
2521 gccccggagc tttgatgagc ttggagagat catccagtct gtatacgaag
atctcgtggc
2581 caatggagtc attgtacctg cccaggaggt gccgccccca accgtgccca
tggactactc
2641 ctgggccagg gagcttggtt tgatccgcaa acctgcctcg ttcatgacca
gcatctgcga
2701 tgagcgagga caggagctca tctacgcggg catgcccatc actgaggtct
tcaaggaaga
2761 gatgggcatt ggcggggtcc tcggcctcct ctggttccag aaaaggttgc
ctaagtactc
2821 ttgccagttc attgagatgt gtctgatggt gacagctgat cacgggccag
ccgtctctgg
2881 agcccacaac accatcattt gtgcgcgagc tgggaaagac ctggtctcca
gcctcacctc
2941 ggggctgctc accatcgggg atcggtttgg gggtgccttg gatgcagcag
ccaagatgtt
3001 cagtaaagcc tttgacagtg gcattatccc catggagttt gtgaacaaga
tgaagaagga
3061 agggaagctg atcatgggca ttggtcaccg agtgaagtcg ataaacaacc
cagacatgcg
3121 agtgcagatc ctcaaagatt acgtcaggca gcacttccct gccactcctc
tgctcgatta
3181 tgcactggaa gtagagaaga ttaccacctc gaagaagcca aatcttatcc
tgaatgtaga
3241 tggtctcatc ggagtcgcat ttgtagacat gcttagaaac tgtgggtcct
ttactcggga
3301 ggaagctgat gaatatattg acattggagc cctcaatggc atctttgtgc
tgggaaggag
3361 tatggggttc attggacact atcttgatca gaagaggctg aagcaggggc
tgtatcgtca
3421 tccgtgggat gatatttcat atgttcttcc ggaacacatg agcatgtaac
agagccagga
3481 accctactgc agtaaactga agacaagatc tcttccccca agaaaaagtg
tacagacagc
3541 tggcagtgga gcctgcttta tttagcaggg gcctggaatg taaacagcca
ctggggtaca
3601 ggcaccgaag accaacatcc acaggctaac acccccttcag tccacacaaa
gaagcttcat
3661 attttttta taagcataga aataaaaacc aagccaatat ttgtgacttt
gctctgctac
3721 ctgctgtatt tattatatgg aagcatctaa gtactgtcag gatggggtct
tcctcattgt
3781 agggcgttag gatgttgctt tctttttcca ttagttaaac attttttttct
cctttggagg
3841 aagggaatga aacatttatg gcctcaagat actatacatt taaagcaccc
caatgtctct
3901 ctttttttttt ttttacttcc ctttcttctt ccttatataa catgaagaac
attgtattaa
3961 tctgattttt aaagatcttt ttgtatgtta cgtgttaagg gcttgtttgg
tatcccactg
4021 aaatgttctg tgttgcagac cagagtctgt ttatgtcagg gggatggggc
cattgcatcc
4081 ttagccattg tcacaaaata tgtggagtag taacttaata tgtaaagttg
taacatacat

```
     4141 acatttaaaa tggaaatgca gaaagctgtg aaatgtcttg tgtcttatgt
tctctgtatt
     4201 tatgcagctg atttgtctgt ctgtaactga agtgtgggtc caaggactcc
taactacttt
     4261 gcatctgtaa tccacaaaga ttctgggcag ctgccacctc agtctcttct
ctgtattatc
     4321 atagtctggt ttaaataaac tatatagtaa caaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
     4381 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
     4441 aaaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001087.2

LOCUS: NP_001087

ACCESSION: NP_001087

```
      1 msakaiseqt gkellykfic ttsaiqnrfk yarvtpdtdw arllqdhpwl
lsqnlvvkpd
       61 qlikrrgklg lvgvnltldg vkswlkprlg qeatvgkatg flknfliepf
vphsqaeefy
      121 vciyatregd yvlfhheggv dvgdvdakaq kllvgvdekl npedikkhll
vhapedkkei
      181 lasfisglfn fyedlyftyl einplvvtkd gvyvldlaak vdatadyick
vkwgdiefpp
      241 pfgreaypee ayiadldaks gaslkltlln pkgriwtmva gggasvvysd
ticdlggvne
      301 lanygeysga pseqqtydya ktilslmtre khpdgkilii ggsianftnv
aatfkgivra
      361 irdyqgplke hevtifvrrg gpnyqeglrv mgevgkttgi pihvfgteth
mtaivgmalg
      421 hrpipnqppt aahtanflln asgststpap srtasfsesr adevapakka
kpampqdsvp
      481 sprslqgkst tlfsrhtkai vwgmqtravq gmldfdyvcs rdepsvaamv
ypftgdhkqk
      541 fywghkeili pvfknmadam rkhpevdvli nfaslrsayd stmetmnyaq
irtiaiiaeg
      601 ipealtrkli kkadqkgvti igpatvggik pgcfkigntg gmldnilask
lyrpgsvayv
      661 srsggmsnel nniisrttdg vyegvaiggd rypgstfmdh vlryqdtpgv
kmivvlgeig
      721 gteeykicrg ikegrltkpi vcwcigtcat mfssevqfgh agacanqase
tavaknqalk
      781 eagvfvprsf delgeiiqsv yedlvangvi vpaqevpppt vpmdysware
lglirkpasf
      841 mtsicdergq eliyagmpit evfkeemgig gvlgllwfqk rlpkyscqfi
emclmvtadh
      901 gpavsgahnt iicaragkdl vssltsgllt igdrfggald aaakmfskaf
dsgiipmefv
      961 nkmkkegkli mgighrvksi nnpdmrvqil kdyvrqhfpa tplldyalev
ekittskkpn
     1021 lilnvdglig vafvdmlrnc gsftreeade yidigalngi fvlgrsmgfi
ghyldqkrlk
     1081 qglyrhpwdd isyvlpehms m
//
```

SNX12

Official Symbol: **SNX12**

Official Name: sorting nexin 12

Gene ID: 29934

Organism: Homo sapiens

Other Aliases: none

Other Designations: sorting nexin-12

Nucleotide sequence:

NCBI Reference Sequence: NM_001256185.1

LOCUS: NM_001256185

ACCESSION : NM_001256185

```
      1 ttttgattgc catttccttg ggacagcctg aagagagaat cgaaagaagt
tcttttcagt
       61 atggttgcat aacatcgagt cggagattgt gaaatgtcct ttgagaaaaa
tgtacgggca
      121 ggaagatgaa taggtgtctg tttgattaag gctctccttc ggaaagatgt
cggacacggc
      181 agtagctgat acccggcgcc ttaactcgaa gccgcaggac ctgaccgacg
cttacgggcc
      241 gccaagtaac ttcctggaga tcgacatctt taatcctcag acggtgggcg
tgggacgcgc
      301 gcgcttcacc acctatgagg ttcgcatgcg gacaaaccta cctatcttca
agctaaagga
      361 gtcctgcgta cggcggcgct acagtgactt tgagtggctg aaaaatgagc
tggagagaga
      421 tagcaagatt gtagtaccac cactgcctgg gaaagccttg aagcggcagc
tccctttccg
      481 aggagatgaa gggatctttg aggagtcttt catcgaagaa aggaggcagg
gcctcgagca
      541 gtttattaac aaaattgctg ggcacccact ggctcagaat gaacgctgcc
tacacatgtt
      601 cctgcaagag gaggcaattg acaggaacta cgtcccgggg aaggtgcgcc
agtaggagcc
      661 cctctcacca cctgccctct actttcctgc tgaaatgaca ttggttttta
cactaagcct
      721 ctctctgtct ttgatctgaa gttggctgcc catctcctgg cctgatagac
tgtctggcat
      781 tgtgctccct ggtacctgac tacaccatgt gggcactctg ctaggatcct
cttctctgag
      841 gagaggtggg aacccacagg cagatgccct ttgcttgggg ttggggtggg
tgggtggggg
      901 gattagactg gaaggcaatt tcttgggcat ttacccatgc cagaaggcta
acctgggggg
```

```
      961 agggggggcgc ttgtgctggt gaggcacttg gatacatact gatgctgcaa
gttcagggga
     1021 tttttcttac tcttaggttt aaccaagaac actgagcagg gaaaaaccct
gcctttccta
     1081 actgcatgta ttttttcctt tttggaaagg tggtagagac tcagaagctt
tccttgtttt
     1141 cttcaggcct gctcccagtt ttcttaacag tttcttttgt tgctttctct
ctcccttgtt
     1201 gctttccatg gcagtaatcc tcctagagtc caagcagtct gttgtatgga
gcagggtgtg
     1261 tgggttttct gggcccatca ttatggctgc ttcagagtca gaagaaagcc
atagggcagt
     1321 aggggagctc ctattgccta gcccctctcc ctttgtggct cccactctag
ctgcctattt
     1381 ttgctcatca gctggtgagt cagtatgggc cagcagttct ccctccctaa
gcccttgcta
     1441 ctttatgggt tagctttgca ggtttggtgg cttgaggggt gggggcaact
caccactgcc
     1501 aggtaactcc ctgaagggtg ggagtggatt atcttctagg ctcttacccg
cggtagggaa
     1561 gggcatcaac actgtcttcc ttccattctc ctttccccca tcccatttag
tgctgccaca
     1621 gggcagaagc acacaaacca accacacagt ctctgacttc tcctaagcac
tttgagttgt
     1681 tgaatggggc tcaggggcaa gagtttttgc tgccctcccc agcgtggtca
cagggttatt
     1741 gaactgcctg cacttgtttc tcatgcaact ccagcatttt ccccagaagt
tgaactatgg
     1801 atagcagctt ggtatggatt tcctaaatct taacatttga agcagcttct
tgaggctggc
     1861 aactatcctg gtttctgtct tggagggggt ggtttgtttg ctggggccca
acgtctgtcc
     1921 caagtggtgg ggtgagagta agttaacttt ggtgccaggt gagaggtggg
ggctctttgc
     1981 ttagactccc tatcatggaa agattggagt tttctatgca gggcactgtg
gaaaaggatt
     2041 gctgattctg actgaccctg atcagagaga ttaggattgt attttgacat
aggatttgga
     2101 acccatctaa atgttgaagt tccctgagac agctctccag ctgctgagcc
tgcgccaggg
     2161 gctaagcagc ccctaatgag aggctctgct ccctttccca cctcgccaat
gttgttgttg
     2221 ctgccttttt gatttgtatc ctctgttata gacatttttt aaaaacgatt
tcctctttca
     2281 ttgtgcacaa gtgctgagag tctgaggccc catttctgct gtgtatatat
atcctgactc
     2341 ggggctttta ttcagcaaac tgttcattct tctgtcagac aatgtcatat
tcaactctgt
     2401 tcatattaaa ccactgtgaa gcaaaaaaaa aaaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001243114.1

LOCUS: NP_001243114

ACCESSION: NP_001243114

```
      1 msdtavadtr rlnskpqdlt daygppsnfl eidifnpqtv gvgrarftty
evrmrtnlpi
       61 fklkescvrr rysdfewlkn elerdskivv pplpgkalkr qlpfrgdegi
feesfieerr
      121 qgleqfinki aghplaqner clhmflqeea idrnyvpgkv rq
//
```

SYNCRIP

Official Symbol: **SYNCRIP**

Official Name: synaptotagmin binding, cytoplasmic RNA interacting protein

Gene ID: 10492

Organism: Homo sapiens

Other Aliases: RP1-3J17.2, GRY-RBP, GRYRBP, HNRPQ1, NSAP1, PP68, hnRNP-Q

Other Designations: NS1-associated protein 1; glycine- and tyrosine-rich RNA-binding protein; heterogeneous nuclear ribonucleoprotein Q

Nucleotide sequence:

NCBI Reference Sequence: NM_001159673.1

LOCUS: NM_001159673

ACCESSION : NM_001159673

```
      1 cggcgtgagc ttcggccgcc attttacaac agctccactc gcgccggaca
cagggagcag
       61 cgagcacgcg tttcccgcaa cccgatacca tcggacagga tttctccgcc
tcagcccaac
      121 ggggagggct agttgcacat agtgatttag atgaaagagc tattgaagct
ttaaaagaat
      181 tcaatgaaga cggtgcattg gcagttcttc aacagtttaa agacagtgat
ctctctcatg
      241 ttcagaacaa aagtgccttt ttatgtggag tcatgaagac ttacaggcag
agagaaaaac
      301 aagggaccaa agtagcagat tctagtaaag gaccagatga ggcaaaaatt
aaggcactct
      361 tggaaagaac aggctacaca cttgatgtga ccactggaca gaggaagtat
ggaggaccac
      421 ctccagattc cgtttattca ggtcagcagc cttctgttgg cactgagata
tttgtgggaa
      481 agatcccaag agatctattt gaggatgaac ttgttccatt atttgagaaa
gctggaccta
      541 tatgggatct tcgtctaatg atggatccac tcactggtct caatagaggt
tatgcgtttg
      601 tcacttttg tacaaaagaa gcagctcagg aggctgttaa actgtataat
aatcatgaaa
```

661 ttcgttctgg aaaacatatt ggtgtctgca tctcagttgc caacaatagg
cttttgtgg
721 gctctattcc taagagtaaa accaaggaac agattcttga agaatttagc
aaagtaacag
781 agggtcttac agacgtcatt ttataccacc aaccggatga caagaaaaaa
aacagaggct
841 tttgctttct tgaatatgaa gatcacaaaa cagctgccca ggcaaggcgt
aggttaatga
901 gtggtaaagt caaggtctgg gggaatgttg gaactgttga atgggctgat
cctatagaag
961 atcctgatcc tgaggttatg gcaaaggtaa aagtgctgtt tgtacgcaac
cttgccaata
1021 ctgtaacaga agagatttta gaaaaggcat ttagtcagtt tgggaaactg
gaacgagtga
1081 agaagttaaa agattatgcg ttcattcatt ttgatgagcg agatggtgct
gtcaaggcta
1141 tggaagaaat gaatggcaaa gacttggagg gagaaaatat tgaaattgtt
tttgccaagc
1201 caccagatca gaaaaggaaa gaaagaaaag ctcagaggca agcagcaaaa
aatcaaatgt
1261 atgacgatta ctactattat ggtccacctc atatgccccc tccaacaaga
ggtcgagggc
1321 gtggaggtag aggtggttat ggatatcctc cagattatta tggatatgaa
gattattatg
1381 attattatgg ttatgattac cataactatc gtggtggata tgaagatcca
tactatggtt
1441 atgaagattt tcaagttgga gctagaggaa ggggtggtag aggagcaagg
ggtgctgctc
1501 catccagagg tcgtggggct gctcctcccc gcggtagagc cggttattca
cagagaggag
1561 gtcctggatc agcaagaggc gttcgaggtg cgagaggagg tgcccaacaa
caaagaggcc
1621 gcgggcaggg aaaaggggtc gaggccggtc ctgacctgtt acaatgaaga
ctgacttgct
1681 atgtgggatt acaccagaag cttgcagtgg agtaatggta aggaaatcaa
gcaaccttaa
1741 atatgtcggc tgtataggag catattctat tgcagaagac cttcctatga
agatcatgga
1801 atcaaatacg ggacattgaa ctaatacttg gactttgata tgaatttctt
taacaatttt
1861 ctctgcagtg caagttatta aactaaagct actctatttt caaaatgtgt
tccaacagaa
1921 atccttcata actcctagca tggtatctta ataaagaata aagttctttt
aaaaatctgc
1981 tctaagtaga tttttcccct tttttaaatt aaggatccca acagtggtat
tttgaaatat
2041 tctcttgaat ttgtgcattt aaattttatt gcagtggtat agatgaatgc
cactgatggt
2101 atccttaaat tttatttctg ctcaccaagg ttaatcatga ttgtctatat
ctttttttata
2161 gtgatcactt ttgaattgtg ttcagatatg cagtttcagg tgtaatcatc
agagctggtt
2221 agtcaggcat tccagatagt ggttcttttc agaacctttt taaaagggtt
ggttaactac
2281 ctcagtagca gaggattgaa ctataccctg tctgtactgt acatagaaaa
tctttgtaga
2341 taaaagcaag gcttgttaaa tatgatatga gggtaagatt ttaatatacc
aaatgtaaca
2401 ttcttagttg cctttagttt cagaggcttg taagacttcc tcatgaccat
cataacaggc

2461 cttgcttttg tcgtattttg tggctgaaaa agcagccttg cttcttcaga
tattgtagtt
2521 atttggatgt ataatagttt agcaagatgt tacttttgta agacatcaga
tgttcaaaaa
2581 agtgcatccg aacttgtact aaatactgca gtgtcccttt ataaaaagtc
agactaaaac
2641 tgacaattgt acagcgaagc ctgacatttg gatattttga agtttttca
taaatcatag
2701 aaattagtat atggctgtag tttagctttt taggtaaaag gtatgtttca
ttagtgcatt
2761 tcttcctgct gatcactgta aacatgtgaa tcagctttcc atttcttatg
caggtcatga
2821 taacttgtag agtagagtac aatcatttgt gctatgtttt taattttcta
aagcaccttg
2881 atgacagtga gtgtccagtg gtgaagcatc ctctattgaa ccaccctcaa
aaatttttt
2941 gccaagtcct aagttgatag cttaaagtaa aaagtgaaaa ttatagtttc
attaggactt
3001 ggtgtaaaga aatcccctcc ccccttcccc aaagggatac tgcagttata
tcacataccc
3061 aataggcacc acgatgaaga tcagagctta tacttaatta aggttttata
cacaccagtt
3121 ccccagtaaa tgcaaattta acaagaaaat cagacatgtc atatgttcaa
aatgctcatg
3181 gcaaacaatc attttgcatt cctgcaaata aaattgtttt atactgtaag
ctggaggcga
3241 gtgtaactta tttttgtaat aaagttttta tttttttttat gtgtcattaa
tataaatgtg
3301 tgttagtgta gaaatcttct ggtttaaaaa cttagaattg cacacatttc
agtatgttta
3361 tttgtactta cataatttta gaatagtggt tgccaatagc ctgtatgttt
cacattaatt
3421 ggtttttttgt tatctaaata aatcatttta gtatgttgta tgtcagttac
tgggatagct
3481 gggacataga gtgtaattta aaatttgtca ataagtattc attggaatat
atgtaaatgt
3541 gccttgccgg ttattgaaac ttatctacaa aatgagtatg gggtgacaaa
aattagttcc
3601 tggtgcttaa tgaaactttc tgccactgat tttatatatt accccgtgct
tttttaaagt
3661 acatctctct caaaacttag tgtaagtttg agggctacac aaaacattta
catttcattc
3721 taacataatg aatataatag gttgtggaaa gtgggtaaac taaatgtagc
cttcagtaaa
3781 attgaatctc agtgtaatcc ttggtgctgg catttctcag ttccgaggag
ttaaatgatc
3841 ccatctaaga ggtcattgcc atgcctattg gcactttact gtcatagcat
ttttaaggga
3901 cactgtcaag gtgtttaagt tctcagaatt acttgttggg attttaggac
aggtttgttt
3961 acttaaagta agaactgcat tgtcaaagtt gaaagaggaa cacttttgtg
agttcacaaa
4021 tgtgttctta agaaaacatt aaaatatgga gctctgggtt ttcaagacta
tttggcattc
4081 ttaatttggg gacttgggag ggaaactgat aaaaagaaat tgaagaattg
atggttatac
4141 ttaaagaagg gtaatgtaaa cagtggtgat gaaatatata cacatcaagt
gaaattactt
4201 gacagtgttc atttgaatga ctttgaattc aagccattat aattactttt
aaaattaaat

4261 atcatttgca ctgttctgat aatgggtgca gtttttgagc aatataatca gagctaaata
4321 tgcatgtagt gattagtgat gtgaacaatt aacgttctga gaagaaatac taactgtggt
4381 attttcaaac ttaaatttct gtagtaaaat cagtatcaaa gtcttatcag atcaaggaaa
4441 aacaggcaat gcatataaac atacttttga atgttgtgtg gcctataaag caataatgca
4501 atttatatgg aatgtcatgg gatatgagaa atggaaatgc aaaaataact aatcctttag
4561 taaaaatgtc aacatgttaa agggggaatg ttaactaatg taggttattg ctatttgtga
4621 tttgtttatg ggttcttggc tttgacagct tcaaagaatg gacagtgata agttaaaaga
4681 aattttgtat attgtcaagg aaagggtctt aaatccgagt caagtccctt ccttggggta
4741 aaaaatgtat tcttaaagca ttctgatgtt aaaaagaaaa cttaagttat ctaaccaaaa
4801 cagacgcaag attttgtttc tgcagactac ttggcaatca aaagtgatca taaatttagg
4861 ttatcagttt tcagaaagtt gctttgtgag aaaattttgt tagatatatt ctcccaagca
4921 tgctttttgt ggaaggtttt cagccattgc cactgaatca gatgttaaaa atgaagggaa
4981 aattgagtgt gcacacacac aactgttgta cactcatgat tgcagttttt agcttaagaa
5041 acttttctac cagttactgt gaatctgact taaaatgtaa agtttcctca tgataaaata
5101 ggaacaacat agaaatggat tgatggggtg atctgagtta ttgtatataa aagtttttaa
5161 agaatagaat gaacatcaag ctagataggc aaaaattgac acattcagaa cagcttttt
5221 gactgcgaag ccaaaagttg tcagaaacag caaaagatcc cttattatta cagagtattt
5281 tacgtagtct ctattttaag gagagaaatt aaatagaagg gcttcatgca tttaggggag
5341 ggtgctaaaa cttctcaagt tcgtcaaact tacaggaata cccaccatga tcatttctc
5401 tctaattatg tataccacaa aattttcatc tggccatagg aattcactgg tgggtgtaaa
5461 attaatgact aaagaaatta agtgacaaat acataaaaga aacagacttg tggggatatt
5521 gttttaaggt gtattaatta ctcagtgatg ataccactca atagggcatg ccactacttt
5581 tcttaagatg ctaattatga agcagtgctc acaggcattt tttaactagc aaattagtag
5641 atggactttt ggggtctgtc actttttaaa agtatttaag acttaaattc tattagcacc
5701 acagtctgcc ttcagtaata cacctaaaat atttttcagg accagaagca ttcagtttga
5761 aaatttgcag atgcaaacca gtattattac taacgctctg ggtcaaagat taggttttta
5821 atattaacag tagtctggta aatatttaga agtctggcat tgagaaacaa aagcttgtac
5881 ctgactagta tttttattta aaaaaattag ttctgttagc ttatttaaat tgtgttttat
5941 ttatccgtag aatttatatt tatttcattc ctttcatctc actgaaaact gtctgcaggc
6001 cctttgattt ggattagatg tgtgaagtac tgtcttttgc caaaaacctc aaattacctg

```
     6061 ttcttttcaa cgtagtgggt ttgtgcttgt ttggagatca gttcaaaaac
tatctgtact
     6121 atctgtactg cctctgatgt taagatttta tgtatagcat aaggaagcta
gctctgacta
     6181 tattttccta agaataaaga cctatttttg tagcatgtct taggatctcc
aggagtccaa
     6241 gaattattgt gggtgtcctc caattcatca ctcttcactt aacagctttt
aagtagacac
     6301 ttggaatctt tagaggtctg tcgccctttg attatccata cattcgaagt
aactagccaa
     6361 tggtgaaaaa ttcctcaaga tatcctcagt tgcaatcaca ttactggaag
atgaatagaa
     6421 taaatgtatt aggctggtct taatttttga tggaaatatt ctgttgtccc
gtacttgcca
     6481 ttggatttga taaagttagt ggtaatttgg aaagaatcgg ggacttgcca
atatatttgt
     6541 gggttttagc ttatacccct aggatttctt ggttgcggga cgagcagttt
tggccacttc
     6601 catcaggaca agacttttta ggtcacttag tgcaggtttt agtttctatt
ttggattaac
     6661 aacatttata ttgattatcg aaaagaagct ttcatcattt cagaacagtc
ctggaagttt
     6721 gactttgagt gtgggagaag tcctaataaa ccattttgga aattaaaaaa aaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001153145.1

LOCUS: NP_001153145

ACCESSION: NP_001153145

```
      1 mktyrqrekq gtkvadsskg pdeakikall ertgytldvt tgqrkyggpp
pdsvysgqqp
       61 svgteifvgk iprdlfedel vplfekagpi wdlrlmmdpl tglnrgyafv
tfctkeaaqe
      121 avklynnhei rsgkhigvci svannrlfvg sipksktkeq ileefskvte
gltdvilyhq
      181 pddkkknrgf cfleyedhkt aaqarrrlms gkvkvwgnvg tvewadpied
pdpevmakvk
      241 vlfvrnlant vteeilekaf sqfgklervk klkdyafihf derdgavkam
eemngkdleg
      301 enieivfakp pdqkrkerka qrqaaknqmy ddyyyygpph mppptrgrgr
ggrggygypp
      361 dyygyedyyd yygydyhnyr ggyedpyygy edfqvgargr ggrgargaap
srgrgaappr
      421 gragysqrgg pgsargvrga rggaqqqrgr gqgkgveagp dllq
//
```

SAR1B

Official Symbol: **SAR1B**

Official Name: SAR1 homolog B (S. cerevisiae)

Gene ID: 51128

Organism: Homo sapiens

Other Aliases: ANDD, CMRD, GTBPB, SARA2

Other Designations: 2310075M17Rik; GTP-binding protein B; GTP-binding protein SAR1b; GTP-binding protein Sara; SAR1a gene homolog 2

Nucleotide sequence:

NCBI Reference Sequence: NM_001033503.2

LOCUS: NM_001033503

ACCESSION : NM_001033503

```
      1 gccggcccgg aaggggctga tgcgaactgg ggccacggca gccatcgcgc
tttgcagttc
       61 ggtctcctgg tgtacggcca acgccaagta ggggattgcg ttccctccag
tcgcagagtt
      121 tccctcttgt cgcccaggct ggagtgaagt ggcacgatct cggcttactg
caagctccgc
      181 ctcccgggtt cacgccattc tcctgcctca gcctcccgag tagctgggac
tacagaccct
      241 atcagatttg gatatgtcct tcatatttga ttggatttac agtggtttca
gcagtgtgct
      301 acagttttta ggattatata agaaaactgg taaactggta tttcttggat
tggataatgc
      361 aggaaaaaca acattgctac acatgctaaa agatgacaga cttggacaac
atgtcccaac
      421 attacatccc acttccgaag aactgaccat tgctggcatg acgtttacaa
cttttgatct
      481 gggtggacat gttcaagctc gaagagtgtg gaaaaactac cttcctgcta
tcaatggcat
      541 tgtatttctg gtggattgtg cagaccacga aaggctgtta gagtcaaaag
aagaacttga
      601 ttcactaatg acagatgaaa ccattgctaa tgtgcctata ctgattcttg
ggaataagat
      661 cgacagacct gaagccatca gtgaagagag gttgcgagag atgtttggtt
tatatggtca
      721 gacaacagga aaggggagta tatctctgaa agaactgaat gcccgaccct
tagaagtttt
      781 catgtgtagt gtgctcaaaa gacaaggtta cggagaaggc ttccgctgga
tggcacagta
      841 cattgattaa cacaaactca cattggttcc aggtctcaac gttcaggctt
actcagagat
      901 ttgattgctc aacatgcata acttgaattc aatagacttt tgctggttat
aaaacagatg
      961 ttttttagat tattaatatt aaatcaactt aatttgaatg agaattgaaa
actgattcaa
     1021 gtaagtttga gtatcacaat gttagctttc taattccata aaagtacttg
gtttttacag
     1081 tttataatct gacatcaccc cagcgccatt tgtaaagagc aactttccag
cagtacattt
     1141 gaagcacttt ttaacaacat gaaactataa accatattta aaagctcatc
atgttaaatt
```

1201 ttttatgtac ttttctggaa ctagttttta aattttagat tatatgtcca cctatcttaa
1261 gtgtacagtt aataattagc ttattcaatg attgcatgat gccttacagt tttcaataac
1321 tttttttctt atgcaaacgt catgcaataa aacaaactct aatgtttggc atccttgttg
1381 ggcaaatgtt tcatttaaat gtgtcttatc tagctagtat actctgaaaa tttgagtatt
1441 taatattagg catatgaagt ggttgttggg aaaggagatt ccttcagaat atttagaata
1501 gcgtttaagg ctctcaaggc ttaggtattt accatgagat ggttgtggtc agttgcacct
1561 gataccatgt atccctgaat catgtgtatt tttattagta atgcagccac taccattgct
1621 gatggggtct gtgtcctagt ccctgtggga ttggccttct gaagaaaagc acatttatgg
1681 tacacaaggt tgattctcag tttggtaggc tctatagttc catcccagct gtctaattct
1741 gaagtgctcc agtcttattg gtgcccaatg gggtaatctg tgcctcttcc ccaaaagcaa
1801 tgacaggcac cagtgtctcc acatttagat atattcccag tctctgtaca tttaacagta
1861 ctctctgggg caagcaataa gtaggacctt gtgtgcagtt tcattgtaga gacatgtata
1921 tctgaggaat aaaacagctt gttctgtgtc agtctgagat atgtggagat tttgttccta
1981 tcctatagct cttttgtatt ctttggcata ttttatatcc tggtagaaga aaacaagtgc
2041 ttgtttccaa ttttcttttt tcttatttgc tctcaggtag ttcttactcc atacaacaaa
2101 gacttttttgt ttgttgggct ttttttttttt ttttttttttt gctctgtttc attttgtttt
2161 agagacgggg tttcaccatg ttgtccaggc tggtctcaaa ctcctgacct caagtgatct
2221 gcccgccttg gcctcccaaa gttctgggat tataggcgtg agctaccatg cctgacctga
2281 cttttgtttt tagatattat gcttgccatg tgatagggcc tgcaagcctc attgctaggc
2341 ttactaagaa aattttagtt tttcaaaagc attataattt cctaagaaac tgaattcttt
2401 ttttatatgt ttgagattcc catcattagt aatataagat gaaaggtaag tgccaaaaat
2461 gtatttttaa agaccctcaa gtttaagatt tatcctgatt ataagccaag ttttatagta
2521 tatttaacca attccatcaa gaataatttt aatatcaaaa attagtgttt tctgtagcca
2581 ttgtccatgt cagagttaca gtccttttg tcattgataa tataacctat gaagcagata
2641 aggattgagg aatatgactg gaaggaatta ctatttagct aagctgacaa ggtcgcttct
2701 taagatgaca tttggtttca gtaatctgac tattctgttt tcactttcat cttctttcta
2761 aatgaaaaca aaagtgctcc ctcccttcct ggaaacctca gtaacactat gggaaaagta
2821 gaacatgaca ttgcagccta ttgatttctt cttccagata ggtttaaagt actccttaag
2881 ttctgactaa atagaactaa gccttattaa aaataactgc ctcttgttca tgttatctgt
2941 accttcaggg acctgccttt cttcaagtat ttcctagagt atctattatg atactgaaga

3001 agctaattat ttgtgttgta aatgggtata aatgaaaaaa aaacatactg
gtttctctag
3061 ccaggaaaaa tgctttctgg tgtaatatat cttgctccag aaccctcatt
ctaattgtaa
3121 cactaggatc aaagaaacaa agtcactttg tggaccacag ctaaactgtg
gatattttcc
3181 caaagacata agatttttat ggcccgagcc tctagaaagg aagccatgtt
taggagcaac
3241 cagctttcct cccagcttta gggggcagag ttcctgagcc agaggactta
ctgtccagct
3301 ttgagaacct ctccagagta tatgcactgg gtactgctct ttttcaagag
aaccaaatta
3361 agaggatggc aagaaacagt agaagcacag aggaaagaca actctgcatg
tgcctgtgtg
3421 aatgtgtgca tccatggagt atttcccagg taaatactag tactggggac
ataggctaat
3481 tgtgtgtccc acactgcaag atgctagggc gtagttaaca ctgtggtata
catacaaatc
3541 aggcactgtc caaaagattt tttaaaatct aaagtctgaa atgtaaaaat
ataaggtctc
3601 aacccacttt tacactttta aagagatccc atacctgttt cactgactgc
cgttaattac
3661 acttttggat cacagctggt taaattgata gatttcagtt tatctcagtg
aatttttaga
3721 atggagatta tagcattttt taattggaga acagacattt cctaaagtat
atgaaaaaaa
3781 attattcact gttggtttaa accagtatct ttgtatgagt gccaaagata
tatgaacaca
3841 gatactgcct gtgcagacct aaattttagt tttgtgtacc tggatccata
tacaaatttt
3901 ttgtggttta tagcataaaa gcagaacgtt gtttctttct tagttttcaa
ccggctcatc
3961 ttttgttttt gttttttgtt tttttgtttt ttgttttttt tgagatggag
tcttgctttg
4021 ttgcccaggc tagagtgcag tggcacaatc tctgcttact gcaacctcca
cctccggggt
4081 tcaagtgatt ctcctgcctc agcctcctga gtagctggga ctacaggcgc
gcaccaccat
4141 gcccggctaa tttttgtatt tttagtagag atggggtttc actatgttgg
ccaggctggt
4201 ctcgaactcc tgacctcgtg atctgcccac ctcagcctcc caaagtgctg
ggattacagg
4261 cgtgagccgc catgcccggc caattttcaa ctggcccata ctttatagtg
atggaaagcg
4321 cataaactac ttgtaaatca ttaaaatagg gtgataactg tgataatagt
gtttcttgca
4381 ttctagaaaa ttattttatt aactacattc aaaacccagc atttcacagg
ttccatcatt
4441 agaaacagta tagttctagt taacatgatt ggagagtttc aggggaaagg
tttacatttt
4501 ctgaaactgt atttggtatg tgactcaatg tggtatttca gtcttgttag
tcacttacat
4561 gactgacgtt tgcaaggatt tattgccaag taaaatttga ccagagtgca
ctgagaatag
4621 ctacataagg ggaaatctct caaaattcct tctgttcatt taatttggag
catattgttt
4681 aaatcatttt aaacatatgt aaaaagttga agcattaaaa aatcttcaag
aaacaatgaa
4741 aaaatagaaa ttagcaaaca taagtttctt aatgcaaaat taatagtgaa
taaaatatag

4801 cctacattaa aagccagagg ctttgctata aatataagag tttagaaaaa
cagtgtgctt
4861 caattaagga ctaaattatc aaaactgcat gtttgttttt tcttttcttt
tctttttttt
4921 ttgagatgaa gtctcactct gttgcccagg ctggagtgca gtggtgcgat
cttggctcac
4981 tgcaacctct gcttcccagg ttcaagcgat tctcgtgcct caccatcccg
agtagctggg
5041 attacaggtg caccacacca tgcccagtta atttttgtat ttttagtaga
gacagggttt
5101 cattatttgg ccaggctggt ctcgagctcc tgatcttaag tgatccactc
gcctcggcct
5161 cccaaagtgc tgggattata cacatgagcc actgtgccca gcctaaaact
acatgttgaa
5221 gcttccggtc atttccatta ttatccttct tttgaaattc aagttagtgc
tttttaacca
5281 aataaaagaa gaaccagctc ttgggatatg tgactctgcc tctgtataaa
gtgactggaa
5341 ttttgttaaa accgtgtttc cacttctgaa ccctgttacc attccccctc
acaaatcccc
5401 acccaacacc tggattttaa agatcctcca gtgtcaaggg aagccacaga
gtctattaaa
5461 gaggcagttc tgaaccaatt aatttttgtc cttataattt agagcattaa
atagctaata
5521 tatttaatgg cactaattgt tgttcacggc tttcatcata cttttaaaca
gaatccaaag
5581 tattcaaagg aaagtaagcg aagttatcca aagccaactt tgtttcaggt
gtgtcccctg
5641 ccccaaatag attttagggc agaaatagaa aactgagttt acacagaact
atttttggaa
5701 aagctgcact ggagtagatg gattcttctt cagcatactt ttttgtttgt
ttgtttgaga
5761 tggagtcttg ctttgtcacc caggctggag tgcagtggtg tgatctccac
tcactgcaac
5821 ctccacctcc cagcttcaag tgattctcct gcctcaacct tccaagtagc
ttggattaca
5881 ggcgtgcgcc accacagctg gctaatattt gtattgttag tagagacagg
gtttcaccat
5941 gttgtccagg cttgtcgaac ttctgacctc acgtgatcca cctgcctcag
cctcccaaag
6001 tgctagatta taggcgtgaa ccactgcgcc cggccagcat gcattttaaa
agtggcttag
6061 atttagtttt aaatattttg gggtgaaagg caggaacagt tctgtttttg
acatacaggt
6121 tttctttggg attgttttca ttttcaagta tagattcatg tcagaatggc
caacttaacg
6181 tgggtttctg tattccctgg tgttgctctt aacctgaact cataatcagt
tgccatactg
6241 aggcaagagc actcagggtg aacatagtca agttacttta aaagtgataa
aagtgttttt
6301 ccatggtgaa accttcagta tttggctgaa tgtaaagtat gttgaagtgg
tatattgatg
6361 gtaagttgtt aatcactaac cttgtttgca cttttgtaca ccactgcttg
cactaggatc
6421 ttggtgtgaa ttttcaattg ttttacagtg tatacagatt attaaggata
atttatataa
6481 agatgtttct gtttaacttt gtgtgtttta caacaaagag ctataataga
tggttaaacg
6541 tttttgaatt gtgtttatat gttagtttga ttatgttcta ttatcttttc
acctgccatg
6601 aatttgagtg ttaggaaggg aaaaataaaa tactaatctg gtcttgaaga a

//

Protein sequence:

NCBI Reference Sequence: NP_001028675.1

LOCUS: NP_001028675

ACCESSION: NP_001028675

```
      1 msfifdwiys gfssvlqflg lykktgklvf lgldnagktt llhmlkddrl
gqhvptlhpt
       61 seeltiagmt fttfdlgghv qarrvwknyl paingivflv dcadherlle
skeeldslmt
      121 detianvpil ilgnkidrpe aiseerlrem fglygqttgk gsislkelna
rplevfmcsv
      181 lkrqgygegf rwmaqyid
//
```

CCDC47

Official Symbol: **CCDC47**

Official Name: coiled-coil domain containing 47

Gene ID: 57003

Organism: Homo sapiens

Other Aliases: GK001, MSTP041

Other Designations: coiled-coil domain-containing protein 47

Nucleotide sequence:

NCBI Reference Sequence: NM_020198.2

LOCUS: NM_020198

ACCESSION : NM_020198

```
      1 attatgtaat tttcccaaaa gccccacctc gcctcagccg ggcgggagag
agggaggtct
       61 cgcgctttcc ccggggttgc gtcccgcccc gcaggctgcg cgcaggcgct
gacgagccgc
      121 tcgcattcta cgtaacggac ggcggaggct acgtgaagag aggcgcggcg
tgactgagct
```

181 acggttctgg ctgcgtccta gaggcatccg gggcagtaaa accgctgcga
tcgcggaggc
241 ggcggccagg ccgagaggca ggccgggcag gggtgtcgga cgcagggcgc
tgggccgggt
301 ttcggcttcg gccacagctt tttttctcaa ggtgcaatga aagccttcca
cactttctgt
361 gttgtccttc tggtgtttgg gagtgtctct gaagccaagt ttgatgattt
tgaggatgag
421 gaggacatag tagagtatga tgataatgac ttcgctgaat ttgaggatgt
catggaagac
481 tctgttactg aatctcctca acgggtcata atcactgaag atgatgaaga
tgagaccact
541 gtggagttgg aagggcagga tgaaaaccaa gaaggagatt ttgaagatgc
agatacccag
601 gagggagata ctgagagtga accatatgat gatgaagaat ttgaaggtta
tgaagacaaa
661 ccagatactt cttctagcaa aaataaagac ccaataacga ttgttgatgt
tcctgcacac
721 ctccagaaca gctgggagag ttattatcta gaaattttga tggtgactgg
tctgcttgct
781 tatatcatga attacatcat tgggaagaat aaaaacagtc gccttgcaca
ggcctggttt
841 aacactcata gggagctttt ggagagcaac tttactttag tgggggatga
tggaactaac
901 aaagaagcca caagcacagg aaagttgaac caggagaatg agcacatcta
taacctgtgg
961 tgttctggtc gagtgtgctg tgagggcatg cttatccagc tgaggttcct
caagagacaa
1021 gacttactga atgtcctggc ccggatgatg aggccagtga gtgatcaagt
gcaaataaaa
1081 gtaaccatga atgatgaaga catggatacc tacgtatttg ctgttggcac
acggaaagcc
1141 ttggtgcgac tacagaaaga gatgcaggat ttgagtgagt tttgtagtga
taaacctaag
1201 tctggagcaa agtatggact gccggactct ttggccatcc tgtcagagat
gggagaagtc
1261 acagacggaa tgatggatac aaagatggtt cactttctta cacactatgc
tgacaagatt
1321 gaatctgttc atttttcaga ccagttctct ggtccaaaaa ttatgcaaga
ggaaggtcag
1381 cctttaaagc tacctgacac taagaggaca ctgttgttta catttaatgt
gcctggctca
1441 ggtaacactt acccaaagga tatggaggca ctgctacccc tgatgaacat
ggtgatttat
1501 tctattgata aagccaaaaa gttccgactc aacagagaag gcaaacaaaa
agcagataag
1561 aaccgtgccc gagtagaaga gaacttcttg aaactgacac atgtgcaaag
acaggaagca
1621 gcacagtctc ggcgggagga gaaaaaaaga gcagagaagg agcgaatcat
gaatgaggaa
1681 gatcctgaga aacagcgcag gctggaggag gctgcattga ggcgtgagca
aaagaagttg
1741 gaaaagaagc aaatgaaaat gaaacaaatc aaagtgaaag ccatgtaaag
ccatcccaga
1801 gatttgagtt ctgatgccac ctgtaagctc tgaattcaca ggaaacatga
aaaacgccag
1861 tccatttctc aaccttaaat ttcagacagt cttgggcaac tgagaaatcc
ttatttcatc
1921 atctactctg tttggggttt ggggttttac agagattgaa gatacctgga
aagggctctg

```
     1981 tttcaagaat ttttttttcc agataatcaa attattttga ttattttata
aaaggaatga
     2041 tctatgaaat ctgtgtaggt tttaaatatt ttaaaaatta taatacaaat
catcagtgct
     2101 tttagtactt cagtgtttaa agaaataccg tgaaatttat aggtagataa
ccagattgtt
     2161 gctttttgtt taaaccaagc agttgaaatg gctataaaga ctgactctaa
accaagattc
     2221 tgcaaataat gattggaatt gcacaataaa cattgcttga tgttttcttg
tatgtctaca
     2281 ttaaacttga gaaaaagtaa aaattagaac actgtatgta gtaatgaaat
ttcagggacc
     2341 cagaacataa tgtagtatat gtttttaggt gggagatgct gataacaaaa
ttaataggaa
     2401 gtctgtaggc attaggatac tgacatgtac atggaaaatt ctagggacag
gagcatcatt
     2461 ttttccttac ctgataccac gaaccagtga caacgtgaat gctgtatttt
aagtggttgt
     2521 atgtttattt tcttgagtaa caaatgcatg aaaaattaat gcttcaccta
ggtaagatca
     2581 ttggtctgtg tgaaatcaca aatgtttttt ccttcttggt tgctgcagcc
tggtggatgt
     2641 tcatggagaa gctctgttct ctatattatg gctgtgtgcc gttgcttctc
cctctgcttt
     2701 tatcttttcc acagttgagg ctgggtatgt tctttcaaag aaatggccat
gaatatgtgt
     2761 aagtatactt ttgaaaatga gctttcctaa actattgaga gttctttcca
cctcttgcgg
     2821 aaccaactct tggaggagag gcccatgtat ctgcacgagc acttagcttg
ttcagatctc
     2881 tgcattttat aaatgcttct taccaagaaa gcatttttag gtcattgctt
gtaccaggta
     2941 atttttgccg gggatgggta agggttgggt tttctggtgg gagtggggtg
gtgggtattt
     3001 tttgttgatg ctttagtgca ggcctgttct gaggcaataa caagttgctg
tgaaaacagc
     3061 atgtgctgct gcctttgtaa ctgcatggaa acttttcaca tgggtttttc
tccaagttaa
     3121 tacagaaata tgtaaactga gagatgcaaa tgtaatattt ttaacagttc
atgaagttgt
     3181 tattaaaata actaacataa aacttaatta ctttaatatt atataattat
agtagtggcc
     3241 ttgttttaca aacctttaaa ttacatttta gaaatcaaag ttgatagtct
tagttatctt
     3301 ttgagtaaga aaagctttcc taaagtccca tacatttgga ccatggcagc
taattttgta
     3361 acttaagcat tcatatgaac tacctatgga catctattaa agtgattgac
aaaatctcaa
     3421 aaaaaaaaaa aaaaaaaaaa aaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_064583.2

LOCUS: NP_064583

ACCESSION: NP_064583

```
      1 mkafhtfcvv llvfgsvsea kfddfedeed iveyddndfa efedvmedsv
tespqrviit
       61 eddedettve legqdenqeg dfedadtqeg dtesepydde efegyedkpd
tsssknkdpi
      121 tivdvpahlq nswesyylei lmvtgllayi mnyiigknkn srlaqawfnt
hrellesnft
      181 lvgddgtnke atstgklnqe nehiynlwcs grvccegmli qlrflkrqdl
lnvlarmmrp
      241 vsdqvqikvt mndedmdtyv favgtrkalv rlqkemqdls efcsdkpksg
akyglpdsla
      301 ilsemgevtd gmmdtkmvhf lthyadkies vhfsdqfsgp kimqeegqpl
klpdtkrtll
      361 ftfnvpgsgn typkdmeall plmnmviysi dkakkfrlnr egkqkadknr
arveenflkl
      421 thvqrqeaaq srreekkrae kerimneedp ekqrrleeaa lrreqkklek
kqmkmkqikv
      481 kam
//
```

PSMD12

Official Symbol: **PSMD12**

Official Name: proteasome (prosome, macropain) 26S subunit, non-ATPase, 12

Gene ID: 5718

Organism: Homo sapiens

Other Aliases: Rpn5, p55

Other Designations: 26S proteasome non-ATPase regulatory subunit 12; 26S proteasome regulatory subunit RPN5; 26S proteasome regulatory subunit p55

Nucleotide sequence:

NCBI Reference Sequence: NM_002816.3

LOCUS: NM_002816

ACCESSION : NM_002816 XM_942494 XM_946044 XM_946047 XM_946049 XM_946052 XM_946055 XM_946058

```
      1 ctgagcgggt gcgcgggcaa cttccggtgt gggtgacgag tggtggccga
agcaggggga
       61 cagcaaggga cgctcaggcg gggaccatgg cggacggcgg ctcggagcgg
gctgacgggc
      121 gcatcgtcaa gatggaggtg gactacagcg ccacggtgga tcagcgccta
cccgagtgtg
      181 cgaagctagc caaggaagga agacttcaag aagtcattga aacccttctc
tctctggaaa
      241 agcagactcg tactgcttcc gatatggtat cgacatcccg tatcttagtt
gcagtagtga
      301 agatgtgcta tgaggctaaa gaatgggatt tacttaatga aaatattatg
cttttgtcca
```

```
      361 aaaggcggag tcagttaaaa caagctgttg ccaaaatggt tcaacagtgc
tgtacttatg
      421 ttgaggaaat cacagacctt cctatcaaac ttcgattaat tgatactcta
cgaatggtta
      481 ccgaaggcaa gatttatgtt gaaattgagc gtgcgcgact gactaaaaca
ttagcaacta
      541 taaaagaaca aaatggtgat gtgaaagagg cagcctccat tttacaggag
ttacaggtgg
      601 aaacctacgg gtcaatggaa aagaaagagc gagtggaatt tattttggag
caaatgaggc
      661 tctgcctagc tgtgaaggat tacattcgaa cacaaatcat cagcaagaaa
attaacacca
      721 aatttttcca ggaagaaaat acagagaaat taaagttgaa gtactataat
ttaatgattc
      781 agctggatca acatgaggga tcctatttgt ctatttgtaa gcactacaga
gcaatatatg
      841 atactccctg tatacaggca gaaagtgaaa aatggcagca ggctctgaag
agtgttgtac
      901 tctatgttat cctggctcct tttgacaatg aacagtcaga tttggttcac
cgaataagtg
      961 gtgacaagaa gttagaagaa attcccaaat acaaggatct tttaaagctt
tttaccacaa
     1021 tggagttgat gcgttggtcc acacttgttg aggactatgg aatggaatta
agaaaaggtt
     1081 cccttgagag tcctgcaacg gatgtttttg gttctacaga ggaaggtgaa
aaaaggtgga
     1141 aagacttgaa gaacagagtt gttgaacata atattagaat aatggccaag
tattatactc
     1201 ggataacaat gaaaaggatg gcacagcttc tggatctatc tgttgatgag
tccgaagcct
     1261 ttctctcaaa tctagtagtt aacaagacca tctttgctaa agtagacaga
ttagcaggaa
     1321 ttatcaactt ccagagaccc aaggatccaa ataatttatt aaatgactgg
tctcagaaac
     1381 tgaactcatt aatgtctctg gttaacaaaa ctacgcatct catagccaaa
gaggagatga
     1441 tacataatct acaataaggg tcttagtgct ttagaaaaaa gttaaaattg
gaagtcatta
     1501 aaaaaagact gttataatgg tgtatatgtt ggggtttttt ttctaagctt
ctttgtctta
     1561 aattttaaaa tagtgaatat gtttgagact ccctttgacc tttcagttcc
ccaagttcat
     1621 tgttaacttt gcatttgcaa ttggtgcaaa aatacagatt tctgtcgtct
gaatacacaa
     1681 aaagttgtgt cataacttac ccagatatgt ttttctatca tttgaaacct
ttttagctac
     1741 tgtttgtttt cattcaacta acaaacatat tccaataata aaagcagtat
atacataaaa
     1801 aaaaaaaaaa aaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_002807.1

LOCUS: NP_002807

ACCESSION: NP_002807 XP_947587 XP_951137 XP_951140 XP_951142 XP_951145 XP_951148 XP_951151

```
      1 madggserad grivkmevdy satvdqrlpe caklakegrl qevietllsl
ekqtrtasdm
       61 vstsrilvav vkmcyeakew dllnenimll skrrsqlkqa vakmvqqcct
yveeitdlpi
      121 klrlidtlrm vtegkiyvei erarltktla tikeqngdvk eaasilqelq
vetygsmekk
      181 ervefileqm rlclavkdyi rtqiiskkin tkffqeente klklkyynlm
iqldqhegsy
      241 lsickhyrai ydtpciqaes ekwqqalksv vlyvilapfd neqsdlvhri
sgdkkleeip
      301 kykdllklft tmelmrwstl vedygmelrk gslespatdv fgsteegekr
wkdlknrvve
      361 hnirimakyy tritmkrmaq lldlsvdese aflsnlvvnk tifakvdrla
giinfqrpkd
      421 pnnllndwsq klnslmslvn ktthliakee mihnlq
            //
```

ATP5F1

Official Symbol: ATP5F1

Official Name: ATP synthase, H+ transporting, mitochondrial Fo complex, subunit B1

Gene ID: 515

Organism: Homo sapiens

Other Aliases: RP11-552M11.5, PIG47

Other Designations: ATP synthase B chain, mitochondrial; ATP synthase subunit b, mitochondrial; ATP synthase, H+ transporting, mitochondrial F0 complex, subunit B1; ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b; ATPase subunit b; H+-ATP synthase subunit b; cell proliferation-inducing protein 47

Nucleotide sequence:

NCBI Reference Sequence: NM_001688.4

LOCUS NM_001688

ACCESSION NM_001688

```
       1 actcccgggc cgccgggggc actaggggag gtggggtttc cttccgcatc
tccacggttc
       61 caactccaac ctagactcaa actggacgcc ggccggagac tccgctccgg
cagcaaaccc
      121 cacgtggtgc acctctgagc ctccgcccct ctcccgaggg aaccgcaact
ctacttctcg
      181 cgagaattgc ttctatggct ccatcctgct ttccggctgt cgccctcatg
cgataggctc
```

241 tcagcgttac ttgactcttc tcgcgataat ttttttaaa aatctcccaa
ggaaagttga
301 aggaagagta caaaattttc atctcgcgag acttgtgagc ggccatcttg
gtcctgccct
361 gacagattct cctatcgggg tcacagggac gctaagattg ctacctggac
tttcgttgac
421 catgctgtcc cgggtggtac tttccgccgc cgccacagcg gccccctctc
tgaagaatgc
481 agccttccta ggtccagggg tattgcaggc aacaaggacc tttcatacag
ggcagccaca
541 ccttgtccct gtaccacctc ttcctgaata cggaggaaaa gttcgttatg
gactgatccc
601 tgaggaattc ttccagtttc tttatcctaa aactggtgta acaggaccct
atgtactcgg
661 aactgggctt atcttgtacg ctttatccaa agaaatatat gtgattagcg
cagagacctt
721 cactgcccta tcagtactag gtgtaatggt ctatggaatt aaaaaatatg
gtccctttgt
781 tgcagacttt gctgataaac tcaatgagca aaaacttgcc caactagaag
aggcgaagca
841 ggcttccatc caacacatcc agaatgcaat tgatacggag aagtcacaac
aggcactggt
901 tcagaagcgc cattaccttt ttgatgtgca aaggaataac attgctatgg
ctttggaagt
961 tacttaccgg gaacgactgt atagagtata taaggaagta aagaatcgcc
tggactatca
1021 tatatctgtg cagaacatga tgcgtcgaaa ggaacaagaa cacatgataa
attgggtgga
1081 gaagcacgtg gtgcaaagca tctccacaca gcaggaaaag gagacaattg
ccaagtgcat
1141 tgcggaccta aagctgctgg caaagaaggc tcaagcacag ccagttatgt
aaatgtatct
1201 atcccaattg agacagctag aaacagttga ctgactaaat ggaaactagt
ctatttgaca
1261 aagtctttct gtgttggtgt ctactgaagt tatagtttac ccttcctaaa
aatgaaaagt
1321 ttgtttcata tagtgagaga acgaaatctc tatcggccag tcagatgttt
ctcatccttc
1381 ttgctctgcc tttgagttgt tccgtgatca cttctgaata agcagtttgc
ctttataaaa
1441 acttgctgcc tgactaaaga ttaacaggtt atagtttaaa tttgtaatta
attctaccat
1501 cttgcaataa agtgacaatt gaatgaaaca gggtttttca agttgtataa
ttctctgaaa
1561 tactcagctt ttgtcatatg ggtaaaaatt aaagatgtca ttgaactact
gtcttgttta
1621 tgagaccatt cagtggtgaa ctgtttctgg ctgataggtt atgagatatg
taaagctttc
1681 tagtactctt aaaataacta aatggagtat tatatatcaa ttcatatcat
tgactttatt
1741 attttagtag tatgcctata gaaaatatta tggactcaga gtgtcataaa
atcactctta
1801 agaatccatg cagcaggcca ggcacagtgg ctcacacctg taatgcctgc
actttggaag
1861 gccgagacag gcggatcact tgaggtcagg agtttgaaac cagccaggcc
aacacagtga
1921 aaccctgtct ctactaaaaa tacaaaaggt tagccgggca tggtggcagg
cgcctgtaat
1981 cccagctact caggaggctg aggcaggaga attgcttgaa cgcaggaggc
aaaggttgca

```
     2041 gtgagctgag atcacgccac tgcactccag cctgggcaac agacctcgac
tccatctaga
      2101 aaaaaaaaaa aaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001679.2

LOCUS NP_001679

ACCESSION NP_001679

```
      1 mlsrvvlsaa ataapslkna aflgpgvlqa trtfhtgqph lvpvpplpey
ggkvryglip
       61 eeffqflypk tgvtgpyvlg tglilyalsk eiyvisaetf talsvlgvmv
ygikkygpfv
      121 adfadklneq klaqleeakq asiqhiqnai dteksqqalv qkrhylfdvq
rnniamalev
      181 tyrerlyrvy kevknrldyh isvqnmmrrk eqehminwve khvvqsistq
qeketiakci
      241 adlkllakka qaqpvm
```

CMPK1

Official Symbol: CMPK1

Official Name: cytidine monophosphate (UMP-CMP) kinase 1, cytosolic

Gene ID: 51727

Organism: Homo sapiens

Other Aliases: RP11-511I2.1, CMK, CMPK, UMK, UMP-CMPK, UMPK

Other Designations: UMP-CMP kinase; UMP/CMP kinase; cytidylate kinase; deoxycytidylate kinase; uridine monophosphate kinase; uridine monophosphate/cytidine monophosphate kinase

Nucleotide sequence:

NCBI Reference Sequence (variant 1): NM_016308.2

LOCUS NM_016308

ACCESSION NM_016308

```
      1 gacagggccg cggacgcccg ggcagccacg gcggcggggc cgcggcgggc
gccggctcag
       61 cccgcccctt tctcccgccg cctccccgcc ccgccccgcg ccgcgccggc
cgctgtcagc
      121 tccctcagcg tccggccgag gcgcggtgta tgctgagccg ctgccgcagc
gggctgctcc
      181 acgtcctggg ccttagcttc ctgctgcaga cccgccggcc gattctcctc
tgctctccac
      241 gtctcatgaa gccgctggtc gtgttcgtcc tcggcggccc cggcgccggc
aaggggaccc
```

301 agtgcgcccg catcgtcgag aaatatggct acacacacct ttctgcagga
gagctgcttc
361 gtgatgaaag gaagaaccca gattcacagt atggtgaact tattgaaaag
tacattaaag
421 aaggaaagat tgtaccagtt gagataacca tcagtttatt aaagagggaa
atggatcaga
481 caatggctgc caatgctcag aagaataaat tcttgattga tgggtttcca
agaaatcaag
541 acaaccttca aggatggaac aagaccatgg atgggaaggc agatgtatct
ttcgttctct
601 tttttgactg taataatgag atttgtattg aacgatgtct tgagagggga
aagagtagtg
661 gtaggagtga tgacaacaga gagagcttgg aaaagagaat tcagacctac
cttcagtcaa
721 caaagccaat tattgactta tatgaagaaa tggggaaagt caagaaaata
gatgcttcta
781 aatctgttga tgaagttttt gatgaagttg tgcagatttt tgacaaggaa
ggctaattct
841 aaacctgaag gcatccttga aatcatgctt gaatattgct ttgatagctg
ctatcatgac
901 ccctttttaa ggcaattcta atctttcata actacatctc aattagtggc
tggaaagtac
961 atggtaaaac aaagtaaatt tttttatgtt ctttttttttg gtcacaggag
tagacagtga
1021 attcaggttt aacttcacct tagttatggt gctcaccaaa cgaagggtat
cagctatttt
1081 ttttaaaatt caaaaagaat atccctttta tagtttgtgc cttctgtgag
caaaactttt
1141 tagtacgcgt atatatccct ctagtaatca caacatttta ggatttaggg
atacccgctt
1201 cctctttttc ttgcaagttt taaatttcca accttaagtg aatttgtgga
ccaaatttca
1261 aaggaacttt ttgtgtagtc agttcttgca caatgtgttt ggtaaacaaa
ctcaaaatgg
1321 attcttagga gcattttagt gtttattaaa taactgacca tttgctgtag
aaagatgaga
1381 aaacttaagc tttgttttac tacaacttgt acaaagttgt atgacagggc
atattctttg
1441 cttccaagat ttgggttggg ggcactaggg gttcagagcc tggcagaatt
gtcagcttta
1501 gtctgacata atctaagggt atggggcaag gatcacatct aatgcttgtg
ttccttatac
1561 tctattatat agtgttattc atgattcagc tgatcttaac aaaattcgta
gcagtggaac
1621 cttgaaatgc atgtggctag atttatgcta aaatgattct cagttagcat
tttagtaaca
1681 cttcaaaggt tttttttttgt ttgttttcta gacttaataa aagcttagga
ttaattagaa
1741 gaagcaatct agttaaattt cccatttgta ttttattttc ttgaatactt
ttttcatagt
1801 tatttgttta aaaagattta aaaatcattg cactttggtc agaaaaataa
taaatatatc
1861 ttataaatgt ttgattccct tccttgctat ttttattcag tagatttttg
tttggcatca
1921 tgttgaagca ccgaaagata aatgattttt aaaaggctat agagtccaaa
ggaatattct
1981 tttacaccaa ttcttccttt aaaaatctct gaggaatttg ttttcgcctt
actttttttt
2041 cttctgtcac aatgctaagt ggtatccgag gttcttaata tgagatttaa
aatcttaaaa

```
     2101 tgtttcttat tttcagcact tacatcattt ggtacacagg gtcaaatagg
gcaaataatt
     2161 ttgtctttgt ataatagatt tgatatttaa agtcactgga aataggacaa
gttaatggat
     2221 gtttttatat tttaatagaa tcatttattt ctatgtgtta tgaaattcac
ttaatgataa
     2281 atttttcaac atacttgcca ttagaaaaca aagtattgct aagtactata
acatattggc
     2341 cactaaaatt catattgaga ttatcttggt ttcttggaag agataggaat
gagttcttat
     2401 ctagtgttgc aggccagcaa atacagaggt ggtttaatca aacagctcta
gtatgaagca
     2461 agagtaaaga ctaaggtttc gagagcattc ctactcacat aagtgaagaa
atctgtcaga
     2521 taggaatcta aatatttata gtgagattgt gaaagcaacc ttaaagtttt
gaagaagact
     2581 gatgagacta ggtgctttgc ttcctttcat caggtatctt tctgtggcat
ttgagaacag
     2641 aaaccaagaa acatggtaat tactaaatta tgaggctttg ctttttgttt
gcttttaagt
     2701 agaaaaacat gttggcaaca ttgagttttg gagttgattg agataatatg
acttaactag
     2761 ttttgtcatt ccatttgtta aagatacagt caccaagaat gttttgagtt
ttttgaaaga
     2821 ccccaattta agccttgctt atttttaaat tatttccatt cagtgatgtt
ggatgtatat
     2881 cagttattta gtaaataatc tcaataaatt ttgtgctgtg gcctttgcta
aaaaaaaaaa
      2941 aaaaaaaaaa aaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_057392.1

LOCUS NP_057392

ACCESSION NP_057392

```
      1 mlsrcrsgll hvlglsfllq trrpillcsp rlmkplvvfv lggpgagkgt
qcarivekyg
       61 ythlsagell rderknpdsq ygeliekyik egkivpveit isllkremdq
tmaanaqknk
      121 flidgfprnq dnlqgwnktm dgkadvsfvl ffdcnneici erclergkss
grsddnresl
      181 ekriqtylqs tkpiidlyee mgkvkkidas ksvdevfdev vqifdkeg
```

COX6B1

Official Symbol: COX6B1

Official Name: cytochrome c oxidase subunit VIb polypeptide 1 (ubiquitous)

Gene ID: 1340

Organism: Homo sapiens

Other Aliases: COX6B, COXG, COXVIb1

Other Designations: COX VIb-1; cytochrome c oxidase subunit 6B1

Nucleotide sequence:

NCBI Reference Sequence: NM_001863.4

LOCUS NM_001863

ACCESSION NM_001863

```
      1 tgggcgtggc ttgaatgact tcagtggcct cctcctggga gggagctgaa
gccgctcgca
       61 agactcccgt agtccccacc tctctcagct tccggctggt agtagttccg
cttcctgtcc
      121 gactgtggtg tctttgctga gggtcacatt gagctgcagg ttgaatccgg
ggtgccttta
      181 ggattcagca ccatggcgga agacatggag accaaaatca agaactacaa
gaccgcccct
      241 tttgacagcc gcttccccaa ccagaaccag actagaaact gctggcagaa
ctacctggac
      301 ttccaccgct gtcagaaggc aatgaccgct aaaggaggcg atatctctgt
gtgcgaatgg
      361 taccagcgtg tgtaccagtc cctctgcccc acatcctggg tcacagactg
ggatgagcaa
      421 cgggctgaag gcacgtttcc cgggaagatc tgaactggct gcatctccct
ttcctctgtc
      481 ctccatcctt ctcccaggat ggtgaagggg gacctggtac ccagtgatcc
ccacccccagg
      541 atcctaaatc atgacttacc tgctaataaa aactcattgg aaaagtgaga
```

Protein sequence:

NCBI Reference Sequence: NP_001854.1

LOCUS NP_001854

ACCESSION NP_001854

```
      1 maedmetkik nyktapfdsr fpnqnqtrnc wqnyldfhrc qkamtakggd
isvcewyqrv
       61 yqslcptswv tdwdeqraeg tfpgki
```

CTSA

Official Symbol: CTSA

Official Name: cathepsin A

Gene ID: 5476

Organism: Homo sapiens

Other Aliases: RP3-337O18.1, GLB2, GSL, NGBE, PPCA, PPGB

Other Designations: beta-galactosidase 2; beta-galactosidase protective protein; carboxypeptidase C; carboxypeptidase L; carboxypeptidase Y-like kininase; carboxypeptidase-L; deamidase; lysosomal carboxypeptidase A; lysosomal protective protein; protective protein cathepsin A; urinary kininase

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_000308.2

LOCUS NM_000308

ACCESSION NM_000308

```
      1 agagtgcacc cgaatccacg ggctcggagg cagcagccat ctctcggcca
tagggcaggc
       61 cagctggcgc cgggggctat tttgggcggc gggcaatgat ggtgaccgca
aggcgacctt
      121 gtaaggcatt tccccccctga ctcccttccc cgagcctctg cccgggggtc
ctagcgccgc
      181 tttctcagcc atcccgccta caacttagcc gtccacaaca ggatcatctg
atcgcgtgcg
      241 cccgggctac gatctgcgag gcccgcggac cttgacccgg cattgaccgc
caccgccccc
      301 caggtccgta gggaccaaag aaggggcggg aggaagactg tcacgtggcg
ccggagttca
      361 cgtgactcgt acacatgact tccagtcccc gggcgcctcc tggagagcaa
ggacgcgggg
      421 gagcagagat gatccgagcc gcgccgccgc cgctgttcct gctgctgctg
ctgctgctgc
      481 tgctagtgtc ctgggcgtcc cgaggcgagg cagcccccga ccaggacgag
atccagcgcc
      541 tccccgggct ggccaagcag ccgtctttcc gccagtactc cggctacctc
aaaggctccg
      601 gctccaagca cctccactac tggtttgtgg agtcccagaa ggatcccgag
aacagccctg
      661 tggtgctttg gctcaatggg ggtcccggct gcagctcact agatgggctc
ctcacagagc
      721 atggcccctt cctggtccag ccagatggtg tcaccctgga gtacaacccc
tattcttgga
      781 atctgattgc caatgtgtta tacctggagt ccccagctgg ggtgggcttc
tcctactccg
      841 atgacaagtt ttatgcaact aatgacactg aggtcgccca gagcaatttt
gaggcccttc
      901 aagatttctt ccgcctcttt ccggagtaca agaacaacaa acttttcctg
accggggaga
      961 gctatgctgg catctacatc cccaccctgg ccgtgctggt catgcaggat
cccagcatga
     1021 accttcaggg gctggctgtg ggcaatggac tctcctccta tgagcagaat
gacaactccc
     1081 tggtctactt tgcctactac catggccttc tggggaacag gctttggtct
tctctccaga
     1141 cccactgctg ctctcaaaac aagtgtaact tctatgacaa caaagacctg
gaatgcgtga
     1201 ccaatcttca ggaagtggcc cgcatcgtgg gcaactctgg cctcaacatc
tacaatctct
     1261 atgccccgtg tgctggaggg gtgcccagcc attttaggta tgagaaggac
actgttgtgg
     1321 tccaggattt gggcaacatc ttcactcgcc tgccactcaa gcggatgtgg
catcaggcac
```

```
     1381 tgctgcgctc aggggataaa gtgcgcatgg acccccccctg caccaacaca
acagctgctt
     1441 ccacctacct caacaacccg tacgtgcgga aggccctcaa catcccggag
cagctgccac
     1501 aatgggacat gtgcaacttt ctggtaaact tacagtaccg ccgtctctac
cgaagcatga
     1561 actcccagta tctgaagctg cttagctcac agaaatacca gatcctatta
tataatggag
     1621 atgtagacat ggcctgcaat ttcatggggg atgagtggtt tgtggattcc
ctcaaccaga
     1681 agatggaggt gcagcgccgg ccctggttag tgaagtacgg ggacagcggg
gagcagattg
     1741 ccggcttcgt gaaggagttc tcccacatcg cctttctcac gatcaagggc
gccggccaca
     1801 tggttcccac cgacaagccc ctcgctgcct tcaccatgtt ctcccgcttc
ctgaacaagc
     1861 agccatactg atgaccacag caaccagctc cacggcctga tgcagcccct
cccagcctct
     1921 cccgctagga gagtcctctt ctaagcaaag tgcccctgca ggccgggttc
tgccgccagg
     1981 actgcccccct tcccagagcc ctgtacatcc cagactgggc ccagggtctc
ccatagacag
     2041 cctgggggca agttagcact ttattcccgc agcagttcct gaatggggtg
gcctggcccc
     2101 ttctctgctt aaagaatgcc ctttatgatg cactgattcc atcccaggaa
cccaacagag
     2161 ctcaggacag cccacaggga ggtggtggac ggactgtaat tgatagattg
attatggaat
     2221 taaattgggt acagcttcaa aaaaaaaaaa aaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_000299.2

LOCUS NP_000299

ACCESSION NP_000299

```
      1 mtssprappg eqgrggaemi raappplfll lllllllvsw asrgeaapdq
deiqrlpgla
       61 kqpsfrqysg ylkgsgskhl hywfvesqkd penspvvlwl nggpgcssld
glltehgpfl
      121 vqpdgvtley npyswnlian vlylespagv gfsysddkfy atndtevaqs
nfealqdffr
      181 lfpeyknnkl fltgesyagi yiptlavlvm qdpsmnlqgl avgnglssye
qndnslvyfa
      241 yyhgllgnrl wsslqthccs qnkcnfydnk dlecvtnlqe varivgnsgl
niynlyapca
      301 ggvpshfrye kdtvvvqdlg niftrlplkr mwhqallrsg dkvrmdppct
nttaastyln
      361 npyvrkalni peqlpqwdmc nflvnlqyrr lyrsmnsqyl kllssqkyqi
llyngdvdma
      421 cnfmgdewfv dslnqkmevq rrpwlvkygd sgeqiagfvk efshiaflti
kgaghmvptd
      481 kplaaftmfs rflnkqpy
```

EPHX1

Official Symbol: EPHX1

Official Name: epoxide hydrolase 1, microsomal (xenobiotic)

Gene ID: 2052

Organism: Homo sapiens

Other Aliases: EPHX, EPOX, HYL1, MEH

Other Designations: epoxide hydratase; epoxide hydrolase 1

Nucleotide sequence: (variant 1)

NCBI Reference Sequence: NM_000120.3

LOCUS NM_000120

ACCESSION NM_000120

```
      1 cagaaggccg tggggagtgg gggccagtgc ctgcagcctg ccctgcctct
ctcacaggcc
       61 cttagagcat cgccaggtgc agagctccac agctctcttt cccaaggagt
aatcagaggg
      121 tgagaacgtg gagcctggtg gacaggtgaa agcactggga tctttctgcc
cagaaagggg
      181 aaagttgcac atttatatcc tagagggaag cgacagcagt gcttctccct
gtgctgaggt
      241 acaggagcca tgtggctaga aatcctcctc acttcagtgc tgggctttgc
catctactgg
      301 ttcatctccc gggacaaaga ggaaactttg ccacttgaag atgggtggtg
ggggccaggc
      361 acgaggtccg cagccaggga ggacgacagc atccgccctt tcaaggtgga
aacgtcagat
      421 gaggagatcc acgacttaca ccagaggatc gataagttcc gtttcacccc
acctttggag
      481 gacagctgct tccactatgg cttcaactcc aactacctga agaaagtcat
ctcctactgg
      541 cggaatgaat ttgactggaa gaagcaggtg gagattctca acagataccc
tcacttcaag
      601 actaagattg aagggctgga catccacttc atccacgtga agccccccca
gctgcccgca
      661 ggccataccc cgaagccctt gctgatggtg cacggctggc ccggctcttt
ctacgagttt
      721 tataagatca tcccactcct gactgacccc aagaaccatg gcctgagcga
tgagcacgtt
      781 tttgaagtca tctgcccttc catccctggc tatggcttct cagaggcatc
ctccaagaag
      841 gggttcaact cggtggccac cgccaggatc ttttacaagc tgatgctgcg
gctgggcttc
      901 caggaattct acattcaagg aggggactgg gggtccctga tctgcactaa
tatggcccag
      961 ctggtgccca gccacgtgaa aggcctgcac ttgaacatgg ctttggtttt
aagcaacttc
     1021 tctaccctga ccctcctcct gggacagcgt ttcgggaggt ttcttggcct
cactgagagg
     1081 gatgtggagc tgctgtaccc cgtcaaggag aaggtattct acagcctgat
gagggagagc
```

```
     1141 ggctacatgc acatccagtg caccaagcct gacaccgtag gctctgctct
gaatgactct
     1201 cctgtgggtc tggctgccta tattctagag aagttttcca cctggaccaa
tacggaattc
     1261 cgatacctgg aggatggagg cctggaaagg aagttctccc tggacgacct
gctgaccaac
     1321 gtcatgctct actggacaac aggcaccatc atctcctccc agcgcttcta
caaggagaac
     1381 ctgggacagg gctggatgac ccagaagcat gagcggatga aggtctatgt
gcccactggc
     1441 ttctctgcct tccctttga gctattgcac acgcctgaaa agtgggtgag
gttcaagtac
     1501 ccaaagctca tctcctattc ctacatggtt cgtgggggcc actttgcggc
ctttgaggag
     1561 ccggagctgc tcgcccagga catccgcaag ttcctgtcgg tgctggagcg
gcaatgaccc
     1621 acccctctcc ccccgcctgc cacctccccc cacaagtgcc ctccaggctt
ttcttgggga
     1681 agataccccc tttctgagga atgagtttgc ctccgtcccc tgcccatgct
gggagcccac
     1741 gctcaccccc tcacccctcc aagctcactc cccaaccccc aactccgtgt
ggtaagcaac
      1801 atggctttga tgataaacga ctttactcta aaaaaaaaaa aaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_000111.1

LOCUS NP_000111

ACCESSION NP_000111

```
      1 mwleilltsv lgfaiywfis rdkeetlple dgwwgpgtrs aareddsirp
fkvetsdeei
       61 hdlhqridkf rftppledsc fhygfnsnyl kkvisywrne fdwkkqveil
nryphfktki
      121 egldihfihv kppqlpaght pkpllmvhgw pgsfyefyki iplltdpknh
glsdehvfev
      181 icpsipgygf seasskkgfn svatarifyk lmlrlgfqef yiqggdwgsl
ictnmaqlvp
      241 shvkglhlnm alvlsnfstl tlllgqrfgr flglterdve llypvkekvf
yslmresgym
      301 hiqctkpdtv gsalndspvg laayilekfs twtntefryl edgglerkfs
lddlltnvml
      361 ywttgtiiss qrfykenlgq gwmtqkherm kvyvptgfsa fpfellhtpe
kwvrfkypkl
      421 isysymvrgg hfaafeepel laqdirkfls vlerq
```

ATP5B

Official Symbol: ATP5B

Official Name: ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide

Gene ID: 506

Organism: Homo sapiens

Other Aliases: ATPMB, ATPSB

Other Designations: ATP synthase subunit beta, mitochondrial; mitochondrial ATP synthase beta subunit; mitochondrial ATP synthetase, beta subunit

Nucleotide sequence:

NCBI Reference Sequence: NM_001686.3

LOCUS NM_001686

ACCESSION NM_001686

```
      1 agttcaccca atggacctgc ctactgcagc gtaggcctcg cctcaacggc
aggagagcag
       61 gcggctgcgg ttgctgcagc cttcagtctc cacccggact acgccatgtt
ggggtttgtg
      121 ggtcgggtgg ccgctgctcc ggcctccggg gccttgcgga gactcacccc
ttcagcgtcg
      181 ctgccccccag ctcagctctt actgcgggcc gctccgacgg cggtccatcc
tgtcagggac
      241 tatgcggcgc aaacatctcc ttcgccaaaa gcaggcgccg ccaccgggcg
catcgtggcg
      301 gtcattggcg cagtggtgga cgtccagttt gatgagggac taccaccaat
tctaaatgcc
      361 ctggaagtgc aaggcaggga gaccagactg gttttggagg tggcccagca
tttgggtgag
      421 agcacagtaa ggactattgc tatggatggt acagaaggct tggttagagg
ccagaaagta
      481 ctggattctg gtgcaccaat caaaattcct gttggtcctg agactttggg
cagaatcatg
      541 aatgtcattg gagaacctat tgatgaaaga ggtcccatca aaaccaaaca
atttgctccc
      601 attcatgctg aggctccaga gttcatggaa atgagtgttg agcaggaaat
tctggtgact
      661 ggtatcaagg ttgtcgatct gctagctccc tatgccaagg gtggcaaaat
tgggcttttt
      721 ggtggtgctg gagttggcaa gactgtactg atcatggagt taatcaacaa
tgtcgccaaa
      781 gcccatggtg gttactctgt gtttgctggt gttggtgaga ggacccgtga
aggcaatgat
      841 ttataccatg aaatgattga atctggtgtt atcaacttaa aagatgccac
ctctaaggta
      901 gcgctggtat atggtcaaat gaatgaacca cctggtgctc gtgcccgggt
agctctgact
      961 gggctgactg tggctgaata cttcagagac caagaaggtc aagatgtact
gctatttatt
     1021 gataacatct ttcgcttcac ccaggctggt tcagaggtgt ctgcattatt
gggccgaatc
     1081 ccttctgctg tgggctatca gcctaccctg gccactgaca tgggtactat
gcaggaaaga
     1141 attaccacta ccaagaaggg atctatcacc tctgtacagg ctatctatgt
gcctgctgat
     1201 gacttgactg accctgcccc tgctactacg tttgcccatt tggatgctac
cactgtactg
     1261 tcgcgtgcca ttgctgagct gggcatctat ccagctgtgg atcctctaga
ctccacctct
```

```
     1321 cgtatcatgg atcccaacat tgttggcagt gagcattacg atgttgcccg
tggggtgcaa
     1381 aagatcctgc aggactacaa atccctccag gatatcattg ccatcctggg
tatggatgaa
     1441 ctttctgagg aagacaagtt gaccgtgtcc cgtgcacgga aaatacagcg
tttcttgtct
     1501 cagccattcc aggttgctga ggtcttcaca ggtcatatgg ggaagctggt
acccctgaag
     1561 gagaccatca aaggattcca gcagattttg gcaggtgaat atgaccatct
cccagaacag
     1621 gccttctata tggtgggacc cattgaagaa gctgtggcaa aagctgataa
gctggctgaa
     1681 gagcattcat cgtgaggggt ctttgtcctc tgtactgtct ctctccttgc
ccctaaccca
     1741 aaaagcttca tttttctgtg taggctgcac aagagccttg attgaagata
tattctttct
      1801 gaacagtatt taaggtttcc aataaaatgt acacccctca gaaaaaaaaa
aaaaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_001677.2

LOCUS NP_001677

ACCESSION NP_001677

```
      1 mlgfvgrvaa apasgalrrl tpsaslppaq lllraaptav hpvrdyaaqt
spspkagaat
       61 grivavigav vdvqfdeglp pilnalevqg retrlvleva qhlgestvrt
iamdgteglv
      121 rgqkvldsga pikipvgpet lgrimnvige pidergpikt kqfapihaea
pefmemsveq
      181 eilvtgikvv dllapyakgg kiglfggagv gktvlimeli nnvakahggy
svfagvgert
      241 regndlyhem iesgvinlkd atskvalvyg qmneppgara rvaltgltva
eyfrdqegqd
      301 vllfidnifr ftqagsevsa llgripsavg yqptlatdmg tmqeritttk
kgsitsvqai
      361 yvpaddltdp apattfahld attvlsraia elgiypavdp ldstsrimdp
nivgsehydv
      421 argvqkilqd ykslqdiiai lgmdelseed kltvsrarki qrflsqpfqv
aevftghmgk
      481 lvplketikg fqqilageyd hlpeqafymv gpieeavaka dklaeehss
```

ATP5D

Official Symbol: ATP5D

Official Name: ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit

Gene ID: 513

Organism: Homo sapiens

Other Aliases: None currently listed

Other Designations: ATP synthase subunit delta, mitochondrial; F-ATPase delta subunit; mitochondrial ATP synthase complex delta-subunit precusor; mitochondrial ATP synthase, delta subunit

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001687.4

LOCUS NM_001687

ACCESSION NM_001687

```
      1 cagacgtccc tgcgcgtcgt cctcctcgcc ctccaggccg cccgcgccgc
gccggagtcc
       61 gctgtccgcc agctacccgc ttcctgccgc ccgccgctgc catgctgccc
gccgcgctgc
      121 tccgccgccc gggacttggc cgcctcgtcc gccacgcccg tgcctatgcc
gaggccgccg
      181 ccgccccggc tgccgcctct ggccccaacc agatgtcctt caccttcgcc
tctcccacgc
      241 aggtgttctt caacggtgcc aacgtccggc aggtggacgt gcccacgctg
accggagcct
      301 tcggcatcct ggcggcccac gtgcccacgc tgcaggtcct gcggccgggg
ctggtcgtgg
      361 tgcatgcaga ggacggcacc acctccaaat actttgtgag cagcggttcc
atcgcagtga
      421 acgccgactc ttcggtgcag ttgttggccg aagaggccgt gacgctggac
atgttggacc
      481 tgggggcagc caaggcaaac ttggagaagg cccaggcgga gctggtgggg
acagctgacg
      541 aggccacgcg ggcagagatc cagatccgaa tcgaggccaa cgaggccctg
gtgaaggccc
      601 tggagtaggc ggtgcgtacc cggtgtcccg aggcccggcc aggggctggg
cagggatgcc
      661 aggtgggccc agccagctcc tggggtcccg gccacctggg gaagccgcgc
ctgccaagga
      721 ggccaccaga gggcagtgca ggcttctgcc tgggccccag gccctgcctg
tgttgaaagc
      781 tctggggact gggccaggga agctcctcct cagctttgag ctgtggctgc
cacccatggg
      841 gctctccttc cgcctctcaa gatcccccca gcctgacggg ccgcttacca
tcccctctgc
      901 cctgcagagc cagccgccaa ggttgacctc agcttcggag ccacctctgg
atgaactgcc
      961 cccagccccc gccccattaa agacccggaa gcctgaaaaa aaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001678.1

LOCUS NP_001678

ACCESSION NP_001678

```
      1 mlpaallrrp glgrlvrhar ayaeaaaapa aasgpnqmsf tfasptqvff
nganvrqvdv
       61 ptltgafgil aahvptlqvl rpglvvvhae dgttskyfvs sgsiavnads
svqllaeeav
      121 tldmldlgaa kanlekaqae lvgtadeatr aeiqiriean ealvkale
```

CAPN1

Official Symbol: CAPN1

Official Name: calpain 1, (mu/I) large subunit

Gene ID: 823

Organism: Homo sapiens

Other Aliases: PIG30, CANP, CANP1, CANPL1, muCANP, muCL

Other Designations: CANP 1; calcium-activated neutral proteinase 1; calpain mu-type; calpain, large polypeptide L1; calpain-1 catalytic subunit; calpain-1 large subunit; cell proliferation-inducing gene 30 protein; cell proliferation-inducing protein 30; micromolar-calpain

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001198868.1

LOCUS NM_001198868

ACCESSION NM_001198868

```
      1 agggacttac ccaaggtcac gcagcgagcc cggtcccccc gcgttccccg
gggagcgctg
       61 agccgggacg cggcggtggg gtggggaagg ggagtggcgc ggccctgcgg
ggtgaggctg
      121 ccgtttgctg agtgtccggc aggggtctgc tcgctgccag cccggccccc
cctcagagca
      181 gctgccgcag cccgaggatg tcggaggaga tcatcacgcc ggtgtactgc
actggggtgt
      241 cagcccaagt gcagaagcag cgggccaggg agctgggcct gggccgccat
gagaatgcca
      301 tcaagtacct gggccaggat tatgagcagc tgcgggtgcg atgcctgcag
agtgggaccc
      361 tcttccgtga tgaggccttc cccccggtac cccagagcct gggttacaag
gacctgggtc
      421 ccaattcctc caagacctat ggcatcaagt ggaagcgtcc cacggaactg
ctgtcaaacc
      481 cccagttcat tgtggatgga gctacccgca cagacatctg ccagggagca
ctgggggact
      541 gctggctctt ggcggccatc gcctccctca ctctcaacga caccctcctg
caccgagtgg
      601 ttccgcacgg ccagagcttc cagaatggct atgccggcat cttccatttc
cagctgtggc
      661 aatttgggga gtgggtggac gtggtcgtgg atgacctgct gcccatcaag
gacgggaagc
      721 tagtgttcgt gcactctgcc gaaggcaacg agttctggag cgccctgctt
gagaaggcct
      781 atgccaaggt aaatggcagc tacgaggccc tgtcaggggg cagcacctca
gagggctttg
```

841 aggacttcac aggcggggtt accgagtggt acgagttgcg caaggctccc
agtgacctct
901 accagatcat cctcaaggcg ctggagcggg gctccctgct gggctgctcc
atagacatct
961 ccagcgttct agacatggag gccatcactt tcaagaagtt ggtgaagggc
catgcctact
1021 ctgtgaccgg ggccaagcag gtgaactacc gaggccaggt ggtgagcctg
atccggatgc
1081 ggaacccctg gggcgaggtg gagtggacgg gagcctggag cgacagctcc
tcagagtgga
1141 acaacgtgga cccatatgaa cgggaccagc tccgggtcaa gatggaggac
ggggagttct
1201 ggatgtcatt ccgagacttc atgcgggagt tcacccgcct ggagatctgc
aacctcacac
1261 ccgacgccct caagagccgg accatccgca aatggaacac cacactctac
gaaggcacct
1321 ggcggcgggg gagcaccgcg gggggctgcc gaaactaccc agccaccttc
tgggtgaacc
1381 ctcagttcaa gatccggctg gatgagacgg atgacccgga cgactacggg
gaccgcgagt
1441 caggctgcag cttcgtgctc gcccttatgc agaagcaccg tcgccgcgag
cgccgcttcg
1501 gccgcgacat ggagactatt ggcttcgcgg tctacgaggt ccctccggag
ctggtgggcc
1561 agccggccgt acacttgaag cgtgacttct tcctggccaa tgcgtctcgg
gcgcgctcag
1621 agcagttcat caacctgcga gaggtcagca cccgcttccg cctgccaccc
ggggagtatg
1681 tggtggtgcc ctccaccttc gagcccaaca aggagggcga cttcgtgctg
cgcttcttct
1741 cagagaagag tgctgggact gtggagctgg atgaccagat ccaggccaat
ctccccgatg
1801 agcaagtgct ctcagaagag gagattgacg agaacttcaa ggccctcttc
aggcagctgg
1861 caggggagga catggagatc agcgtgaagg agttgcggac aatcctcaat
aggatcatca
1921 gcaaacacaa agacctgcgg accaagggct tcagcctaga gtcgtgccgc
agcatggtga
1981 acctcatgga tcgtgatggc aatgggaagc tgggcctggt ggagttcaac
atcctgtgga
2041 accgcatccg gaattacctg tccatcttcc ggaagtttga cctggacaag
tcgggcagca
2101 tgagtgccta cgagatgcgg atggccattg agtcggcagg cttcaagctc
aacaagaagc
2161 tgtacgagct catcatcacc cgctactcgg agcccgacct ggcggtcgac
tttgacaatt
2221 tcgtttgctg cctggtgcgg ctagagacca tgttccgatt tttcaaaact
ctggacacag
2281 atctggatgg agttgtgacc tttgacttgt ttaagtggtt gcagctgacc
atgtttgcat
2341 gaggcaggga ctcggtcccc cttgccgtgc tcccctccct cctcgtctgc
caagcctcgc
2401 ctcctaccac accacaccag gccaccccag ctgcaagtgc cttccttgga
gcagagaggc
2461 agcctcgtcc tcctgtcccc tctcctccca gccaccatcg ttcatctgct
ccgggcagaa
2521 ctgtgtggcc cctgcctgtg ccagccatgg gctcgggatg gactccctgg
gccccaccca
2581 ttgccaagcc aggaaggcag ctttcgcttg ttcctgcctc gggacagccc
cgggtttccc

```
     2641 cagcatcctg atgtgtcccc tctccccact tcagaggcca cccactcagc
accaccggcc
     2701 tggccttgcc tgcagactat aaactataac cactagctcg acacagtctg
cagtccaggc
     2761 gtgtggagcc gcctcccggc tcggggaggc cccggggctg ggaacgcctg
tgccttcctg
     2821 cgccgaagcc aacgcccccc ctgtccttcc ctggccctgc tgccgaccag
gagctgccca
     2881 gcctgtgggc ggtcggcctt ccctccttcg ctccttttt atattagtga
ttttaaaggg
     2941 gactcttcag ggacttgtgt actggttatg ggggtgccag aggcactagg
cttggggtgg
     3001 ggaggtcccg tgttccatat agaggaaccc caaataataa aaggccccac
atctgtctgt
     3061 gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa
      3121 aaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001185797.1

LOCUS NP_001185797

ACCESSION NP_001185797

```
      1 mseeiitpvy ctgvsaqvqk qrarelglgr henaikylgq dyeqlrvrcl
qsgtlfrdea
       61 fppvpqslgy kdlgpnsskt ygikwkrpte llsnpqfivd gatrtdicqg
algdcwllaa
      121 iasltlndtl lhrvvphgqs fqngyagifh fqlwqfgewv dvvvddllpi
kdgklvfvhs
      181 aegnefwsal lekayakvng syealsggst segfedftgg vtewyelrka
psdlyqiilk
      241 alergsllgc sidissvldm eaitfkklvk ghaysvtgak qvnyrgqvvs
lirmrnpwge
      301 vewtgawsds ssewnnvdpy erdqlrvkme dgefwmsfrd fmreftrlei
cnltpdalks
      361 rtirkwnttl yegtwrrgst aggcrnypat fwvnpqfkir ldetddpddy
gdresgcsfv
      421 lalmqkhrrr errfgrdmet igfavyevpp elvgqpavhl krdfflanas
rarseqfinl
      481 revstrfrlp pgeyvvvpst fepnkegdfv lrffseksag tvelddqiqa
nlpdeqvlse
      541 eeidenfkal frqlagedme isvkelrtil nriiskhkdl rtkgfslesc
rsmvnlmdrd
      601 gngklglvef nilwnrirny lsifrkfdld ksgsmsayem rmaiesagfk
lnkklyelii
      661 trysepdlav dfdnfvcclv rletmfrffk tldtdldgvv tfdlfkwlql tmfa
```

CAPZA2

Official Symbol: CAPZA2

Official Name: capping protein (actin filament) muscle Z-line, alpha 2

Gene ID: 830

Organism: Homo sapiens

Other Aliases: CAPPA2, CAPZ

Other Designations: F-actin capping protein alpha-2 subunit; F-actin-capping protein subunit alpha-2; capZ alpha-2

Nucleotide sequence:

NCBI Reference Sequence: NM_006136.2

LOCUS NM_006136

ACCESSION NM_006136

```
      1 cccctccctt agcgggggcg cgcggcgctg aggaccgcac ggaaacgggg
aagtcaggtg
       61 gccgctgccg ccgccgccgc cgcggtttgt cgccagaagg aagatggcgg
atctggagga
      121 gcagttgtct gatgaagaga aggtgcgtat agcagcaaaa ttcatcattc
atgcccctcc
      181 tggagaattt aatgaggttt tcaatgatgt tcggttactg cttaataatg
acaatcttct
      241 cagggaagga gcagcccatg catttgcaca gtataacttg gaccagttta
ctccagtaaa
      301 aattgaaggt tatgaagatc aggtattgat aacagaacat ggcgacttgg
gaaatggaaa
      361 gtttttggat ccaaagaaca gaatctgttt taaatttgat cacttaagga
aggaggcaac
      421 tgatccaaga ccctgtgaag tagaaaatgc agttgaatca tggagaactt
cagtagaaac
      481 tgctctgaga gcttacgtaa aagaacatta cccgaatgga gtctgcactg
tgtatggcaa
      541 aaaaatagat ggacagcaaa ccattattgc atgcatagaa agccatcagt
tccaagcaaa
      601 aaatttttgg aatggtcgtt ggaggtcaga atggaagttt acaatcactc
cttcaaccac
      661 tcaagtggtt ggcatcttga aaattcaggt tcattattat gaagatggta
atgttcagct
      721 agtgagtcat aaagatatac aagattccct aacagtgtct aatgaagtgc
aaacagcaaa
      781 agaatttata aagattgtag aagctgcaga aaatgaatac cagactgcca
tcagtgagaa
      841 ttatcagaca atgtcggaca ctactttcaa agccttacgt cgacagttgc
cagttacacg
      901 cactaagatt gattggaaca agatccttag ctacaagatt ggcaaagaga
tgcagaatgc
      961 ataagatgaa cattgcatga ccggatcatt ttagtgtctt tgcgttaaaa
aatcattgca
     1021 aaagtattct gaactgtcaa gctgcccagt cagatgggct gttgccattt
aaaatcactg
     1081 taattaatta gtttgattag agcacaaagc ttagctaatc aaccattatt
tttcattttg
     1141 tttgttctaa gaggattgaa aatcagttta gtttaaatgt ctttctgtta
ggcctttctt
     1201 tcttacaatg aagagatgat tcttctagtt tatggttaaa agtttttgaa
gtgtctcaaa
```

```
     1261 aatattttac taactgtaac cctaaaattg atgtcttttg gtttatgaaa
tcagtaattt
     1321 ttgatatttc cccagttctt tttaatgggg tcaataatgg acattctagt
ttaaggtggt
     1381 tgatggattt agccatatat gctgctaaag aaattgtcta ccttttcttc
ctcacctgtt
     1441 ccatttatgt aaagttgaga ttagagggaa agcattttct atatcaattg
tgtttaaacc
     1501 tttcaagaag gttatttagc tagcttagtg ttgaactaaa tttttttaa
acaaggcaag
     1561 gtctaatgct gttttgagat tctgaaatta atgaaaatac ttatttcaga
aatgcattta
     1621 atgctttttt tcttgtgaca gttacgcaaa tcagcttgaa ttccatatgt
ccctgagtta
     1681 tttttatcat aaagccacaa atgtattata acaaggcaaa ttgtaatata
tataatcctg
     1741 aactcatgac catgtctcgg tttatttttt ttttcttgga ttgaaaagta
ctgaaattca
     1801 atgtgacatt aaaatgcaaa ttttcctatt tatttgagta gaaaatcact
taccagtgag
     1861 catatatatt ttaaaatact ttctttggat attgtaattc ttaactggtt
gtaaattaga
     1921 aaagctggga ttacatatgg tgtgcggtta cagtctaaat tttttcatcc
tcctatgcat
     1981 cataagcatg tttgtaatat tttcaaaaat agttctactg atgctacagg
aatttcaagc
     2041 ctgtggtgaa tgttagtatt taccataggg agtgaagtgg agttatggtt
tcattcaata
     2101 gagtattgct gattatactt gagtggaatc ctttcctcac gtactcccac
agacgtctgg
     2161 gcctggaaat tttttttta ttttatttta ttgtttttt ttttagaaaa
acaccacttt
     2221 tattatgtac aataaaatat ttcattagct tgaattgtat agatttttaa
aaattcaatg
     2281 aaagcatgtt gtttaatttc tttttaaaat cactgttggg ctttgaaagc
attgagaata
      2341 taatatgaaa ttatgaaaaa aaaaaaaaaa aaa
```

Protein sequence:

NCBI Reference Sequence: NP_006127.1

LOCUS NP_006127

ACCESSION NP_006127

```
      1 madleeqlsd eekvriaakf iihappgefn evfndvrlll nndnllrega
ahafaqynld
       61 qftpvkiegy edqvlitehg dlgngkfldp knricfkfdh lrkeatdprp
cevenavesw
      121 rtsvetalra yvkehypngv ctvygkkidg qqtiiacies hqfqaknfwn
grwrsewkft
      181 itpsttqvvg ilkiqvhyye dgnvqlvshk diqdsltvsn evqtakefik
iveaaeneyq
      241 taisenyqtm sdttfkalrr qlpvtrtkid wnkilsykig kemqna
```

CCT7

Official Symbol: CCT7

Official Name: chaperonin containing TCP1, subunit 7 (eta)

Gene ID: 10574

Organism: Homo sapiens

Other Aliases: CCTETA, CCTH, NIP7-1, TCP1ETA

Other Designations: CCT-eta; HIV-1 Nef interacting protein; HIV-1 Nef-interacting protein; T-complex protein 1 subunit eta; TCP-1-eta; chaperonin containing t-complex polypeptide 1, eta subunit

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_006429.3

LOCUS NM_006429

ACCESSION NM_006429

```
      1 atagagtagc ggaagtggtc cgttctcttc ctctcccggc ccaagcttct
gggtatttct
       61 attgcgcgag gcattgtggg ttgctgggcg gcccggtctc ggagaagagg
ggagagtggc
      121 gggccgctga ataagcttcc aaaatgatgc ccacaccagt tatcctattg
aaagagggga
      181 ctgatagctc ccaaggcatc ccccagcttg tgagtaacat cagtgcctgc
caggtgattg
      241 ctgaggctgt aagaactacc ctgggtcccc gtggcatgga caagcttatt
gtagatggca
      301 gaggcaaagc aacaatttct aatgatgggg ccacaattct gaaacttctt
gatgttgtcc
      361 atcctgcagc aaagactttg gtagacattg ccaaatccca agatgctgag
gtgggtgatg
      421 gcaccacctc agtgaccttg ctggctgcag agtttctgaa gcaggtgaaa
ccctatgtgg
      481 aggaaggttt acacccccag atcatcattc gagctttccg cacagccacc
cagctggcag
      541 ttaacaagat caaagagatt gctgtgaccg tgaagaaggc agataaagtg
gagcagagga
      601 agctgctgga aaagtgtgcc atgaccgctc tgagctccaa gctgatctcc
cagcagaaag
      661 ctttctttgc taagatggtg gtggatgcag tgatgatgct cgatgatttg
ctgcagctta
      721 aaatgattgg aatcaagaag gtacagggtg gagccctcga ggattctcag
ctggtagctg
      781 gtgttgcatt caagaagact ttctcttacg ctgggtttga aatgcaaccc
aaaaagtacc
      841 acaatcccaa gattgccctt ttgaatgtcg agctcgagtt gaaagctgag
aaagacaatg
      901 ctgagataag agtccacaca gttgaggatt atcaggcaat tgttgatgct
gagtggaaca
      961 ttctctatga caagttagag aagatccatc attctggagc caaagttgtc
ttgtccaaac
     1021 tccccattgg ggatgtggcc acccagtact ttgctgacag ggacatgttc
tgtgctggcc
```

```
     1081 gagtacctga ggaggatctg aagaggacaa tgatggcctg tggaggctca
atccagacca
     1141 gtgtgaatgc tctgtcagca gatgtgctgg gtcgatgcca ggtgtttgaa
gagacccaga
     1201 ttggaggcga gaggtacaat ttttttactg gctgccccaa ggccaagaca
tgcaccttca
     1261 ttctccgtgg cggcgccgag cagtttatgg aggagacaga gcggtccctg
catgatgcca
     1321 tcatgatcgt caggagggcc atcaagaatg attcagtggt ggctggtggc
ggggccattg
     1381 agatggaact ctccaagtac ctgcgggatt actcaaggac tattccagga
aaacagcagc
     1441 tgttgattgg ggcatatgcc aaggccttgg agattatccc acgccagctg
tgtgacaatg
     1501 ctggctttga tgccacaaac attctcaaca agctgcgggc tcggcatgcc
caggggggta
     1561 catggtatgg agtagacatc aacaacgagg acattgctga caactttgaa
gctttcgtgt
     1621 gggagccagc tatggtgcgg atcaatgcgc tgacagcagc ctctgaggct
gcgtgcctga
     1681 tcgtgtctgt agatgaaacc atcaagaacc cccgctcgac tgtggatgct
cccacagcag
     1741 caggccgggg ccgtggtcgt ggccgccccc actgagaggc accccaccca
tcacatggct
     1801 ggctggctgc tgggtgcact taccctcctt ggcttggtta cttcatttta
caaggaaggg
     1861 gtagtaattg gcccactctc ttcttactgg aggctattta aataaaatgt
aagacttcag
     1921 ataactttgt aaattaaaaa aaaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_006420.1

LOCUS NP_006420

ACCESSION NP_006420

```
      1 mmptpvillk egtdssqgip qlvsnisacq viaeavrttl gprgmdkliv
dgrgkatisn
       61 dgatilklld vvhpaaktlv diaksqdaev gdgttsvtll aaeflkqvkp
yveeglhpqi
      121 iirafrtatq lavnkikeia vtvkkadkve qrkllekcam talssklisq
qkaffakmvv
      181 davmmlddll qlkmigikkv qqgaledsql vagvafkktf syagfemqpk
kyhnpkiall
      241 nvelelkaek dnaeirvhtv edyqaivdae wnilydklek ihhsgakvvl
sklpigdvat
      301 qyfadrdmfc agrvpeedlk rtmmacggsi qtsvnalsad vlgrcqvfee
tqiggerynf
      361 ftgcpkaktc tfilrggaeq fmeeterslh daimivrrai kndsvvaggg
aiemelskyl
      421 rdysrtipgk qqlligayak aleiiprqlc dnagfdatni lnklrarhaq
ggtwygvdin
      481 nediadnfea fvwepamvri naltaaseaa clivsvdeti knprstvdap
taagrgrgrg
      541 rph
```

CTSB

Official Symbol: CTSB

Official Name: cathepsin B

Gene ID: 1508

Organism: Homo sapiens

Other Aliases: APPS, CPSB

Other Designations: APP secretase; amyloid precursor protein secretase; cathepsin B1; cysteine protease

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001908.3

LOCUS NM_001908

ACCESSION NM_001908

```
      1 ggggcggggc cgggagggta cttagggccg gggctggccc aggctacggc
ggctgcaggg
       61 ctccggcaac cgctccggca acgccaaccg ctccgctgcg cgcaggctgg
gctgcaggct
      121 ctcggctgca gcgctgggtg gatctaggat ccggcttcca acatgtggca
gctctgggcc
      181 tccctctgct gcctgctggt gttggccaat gcccggagca ggccctcttt
ccatcccctg
      241 tcggatgagc tggtcaacta tgtcaacaaa cggaatacca cgtggcaggc
cgggcacaac
      301 ttctacaacg tggacatgag ctacttgaag aggctatgtg gtaccttcct
gggtgggccc
      361 aagccacccc agagagttat gtttaccgag gacctgaagc tgcctgcaag
cttcgatgca
      421 cgggaacaat ggccacagtg tcccaccatc aaagagatca gagaccaggg
ctcctgtggc
      481 tcctgctggg ccttcggggc tgtggaagcc atctctgacc ggatctgcat
ccacaccaat
      541 gcgcacgtca gcgtggaggt gtcggcggag gacctgctca catgctgtgg
cagcatgtgt
      601 ggggacggct gtaatggtgg ctatcctgct gaagcttgga acttctggac
aagaaaaggc
      661 ctggtttctg gtggcctcta tgaatcccat gtagggtgca gaccgtactc
catccctccc
      721 tgtgagcacc acgtcaacgg ctcccggccc ccatgcacgg gggagggaga
taccccaag
      781 tgtagcaaga tctgtgagcc tggctacagc ccgacctaca aacaggacaa
gcactacgga
      841 tacaattcct acagcgtctc caatagcgag aaggacatca tggccgagat
ctacaaaaac
      901 ggccccgtgg agggagcttt ctctgtgtat tcggacttcc tgctctacaa
gtcaggagtg
      961 taccaacacg tcaccggaga gatgatgggt ggccatgcca tccgcatcct
gggctgggga
     1021 gtggagaatg gcacacccta ctggctggtt gccaactcct ggaacactga
ctggggtgac
```

1081 aatggcttct ttaaaatact cagaggacag gatcactgtg gaatcgaatc
agaagtggtg
1141 gctggaattc cacgcaccga tcagtactgg gaaaagatct aatctgccgt
gggcctgtcg
1201 tgccagtcct gggggcgaga tcggggtaga aatgcatttt attctttaag
ttcacgtaag
1261 atacaagttt cagacagggt ctgaaggact ggattggcca aacatcagac
ctgtcttcca
1321 aggagaccaa gtcctggcta catcccagcc tgtggttaca gtgcagacag
gccatgtgag
1381 ccaccgctgc cagcacagag cgtccttccc cctgtagact agtgccgtag
ggagtacctg
1441 ctgccccagc tgactgtggc cccctccgtg atccatccat ctccagggag
caagacagag
1501 acgcaggaat ggaaagcgga gttcctaaca ggatgaaagt tcccccatca
gttcccccag
1561 tacctccaag caagtagctt tccacatttg tcacagaaat cagaggagag
acggtgttgg
1621 gagccctttg gagaacgcca gtctcccagg ccccctgcat ctatcgagtt
tgcaatgtca
1681 caacctctct gatcttgtgc tcagcatgat tctttaatag aagttttatt
ttttcgtgca
1741 ctctgctaat catgtgggtg agccagtgga acagcgggag acctgtgcta
gttttacaga
1801 ttgcctcctt atgacgcggc tcaaaaggaa accaagtggt caggagttgt
ttctgaccca
1861 ctgatctcta ctaccacaag gaaaatagtt taggagaaac cagcttttac
tgtttttgaa
1921 aaattacagc ttcaccctgt caagttaaca aggaatgcct gtgccaataa
aagttttctc
1981 caacttgaag tctactctga tgggatctca gatcctttgt cactgcctat
agacttgtag
2041 ctgctgtctc tctttgtccc tgcagagaat cacgtcctgg aactgcatgt
tcttgcgact
2101 cttgggactt catcttaact tctcgctgcc ccagccatgt tttcaaccat
ggcatccctc
2161 ccccaattag ttccctgtca tcctcgtcaa ccttctctgt aagtgcctgg
taagcttgcc
2221 cttgcttaag aactcaaaac atagctgtgc tctatttttt tgttgttgtt
gtgactgaca
2281 gagtgagatt ccgtctccca ggctggagtg cagtggcgcc ttctcagctc
actgcaacct
2341 gcagcctcct agattcaagc gattctcctg cttcagcctt ccgagtagct
gggatgacag
2401 gcactcacca atatgcctgg gtaatttttg tatttttaag tacatacagg
atttcaccat
2461 gttggccagg ctagtttcaa actcccggcc tcaggtggtc tgcctgcctc
agcctcccaa
2521 agtgttggga ttacaggcgt gagccactgg gccctgcctg tatttttat
cagccacaaa
2581 tccagcaaca agctgaggat tcagctcata aaacaggctt ggtgtcttgg
tgatctcaca
2641 taaccaagat gctaccccgt ggggaaccac atccccctgg atgccctcca
gccttggttt
2701 gggctggagt cagggcctgt atacagtatt ttgaatttgt atgccactgg
tttgcattgc
2761 tggtcaggaa ctctagtgct ttgcatagcc ctggtttaga aacatgttat
agcagttctt
2821 ggtatagagc aaactagaag aaccagcaat cattccactg tcctgccaag
gtacacctca

```
     2881 gtactcccct tcccaactga agtggtatga ggctagctct ttccaaaagc
attcaagttt
     2941 ggcttctgat gtgactcaga atttaggaac cagatgctag atcaaataag
ctctgaaaat
     3001 ctgaggaaca ttgtaggaaa ggtttgttaa gcatctctta agtgccatga
tgagcataac
     3061 agccggccgt cgtggctcac gcctgtaatc ccagcacttt gggaggccaa
ggtgggagga
     3121 tgacaaggtc aggagttcaa gaccagcctg gccaacatgc tgaaacctca
cctctactaa
     3181 aaatacaaaa attagctggg catggtggca catgcctgta atcccagcta
cttgggaggc
     3241 tgaggcagga gaatcgcttg aacccgggag gcggaggttg cagtgagcca
agacagtgcc
     3301 agtgcactcc agcctcggtg acagcgcaag gctccgtctc aataattaaa
aaaaaaaaaa
     3361 aaaaaaaaaa ggccgggcgc agtggctcaa gcctgtaatc ccagcacttt
gggaggctga
     3421 ggcgggcaga tcacctgagg tcaggagttt tgagatcagc cttggcaaca
cggtgaaacc
     3481 ccatctctac taaaaataca aaattagcca agcatgctgg cacatgcctg
taatcccagc
     3541 tactcgggag gctgaggtac gagaatcgct tgaacctggg aggcagagga
tgcagtgagc
     3601 cgagatcacg ccattgcact ccagcctggg ggacaagagt gaatctgtgt
ctcaccaaaa
     3661 aaaaaaagaa aaagaaagat gcttaacaaa ggttaccata agccacaaat
tcataaccac
     3721 ttatccttcc agtttcaagt agaatatatt cataacctca ataaagttct
ccctgctccc
      3781 aaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_001899.1

LOCUS NP_001899

ACCESSION NP_001899

```
      1 mwqlwaslcc llvlanarsr psfhplsdel vnyvnkrntt wqaghnfynv
dmsylkrlcg
       61 tflggpkppq rvmftedlkl pasfdareqw pqcptikeir dqgscgscwa
fgaveaisdr
      121 icihtnahvs vevsaedllt ccgsmcgdgc nggypaeawn fwtrkglvsg
glyeshvgcr
      181 pysippcehh vngsrppctg egdtpkcski cepgysptyk qdkhygynsy
svsnsekdim
      241 aeiykngpve gafsvysdfl lyksgvyqhv tgemmgghai rilgwgveng
tpywlvansw
      301 ntdwgdngff kilrgqdhcg iesevvagip rtdqyweki
```

FKBP2

Official Symbol: FKBP2

Official Name: FK506 binding protein 2, 13kDa

Gene ID: 2286

Organism: Homo sapiens

Other Aliases: FKBP-13, PPIase

Other Designations: 13 kDa FK506-binding protein; 13 kDa FKBP; FK506-binding protein 2 (13kD); FKBP-2; PPIase **FKBP2**; immunophilin FKBP13; peptidyl-prolyl cis-trans isomerase **FKBP2**; proline isomerase; rapamycin-binding protein; rotamase

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_004470.3

LOCUS NM_004470

ACCESSION NM_004470

```
      1 gccggaagtg acgcagggca gcggcgtcgc gggggcgggg ctcgggaaag
acccgtgcca
       61 gcgggcgtgt ggccgcgggt ttcgcacggt ccaataaggg agggcggcgt
ggcccggcct
      121 ggtagcgacg aggacgcgcc tgcgcagagg cggcagcacc accggggttg
actccggggg
      181 cgcggcgagg agagacatga ggctgagctg gttccgggtc ctgacagtac
tgtccatctg
      241 cctgagcgcc gtggccacgg ccacgggggc cgagggcaaa aggaagctgc
agatcggggt
      301 caagaagcgg gtggaccact gtcccatcaa atcgcgcaaa ggggatgtcc
tgcacatgca
      361 ctacacgggg aagctggaag atgggacaga gtttgacagc agcctgcccc
agaaccagcc
      421 ctttgtcttc tcccttggca caggccaggt catcaagggc tgggaccagg
ggctgctggg
      481 gatgtgtgag ggggaaaagc gcaagctggt gatcccatcc gagctagggt
atggagagcg
      541 gggagctccc ccaaagattc caggcggtgc aaccctggtg ttcgaggtgg
agctgctcaa
      601 aatagagcga cgaactgagc tgtaaccaga ctggggaggg gcaggggggag
aggcccccat
      661 cagggaccag actgttccaa aaaaaaaaca aaaaacaaaa acaaacaaaa
aaacacttaa
      721 aagcccaagg aaaaaaaaaa aaaaaaa
```

Protein sequence (variant 1):

NCBI Reference Sequence: NP_004461.2

LOCUS NP_004461

ACCESSION NP_004461

```
      1 mrlswfrvlt vlsiclsava tatgaegkrk lqigvkkrvd hcpiksrkgd
vlhmhytgkl
```

```
       61 edgtefdssl pqnqpfvfsl gtgqvikgwd qgllgmcege krklvipsel
gygergappk
      121 ipggatlvfe vellkierrt el
```

FLNC

Official Symbol: FLNC

Official Name: filamin C, gamma

Gene ID: 2318

Organism: Homo sapiens

Other Aliases: ABP-280, ABP280A, ABPA, ABPL, FLN2, MFM5, MPD4

Other Designations: ABP-280-like protein; ABP-L, gamma filamin; FLN-C; actin binding protein 280; actin-binding-like protein; filamin 2; filamin-2; filamin-C

Nucleotide sequence (variant 1):

NCBI Reference Sequence: NM_001458.4

LOCUS NM_001458

ACCESSION NM_001458

```
      1 ccctggaggg agagagagcc agagagcggc cgagcgccta ggaggcccgc
cgagcctcgc
       61 cgagccccgc cagccccggc gcgagagaag ttggagagga gagcagcgca
gcgcagcgag
      121 tcccgtggtc gcgccccaac agcgcccgac agcccccgat agcccaaacc
gcggccctag
      181 ccccggccgc acccccagcc cgcgccagca tgatgaacaa cagcggctac
tcagacgccg
      241 gcctcggcct gggcgatgag acagacgaga tgccgtccac ggagaaggac
ctggcggagg
      301 acgcgccgtg gaagaagatc cagcagaaca cattcacgcg ctggtgcaat
gagcacctca
      361 agtgcgtggg caagcgcctg accgacctgc agcgcgacct cagcgacggg
ctccggctca
      421 tcgcgctgct cgaggtgctc agccagaagc gcatgtaccg caagttccat
ccgcgcccca
      481 acttccgcca aatgaagctg gagaacgtgt ccgtggccct cgagttcctc
gagcgcgagc
      541 acatcaagct cgtgtccata gacagcaagg ccatcgtgga tgggaacctg
aagctgatcc
      601 tgggcctgat ctggacgctg atcctgcact actccatctc catgcccatg
tgggaggatg
      661 aagatgatga ggatgcccgc aaacagacgc ccaagcagcg gctgcttggc
tggatccaga
      721 acaaggtgcc ccagctgccc atcaccaact tcaaccgtga ctggcaggac
ggcaaagctc
      781 tgggcgccct ggtggacaac tgcgcccccg gtctctgccc cgactgggag
gcctgggacc
```

841 ccaaccagcc cgtggagaac gcccgggagg ccatgcagca ggccgacgac tggcttgggg
901 tgccccaggt cattgcccct gaggagattg tggaccccaa cgtggatgag cattctgtta
961 tgacctacct gtcccagttc cccaaggcca agctcaaacc tggtgcccct gttcgatcca
1021 agcagctgaa ccccaagaaa gccatcgcct atgggcctgg catcgagcca cagggcaaca
1081 ccgtgctgca gcctgcccac ttcaccgtgc agacggtgga cgcgggcgtg ggcgaggtgc
1141 tggtctacat cgaggaccct gaaggccaca ccgaggaggc taaggtggtt cccaacaatg
1201 acaaggatcg cacctatgct gtctcctatg tgcccaaggt cgctgggtta cacaaggtga
1261 ccgtgctctt tgctggccag aacattgaac gcagtccctt tgaggtgaac gtgggcatgg
1321 ccctgggaga tgccaacaag gtgtcagccc gtggccctgg cctggaacct gtgggcaatg
1381 tggccaacaa acccacctac tttgacatct acactgcggg ggccggcact ggcgatgttg
1441 ctgtggtgat cgtggaccca cagggccggc gggacacagt ggaggtggcc ctggaggaca
1501 agggtgacag cacgttccgc tgcacataca gacctgccat ggaggggcca cataccgtgc
1561 atgtggcctt tgcgggtgcc cccatcaccc gcagtccctt ccctgtccat gtgtcggaag
1621 cctgtaaccc caacgcctgc cgcgcctctg ggcgaggcct gcagcccaag ggtgttcgcg
1681 tgaaagaggt ggctgacttc aaggtgttta ccaagggtgc cggcagcggg gagctcaagg
1741 tcacggtcaa ggggccaaag ggcacagagg agccagtgaa ggtgcgggag gctggggatg
1801 gtgtgttcga gtgcgagtac tacccggtgg tgcctgggaa gtatgtggtg accatcacgt
1861 ggggcggcta cgccatccct cgcagcccct ttgaggtaca ggtgagccca gaggcaggag
1921 tgcaaaaggt ccgggcctgg ggtcctggtt tggagactgg ccaggtgggc aagtcagccg
1981 attttgtggt ggaagccatt ggcaccgagg tggggacact gggcttctcc atcgaggggc
2041 cctcacaagc caagatcgaa tgtgacgaca agggggatgg ctcctgcgat gtgcggtact
2101 ggcccacgga gcctggggag tacgctgtgc acgtcatctg tgacgatgag gacatccgag
2161 actcaccctt cattgcccac atcctgcccg ccccacctga ctgcttccca gataaggtga
2221 aggcctttgg gcctggcctg gagcctaccg gctgcatcgt ggacaagccc gctgagttca
2281 ccattgatgc tcgtgcagct ggcaagggag acctgaagct ctatgcccag gacgccgacg
2341 gctgtcccat cgacatcaag gtgatcccca acggcgacgg caccttccgc tgctcctacg
2401 tgcccaccaa gcccattaag cacaccatca tcatctcctg gggaggcgta aacgtgccca
2461 agagcccctt ccgggtgaac gtgggcgagg gcagccaccc cgagcgggta aaggtgtacg
2521 gccccggagt ggagaagaca ggcctcaagg ccaatgagcc cacctacttc acggtggact
2581 gcagcgaggc ggggcaaggc gacgtgagca tcggcatcaa gtgcgcccca ggcgtggtgg

2641 gccctgcaga ggctgacatt gacttcgaca tcatcaagaa tgacaacgac accttcaccg
2701 tcaagtacac gccaccaggg gcgggccgct acaccatcat ggtgctgttt gccaaccagg
2761 agatccccgc cagccccttc cacatcaagg tggacccatc ccacgatgcc agcaaagtca
2821 aggccgaggg ccctgggctg aatcgcacag gtgtggaagt cgggaagccc acccacttca
2881 cggtgctgac caagggagcc ggcaaggcca agctggatgt gcagtttgca gggacagcca
2941 agggcgaggt tgtgcgggac tttgagatca tagacaacca tgactactcc tacactgtca
3001 agtacaccgc tgtccagcag ggcaacatgg cagtgacagt gacttatggc ggggaccctg
3061 tccccaagag ccccttgtg gtgaatgtgg cacccccgct ggacctcagc aaaatcaaag
3121 ttcagggcct taatagcaag gtggctgtgg gacaggaaca agcattctct gtgaacacac
3181 gaggggctgg cggtcagggc caactggatg tgcggatgac ttcgccctct cgccggccca
3241 tcccctgcaa gctggagcca ggcggtggag cggaagccca ggctgtgcgc tacatgcccc
3301 cggaggaggg gccctacaag gtggatatca cctacgatgg tcacccggtg cctggcagcc
3361 cgtttgctgt ggagggtgtc ctgccccctg atccctccaa ggtctgtgct tatggcccgg
3421 gtctcaaggg tggactggta ggcacccccg cgccattctc catcgacacc aaggggggctg
3481 gcacaggtgg cctggggctg accgtagagg gcccctgcga ggccaagatc gagtgccagg
3541 acaatggtga tggctcatgt gctgtcagct acctgcccac ggagcctggc gagtacacca
3601 tcaacatcct gtttgctgag gcccacatcc ctggctcgcc cttcaaagcc accattcggc
3661 ctgtgtttga cccgagcaag gtgcgggcca gtggaccggg cctggagcgc ggcaaggtcg
3721 gtgaggcagc caccttcact gtggactgct cagaggcagg cgaggcggag ctgaccattg
3781 agatcctgtc ggatgccggg gtcaaggccg aggtgctgat ccacaacaac gcggatggca
3841 cctaccacat cacctacagc cctgccttcc ctggcaccta caccattacc atcaagtatg
3901 gcgggcatcc cgtgcccaaa ttccccaccc gtgtccatgt gcagcctgcg gtcgatacca
3961 gtggcgtcaa ggtctcaggg cctggtgttg agccacacgg tgtcctgcgg gaggtgacca
4021 ctgagttcac tgtggatgca agatccctaa cagccacagg cggcaaccac gtgacggctc
4081 gtgtgctcaa cccctcgggg gccaagacag acacctatgt gacagacaat ggggacggca
4141 cctaccgagt gcagtacacc gcctacgagg agggcgtgca tctggtggag gtcctgtatg
4201 atgaggtcgc tgtgcccaag agccccttcc gagtgggcgt gaccgagggc tgtgatccca
4261 cccgcgtccg agccttcggg ccaggcctgg agggtggctt ggtcaacaag gccaaccgat
4321 tcactgtgga gaccagggga gcgggcaccg ggggccttgg cctagccatc gagggtccct
4381 cggaagccaa gatgtcctgc aaggacaaca aggatggtag ctgcaccgtg gagtacatcc

4441 ccttcactcc tggagactat gacgtcaaca tcaccttcgg ggggcggccc atcccaggga
4501 gcccgttccg cgtgccagtg aaggatgtgg tggaccctgg gaaggtgaag tgctcagggc
4561 cagggctggg ggctggtgtc agggcccggg ttcctcagac cttcacagtg gactgcagtc
4621 aagctggccg ggcgcccctg caggtggctg tgctgggccc cacaggtgtg gccgagcctg
4681 tggaggtgcg ggacaatgga gatggcaccc acactgtcca ctacacccca gccactgacg
4741 ggccctacac ggtagccgtc aagtatgctg accaggaggt gccacgcagc cccttcaaga
4801 tcaaggtcct cccagctcat gatgccagca aggtgcgggc cagcggccca ggcctcaacg
4861 cctctggcat ccctgccagc ctgcctgtgg agttcaccat cgacgcacgg gacgcgggcg
4921 aggggttgct cactgtccag atcttggacc ccgagggtaa gcccaagaag gccaacatcc
4981 gggacaatgg ggatggcacg tacactgtgt cctacctgcc ggacatgagt ggccggtaca
5041 ccatcaccat caagtatggc ggtgatgaga tcccctactc gcccttccgc atccatgctc
5101 tgcccactgg ggatgccagc aagtgcctcg tcacagtgtc cattggaggc catggcctgg
5161 gtgcctgcct gggccctcga atccagattg ggcaggagac ggtgatcacg gtggatgcca
5221 aggcagccgg tgaggggaag gtgacatgca cggtgtccac gccggatggg gcagagctcg
5281 atgtggatgt ggttgagaac catgacggta cctttgacat ctactacaca gcgcccgagc
5341 cgggcaagta cgtcatcacc atccgcttcg ggggtgagca catccccaac agccccttcc
5401 acgtgctggc gtgtgacccc ctgccgcacg aggaggagcc ctctgaagtg ccacagctgc
5461 gccagcccta cgctcctccc cggcccggcg cccgccccac acactgggcc acagaggagc
5521 cagtggtgcc tgtggagcca atggagtcca tgctgaggcc cttcaacctg gtcatccccc
5581 tcgcggtgca gaaaggggag ctcacaggag aggtgcggat gccctcgggg aagacggcac
5641 ggcccaacat caccgacaac aaggacggca ccatcacggt gaggtatgca cccactgaga
5701 aaggcctgca ccagatgggg atcaagtatg acggcaacca catccctggg agccccttac
5761 agttctatgt ggatgccatc aacagccgcc atgtcagtgc ctatgggcca ggcctgagcc
5821 atggcatggt caacaagcca gccaccttca ctattgtcac caaagatgct ggagaagggg
5881 gtctgtcact ggccgtggag ggcccatcca aggcagagat cacctgtaag gacaacaagg
5941 atggcacctg caccgtgtcc tatctgccga ctgcgcctgg agactacagc atcatcgtgc
6001 gcttcgatga caagcacatc ccggggagcc ccttcacagc caagatcaca ggtgatgact
6061 ccatgaggac ctcacagctg aatgtgggca cctccacgga cgtgtcactg aagatcaccg
6121 agagtgatct gagccagctg accgccagca tccgtgcccc ctcgggcaac gaggagccct
6181 gcctgctgaa gcgcctgccc aaccggcaca ttgggatctc cttcaccccc aaggaggtcg

6241 gggagcacgt ggtgagcgtg cgcaagagtg gcaagcatgt caccaacagc
cccttcaaga
6301 tcctggtggg gccatctgag atcggggacg ccagcaaggt gcgggtctgg
ggcaaggggc
6361 tttccgaggg acacacattc caggtggcag agttcatcgt ggacactcgc
aatgcaggtt
6421 atgggggctt ggggctgagt attgaaggcc caagcaaggt ggacatcaac
tgtgaggaca
6481 tggaggacgg gacatgcaaa gtcacctact gccccaccga gcccggcacc
tacatcatca
6541 acatcaagtt tgctgacaag cacgtgcctg gaagcccctt cactgtgaag
gtgaccggcg
6601 agggccgcat gaaggagagc atcacccggc ggagacaggc accttccatc
gccaccatcg
6661 gcagcacctg tgacctcaac ctcaagatcc caggaaactg gttccagatg
gtgtctgccc
6721 aggagcgcct gacacgcacc ttcacacgca gcagccacac ctacacccgc
acggagcgca
6781 cggagatcag caagacgcgg ggcggggaga caaagcgcga ggtgcgggtg
gaggagtcca
6841 cccaggtcgg cggggacccc ttccctgctg tgtttgggga cttcctgggc
cgggagcgcc
6901 tgggatcctt cggcagcatc acccggcagc aggagggtga ggccagctct
caggacatga
6961 ctgcacaggt gaccagccca tcgggcaagg tggaagccgc agagatcgtc
gagggcgagg
7021 acagcgccta cagcgtgcgc tttgtgcccc aggaaatggg gccccatacg
gtcgctgtca
7081 agtaccgtgg ccagcacgtg cccggcagcc cctttcagtt cactgtgggg
ccgctgggtg
7141 aaggtggtgc ccacaaggtg cgggccggag gcacagggct ggagcgaggt
gtggccggcg
7201 tgccagccga gttcagcatc tggacccggg aggctggcgc tgggggcctg
tccattgctg
7261 tggagggtcc tagcaaagcg gagattgcat ttgaggatcg caaagatggc
tcctgcggcg
7321 tctcctatgt cgtccaggaa ccaggtgact atgaggtctc catcaagttc
aatgatgagc
7381 acatcccaga cagccccttt gtggtgcctg tggcctccct ctcggatgac
gctcgccgtc
7441 tcactgtcac cagcctccag gagacggggc tcaaggtgaa ccagccagcg
tcctttgccg
7501 tgcagctgaa cggtgcccgg ggcgtgattg atgcccgggt gcacacaccc
tcgggggctg
7561 tggaggagtg ctacgtctct gagctggaca gtgacaagca caccatccgc
ttcatccccc
7621 acgagaatgg cgtccactcc atcgatgtca agttcaacgg tgcccacatc
cctggaagtc
7681 ccttcaagat ccgcgttggg gagcagagcc aggctgggga cccaggcttg
gtgtcagcct
7741 acggtcctgg gctcgaggga ggcactaccg gtgtgtcatc agagttcatc
gtgaacaccc
7801 tgaatgccgg ctcgggggcc ttgtctgtca ccattgatgg cccctccaag
gtgcagctgg
7861 actgtcggga gtgtcctgag ggccatgtgg tcacttatac tcccatggcc
cctggcaact
7921 acctcattgc catcaagtac ggtggccccc agcacatcgt gggcagcccc
ttcaaggcca
7981 aggtcactgg tccgaggctg tccggaggcc acagccttca cgaaacatcc
acggttctgg

```
     8041 tggagactgt gaccaagtcc tcctcaagcc ggggctccag ctacagctcc
atccccaagt
     8101 tctcctcaga tgccagcaag gtggtgactc ggggccctgg gctgtcccag
gccttcgtgg
     8161 gccagaagaa ctccttcacc gtggactgca gcaaagcagg caccaacatg
atgatggtgg
     8221 gcgtgcacgg ccccaagacc ccctgtgagg aggtgtacgt gaagcacatg
gggaaccggg
     8281 tgtacaatgt cacctacact gtcaaggaga aaggggacta catcctcatt
gtcaagtggg
     8341 gtgacgaaag tgtccctgga agccccttca aagtcaaggt cccttgaatc
ccaaaagtgc
     8401 ctccccagcc tcagccccca cctccagcca cacacacatt acacacacac
acacacacac
     8461 acaaatgtgc cacacccaga cacgcacaga atcagacact acaaacacct
gccttggggg
     8521 tgaagtgaag gcccagcctc cccaccccac cgcgccccag gggttggagg
accttgtctg
     8581 tgtcaggaca gtgtccctcc ctgggaatgt gacatgaggg ccgactgggg
ccaggctcag
     8641 gggcagaggc tgggacacaa ggggctggcg agggctgcga ggccagggaa
gccctgagtt
     8701 tctggcgggg ctgagcagtg ggggagcatt gtgttgtggg tgtctgtgtg
tgaggtcacc
     8761 ctcaaactgc accgccggcc agataccctc ctgaccccga ggacttggtc
tggtctctct
     8821 ggtggctaca accccagagt tttaaggact tggaaaggaa agcacaatca
gagaagaaaa
     8881 cagcccccga accagcagga gtggcctggc acatggaccg gcctgagcga
tgtgcactcc
     8941 acccaagcca ggctcccagg gggcctgatt tctctctcac tgtctctttt
tttaaaatgg
     9001 ttgcacggct ctgccccatg gggggccttt tttacacact gcgaggccca
gctttctagg
     9061 ggacttttgc acatgtcatg cagctcagct gggagctgct taggtggaaa
actccaaata
     9121 aagtgcggct gtcgcagaaa aaaaaaaaa
```

Protein sequence (variant):

NCBI Reference Sequence: NP_001449.3

LOCUS NP_001449

ACCESSION NP_001449

```
     1 mmnnsgysda glglgdetde mpstekdlae dapwkkiqqn tftrwcnehl
kcvgkrltdl
    61 qrdlsdglrl iallevlsqk rmyrkfhprp nfrqmklenv svaleflere
hiklvsidsk
   121 aivdgnlkli lgliwtlilh ysismpmwed eddedarkqt pkqrllgwiq
nkvpqlpitn
   181 fnrdwqdgka lgalvdncap glcpdweawd pnqpvenare amqqaddwlg
vpqviapeei
   241 vdpnvdehsv mtylsqfpka klkpgapvrs kqlnpkkaia ygpgiepqgn
tvlqpahftv
   301 qtvdagvgev lvyiedpegh teeakvvpnn dkdrtyavsy vpkvaglhkv
tvlfagqnie
   361 rspfevnvgm algdankvsa rgpglepvgn vankptyfdi ytagagtgdv
avvivdpqgr
```

421 rdtvevaled kgdstfrcty rpamegphtv hvafagapit rspfpvhvse
acnpnacras
481 grglqpkgvr vkevadfkvf tkgagsgelk vtvkgpkgte epvkvreagd
gvfeceyypv
541 vpgkyvvtit wggyaiprsp fevqvspeag vqkvrawgpg letgqvgksa
dfvveaigte
601 vgtlgfsieg psqakiecdd kgdgscdvry wptepgeyav hvicddedir
dspfiahilp
661 appdcfpdkv kafgpglept gcivdkpaef tidaraagkg dlklyaqdad
gcpidikvip
721 ngdgtfrcsy vptkpikhti iiswggvnvp kspfrvnvge gshpervkvy
gpgvektglk
781 aneptyftvd cseagqgdvs igikcapgvv gpaeadidfd iikndndtft
vkytppgagr
841 ytimvlfanq eipaspfhik vdpshdaskv kaegpglnrt gvevgkpthf
tvltkgagka
901 kldvqfagta kgevvrdfei idnhdysytv kytavqqgnm avtvtyggdp
vpkspfvvnv
961 appldlskik vqglnskvav gqeqafsvnt rgaggqgqld vrmtspsrrp
ipcklepggg
1021 aeaqavrymp peegpykvdi tydghpvpgs pfavegvlpp dpskvcaygp
glkgglvgtp
1081 apfsidtkga gtgglgltve gpceakiecq dngdgscavs ylptepgeyt
inilfaeahi
1141 pgspfkatir pvfdpskvra sgpglergkv geaatftvdc seageaelti
eilsdagvka
1201 evlihnnadg tyhityspaf pgtytitiky gghpvpkfpt rvhvqpavdt
sgvkvsgpgv
1261 ephgvlrevt teftvdarsl tatggnhvta rvlnpsgakt dtyvtdngdg
tyrvqytaye
1321 egvhlvevly devavpkspf rvgvtegcdp trvrafgpgl egglvnkanr
ftvetrgagt
1381 gglglaiegp seakmsckdn kdgsctveyi pftpgdydvn itfggrpipg
spfrvpvkdv
1441 vdpgkvkcsg pglgagvrar vpqtftvdcs qagraplqva vlgptgvaep
vevrdngdgt
1501 htvhytpatd gpytvavkya dqevprspfk ikvlpahdas kvrasgpgln
asgipaslpv
1561 eftidardag eglltvqild pegkpkkani rdngdgtytv sylpdmsgry
titikyggde
1621 ipyspfriha lptgdaskcl vtvsigghgl gaclgpriqi gqetvitvda
kaagegkvtc
1681 tvstpdgael dvdvvenhdg tfdiyytape pgkyvitirf ggehipnspf
hvlacdplph
1741 eeepsevpql rqpyapprpg arpthwatee pvvpvepmes mlrpfnlvip
favqkgeltg
1801 evrmpsgkta rpnitdnkdg titvryapte kglhqmgiky dgnhipgspl
qfyvdainsr
1861 hvsaygpgls hgmvnkpatf tivtkdageg glslavegps kaeitckdnk
dgtctvsylp
1921 tapgdysiiv rfddkhipgs pftakitgdd smrtsqlnvg tstdvslkit
esdlsqltas
1981 irapsgneep cllkrlpnrh igisftpkev gehvvsvrks gkhvtnspfk
ilvgpseigd
2041 askvrvwgkg lseghtfqva efivdtrnag ygglglsieg pskvdinced
medgtckvty
2101 cptepgtyii nikfadkhvp gspftvkvtg egrmkesitr rrqapsiati
gstcdlnlki
2161 pgnwfqmvsa qerltrtftr sshtytrter teisktrgge tkrevrvees
tqvggdpfpa

```
     2221 vfgdflgrer lgsfgsitrq qegeassqdm taqvtspsgk veaaeivege
dsaysvrfvp
     2281 qemgphtvav kyrgqhvpgs pfqftvgplg eggahkvrag gtglergvag
vpaefsiwtr
     2341 eagagglsia vegpskaeia fedrkdgscg vsyvvqepgd yevsikfnde
hipdspfvvp
     2401 vaslsddarr ltvtslqetg lkvnqpasfa vqlngargvi darvhtpsga
veecyvseld
     2461 sdkhtirfip hengvhsidv kfngahipgs pfkirvgeqs qagdpglvsa
ygpgleggtt
     2521 gvssefivnt lnagsgalsv tidgpskvql dcrecpeghv vtytpmapgn
yliaikyggp
     2581 qhivgspfka kvtgprlsgg hslhetstvl vetvtkssss rgssyssipk
fssdaskvvt
     2641 rgpglsqafv gqknsftvdc skagtnmmmv gvhgpktpce evyvkhmgnr
vynvtytvke
      2701 kgdyilivkw gdesvpgspf kvkvp
```

HPX

Official Symbol: HPX

Official Name: hemopexin

Gene ID: 3263

Organism: Homo sapiens

Other Aliases: HX

Other Designations: beta-1B-glycoprotein

Nucleotide sequence:

NCBI Reference Sequence: NM_000613.2

LOCUS NM_000613

ACCESSION NM_000613

```
      1 aactctatat agggagttca actggtcacc cagagctgtc ctgtggcctc
tgcagctcag
       61 catggctagg gtactgggag cacccgttgc actggggttg tggagcctat
gctggtctct
      121 ggccattgcc acccctcttc ctccgactag tgcccatggg aatgttgctg
aaggcgagac
      181 caagccagac ccagacgtga ctgaacgctg ctcagatggc tggagctttg
atgctaccac
      241 cctggatgac aatggaacca tgctgttttt taaaggggag tttgtgtgga
agagtcacaa
      301 atgggaccgg gagttaatct cagagagatg gaagaatttc cccagccctg
tggatgctgc
      361 attccgtcaa ggtcacaaca gtgtctttct gatcaagggg gacaaagtct
gggtataccc
      421 tcctgaaaag aaggagaaag gatacccaaa gttgctccaa gatgaatttc
ctggaatccc
```

```
      481 atccccactg gatgcagctg tggaatgtca ccgtggagaa tgtcaagctg
aaggcgtcct
      541 cttcttccaa ggtgaccgcg agtggttctg ggacttggct acgggaacca
tgaaggagcg
      601 ttcctggcca gctgttggga actgctcctc tgccctgaga tggctgggcc
gctactactg
      661 cttccagggt aaccaattcc tgcgcttcga ccctgtcagg ggagaggtgc
ctcccaggta
      721 cccgcgggat gtccgagact acttcatgcc ctgccctggc agaggccatg
gacacaggaa
      781 tgggactggc catgggaaca gtacccacca tggccctgag tatatgcgct
gtagcccaca
      841 tctagtcttg tctgcactga cgtctgacaa ccatggtgcc acctatgcct
tcagtgggac
      901 ccactactgg cgtctggaca ccagccggga tggctggcat agctggccca
ttgctcatca
      961 gtggccccag ggtccttcag cagtggatgc tgccttttcc tgggaagaaa
aactctatct
     1021 ggtccagggc acccaggtat atgtcttcct gacaaaggga ggctataccc
tagtaagcgg
     1081 ttatccgaag cggctggaga aggaagtcgg gacccctcat gggattatcc
tggactctgt
     1141 ggatgcggcc tttatctgcc ctgggtcttc tcggctccat atcatggcag
gacggcggct
     1201 gtggtggctg gacctgaagt caggagccca agccacgtgg acagagcttc
cttggcccca
     1261 tgagaaggta gacggagcct tgtgtatgga aaagtccctt ggccctaact
catgttccgc
     1321 caatggtccc ggcttgtacc tcatccatgg tcccaatttg tactgctaca
gtgatgtgga
     1381 gaaactgaat gcagccaagg cccttccgca accccagaat gtgaccagtc
tcctgggctg
     1441 cactcactga ggggccttct gacatgagtc tggcctggcc ccacctccta
gttcctcata
     1501 ataaagacag attgcttctt cgcttctcac tgaggggcct tctgacatga
gtctggcctg
     1561 gccccacctc cccagtttct cataataaag acagattgct tcttcacttg
aatcaaggga
      1621 cctaaaaaaa aaaaa
```

Protein sequence:

NCBI Reference Sequence: NP_000604.1

LOCUS NP_000604

ACCESSION NP_000604

```
      1 marvlgapva lglwslcwsl aiatplppts ahgnvaeget kpdpdvterc
sdgwsfdatt
       61 lddngtmlff kgefvwkshk wdreliserw knfpspvdaa frqghnsvfl
ikgdkvwvyp
      121 pekkekgypk llqdefpgip spldaavech rgecqaegvl ffqgdrewfw
dlatgtmker
      181 swpavgncss alrwlgryyc fqgnqflrfd pvrgevppry prdvrdyfmp
cpgrghghrn
      241 gtghgnsthh gpeymrcsph lvlsaltsdn hgatyafsgt hywrldtsrd
gwhswpiahq
      301 wpqgpsavda afsweeklyl vqgtqvyvfl tkggytlvsg ypkrlekevg
tphgiildsv
```

361 daaficpgss rlhimagrrl wwldlksgaq atwtelpwph ekvdgalcme kslgpnscsa
421 ngpglylihg pnlycysdve klnaakalpq pqnvtsllgc th

TLN1

Official Symbol: TLN1

Official Name: talin 1

Gene ID: 7094

Organism: Homo sapiens

Other Aliases: RP11-112J3.1, ILWEQ, TLN

Other Designations: talin-1

Nucleotide sequence:

NCBI Reference Sequence: NM_006289.3

LOCUS NM_006289

ACCESSION NM_006289

1 ggaagagttc tagcctgaaa gggaactcgg gcgtcgtcct ggcgtcctct ccggattgcg
61 ccgcaccctc gcttcctgcc agggaggggc tgccgggctt tggcggctcc cgagcatcga
121 gaacggggcc agagcagctt cctgcctgcc ccccgcgacc atagagcgcg ggcccagggc
181 gccgcccgcg ggtgggggac gttcccagga cggaagtggc cgagagagtg tcgaagggag
241 ggcgaggccg gagcccgagg gcgacccgag aagcggcggg gcggcgggcc ggcgggcggg
301 gcgcagagcc aggcagcgca ggtatagcca ggctggagaa aagaagctgc caccatggtt
361 gcactttcac tgaagatcag cattgggaat gtggtgaaga cgatgcagtt tgagccgtct
421 accatggtgt acgacgcctg ccgcatcatt cgtgagcgga tcccagaggc cccagctggt
481 cctcccagcg actttgggct ctttctgtca gatgatgacc ccaaaaaggg tatatggctg
541 gaggctggga aagctttgga ctactacatg ctccgaaatg gggacactat ggagtacagg
601 aagaaacaga gacccctgaa gatccgtatg ctggatggaa ctgtgaagac gatcatggtg
661 gatgactcta agactgtcac tgacatgctc atgaccatct gtgcccgcat tggcatcacc
721 aatcatgatg aatattcatt ggttcgagag ctgatggaag agaaaaagga ggaaataaca
781 gggaccttaa gaaaggacaa gacattgctg cgagatgaaa agaagatgga gaaactaaag
841 cagaaattgc acacagatga tgagttgaac tggctggacc atggtcggac actgagggag

```
      901 cagggtgtag aggagcacga gacgctgctg ctgcggagga agttctttta
ctcagaccag
      961 aatgtggatt cccgggaccc tgtacagctg aacctcctgt atgtgcaggc
acgagatgac
     1021 atcctgaatg gctcccaccc tgtctccttt gacaaggcct gtgagtttgc
tggcttccaa
     1081 tgccagatcc agtttgggcc ccacaatgag cagaagcaca aggctggctt
ccttgacctg
     1141 aaggacttcc tgcccaagga gtatgtgaag cagaagggag agcgtaagat
cttccaggca
     1201 cacaagaatt gtgggcagat gagtgagatt gaggccaagg tccgctacgt
gaagctagcc
     1261 cgttctctca agacttacgg tgtctccttc ttcctggtga aggaaaaaat
gaaagggaag
     1321 aacaagctag tgcccaggct tctgggcatc accaaggagt gtgtgatgcg
agtggatgag
     1381 aagaccaagg aagtgatcca ggagtggaac ctcaccaaca tcaaacgctg
ggctgcgtct
     1441 cccaaaagct tcaccctgga ttttggagat taccaagatg gctattactc
agtacagaca
     1501 actgaagggg agcagattgc acagctcatt gccggctaca tcgatatcat
cctgaagaag
     1561 aaaaaaagca aggatcactt tgggctggaa ggagatgagg agtctactat
gctggaggac
     1621 tcagtgtccc ccaaaaagtc aacagtcctg cagcagcaat acaaccgggt
ggggaaagtg
     1681 gagcatggct ctgtggccct gcctgccatc atgcgctctg gagcctctgg
tcctgagaat
     1741 ttccaggtgg gcagcatgcc ccctgcccag cagcagatta ccagcggcca
gatgcaccga
     1801 ggacacatgc ctcctctgac ttcagcccag caggcactca ctggaaccat
taactccagc
     1861 atgcaggccg tgcaggctgc ccaggccacc ctggatgact ttgacactct
gccgcctctt
     1921 ggccaggatg ctgcctctaa ggcctggcgt aaaaacaaga tggatgaatc
aaagcatgag
     1981 atccactctc aggtagatgc catcacagct ggtactgcgt ctgtggtgaa
cctgacagca
     2041 ggggaccctg ctgagacaga ctataccgca gtgggctgtg cagtcaccac
aatctcctcc
     2101 aacctgacgg agatgtcccg tggggtgaag ctgctggctg ccttgctgga
ggacgaaggc
     2161 ggcagtggtc ggcccctgtt gcaggcagca aagggccttg cgggagcagt
gtcagaactg
     2221 ctgcgcagtg cccaaccagc cagtgctgag ccccgtcaga acctgctgca
agcagctggg
     2281 aacgtgggcc aggccagtgg ggagctgttg caacaaattg gggaaagtga
tactgacccc
     2341 cacttccagg atgcgctaat gcagctcgcc aaagctgtgg caagtgctgc
agctgccctg
     2401 gtcctcaagg ccaagagtgt ggcccagcgg acagaggact cgggacttca
gacccaagtt
     2461 attgctgcag caacacagtg tgccctatcc acttcccaac tagtggcctg
tactaaggtg
     2521 gtggcaccta caatcagctc acctgtctgc caagagcaac tggtggaggc
tggacgactg
     2581 gtagccaaag ccgtggaggg ctgtgtgtct gcctcccagg cagctacaga
ggatgggcaa
     2641 ctgttgcgag gggtaggagc agcagccaca gctgtcaccc aggccctaaa
tgagctgctg
```

2701 cagcatgtga aagcccatgc cacaggggct gggcctgctg gccgttatga
ccaggctact
2761 gacaccatcc taaccgtcac tgagaacatc tttagctcca tgggtgatgc
tggggagatg
2821 gtgcgacagg cccgcatcct ggcccaagcc acatctgacc tggtcaatgc
catcaaggct
2881 gatgctgagg gggaaagtga tctggagaac tcccgcaagc tcttaagtgc
tgccaagatc
2941 ctagctgatg ccacagccaa gatggtagag gctgccaagg gagcagctgc
ccaccctgac
3001 agtgaggagc agcagcagcg gctgcgggag gcagctgagg ggctgcgcat
ggccaccaat
3061 gcagctgcgc agaatgccat caagaaaaag ctggtgcagc gcctggagca
tgcagccaag
3121 caggctgcag cctcagccac acagaccatc gctgcagctc agcacgcagc
ctctaccccc
3181 aaggcctctg ccggccccca gcccctgctg gtgcagagct gcaaggcagt
ggcagagcag
3241 attccactgc tggtgcaggg cgtccgagga agccaagccc agcctgacag
ccccagcgct
3301 cagcttgccc tcattgctgc cagccagagc ttcctgcagc caggtgggaa
gatggtggca
3361 gctgcaaagg cctcagtgcc aacgattcag gaccaggctt cagccatgca
gctgagtcag
3421 tgtgccaaga acctgggcac cgcgctggct gaactccgga cggctgccca
gaaggctcag
3481 gaagcatgtg gacctttgga gatggattct gcactgagtg tggtacagaa
tctagagaaa
3541 gatctacagg aagtgaaggc agcagctcga gatggcaagc ttaaaccctt
acctggggag
3601 acaatggaga agtgtaccca ggacctgggc aacagcacca aagccgtgag
ctcagccatc
3661 gcccagctac tgggagaggt tgcccagggc aatgagaatt atgcaggtat
tgcagctcgg
3721 gatgtggcag gtgggctgcg gtcactggcc caggccgcta ggggagtcgc
tgcactgacg
3781 tcagatcctg cagtgcaggc cattgtactt gatacggcca gtgatgtgct
ggacaaggcc
3841 agcagcctca ttgaggaggc gaaaaaggca gctggccatc caggggaccc
tgagagccag
3901 cagcggcttg cccaggtggc taaagcagtg acccaggctc tgaaccgctg
tgtcagctgc
3961 ctacctggcc agcgcgatgt ggataatgcc ctgagggcag ttggagatgc
cagcaagcga
4021 ctcctgagtg actcgcttcc tcctagcact gggacatttc aagaagctca
gagccggttg
4081 aatgaagctg ctgctgggct gaatcaggca gccacagaac tggtgcaggc
ctctcgggga
4141 acccctcagg acctggctcg agcctcaggc cgatttggac aggacttcag
caccttcctg
4201 gaagctggtg tggagatggc aggccaggct ccgagccagg aggaccgagc
ccaagttgtg
4261 tccaacttga agggcatctc catgtcttca agcaaacttc ttctggctgc
caaggccctg
4321 tccacggacc ctgctgcccc taacctcaag agtcagctgg ctgcagctgc
cagggcagta
4381 actgacagca tcaatcagct catcactatg tgcacccagc aggcacccgg
ccagaaggag
4441 tgtgataacg ccctgcggga attggagacg gtccgggaac tcctggagaa
cccagtccag

4501 cccatcaatg acatgtccta ctttggttgc ctggacagtg taatggagaa
ctcaaaggtg
4561 ctgggcgagg ccatgactgg catctcccaa aatgccaaga acggaaacct
gccagagttt
4621 ggagatgcca tttccacagc ctcaaaggca ctttgtggct tcaccgaggc
agctgcacag
4681 gctgcatatc tggttggtgt ctctgacccc aatagccaag ctggacagca
agggctagtg
4741 gagcccacac agtttgcccg tgcaaaccag gcaattcaga tggcctgcca
gagtttggga
4801 gagcctggct gtacccaggc ccaggtgctc tctgcagcca ccattgtggc
taaacacacc
4861 tctgcactgt gtaacagctg tcgcctggct tctgcccgta ccaccaatcc
tactgccaag
4921 cgccagtttg tacagtcagc caaggaggtg gccaacagca cagctaatct
tgtcaagacc
4981 atcaaggcgc tagatggggc cttcacagag gagaaccgtg cccagtgccg
agcagcaaca
5041 gcccctctgc tggaggctgt ggacaatctg agtgcctttg cgtccaaccc
tgagttctcc
5101 agcattcctg cccagatcag ccctgagggt cgggctgcca tggagcccat
tgtgatctct
5161 gccaagacaa tgttagagag tgccgggggа ctcatccaga cagcccgggc
cctcgcagtc
5221 aatccccggg acccccccgag ctggtcggtg ctggccggcc actcccgtac
tgtctcagac
5281 tccatcaaga agctaattac aagcatgagg gacaaggctc cagggcagct
ggagtgtgaa
5341 acggccattg cagctctgaa cagttgtcta cgggacctag accaggcttc
cctcgctgca
5401 gtcagccagc agcttgctcc ccgtgaggga atctctcaag aggccttgca
cactcagatg
5461 ctcactgcag tccaagagat ctcccatctc attgagccgc tggccaatgc
tgcccgggct
5521 gaagcctccc agctgggaca caaggtgtcc cagatggcgc agtactttga
gccgctcacc
5581 ctggctgcag tgggtgctgc ctccaagacc ctgagccacc cgcagcagat
ggcactcctg
5641 gaccagacta aaacattggc agagtctgcc ctgcagttgc tatacactgc
caaggaggct
5701 ggtggtaacc caaagcaagc agctcacacc caggaagccc tggaggaggc
tgtgcagatg
5761 atgaccgagg ccgtagagga cctgacaaca accctcaacg aggcagccag
tgctgctggg
5821 gtcgtgggtg gcatggtgga ctccatcacc caggccatca accagctaga
tgaaggacca
5881 atgggtgaac cagaaggttc cttcgtggat taccaaacaa ctatggtgcg
gacagccaag
5941 gccattgcag tgaccgttca ggagatggtt accaagtcaa acaccagccc
agaggagctg
6001 ggccctcttg ctaaccagct gaccagtgac tatggccgtc tggcctcgga
ggccaagcct
6061 gcagcggtgg ctgctgaaaa tgaagagata ggttcccata tcaaacaccg
ggtacaggag
6121 ctgggccatg gctgtgccgc tctggtcacc aaggcaggcg ccctgcagtg
cagccccagt
6181 gatgcctaca ccaagaagga gctcatagag tgtgcccgga gagtctctga
gaaggtctcc
6241 cacgtcctgg ctgcgctcca ggctgggaat cgtggcaccc aggcctgcat
cacagcagcc

6301 agcgctgtgt ctggtatcat tgctgacctc gacaccacca tcatgttcgc
cactgctggc
6361 acgctcaatc gtgagggtac tgaaactttc gctgaccacc gggagggcat
cctgaagact
6421 gcgaaggtgc tggtggagga caccaaggtc ctggtgcaaa acgcagctgg
gagccaggag
6481 aagttggcgc aggctgccca gtcctccgtg gcgaccatca cccgcctcgc
tgatgtggtc
6541 aagctgggtg cagccagcct gggagctgag gaccctgaga cccaggtggt
actaatcaac
6601 gcagtgaaag atgtagccaa agccctggga gacctcatca gtgcaacgaa
ggctgcagct
6661 ggcaaagttg gagatgaccc tgctgtgtgg cagctaaaga actctgccaa
ggtgatggtg
6721 accaatgtga catcattgct taagacagta aaagccgtgg aagatgaggc
caccaaaggc
6781 actcgggccc tggaggcaac cacagaacac atacggcagg agctggcggt
tttctgttcc
6841 ccagagccac ctgccaagac ctctacccca gaagacttca tccgaatgac
caagggtatc
6901 accatggcaa ccgccaaggc cgttgctgct ggcaattcct gtcgccagga
agatgtcatt
6961 gccacagcca atctgagccg ccgtgctatt gcagatatgc ttcgggcttg
caaggaagca
7021 gcttaccacc cagaagtggc ccctgatgtg cggcttcgag ccctgcacta
tggccgggag
7081 tgtgccaatg gctacctgga actgctggac catgtactgc tgaccctgca
gaagccaagc
7141 ccagaactga agcagcagtt gacaggacat tcaaagcgtg tggctggttc
cgtcactgag
7201 ctcatccagg ctgctgaagc catgaaggga acagaatggg tagacccaga
ggaccccaca
7261 gtcattgctg agaatgagct cctgggagct gcagccgcca ttgaggctgc
agccaaaaag
7321 ctagagcagc tgaagccccg ggccaaaccc aaggaggcag atgagtcctt
gaactttgag
7381 gagcagatac tagaagctgc caagtccatt gcagcagcca ccagtgcact
ggtaaaggct
7441 gcgtcggctg cccagagaga actagtggcc caagggaagg tgggtgccat
tccagccaat
7501 gcactggacg atgggcagtg gtcccagggc ctcatttctg ctgcccggat
ggtggctgcg
7561 gccaccaaca atctgtgtga ggcagccaat gcagctgtac aaggccatgc
cagccaggag
7621 aagctcatct catcagccaa gcaggtagct gcctccacag cccagctcct
tgtggcctgc
7681 aaggtcaagg ctgaccagga ctcggaggca atgaaacgac ttcaggctgc
tggcaacgca
7741 gtgaagcgag cctcagataa tctggtgaaa gcagcacaga aggctgcagc
ctttgaagag
7801 caggagaatg agacagtggt ggtgaaagag aagatggttg gcggcattgc
ccagatcatc
7861 gcagcacagg aagaaatgct tcggaaggaa cgagagctgg aagaggcgcg
gaagaaactg
7921 gcccagatcc ggcagcagca gtacaagttt ctgccttcag agcttcgaga
tgagcactaa
7981 agaagcctct tctatttaat gcagacccgg cccagagact gtgcgtgcca
ctaccaaagc
8041 cttctgggct gtcggggccc aacctgccca accccagcac tccccaaagt
gcctgccaaa

```
     8101 ccccagggcc tggccccgcc cagtcccgca gtacatcccc tgtcccctcc
ccaaccccaa
     8161 gtgccttcat gccctagggc cccccaagtg cctgcccctc cccagagtat
taacgctcca
     8221 agagtattat taacgctgct gtacctcgat ctgaatctgc cggggcccca
gcccactcca
     8281 ccctgccagc agcttccggc cagtccccac agcctcatca gctctcttca
ccgttttttg
     8341 atactatctt cccccacccc cagctaccca taggggctgc agagttataa
gccccaaaca
      8401 ggtcatgctc caataaaaat gattctacct acaa
```

Protein sequence:

NCBI Reference Sequence: NP_006280.3

LOCUS NP_006280

ACCESSION NP_006280

```
      1 mvalslkisi gnvvktmqfe pstmvydacr iireripeap agppsdfglf
lsdddpkkgi
       61 wleagkaldy ymlrngdtme yrkkqrplki rmldgtvkti mvddsktvtd
mlmticarig
      121 itnhdeyslv relmeekkee itgtlrkdkt llrdekkmek lkqklhtdde
lnwldhgrtl
      181 reqgveehet lllrrkffys dqnvdsrdpv qlnllyvqar ddilngshpv
sfdkacefag
      241 fqcqiqfgph neqkhkagfl dlkdflpkey vkqkgerkif qahkncgqms
eieakvryvk
      301 larslktygv sfflvkekmk gknklvprll gitkecvmrv dektkeviqe
wnltnikrwa
      361 aspksftldf gdyqdgyysv qttegeqiaq liagyidiil kkkkskdhfg
legdeestml
      421 edsvspkkst vlqqqynrvg kvehgsvalp aimrsgasgp enfqvgsmpp
aqqqitsgqm
      481 hrghmpplts aqqaltgtin ssmqavqaaq atlddfdtlp plgqdaaska
wrknkmdesk
      541 heihsqvdai tagtasvvnl tagdpaetdy tavgcavtti ssnltemsrg
vkllaalled
      601 eggsgrpllq aakglagavs ellrsaqpas aeprqnllqa agnvgqasge
llqqigesdt
      661 dphfqdalmq lakavasaaa alvlkaksva qrtedsglqt qviaaatqca
lstsqlvact
      721 kvvaptissp vcqeqlveag rlvakavegc vsasqaated gqllrgvgaa
atavtqalne
      781 llqhvkahat gagpagrydq atdtiltvte nifssmgdag emvrqarila
qatsdlvnai
      841 kadaegesdl ensrkllsaa kiladatakm veaakgaaah pdseeqqqrl
reaaeglrma
      901 tnaaaqnaik kklvqrleha akqaaasatq tiaaaqhaas tpkasagpqp
llvqsckava
      961 eqipllvqgv rgsqaqpdsp saqlaliaas qsflqpggkm vaaakasvpt
iqdqasamql
     1021 sqcaknlgta laelrtaaqk aqeacgplem dsalsvvqnl ekdlqevkaa
ardgklkplp
     1081 getmekctqd lgnstkavss aiaqllgeva qgnenyagia ardvagglrs
laqaargvaa
     1141 ltsdpavqai vldtasdvld kasslieeak kaaghpgdpe sqqrlaqvak
avtqalnrcv
```

```
     1201 sclpgqrdvd nalravgdas krllsdslpp stgtfqeaqs rlneaaagln
qaatelvqas
     1261 rgtpqdlara sgrfgqdfst fleagvemag qapsqedraq vvsnlkgism
sssklllaak
     1321 alstdpaapn lksqlaaaar avtdsinqli tmctqqapgq kecdnalrel
etvrellenp
     1381 vqpindmsyf gcldsvmens kvlgeamtgi sqnakngnlp efgdaistas
kalcgfteaa
     1441 aqaaylvgvs dpnsqagqqg lveptqfara nqaiqmacqs lgepgctqaq
vlsaativak
     1501 htsalcnscr lasarttnpt akrqfvqsak evanstanlv ktikaldgaf
teenraqcra
     1561 ataplleavd nlsafasnpe fssipaqisp egraamepiv isaktmlesa
ggliqtaral
     1621 avnprdppsw svlaghsrtv sdsikklits mrdkapgqle cetaiaalns
clrdldqasl
     1681 aavsqqlapr egisqealht qmltavqeis hlieplanaa raeasqlghk
vsqmaqyfep
     1741 ltlaavgaas ktlshpqqma lldqtktlae salqllytak eaggnpkqaa
htqealeeav
     1801 qmmteavedl tttlneaasa agvvggmvds itqainqlde gpmgepegsf
vdyqttmvrt
     1861 akaiavtvqe mvtksntspe elgplanqlt sdygrlasea kpaavaaene
eigshikhrv
     1921 qelghgcaal vtkagalqcs psdaytkkel iecarrvsek vshvlaalqa
gnrgtqacit
     1981 aasavsgiia dldttimfat agtlnregte tfadhregil ktakvlvedt
kvlvqnaags
     2041 qeklaqaaqs svatitrlad vvklgaaslg aedpetqvvl inavkdvaka
lgdlisatka
     2101 aagkvgddpa vwqlknsakv mvtnvtsllk tvkavedeat kgtraleatt
ehirqelavf
     2161 cspeppakts tpedfirmtk gitmatakav aagnscrqed viatanlsrr
aiadmlrack
     2221 eaayhpevap dvrlralhyg recangylel ldhvlltlqk pspelkqqlt
ghskrvagsv
     2281 teliqaaeam kgtewvdped ptviaenell gaaaaieaaa kkleqlkpra
kpkeadesln
     2341 feeqileaak siaaatsalv kaasaaqrel vaqgkvgaip analddgqws
qglisaarmv
     2401 aaatnnlcea anaavqghas qeklissakq vaastaqllv ackvkadqds
eamkrlqaag
     2461 navkrasdnl vkaaqkaaaf eeqenetvvv kekmvggiaq iiaaqeemlr
kereleeark
      2521 klaqirqqqy kflpselrde h
```

PSME2,

Official Symbol: PSME2 and Name: proteasome (prosome, macropain) activator subunit 2 (PA28 beta) [Homo sapiens]

Other Aliases: PA28B, PA28beta, REGbeta

Other Designations: 11S regulator complex beta subunit; 11S regulator complex subunit beta; MCP activator, 31-kD subunit; REG-beta; activator of multicatalytic protease subunit 2; cell migration-inducing protein 22; proteasome activator 28 subunit beta; proteasome activator 28-beta; proteasome activator complex subunit 2; proteasome activator hPA28 subunit beta

LOCUS NM_002818

ACCESSION NM_002818

VERSION NM_002818.2 GI:30410791

```
        1 tggggagtga aagcgaaagc ccgggcgact agccgggaga ccagagatct
agcgactgaa
       61 gcagcatggc caagccgtgt ggggtgcgcc tgagcgggga agcccgcaaa
caggtggagg
      121 tcttcagaca gaatcttttc caggaggctg aggaattcct ctacagattc
ttgccacaga
      181 aaatcatata cctgaatcag ctcttgcaag aggactccct caatgtggct
gacttgactt
      241 ccctccgggc cccactggac atccccatcc cagaccctcc acccaaggat
gatgagatgg
      301 aaacagataa gcaggagaag aaagaagtcc ataagtgtgg atttctccct
gggaatgaga
      361 aagtcctgtc cctgcttgcc ctggttaagc cagaagtctg gactctcaaa
gagaaatgca
      421 ttctggtgat tacatggatc caacacctga tccccaagat tgaagatgga
aatgattttg
      481 gggtagcaat ccaggagaag gtgctggaga gggtgaatgc cgtcaagacc
aaagtggaag
      541 ctttccagac aaccatttcc aagtacttct cagaacgtgg ggatgctgtg
gccaaggcct
      601 ccaaggagac tcatgtaatg gattaccggg ccttggtgca tgagcgagat
gaggcagcct
      661 atggggagct cagggccatg gtgctggacc tgagggcctt ctatgctgag
ctttatcata
      721 tcatcagcag caacctggag aaaattgtca acccaaaggg tgaagaaaag
ccatctatgt
      781 actgaacccg ggactagaag gaaaataaat gatctatatg ttgtgtgga
```

LOCUS NP_002809

ACCESSION NP_002809

VERSION NP_002809.2 GI:30410792

```
        1 makpcgvrls gearkqvevf rqnlfqeaee flyrflpqki iylnqllqed
slnvadltsl
       61 rapldipipd pppkddemet dkqekkevhk cgflpgnekv lsllalvkpe
vwtlkekcil
      121 vitwiqhlip kiedgndfgv aiqekvlerv navktkveaf qttiskyfse
rgdavakask
      181 ethvmdyral vherdeaayg elramvldlr afyaelyhii ssnlekivnp
kgeekpsmy
```

Q9BQE5

Official Symbol: APOL2 and Name: apolipoprotein L, 2

Other Aliases: APOL-II, APOL3

Other Designations: apolipoprotein L-II; apolipoprotein L2

LOCUS NM_030882

ACCESSION NM_030882

VERSION NM_030882.2 GI:22035654

```
        1 gtgctgggga gcagcgtgtt tactgtgctt ggtcatgagc tgctgggaag
ttgtgacttt
       61 cactttccct ttcgaattcc agggtatatc tgggaggccg gaggacgtgt
ctggttatta
      121 cacagatgca cagctggacg tgggatccac acagctcaga acagttggat
cttgctcagt
      181 ctctgtcaga ggaagatccc ttggacaaga ggaccctgcc ttggtgtgag
agtgagggaa
      241 gaggaagctg gaacgagggt taaggaaaac cttccagtct ggacagtgac
tggagagctc
      301 caaggaaagc ccctcggtaa cccagccgct ggcaccatga acccagagag
cagtatcttt
      361 attgaggatt accttaagta tttccaggac caagtgagca gagagaatct
gctacaactg
      421 ctgactgatg atgaagcctg gaatggattc gtggctgctg ctgaactgcc
cagggatgag
      481 gcagatgagc tccgtaaagc tctgaacaag cttgcaagtc acatggtcat
gaaggacaaa
      541 aaccgccacg ataaagacca gcagcacagg cagtggtttt tgaaagagtt
tcctcggttg
      601 aaaagggagc ttgaggatca cataaggaag ctccgtgccc ttgcagagga
ggttgagcag
      661 gtccacagag gcaccaccat tgccaatgtg gtgtccaact ctgttggcac
tacctctggc
      721 atcctgaccc tcctcggcct gggtctggca cccttcacag aaggaatcag
ttttgtgctc
      781 ttggacactg gcatgggtct gggagcagca gctgctgtgg ctgggattac
ctgcagtgtg
      841 gtagaactag taaacaaatt gcgggcacga gcccaagccc gcaacttgga
ccaaagcggc
      901 accaatgtag caaaggtgat gaaggagttt gtgggtggga acacacccaa
tgttcttacc
      961 ttagttgaca attggtacca agtcacacaa gggattggga ggaacatccg
tgccatcaga
     1021 cgagccagag ccaaccctca gttaggagcg tatgccccac ccccgcatat
cattgggcga
     1081 atctcagctg aaggcggtga acaggttgag agggttgttg aaggccccgc
ccaggcaatg
     1141 agcagaggaa ccatgatcgt gggtgcagcc actggaggca tcttgcttct
gctggatgtg
     1201 gtcagccttg catatgagtc aaagcacttg cttgaggggg caaagtcaga
gtcagctgag
     1261 gagctgaaga agcgggctca ggagctggag gggaagctca actttctcac
caagatccat
     1321 gagatgctgc agccaggcca agaccaatga ccccagagca gtgcagccac
cagggcagaa
     1381 atgccgggca caggccagga caaaatgcag actttttttt ttttttttt
ttttttttga
     1441 gatggagtct cgctctatcg cccaggatgg agtgcagtgg ctcaatctcg
gctcactgca
     1501 aactccgcct cccgggttca caccattctc cggcctcagt ctcccgagta
gctgggacta
```

```
     1561 caggcacctg ccaccacgcc cggctaattt ttttgtattt tcactggaga
cggggtttca
     1621 ctgtgttagc cacgatggtc tccatctcct gacctcgtga tctgcccacc
tcggcctccc
     1681 aaagtgctgg gattacaggc gtgagccacc gcgcctggcc aaaatgcaga
cattttatta
     1741 gggggataag gagggcaagg taaagcttat ggaactgagt gttagtgact
ttggcatttg
     1801 tgtagctgag cacagcaagg gaggggttaa tgcagatggc aagtgcacca
aggagaaggc
     1861 aggaacactg gagcctgcaa taagggagga gagaggactg gagagtgtgg
ggaatgggaa
     1921 gaagtagttt actttggact aaagaatata ttgggcgaag aatagagggg
gagcttgcag
     1981 gaaccagcaa tgagaaggcc aggaaaagaa agagctgaaa atggagaaaa
ccagagttag
     2041 aactgttgga tacaggagaa gaaacagcag ctccactacc gacccccccc
caggtttgat
     2101 gtccttccaa gaataaagtc tttccctggt gatggtctct cgctctgtct
ttccagcatc
     2161 cactctccct tgtccttctg ggggtgtatc acagtcagcc agtggcttct
tcatgatggt
     2221 ggttggggtg gttgtcatgt gacgggtccc ctccaggtta ctaaagggtg
catgtcccct
     2281 gcttgaaccc tgagaggcag gtggtaggcc atggccacaa tccccagctg
aggagcaggt
     2341 gtccctgaga acccaaactt cccagagagt atctgagaac caaccaatga
aaacagtccc
     2401 atcgctctta gccggtaagt aaacagtcag aagattagca tgaaagcagt
ttagcattgg
     2461 gaggaagcac agatctctag agctgtcctg tcgctgccca ggattgacct
gtgtgtaagt
     2521 cccaataaac tcacctactc accaa
```

LOCUS NP_112092

ACCESSION NP_112092

VERSION NP_112092.1 GI:13562090

```
        1 mnpessifie dylkyfqdqv srenllqllt ddeawngfva aaelprdead
elrkalnkla
       61 shmvmkdknr hdkdqqhrqw flkefprlkr eledhirklr alaeeveqvh
rgttianvvs
      121 nsvgttsgil tllglglapf tegisfvlld tgmglgaaaa vagitcsvve
lvnklraraq
      181 arnldqsgtn vakvmkefvg gntpnvltlv dnwyqvtqgi grnirairra
ranpqlgaya
      241 ppphiigris aeggeqverv vegpaqamsr gtmivgaatg gillllqvvs
layeskhlle
      301 gaksesaeel kkraqelegk lnfltkihem lqpgqdq
```

Q9Y262

Official Symbol: EIF3L and Name: eukaryotic translation initiation factor 3, subunit L

Other Aliases: AL022311.1, EIF3EIP, EIF3S11, EIF3S6IP, HSPC021, HSPC025, MSTP005

Other Designations: eIEF associated protein HSPC021; eukaryotic translation initiation factor 3 subunit 6-interacting protein; eukaryotic translation initiation factor 3 subunit E-interacting protein; eukaryotic translation initiation factor 3 subunit L

LOCUS NM_001242923

ACCESSION NM_001242923

VERSION NM_001242923.1 GI:339275830

```
        1 gctgaacttc cggcctcagg acgcaggcgc gggccgctca tttcgctctt
tccggcggtg
       61 ctcgcaagcg aggcagccat gtcttatccc gctgatgatt atgagtctga
ggcggcttat
      121 gacccctacg cttatcccag cgactatgat atgcacacag gagatccaaa
gcaggacctt
      181 gcttatgaac gtcagtatga acagcaaacc tatcaggtga tccctgaggt
gatcaaaaac
      241 ttcatccagt atttccacaa aactgtctca gatttgattg accagaaagt
gtatgagcta
      301 caggccagtc gtgtctccag tgatgtcatt gaccagaagg tgtatgagat
ccaggacatc
      361 tatgagaaca gctggaccaa gctgactgaa agattcttca agaatacacc
ttggcccgag
      421 gctgaagcca ttgctccaca ggttggcaat gatgctgtct tcctgatttt
atacaaagaa
      481 ttatactaca ggcacatata tgccaaagtc agttttcagt cattcagtca
gtaccgctgt
      541 aagactgcca agaagtcaga ggaggagatt gactttcttc gttccaatcc
caaaatctgg
      601 aatgttcata gtgtcctcaa tgtccttcat tccctggtag acaaatccaa
catcaaccga
      661 cagttggagg tatacacaag cggaggtgac cctgagagtg tggctgggga
gtatgggcgg
      721 cactccctct acaaaatgct tggttacttc agcctggtcg ggcttctccg
cctgcactcc
      781 ctgttaggag attactacca ggccatcaag gtgctggaga acatcgaact
gaacaagaag
      841 agtatgtatt cccgtgtgcc agagtgccag gtcaccacat actattatgt
tgggtttgca
      901 tatttgatga tgcgtcgtta ccaggatgcc atccgggtct tcgccaacat
cctcctctac
      961 atccagagga ccaagagcat gttccagagg accacgtaca agtatgagat
gattaacaag
     1021 cagaatgagc agatgcatgc gctgctggcc attgccctca cgatgtaccc
catgcgtatt
     1081 gatgagagca ttcacctcca gctgcgggag aaatatgggg acaagatgtt
gcgcatgcag
     1141 aaaggtgacc cacaagtcta tgaagaactt ttcagttact cctgccccaa
gttcctgtcg
     1201 cctgtagtgc ccaactatga taatgtgcac cccaactacc acaaagagcc
cttcctgcag
     1261 cagctgaagg tgttttctga tgaagtacag cagcaggccc agctttcaac
catccgcagc
     1321 ttcctgaagc tctacaccac catgcctgtg gccaagctgg ctggcttcct
ggacctcaca
```

```
     1381 gagcaggagt tccggatcca gcttcttgtc ttcaaacaca agatgaagaa
cctcgtgtgg
     1441 accagcggta tctcagccct ggatggtgaa tttcagtcag cctcagaggt
tgacttctac
     1501 attgataagg acatgatcca catcgcggac accaaggtcg ccaggcgtta
tggggatttc
     1561 ttcatccgtc agatccacaa atttgaggag cttaatcgaa ccctgaagaa
gatgggacag
     1621 agaccttgat gatattcaca cacattcagg aacctgtttt gatgtattat
aggcaggaag
     1681 tgtttttgct accgtgaaac ctttacctag atcagccatc agcctgtcaa
ctcagttaac
     1741 aagttaagga ccgaagtgtt tcaagtggat ctcagtaaag gatctttgga
gccagatttg
     1801 tcgtctcatt attgtaggag agaatttgtg ggttgtggca gtaatacatt
tcccatgtgt
     1861 cctgatgctt tcaggataca tcagttgtta gtgtttaaat tgagttattt
ttatttttgtg
     1921 cttttgagat ggagtctcac tctgtct
```

LOCUS NP_001229852

ACCESSION NP_001229852

VERSION NP_001229852.1 GI:339275831

```
        1 msypaddyes eaaydpyayp sdydmhtgdp kqdlayerqy eqqtyqvipe
viknfiqyfh
       61 ktvsdlidqk vyelqasrvs sdvidqkvye iqdiyenswt klterffknt
pwpeaeaiap
      121 qvgndavfli lykelyyrhi yakvsfqsfs qyrcktakks eeeidflrsn
pkiwnvhsvl
      181 nvlhslvdks ninrqlevyt sggdpesvag eygrhslykm lgyfslvgll
rlhsllgdyy
      241 qaikvlenie lnkksmysrv pecqvttyyy vgfaylmmrr yqdairvfan
illyiqrtks
      301 mfqrttykye minkqneqmh allaialtmy pmridesihl qlrekygdkm
lrmqkgdpqv
      361 yeelfsyscp kflspvvpny dnvhpnyhke pflqqlkvfs devqqqaqls
tirsflklyt
      421 tmpvaklagf ldlteqefri qllvfkhkmk nlvwtsgisa ldgefqsase
vdfyidkdmi
      481 hiadtkvarr ygdffirqih kfeelnrtlk kmgqrp
```

Official Symbol: RAB1B and Name: RAB1B, member RAS oncogene family [Homo sapiens]

Other Designations: ras-related protein Rab-1B; small GTP-binding protein

LOCUS NM_030981

ACCESSION NM_030981 XM_001134089

VERSION NM_030981.2 GI:116014337

```
        1 gcgggcgggg cctctggggc ggagcggcca ccatcttgga acgggaggcg
gagcagagtc
       61 gactgggagc gaccgagcgg gccgccgccg ccgccatgaa ccccgaatat
gactacctgt
      121 ttaagctgct tttgattggc gactcaggcg tgggcaagtc atgcctgctc
ctgcggtttg
      181 ctgatgacac gtacacagag agctacatca gcaccatcgg ggtggacttc
aagatccgaa
      241 ccatcgagct ggatggcaaa actatcaaac ttcagatctg ggacacagcg
ggccaggaac
      301 ggttccggac catcacttcc agctactacc ggggggctca tggcatcatc
gtggtgtatg
      361 acgtcactga ccaggaatcc tacgccaacg tgaagcagtg gctgcaggag
attgaccgct
      421 atgccagcga gaacgtcaat aagctcctgg tgggcaacaa gagcgacctc
accaccaaga
      481 aggtggtgga caacaccaca gccaaggagt ttgcagactc tctgggcatc
cccttcttgg
      541 agacgagcgc caagaatgcc accaatgtcg agcaggcgtt catgaccatg
gctgctgaaa
      601 tcaaaaagcg gatggggcct ggagcagcct ctgggggcga gcggcccaat
ctcaagatcg
      661 acagcacccc tgtaaagccg gctggcggtg gctgttgcta ggaggggcac
atggagtggg
      721 acaggagggg gcaccttctc cagatgatgt ccctggaggg ggcaggaggt
acctccctct
      781 ccctctcctg gggcatttga gtctgtggct ttggggtgtc ctgggctccc
catctccttc
      841 tggcccatct gcctgctgcc ctgagccccg gttctgtcag ggtccctaag
ggaggacact
      901 cagggcctgt ggccaggcag ggcggaggcc tgctgtgctg ttgcctctag
gtgactttcc
      961 aagatgcccc cctacacacc tttctttgga acgagggctc ttctgtcggt
gtccctccca
     1021 cccccatgta tgctgcactg ggttctctcc ttcttcttcc tgctgtcctg
cccaagaact
     1081 gagggtctcc ccggcctcta ctgccctggc tgcagtcagt gcccagggcg
aggaatgtgg
     1141 ccaggggatc caggacctgg gatccagggc cctgggctgg acctcaggac
aggcatggag
     1201 gccacagggg cccagcagcc cacccttcc tctccccact gcctcctctc
ccttcctaca
     1261 ctcccagctc gagccgtcca gctgcggtgg gatctgagta tatctagggc
gggtgggcgg
     1321 gtagcagtgc tgggcctgtg tcttgagcct ggagggagtc tgctcctgcc
gccctctgcc
     1381 ctgccagaga cagacccatg cgctgcctgc ccaccgtgcc cctttgtccc
catgtcaggc
     1441 ggaggcggaa ggcccaccgt gccagaggct gggcaccagc cttaaccctc
actctgctag
     1501 cacctcctcc ctttccccaa ggtagcacat ctggctcact ccccactccg
tctctggagc
     1561 ccaccaggga aggccctcat cccctgccgc tacttctctg gggaatgtgg
gttccatcca
     1621 ggattggggg cctctctgct cacccactct gcacccagga tcctagtccc
ctgccctctg
     1681 gcacagctgc ttcctgcaag aaagcaagtc tttggtctcc ctgagaagcc
atgtccctcg
     1741 tgctgtctct tgcctgtccc acctgtgccc tgccctccag cttgtattta
agtccctggg
```

```
     1801 ctgccccctt ggggtgcccc ccgctcccag gttcccctct ggtgtcatgt
caggcatttt
     1861 gcaaggaaaa gccacttggg gaaagatgga aaaggacaaa aaaaattaat
aaatttccat
     1921 tggccctcgg gtgagctgag ggtttttgca aggaa
```

LOCUS NP_112243

ACCESSION NP_112243 XP_001134089

VERSION NP_112243.1 GI:13569962

```
        1 mnpeydylfk llligdsgvg kscllllrfad dtytesyist igvdfkirti
eldgktiklq
       61 iwdtagqerf rtitssyyrg ahgiivvydv tdqesyanvk qwlqeidrya
senvnkllvg
      121 nksdlttkkv vdnttakefa dslgipflet saknatnveq afmtmaaeik
krmgpgaasg
      181 gerpnlkids tpvkpagggc c
```

Official Symbol: RPS6 provided by HGNC

Official Full Name: ribosomal protein S6provided by HGNC

Also known as: S6

LOCUS NM_001010

ACCESSION NM_001010

VERSION NM_001010.2 GI:17158043

```
        1 cctcttttcc gtggcgcctc ggaggcgttc agctgcttca agatgaagct
gaacatctcc
       61 ttcccagcca ctggctgcca gaaactcatt gaagtggacg atgaacgcaa
acttcgtact
      121 ttctatgaga agcgtatggc cacagaagtt gctgctgacg ctctgggtga
agaatggaag
      181 ggttatgtgg tccgaatcag tggtgggaac gacaaacaag gtttccccat
gaagcagggt
      241 gtcttgaccc atggccgtgt ccgcctgcta ctgagtaagg ggcattcctg
ttacagacca
      301 aggagaactg gagaaagaaa gagaaaatca gttcgtggtt gcattgtgga
tgcaaatctg
      361 agcgttctca acttggttat tgtaaaaaaa ggagagaagg atattcctgg
actgactgat
      421 actacagtgc ctcgccgcct gggccccaaa agagctagca gaatccgcaa
acttttcaat
      481 ctctctaaag aagatgatgt ccgccagtat gttgtaagaa agcccttaaa
taaagaaggt
      541 aagaaaccta ggaccaaagc acccaagatt cagcgtcttg ttactccacg
tgtcctgcag
      601 cacaaacggc ggcgtattgc tctgaagaag cagcgtacca agaaaaataa
agaagaggct
      661 gcagaatatg ctaaactttt ggccaagaga atgaaggagg ctaaggagaa
gcgccaggaa
```

```
      721 caaattgcga agagacgcag actttcctct ctgcgagctt ctacttctaa
gtctgaatcc
      781 agtcagaaat aagatttttt gagtaacaaa taaataagat cagactctg
```

LOCUS NP_001001

ACCESSION NP_001001

VERSION NP_001001.2 GI:17158044

```
        1 mklnisfpat gcqklievdd erklrtfyek rmatevaada lgeewkgyvv
risggndkqg
       61 fpmkqgvlth grvrlllskg hscyrprrtg erkrksvrgc ivdanlsvln
lvivkkgekd
      121 ipgltdttvp rrlgpkrasr irklfnlske ddvrqyvvrk plnkegkkpr
tkapkiqrlv
      181 tprvlqhkrr rialkkqrtk knkeeaaeya kllakrmkea kekrqeqiak
rrrlsslras
      241 tsksessqk
```

Official SymbolRRP1

Official Full Nameribosomal RNA processing 1 homolog (S. cerevisiae)

Also known asNNP-1; NOP52; RRP1A; D21S2056E

LOCUS NM_003683

ACCESSION NM_003683

VERSION NM_003683.5 GI:134304836

```
        1 gccggggcca ggaggaggcg ggcgcaggag gcgcgtgctc agtgtgctgg
gtaccaggcg
       61 actccgggac agggggtctc ggccgtcggc gtcatggttt cgcgcgtgca
gctcccgcct
      121 gagatccagc tggctcagcg cctggcgggg aatgagcagg tgacccggga
ccgggcggtg
      181 aggaagctcc ggaaatacat cgtcgccagg actcagcggg ccgcaggtgg
ttttacgcac
      241 gacgagctgc tgaaggtgtg gaaaggactg ttttattgca tgtggatgca
ggacaagcca
      301 ctcctccagg aagaattagg aaggactatt tcccagctcg ttcatgcttt
tcagaccacg
      361 gaggcgcagc acctgttcct tcaggccttc tggcagacca tgaatcgcga
gtggacgggc
      421 attgacaggc tgcgcctgga taaattctac atgctcatgc ggatggtcct
gaacgagtcc
      481 ttgaaggttc tgaagatgca aggctgggaa gaaagacaga tcgaggagct
gctagagctg
      541 ctgatgactg agatcctgca ccccagcagc caggccccca acggtgtgaa
gagccacttc
      601 atcgagatct tcctggagga gctgaccaaa gtgggcgccg aggagcttac
ggcagaccag
      661 aacctgaagt tcatcgaccc cttctgcaga attgctgccc ggaccaagga
ttccttggtt
```

```
      721 ttgaacaaca tcactcgagg catctttgag acgattgtgg agcaggcccc
gcttgccatt
      781 gaagacctcc tgaatgaact ggacacacag gatgaggagg tggcgtcgga
cagtgatgag
      841 tcctctgagg gtggtgagcg tggagacgcg ctgtcccaga agaggtctga
gaagccgccc
      901 gcaggctcca tctgcagggc tgaacctgag gctggtgagg agcaggcagg
tgacgacagg
      961 gacagtggcg gccccgttct ccagtttgac tacgaggcag ttgctaacag
actgtttgaa
     1021 atggccagcc gccagagcac cccttctcag aacagaaagc gtctctacaa
agtgatccgg
     1081 aagctgcagg acctggcagg aggcattttc cctgaagatg agatcccaga
gaaggcctgc
     1141 aggcgcctgc ttgaagggag gcggcagaag aagacgaaga agcagaagcg
tctgctcagg
     1201 ttgcagcagg agagagggaa aggtgagaag gagcccccga gcccgggcat
ggagaggaag
     1261 aggagcagga ggaggggtgt aggggccgac cccgaggcgc gggcagaggc
tggtgagcag
     1321 ccaggcacag ctgagcgggc cctgctccga gatcagccca ggggccgtgg
ccagagaggg
     1381 gctcgccaga gaaggaggac acctcggccc ctgaccagtg cccgagcaaa
ggcggccaat
     1441 gtccaggagc cggagaagaa gaagaaacgc agggagtgat gtggccgggc
caaggacagg
     1501 cagggaggga ggccaggcct cgcttgcacc gcgggacgag gctgaccggg
ctgttctgta
     1561 gactcaggac cgtggctcca gaactcctgt gccaggcggg agggaagggc
ggcactggag
     1621 agatgggccc atcattaggg gccagcatcc caggaactgg acctttcccc
agagcctccg
     1681 cctgtggctg tgatgacctt gggccagaag gtcaaactcc gaagactgaa
actctgcctg
     1741 cagcaggact ggccgcccct gctgtggggg gttcagaaaa taaaatgccg
cgcagccctt
     1801 gccagggaaa aaaaaaaaaa
```

LOCUS NP_003674

ACCESSION NP_003674

VERSION NP_003674.1 GI:4503247

```
        1 mvsrvqlppe iqlaqrlagn eqvtrdravr klrkyivart qraaggfthd
ellkvwkglf
       61 ycmwmqdkpl lqeelgrtis qlvhafqtte aqhlflqafw qtmnrewtgi
drlrldkfym
      121 lmrmvlnesl kvlkmqgwee rqieellell mteilhpssq apngvkshfi
eifleeltkv
      181 gaeeltadqn lkfidpfcri aartkdslvl nnitrgifet iveqaplaie
dllneldtqd
      241 eevasdsdes seggergdal sqkrsekppa gsicraepea geeqagddrd
sggpvlqfdy
      301 eavanrlfem asrqstpsqn rkrlykvirk lqdlaggifp edeipekacr
rllegrrqkk
      361 tkkqkrllrl qqergkgeke ppspgmerkr srrrgvgadp earaeageqp
gtaerallrd
      421 qprgrgqrga rqrrrtprpl tsarakaanv qepekkkkrr e
```

Summary Official Symbol: SEPT11

Official Full Name: septin 11

LOCUS NM_018243

ACCESSION NM_018243

VERSION NM_018243.2 GI:38605734

```
        1 ggcgtggggg gagcagatgc cgctggctgc cagcgggacg ccggcgagca
gagcgcagcc
       61 gcgagggagg cgcgagggag gcgagccgga gcccgagcac tagcagcagc
cggagtcggc
      121 gtaaagcacc cgggcgcagc cggagccggt gccgcagctg cgatggccgt
ggccgtgggg
      181 agaccgtcta atgaagagct tcgaaacttg tctttgtctg gccatgtggg
atttgacagc
      241 ctccctgacc agctggtcaa caagtctact tctcaaggat tctgtttcaa
catcctttgt
      301 gttggtgaga caggcattgg caaatccacg ttaatggaca ctttgttcaa
caccaaattt
      361 gaaagtgacc cagctactca caatgaacca ggtgttcggt taaaagccag
aagttatgag
      421 cttcaggaaa gcaatgtacg gctgaagtta accattgttg acaccgtggg
atttggagac
      481 cagataaata aagatgacag ctataagccg atagtagaat atattgatgc
ccagttcgag
      541 gcctacctgc aagaggaatt gaagattaaa cgttctctct tcaactacca
tgacacgagg
      601 atccatgcct gcctctactt tattgcccct actggacatt cactaaagtc
cctggatctg
      661 gtcaccatga aaaagctgga cagtaaggtg aacatcattc caataattgc
aaaagctgac
      721 accattgcca agaatgaact gcacaaattc aagagtaaga tcatgagtga
actggtcagc
      781 aatggggtcc agatatatca gtttcccact gatgaagaaa cggtggcaga
gattaacgca
      841 acaatgagtg tccatctccc atttgcagtg gttggcagca ccgaagaggt
gaagattggc
      901 aacaagatgg caaaggccag gcagtacccc tggggtgtgg tgcaggttga
gaatgaaaat
      961 cattgcgatt ttgtgaaact tcgagagatg ctgatccgcg tgaacatgga
ggacttgcga
     1021 gagcagactc acacccgcca ctatgaattg taccgacgct gtaagcttga
agagatgggg
     1081 ttcaaggaca ctgaccctga cagcaaaccc ttcagtcttc aggagacata
tgaagcaaaa
     1141 aggaatgaat tcctgggaga actgcagaag aaagaagaag aaatgagaca
aatgtttgtt
     1201 atgagagtga aggagaaaga agctgaactt aaggaggcag agaaagagct
tcacgagaag
     1261 tttgaccttc taaagcggac acaccaagaa gaaaagaaga aagtggaaga
caagaagaag
```

1321 gagcttgagg aggaggtgaa caacttccag aagaagaaag cagcggctca
gttactacag
1381 tcccaggccc agcaatctgg ggcccagcaa accaagaaag acaaggataa
gaaaaatgca
1441 agcttcacat aaagcctggc aagccaagga tgttcccgca ttcacctgct
tttgcagtaa
1501 tatcgtatct ctgccatgtg tgttctttag ttttatttta ttttatttta
tttttttacc
1561 cttcctcaaa caccagtaac tattattaac tcgttttgct gaatgttgtt
gggtggtaga
1621 aaatgataga acaagggaat aaccgcgaat gctctgtgca gctggactct
gtttccggaa
1681 agtaaatgat ttgcttttta tgcctgttct gaatggcagc acgaagcagg
cctgttactt
1741 gtatgtcgct ttggacagag gaaagtgggg taaaatgcta cctgtacgtc
tgacatgaaa
1801 acttctcacc gcctcagcag ctgaactaaa aacctgaata gccatgacaa
gagtttgcat
1861 tttcttgatg attcatctcc atgagtgcac aatccctgaa ctcactgtct
tttctccaca
1921 cttgtcctaa gccaaggtag atttgtacgt agacagactg gtgagcaagc
attatatttt
1981 atttttaccc ttgcatgaca ttttcatttt aatcaataac attatttggc
ctgagcttgt
2041 gggtctgttc agactgtctc ctctcatggt ttgaaactgc atctgaatgc
ctgccttcaa
2101 tcctggccaa gttggagtag actggtatga gaaaactatg attagttcac
atttactggt
2161 gcatccttga tcctctcaca gatagaggtc ttaaaggttg gatcatgtaa
cattgcttag
2221 tagaagaatc ttcttctaag gatgatgggc tttctacagc ctgcttacca
ctaacagtaa
2281 ggaatctttc ataaacacac ctcagtttgt tcccagtggg cttagaggga
ggacctgatg
2341 actgattcca ggatacttgt acttctaata acatttttca tgaatcatga
gaaaatttcc
2401 acagatactt cccttagaaa atttgctata aactctgtat cattggtagc
acaaatttga
2461 gcgaggcctt gtcaatttta aggtggaaat aggaaggacc acaacatgac
ccgtaagtca
2521 agaaggtaga catttcatat ccagcttcct tgcttagtct cctttcagta
tttggcaata
2581 aaagaaagaa gaaatagaac agctgaagtc tcaaatcatt gtctggaatt
ttcctcacct
2641 tggctagctc cacctgctct ttgtctaagg cccttgcctc atcagggatt
agaactggcc
2701 catatgccag aacctgtact aaatgcctaa tttgtatgga agagtgcata
tttaatctct
2761 tttctatact gctcctttct gatgcttatc ctttcatctg tgtgattgtt
ttttcccctc
2821 tactaacaag atcctcccag ctttctctct acatgtagaa aggataacat
ttctcatgaa
2881 cccactgccc ctctgcattt tcctcactgg ttagagatta agtaaatagg
atagaatatg
2941 ctgcgtctcc cctgacacac actttctttt ttgaatgagc aagtctccat
tttgatttca
3001 gcaaagattt tttctccttt tctttgtcct caaccatact tagaggaaag
aaggaatggt
3061 cttccatgaa ctgattatgc ttaattaagc aaagtaagga aattagtttc
atggaagcct

3121 aaacaaagct ggaatagaaa ctacacacta gacacagcag tagtcatagt
cttcacaggt
3181 ttaggagcta ctggaccaac attcttgttt ttgcttttgt ttttttaaat
aattctagtc
3241 tggagctaac tgtggagcag ccaaatagta gctggcatgt tgattcaaac
catgggctga
3301 atttgctcat aggctgtgca tcagacaaaa gcttgaatat ttgtgttgta
tgcttgttcc
3361 aaccaccgct tgtgtgagca tttttgtggc ttgtacagaa agtacacttt
taaattgtct
3421 cttgcatcac taaaattttt ttaaaatgag cataacaacg aaaggcatcc
agctgacttt
3481 ttgattccaa gattattgat tggattgact tttttgcatt aaatttttcc
cagcaaaata
3541 aatcatatgg cgagtcaggg aataaaaagt caaaagaaac aaatagaagc
ttttttttt
3601 aaaaaatgta ttgcttctga actttttct gccactgctc cctagccctg
tttagtttgt
3661 tattgctgct tttctttttt ctttctgtat ctatgccttt ttttcacagt
agtccttggc
3721 tctgcacgga ataaatgata ccctcaaatc taattggatg tgctttcgcc
tttgcatgta
3781 agtacggtag taagaaacct ttgagatctt tctgactttt caaaattaga
gaaagcaaat
3841 gggatggata gatttttttt ttcttttcaa ggggggcagg aaggtaatgg
tttgagtagc
3901 ctttgtttaa aaaaaagact aaatatattt aaaaggccac atttatattt
ttttcacaag
3961 aaccacataa taaattccac ttcttgacct gaatttggaa atccgaaatt
actaatccag
4021 gccaggtgtg gtggctcatg cctgtaatcc cagcactttg agaggccgag
gtgggcagat
4081 cacttgaggc ctggagttca agaccacctt ggcgaacacg gtgaaacccc
gtctctacaa
4141 aaaatacaaa aattagccag gcgtggtggc acgtgcctgt agtcccagct
acttgggagg
4201 ctaagtcagg agaattgctt gaacttggga gatggaggtt gcagtgagcc
aagattgcac
4261 cactgcattc caacctgggt gatgaagtga gactctccaa aaaaaaaaaa
gaaattatta
4321 atccctgcct gtgctctaca tagcctcatg ggcatcattg gatagctcag
agggcccttg
4381 attctggcaa ggcaaataaa gccagaatga gaaattacca tcttctacta
gagaaaacca
4441 agagaaaaat ttttatgcta ggatgccttt atgaccactt aatttttaa
tcttagttta
4501 atggtctctc cctggtgcta actgctgaca gtggccacct ctttttgggg
gattgagggg
4561 cctacataac tagctggcct taccccatat cttttgttca aacataatac
catctttttg
4621 cttcttctga actttagatc tccataacac atgtactgta gaatgtgatg
gaaaagcatt
4681 gatgagaatt tattggcagt tcagattgtg ttttcccaac ttaggctctt
tattaattgg
4741 ttaaggtttt ctccaaaaag ggcatttcaa caatgggaat tatttaatgt
aacagtgggc
4801 acagattact tatcttcctt ctctgctttg tgactcacca gcagtaacac
acacaatcca
4861 catcttgtgc acctcaaatg aacagacttg gtttccttgc tttcttgaca
tttccatgac

```
     4921 tgtttcacat acaaactatt gggtgaggtt tttcagctgt taccgaccca
cgtcctgctg
     4981 tctctgtgtg gtcctacaaa aactgtccat tcccacccct ttgctttgcc
atttgcaaga
     5041 gtctggaatt gtcaggtctc agcttcgaaa agtcctggtt ccactgacag
gacacattct
     5101 ttagtgggaa ttaagaccta caaagtctag tttgtatgta ggtatgaagg
gaatttttta
     5161 aataaattga aaagctgtga acagcattag aactttgtct atttcttaat
tttaaaatat
     5221 gctgatatgc cttaaactgt agttgtagat ccttgtcatt ttgctgtttg
aaaataacca
     5281 atgtgttttc taaaactgtc gtgtaatcta ctttcattgt taatgcagaa
ttgtcatata
     5341 tgtaagctgc atgttagaca tttgtctttt ttaaactaaa gtaattgtat
tgatgtgaag
     5401 catatcattt tttcaaatat gaaagtgatc acttagcaac atgcttggta
atttggcatc
     5461 tgttaaggta ggagagtggt gaacagataa tctatgcata tatcactagt
gccaagacat
     5521 aaagcggggg aaaatatatt tttacccaaa cattaaaaaa aaaaaaaaaa
aaaaaaaaaa
     5581 aa
```

LOCUS NP_060713

ACCESSION NP_060713

VERSION NP_060713.1 GI:8922712

```
       1 mavavgrpsn eelrnlslsg hvgfdslpdq lvnkstsqgf cfnilcvget
gigkstlmdt
      61 lfntkfesdp athnepgvrl karsyelqes nvrlkltivd tvgfgdqink
ddsykpivey
     121 idaqfeaylq eelkikrslf nyhdtrihac lyfiaptghs lksldlvtmk
kldskvniip
     181 iiakadtiak nelhkfkski mselvsngvq iyqfptdeet vaeinatmsv
hlpfavvgst
     241 eevkignkma karqypwgvv qvenenhcdf vklremlirv nmedlreqth
trhyelyrrc
     301 kleemgfkdt dpdskpfslq etyeakrnef lgelqkkeee mrqmfvmrvk
ekeaelkeae
     361 kelhekfdll krthqeekkk vedkkkelee evnnfqkkka aaqllqsqaq
qsgaqqtkkd
     421 kdkknasft
```

Official Symbol: SEPT7 and Name: septin 7

Other Aliases: CDC10, CDC3, NBLA02942, SEPT7A

Other Designations: CDC10 (cell division cycle 10, S. cerevisiae, homolog); CDC10 protein homolog; septin-7

LOCUS NM_001011553

ACCESSION NM_001011553

VERSION NM_001011553.3 GI:339639595

```
        1 gagatggaag ccagcctccg ctaggcccgg aagcctcgtc tgaggggcg
ggggacggag
       61 gagggagcgg gagtcgagcg agagcctgtg gaggagtccg cctgctgtag
cgtgcgtaag
      121 caaggcagct acgccgggcg gctacgctgc ggaatcggcg taggcgcctt
tggagaatcg
      181 gcgggctgcg ctccgctggg gctggtcgcg gaggggggga ggggatgtcg
gtcagtgcga
      241 gatccgctgc tgctgaggag aggagcgtca acagcagcac catggctcaa
cagaagaacc
      301 ttgaaggcta tgtgggattt gccaatctcc caaatcaagt atacagaaaa
tcggtgaaga
      361 gaggttttga attcacgctt atggtagtgg gtgaatctgg attgggaaag
tcgacattaa
      421 tcaactcatt attcctcaca gatttgtatt ctccagagta tccaggtcct
tctcatagaa
      481 ttaaaaagac tgtacaggtg gaacaatcca aagttttaat caaagaaggt
ggtgttcagt
      541 tgctgctcac aatagttgat accccaggat ttggagatgc agtggataat
agtaattgct
      601 ggcagcctgt tatcgactac attgatagta aatttgagga ctacctaaat
gcagaatcac
      661 gagtgaacag acgtcagatg cctgataaca gggtgcagtg ttgtttatac
ttcattgctc
      721 cttcaggaca tggacttaaa ccattggata ttgagtttat gaagcgtttg
catgaaaaag
      781 tgaatatcat cccacttatt gccaaagcag acacactcac accagaggaa
tgccaacagt
      841 ttaaaaaaca gataatgaaa gaaatccaag aacataaaat taaaatatac
gaatttccag
      901 aaacagatga tgaagaagaa aataaacttg ttaaaaagat aaaggaccgt
ttacctcttg
      961 ctgtggtagg tagtaatact atcattgaag ttaatggcaa aagggtcaga
ggaaggcagt
     1021 atccttgggg tgttgctgaa gttgaaaatg gtgaacattg tgattttaca
atcctaagaa
     1081 atatgttgat aagaacacac atgcaggact tgaaagatgt tactaataat
gtccactatg
     1141 agaactacag aagcagaaaa cttgcagctg tgacttataa tggagttgat
aacaacaaga
     1201 ataaagggca gctgactaag agccctctgg cacaaatgga agaagaaaga
agggagcatg
     1261 tagctaaaat gaagaagatg gagatggaga tggagcaggt gtttgagatg
aaggtcaaag
     1321 aaaaagttca aaaactgaag gactctgaag ctgagctcca gcggcgccat
gagcaaatga
     1381 aaaagaattt ggaagcacag cacaaagaat tggaggaaaa acgtcgtcag
ttcgaggatg
     1441 agaaagcaaa ctgggaagct caacaacgta ttttagaaca acagaactct
tcaagaacct
     1501 tggaaaagaa caagaagaaa gggaagatct tttaaactct ctattgacca
ccagttaacg
     1561 tattagttgc caatatgcca gcttggacat cagtgtttgt tggatccgtt
tgaccaattt
     1621 gcaccagttt tatccataat gatggattta acagcatgac aaaaattatt
ttttttttg
     1681 ttcttgatgg agattaagat gccttgaatt gtctagggtg ttctgtactt
agaaagtaag
```

1741 agctctaagt acctttccta cattttcttt ttttattaaa cagatatctt
cagtttaatg
1801 caagagaaca ttttactgtt gtacaatcat gttctggtgg tttgattgtt
tacaggatat
1861 tccaaaataa aaggactctg gaagattttc attgaggata aattgccata
atatgatgca
1921 aactgtgctt ctctatgata attacaatac aaaggttcca ttcagtgcag
catatacaat
1981 aatgtaattt agtctaacac agttgaccct atttttgac acttccattg
tttaaaaata
2041 cacatggaaa aaaaaaaaac cctatatgct tactgtgcac ctagagcttt
tttataacaa
2101 cgtctttttg tttgtttgtt ttggattctt taaatatata ttattctcat
ttagtgccct
2161 ctttagccag aatctcatta ctgcttcatt tttgtaataa catttaattt
agatattttc
2221 catatattgg cactgctaaa atagaatata gcatctttca tatggtagga
accaacaagg
2281 aaactttcct ttaactccct ttttacactt tatggtaagt agcaggggggg
gaaatgcatt
2341 tatagatcat ttctaggcaa aattgtgaag ctaatgacca acctgtttct
acctatatgc
2401 agtctcttta ttttactaga aatgggaatc atggcctctt gaagagaaaa
aagtcaccat
2461 tctgcattta gctgtattca tatattgcat ttctgtattt tttgtttgta
ttgtaaaaaa
2521 ttcacataat aaacgatgtt gtgatgtaat attgtgtgag gtcttaaata
tcctacagtc
2581 gatgtacaag agtagagtat gtttgggaag aaacttttca gcttaagttt
gcctcctcta
2641 caatgacatc ttttatatgc ttgtctcatt tagaatgcat atgtgctgat
tttctaattt
2701 aagagatacc atatctctct attcatttct atctctcatt tgtatgctta
tttttctgag
2761 aacatttttt ttttccccca gacagggtct tgcttcattg cccaggctgg
agtgcggtgg
2821 cacaaacacg acttgactgc agcctcaacc ctctgggctc aagcagtcct
cctgcctcag
2881 ccccctgagt atctgggatt gcaggcgtgc accaccacgc ctggctaatt
tttgtatttt
2941 ttgcagcctc ccaaagttct gggattacag gcatgagccg tcatgcctgg
cctctgagaa
3001 cagtttctga ctcattcaga ttaggtatac tctcaagtcc ctggaaactg
aaatttttt
3061 taactgtaaa gagggtagtg tcatttcttt tcttaaggtc aagtgacata
gattttaatg
3121 taatgcataa tttaggtaag aaattaatta atgtagccta gtttattatc
ttgaaatgtt
3181 ttaccctatt tacttttaa aattaatgac ctaagcggag ggaataatta
taagtcaata
3241 gcagagagat tgttgtttgg gtgtttattt ttttcagttt ttgttttgag
agattgggtt
3301 aacacctcta gccaaaattg tttggtttta gggaggctaa caataaccta
ctgaatttgg
3361 aaaatgcaaa ggtaaaaaat gtatatagac tgcctgctga actggttaag
tactactgct
3421 tctgggaaat actatttcaa aattctatgt attataataa taaatttgta
agacattcat
3481 tattctacca tcctaatgaa aactttcaga agtctttctt tatccatggc
atgcccaggg

```
     3541 ttttacctga atctgataca ggatctatat aactttacta ggacttttga
ttgttgactc
     3601 caggcttagg tatatcagaa ggttcttttt gccatttggc ctgtggatgt
ctgagaagat
     3661 cattcacaat acatgtaaaa ttcaggtagg cctaaggaaa ggccagcctg
tagaaagcaa
     3721 aatggcagtg tctgttctcc actgttggag gcattatgta atttaagtat
cctgttagcc
     3781 actgtctttc tgctaattaa gtggggctga acaagtaagc actaataata
ccagtgaacc
     3841 acttgggcac cttgtgggta gagttttgct gccacctagt ggaatgggat
atcattgctt
     3901 ccatatcagg ttcacaagca agttaagtgg gcacagttta tttctgtgta
gctcaggctg
     3961 taatcttgaa agctgaggag atacccatgc ctctcagact cattagctgg
gtgtcacatt
     4021 accacctgca cattctgacc caccgcatct taatatgttt tgtcctcttg
gagaaactag
     4081 gagtagaagt caggatatgg taggtaaggg ggaaaaagga aagacggctt
gatagctatg
     4141 aatgcatgag gagcgaaatg ttgactcagt tatctagatc atggtctcca
aacctgatgc
     4201 tatttcctta caaaaatatt tgttgagcat gtgtccataa ttatatgtat
tgaacaatga
     4261 aaatatgtgt caacaaatgt actgctacac taatgtgaac attatggaac
aaaatttgaa
     4321 agagtgaaat aaaaggttta cactttcaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaa
```

LOCUS NP_001011553

ACCESSION NP_001011553

```
        1 msvsarsaaa eersvnsstm aqqknlegyv gfanlpnqvy rksvkrgfef
tlmvvgesgl
       61 gkstlinslf ltdlyspeyp gpshrikktv qveqskvlik eggvqlllti
vdtpgfgdav
      121 dnsncwqpvi dyidskfedy lnaesrvnrr qmpdnrvqcc lyfiapsghg
lkpldiefmk
      181 rlhekvniip liakadtltp eecqqfkkqi mkeiqehkik iyefpetdde
eenklvkkik
      241 drlplavvgs ntiievngkr vrgrqypwgv aevengehcd ftilrnmlir
thmqdlkdvt
      301 nnvhyenyrs rklaavtyng vdnnknkgql tksplaqmee errehvakmk
kmememeqvf
      361 emkvkekvqk lkdseaelqr rheqmkknle aqhkeleekr rqfedekanw
eaqqrileqq
      421 nssrtleknk kkgkif
```

Official Symbol: SH3BGRL and Name: SH3 domain binding glutamic acid-rich protein like [Homo sapiens]

Other Aliases: SH3BGR

Other Designations: SH3 domain-binding glutamic acid-rich-like protein; SH3-binding domain glutamic acid-rich protein like

LOCUS NM_003022 2090 bp mRNA linear PRI 27-JUN-2012

ACCESSION NM_003022

VERSION NM_003022.2 GI:211938420

```
        1 ttttagtctc cgcgtgaaaa ggtccttcat gctaatctaa ttcaattcag
ttctccttt
       61 tctttcttct tctcctcgcc ctcttctcag ggaaggaatc gtcaaaaatc
aatgtttcaa
      121 cagatctcgc ggtgctattc gagattccct attaaaaaaa aatagaattg
atgcaaacag
      181 cctgttcctt ccggggtttt gggctggaac tgcagcgctt agagagctcg
gtggaagctg
      241 ctaaaggcgg aggcggggct ctggcgagtt ctccttccac cttcccccac
ccttctctgc
      301 caaccgctgt ttcagcccct agctggattc cagccattgc tgcagctgct
ccacagccct
      361 tttcaggacc caaacaaccg cagccgctgt tcccaggatg gtgatccgtg
tatatattgc
      421 atcttcctct ggctctacag cgattaagaa gaaacaacaa gatgtgcttg
gtttcctaga
      481 agccaacaaa ataggatttg aagaaaaaga tattgcagcc aatgaagaga
atcggaagtg
      541 gatgagagaa aatgtacctg aaaatagtcg accagccaca ggttaccccc
tgccacctca
      601 gattttcaat gaaagccagt atcgcgggga ctatgatgcc ttctttgaag
ccagagaaaa
      661 taatgcagtg tatgccttct taggcttgac agccccacct ggttcaaagg
aagcagaagt
      721 gcaagcaaag cagcaagcat gaaccttaag cactgtgctt taagcatcct
gaaaaatgag
      781 tctccattgc ttttataaaa tagcagaatt agctttgctt caaaagaaat
aggcttaatg
      841 ttgaaataat agattagttg ggttttcaca tgcaaacatt caaaatgaat
acaaaattaa
      901 aatttgaaca ttatggtgat tatggtgagg agaatgggat attaacataa
aattatatta
      961 ataagtagat atcgtagaaa tagtgttgtt acctgccaag ccatcctgta
tacaccaatg
     1021 attttacaaa gaaaacaccc ttccctcctt ctgccattac tatggcaact
taagtgtatc
     1081 tgcagctcta cattaaaaag gagaaagaga aataacctgt ctctcattcc
taagttgcct
     1141 cattaatttt catgaacaag aatatgtacc tttttgatgc tatattactg
cgattaaaaa
     1201 gttcttgcag gtaatgttta tgatatgtta aacgttgtaa tttcctatcg
taattataac
     1261 attcccattc ttttgtagat gaaacttcta catattgaac cacagatttt
ctgagcttct
     1321 aaatgtagcc tttcattgca catttcagtg atcagaatag atatccttttt
acacgcacaa
     1381 aagcaataga ttcattcagt ggacaagttc cttgtttaac tacacagcta
tgatggaatg
     1441 atatatccaa gttccttgcc tcagtgaaat atgcatatgt atatcatgaa
agtgggatgc
```

```
     1501 caagtaagct taaaatggca ttctctagca aagagattag acttttaaat
aactcttata
     1561 aaacaggttg gcgatcattt cccaagattg gtttcccttg agtttttgct
aaaacaaatc
     1621 ttagtagttt tgcccgttta aaacaactca caatcgtaaa tgctactatt
cctaagatat
     1681 cttacctttt tatttcagtt tagccatgta ttgtatgagt gtattagtct
aagcagtgag
     1741 aatcttttct atgcctctat tccagcaaaa agtagaagta tcaaataaaa
agggcaactt
     1801 ttaaaatatt aagcctgaag acttctaaaa agacaagaaa catggcctaa
ataaccaaca
     1861 tagatttaca tagtaagttt cacactacct tattaccaaa agcaaacacc
tcttacttta
     1921 aactacatta tcatgtatat ctattgtatg ctggtcttta ctttttgcca
aaatcaacat
     1981 ataatgaaga gatgcctttg tttcatgaga ttcaaacttg atgctatgct
ttaaaataaa
     2041 ctcagtactt ttagaaacat aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
```

LOCUS NP_003013

ACCESSION NP_003013

VERSION NP_003013.1 GI:4506925

```
        1 mvirvyiass sgstaikkkq qdvlgflean kigfeekdia aneenrkwmr
envpensrpa
       61 tgyplppqif nesqyrgdyd affearenna vyaflgltap pgskeaevqa kqqa
```

Official Symbol: SNRPB and Name: small nuclear ribonucleoprotein polypeptides B and B1

Other Aliases: COD, SNRPB1, Sm-B/B', SmB/B', SmB/SmB', snRNP-B

Other Designations: B polypeptide of Sm protein; Sm protein B/B'; sm-B/Sm-B'; small nuclear ribonucleoprotein polypeptide B; small nuclear ribonucleoprotein polypeptides B and B'; small nuclear ribonucleoprotein-associated proteins B and B'

LOCUS NM_003091

ACCESSION NM_003091

VERSION NM_003091.3 GI:38149990

```
        1 aactccaggg ctagtgagct ggaccggaag taggtttcta cccgaccgca
ttttacgtgg
       61 tgctgcattt ccggtagcgg cggcgggaaa tcggctgtgg gagagaggct
aggcctctga
      121 ggaggcgaat ccggcgggta tcagagccat cagaaccgcc accatgacgg
tgggcaagag
      181 cagcaagatg ctgcagcata ttgattacag gatgaggtgc atcctgcagg
acggccggat
      241 cttcattggc accttcaagg cttttgacaa gcacatgaat ttgatcctct
gtgactgtga
```

```
      301 tgagttcaga aagatcaagc caaagaactc caaacaagca gaaagggaag
agaagcgagt
      361 cctcggtctg gtgctgctgc gaggggagaa tctggtctca atgacagtag
agggacctcc
      421 tcccaaagat actggtattg ctcgagttcc acttgctgga gctgccgggg
gcccagggat
      481 cggcagggct gctggcagag gaatcccagc tggggttccc atgccccagg
ctcctgcagg
      541 acttgctggg ccagtccgtg gggttggcgg gccatcccaa caggtgatga
ccccacaagg
      601 aagaggtact gttgcagccg ctgcagctgc tgccacagcc agtattgccg
gggctccaac
      661 ccagtaccca cctggccgtg ggggtcctcc cccacctatg ggccgaggag
cacccccctcc
      721 aggcatgatg ggcccacctc ctggtatgag acctcctatg ggtcccccaa
tggggatccc
      781 ccctggaaga gggactccaa tgggcatgcc ccctccggga atgcggcctc
ctccccctgg
      841 gatgcgaggc cttctttgac ccttggccac agagtatgga agtagctccg
cagaggcgtg
      901 ggctcgattc ctcagggcca cgttaccaca gacctgtttg tttcttatgc
tgttgttcgt
      961 ggagtctcat gggattgtct ggtttccctt acagggcccc ctccccgggg
aatgcgccca
     1021 ccaaggccct agactcatct tggccctcct cagctccctg cctgtttccc
gtaaggctgt
     1081 acatagtcct tttatctcct tgtggcctat gaaactggtt tataataaac
tcttaagaga
     1141 acattataat tgc
```

LOCUS NP_003082 231 aa linear PRI 27-JUN-2012

ACCESSION NP_003082

VERSION NP_003082.1 GI:4507125

```
        1 mtvgksskml qhidyrmrci lqdgrifigt fkafdkhmnl ilcdcdefrk
ikpknskqae
       61 reekrvlglv llrgenlvsm tvegpppkdt giarvplaga aggpgigraa
grgipagvpm
      121 pqapaglagp vrgvggpsqq vmtpqgrgtv aaaaaaatas iagaptqypp
grggpppppmg
      181 rgapppgmmg pppgmrppmg ppmgippgrg tpmgmpppgm rpppgmrgl l
```

Official Symbol: SOD1 and Name: superoxide dismutase 1, soluble

Other Aliases: ALS, ALS1, IPOA, SOD, hSod1, homodimer

Other Designations: Cu/Zn superoxide dismutase; SOD, soluble; indophenoloxidase A; superoxide dismutase [Cu-Zn]; superoxide dismutase, cystolic

LOCUS NM_000454

ACCESSION NM_000454

VERSION NM_000454.4 GI:48762945

```
        1 gtttggggcc agagtgggcg aggcgcggag gtctggccta taaagtagtc
gcggagacgg
       61 ggtgctggtt tgcgtcgtag tctcctgcag cgtctggggt ttccgttgca
gtcctcggaa
      121 ccaggacctc ggcgtggcct agcgagttat ggcgacgaag gccgtgtgcg
tgctgaaggg
      181 cgacggccca gtgcagggca tcatcaattt cgagcagaag gaaagtaatg
gaccagtgaa
      241 ggtgtgggga agcattaaag gactgactga aggcctgcat ggattccatg
ttcatgagtt
      301 tggagataat acagcaggct gtaccagtgc aggtcctcac tttaatcctc
tatccagaaa
      361 acacggtggg ccaaaggatg aagagaggca tgttggagac ttgggcaatg
tgactgctga
      421 caaagatggt gtggccgatg tgtctattga agattctgtg atctcactct
caggagacca
      481 ttgcatcatt ggccgcacac tggtggtcca tgaaaaagca gatgacttgg
gcaaaggtgg
      541 aaatgaagaa agtacaaaga caggaaacgc tggaagtcgt ttggcttgtg
gtgtaattgg
      601 gatcgcccaa taaacattcc cttggatgta gtctgaggcc ccttaactca
tctgttatcc
      661 tgctagctgt agaaatgtat cctgataaac attaaacact gtaatcttaa
aagtgtaatt
      721 gtgtgacttt ttcagagttg ctttaaagta cctgtagtga gaaactgatt
tatgatcact
      781 tggaagattt gtatagtttt ataaaactca gttaaaatgt ctgtttcaat
gacctgtatt
      841 ttgccagact taaatcacag atgggtatta aacttgtcag aatttctttg
tcattcaagc
      901 ctgtgaataa aaaccctgta tggcacttat tatgaggcta ttaaaagaat
ccaaattcaa
      961 actaaaaaaa aaaaaaaaaa a
```

LOCUS NP_000445

ACCESSION NP_000445

VERSION NP_000445.1 GI:4507149

```
        1 matkavcvlk gdgpvqgiin feqkesngpv kvwgsikglt eglhgfhvhe
fgdntagcts
       61 agphfnplsr khggpkdeer hvgdlgnvta dkdgvadvsi edsvislsgd
hciigrtlvv
      121 hekaddlgkg gneestktgn agsrlacgvi giaq
```

KARS

Official Symbol: **KARS**

Official Name: lysyl-tRNA synthetase

Gene ID: 3735

Organism: Homo sapiens

Other Aliases: CMTRIB, KARS2, KRS

Other Designations: lysRS; lysine tRNA ligase; lysine--tRNA ligase

Nucleotide sequence:

NCBI Reference Sequence: NM_001130089.1

LOCUS: NM_001130089

ACCESSION : NM_001130089

```
        1 aaattaacgt actatcctcc ttacttttgg gtcgggccct ccgggaagat
ggcggccgtg
       61 caggcggccg aggtgaaagt ggatggcagc gagccgaaac tgagcaagaa
gtggtggtaa
      121 tcattagttc cagggtgctc tgccatgttg acgcaagctg ctgtaaggct
tgttaggggg
      181 tccctgcgca aaacctcctg ggcagagtgg ggtcacaggg aactgcgact
gggtcaactt
      241 gctcctttca cagcgcctca caaggacaag tcattttctg atcaaagaag
tgagctgaag
      301 agacgcctga aagctgagaa gaaagtagca gagaaggagg ccaaacagaa
agagctcagt
      361 gagaaacagc taagccaagc cactgctgct gccaccaacc acaccactga
taatggtgtg
      421 ggtcctgagg aagagagcgt ggacccaaat caatactaca aaatccgcag
tcaagcaatt
      481 catcagctga aggtcaatgg ggaagaccca tacccacaca agttccatgt
agacatctca
      541 ctcactgact tcatccaaaa atatagtcac ctgcagcctg gggatcacct
gactgacatc
      601 accttaaagg tggcaggtag gatccatgcc aaaagagctt ctgggggaaa
gctcatcttc
      661 tatgatcttc gaggagaggg ggtgaagttg caagtcatgg ccaattccag
aaattataaa
      721 tcagaagaag aatttattca tattaataac aaactgcgtc ggggagacat
aattggagtt
      781 caggggaatc ctggtaaaac caagaagggt gagctgagca tcattccgta
tgagatcaca
      841 ctgctgtctc cctgtttgca tatgttacct catcttcact ttggcctcaa
agacaaggaa
      901 acaaggtatc gccagagata cttggacttg atcctgaatg actttgtgag
gcagaaattt
      961 atcatccgct ctaagatcat cacatatata agaagtttct tagatgagct
gggattccta
     1021 gagattgaaa ctcccatgat gaacatcatc ccaggggggag ccgtggccaa
gcctttcatc
     1081 acttatcaca acgagctgga catgaactta tatatgagaa ttgctccaga
actctatcat
     1141 aagatgcttg tggttggtgg catcgaccgg gtttatgaaa ttggacgcca
gttccggaat
     1201 gaggggattg atttgacgca caatcctgag ttcaccacct gtgagttcta
catggcctat
     1261 gcagactatc acgatctcat ggaaatcacg gagaagatgg tttcagggat
ggtgaagcat
```

```
     1321 attacaggca gttacaaggt cacctaccac ccagatggcc cagagggcca
agcctacgat
     1381 gttgacttca ccccaccctt ccggcgaatc aacatggtag aagagcttga
gaaagccctg
     1441 gggatgaagc tgccagaaac gaacctcttt gaaactgaag aaactcgcaa
aattcttgat
     1501 gatatctgtg tggcaaaagc tgttgaatgc cctccacctc ggaccacagc
caggctcctt
     1561 gacaagcttg ttggggagtt cctggaagtg acttgcatca atcctacatt
catctgtgat
     1621 cacccacaga taatgagccc tttggctaaa tggcaccgct ctaaagaggg
tctgactgag
     1681 cgctttgagc tgtttgtcat gaagaaagag atatgcaatg cgtatactga
gctgaatgat
     1741 cccatgcggc agcggcagct ttttgaagaa caggccaagg ccaaggctgc
aggtgatgat
     1801 gaggccatgt tcatagatga aaacttctgt actgccctgg aatatgggct
gccccccaca
     1861 gctggctggg gcatgggcat tgatcgagtc gccatgtttc tcacggactc
caacaacatc
     1921 aaggaagtac ttctgtttcc tgccatgaaa cccgaagaca agaaggagaa
tgtagcaacc
     1981 actgatacac tggaaagcac aacagttggc acttctgtct agaaaataat
aattgcaagt
     2041 tgtataactc aggcgtcttt gcatttctgc gaaagatcaa ggtctgcaag
ggaattcttg
     2101 tgtgctgctt tccatttgac accgcagttc tgttcagcca tcagaagaga
gacaaggaat
     2161 taaaaatttc tttttaatcc tgttaccaaa taaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001123561.1

LOCUS: NP_001123561

ACCESSION: NP_001123561

```
      1 mltqaavrlv rgslrktswa ewghrelrlg qlapftaphk dksfsdqrse
lkrrlkaekk
       61 vaekeakqke lsekqlsqat aaatnhttdn gvgpeeesvd pnqyykirsq
aihqlkvnge
      121 dpyphkfhvd isltdfiqky shlqpgdhlt ditlkvagri hakrasggkl
ifydlrgegv
      181 klqvmansrn ykseeefihi nnklrrgdii gvqgnpgktk kgelsiipye
itllspclhm
      241 lphlhfglkd ketryrqryl dlilndfvrq kfiirskiit yirsfldelg
fleietpmmn
      301 iipggavakp fityhneldm nlymriapel yhkmlvvggi drvyeigrqf
rnegidlthn
      361 pefttcefym ayadyhdlme itekmvsgmv khitgsykvt yhpdgpegqa
ydvdftppfr
      421 rinmveelek algmklpetn lfeteetrki lddicvakav ecpppprttar
lldklvgefl
```

```
      481 evtcinptfi cdhpqimspl akwhrskegl terfelfvmk keicnaytel
ndpmrqrqlf
      541 eeqakakaag ddeamfiden fctaleyglp ptagwgmgid rvamfltdsn
nikevllfpa
      601 mkpedkkenv attdtlestt vgtsv
      //
```

KIF5B

Official Symbol: **KIF5B**

Official Name: kinesin family member 5B

Gene ID: 3799

Organism: Homo sapiens

Other Aliases: KINH, KNS, KNS1, UKHC

Other Designations: conventional kinesin heavy chain; kinesin 1 (110-120kD); kinesin heavy chain; kinesin-1 heavy chain; ubiquitous kinesin heavy chain

Nucleotide sequence:

NCBI Reference Sequence: NM_004521.2

LOCUS: NM_004521

ACCESSION : NM_004521

```
      1 ctcctcccgc accgccctgt cgcccaacgg cggcctcagg agtgatcggg
cagcagtcgg
       61 ccggccagcg gacggcagag cgggcggacg ggtaggcccg gcctgctctt
cgcgaggagg
      121 aagaaggtgg ccactctccc ggtccccaga acctccccag cccccgcagt
ccgcccagac
      181 cgtaaagggg gacgctgagg agccgcggac gctctccccg gtgccgccgc
cgctgccgcc
      241 gccatggctg ccatgatgga tcggaagtga gcattagggt taacggctgc
cggcgccggc
      301 tcttcaagtc ccggctcccc ggccgcctcc acccggggaa gcgcagcgcg
gcgcagctga
      361 ctgctgcctc tcacggccct cgcgaccaca agccctcagg tccggcgcgt
tccctgcaag
      421 actgagcggc ggggagtggc tcccggccgc cggccccggc tgcgagaaag
atggcggacc
      481 tggccgagtg caacatcaaa gtgatgtgtc gcttcagacc tctcaacgag
tctgaagtga
      541 accgcggcga caagtacatc gccaagtttc agggagaaga cacggtcgtg
atcgcgtcca
      601 agccttatgc atttgatcgg gtgttccagt caagcacatc tcaagagcaa
gtgtataatg
      661 actgtgcaaa gaagattgtt aaagatgtac ttgaaggata taatggaaca
atatttgcat
```

721 atggacaaac atcctctggg aagacacaca caatggaggg taaacttcat
gatccagaag
781 gcatgggaat tattccaaga atagtgcaag atatttttaa ttatatttac
tccatggatg
841 aaaatttgga atttcatatt aaggtttcat attttgaaat atatttggat
aagataaggg
901 acctgttaga tgtttcaaag accaaccttt cagttcatga agacaaaaac
cgagttccct
961 atgtaaaggg gtgcacagag cgttttgtat gtagtccaga tgaagttatg
gataccatag
1021 atgaaggaaa atccaacaga catgtagcag ttacaaatat gaatgaacat
agctctagga
1081 gtcacagtat atttcttatt aatgtcaaac aagagaacac acaaacggaa
caaaagctga
1141 gtggaaaact ttatctggtt gatttagctg gtagtgaaaa ggttagtaaa
actggagctg
1201 aaggtgctgt gctggatgaa gctaaaaaca tcaacaagtc actttctgct
cttggaaatg
1261 ttatttctgc tttggctgag ggtagtacat atgttccata tcgagatagt
aaaatgacaa
1321 gaatccttca agattcatta ggtggcaact gtagaaccac tattgtaatt
tgctgctctc
1381 catcatcata caatgagtct gaaacaaaat ctacactctt atttggccaa
agggccaaaa
1441 caattaagaa cacagtttgt gtcaatgtgg agttaactgc agaacagtgg
aaaaagaagt
1501 atgaaaaaga aaaagaaaaa aataagatcc tgcggaacac tattcagtgg
cttgaaaatg
1561 agctcaacag atggcgtaat ggggagacgg tgcctattga tgaacagttt
gacaaagaga
1621 aagccaactt ggaagctttc acagtggata aagatattac tcttaccaat
gataaaccag
1681 caaccgcaat tggagttata ggaaatttta ctgatgctga aagaagaaag
tgtgaagaag
1741 aaattgctaa attatacaaa cagcttgatg acaaggatga agaaattaac
cagcaaagtc
1801 aactggtaga gaaactgaag acgcaaatgt tggatcagga ggagcttttg
gcatctacca
1861 gaagggatca agacaatatg caagctgagc tgaatcgcct tcaagcagaa
aatgatgcct
1921 ctaaagaaga agtgaaagaa gttttacagg ccctagaaga acttgctgtc
aattatgatc
1981 agaagtctca ggaagttgaa gacaaaacta aggaatatga attgcttagt
gatgaattga
2041 atcagaaatc ggcaacttta gcgagtatag atgctgagct tcagaaactt
aaggaaatga
2101 ccaaccacca gaaaaaacga gcagctgaga tgatggcatc tttactaaaa
gaccttgcag
2161 aaataggaat tgctgtggga aataatgatg taaagcagcc tgagggaact
ggcatgatag
2221 atgaagagtt cactgttgca agactctaca ttagcaaaat gaagtcagaa
gtaaaaacca
2281 tggtgaaacg ttgcaagcag ttagaaagca cacaaactga gagcaacaaa
aaaatggaag
2341 aaaatgaaaa ggagttagca gcatgtcagc ttcgtatctc tcaacatgaa
gccaaaatca
2401 agtcattgac tgaatacctt caaaatgtgg aacaaaagaa aagacagttg
gaggaatctg
2461 tcgatgccct cagtgaagaa ctagtccagc ttcgagcaca agagaaagtc
catgaaatgg

2521 aaaaggagca cttaaataag gttcagactg caaatgaagt taagcaagct
gttgaacagc
2581 agatccagag ccatagagaa actcatcaaa aacagatcag tagtttgaga
gatgaagtag
2641 aagcaaaagc aaaacttatt actgatcttc aagaccaaaa ccagaaaatg
atgttagagc
2701 aggaacgtct aagagtagaa catgagaagt tgaaagccac agatcaggaa
aagagcagaa
2761 aactacatga acttacggtt atgcaagata gacgagaaca agcaagacaa
gacttgaagg
2821 gtttggaaga gacagtggca aaagaacttc agactttaca caacctgcgc
aaactctttg
2881 ttcaggacct ggctacaaga gttaaaaaga gtgctgagat tgattctgat
gacaccggag
2941 gcagcgctgc tcagaagcaa aaaatctcct ttcttgaaaa taatcttgaa
cagctcacta
3001 aagtgcacaa acagttggta cgtgataatg cagatctccg ctgtgaactt
cctaagttgg
3061 aaaagcgact tcgagctaca gctgagagag tgaaagcttt ggaatcagca
ctgaaagaag
3121 ctaaagaaaa tgcatctcgt gatcgcaaac gctatcagca agaagtagat
cgcataaagg
3181 aagcagtcag gtcaaagaat atggccagaa gagggcattc tgcacagatt
gctaaaccta
3241 ttcgtcccgg gcaacatcca gcagcttctc caactcaccc aagtgcaatt
cgtggaggag
3301 gtgcatttgt tcagaacagc cagccagtgg cagtgcgagg tggaggaggc
aaacaagtgt
3361 aatcgtttat acatacccac aggtgttaaa aagtaatcga agtacgaaga
ggacatggta
3421 tcaagcagtc attcaatgac tataacctct actcccttgg gattgtagaa
ttataacttt
3481 taaaaaaaat gtataaatta tacctggcct gtacagctgt ttcctaccta
ctcttcttgt
3541 aaactctgct gcttcccaac acaactagag tgcaattttg gcatcttagg
agggaaaaag
3601 gacagtttac aactgtggcc ctatttatta cacagtttgt ctatcgtgtc
ttaaatttag
3661 tctttactgt gccaagctaa ctgtacctta taggactgta ctttttgtat
tttttgtgta
3721 tgtttatttt ttaatctcag tttaaattac ctagctgcta ctgcttcttg
ttttttcttt
3781 cctattaaaa cgtcttcctt tttttttctt aagagaaaat ggaacattta
ggttaaatgt
3841 ctttaaattt taccacttaa caacactaca tgcccataaa atatatccag
tcagtactgt
3901 attttaaaat cccttgaaat gatgatatca gggttaaaat tacttgtatt
gtttctgaag
3961 tttgctcctg aaaactactg tttgagcact gaaacgttac aaatgcctaa
taggcatttg
4021 agactgagca aggctacttg ttatctcatg aaatgcctgt tgccgagtta
ttttgaatag
4081 aaatatttta aagtatcaaa agcagatctt agtttaaggg agtttggaaa
aggaattata
4141 tttctctttt tcctgattct gtactcaaca agtcttgatg gaattaaaat
actctgcttt
4201 attctggtga gcctgctagc taatataagt attggacagg taataatttg
tcatctttaa
4261 tattagtaaa atgaattaag atattatagg attaaacata attttatacg
gttagtactt

```
     4321 tattggccga cctaaattta tagcgtgtgg aaattgagaa aaatgaagaa
acaggacaga
     4381 tatatgatga attaaaaata tatataggtc aattttggtc tgaaatccct
gaggtgtttt
     4441 taacctgcta cactaatttg tacactaatt tatttcttta gtctagaaat
agtaaattgt
     4501 ttgcaagtca ctaataatca ttagataaat tattttcttg gccatagccg
ataattttgt
     4561 aatcagtact aagtgtatac gtatttttgc cacttttttcc tcagatgatt
aaagtaagtc
     4621 aacagcttat tttaggaaac tgtaaaagta atagggaaag agatttcact
atttgcttca
     4681 tcagtggtag gggggcggtg actgcaactg tgttagcaga aattcacaga
gaatggggat
     4741 ttaaggttag cagagaaact tggaaagttc tgtgttagga tcttgctggc
agaattaact
     4801 ttttgcaaaa gttttataca cagatatttg tattaaattt ggagccatag
tcagaagact
     4861 cagatcataa ttggcttatt tttctatttc cgtaactatt gtaatttcca
cttttgtaat
     4921 aattttgatt taaaatataa atttatttat ttatttttttt aatagtcaaa
aatctttgct
     4981 gttgtagtct gcaacctcta aaatgattgt gttgctttta ggattgatca
gaagaaacac
     5041 tccaaaaatt gagatgaaat gttggtgcag ccagttataa gtaatatagt
taacaagcaa
     5101 aaaaagtgct gccacctttt atgatgattt tctaaatgga gaaacatttg
gctgcatcca
     5161 catagacctt tatgttttgt tttcagttga aaacttgcct cctttggcaa
cattcgtaaa
     5221 tgaagcagaa tttttttttc tctttttttcc aaatatgtta gttttgttct
tgtaagatgt
     5281 atcatgggta ttggtgctgt gtaatgaaca acgaatttta attagcatgt
ggttcagaat
     5341 atacaatgtt aggtttttaa aaagtatctt gatggttctt ttctatttat
aatttcagac
     5401 tttcataaag tgtaccaaga atttcataaa tttgttttca gtgaactgct
ttttgctatg
     5461 gtaggtcatt aaacacagca cttactctta aaaatgaaaa tttctgatca
tctaggatat
     5521 tgacacattt caatttgcag tgtctttttg actggatata ttaacgttcc
tctgaatggc
     5581 attgatagat ggttcagaag agaaactcaa tgaaataaag agaatattta
ttcatggcga
     5641 ttaattaaat tatttgccta acttaagaaa actactgtgc gtaactctca
gtttgtgctt
     5701 aactccattt gacatgaggt gacagaagag agtctgagtc tacctgtgga
atatgttggt
     5761 ttattttcag tgcttgaaga tacattcaca aatacttggt ttgggaagac
accgtttaat
     5821 tttaagttaa cttgcatgtt gtaaatgcgt tttatgttta aataaagagg
aaaatttttt
     5881 gaaatgtaaa aaaaaaaaaa aaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_004512.1

LOCUS: NP_004512

ACCESSION: NP_004512

```
      1 madlaecnik vmcrfrplne sevnrgdkyi akfqgedtvv iaskpyafdr
vfqsstsqeq
       61 vyndcakkiv kdvlegyngt ifaygqtssg kthtmegklh dpegmgiipr
ivqdifnyiy
      121 smdenlefhi kvsyfeiyld kirdlldvsk tnlsvhedkn rvpyvkgcte
rfvcspdevm
      181 dtidegksnr hvavtnmneh ssrshsifli nvkqentqte qklsgklylv
dlagsekvsk
      241 tgaegavlde akninkslsa lgnvisalae gstyvpyrds kmtrilqdsl
ggncrttivi
      301 ccspssynes etkstllfgq raktikntvc vnveltaeqw kkkyekekek
nkilrntiqw
      361 lenelnrwrn getvpideqf dkekanleaf tvdkditltn dkpataigvi
gnftdaerrk
      421 ceeeiaklyk qlddkdeein qqsqlveklk tqmldqeell astrrdqdnm
qaelnrlqae
      481 ndaskeevke vlqaleelav nydqksqeve dktkeyells delnqksatl
asidaelqkl
      541 kemtnhqkkr aaemmasllk dlaeigiavg nndvkqpegt gmideeftva
rlyiskmkse
      601 vktmvkrckq lestqtesnk kmeenekela acqlrisqhe akikslteyl
qnveqkkrql
      661 eesvdalsee lvqlraqekv hemekehlnk vqtanevkqa veqqiqshre
thqkqisslr
      721 deveakakli tdlqdqnqkm mleqerlrve heklkatdqe ksrklheltv
mqdrreqarq
      781 dlkgleetva kelqtlhnlr klfvqdlatr vkksaeidsd dtggsaaqkq
kisflennle
      841 qltkvhkqlv rdnadlrcel pklekrlrat aervkalesa lkeakenasr
drkryqqevd
      901 rikeavrskn marrghsaqi akpirpgqhp aaspthpsai rgggafvqns
qpvavrgggg
      961 kqv
      //
```

KPNA3

Official Symbol: **KPNA3**

Official Name: karyopherin alpha 3 (importin alpha 4)

Gene ID: 3839

Organism: Homo sapiens

Other Aliases: RP11-432M24.3, IPOA4, SRP1, SRP1gamma, SRP4, hSRP1

Other Designations: SRP1-gamma; importin alpha 4; importin alpha Q2; importin alpha-3; importin subunit alpha-3; importin-alpha-Q2; karyopherin subunit alpha-3; qip2

Nucleotide sequence:

NCBI Reference Sequence: NM_002267.3

LOCUS: NM_002267

ACCESSION : NM_002267

```
      1 gccccgcgcc tgaggggcag taaaagtcgc caggtccggc tccatttctg
gcacaaaact
       61 tgcagcaccg aggggttgtg gagagccctt gcaggggaag agggcagggt
catcccgaga
      121 accaacgggc acgtatagcc cggcgaacgc ccaagccggt caccgccccc
ggtcacgtgt
      181 cgccagcctc cgcggccgcg cgccgctctc agcaccgttc ccgccccacc
cggcccggca
      241 gtcggcccgc gcctcccccg gcgctactgc cacctcgcgc tcggaggcgt
cacagaacgt
      301 gctcttctct cccctccccc ctcccgctct ccccctcctc cccctcccgc
tccaagattc
      361 gccgccgccg ccgccgcagc cgcaggagta gccgccgccg gagccgcgcg
cagccatggc
      421 cgagaacccc agcttggaga accaccgcat caagagcttc aagaacaagg
gccgcgatgt
      481 ggaaacaatg cgaagacata gaaatgaagt gacagtggaa ctgcggaaga
acaaaagaga
      541 tgaacactta ttgaaaaaga gaaatgttcc ccaagaagaa agtctagaag
attcagatgt
      601 tgatgctgat tttaaagcac aaaatgtaac cctagaagct atattgcaga
atgccacaag
      661 tgataaccca gtggtccaat tgagtgctgt ccaggcagca agaaaactgt
tatccagtga
      721 cagaaatcca ccgattgatg acttaataaa atctgggatt ttaccaattc
tagtcaaatg
      781 tctagaaagg gatgataatc cttcattaca gtttgaagct gcttgggcat
taactaacat
      841 agcatcagga acttctgcac agactcaagc tgttgtgcag tctaatgcag
tacctctttt
      901 tctgagactt cttcgttcac cacatcagaa tgtttgtgaa caagcagtat
gggctttggg
      961 aaacattata ggtgatggtc ctcaatgtag agattatgtc atatcactgg
gagttgtcaa
     1021 acctcttctg tccttcatca gtccctccat ccccatcacc ttccttcgga
acgtcacatg
     1081 ggtcattgtc aatctctgca ggaataagga tcccccgccg cctatggaga
cagttcagga
     1141 gattttgcca gctttatgtg tcctcatata ccatacagat ataaacattc
ttgtagacac
     1201 tgtttgggct ctgtcatact tgacagatgg aggtaatgaa cagatacaga
tggttattga
     1261 ttcaggagtt gtgccctttc ttgtgcccct tctgagccat caggaagtca
aagttcaaac
     1321 agcagccctc agagcagttg gcaacatagt gactggcacc gacgagcaga
cccaggttgt
     1381 tctcaattgt gatgtcctgt cacacttccc aaatctctta tcacacccaa
aagagaagat
     1441 aaataaggaa gcagtgtggt tcctttccaa cataacagca ggcaaccagc
aacaagttca
     1501 agctgtaata gatgctggat taattcctat gataattcat cagcttgcta
agggggactt
```

1561 tggaacacaa aaagaagctg cttgggcaat cagcaactta acaataagtg
gcagaaaaga
1621 tcaggttgag taccttgtac agcagaatgt aataccaccg ttctgtaatt
tactgtcagt
1681 gaaagattct caagtggttc aggtggttct agatggtcta aaaaacattc
tgataatggc
1741 cggtgatgaa gcaagcacaa tagctgaaat aatagaggaa tgtggaggtt
tggagaaaat
1801 tgaagtttta cagcaacatg aaaatgaaga catatataaa ttagcatttg
aaatcataga
1861 tcagtatttc tctggtgatg atattgatga agatccctgc ctcattcctg
aagcaacaca
1921 aggaggtacc tacaattttg atccaacagc caaccttcaa acaaaagaat
ttaattttta
1981 aattcagttg agtgcagcat ctttcccaca ttcaatatga agcaccacca
gatggctacc
2041 aaatgataag aacaacagca acaaaaggct ccaaaacaca catgcctctt
tgttttgatg
2101 cttctaaagc aagccatgtc tcagtcactt tgcagttgcc aaaagtcact
atcacatgga
2161 ctgtaaatgc atatgcatga tttcctaaac tgttttagaa ctctccttaa
caatctcaac
2221 taccctattt ttccctgttc cctggtgcca caggctgaca actgcagtct
ccagtttaga
2281 ataaatattc catagtggtg acatgtcagc tgcccactga tactcctttg
gaaaatggtg
2341 cgctgtggat caagacactt tggtatgatg catatacaag ttggaagact
aaagaggtgc
2401 agtgtgatct gagcctccat cattgtcctc cacaaacata ttttcatatt
ctttatgtgg
2461 aagaatagat tttaaagtac aagccaaatg attttcattg gtggaactga
cacaaaaaaa
2521 gtaacttaaa aacaagaaac ttggttattg aataaacaga taagtttaaa
aaaaaaaaaa
2581 aactacttca tctaccagta attgatgtgt ttattatctg cctcagaagc
cagggttgga
2641 ggaagaactt tagatatgga tattaatgct tttgccatta tacctaattt
ttgagaacag
2701 caagccctat ttgaccactc tcttcagcct gtgtgttcct gctgttttga
agtaatcaaa
2761 tgctgtgcat ggtattttac ctgagctgca acctgttatg gacttgaact
tctgtttaag
2821 ttgaaagcaa gagtccctga gtataaagga aaaacagcaa aacaaaaagc
aaacaaaaaa
2881 aaactgcaaa agtctaaaat acccattggt gatgtttttt aaaaaaatct
tgctttcagc
2941 tttcaggagt taatattctt tgttttaatt tgataattgg atatggttga
tttatattgg
3001 gtttaaactg tggagctttc atgtttactg taatttagtc ttaaaatatt
ttttacttag
3061 taaccagtgc ttttgataat gtggttggca acaaaccagc aactatttag
aagtgtcata
3121 agagttcatt ctttgagtat tgggaaagtt aattcagatc ctactcaaaa
agcatcttca
3181 catattaaaa gattcagaca gggatctgtg tagaggagta atttgcagtt
atttaacata
3241 aacctgattt gcagtgatct ctaagtaatt ctgcaaaatc cggtattact
atgtcaagtt
3301 attgcttttg gtaaattgtc tgacccagtt attaatgaaa gaatatggat
ttaaaaattt

```
     3361 ttaaactaaa taatttgtgc tgtcacagaa atggtattgt tgctcttgtt
tactgggtat
     3421 aatttcccaa tgcattgatg tgaagggata gaaaatctaa actaatttag
ttatccattg
     3481 gggggtgtat ttactgtgat gaagatgaga cagatgccat cagagctttg
tgaatcagct
     3541 ggggtgtttt cactgataaa caacacatag caggtgtgca ttcattacaa
atatatgtat
     3601 ctgccaaggt ggagccactt taaagagtga gttttgtctt gtatctaaag
tggatacaag
     3661 cgtatgttta aactgcaaga tttttacttg ctagagaatc tgttttaata
tagtggtttg
     3721 gcctctgatt atttataggt tttataaatt ttagaatcaa tttctcttta
aggtggctca
     3781 gatttttcaa ctcttgtgca cataaaattg agttgaagtt cattgtgcct
ttttttcttt
     3841 atccaaattt tgagttaaag cttcatatgg taactgcatc ctgttcggac
actatagtct
     3901 aaatttttga aactgtgtgg tgttcgctaa aagtaggaat aacaacgtaa
aagctaatta
     3961 aggtcacaaa cttcggtgaa acccttaaaa gtccaaatct tcttgatatt
gtgaaccgta
     4021 ccccttccag tttagtttct tctggacttt ccttacttaa ctgacagtta
ccttttaaaa
     4081 tttgcacaca ttatgattaa aattgggcct ctactgtgat gattcctatt
tcctctcatg
     4141 ttttaaagtg caaactaaca tttaagtgaa cattagcatc aagtagtgca
gacacttgta
     4201 tgcatttcct tgattcaatt tgtgacctta ccagttttga attggaattg
caccatttcg
     4261 tagataaagg aaactaagta tattgctgca cttttaagtt ttcaaaacag
tgtttaaaaa
     4321 ttgcattgtt attttttttt aaactcagtt taaaaagact aaaacgttct
ttcaaaagag
     4381 gcatctaaat gtgttcctaa ttttgtatat gggcttaggt tttgtaacca
ataaaaaaag
     4441 ctgctatcaa atatgataaa acattgaaaa cttaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_002258.2

LOCUS: NP_002258

ACCESSION: NP_002258

```
      1 maenpslenh riksfknkgr dvetmrrhrn evtvelrknk rdehllkkrn
vpqeesleds
       61 dvdadfkaqn vtleailqna tsdnpvvqls avqaarklls sdrnppiddl
iksgilpilv
      121 kclerddnps lqfeaawalt niasgtsaqt qavvqsnavp lflrllrsph
qnvceqavwa
      181 lgniigdgpq crdyvislgv vkpllsfisp sipitflrnv twvivnlcrn
kdppppmetv
      241 qeilpalcvl iyhtdinilv dtvwalsylt dggneqiqmv idsgvvpflv
pllshqevkv
```

```
      301 qtaalravgn ivtgtdeqtq vvlncdvlsh fpnllshpke kinkeavwfl
snitagnqqq
      361 vqavidagli pmiihqlakg dfgtqkeaaw aisnltisgr kdqveylvqq
nvippfcnll
      421 svkdsqvvqv vldglknili magdeastia eiieecggle kievlqqhen
ediyklafei
      481 idqyfsgddi dedpclipea tqggtynfdp tanlqtkefn f
      //
```

LGALS1

Official Symbol: **LGALS1**

Official Name: lectin, galactoside-binding, soluble, 1

Gene ID: 3956

Organism: Homo sapiens

Other Aliases: GAL1, GBP

Other Designations: 14 kDa laminin-binding protein; 14 kDa lectin; HBL; HLBP14; HPL; S-Lac lectin 1; beta-galactoside-binding lectin L-14-I; beta-galactoside-binding protein 14kDa; gal-1; galaptin; galectin 1; galectin-1; lactose-binding lectin 1; putative MAPK-activating protein PM12

Nucleotide sequence:

NCBI Reference Sequence: NM_002305.3

LOCUS: NM_002305

ACCESSION : NM_002305

```
      1 agttaaaagg gtgggagcgt ccgggggccc atctctctcg ggtggagtct
tctgacagct
       61 ggtgcgcctg cccgggaaca tcctcctgga ctcaatcatg gcttgtggtc
tggtcgccag
      121 caacctgaat ctcaaacctg gagagtgcct tcgagtgcga ggcgaggtgg
ctcctgacgc
      181 taagagcttc gtgctgaacc tgggcaaaga cagcaacaac ctgtgcctgc
acttcaaccc
      241 tcgcttcaac gcccacggcg acgccaacac catcgtgtgc aacagcaagg
acggcggggc
      301 ctgggggacc gagcagcggg aggctgtctt tcccttccag cctggaagtg
ttgcagaggt
      361 gtgcatcacc ttcgaccagg ccaacctgac cgtcaagctg ccagatggat
acgaattcaa
      421 gttccccaac cgcctcaacc tggaggccat caactacatg gcagctgacg
gtgacttcaa
      481 gatcaaatgt gtggcctttg actgaaatca gccagcccat ggcccccaat
aaaggcagct
      541 gcctctgctc cctctgaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_002296.1

LOCUS: NP_002296

ACCESSION: NP_002296

```
      1 macglvasnl nlkpgeclrv rgevapdaks fvlnlgkdsn nlclhfnprf
nahgdantiv
       61 cnskdggawg teqreavfpf qpgsvaevci tfdqanltvk lpdgyefkfp
nrlnleainy
      121 maadgdfkik cvafd
      //
```

MACF1

Official Symbol: **MACF1**

Official Name: microtubule-actin crosslinking factor 1

Gene ID: 23499

Organism: Homo sapiens

Other Aliases: ABP620, ACF7, MACF, OFC4

Other Designations: 620 kDa actin binding protein; actin cross-linking family protein 7; macrophin 1; microtubule-actin cross-linking factor 1; trabeculin-alpha

Nucleotide sequence:

NCBI Reference Sequence: NM_012090.4

LOCUS: NM_012090

ACCESSION : NM_012090 NM_033024

```
      1 attgtgggag ccgctcccct cggctccgcc acgctcccct cgactgcgct
ccagcctggg
       61 gcgcgcccgg ccgccgccgc cttcgctgcc gccacgggcc cgtcttcttc
ctccttcggc
      121 tcccaggatg aagaaactga gtctcagaga ggtgaagtga cttgcccaag
atcacagcaa
      181 ttatcacttc tccctgggct cccaggccct cctgcagcag cccccgcctg
ggccatgtct
```

241 tcctcagatg aagagacgct cagtgagcgg tcatgtcgga gtgagcggtc
ttgtcggagt
301 gagcgatctt acaggagcga gcggtcgggg agcctgtctc cctgtccccc
aggggacacc
361 ttgccctgga acctgccact gcatgagcag aaaaagcgga aaagccagga
ttcggtgctg
421 gaccctgcag agcgtgctgt ggtcagagtc gctgatgaac gggaccgggt
tcagaagaaa
481 acgttcacca agtgggtcaa caagcactta atgaaggtcc gcaagcacat
caatgatctt
541 tatgaagatc tgcgggatgg ccataacctg atctctctgt tggaggtcct
ctcaggcatc
601 aaactgcccc gggagaaggg caggatgcgt tttcataggc tgcagaatgt
gcagattgcc
661 ctggacttcc taaagcagcg acaggtgaaa ctagtgaata ttcgcaatga
tgacatcaca
721 gatggcaacc ccaagttgac cctgggtctg atctggacca ttattttgca
tttccagatc
781 tctgacatct acattagtgg agaatcaggg gatatgtcag ccaaggagaa
actactcctg
841 tggacccaga aggtgacagc tggttacaca ggaatcaaat gcaccaactt
ttcctcctgc
901 tggagtgatg ggaagatgtt caatgcactc attcaccgat accgacccga
tctagtagac
961 atggagaggg tgcaaatcca aagtaaccga gagaatctgg aacaggcttt
tgaagtggca
1021 gaaagactgg gggtcactcg cctgctggat gcagaagatg tggatgtgcc
atctccagat
1081 gaaaagtctg taatcactta tgtgtcttcg atttatgatg ccttccctaa
agttcctgag
1141 ggtggagaag ggatcagtgc tacggaagtg gactccaggt ggcaagaata
ccaaagccga
1201 gtggactccc tcattccctg gatcaaacag catacaatac tgatgtcaga
taaaactttt
1261 ccccaaaacc ctgttgaact aaaggcactt tataaccaat atatacactt
caaagaaaca
1321 gaaattctgg ccaaggagag agaaaaagga agaattgagg aattatataa
attactagag
1381 gtgtggattg aatttggccg aattaaactg cctcaaggtt atcaccctaa
tgatgtggaa
1441 gaagagtggg gaaagctcat catagagatg ctggaacgag agaaatcact
tcggccggct
1501 gtggagaggc tggaattgct gctacagatt gcaaacaaaa tccagaatgg
tgctttgaac
1561 tgtgaagaaa aactgacact agctaagaat acactgcagg ctgatgctgc
tcacctggaa
1621 tcaggacaac cggtacaatg tgagtcagat gtcattatgt acattcagga
gtgtgaaggt
1681 ctcatcaggc agctgcaggt ggatctccag atcctgcggg atgagaatta
ctaccagcta
1741 gaagagctgg cttttagggt catgcgtctt caggatgagc tggtcacctt
gcgtctagag
1801 tgtacaaacc tgtaccggaa gggtcatttc acttcacttg aattggttcc
accctctact
1861 ttaaccacca ctcatctgaa agcagaaccc ttaaccaagg caacccattc
ttcttctacc
1921 tcctggttcc gaaagcctat gactcgggct gaacttgtgg ccatcagctc
ctctgaagat
1981 gaaggcaatc tccgatttgt gtatgaacta ctgtcttggg tagaagagat
gcagatgaaa

2041 ctggagcgag cagagtgggg caatgacctg cctagtgtgg agttgcagct
agaaacacag
2101 cagcacatcc atacgagtgt agaagagctg ggctcaagtg tcaaggaggc
caggttgtat
2161 gagggaaaga tgtcccagaa tttccatacc agctatgctg aaactcttgg
aaagctggag
2221 acacagtatt gtaaattgaa ggaaacttct agcttccgga tgaggcacct
tcagagcctg
2281 cataaatttg tttccagagc tacagctgag ttgatctggt tgaatgagaa
ggaggaggag
2341 gaactagcat atgactggag tgacaacaat tccaatatct cagccaagag
aaattacttc
2401 tctgagttga caatggaact ggaggagaaa caggatgtgt ttcgttctct
acaagataca
2461 gcagaactac tttcacttga gaaccaccca gccaagcaga cagtggaggc
ttacagtgct
2521 gctgtccagt cccagttgca gtggatgaag cagctgtgcc tgtgtgttga
gcagcatgtg
2581 aaagagaata ctgcttattt tcagttcttc agtgatgcac gagagctgga
gtcattcttg
2641 aggaacctcc aagattccat taaacgaaaa tattcctgtg accacaacac
cagcttatcc
2701 cgccttgaag acctgctcca ggactccatg gatgaaaagg agcagcttat
acagtccaag
2761 agttccgttg ccagtctcgt tgggagatca aaaaccatcg ttcagctaaa
accacgcagt
2821 ccagaccatg tgttaaagaa caccatttct gtcaaggctg tctgtgacta
caggcagatc
2881 gagattacta tttgcaaaaa tgatgaatgt gtgctagaag ataattctca
gcggaccaaa
2941 tggaaagtga tcagccccac agggaacgag gcaatggtgc cgtcagtctg
cttcctcatc
3001 cccccaccca ataaggatgc cattgagatg gccagcaggg tcgaacaatc
ttatcagaag
3061 gttatggccc tttggcatca gctgcatgtt aacaccaaaa gccttatctc
ttggaactat
3121 ctgcgtaaag accttgacct tgtacagacc tggaacctag aaaagcttcg
atcctcagca
3181 ccaggggagt gccatcagat tatgaagaac cttcaggccc actatgaaga
ctttctgcag
3241 gatagtcgtg actctgtgct gttctcagtg gctgatcgct tgcgcttgga
agaggaggtg
3301 gaagcttgta aagcccgctt ccagcacctg atgaagtcca tggagaatga
ggacaaagag
3361 gagactgtgg ccaagatgta catttcagag ttgaagaaca tccggctacg
cctggaggag
3421 tatgaacaga gggtggtcaa acgaattcag tctctagcca gctctaggac
tgacagagat
3481 gcctggcagg acaatgcatt aaggattgca gagcaagagc acacccagga
ggatttacag
3541 caattgaggt cagacttgga tgcagtttct atgaaatgtg acagctttct
ccatcagtct
3601 ccatctagtt caagtgtccc aactctgcgc tcagaactga atctgctggt
ggagaagatg
3661 gaccatgtct atggtctctc tactgtatat ctgaataagt taaagacagt
tgatgttata
3721 gtacgtagca tacaggatgc tgaactcttg gtcaaaggtt atgagattaa
gctgagtcaa
3781 gaagaagtag tactggcaga tctctcagct ctggaggccc attggtcgac
attacggcac

3841 tggcttagtg atgtgaagga caagaattca gtgttttcag tcctggatga
ggaaattgcc
3901 aaggccaagg tagtggcaga gcagatgagt cgtctgacac cagagcgaaa
tctggatttg
3961 gagcgctatc aggaaaaagg ctcccagctg caggagcgtt ggcaccgagt
cattgcccag
4021 ctcgagattc gccaatctga gctagaaagt atccaggaag ttctgggaga
ttaccgagcc
4081 tgccatggaa ctctcatcaa gtggattgag gaaaccactg cccagcagga
aatgatgaag
4141 ccaggccagg cagaggatag cagagtgctt tcggagcagc tcagccagca
gacggcccta
4201 tttgcagaaa ttgagagaaa tcagacaaaa ctggatcaat gtcaaaaatt
ttcccagcag
4261 tactctacta ttgtaaagga ctatgaattg caactgatga catacaaggc
ctttgtggaa
4321 tcgcagcaga aatcccctgg caagcgccgt cgcatgcttt cctcttcaga
tgccatcact
4381 caagagttca tggacttaag gactcgctac acggcattgg tgactttaac
aactcagcac
4441 gtgaaataca tcagtgatgc actccggcgt ctggaggagg aggagaaagt
ggtagaagag
4501 gagaaacaag aacatgtgga gaaggttaaa gaacttttgg gctgggtgtc
taccctagcg
4561 aggaatacac aaggaaaagc tacctcatcc gagaccaaag aatcaacaga
cattgaaaaa
4621 gctattttgg aacagcaggt tctgtcagaa gagctgacaa caaagaaaga
acaagtctct
4681 gaagctatta aaacatcaca gatcttcttg gccaagcatg gtcataagct
ctcagaaaaa
4741 gagaagaaac aaatatctga gcaattgaat gccctaaaca aggcttacca
tgacctttgt
4801 gatggttctg caaatcagct tcagcagctt cagagccagt tggctcacca
gacagaacaa
4861 aagaccctgc agaaacaaca aaatacctgt caccagcaac tggaggatct
ttgcagttgg
4921 gtaggacagg cagaaagagc actggcaggc caccaaggca gaaccaccca
gcaggatctc
4981 tctgctttgc agaagaacca aagtgacttg aaggatttac aggatgacat
tcagaatcgt
5041 gccacctcat ttgccactgt tgtcaaggac attgaggggt tcatggaaga
gaatcagacc
5101 aagctgagcc cacgtgagtt gacagctctt cgggaaaagc ttcatcaggc
taaggagcaa
5161 tatgaggcgc tccaggaaga gacacgtgtg gcccagaagg aactggagga
agcagtgacc
5221 tccgccttac agcaggagac tgaaaagagt aaagcagcaa aggaactggc
agagaacaag
5281 aagaagatcg atgctctcct ggattgggta acttcagtag gatcatctgg
tggacagctg
5341 ctgaccaacc ttccaggaat ggagcagctc tcgggagcta gcttggagaa
aggagccttg
5401 gacaccactg atggttacat gggggtgaat caagccccag agaaactgga
caagcaatgt
5461 gagatgatga aggcccgtca ccaagaattg ctgtcccagc agcaaaattt
cattctggcc
5521 acccagtcag ctcaggcctt cttggatcag catggccaca atctcacacc
tgaggagcaa
5581 cagatgctgc aacagaagct gggagagcta aaggaacaat actctacttc
cctggcccaa

5641 tcagaggcag aactgaagca ggtgcagaca cttcaggatg agttgcagaa
atttctgcag
5701 gatcataaag agtttgaaag ctggttggaa cgatccgaga aagagctgga
gaacatgcat
5761 aagggaggca gcagccccga gacccttccc tccctgctaa agcggcaagg
aagcttctca
5821 gaggatgtca tttcccacaa aggagacttg agatttgtga ctatctcagg
acagaaagtc
5881 ttggacatgg aaaacagttt taaggaaggc aaagaaccat cagaaattgg
aaacttagta
5941 aaggacaagt tgaaggatgc aacagaaaga tacactgctc tccactcaaa
gtgtacacga
6001 ttaggatctc acctgaatat gctgttaggc cagtatcatc aattccaaaa
cagtgctgac
6061 agcctgcagg cctggatgca ggcttgtgag gccaacgtgg agaagctcct
ctcagatact
6121 gttgcctctg accctggagt tctccaggag cagcttgcaa caacaaagca
gttgcaggag
6181 gaattggctg agcaccaagt acctgtggaa aaactccaaa aagtagctcg
tgacataatg
6241 gaaattgaag gggagccagc cccagaccac aggcatgttc aagaaactac
agattccata
6301 ctcagccact tccaaagcct ctcctatagc ctggctgagc gatcttctct
gctgcagaaa
6361 gcaattgccc aatctcagag tgtccaggaa agcctggaga gcctgttgca
gtctattggg
6421 gaagttgaac aaaacctgga agggaagcag gtgtcatcac tctcatcagg
agtcatccag
6481 gaagccttag ccacaaatat gaaattgaag caggacattg ctcggcaaaa
gagcagcttg
6541 gaggccaccc gtgagatggt gacccgattc atggagacag cagacagtac
tacagcagca
6601 gtgctgcagg gcaaactggc agaggtgagc cagcggttcg aacagctctg
tctacagcag
6661 caagaaaagg agagctccct aaagaagctt ctaccccagg cagagatgtt
tgaacacctc
6721 tctggtaagc tgcagcagtt catggaaaac aaaagtcgga tgctggcctc
tggaaatcag
6781 ccagatcaag atattacaca tttcttccaa cagatccagg agctcaattt
ggaaatggaa
6841 gaccaacagg agaacctaga tactcttgag cacctggtca ctgaactgag
ctcttgtggc
6901 tttgcgctgg acttgtgcca gcatcaggac agggtacaga atctaagaaa
agacttcaca
6961 gagctacaga agacagttaa agagagagag aaagatgcat catcttgcca
ggaacagttg
7021 gatgaattcc ggaagctggt caggaccttc cagaaatggt tgaaagaaac
tgaagggagt
7081 attccaccta cggaaacttc tatgagtgct aaagagttag aaaagcagat
tgaacacctg
7141 aagagtctac tagatgactg ggcaagtaag ggaactctgg tggaagaaat
caattgcaaa
7201 ggtacttctt tagaaaatct catcatggaa atcacagcac ctgattccca
aggcaagaca
7261 ggttccatac tgccctctgt aggaagctct gtaggcagtg taaacggata
ccacacctgc
7321 aaagatctga cggagatcca gtgtgacatg tcagatgtaa acttgaagta
tgagaaacta
7381 gggggagtac ttcatgaacg ccaggaaagc cttcaggcta tcctcaacag
aatggaggag

7441 gttcacaagg aggcaaactc tgtgctgcag tggctggaat caaaagagga
agtcctgaaa
7501 tccatggatg ccatgtcatc tccaaccaag acagaaacag tgaaagccca
agctgaatct
7561 aacaaggcct tcctggctga gttggaacag aattctccaa aaattcaaaa
agtaaaggaa
7621 gccctggctg gattactggt gacatatccc aactcacagg aagcagaaaa
ttggaagaaa
7681 attcaggaag aactcaattc ccgatgggaa agggccactg aggttactgt
ggctcggcaa
7741 aggcagctag aggaatctgc aagtcatctg gcctgcttcc aggctgcaga
atcccagctc
7801 cggccgtggc tgatggagaa agaactgatg atgggagtgc tggggcccct
gtctattgac
7861 cccaacatgt tgaatgcaca aaagcaacag gtccagttta tgctaaagga
atttgaagca
7921 cgcaggcaac agcatgagca actgaatgag gcagctcagg gcatcctaac
aggccctgga
7981 gatgtctctc tgtccaccag ccaagtacag aaagaactcc agagcatcaa
tcagaaatgg
8041 gttgagctga ctgacaaact caactcccgt tccagccaaa ttgaccaagc
tattgttaag
8101 agcacccagt accaggaact gctccaggac ttatcagaga aggtgagggc
agttggacaa
8161 cggctgagtg tccagtcagc tatcagcacc caaccagagg ctgtaaagca
gcaattggaa
8221 gagaccagtg aaattcgatc tgacttggag cagttagacc acgaggttaa
ggaggctcag
8281 acactgtgcg atgaactctc agtgctcatt ggtgagcagt acctcaagga
tgaactgaag
8341 aagcgtttgg agacagttgc cctgcctctc caaggtttag aagaccttgc
agccgatcgc
8401 attaacagac tccaggcagc tcttgccagc acccagcagt tccagcaaat
gtttgatgag
8461 ttgaggacct ggttggatga taaacaaagc cagcaagcaa aaaactgccc
aatttctgca
8521 aaattggagc ggctacagtc tcagctacag gagaatgaag agtttcagaa
aagtcttaat
8581 caacacagtg gctcctatga ggtgattgtg gctgaagggg aatctctact
tctttctgta
8641 cctcctggag aagagaaaag gactctacaa aaccagttgg ttgagctcaa
aaaccattgg
8701 gaagagctta gtaaaaaaac tgcagacaga caatccaggc tcaaggattg
tatgcagaaa
8761 gctcagaaat atcagtggca tgtggaagac cttgtgccat ggatagaaga
ttgtaaagct
8821 aagatgtctg agttgcgagt cactctggat ccagtgcagc tagagtccag
tctcctaaga
8881 tcaaaggcta tgctgaatga ggtggagaag cgccgctccc tgctggaaat
attgaatagt
8941 gctgctgaca ttctgatcaa ttcttcagaa gcagatgagg atggaatccg
ggatgagaag
9001 gctgggatca accagaacat ggatgctgtt acagaagagc tgcaggccaa
aacagggtca
9061 ctcgaagaaa tgactcagag gctcagggag ttccaggaaa gctttaagaa
tattgaaaag
9121 aaggttgaag gagccaaaca ccaacttgag atctttgatg ctctgggttc
tcaagcctgt
9181 agcaacaaga acctggagaa gctaagagct caacaggaag tgctgcaggc
cctagagcct

9241 caggtagact atctgaggaa ctttactcag ggtctggtag aagatgcccc
agatggatct
9301 gatgcttctc aacttctcca ccaagctgag gtcgcccagc aagagttcct
cgaagttaag
9361 caaagagtga acagtggttg tgtgatgatg gaaaacaagc tggaggggat
tggccagttt
9421 cactgccggg tccgagagat gttctctcaa ttggcagacc tggatgatga
gctagatggc
9481 atgggtgcta ttggcagaga cactgatagc ctccagtccc aaatcgagga
tgtccggcta
9541 ttccttaaca aaattcacgt cctcaaatta gacatagagg cctctgaagc
agagtgtcga
9601 catatgctag aagaagaggg gactctggat ttgttaggtc tcaaaaggga
gctagaagcc
9661 ctgaacaaac agtgtggcaa actgacagag agggggaaag ctcgtcagga
acagctggaa
9721 ctgacactag gccgtgtaga ggacttctac aggaaattga aaggactcaa
tgacgcgacc
9781 acagcagcag aggaggcaga ggccctccag tgggtagtgg ggaccgaagt
ggaaatcatc
9841 aaccaacaat tagcagattt taaaatgttt cagaaagaac aagtggatcc
tcttcagatg
9901 aaattgcagc aggtgaatgg acttggccag ggattaattc agagtgcagg
aaaagactgt
9961 gatgtacagg gtttagaaca tgacatggaa gagatcaatg ctcgatggaa
tacattgaat
10021 aaaaaggtcg cacaaagaat tgcacagcta caggaagctt tgttgcattg
tgggaagttt
10081 caagatgcct tggagccatt gctcagctgg ttggcagata ccgaggagct
catagccaat
10141 cagaaacctc catctgctga gtataaagtg gtgaaagcac agatccaaga
acagaagttg
10201 ctccagcggc tcctagatga tcgaaaggcc acagtagaca tgcttcaagc
agaaggaggc
10261 agaatagccc agtcagcaga gctggctgat agagagaaaa tcactggaca
gctggagagt
10321 cttgaaagta gatggactga actactcagt aaggcagcag ccaggcaaaa
acagctggaa
10381 gacatcctgg ttctggccaa acagttccat gagacagctg agcctatttc
tgacttctta
10441 tctgtcacag agaaaaagct tgctaactca gaacctgttg gcactcagac
tgccaaaata
10501 cagcagcaga tcattcggca caaggctctg gaagaagaca tagaaaacca
tgcaacagat
10561 gtgcaccagg cagtcaaaat tgggcagtcc ctctcctccc tgacatctcc
tgcagaacag
10621 ggtgtgctgt cagaaaagat agactcattg caggcccgat acagtgaaat
tcaagaccgc
10681 tgttgtcgga aggcagccct acttgaccaa gctctgtcta atgctaggct
gtttggggag
10741 gatgaggtgg aggtgctcaa ctggctggct gaggttgagg acaagctcag
ttcagtgttc
10801 gtaaaggatt tcaaacagga tgtcctgcac aggcagcatg ctgaccacct
ggctttaaat
10861 gaagaaattg ttaatagaaa gaagaatgta gatcaagcta ttaaaaatgg
tcaggctctt
10921 ctaaaacaaa ccacaggtga ggaggtgtta cttatccagg aaaaactaga
tggtataaag
10981 actcgttacg cagacatcac agttactagc tccaaggccc tcagaacttt
agagcaagcc

11041 cggcagctgg ccaccaagtt ccagtctact tatgaggaac tgaccgggtg gctgagggag
11101 gtggaggagg agctggcaac cagtggagga cagtctccca caggggaaca gataccccag
11161 tttcagcaga gacagaagga attaaagaag gaggtcatgg agcacaggct ggtgttggac
11221 acagtgaatg aggtgagccg tgctctctta gagctggtgc cctggagagc cagagaaggg
11281 ctggataaac ttgtgtccga tgctaacgag cagtacaaac tagtcagtga cactattgga
11341 caaagggtgg atgaaattga tgctgctatt cagagatcac aacagtatga gcaagctgcc
11401 gatgcagaac tagcttgggt tgctgaaaca aaacggaaac tgatggctct gggtccaatt
11461 cgcctggaac aggaccagac cacagctcag cttcaggtac agaaggcttt ctccattgac
11521 attattcgac acaaagattc aatggatgaa ctcttcagtc accgtagtga aatctttggc
11581 acatgtgggg aggagcaaaa aactgtatta caggaaaaga cagagtctct aatacagcaa
11641 tatgaagcca ttagcctact caattcagag cgttatgccc gcctagagcg ggcccaggtc
11701 ttagtaaacc agttttggga aacttatgaa gagctcagcc cctggattga ggaaactcgg
11761 gcactaatag cacagttacc ctctccagcc attgatcatg agcagctcag gcagcaacaa
11821 gaggaaatga ggcaattaag ggaatctatt gctgaacaca aacctcatat tgacaaacta
11881 ctaaagatag gcccacaact aaaggaatta aaccctgagg aaggggaaat ggtggaagaa
11941 aaataccaga aagcagaaaa catgtatgcc caaataaagg aggaggtgcg ccagcgagcc
12001 ctggctctgg atgaagccgt gtcccagtcc acacagatta cagagtttca tgataaaatt
12061 gagcctatgt tggagacact ggagaatctt tcctctcgcc tgcgtatgcc accactgatc
12121 cctgctgaag tagacaagat cagagagtgc atcagtgaca ataagagtgc caccgtggag
12181 ctagaaaaac tgcagccatc ctttgaggcc ttgaagcgcc gtggagagga gcttattgga
12241 cgatctcagg gagcagacaa ggatctggct gcaaaagaaa tccaggataa attggatcaa
12301 atggtattct tctgggagga catcaaagct cgggctgaag aacgagaaat caaatttctt
12361 gatgtccttg aattagcaga gaagttctgg tatgacatgg cagctctcct gaccaccatc
12421 aaagacaccc aggatattgt ccatgacttg gaaagcccag gcattgatcc ttccatcatc
12481 aaacaacagg ttgaagctgc tgagactatt aaggaagaga cagatggtct gcatgaagag
12541 ctggagttta ttcggatcct tggagcagat ttgatttttg cctgtggaga aactgagaag
12601 cctgaagtga ggaagagcat tgatgagatg aataatgctt gggagaactt aaacaaaaca
12661 tggaaagaga ggctagaaaa acttgaggat gctatgcaag ctgctgtgca gtatcaggac
12721 actcttcagg ctatgtttga ctggctagat aacactgtga ttaaactctg caccatgccc
12781 cctgttggca ctgacctcaa tactgttaaa gatcagttaa atgaaatgaa ggagttcaaa

12841 gtagaagttt accaacagca aattgagatg gagaagctta atcaccaggg
tgaactgatg
12901 ttaaagaaag ctactgatga gacggacaga gacattatac gagaaccact
gacagaactc
12961 aaacacctct gggagaacct gggtgagaaa attgcccacc gacagcacaa
actagaaggg
13021 gctctgttgg cccttggtca gttccagcat gccttagagg aactaatgag
ttggctgact
13081 cataccgaag agttgttaga tgctcagaga ccaataagtg gagacccaaa
agtcattgaa
13141 gttgagctcg caaagcacca tgtcctaaaa aatgatgttt tggctcatca
agccacagtg
13201 gaaacagtca acaaagctgg caatgagctt cttgaatcca gtgctggaga
tgatgccagc
13261 agcttaagga gccgtttgga agccatgaac caatgctggg agtcagtgtt
acagaaaaca
13321 gaggagaggg agcagcagct tcagtcaact ctgcagcagg cccagggctt
ccacagtgaa
13381 attgaagatt tcctcttgga acttactaga atggagagcc agctttctgc
atctaagccc
13441 acaggaggac ttcctgaaac tgctagggaa cagcttgata cacatatgga
actctattcc
13501 cagctgaaag ccaaggaaga gacttataat caactacttg acaagggcag
actcatgctt
13561 ctaagccgtg acgactctgg gtctggctcc aagacagaac agagtgtagc
acttttggag
13621 cagaagtggc atgtggtcag cagtaagatg gaagaaagaa agtcaaagct
ggaagaggcc
13681 ctcaacttgg caacagaatt ccagaattcc ctacaagaat ttatcaactg
gctcactcta
13741 gcagagcaga gtttaaacat cgcttctcca ccaagcctga ttctaaatac
tgtcctttcc
13801 cagatagaag agcacaaggt ttttgctaat gaagtaaatg ctcatcgaga
ccagatcatt
13861 gagctggatc aaactgggaa tcaattaaag ttccttagcc aaaagcagga
tgttgttctg
13921 atcaagaatt tgttggtgag cgtgcagtct cgatgggaga aggttgtcca
gcgatctatt
13981 gaaagagggc gatcactaga tgatgccagg aagcgggcaa aacaattcca
tgaagcttgg
14041 aaaaaactga ttgactggct agaagatgca gagagtcacc tggactcaga
actagagata
14101 tccaatgacc cagacaaaat taaacttcag ctttctaagc ataaggagtt
tcagaagact
14161 cttggtggca agcagcctgt gtatgatacc acaattagaa ctggcagagc
actgaaagaa
14221 aagactttgc ttcccgaaga tagtcagaaa cttgacaatt tcctaggaga
agtcagagac
14281 aaatgggata ctgtttgtgg caagtctgtg gagcggcagc acaagttgga
ggaagccctg
14341 ctcttttcgg gtcagttcat ggatgctttg caggcattgg ttgactggtt
atacaaggtg
14401 gagccacagc tggctgagga ccagcccgtg cacggggacc ttgacctcgt
catgaacctc
14461 atggatgcac acaaggtttt ccagaaggaa ctgggaaagc gaacaggaac
cgttcaggtc
14521 ctgaagcggt caggccgaga gctgattgag aatagtcgag atgacaccac
ttgggtaaaa
14581 ggacagctcc aggaactgag cactcgctgg gacactgtct gtaaactctc
tgtttccaaa

14641 caaagccggc ttgagcaggc cttaaaacaa gcggaagtgt ttcgagacac
agtccacatg
14701 ctgttggagt ggctttctga agcagagcaa acgcttcgct ttcggggagc
acttcctgat
14761 gacacagagg ccctgcagtc tctcattgac acccataagg aattcatgaa
gaaagtagaa
14821 gaaaagcgag tggacgttaa ctcagcagta gccatgggag aagtcatcct
ggctgtctgc
14881 caccccgatt gcatcacaac catcaaacac tggatcacca tcatccgagc
tcgcttcgag
14941 gaggtcctga catgggctaa gcagcaccag cagcgtcttg aaacggcctt
gtcagaactg
15001 gtggctaatg ctgagctcct ggaagaactt ctggcatgga tccagtgggc
tgagaccacc
15061 ctcattcagc gggatcagga gccaatcccg cagaacattg accgagttaa
agcccttatc
15121 gctgagcatc agacatttat ggaggagatg actcgcaaac agcctgacgt
ggaccgggtc
15181 accaagacat acaaaaggaa aaacatagag cctactcacg cgcctttcat
agagaaatcc
15241 cgcagcggag gcaggaaatc cctaagtcag ccaacccctc ctcccatgcc
aatcctttca
15301 cagtctgaag caaaaaaccc acggatcaac cagctttctg cccgctggca
gcaggtgtgg
15361 ctgttagcac tggagcggca aaggaaactg aatgatgcct tggatcggct
ggaggagttg
15421 aaagaatttg ccaactttga ctttgatgtc tggaggaaaa agtatatgcg
ttggatgaat
15481 cacaaaaagt ctcgagtgat ggatttcttc cggcgcattg ataaggacca
ggatgggaag
15541 ataacacgtc aggagtttat cgatggcatt ttagcatcca agttccccac
caccaagtta
15601 gagatgactg ctgtggctga cattttcgac cgagatgggg atggttacat
tgattattat
15661 gaatttgtgg ctgctcttca tcccaacaag gatgcgtatc gaccaacaac
cgatgcagat
15721 aaaatcgaag atgaggttac aagacaagtg gctcagtgca aatgtgcaaa
aaggtttcag
15781 gtggagcaga tcggagagaa taaataccgg tttggggatt ctcagcagtt
gcggctggtc
15841 cgtattctgc gcagcaccgt gatggttcgc gttggtggag gatggatggc
cttggatgaa
15901 tttttagtga aaaatgatcc ctgccgagca cgaggtagaa ctaacattga
acttagagag
15961 aaattcatcc taccagaggg agcatcccag ggaatgaccc ccttccgctc
acggggtcga
16021 aggtccaaac catcttcccg ggcagcttcc cctactcgtt ccagctccag
tgctagtcag
16081 agtaaccaca gctgtacatc catgccatct tctccagcca ccccagccag
tggaaccaag
16141 gttatcccat catcaggtag caagttgaaa cgaccaacac caacttttca
ttctagtcgg
16201 acatcccttg ctggtgatac cagcaatagt tcttccccgg cctccacagg
tgccaaaact
16261 aatcgggcag accctaaaaa gtctgccagt cgccctggga gtcgggctgg
gagtcgagcc
16321 gggagtcgag ccagcagccg gcgaggaagt gacgcttctg actttgacct
cttagagacg
16381 cagtctgctt gttccgacac ttcagaaagc agcgctgcag ggggccaagg
caactccagg

```
    16441 agagggctaa acaaaccttc caaaatccca accatgtcta agaagaccac
cactgcctcc
    16501 cccaggactc caggtcccaa gcgataacac tgtctaagca cccccaagcc
actatccact
    16561 ttgaatcctg ctccatacat tgggtgtata tttattctga acgggagaag
ttatattgtt
    16621 aaaagtgtaa aagaataatt gtgttatgaa gctgccttat ttttttttctt
tttgtaagtt
    16681 actattttca tgtgaatatt tatgtagata aaatttgcct cctggtaacc
ctgtaatgga
    16741 tggggcccag aaatgaaata tttgagaaaa acaagtgaaa aggtcaagat
acaaatgtgt
    16801 attaaaaaaa aaaaagccta ttaatagggt ttctgcgcgg tgcagggttg
taaacctgct
    16861 ttatctttta ggattattcc taaatgcatc ttctttataa acttgacttg
ctatctcagc
    16921 aagataaatt atattaaaaa aataagaatc ctgcagtgtt taaggaactc
tttttttgta
    16981 aatcacggac acctcaatta gcaagaactg aggggagggc tttttccatt
gtttaatgtt
    17041 ttgtgatttt tagctaaaga gagggaacct catctaagta acatttgcac
atgatacagc
    17101 aaaaggagtt cattgcaata ctgtctttgg atattgtttc agtactgggt
gtttaaagga
    17161 caaatagctg ctagaattca ggggtaaatg taagtgttca gaaaacgtca
gaacatttgg
    17221 ggttttaaac tgatttgttg ctccctatcc agcctagaca ccagtaactc
ttgtgttcac
    17281 caggacccag acccttggca agggataggc tcgttggtga cattgtgaat
ttcagatttg
    17341 ttttatccac tttttttgct atttatttaa atggtcgatc aacttcccac
aaactgagga
    17401 atgaattcca cgagcctgtt ctgaaaatgt ggacgtaaga caaacacgtg
ctcgtccttt
    17461 aatggagttc accagcacac ttgttaacca gtcctgtttg ctttcgtctt
tttttgtgcg
    17521 taataaagtc aactgaccaa gtgaccatga aaaggggctg tctggggctc
ctgtttttta
    17581 gctgctgttc ttcagctccg accatgttgc tgtgtgatta tctcaattgg
ttttaattga
    17641 ggcagaaact gaagctctac caatgaactg tttagaaaca agacacactt
ttgtattaaa
    17701 attgcttgca gtaacaaata ttttgtattt cctgattttc ttttcaacta
ttaccttatc
    17761 tataaatgtt accctggggt ataatcatgt tgtaggtact taaatgcatt
ccgcaaatca
    17821 aaatatcttg atggataaat tatagagctt aatagatctt gttttatttc
aaaaaaaaaa
    17881 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_036222.3

LOCUS: NP_036222

ACCESSION: NP_036222 NP_148984

```
      1 msssdeetls erscrsersc rsersyrser sgslspcppg dtlpwnlplh
eqkkrksqds
       61 vldpaeravv rvaderdrvq kktftkwvnk hlmkvrkhin dlyedlrdgh
nlisllevls
      121 giklprekgr mrfhrlqnvq ialdflkqrq vklvnirndd itdgnpkltl
gliwtiilhf
      181 qisdiyisge sgdmsakekl llwtqkvtag ytgikctnfs scwsdgkmfn
alihryrpdl
      241 vdmervqiqs nrenleqafe vaerlgvtrl ldaedvdvps pdeksvityv
ssiydafpkv
      301 peggegisat evdsrwqeyq srvdslipwi kqhtilmsdk tfpqnpvelk
alynqyihfk
      361 eteilakere kgrieelykl levwiefgri klpqgyhpnd veeewgklii
emlerekslr
      421 paverlelll qiankiqnga lnceekltla kntlqadaah lesgqpvqce
sdvimyiqec
      481 eglirqlqvd lqilrdenyy qleelafrvm rlqdelvtlr lectnlyrkg
hftslelvpp
      541 stlttthlka epltkathss stswfrkpmt raelvaisss edegnlrfvy
ellswveemq
      601 mkleraewgn dlpsvelqle tqqhihtsve elgssvkear lyegkmsqnf
htsyaetlgk
      661 letqycklke tssfrmrhlq slhkfvsrat aeliwlneke eeelaydwsd
nnsnisakrn
      721 yfseltmele ekqdvfrslq dtaellslen hpakqtveay saavqsqlqw
mkqlclcveq
      781 hvkentayfq ffsdareles flrnlqdsik rkyscdhnts lsrledllqd
smdekeqliq
      841 skssvaslvg rsktivqlkp rspdhvlknt isvkavcdyr qieiticknd
ecvlednsqr
      901 tkwkvisptg neamvpsvcf lipppnkdai emasrveqsy qkvmalwhql
hvntkslisw
      961 nylrkdldlv qtwnleklrs sapgechqim knlqahyedf lqdsrdsvlf
svadrlrlee
     1021 eveackarfq hlmksmened keetvakmyi selknirlrl eeyeqrvvkr
iqslassrtd
     1081 rdawqdnalr iaeqehtqed lqqlrsdlda vsmkcdsflh qspssssvpt
lrselnllve
     1141 kmdhvyglst vylnklktvd vivrsiqdae llvkgyeikl sqeevvladl
saleahwstl
     1201 rhwlsdvkdk nsvfsvldee iakakvvaeq msrltpernl dleryqekgs
qlqerwhrvi
     1261 aqleirqsel esiqevlgdy rachgtlikw ieettaqqem mkpgqaedsr
vlseqlsqqt
     1321 alfaeiernq tkldqcqkfs qqystivkdy elqlmtykaf vesqqkspgk
rrrmlsssda
     1381 itqefmdlrt rytalvtltt qhvkyisdal rrleeeekvv eeekqehvek
vkellgwvst
     1441 larntqgkat ssetkestdi ekaileqqvl seelttkkeq vseaiktsqi
flakhghkls
     1501 ekekkqiseq lnalnkayhd lcdgsanqlq qlqsqlahqt eqktlqkqqn
tchqqledlc
     1561 swvgqaeral aghqgrttqq dlsalqknqs dlkdlqddiq nratsfatvv
kdiegfmeen
     1621 qtklsprelt alreklhqak eqyealqeet rvaqkeleea vtsalqqete
kskaakelae
     1681 nkkkidalld wvtsvgssgg qlltnlpgme qlsgaslekg aldttdgymg
vnqapekldk
     1741 qcemmkarhq ellsqqqnfi latqsaqafl dqhghnltpe eqqmlqqklg
elkeqystsl
```

```
     1801 aqseaelkqv qtlqdelqkf lqdhkefesw lersekelen mhkggsspet
lpsllkrqgs
     1861 fsedvishkg dlrfvtisgq kvldmensfk egkepseign lvkdklkdat
erytalhskc
     1921 trlgshlnml lgqyhqfqns adslqawmqa ceanveklls dtvasdpgvl
qeqlattkql
     1981 qeelaehqvp veklqkvard imeiegepap dhrhvqettd silshfqsls
yslaerssll
     2041 qkaiaqsqsv qeslesllqs igeveqnleg kqvsslssgv iqealatnmk
lkqdiarqks
     2101 sleatremvt rfmetadstt aavlqgklae vsqrfeqlcl qqqekesslk
kllpqaemfe
     2161 hlsgklqqfm enksrmlasg nqpdqdithf fqqiqelnle medqqenldt
lehlvtelss
     2221 cgfaldlcqh qdrvqnlrkd ftelqktvke rekdasscqe qldefrklvr
tfqkwlkete
     2281 gsipptetsm sakelekqie hlkslldwa skgtlveein ckgtslenli
meitapdsqg
     2341 ktgsilpsvg ssvgsvngyh tckdlteiqc dmsdvnlkye klggvlherq
eslqailnrm
     2401 eevhkeansv lqwleskeev lksmdamssp tktetvkaqa esnkaflael
eqnspkiqkv
     2461 kealagllvt ypnsqeaenw kkiqeelnsr weratevtva rqrqleesas
hlacfqaaes
     2521 qlrpwlmeke lmmgvlgpls idpnmlnaqk qqvqfmlkef earrqqheql
neaaqgiltg
     2581 pgdvslstsq vqkelqsinq kwveltdkln srssqidqai vkstqyqell
qdlsekvrav
     2641 gqrlsvqsai stqpeavkqq leetseirsd leqldhevke aqtlcdelsv
ligeqylkde
     2701 lkkrletval plqgledlaa drinrlqaal astqqfqqmf delrtwlddk
qsqqakncpi
     2761 saklerlqsq lqeneefqks lnqhsgsyev ivaegeslll svppgeekrt
lqnqlvelkn
     2821 hweelskkta drqsrlkdcm qkaqkyqwhv edlvpwiedc kakmselrvt
ldpvqlessl
     2881 lrskamlnev ekrrslleil nsaadilins seadedgird ekaginqnmd
avteelqakt
     2941 gsleemtqrl refqesfkni ekkvegakhq leifdalgsq acsnknlekl
raqqevlqal
     3001 epqvdylrnf tqglvedapd gsdasqllhq aevaqqefle vkqrvnsgcv
mmenklegig
     3061 qfhcrvremf sqladlddel dgmgaigrdt dslqsqiedv rlflnkihvl
kldieaseae
     3121 crhmleeegt ldllglkrel ealnkqcgkl tergkarqeq leltlgrved
fyrklkglnd
     3181 attaaeeaea lqwvvgteve iinqqladfk mfqkeqvdpl qmklqqvngl
gqgliqsagk
     3241 dcdvqglehd meeinarwnt lnkkvaqria qlqeallhcg kfqdalepll
swladteeli
     3301 anqkppsaey kvvkaqiqeq kllqrllddr katvdmlqae ggriaqsael
adrekitgql
     3361 eslesrwtel lskaaarqkq ledilvlakq fhetaepisd flsvtekkla
nsepvgtqta
     3421 kiqqqiirhk aleedienha tdvhqavkig qslssltspa eqgvlsekid
slqaryseiq
     3481 drccrkaall dqalsnarlf gedevevlnw laevedklss vfvkdfkqdv
lhrqhadhla
     3541 lneeivnrkk nvdqaikngq allkqttgee vlliqekldg iktryaditv
tsskalrtle
```

3601 qarqlatkfq styeeltgwl reveeelats ggqsptgeqi pqfqqrqkel
kkevmehrlv
3661 ldtvnevsra llelvpwrar egldklvsda neqyklvsdt igqrvdeida
aiqrsqqyeq
3721 aadaelawva etkrklmalg pirleqdqtt aqlqvqkafs idiirhkdsm
delfshrsei
3781 fgtcgeeqkt vlqektesli qqyeaislln seryarlera qvlvnqfwet
yeelspwiee
3841 traliaqlps paidheqlrq qqeemrqlre siaehkphid kllkigpqlk
elnpeegemv
3901 eekyqkaenm yaqikeevrq ralaldeavs qstqitefhd kiepmletle
nlssrlrmpp
3961 lipaevdkir ecisdnksat veleklqpsf ealkrrgeel igrsqgadkd
laakeiqdkl
4021 dqmvffwedi karaeereik fldvlelaek fwydmaallt tikdtqdivh
dlespgidps
4081 iikqqveaae tikeetdglh eelefirilg adlifacget ekpevrksid
emnnawenln
4141 ktwkerlekl edamqaavqy qdtlqamfdw ldntviklct mppvgtdlnt
vkdqlnemke
4201 fkvevyqqqi emeklnhqge lmlkkatdet drdiireplt elkhlwenlg
ekiahrqhkl
4261 egallalgqf qhaleelmsw lthteellda qrpisgdpkv ievelakhhv
lkndvlahqa
4321 tvetvnkagn ellessagdd asslrsrlea mnqcwesvlq kteereqqlq
stlqqaqgfh
4381 seiedfllel trmesqlsas kptgglpeta reqldthmel ysqlkakeet
ynqlldkgrl
4441 mllsrddsgs gskteqsval leqkwhvvss kmeerkskle ealnlatefq
nslqefinwl
4501 tlaeqslnia sppslilntv lsqieehkvf anevnahrdq iieldqtgnq
lkflsqkqdv
4561 vliknllvsv qsrwekvvqr siergrsldd arkrakqfhe awkklidwle
daeshldsel
4621 eisndpdkik lqlskhkefq ktlggkqpvy dttirtgral kektllpeds
qkldnflgev
4681 rdkwdtvcgk sverqhklee allfsgqfmd alqalvdwly kvepqlaedq
pvhgdldlvm
4741 nlmdahkvfq kelgkrtgtv qvlkrsgrel iensrddttw vkgqlqelst
rwdtvcklsv
4801 skqsrleqal kqaevfrdtv hmllewlsea eqtlrfrgal pddtealqsl
idthkefmkk
4861 veekrvdvns avamgevila vchpdcitti khwitiirar feevltwakq
hqqrletals
4921 elvanaelle ellawiqwae ttliqrdqep ipqnidrvka liaehqtfme
emtrkqpdvd
4981 rvtktykrkn iepthapfie ksrsggrksl sqptpppmpi lsqseaknpr
inqlsarwqq
5041 vwllalerqr klndaldrle elkefanfdf dvwrkkymrw mnhkksrvmd
ffrridkdqd
5101 gkitrqefid gilaskfptt klemtavadi fdrdgdgyid yyefvaalhp
nkdayrpttd
5161 adkiedevtr qvaqckcakr fqveqigenk yrfgdsqqlr lvrilrstvm
vrvgggwmal
5221 deflvkndpc rargrtniel rekfilpega sqgmtpfrsr grrskpssra
asptrssssa
5281 sqsnhsctsm psspatpasg tkvipssgsk lkrptptfhs srtslagdts
nssspastga
5341 ktnradpkks asrpgsrags ragsrassrr gsdasdfdll etqsacsdts
essaaggqgn
5401 srrglnkpsk iptmskkttt asprtpgpkr

//

MAP1B

Official Symbol: **MAP1B**

Official Name: microtubule-associated protein 1B

Gene ID: 4131

Organism: Homo sapiens

Other Aliases: FUTSCH, MAP5

Other Designations: MAP-1B

Nucleotide sequence:

NCBI Reference Sequence: NM_019217.1

LOCUS: NM_019217

ACCESSION : NM_019217 XM_001061557 XM_215469

```
      1 cgcgcaggga gagagcggag ggggaggcga cgcgcgccgg gaggaggggg
gacgcagtgg
       61 gcggagcgga gacagcacct tcggagataa tcctttctcc tgccgcagag
cagaggagcg
      121 gcgggagagg aacacttctc ccaggcttta gcagagccgg caggatggcg
accgtggtgg
      181 tggaagccac cgagccggag ccatcgggca gcatcggcaa cccggcggcg
accacctcgc
      241 ccagcctgtc gcaccgcttc ctagacagca agttctactt gctggtggtg
gtcggcgaga
      301 cggtgaccga agagcacctg aggcgtgcca tcggcaacat cgagctgggg
atccgatcgt
      361 gggacacaaa cctgatcgag tgcaacttgg accaagagct caaacttttc
gtgtctcgac
      421 actccgcgag attctctcct gaagttccag gacaaaagat cctccatcac
cgaagtgacg
      481 tcttagaaac tgtagttctg atcaaccctt cggatgaagc agtcagcacc
gaggtgcgtt
      541 tgatgatcac tgacgccgcc cgccataaac tgctggtgct caccggacag
tgctttgaga
      601 acactggaga gctcatcctc cagtcaggct ctttctcctt ccagaacttc
atagagattt
      661 tcaccgacca agagattggg gagctcctaa gcaccaccca tcctgccaac
aaagccagcc
      721 tcaccctctt ctgccctgag gaaggagact ggaagaactc caaccttgac
agacacaatc
      781 tccaagactt catcaacatc aagctcaact cagcttctat cttgccagaa
atggagggac
```

841 tttctgagtt caccgagtac ctctcggagt ctgtcgaagt cccctccccc
tttgacatcc
901 tggagccccc gacctcgggc ggatttctga agctctccaa gccttgttgt
tacatttttc
961 cgggggggccg cggggactct gccctgttcg cagtgaacgg attcaacatg
ctcattaacg
1021 gaggatcaga aagaaagtcc tgcttctgga agctcattcg gcacttggac
cgggtggact
1081 ccatcctgct cacccacatt ggggatgaca acttgcccgg gatcaacagc
atgttgcaac
1141 gcaagattgc agagctggaa gaggagcggt cccagggctc caccagcaac
agtgactgga
1201 tgaaaaacct catctcccct gacttggggg ttgtgtttct caatgtacct
gaaaatctga
1261 aaaacccaga acccaacatc aagatgaaga gaagtacaga agaagcatgc
ttcaccctcc
1321 agtacctaaa caaactgtcc atgaaaccag agcctttatt tagaagtgta
ggcaatgcca
1381 ttgagcctgt catcctgttc caaaaaatgg gagtgggtaa actggagatg
tacgtgctta
1441 acccagtcaa aagcagcaag gaaatgcagt atttcatgca gcagtggact
ggaaccaaca
1501 aagacaaggc tgaacttatc ctgcccaatg gtcaagaagt agacatcccg
atttcctacc
1561 tgacttccgt ctcgtctttg attgtgtggc acccagccaa ccctgctgag
aaaatcatcc
1621 gggttctgtt tcctggaaac agcacccagt acaacatcct agaagggctg
gaaaaactca
1681 aacatctaga cttcctaaag cagccactgg ccacccaaaa agatctcact
ggccaggtgt
1741 ccacccccccc agtgaaacag gtcaagttga aacagcgggc tgacagccga
gagagtctga
1801 agccagccac aaaaccactt tccagtaaat cagtgaggaa ggagtccaaa
gaggaggccc
1861 ctgaagccac aaaagccagc caagtggaaa aaacacccaa agttgaaagc
aaagagaaag
1921 tgatagtgaa aaaagacaag ccaggaaagg tagaaagtaa gccatcggtg
acggaaaagg
1981 aggtgcccag caaagaggag cagtcgcccg tcaaagctga ggtggctgag
aaggcggcca
2041 cggagagcaa acccaaagtc accaaagaca aagtggtaaa aaaggaaata
aagacaaaac
2101 ccgaagaaaa gaaagaggag aagcccaaga aggaagtggc taaaaaggaa
gacaaaactc
2161 ccctcaagaa agacgagaag cccaaaaagg aagaggcgaa gaaggagatc
aagaaagaaa
2221 tcaaaaagga agagaaaaag gagctgaaga aagaggtgaa gaaggaaacg
cccctgaagg
2281 acgccaagaa ggaggtgaag aaagacgaga agaaagaagt taaaaaggaa
gagaaggaac
2341 ccaaaaagga gattaagaag atctccaagg acataaagaa atccactcct
ctgtcagaca
2401 caaagaaacc ggctgcattg aaaccaaaag tagcaaagaa agaagagccc
accaagaagg
2461 agcctattgc tgctgggaaa ctcaaggaca aggggaaggt caaagtcatt
aaaaaggaag
2521 gcaagaccac agaggccgct gccacagctg ttggcactgc tgccgtggct
gcagcagccg
2581 gagtagcggc cagcggtcct gccaaagaac ttgaagctga gcggtccctc
atgtcgtccc

2641 ctgaggatct aaccaaggac tttgaggagc taaaggctga ggagatcgat
gtagcgaagg
2701 acatcaagcc tcagctggag ctcattgaag atgaagagaa actgaaggaa
accgagccgg
2761 gagaagccta cgtcattcag aaagagacgg aagtcagcaa aggttctgct
gagtcacctg
2821 atgaagggat caccaccact gaggggggaag gggagtgcga gcaaaccccc
gaggagctgg
2881 agccagttga gaagcagggc gtggatgaca tcgagaagtt cgaggatgaa
ggcgctggtt
2941 ttgaagaatc ctcagaggcc ggagactacg aagagaaggc agaaactgag
gaggccgagg
3001 agccggaaga agacggggaa gacaatgtga gcgggagcgc ctcgaagcac
agccccacag
3061 aagacgaaga aatcgctaag gctgaggcgg acgtacacat caaggagaag
agggagtctg
3121 tggccagcgg cgatgaccgg gccgaagaag acatggatga agcgcttgag
aaaggagaag
3181 ctgaacagtc tgaggaggag ggtgaggagg aggaggacaa agcagaggac
gccagagagg
3241 aagaccatga gcccgacaaa actgaggctg aagattatgt gatggctgtg
gttgacaagg
3301 ccgcggaggc cggagtcacc gaggatcagt atgatttcct ggggacaccg
gccaagcaac
3361 ctggagtcca gtctcctagc cgagaacccg cgtcttcaat tcatgatgag
accctacccg
3421 gaggctccga gagcgaggcc actgcttcag atgaggagaa tcgagaagac
cagcctgagg
3481 aattcactgc tacctccgga tatactcagt ccaccatcga gatatctagt
gagccgactc
3541 caatggatga gatgtccact cctcgagatg tgatgaccga cgagaccaac
aatgaggaga
3601 cagagtcccc gtctcaggag ttcgtgaaca ttaccaaata cgagtcttcg
ctgtactctc
3661 aggagtactc caaacctgtg gttgcatcat tcaatggatt gtcagacggg
tcaaagacag
3721 acgccactga cggtagggat tacaacgctt ccgcctccac catatcacca
ccttcgtcca
3781 tggaagaaga caaattcagc aagtctgctc ttcgtgacgc ttaccgccca
gaagagacgg
3841 acgtgaaaac cggtgccgag ttggacatca aagatgtttc ggatgagaga
cttagcccag
3901 ccaagagtcc atccctgagt ccttctccac catcacccat agagaagact
cccctgggtg
3961 aacgtagcgt gaatttctct ctgacaccca acgagatcaa agcctctgca
gagggagagg
4021 caacagcagt agtgtccccc ggagtgaccc aagcagtagt tgaagaacac
tgtgccagtc
4081 ctgaggagaa gaccttggag gtagtgtcac cgtctcagtc tgtgacaggc
agtgcgggcc
4141 acacacctta ctaccaatct cccaccgacg aaaagtccag tcacctacct
acagaagtca
4201 ctgagaacgc gcaggcagtc ccggtgagct ttgaattcac tgaggccaaa
gatgagaacg
4261 agaggtcgtc catcagcccc atggatgaac ctgtgcctga ctcagagtct
cctatcgaga
4321 aagttctgtc tccgttacgc agccctcccc ttattggatc cgagtccgca
tatgaagact
4381 tcctgagtgc ggatgacaag gctcttggca gacgttcaga aagccccttt
gaagggaaga

4441 atggaaagca aggcttctca gacaaagaaa gcccagtttc tgacctgact
tccgatcttt
4501 accaagacaa gcaggaagag aaaagcgcgg gcttcatacc gataaaggaa
gactttagtc
4561 cagaaaagaa agccagcgat gctgaaatca tgagttctca atcagctctg
gctttggatg
4621 aaaggaaact gggaggagat ggatctccaa cgcaagtaga tgtcagtcag
tttggctctt
4681 tcaaagaaga caccaagatg tccatttcgg aaggcaccgt ttcagacaag
tccgccacgc
4741 ctgtggatga gggcgtggcc gaagacacct attcacacat ggaaggtgtg
gcctcagtgt
4801 caaccgcctc tgtggctacc agctcgtttc cagagccaac cacagatgac
gtgtctcctt
4861 ctctccacgc tgaagtgggc tctccacatt ccacagaggt ggatgactcc
ctgtcggtgt
4921 cggtggtgca aacaccaact actttccagg aaacagaaat gtctccgtct
aaagaagagt
4981 gcccaagacc aatgtcgatt tctcctcctg acttctcccc taagacagcc
aaatccagga
5041 caccagttca agatcaccga tccgaacagt cttcaatgtc tattgaattc
ggtcaggaat
5101 cccccgagca ttctcttgct atggacttta gtcggcagtc tccagaccac
cctactgtgg
5161 gtgctggtat gcttcacatc accgaaaatg ggccaactga ggtggactac
agtccctccg
5221 atatccagga ctctagtttg tcacataaga ttccgccgac agaagagcca
tcctacaccc
5281 aggataatga tctgtccgag ctcatctctg tgtctcaggt ggaggcttcc
ccatccacct
5341 cttctgctca cactccttct cagatagcct ctcctcttca ggaagacact
ctctctgatg
5401 tcgttcctcc cagagatatg tccttatatg cctcgcttgc gtctgagaaa
gtgcagagcc
5461 tggaaggaga gaaactctct ccaaaatccg atatttctcc gctcacccct
cgagagtcct
5521 cacctacata ttcacctggc ttttcagatt ctacctctgg agctaaagag
agtacagcgg
5581 cttaccaaac ctcctcttcc ccaccaatag atgcagcagc cgcagagccc
tacggcttcc
5641 gctcctcaat gttatttgat acaatgcagc atcacctggc cttgagtaga
gatttgacca
5701 catctagtgt ggagaaggac aatggaggga agacacccgg tgactttaac
tatgcctatc
5761 aaaagcccga gagcaccacc gaatccccag atgaagaaga ttatgactat
gaatctcacg
5821 agaaaaccat ccaggcccac gatgtgggtg gttactacta tgagaagaca
gagagaacca
5881 taaaatcccc atgtgacagt ggatactcct atgagaccat tgagaagacc
accaagaccc
5941 cagaagatgg tggctactcc tgtgaaatta ccgagaaaac cactcggacc
cctgaagagg
6001 gcgggtactc gtatgagatc agcgagaaga caacacgaac ccctgaagta
agtggctaca
6061 cctatgagaa gaccgagagg tccagaaggc tcctcgatga cattagcaat
ggctacgatg
6121 acactgagga tggtggccac acacttggcg actgtagcta ttcctacgaa
accactgaga
6181 aaattaccag ctttcctgaa tctgaaagct attcctatga gacaactaca
aaaacaacac

6241 ggagtccaga cacctctgca tactgttacg agaccatgga gaagatcacc
aagaccccac
6301 aggcatccac atactcctat gagacctcag accgatgcta cactccagaa
aggaagtccc
6361 cctcggaggc acgccaggat gttgacttgt gtctggtgtc ctcctgtgaa
ttcaagcatc
6421 ccaagaccga gctctcacct tccttcatta atccaaaccc tctcgagtgg
tttgctgggg
6481 aagagcccac tgaagaatct gagaagcctc tcactcagtc tggaggagcc
cccccacctt
6541 caggaggaaa acaacagggc agacaatgcg atgaaactcc acccacctca
gtcagtgagt
6601 cagctccatc ccagacggac tctgatgttc ccccagagac agaagagtgc
ccctccatca
6661 cagctgatgc caacattgac tctgaagatg agtcagaaac catccccaca
gacaaaacgg
6721 ttacgtacaa acacatggac ccgcctccag cccccatgca agaccgaagc
ccttctcctc
6781 gccaccctga tgtgtccatg gtggatccag aggccttggc tattgagcag
aacctaggca
6841 aggctctgaa aaaggatctg aaggagaagg ccaagaccaa gaaaccaggc
acaaagacca
6901 agtcctcttc acctgtcaaa aagggtgatg ggaagtccaa gccttcagca
gcttccccca
6961 aaccaggagc cttgaaggaa tcctctgaca aggtgtccag agtggcttct
cccaagaaga
7021 aagagtctgt ggagaaagct atgaagacca ccaccactcc tgaggtcaaa
gccacacgag
7081 gggaagagaa ggacaaggaa actaagaatg cagccaatgc ttctgcatcc
aagtcagtga
7141 agactgcaac agcaggacca ggaaccacta agacggccaa gtcgtccacc
gtgcctcccg
7201 gcctccctgt gtatttggac ctctgctata ttcccaacca cagcaacagt
aagaatgtcg
7261 atgttgagtt tttcaagaga gtgaggtcat cttactacgt ggtgagtggg
aacgaccctg
7321 ccgcggagga gcccagccgg gctgtcctgg atgccttgtt ggaagggaag
gctcagtggg
7381 gaagcaacat gcaggtgact ctgatcccaa cacatgactc tgaggtgatg
agggagtggt
7441 accaggagac ccacgagaag cagcaagacc tcaacatcat ggtcctagca
agcagtagta
7501 cagtggtcat gcaagacgag tccttccctg catgcaagat agaactgtag
aaaccgcagc
7561 cgaccacacc acaggatttg aactgtgttt ccagaaattc ctgaatttga
aactaccttt
7621 tcttaaacgt ccattcatct aattacgtca ctgaacaagg acctgccaga
tgctatacag
7681 tgtcatggtg atgcaaatca ctgatatttc tcaatttttg ccgaccgcta
aggaaagtaa
7741 ccatattccc acaatagatt tcaagttact gcaaaattac ctacccccgt
tcatctctgc
7801 tgaaatacgt ggaagccagg cactcgcaca cccaactgac ttctgctagg
tattggcatt
7861 tatcttagag agagaaagag agagggaaag agagtcagcg ggaggggagg
gagagcggcg
7921 ggagagaggg acaaagagac tgccggagag agagagagtg agagtgagag
aggaagaatc
7981 agaaagaaaa agaatgcaag acaaaaaagg tagagagttc tgatgagatg
cccagggaga

8041 aagagtgggc aggatggggt atagagaaga caccagcaac tgggtctgcg
ttttcccaga
8101 ccacagcgat tcattctgtg gtctacacag gtggagtttt ccattttcac
cagagtcatc
8161 agaaagagtc gatctcctaa agcttgtttc taaagaacag gaaggacgaa
gcctgtcacg
8221 agggcatgag atttttcacg ccttaattaa atgcctttgc tataaaggct
gcccacgcat
8281 aatgtagcaa gtgtaggctg gagaagcgag tggttggagc cccgttcaga
actctcagac
8341 ttttcaaaca cgtgagtagg ttgatctaag gcatgctccc agcatttgtc
tacccaagtc
8401 cacatcgagt caacccgcat gcagcaacac ccaaggccac cccagttaac
tgaagcaaat
8461 accaaagcag ttgggagaac atatgggaga cattttgcct taggaagtga
cttgaatgta
8521 caaagttacc cgatgcactt atttttaac gtgagacggc aagtttttaa
aacatccgtg
8581 taggattgta gatactccag gcgaaggagc atgcggggag ggaaggtact
agaaactcgt
8641 ctgactggcc caacagttta gtgcagagtc atgcttggtg aacgtcacca
cttctgatgt
8701 acccgtggcc tctcagccag ataagttgca ccctaagtag cttctttaac
gcttcaagtt
8761 taagactgaa atggcttctc taatcagaac cagggaaaca atgaatctca
cggtggaagg
8821 ggttctcggc aagtgtacag tgtctgcctt cctttgtctt gcattgacta
ttttaatttt
8881 tccattaatt ccaacacgtg ggaacacatg tacagaagat ttttttttt
agatacatga
8941 gaacttttca tagatgaact ttctaacgaa tgttttcatt tacagaaaaa
tgcaaagaaa
9001 aatttgaagc gatggtcttt ttttttaatt attattttaa gtgttttgta
agacaaaaaa
9061 attgaagttt tttgaggttc tggaaagatt tgaagcctga tattgaagtc
gtgatgatat
9121 ttatttaaaa acccgtcact actggaaacg gtggtacctc ccacccttg
actctcatat
9181 tatgaaagtg tgagtccgtg ctgtttgaga gtgggtaggt ggcagggtag
gctactgttc
9241 agggtttcac agtgctattc cctcctcctt tcaagatttt ttcaccgtga
ggtggaagag
9301 ccaagttcag aagcacccta gcgccagctt gcttgggcct tttctggaaa
acattcattg
9361 aaagaaataa cagattgaaa acaaatgaat tctcagctcc tacgttcacc
atgtagagag
9421 ttcagacaca atgtccgtca ctgtcatcac tgaaccacaa actcgtaacg
ccagatcatg
9481 aggaaatctt tcgccaagtt tcaaacggca gatccatgta ccaggggttc
agagttggca
9541 atcttctcag tgacagccat gacagctcgt tcacgctgag tttcctgcag
actcttaaga
9601 tctccggagt agtgaacaat gacctcattt tattttctat gttagttatt
tatttcaaaa
9661 gttacatttt agtttacttt tcgtctgtga agtctatgtt tcgcactgct
gtttactctg
9721 agggtttaac aatatttctc cagggtcccc tcaccgagga cccatgcagt
ctacttaatg
9781 ctgtgaatca cattttccaa atgtttaatt ttttaaagaa aattaatatt
ctatttttgt

9841 taggcgtctc taggaatgca gcttttattt attttcctat ttctttccaa
ctcgtaaaag
9901 ccacacatta aaggtacaga aatgtcctat atggtggaag agacattgag
aacggagttc
9961 attgagtgtc agctcacttg tttctcactt caggcaagct gaacacacac
aagatggcaa
10021 tctcatggta gctgttgggt tggtccacac aatacactca aagagaaaca
gtttctagcc
10081 ttcttgccaa atccagacct ctggttgact tttctttcct aaaagatgga
gttactgccc
10141 agttctacag cttaaattta tttagccttt tatattattt tgttttaaag
atgcaaagac
10201 cttagaagaa ccatagccag ccttcagtta taaccactcg accccatacg
tcttattgtt
10261 tatacagaat ccgatgggga aaacactttt ttttttcaaa actgtcactg
atccactggc
10321 tgtgattcct acacaatctg gcagcaccat cctgacccct gtcctcagtg
atggccatgt
10381 agccgaggac accagggtga agtacattgg ctttgcaggt agctcctgat
ctcccaggca
10441 cccctcctac ctgtgtggct gatctgatct tgcttgttcc atacatacac
aattagttta
10501 gaggtagcca ccacgatgga ctatgtacat ttgtggtgag agctcaaaac
cgacacagac
10561 ttctagaaga tttgtgcata caatccttga tccaattgta tagattgact
atttgagtgg
10621 aaggcgtttc caccctgttt aaatgatgga attctattgt gctagcactc
ccgatcatga
10681 ccctttttggt agtatttgta aacaaaattc tacagagact aaatcttaga
gataatcctc
10741 catttcaatt ttaatcaatt ctgtcctcct tttttccaaa ccccgaaccc
catgcatgct
10801 ttcccagtct tgtgatggga ctggacacag tgaccgaccc cctcgcctca
aacaccattt
10861 tccatggaat tcaaaagaaa aaaatttttt ttcttaacct tacatatcat
agtgaatggt
10921 ttccccggtg tatatgaatg ttttaagtgt ttccaatagc ttatgaaatt
taggagcttt
10981 ctaatactcg ttttataaat ttaatcattt gctaatggaa attttaccac
ctcccatttg
11041 tgttacaaat cttagctcct ggagcggcac tacaattcag gagttgtttt
ttctcacctc
11101 ctctgtcatt tgtcacagga ggtccctgct tggcaatgac atttgtgagt
taggataatg
11161 acgttccttc tctccttttt ttttcctttc atacttcaga tttaggagaa
aaagattctg
11221 tttccacgtg agaggaactg taagctttta tcacgtaacc agctgaacaa
cacaccaaaa
11281 gcagcctagg gatgagcacc gcgctttggt agcgattagg ttttattcac
ctggtattaa
11341 aactattcac tatttcaaaa atccggaact tttaagaatt catttcaaag
gcagcatcaa
11401 aaactgaaaa ggaagggaaa aaaaaacaac agctaataat cggcttctcc
gcacgcgtgg
11461 agctcgcgaa actggagccc cggagaagtg gctctgctca gccgcccgcc
cacgccgcgg
11521 cggtccttgc tttccccgca tgcgcccgca ggcagcgtgc agtcctaagc
ccggctgtgg
11581 agaagctcac tctctctctt gttctgaatg gtgtttgtgt cggtctgcct
ctgtgtatgg

```
    11641 tattatgtct tataatcctg catcacttcc atcctatcca gtcatatcta
atgtagaaaa
    11701 attagtttcc agtgaaagta atatgtagtg cttttatggt atttgtgtgc
aatatcccct
    11761 cttctattga ggatatttga tgtaaaggaa aaaaaaaaag aaaaaagaaa
ctgagttcca
    11821 caataaaata caaagtggca aaagttcact tgtgtgttga gacatcaaaa
aaaaaaaaaa
    11881 aaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_062090.1

LOCUS: NP_062090

ACCESSION: NP_062090 XP_001061557 XP_215469

```
      1 matvvveate pepsgsignp aattspslsh rfldskfyll vvvgetvtee
hlrraignie
       61 lgirswdtnl iecnldqelk lfvsrhsarf spevpgqkil hhrsdvletv
vlinpsdeav
      121 stevrlmitd aarhkllvlt gqcfentgel ilqsgsfsfq nfieiftdqe
igellstthp
      181 ankasltlfc peegdwknsn ldrhnlqdfi niklnsasil pemeglseft
eylsesvevp
      241 spfdileppt sggflklskp ccyifpggrg dsalfavngf nmlinggser
kscfwklirh
      301 ldrvdsillt higddnlpgi nsmlqrkiae leeersqgst snsdwmknli
spdlgvvfln
      361 vpenlknpep nikmkrstee acftlqylnk lsmkpeplfr svgnaiepvi
lfqkmgvgkl
      421 emyvlnpvks skemqyfmqq wtgtnkdkae lilpngqevd ipisyltsvs
slivwhpanp
      481 aekiirvlfp gnstqynile gleklkhldf lkqplatqkd ltgqvstppv
kqvklkqrad
      541 sreslkpatk plssksvrke skeeapeatk asqvektpkv eskekvivkk
dkpgkveskp
      601 svtekevpsk eeqspvkaev aekaateskp kvtkdkvvkk eiktkpeekk
eekpkkevak
      661 kedktplkkd ekpkkeeakk eikkeikkee kkelkkevkk etplkdakke
vkkdekkevk
      721 keekepkkei kkiskdikks tplsdtkkpa alkpkvakke eptkkepiaa
gklkdkgkvk
      781 vikkegktte aaatavgtaa vaaaagvaas gpakeleaer slmsspedlt
kdfeelkaee
      841 idvakdikpq leliedeekl ketepgeayv iqketevskg saespdegit
ttegegeceq
      901 tpeelepvek qgvddiekfe degagfeess eagdyeekae teeaeepeed
gednvsgsas
      961 khsptedeei akaeadvhik ekresvasgd draeedmdea lekgeaeqse
eegeeeedka
     1021 edareedhep dkteaedyvm avvdkaaeag vtedqydflg tpakqpgvqs
psrepassih
     1081 detlpggses eatasdeenr edqpeeftat sgytqstiei sseptpmdem
stprdvmtde
```

```
     1141 tnneetesps qefvnitkye sslysqeysk pvvasfngls dgsktdatdg
rdynasasti
     1201 sppssmeedk fsksalrday rpeetdvktg aeldikdvsd erlspaksps
lspsppspie
     1261 ktplgersvn fsltpneika saegeatavv spgvtqavve ehcaspeekt
levvspsqsv
     1321 tgsaghtpyy qsptdekssh lptevtenaq avpvsfefte akdenerssi
spmdepvpds
     1381 espiekvlsp lrsppligse sayedflsad dkalgrrses pfegkngkqg
fsdkespvsd
     1441 ltsdlyqdkq eeksagfipi kedfspekka sdaeimssqs alalderklg
gdgsptqvdv
     1501 sqfgsfkedt kmsisegtvs dksatpvdeg vaedtyshme gvasvstasv
atssfpeptt
     1561 ddvspslhae vgsphstevd dslsvsvvqt pttfqetems pskeecprpm
sisppdfspk
     1621 taksrtpvqd hrseqssmsi efgqespehs lamdfsrqsp dhptvgagml
hitengptev
     1681 dyspsdiqds slshkippte epsytqdndl selisvsqve aspstssaht
psqiasplqe
     1741 dtlsdvvppr dmslyaslas ekvqslegek lspksdispl tpressptys
pgfsdstsga
     1801 kestaayqts ssppidaaaa epygfrssml fdtmqhhlal srdlttssve
kdnggktpgd
     1861 fnyayqkpes ttespdeedy dyeshektiq ahdvggyyye ktertikspc
dsgysyetie
     1921 kttktpedgg ysceitektt rtpeeggysy eisekttrtp evsgytyekt
ersrrllddi
     1981 sngyddtedg ghtlgdcsys yettekitsf pesesysyet ttkttrspdt
saycyetmek
     2041 itktpqasty syetsdrcyt perkspsear qdvdlclvss cefkhpktel
spsfinpnpl
     2101 ewfageepte esekpltqsg gapppsggkq qgrqcdetpp tsvsesapsq
tdsdvppete
     2161 ecpsitadan idsedeseti ptdktvtykh mdpppapmqd rspsprhpdv
smvdpealai
     2221 eqnlgkalkk dlkekaktkk pgtktksssp vkkgdgkskp saaspkpgal
kessdkvsrv
     2281 aspkkkesve kamkttttpe vkatrgeekd ketknaanas asksvktata
gpgttktaks
     2341 stvppglpvy ldlcyipnhs nsknvdveff krvrssyyvv sgndpaaeep
sravldalle
     2401 gkaqwgsnmq vtlipthdse vmrewyqeth ekqqdlnimv lassstvvmq
desfpackie
     2461 l
      //
```

MDH1

Official Symbol: **MDH1**

Official Name: malate dehydrogenase 1, NAD (soluble)

Gene ID: 4190

Organism: Homo sapiens

Other Aliases: MDH-s, MDHA, MGC:1375, MOR2

Other Designations: cytosolic malate dehydrogenase; malate dehydrogenase, cytoplasmic; soluble malate dehydrogenase

Nucleotide sequence:

NCBI Reference Sequence: NM_001199111.1

LOCUS: NM_001199111

ACCESSION : NM_001199111

```
      1 ccttcgcgcc ctttggcaag ctcggactca tcttctgggg attgccgcag
tgacccagta
       61 atgggaaggg attgatttcc accttgcggg gtatggggcg ctcttaggag
gactctggag
      121 aagtagttgt cctgggagag gagcgatctt aatcctgctg catgacggga
ggacaaaatg
      181 cgacgctgca gctattttcc aaaggacgtt acggtgtttg ataaggacga
taagtctgaa
      241 ccaatcagag tccttgtgac tggagcagct ggtcaaattg catattcact
gctgtacagt
      301 attggaaatg gatctgtctt tggtaaagat cagcctataa ttcttgtgct
gttggatatc
      361 acccccatga tgggtgtcct ggacggtgtc ctaatggaac tgcaagactg
tgcccttccc
      421 ctcctgaaag atgtcatcgc aacagataaa gaagacgttg ccttcaaaga
cctggatgtg
      481 gccattcttg tgggctccat gccaagaagg gaaggcatgg agagaaaaga
tttactgaaa
      541 gcaaatgtga aaatcttcaa atcccagggt gcagccttag ataaatacgc
caagaagtca
      601 gttaaggtta ttgttgtggg taatccagcc aataccaact gcctgactgc
ttccaagtca
      661 gctccatcca tccccaagga gaacttcagt tgcttgactc gtttggatca
caaccgagct
      721 aaagctcaaa ttgctcttaa acttggtgtg actgctaatg atgtaaagaa
tgtcattatc
      781 tggggaaacc attcctcgac tcagtatcca gatgtcaacc atgccaaggt
gaaattgcaa
      841 ggaaaggaag ttggtgttta tgaagctctg aaagatgaca gctggctcaa
gggagaattt
      901 gtcacgactg tgcagcagcg tggcgctgct gtcatcaagg ctcgaaaact
atccagtgcc
      961 atgtctgctg caaaagccat ctgtgaccac gtcagggaca tctggtttgg
aaccccagag
     1021 ggagagtttg tgtccatggg tgttatctct gatggcaact cctatggtgt
tcctgatgat
     1081 ctgctctact cattccctgt tgtaatcaag aataagacct ggaagtttgt
tgaaggtctc
     1141 cctattaatg atttctcacg tgagaagatg gatcttactg caaaggaact
gacagaagaa
```

```
     1201 aaagaaagtg cttttgaatt tctttcctct gcctgactag acaatgatgt
tactaaatgc
     1261 ttcaaagctg aagaatctaa atgtcgtctt tgactcaagt accaaataat
aataatgcta
     1321 tacttaaatt acttgtgaaa aacaacacat tttaaagatt acgtgcttct
tggtacaggt
     1381 ttgtgaatga cagtttatcg tcatgctgtt agtgtgcatt ctaaataaat
atatattcaa
     1441 atgaaaaaaa aaaaaaaaaa a
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001186040.1

LOCUS: NP_001186040

ACCESSION: NP_001186040

```
      1 mrrcsyfpkd vtvfdkddks epirvlvtga agqiayslly signgsvfgk
dqpiilvlld
       61 itpmmgvldg vlmelqdcal pllkdviatd kedvafkdld vailvgsmpr
regmerkdll
      121 kanvkifksq gaaldkyakk svkvivvgnp antncltask sapsipkenf
scltrldhnr
      181 akaqialklg vtandvknvi iwgnhsstqy pdvnhakvkl qgkevgvyea
lkddswlkge
      241 fvttvqqrga avikarklss amsaakaicd hvrdiwfgtp egefvsmgvi
sdgnsygvpd
      301 dllysfpvvi knktwkfveg lpindfsrek mdltakelte ekesafefls sa
      //
```

NHP2L1

Official Symbol: **NHP2L1**

Official Name: HP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)

Gene ID: 4809

Organism: Homo sapiens

Other Aliases: CTA-216E10.8, 15.5K, FA-1, FA1, NHPX, OTK27, SNRNP15-5, SNU13, SPAG12, SSFA1

Other Designations: NHP2-like protein 1; U4/U6.U5 tri-snRNP 15.5 kDa protein; [U4/U6.U5] tri-snRNP 15.5 kD RNA binding protein; high mobility group-like nuclear protein 2 homolog 1; non-histone chromosome protein 2-like 1; small nuclear ribonucleoprotein 15.5kDa (U4/U6.U5); sperm specific antigen 1

Nucleotide sequence:

NCBI Reference Sequence: NM_001003796.1

LOCUS: NM_001003796

ACCESSION : NM_001003796

```
      1 gccgcgcggg gccatttccg ctgctgcttc tgtgagtttt tccggtgcac
gcgagtgctt
       61 ctgaaacgtc agctgcgctc ccctaggagt gctgagcccg cggaaccgca
gccatgactg
      121 aggctgatgt gaatccaaag gcctatcccc ttgccgatgc ccacctcacc
aagaagctac
      181 tggacctcgt tcagcagtca tgtaactata agcagcttcg gaaaggagcc
aatgaggcca
      241 ccaaaaccct caacaggggc atctctgagt tcatcgtgat ggctgcagac
gccgagccac
      301 tggagatcat tctgcacctg ccgctgctgt gtgaagacaa gaatgtgccc
tacgtgtttg
      361 tgcgctccaa gcaggccctg gggagagcct gtggggtctc caggcctgtc
atcgcctgtt
      421 ctgtcaccat caaagaaggc tcgcagctga aacagcagat ccaatccatt
cagcagtcca
      481 ttgaaaggct cttagtctaa acctgtggcc tctgccacgt gctccctgcc
agcttccccc
      541 ctgaggttgt gtatcatatt atctgtgtta gcatgtagta ttttcagcta
ctctctattg
      601 ttataaaatg tagtactaaa tctggtttct ggatttttgt gttgtttttg
ttctgtttta
      661 cagggttgct atccccctc ctttcctccc tccctctgcc atccttcatc
cttttatcct
      721 ccctttttgg aacaagtgtt cagagcagac agaagcaggg tggtggcacc
gttgaaaggc
      781 agaaagagcc aggagaaagc tgatggagcc aggacagaga tctggttcca
gctttcagcc
      841 actagcttcc tgttgtgtgc ggggtgtggt ggaattaaac agcattcatt
gtgtgtccct
      901 gtgcctggca cacagaatca ttcatacgtg ttcaagtgat caaggggttt
catttgctct
      961 tgggggatta ggtatcattt ggggaggaag catgtgttct gtgaggttgt
tcggctatgt
     1021 ccaagtgtcg tttactaatg tacccctgct gtttgctttt ggtaatgtga
tgttgatgtt
     1081 ctccccctac ccacaaccat gcccttgagg gtagcagggc agcagcatac
caaagagatg
     1141 tgctgcagga ctccggaggc agcctgggtg ggtgagccat ggggcagttg
acctgggtct
     1201 tgaaagagtc gggagtgaca agctcagaga gcatgaactg atgctggcat
gaaggattcc
     1261 aggaagatca tggagacctg gctggtagct gtaacagaga tggtggagtc
caaggaaaca
     1321 gcctgtctct ggtgaatggg actttctttg gtggacactt ggcaccagct
ctgagagccc
     1381 ttcccctgtg tcctgccacc atgtgggtca gatgtactct ctgtcacatg
aggagagtgc
     1441 tagttcatgt gttctccatt cttgtgagca tcctaataaa tctgttccat
tttgatgaca
     1501 aaaaaaaaaa aaaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001003796.1

LOCUS: NP_001003796

ACCESSION: NP_001003796

```
      1 mteadvnpka ypladahltk klldlvqqsc nykqlrkgan eatktlnrgi
sefivmaada
       61 epleiilhlp llcedknvpy vfvrskqalg racgvsrpvi acsvtikegs
qlkqqiqsiq
      121 qsierllv
      //
```

OLA1

Official Symbol: **OLA1**

Official Name: Obg-like ATPase 1

Gene ID: 29789

Organism: Homo sapiens

Other Aliases: PTD004, DOC45, GBP45, GTBP9, GTPBP9

Other Designations: DNA damage-regulated overexpressed in cancer 45 protein; GTP-binding protein 9 (putative); GTP-binding protein PTD004; homologous yeast-44.2 protein; obg-like ATPase 1

Nucleotide sequence:

NCBI Reference Sequence: NM_001011708.1

LOCUS: NM_001011708

ACCESSION : NM_001011708

```
      1 ggtctgcgcg caggtgccgc tcggcgcccg gcccgcccgt tccgccgctg
tcgccgccgt
       61 cgtgcgtgcc gctcggcgga ggggacgggc ctgcgttctc tcctccttcc
tccccgcctc
      121 cagctgccgg caggaccttt ctctcgctgc cgctgggacc ccgtgtcatc
gcccaggccg
      181 agcacggaaa tctactttct tcaatgtgtt aaccaatagt caggcttcag
cagaaaactt
      241 cccgttctgc actattgatc ctaatgagag cagagtacct gtgccagatg
aaaggtttga
```

```
      301 ctttctttgt caataccaca aaccagcaag caaaattcct gcctttctaa
atgtggtgga
      361 tattgctggc cttgtgaaag gagctcacaa tgggcagggc ctggggaatg
ctttttatc
      421 tcatattagt gcctgtgatg gcatctttca tctaacacgt gcttttgaag
atgatgatat
      481 cacgcacgtt gaaggaagtg tagatcctat tcgagatata gaaataatac
atgaagagct
      541 tcagcttaaa gatgaggaaa tgattgggcc cattatagat aaactagaaa
aggtggctgt
      601 gagaggagga gataaaaaac taaaacctga atatgatata atgtgcaaag
taaaatcctg
      661 ggttatagat caaaagaaac ctgttcgctt ctatcatgat tggaatgaca
aagagattga
      721 agtgttgaat aaacacttat ttttgacttc aaaaccaatg gtctacttgg
ttaatctttc
      781 tgaaaaagac tacattagaa agaaaaacaa atggttgata aaaattaaag
agtgggtgga
      841 caagtatgac ccaggtgctt tggtcattcc ttttagtggg gccttggaac
tcaagttgca
      901 agaattgagt gctgaggaga gacagaagta tctggaagcg aacatgacac
aaagtgcttt
      961 gccaaagatc attaaggctg ggtttgcagc actccaacta gaatactttt
tcactgcagg
     1021 cccagatgaa gtgcgtgcat ggaccatcag gaaagggact aaggctcctc
aggctgcagg
     1081 aaagattcac acagattttg aaaagggatt cattatggct gaagtaatga
aatacgaaga
     1141 ttttaaagag gaaggttctg aaaatgcagt caaggctgct ggaaagtaca
gacaacaagg
     1201 cagaaattat attgttgaag atggagatat tatcttcttc aaatttaaca
cacctcaaca
     1261 accgaagaag aaataaaatt tagttattgc tcagataaac atacaacttc
caaaaggcat
     1321 ctgattttta aaaaattaaa atttctgaaa accaatgcga caaataaagt
tggggagatg
     1381 ggaatctttg acaaacaaat tatttttatt tgttttaaaa ttaaaatact
gtgtaccccc
     1441 ccccccccat gaaatgcagg ttcactaaat gtgaacagct ttgcttttca
cgtgattaag
     1501 accctactcc aaattgtaga agcttttcag gaaccatatt actctcatga
tacttcatta
     1561 atctccatca tgtatgccaa gcctgacaca tttgacagtg aggacaatgt
ggcttgctcc
     1621 tttttgaatc tacagataat gcatgtttta cagtactcca gatgtctaca
ctcaataaaa
     1681 catttgacaa aaccagcctt ggtgtgtttg gggatgtctg tattgactga
ctgtggtgtg
     1741 ctgaatgcga tacggcacct ggtggttgct gattacagaa ttttaaggcg
tgtgtatgac
     1801 acagtaactg gcagtgtggg gcagctgcaa ttgtatctta aaagtgagtc
tttcatggga
     1861 atcagaagaa tagtatacca gggatttgtc tcaaaaaagt taattaattt
atagatcaac
     1921 atttattgaa agatgatagc aatgatatta tgagggatat gagataggtg
actgaccaca
     1981 aagaagaaaa ttctgtctca aaaattaaag aattttgttt gttattgttg
ttctgaccat
     2041 attgaaaaga ggttcacttt taatctttcc tttgaaatta ttaaattgta
aaaactgacc
```

2101 cattgatgtc tggtgggtta tgttttgctt caattcagca atgtgtataa
aagttcctac
2161 actgattttg aaatactaag atcagacttt gaaaatttta aaaattcttt
cattcttact
2221 tttcagatat ttgagagcag tcgaagtggt agtatgggta agttaagcac
ttcttaacat
2281 tgtagcagaa aattccaaaa gaaagactag aacaaattgg taatttggag
acccttttgcc
2341 acttagcctt ctcttgtgat gatggcctga aagctctctc ggccctgagc
ctccctttct
2401 ctcccatgtt tccattcctg tgagacctcc cttccttgtc cttgtcctct
gccctttttgt
2461 gctcttctgt gatcacagga ccatttccat ggaatgcaga ttcataccca
gccagccttt
2521 gcctctactg tatcgctttg ggactttaaa gatgcagaag tattagaaca
cacacaaact
2581 taaaatggaa agttaccaaa tgtagcacta gaggcaagaa aaggggtgat
tttttaaaga
2641 tttggctata cttaagatat ttaaagagga tgaggttgag tcgtgagtat
aatttagaag
2701 ctctccagtg agtatttttt tagtatcaaa tagtgatctt cttttgctaa
cataaaaatg
2761 agtaaactac agcagagaaa taagaataaa actaatgaag ggtttttttaa
aagaactaaa
2821 taaaacaaat aaaactaggc tacaggtata gctgaccaag tttgctgtga
aaaatccttg
2881 ttatatctat ttctgtattg gagtaatgtg tagattagtg gattgttagg
gatttcagca
2941 gtataagaaa gaatcaaata tcttgctttt gggtcttctc caaagtaaat
tgcatggact
3001 tcttaaagaa gtttgaaatc agtaaaaaga cattaggatc atgcaatttt
actttgtcta
3061 gcttctaatg tggtaaattg cacattgtgc aacagaacca ttaataaaag
aaaggagtta
3121 gttttgaggt aactgcattc attaggggct gaaaataaat gttaaaaatg
tttaatttgg
3181 ttagaagttc tacaacattg tatgtacata tacaaaattt cattttctct
gaaagtcgag
3241 aattccaaga actgttacat ataatcatat atatgcctat atctaaacgt
tttccctcct
3301 ttaatattga taaagttaat atattttagg atgttatgta aatatttttg
catgtatctg
3361 tagctccatc ttaatcaaat ttatggttaa aagagatttg aggccaggcg
tggaggctca
3421 cacctataat cccaacactt tgggaggccg aggtgggagg atcacttgag
cacaggagtt
3481 tgagaccagc ctgtgcaaca tcatgaaaca tcatctctac ataaaaaaat
acatatatac
3541 aaaaattagc tgggcatcat ggcacacacc tgtagtccca gctacttagg
aggctgagat
3601 gggaggattg atcgatcgct tgagcctggg aggtggaggc tactgtgagc
catgatcatg
3661 ccactgcact ccagcctggg caacagagtg agaccctgtc tccaaagaga
aagagaaact
3721 tgagaaggct tgtgcccaaa gcttgaagaa agattgtaat agcatttagt
tgtgttttat
3781 catgttatct tcataacatg cattttgcat atggctcaga gcagagccgg
atctcccaag
3841 gaagcacaaa tagtttttgt cgctaactta gttatgagtg aagcctctgt
tcacttataa

```
     3901 cttgccagtt tcattggtgg aataagtccc cttactcatg attcatcaat
attcctatat
     3961 taaaattcca gtttgatgtt tgtctctagt tgtctgtgta taatatttcc
aaagctcctt
     4021 tgtaaacccc gtctcctttc attcagagag tcagagacgg tatctcctcc
tcctgcctct
     4081 tctattgttc tgattagtct acttaaattc acttgatgct ctttttattt
tacaagagac
     4141 ttgagaaatg tcggttgttc ctaatggcaa actactttag ctgggaaaat
tcatttcaac
     4201 ttattttaag cttgtaaaat acactgtggt gataaaataa aagagctggg tttga
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001011708.1

LOCUS: NP_001011708

ACCESSION: NP_001011708

```
      1 migpiidkle kvavrggdkk lkpeydimck vkswvidqkk pvrfyhdwnd
keievlnkhl
       61 fltskpmvyl vnlsekdyir kknkwlikik ewvdkydpga lvipfsgale
lklqelsaee
      121 rqkyleanmt qsalpkiika gfaalqleyf ftagpdevra wtirkgtkap
qaagkihtdf
      181 ekgfimaevm kyedfkeegs enavkaagky rqqgrnyive dgdiiffkfn
tpqqpkkk
      //
```

POFUT1

Official Symbol: **POFUT1**

Official Name: protein O-fucosyltransferase 1

Gene ID: 23509

Organism: Homo sapiens

Other Aliases: FUT12, O-FUT, O-Fuc-T, O-FucT-1

Other Designations: GDP-fucose protein O-fucosyltransferase 1; o-fucosyltransferase protein; peptide-O-fucosyltransferase 1

Nucleotide sequence:

NCBI Reference Sequence: NM_015352.1

LOCUS: NM_015352

ACCESSION : NM_015352

```
      1 cttccctccc cgactgtgcg ccgcggctgg ctcgggttcc cgggccgaca
tgggcgccgc
       61 cgcgtgggca cggccgctga gcgtgtcttt cctgctgctg cttctgccgc
tcccggggat
      121 gcctgcgggc tcctgggacc cggccggtta cctgctctac tgcccctgca
tggggcgctt
      181 tgggaaccag gccgatcact tcttgggctc tctggcattt gcaaagctgc
taaaccgtac
      241 cttggctgtc cctccttgga ttgagtacca gcatcacaag cctcctttca
ccaacctcca
      301 tgtgtcctac cagaagtact tcaagctgga gcccctccag gcttaccatc
gggtcatcag
      361 cttggaggat ttcatggaga agctggcacc cacccactgg ccccctgaga
agcgggtggc
      421 atactgcttt gaggtggcag cccagcgaag cccagataag aagacgtgcc
ccatgaagga
      481 aggaaacccc tttggcccat tctgggatca gtttcatgtg agtttcaaca
agtcggagct
      541 ttttacaggc atttccttca gtgcttccta cagagaacaa tggagccaga
gattttctcc
      601 aaaggaacat ccggtgcttg ccctgccagg agccccagcc cagttccccg
tcctagagga
      661 acacaggcca ctacagaagt acatggtatg gtcagacgaa atggtgaaga
cgggagaggc
      721 ccagattcat gcccaccttg tccggcccta tgtgggcatt catctgcgca
ttggctctga
      781 ctggaagaac gcctgtgcca tgctgaagga cgggactgca ggctcgcact
tcatggcctc
      841 tccgcagtgt gtgggctaca gccgcagcac agcggccccc ctcacgatga
ctatgtgcct
      901 gcctgacctg aaggagatcc agagggctgt gaagctctgg gtgaggtcgc
tggatgccca
      961 gtcggtctac gttgctactg attccgagag ttatgtgcct gagctccaac
agctcttcaa
     1021 agggaaggtg aaggtggtga gcctgaagcc tgaggtggcc caggtcgacc
tgtacatcct
     1081 cggccaagcc gaccacttta ttggcaactg tgtctcctcc ttcactgcct
ttgtgaagcg
     1141 ggagcgggac ctccagggga ggccgtcttc tttcttcggc atggacaggc
cccctaagct
     1201 gcgggacgag ttctgattct ggccggagca ccagaccctc tgatcctgga
gggaccagag
     1261 tctgagctgg tccttccagc caggcctggc agccagaggt gctccgggat
tgcaaactcc
     1321 tcttctcacc tgccaaagat ggagaagagt gccagggacc cctcaaggag
ggagacgctc
     1381 catatcccag ggcataggac ttgcaggttc ctaggagcag gagcatctcc
catcgcacgt
     1441 gctttctgct cttctgggaa tttctcacac tggcaaagca gtccagcctc
cgtcttctgg
     1501 tccactctgc tctgagcagc ctgggatgct gaactcttca gagagatttt
tttatagaga
```

1561 gatttctata attttgatac aaggtcatga ctatcctaga actctctgtg gtttttgaaa
1621 atcattgaat tctattaatg taggtaccta aagtgacctt aactgaatgt ggatgaggct
1681 ggggctggtg tgggtctttt ggctgctttt caaggtgtcc cccaatgtgg ccctcaagag
1741 ccatccccac tgcctggcca gagccattgt tgtcccctac ttcctaggcc atttctgggg
1801 cttgggggat gaatgctgtc ctgtgctgta aacactatgc aaatggaagt tatcggttgt
1861 ggtgctgtgc agcgctctgt gggcgactaa gtgccactca cgcagcatgt tcctggcaag
1921 gagcacatac catcaagcca cactatcatg gtattgttct cacagtcttt tggtggttga
1981 tggccactgc aaacctggca ccatcagatc tcttctgatc tcttgcccca gtggggcctg
2041 gttggtagaa tgttggcatt cggttgatat ccaaagcctg ttctcccagc cgtcctcctg
2101 cagctggagc cttcaggccg tattctcacg agggaacgtt tgccaaggct ctgacctcac
2161 agaagatgcc cagggcccag aagccatcag aattatcagt ggagaagcac cttttgactc
2221 ttcccttcca atgtaatctc tgccaacacc atgaggctta aggtgctcta agtcatgagt
2281 gttttggtct caaatgctgc agttttaata atctgtgact cctgagagcc catggttttt
2341 tgaccttgtg gttctaaaat tccttgtctg acccctgtag atcttttcct tgccatgtca
2401 cctcccttgg cctttgatcc tggaaaggtg gcagagcctc cactgagcca ggcccagagc
2461 tccttgcagt gccttcttcc ttgtttacct gtgggaggaa acacttttt tgtcaggggc
2521 agcctggttc agagctcaga ggtcacactg tatcaaagat ctcaaacagc aaagtcagca
2581 tttgctgtat agagctgcca cccaactcta agcaggagaa actgtacaga aagggctttg
2641 ctatttttcc cttttgggaa aacaatgaag tgttttaagt cctgggtgga ctgagagatg
2701 gtttgcctgt ccagacttgc tctcaagcct catccagaga aggagctgca gatgagggag
2761 cccgtacact ccctgccacc actaggttgt aagcctgtag ctggctggct gatttcattt
2821 tggaattcat ttgccatcca cagccttaca ctaggcacac actttagagt ctggggctcc
2881 agtggggccc gcctaatttt ttttcccccc aagacagggc cttgctctgt ctcccaggct
2941 ggagtgcagt ggcatgatca tggcttactg cagccttgat ctcccaggct caagcgatcc
3001 ttctgcctca gcctctctgg tagctgagac tgcatgccca gctccaaatc accttgattc
3061 atatcagcag taataatcac ttgtgttctg aaagaaaggg caccagaagt tctagcaaaa
3121 ttcagttgtg ttctgtgagc tagcactttt tcctctgacc caatttttctt acctataaaa
3181 tggtgataaa aaccgacagg ttgttcaaag gcccagatca gctaaagcat gtatataaga
3241 gcacgttgta aacttgaaag agacaaaggc acaaatgtgg ctgttgatta atttgactgc
3301 ttctcgttgc tcgtcacctc catgccaggc actgtgcttg ctaattgctt tatgggggca

3361 ttctcttatt tattccccag ccctgggaaa taggagctgt cattatcctt ctctttctgc
3421 acaaggaaaa attaatgccc tgagaattgt cataattttc ccaaggctgc ccagctggtg
3481 gtgttaagcc agaatttgac ctcccagagc cagtttccat tagctgccat gctctgctgc
3541 ctctaattca cagaatgcac tttctaccct gtgtgccatg gagacctcct atggaaaaat
3601 gatcagccac cttaccttct actgggtacc tgctgtgagt ctgcctatgc cagaaggatt
3661 aaggagggga ggttacccaa gaaacaaagc ctacatgccg cttacagccc ccgttggatg
3721 gttgctcagt acaacagtct tgcattcagc aggtgtttgt tcatcaccta ctatgtgtca
3781 ggctctatgc taggtactgg ggatacagga gagaatcaag cgtaaagtct ttgttctcaa
3841 ggaatttgca ttctagaaag tagaagatgt aataaatgta ctgtgggaca tgttaataag
3901 tgctataaag aaatataaag ggtttgggag caaaaagagg gagtggatct attttagatg
3961 agcccaggta agacctctct gaagagctgt catgaaggag ggagggagca cattcctggc
4021 agagaaaaca gcacgtgcaa aggccccgag actggagtgt gttcctgaag agcagccagg
4081 aggccagcat ggctggagag gcaggcatag gcagggaacc gagcagcagg tcagagcagg
4141 cgagctgaca ttctgcagcc tggacggcca tggcaggaag cttttagttg gagagataca
4201 ggaagcctcc tagggttctg agcagaagag gggcatgagc tgattcacat tctgaaggac
4261 ctctctagct ggccagtgct gaggaggttg gagagagaaa gggtgaaagc agagagacca
4321 gtgcagggct gttaacaggg ttgcaggcga gagactgggg tgctgggctc ccctagacta
4381 ggactccagt gccctcctct cccaagagac aaaggccatt gcattgaagg aggtgggaaa
4441 tgattagatt ctgaacatat gtaattattt ttcagtcttt ttcaaagata caaatattta
4501 catagtttta atcatgtaat atatacaatt taatgtccta gtgttttact taatagtgta
4561 tcatgttttc cctgttggta tgtagcctgg ataaatgctc ttaattataa aaaattctgt
4621 cgaggagtgt tccatagttt attgttttcc tattatgaga atttaggcca agtgtggtgg
4681 ctcatgcctg taatcccagc actttgcgag gccgaggtgg gcagatcact tgaggtgagg
4741 agttcaagac cagcctggcc aacatggtga attatctcta ctaaaaatac aaaaaaataa
4801 taataatagc caggcgtggt ggcacatgcc tgtattccca gctgcttggg aggctgaggc
4861 aggagaatgg cttgaacctg ggaggtggag gttgcagtga gccgagatgg tgccactgca
4921 ttccagcctg ggcaacagag cgagactcca tctcaaaaaa aaggagactt catgtgcccc
4981 caatttttca ctattgttat ttgaaaaaat atttttattt gtaagagttt ttctttattt
5041 aaaatgttca ttaataaagt tgttggacgg gaagcaaaaa aaaaaagttg tttaagataa
5101 attcccagaa gtgaatttgt tagatcaaac acttaaaact ttttgttatg gaagaattca

```
     5161 aatataaata aaaaattgtg agtaataaaa tgaactcaca gtttcaacaa
tgacccacaa
     5221 aaaaaaaaaa aaaaaaaaaa aaaaaaaaa
      //
```

Protein sequence:

NCBI Reference Sequence: NP_056167.1

LOCUS: NP_056167

ACCESSION: NP_056167

```
      1 mgaaawarpl svsflllllp lpgmpagswd pagyllycpc mgrfgnqadh
flgslafakl
       61 lnrtlavppw ieyqhhkppf tnlhvsyqky fkleplqayh rvisledfme
klapthwppe
      121 krvaycfeva aqrspdkktc pmkegnpfgp fwdqfhvsfn kselftgisf
sasyreqwsq
      181 rfspkehpvl alpgapaqfp vleehrplqk ymvwsdemvk tgeaqihahl
vrpyvgihlr
      241 igsdwknaca mlkdgtagsh fmaspqcvgy srstaapltm tmclpdlkei
qravklwvrs
      301 ldaqsvyvat dsesyvpelq qlfkgkvkvv slkpevaqvd lyilgqadhf
ignсvssfta
      361 fvkrerdlqg rpssffgmdr ppklrdef
      //
```

PRKDC

Official Symbol: **PRKDC**

Official Name: protein kinase, DNA-activated, catalytic polypeptide

Gene ID: 5591

Organism: Homo sapiens

Other Aliases: DNA-PKcs, DNAPK, DNPK1, HYRC, HYRC1, XRCC7, p350

Other Designations: DNA-PK catalytic subunit; DNA-dependent protein kinase catalytic subunit; hyper-radiosensitivity of murine scid mutation, complementing 1; p460

Nucleotide sequence:

NCBI Reference Sequence: NM_001081640.1

LOCUS: NM_001081640

ACCESSION : NM_001081640

```
      1 ggggcatttc cgggtccggg ccgagcgggc gcacgcgcgg gagcgggact
cggcggcatg
       61 gcgggctccg gagccggtgt gcgttgctcc ctgctgcggc tgcaggagac
cttgtccgct
      121 gcggaccgct gcggtgctgc cctggccggt catcaactga tccgcggcct
ggggcaggaa
      181 tgcgtcctga gcagcagccc cgcggtgctg gcattacaga catctttagt
tttttccaga
      241 gatttcggtt tgcttgtatt tgtccggaag tcactcaaca gtattgaatt
tcgtgaatgt
      301 agagaagaaa tcctaaagtt tttatgtatt ttcttagaaa aaatgggcca
gaagatcgca
      361 ccttactctg ttgaaattaa gaacacttgt accagtgttt atacaaaaga
tagagctgct
      421 aaatgtaaaa ttccagccct ggaccttctt attaagttac ttcagacttt
tagaagttct
      481 agactcatgg atgaatttaa aattggagaa ttatttagta aattctatgg
agaacttgca
      541 ttgaaaaaaa aaataccaga tacagtttta gaaaaagtat atgagctcct
aggattattg
      601 ggtgaagttc atcctagtga gatgataaat aatgcagaaa acctgttccg
cgcttttctg
      661 ggtgaactta agacccagat gacatcagca gtaagagagc ccaaactacc
tgttctggca
      721 ggatgtctga aggggttgtc ctcacttctg tgcaacttca ctaagtccat
ggaagaagat
      781 ccccagactt caagggagat ttttaatttt gtactaaagg caattcgtcc
tcagattgat
      841 ctgaagagat atgctgtgcc ctcagctggc ttgcgcctat ttgccctgca
tgcatctcag
      901 tttagcacct gccttctgga caactacgtg tctctatttg aagtcttgtt
aaagtggtgt
      961 gcccacacaa atgtagaatt gaaaaaagct gcactttcag ccctggaatc
ctttctgaaa
     1021 caggtttcta atatggtggc gaaaaatgca gaaatgcata aaaataaact
gcagtacttt
     1081 atggagcagt tttatggaat catcagaaat gtggattcga acaacaagga
gttatctatt
     1141 gctatccgtg gatatggact ttttgcagga ccgtgcaagg ttataaacgc
aaaagatgtt
     1201 gacttcatgt acgttgagct cattcagcgc tgcaagcaga tgttcctcac
ccagacagac
     1261 actggtgacg accgtgttta tcagatgcca agcttcctcc agtctgttgc
aagcgtcttg
     1321 ctgtaccttg acacagttcc tgaggtgtat actccagttc tggagcacct
cgtggtgatg
     1381 cagatagaca gtttcccaca gtacagtcca aaaatgcagc tggtgtgttg
cagagccata
     1441 gtgaaggtgt tcctagcttt ggcagcaaaa gggccagttc tcaggaattg
cattagtact
     1501 gtggtgcatc agggtttaat cagaatatgt tctaaaccag tggtccttcc
aaagggccct
     1561 gagtctgaat ctgaagacca ccgtgcttca ggggaagtca gaactggcaa
atggaaggtg
     1621 cccacataca aagactacgt ggatctcttc agacatctcc tgagctctga
ccagatgatg
     1681 gattctattt tagcagatga agcatttttc tctgtgaatt cctccagtga
aagtctgaat
```

1741 catttacttt atgatgaatt tgtaaaatcc gttttgaaga ttgttgagaa attggatctt
1801 acacttgaaa tacagactgt tggggaacaa gagaatggag atgaggcgcc tggtgtttgg
1861 atgatcccaa cttcagatcc agcggctaac ttgcatccag ctaaacctaa agatttttcg
1921 gctttcatta acctggtgga attttgcaga gagattctcc ctgagaaaca agcagaattt
1981 tttgaaccat gggtgtactc attttcatat gaattaattt tgcaatctac aaggttgccc
2041 ctcatcagtg gtttctacaa attgctttct attacagtaa gaaatgccaa gaaaataaaa
2101 tatttcgagg gagttagtcc aaagagtctg aaacactctc ctgaagaccc agaaaagtat
2161 tcttgctttg ctttatttgt gaaatttggc aaagaggtgg cagttaaaat gaagcagtac
2221 aaagatgaac ttttggcctc ttgtttgacc tttcttctgt ccttgccaca caacatcatt
2281 gaactcgatg ttagagccta cgttcctgca ctgcagatgg ctttcaaact gggcctgagc
2341 tatacccccc tggcagaagt aggcctgaat gctctagaag aatggtcaat ttatattgac
2401 agacatgtaa tgcagcctta ttacaaagac attctcccct gcctggatgg atacctgaag
2461 acttcagcct tgtcagatga gaccaagaat aactgggaag tgtcagctct ttctcgggct
2521 gcccagaaag gatttaataa agtggtgtta aagcatctga agaagacaaa gaacctttca
2581 tcaaacgaag caatatcctt agaagaaata agaattagag tagtacaaat gcttggatct
2641 ctaggaggac aaataaacaa aaatcttctg acagtcacgt cctcagatga gatgatgaag
2701 agctatgtgg cctgggacag agagaagcgg ctgagctttg cagtgccctt tagagagatg
2761 aaacctgtca ttttcctgga tgtgttcctg cctcgagtca cagaattagc gctcacagcc
2821 agtgacagac aaactaaagt tgcagcctgt gaacttttac atagcatggt tatgtttatg
2881 ttgggcaaag ccacgcagat gccagaaggg ggacagggag ccccacccat gtaccagctc
2941 tataagcgga cgtttcctgt gctgcttcga cttgcgtgtg atgttgatca ggtgacaagg
3001 caactgtatg agccactagt tatgcagctg attcactggt tcactaacaa caagaaattt
3061 gaaagtcagg atactgttgc cttactagaa gctatattgg atggaattgt ggaccctgtt
3121 gacagtactt taagagattt ttgtggtcgg tgtattcgag aattccttaa atggtccatt
3181 aagcaaataa caccacagca gcaggagaag agtccagtaa acaccaaatc gcttttcaag
3241 cgactttata gccttgcgct tcaccccaat gctttcaaga ggctgggagc atcacttgcc
3301 tttaataata tctacaggga attcagggaa gaagagtctc tggtggaaca gtttgtgttt
3361 gaagccttgg tgatatacat ggagagtctg gccttagcac atgcagatga gaagtcctta
3421 ggtacaattc aacagtgttg tgatgccatt gatcacctat gccgcatcat tgaaaagaag
3481 catgtttctt taaataaagc aaagaaacga cgtttgccgc gaggatttcc accttccgca

3541 tcattgtgtt tattggatct ggtcaagtgg cttttagctc attgtgggag
gccccagaca
3601 gaatgtcgac acaaatccat tgaactcttt tataaattcg ttccttttatt
gccaggcaac
3661 agatccccta atttgtggct gaaagatgtt ctcaaggaag aaggtgtctc
ttttctcatc
3721 aacacctttg agggggggtgg ctgtggccag ccctcgggca tcctggccca
gcccaccctc
3781 ttgtaccttc gggggccatt cagcctgcag gccacgctat gctggctgga
cctgctcctg
3841 gccgcgttgg agtgctacaa cacgttcatt ggcgagagaa ctgtaggagc
gctccaggtc
3901 ctaggtactg aagcccagtc ttcacttttg aaagcagtgg ctttcttctt
agaaagcatt
3961 gccatgcatg acattatagc agcagaaaag tgctttggca ctggggcagc
aggtaacaga
4021 acaagcccac aagagggaga aaggtacaac tacagcaaat gcaccgttgt
ggtccggatt
4081 atggagttta ccacgactct gctaaacacc tccccggaag gatggaagct
cctgaagaag
4141 gacttgtgta atacacacct gatgagagtc ctggtgcaga cgctgtgtga
gcccgcaagc
4201 ataggtttca acatcggaga cgtccaggtt atggctcatc ttcctgatgt
ttgtgtgaat
4261 ctgatgaaag ctctaaagat gtccccatac aaagatatcc tagagaccca
tctgagagag
4321 aaaataacag cacagagcat tgaggagctt tgtgccgtca acttgtatgg
ccctgacgcg
4381 caagtggaca ggagcaggct ggctgctgtt gtgtctgcct gtaaacagct
tcacagagct
4441 gggcttctgc ataatatatt accgtctcag tccacagatt tgcatcattc
tgttggcaca
4501 gaacttcttt ccctggttta taaaggcatt gcccctggag atgagagaca
gtgtctgcct
4561 tctctagacc tcagttgtaa gcagctggcc agcggacttc tggagttagc
ctttgctttt
4621 ggaggactgt gtgagcgcct tgtgagtctt ctcctgaacc cagcggtgct
gtccacggcg
4681 tccttgggca gctcacaggg cagcgtcatc cacttctccc atggggagta
tttctatagc
4741 ttgttctcag aaacgatcaa cacggaatta ttgaaaaatc tggatcttgc
tgtattggag
4801 ctcatgcagt cttcagtgga taataccaaa atggtgagtg ccgttttgaa
cggcatgtta
4861 gaccagagct tcagggagcg agcaaaccag aaacaccaag gactgaaact
tgcgactaca
4921 attctgcaac actggaagaa gtgtgattca tggtgggcca aagattcccc
tctcgaaact
4981 aaaatggcag tgctggcctt actggcaaaa attttacaga ttgattcatc
tgtatctttt
5041 aatacaagtc atggttcatt ccctgaagtc tttacaacat atattagtct
acttgctgac
5101 acaaagctgg atctacattt aaagggccaa gctgtcactc ttcttccatt
cttcaccagc
5161 ctcactggag gcagtctgga ggaacttaga cgtgttctgg agcagctcat
cgttgctcac
5221 ttccccatgc agtccaggga atttcctcca ggaactccgc ggttcaataa
ttatgtggac
5281 tgcatgaaaa agtttctaga tgcattggaa ttatctcaaa gccctatgtt
gttggaattg

5341 atgacagaag ttctttgtcg ggaacagcag catgtcatgg aagaattatt
tcaatccagt
5401 ttcaggagga ttgccagaag gggttcatgt gtcacacaag taggccttct
ggaaagcgtg
5461 tatgaaatgt tcaggaagga tgacccccgc ctaagtttca cacgccagtc
ctttgtggac
5521 cgctccctcc tcactctgct gtggcactgt agcctggatg ctttgagaga
attcttcagc
5581 acaattgtgg tggatgccat tgatgtgttg aagtccaggt ttacaaagct
aaatgaatct
5641 acctttgata ctcaaatcac caagaagatg ggctactata agattctaga
cgtgatgtat
5701 tctcgccttc ccaaagatga tgttcatgct aaggaatcaa aaattaatca
agttttccat
5761 ggctcgtgta ttacagaagg aaatgaactt acaaagacat tgattaaatt
gtgctacgat
5821 gcatttacag agaacatggc aggagagaat cagctgctgg agaggagaag
actttaccat
5881 tgtgcagcat acaactgcgc catatctgtc atctgctgtg tcttcaatga
gttaaaattt
5941 taccaaggtt ttctgtttag tgaaaaacca gaaaagaact tgcttatttt
tgaaaatctg
6001 atcgacctga agcgccgcta taattttcct gtagaagttg aggttcctat
ggaaagaaag
6061 aaaaagtaca ttgaaattag gaaagaagcc agagaagcag caaatgggga
ttcagatggt
6121 ccttcctata tgtcttccct gtcatatttg gcagacagta ccctgagtga
ggaaatgagt
6181 caatttgatt tctcaaccgg agttcagagc tattcataca gctcccaaga
ccctagacct
6241 gccactggtc gttttcggag acgggagcag cgggacccca cggtgcatga
tgatgtgctg
6301 gagctggaga tggacgagct caatcggcat gagtgcatgg cgcccctgac
ggccctggtc
6361 aagcacatgc acagaagcct gggcccgcct caaggagaag aggattcagt
gccaagagat
6421 cttccttctt ggatgaaatt cctccatggc aaactgggaa atccaatagt
accattaaat
6481 atccgtctct tcttagccaa gcttgttatt aatacagaag aggtctttcg
cccttacgcg
6541 aagcactggc ttagcccctt gctgcagctg gctgcttctg aaaacaatgg
aggagaagga
6601 attcactaca tggtggttga gatagtggcc actattcttt catggacagg
cttggccact
6661 ccaacagggg tccctaaaga tgaagtgtta gcaaatcgat tgcttaattt
cctaatgaaa
6721 catgtctttc atccaaaaag agctgtgttt agacacaacc ttgaaattat
aaagaccctt
6781 gtcgagtgct ggaaggattg tttatccatc ccttataggt taatatttga
aaagttttcc
6841 ggtaaagatc ctaattctaa agacaactca gtagggattc aattgctagg
catcgtgatg
6901 gccaatgacc tgcctcccta tgacccacag tgtggcatcc agagtagcga
atacttccag
6961 gctttggtga ataatatgtc ctttgtaaga tataaagaag tgtatgccgc
tgcagcagaa
7021 gttctaggac ttatacttcg atatgttatg gagagaaaaa acatactgga
ggagtctctg
7081 tgtgaactgg ttgcgaaaca attgaagcaa catcagaata ctatggagga
caagtttatt

7141 gtgtgcttga acaaagtgac caagagcttc cctcctcttg cagacaggtt
catgaatgct
7201 gtgttctttc tgctgccaaa atttcatgga gtgttgaaaa cactctgtct
ggaggtggta
7261 ctttgtcgtg tggagggaat gacagagctg tacttccagt taaagagcaa
ggacttcgtt
7321 caagtcatga gacatagaga tgatgaaaga caaaaagtat gtttggacat
aatttataag
7381 atgatgccaa agttaaaacc agtagaactc cgagaacttc tgaaccccgt
tgtggaattc
7441 gtttcccatc cttctacaac atgtagggaa caaatgtata atattctcat
gtggattcat
7501 gataattaca gagatccaga aagtgagaca gataatgact cccaggaaat
atttaagttg
7561 gcaaaagatg tgctgattca aggattgatc gatgagaacc ctggacttca
attaattatt
7621 cgaaatttct ggagccatga aactaggtta ccttcaaata ccttggaccg
gttgctggca
7681 ctaaattcct tatattctcc taagatagaa gtgcactttt taagtttagc
aacaaatttt
7741 ctgctcgaaa tgaccagcat gagcccagat tatccaaacc ccatgttcga
gcatcctctg
7801 tcagaatgcg aatttcagga atataccatt gattctgatt ggcgtttccg
aagtactgtt
7861 ctcactccga tgtttgtgga gacccaggcc tcccagggca ctctccagac
ccgtacccag
7921 gaagggtccc tctcagctcg ctggccagtg gcagggcaga taagggccac
ccagcagcag
7981 catgacttca cactgacaca gactgcagat ggaagaagct catttgattg
gctgaccggg
8041 agcagcactg acccgctggt cgaccacacc agtccctcat ctgactcctt
gctgtttgcc
8101 cacaagagga gtgaaaggtt acagagagca cccttgaagt cagtggggcc
tgattttggg
8161 aaaaaaaggc tgggccttcc aggggacgag gtggataaca aagtgaaagg
tgcggccggc
8221 cggacggacc tactacgact gcgcagacgg tttatgaggg accaggagaa
gctcagtttg
8281 atgtatgcca gaaaaggcgt tgctgagcaa aaacgagaga aggaaatcaa
gagtgagtta
8341 aaaatgaagc aggatgccca ggtcgttctg tacagaagct accggcacgg
agaccttcct
8401 gacattcaga tcaagcacag cagcctcatc accccgttac aggccgtggc
ccagagggac
8461 ccaataattg caaaacagct ctttagcagc ttgttttctg gaattttgaa
agagatggat
8521 aaatttaaga cactgtctga aaaaaacaac atcactcaaa agttgcttca
agacttcaat
8581 cgttttctta ataccacctt ctctttcttt ccacccttg tctcttgtat
tcaggacatt
8641 agctgtcagc acgcagccct gctgagcctc gacccagcgg ctgttagcgc
tggttgcctg
8701 gccagcctac agcagcccgt gggcatccgc ctgctagagg aggctctgct
ccgcctgctg
8761 cctgctgagc tgcctgccaa gcgagtccgt gggaaggccc gcctccctcc
tgatgtcctc
8821 agatgggtgg agcttgctaa gctgtataga tcaattggag aatacgacgt
cctccgtggg
8881 atttttacca gtgagatagg aacaaagcaa atcactcaga gtgcattatt
agcagaagcc

8941 agaagtgatt attctgaagc tgctaagcag tatgatgagg ctctcaataa
acaagactgg
9001 gtagatggtg agcccacaga agccgagaag gatttttggg aacttgcatc
ccttgactgt
9061 tacaaccacc ttgctgagtg gaaatcactt gaatactgtt ctacagccag
tatagacagt
9121 gagaaccccc cagacctaaa taaaatctgg agtgaaccat tttatcagga
aacatatcta
9181 ccttacatga tccgcagcaa gctgaagctg ctgctccagg gagaggctga
ccagtccctg
9241 ctgacattta ttgacaaagc tatgcacggg gagctccaga aggcgattct
agagcttcat
9301 tacagtcaag agctgagtct gctttacctc ctgcaagatg atgttgacag
agccaaatat
9361 tacattcaaa atggcattca gagttttatg cagaattatt ctagtattga
tgtcctctta
9421 caccaaagta gactcaccaa attgcagtct gtacaggctt taacagaaat
tcaggagttc
9481 atcagcttta taagcaaaca aggcaattta tcatctcaag ttccccttaa
gagacttctg
9541 aacacctgga caaacagata tccagatgct aaaatggacc caatgaacat
ctgggatgac
9601 atcatcacaa atcgatgttt ctttctcagc aaaatagagg agaagcttac
ccctcttcca
9661 gaagataata gtatgaatgt ggatcaagat ggagacccca gtgacaggat
ggaagtgcaa
9721 gagcaggaag aagatatcag ctccctgatc aggagttgca agttttccat
gaaaatgaag
9781 atgatagaca gtgcccggaa gcagaacaat ttctcacttg ctatgaaact
actgaaggag
9841 ctgcataaag agtcaaaaac cagagacgat tggctggtga gctgggtgca
gagctactgc
9901 cgcctgagcc actgccggag ccggtcccag ggctgctctg agcaggtgct
cactgtgctg
9961 aaaacagtct ctttgttgga tgagaacaac gtgtcaagct acttaagcaa
aaatattctg
10021 gctttccgtg accagaacat tctcttgggt acaacttaca ggatcatagc
gaatgctctc
10081 agcagtgagc cagcctgcct tgctgaaatc gaggaggaca aggctagaag
aatcttagag
10141 ctttctggat ccagttcaga ggattcagag aaggtgatcg cgggtctgta
ccagagagca
10201 ttccagcacc tctctgaggc tgtgcaggcg gctgaggagg aggcccagcc
tccctcctgg
10261 agctgtgggc ctgcagctgg ggtgattgat gcttacatga cgctggcaga
tttctgtgac
10321 caacagctgc gcaaggagga agagaatgca tcagttattg attctgcaga
actgcaggcg
10381 tatccagcac ttgtggtgga gaaaatgttg aaagctttaa aattaaattc
caatgaagcc
10441 agattgaagt ttcctagatt acttcagatt atagaacggt atccagagga
gactttgagc
10501 ctcatgacaa aagagatctc ttccgttccc tgctggcagt tcatcagctg
gatcagccac
10561 atggtggcct tactggacaa agaccaagcc gttgctgttc agcactctgt
ggaagaaatc
10621 actgataact acccgcaggc tattgtttat cccttcatca taagcagcga
aagctattcc
10681 ttcaaggata cttctactgg tcataagaat aaggagtttg tggcaaggat
taaaagtaag

10741 ttggatcaag gaggagtgat tcaagatttt attaatgcct tagatcagct
ctctaatcct
10801 gaactgctct ttaaggattg gagcaatgat gtaagagctg aactagcaaa
aacccctgta
10861 aataaaaaaa acattgaaaa aatgtatgaa agaatgtatg cagccttggg
tgacccaaag
10921 gctccaggcc tgggggcctt tagaaggaag tttattcaga cttttggaaa
agaatttgat
10981 aaacattttg ggaaaggagg ttctaaacta ctgagaatga agctcagtga
cttcaacgac
11041 attaccaaca tgctactttt aaaaatgaac aaagactcaa agccccctgg
gaatctgaaa
11101 gaatgttcac cctggatgag cgacttcaaa gtggagttcc tgagaaatga
gctggagatt
11161 cccggtcagt atgacggtag gggaaagcca ttgccagagt accacgtgcg
aatcgccggg
11221 tttgatgagc gggtgacagt catggcgtct ctgcgaaggc ccaagcgcat
catcatccgt
11281 ggccatgacg agagggaaca ccctttcctg gtgaagggtg gcgaggacct
gcggcaggac
11341 cagcgcgtgg agcagctctt ccaggtcatg aatgggatcc tggcccaaga
ctccgcctgc
11401 agccagaggg ccctgcagct gaggacctat agcgttgtgc ccatgacctc
cagtgatccc
11461 agggcaccgc cgtgtgaata taaagattgg ctgacaaaaa tgtcaggaaa
acatgatgtt
11521 ggagcttaca tgctaatgta taagggcgct aatcgtactg aaacagtcac
gtcttttaga
11581 aaacgagaaa gtaaagtgcc tgctgatctc ttaaagcggg ccttcgtgag
gatgagtaca
11641 agccctgagg ctttcctggc gctccgctcc cacttcgcca gctctcacgc
tctgatatgc
11701 atcagccact ggatcctcgg gattggagac agacatctga acaactttat
ggtggccatg
11761 gagactggcg gcgtgatcgg gatcgacttt gggcatgcgt ttggatccgc
tacacagttt
11821 ctgccagtcc ctgagttgat gccttttcgg ctaactcgcc agtttatcaa
tctgatgtta
11881 ccaatgaaag aaacgggcct tatgtacagc atcatggtac acgcactccg
ggccttccgc
11941 tcagaccctg gcctgctcac caacaccatg gatgtgtttg tcaaggagcc
ctcctttgat
12001 tggaaaaatt ttgaacagaa aatgctgaaa aaaggagggt catggattca
agaaataaat
12061 gttgctgaaa aaaattggta cccccgacag aaaatatgtt acgctaagag
aaagttagca
12121 ggtgccaatc cagcagtcat tacttgtgat gagctactcc tgggtcatga
gaaggcccct
12181 gccttcagag actatgtggc tgtggcacga ggaagcaaag atcacaacat
tcgtgcccaa
12241 gaaccagaga gtgggctttc agaagagact caagtgaagt gcctgatgga
ccaggcaaca
12301 gaccccaaca tccttggcag aacctgggaa ggatgggagc cctggatgtg
aggtctgtgg
12361 gagtctgcag atagaaagca ttacattgtt taaagaatct actatacttt
ggttggcagc
12421 attccatgag ctgattttcc tgaaacacta aagagaaatg tcttttgtgc
tacagtttcg
12481 tagcatgagt ttaaatcaag attatgatga gtaaatgtgt atgggttaaa
tcaaagataa

```
    12541 ggttatagta acatcaaaga ttaggtgagg tttatagaaa gatagatatc
caggcttacc
    12601 aaagtattaa gtcaagaata taatatgtga tcagctttca aagcatttac
aagtgctgca
    12661 agttagtgaa acagctgtct ccgtaaatgg aggaaatgtg gggaagcctt
ggaatgccct
    12721 tctggttctg gcacattgga aagcacactc agaaggcttc atcaccaaga
ttttgggaga
    12781 gtaaagctaa gtatagttga tgtaacattg tagaagcagc ataggaacaa
taagaacaat
    12841 aggtaaagct ataattatgg cttatattta gaaatgactg catttgatat
tttaggatat
    12901 ttttctaggt tttttccttt cattttattc tcttctagtt ttgacatttt
atgatagatt
    12961 tgctctctag aaggaaacgt ctttatttag gagggcaaaa attttggtca
tagcattcac
    13021 ttttgctatt ccaatctaca actggaagat acataaaagt gctttgcatt
gaatttggga
    13081 taacttcaaa aatcccatgg ttgttgttag ggatagtact aagcatttca
gttccaggag
    13141 aataaaagaa attcctattt gaaatgaatt cctcatttgg aggaaaaaaa
gcatgcattc
    13201 tagcacaaca agatgaaatt atggaataca aaagtggctc cttcccatgt
gcagtccctg
    13261 tccccccccg ccagtcctcc acacccaaac tgtttctgat tggcttttag
ctttttgttg
    13321 ttttttttt tccttctaac acttgtattt ggaggctctt ctgtgatttt
gagaagtata
    13381 ctcttgagtg tttaataaag tttttttcca aaagta
      //
```

Protein sequence:

NCBI Reference Sequence: NP_001075109.1

LOCUS: NP_001075109

ACCESSION: NP_001075109

```
      1 magsgagvrc sllrlqetls aadrcgaala ghqlirglgq ecvlssspav
lalqtslvfs
       61 rdfgllvfvr kslnsiefre creeilkflc iflekmgqki apysveiknt
ctsvytkdra
      121 akckipaldl likllqtfrs srlmdefkig elfskfygel alkkkipdtv
lekvyellgl
      181 lgevhpsemi nnaenlfraf lgelktqmts avrepklpvl agclkglssl
lcnftksmee
      241 dpqtsreifn fvlkairpqi dlkryavpsa glrlfalhas qfstclldny
vslfevllkw
      301 cahtnvelkk aalsalesfl kqvsnmvakn aemhknklqy fmeqfygiir
nvdsnnkels
      361 iairgyglfa gpckvinakd vdfmyveliq rckqmfltqt dtgddrvyqm
psflqsvasv
      421 llyldtvpev ytpvlehlvv mqidsfpqys pkmqlvccra ivkvflalaa
kgpvlrncis
      481 tvvhqgliri cskpvvlpkg pesesedhra sgevrtgkwk vptykdyvdl
frhllssdqm
```

541 mdsiladeaf fsvnsssesl nhllydefvk svlkivekld ltleiqtvge
qengdeapgv
601 wmiptsdpaa nlhpakpkdf safinlvefc reilpekqae ffepwvysfs
yelilqstrl
661 plisgfykll sitvrnakki kyfegvspks lkhspedpek yscfalfvkf
gkevavkmkq
721 ykdellascl tfllslphni ieldvrayvp alqmafklgl sytplaevgl
naleewsiyi
781 drhvmqpyyk dilpcldgyl ktsalsdetk nnwevsalsr aaqkgfnkvv
lkhlkktknl
841 ssneaislee irirvvqmlg slggqinknl ltvtssdemm ksyvawdrek
rlsfavpfre
901 mkpvifldvf lprvtelalt asdrqtkvaa cellhsmvmf mlgkatqmpe
ggqgappmyq
961 lykrtfpvll rlacdvdqvt rqlyeplvmq lihwftnnkk fesqdtvall
eaildgivdp
1021 vdstlrdfcg rcireflkws ikqitpqqqe kspvntkslf krlyslalhp
nafkrlgasl
1081 afnniyrefr eeeslveqfv fealviymes lalahadeks lgtiqqccda
idhlcriiek
1141 khvslnkakk rrlprgfpps aslclldlvk wllahcgrpq tecrhksiel
fykfvpllpg
1201 nrspnlwlkd vlkeegvsfl intfegggcg qpsgilaqpt llylrgpfsl
qatlcwldll
1261 laalecyntf igertvgalq vlgteaqssl lkavaffles iamhdiiaae
kcfgtgaagn
1321 rtspqegery nyskctvvvr imeftttlln tspegwkllk kdlcnthlmr
vlvqtlcepa
1381 sigfnigdvq vmahlpdvcv nlmkalkmsp ykdilethlr ekitaqsiee
lcavnlygpd
1441 aqvdrsrlaa vvsackqlhr agllhnilps qstdlhhsvg tellslvykg
iapgderqcl
1501 psldlsckql asgllelafa fgglcerlvs lllnpavlst aslgssqgsv
ihfshgeyfy
1561 slfsetinte llknldlavl elmqssvdnt kmvsavlngm ldqsfreran
qkhqglklat
1621 tilqhwkkcd swwakdsple tkmavlalla kilqidssvs fntshgsfpe
vfttyislla
1681 dtkldlhlkg qavtllpfft sltggsleel rrvleqliva hfpmqsrefp
pgtprfnnyv
1741 dcmkkfldal elsqspmlle lmtevlcreq qhvmeelfqs sfrriarrgs
cvtqvglles
1801 vyemfrkddp rlsftrqsfv drslltllwh csldalreff stivvdaidv
lksrftklne
1861 stfdtqitkk mgyykildvm ysrlpkddvh akeskinqvf hgscitegne
ltktliklcy
1921 daftenmage nqllerrrly hcaayncais viccvfnelk fyqgflfsek
peknllifen
1981 lidlkrrynf pvevevpmer kkkyieirke areaangdsd gpsymsslsy
ladstlseem
2041 sqfdfstgvq sysyssqdpr patgrfrrre qrdptvhddv lelemdelnr
hecmapltal
2101 vkhmhrslgp pqgeedsvpr dlpswmkflh gklgnpivpl nirlflaklv
inteevfrpy
2161 akhwlspllq laasenngge gihymvveiv atilswtgla tptgvpkdev
lanrllnflm
2221 khvfhpkrav frhnleiikt lvecwkdcls ipyrlifekf sgkdpnskdn
svgiqllgiv
2281 mandlppydp qcgiqsseyf qalvnnmsfv rykevyaaaa evlglilryv
merknilees

```
     2341 lcelvakqlk qhqntmedkf ivclnkvtks fppladrfmn avffllpkfh
gvlktlclev
     2401 vlcrvegmte lyfqlkskdf vqvmrhrdde rqkvcldiiy kmmpklkpve
lrellnpvve
     2461 fvshpsttcr eqmynilmwi hdnyrdpese tdndsqeifk lakdvliqgl
idenpglqli
     2521 irnfwshetr lpsntldrll alnslyspki evhflslatn fllemtsmsp
dypnpmfehp
     2581 lsecefqeyt idsdwrfrst vltpmfvetq asqgtlqtrt qegslsarwp
vagqiratqq
     2641 qhdftltqta dgrssfdwlt gsstdplvdh tspssdsllf ahkrserlqr
aplksvgpdf
     2701 gkkrlglpgd evdnkvkgaa grtdllrlrr rfmrdqekls lmyarkgvae
qkrekeikse
     2761 lkmkqdaqvv lyrsyrhgdl pdiqikhssl itplqavaqr dpiiakqlfs
slfsgilkem
     2821 dkfktlsekn nitqkllqdf nrflnttfsf fppfvsciqd iscqhaalls
ldpaavsagc
     2881 laslqqpvgi rlleeallrl lpaelpakrv rgkarlppdv lrwvelakly
rsigeydvlr
     2941 giftseigtk qitqsallae arsdyseaak qydealnkqd wvdgepteae
kdfwelasld
     3001 cynhlaewks leycstasid senppdlnki wsepfyqety lpymirsklk
lllqgeadqs
     3061 lltfidkamh gelqkailel hysqelslly llqddvdrak yyiqngiqsf
mqnyssidvl
     3121 lhqsrltklq svqalteiqe fisfiskqgn lssqvplkrl lntwtnrypd
akmdpmniwd
     3181 diitnrcffl skieekltpl pednsmnvdq dgdpsdrmev qeqeedissl
irsckfsmkm
     3241 kmidsarkqn nfslamkllk elhkesktrd dwlvswvqsy crlshcrsrs
qgcseqvltv
     3301 lktvsllden nvssylskni lafrdqnill gttyriiana lssepaclae
ieedkarril
     3361 elsgssseds ekviaglyqr afqhlseavq aaeeeaqpps wscgpaagvi
daymtladfc
     3421 dqqlrkeeen asvidsaelq aypalvvekm lkalklnsne arlkfprllq
iierypeetl
     3481 slmtkeissv pcwqfiswis hmvalldkdq avavqhsvee itdnypqaiv
ypfiissesy
     3541 sfkdtstghk nkefvariks kldqggviqd finaldqlsn pellfkdwsn
dvraelaktp
     3601 vnkkniekmy ermyaalgdp kapglgafrr kfiqtfgkef dkhfgkggsk
llrmklsdfn
     3661 ditnmlllkm nkdskppgnl kecspwmsdf kveflrnele ipgqydgrgk
plpeyhvria
     3721 gfdervtvma slrrpkriii rghderehpf lvkggedlrq dqrveqlfqv
mngilaqdsa
     3781 csqralqlrt ysvvpmtssd prappceykd wltkmsgkhd vgaymlmykg
anrtetvtsf
     3841 rkreskvpad llkrafvrms tspeaflalr shfasshali cishwilgig
drhlnnfmva
     3901 metggvigid fghafgsatq flpvpelmpf rltrqfinlm lpmketglmy
simvhalraf
     3961 rsdpglltnt mdvfvkepsf dwknfeqkml kkggswiqei nvaeknwypr
qkicyakrkl
     4021 aganpavitc delllgheka pafrdyvava rgskdhnira qepesglsee
tqvkclmdqa
     4081 tdpnilgrtw egwepwm
      //
```

PSMD6

Official Symbol: **PSMD6**

Official Name: proteasome (prosome, macropain) 26S subunit, non-ATPase, 6

Gene ID: 9861

Organism: Homo sapiens

Other Aliases: Rpn7, S10, SGA-113M, p44S10

Other Designations: 26S proteasome non-ATPase regulatory subunit 6; 26S proteasome regulatory subunit RPN7; 26S proteasome regulatory subunit S10; breast cancer-associated protein SGA-113M; p42A; phosphonoformate immuno-associated protein 4; proteasome regulatory particle subunit p44S10

Nucleotide sequence:

NCBI Reference Sequence: NM_014814.1

LOCUS: NM_014814

ACCESSION : NM_014814

```
      1 gtcagccgct gtcccttag ccgcgatgcc gctggagaac ctggaggagg
agggtctgcc
       61 caagaacccc gacttgcgta tcgcgcagct gcgcttcctg ctcagcctgc
ccgagcaccg
      121 cggagacgct gccgtgcgcg acgagctgat ggcggccgtc cgcgataaca
acatggctcc
      181 ttactatgaa gccttgtgca aatccctcga ctggcagata gacgtggacc
tactcaataa
      241 aatgaagaag gcaaatgaag atgagttgaa gcgtttggat gaggagctgg
aagatgcaga
      301 gaagaatcta ggagagagcg aaattcgcga tgcaatgatg gcaaaggccg
agtacctctg
      361 ccggataggt gacaaagagg gagctctgac agcctttcgc aagacatatg
acaaaactgt
      421 ggccctgggt caccgattgg atattgtatt ctatctcctt aggattggct
tattttatat
      481 ggataatgat ctcatcacac gaaacacaga aaaggccaaa agcttaatag
aagaaggagg
      541 agactgggac aggagaaacc gcctaaaagt gtatcagggt ctttattgtg
tggctattcg
      601 tgatttcaaa caggcagctg aactcttcct tgacactgtt tcaacattta
catcctatga
      661 actcatggat tataaaacat ttgtgactta tactgtctat gtcagtatga
ttgccttaga
      721 aagaccagat ctcagggaaa aggtcattaa aggagcagag attcttgaag
tgttgcacag
```

```
      781 tcttccagca gttcggcagt atctgttttc actctatgaa tgccgttact
ctgttttctt
      841 ccaatcatta gcggttgtgg aacaggaaat gaaaaaggac tggctttttg
cccctcatta
      901 tcgatactat gtaagagaaa tgagaattca tgcatacagt cagctgctgg
aatcatatag
      961 gtcattaacc cttggctata tggcagaagc gtttggtgtt ggtgtggaat
tcattgatca
     1021 ggaactgtcc aggtttattg ctgccgggag actacactgc aaaatagata
aagtgaatga
     1081 aatagtagaa accaacagac ctgatagcaa gaactggcag taccaagaaa
ctatcaagaa
     1141 aggagatctg ctactaaaca gagttcaaaa actttccaga gtaattaata
tgtaaagcca
     1201 tgtaactaac aaaggatttg ctttagagat aattatttgg aatttttata
gcttacttca
     1261 caatgtgccc aggtcagctg tataaaataa atactgcatt gttgtttc
      //
```

Protein sequence:

NCBI Reference Sequence: NP_055629.1

LOCUS: NP_055629

ACCESSION: NP_055629

```
      1 mplenleeeg lpknpdlria qlrfllslpe hrgdaavrde lmaavrdnnm
apyyealcks
       61 ldwqidvdll nkmkkanede lkrldeeled aeknlgesei rdammakaey
lcrigdkega
      121 ltafrktydk tvalghrldi vfyllriglf ymdndlitrn tekaksliee
ggdwdrrnrl
      181 kvyqglycva irdfkqaael fldtvstfts yelmdyktfv tytvyvsmia
lerpdlrekv
      241 ikgaeilevl hslpavrqyl fslyecrysv ffqslavveq emkkdwlfap
hyryyvremr
      301 ihaysqlles yrsltlgyma eafgvgvefi dqelsrfiaa grlhckidkv
neivetnrpd
      361 sknwqyqeti kkgdlllnrv qklsrvinm
      //
```

ITGB1

Official Symbol: **ITGB1**

Official Name: integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) [Homo sapiens]

Gene ID: 3688

Organism: Homo sapiens

Other Aliases: RP11-479G22.2, CD29, FNRB, GPIIA, MDF2, MSK12, VLA-BETA, VLAB

Other Designations: integrin VLA-4 beta subunit; integrin beta-1; very late activation protein, beta polypeptide

Nucleotide sequence:

NCBI Reference Sequence: NM_002211.3

LOCUS: NM_002211

ACCESSION : NM_002211

```
        1 atcagacgcg cagaggaggc ggggccgcgg ctggtttcct gccgggggggc
ggctctgggc
       61 cgccgagtcc cctcctcccg cccctgagga ggaggagccg ccgccacccg
ccgcgcccga
      121 cacccgggag gccccgccag cccgcgggag aggcccagcg ggagtcgcgg
aacagcaggc
      181 ccgagcccac cgcgccgggc cccggacgcc gcgcggaaaa gatgaattta
caaccaattt
      241 tctggattgg actgatcagt tcagtttgct gtgtgtttgc tcaaacagat
gaaaatagat
      301 gtttaaaagc aaatgccaaa tcatgtggag aatgtataca agcagggcca
aattgtgggt
      361 ggtgcacaaa ttcaacattt ttacaggaag gaatgcctac ttctgcacga
tgtgatgatt
      421 tagaagcctt aaaaaagaag ggttgccctc cagatgacat agaaaatccc
agaggctcca
      481 aagatataaa gaaaaataaa aatgtaacca accgtagcaa aggaacagca
gagaagctca
      541 agccagagga tattactcag atccaaccac agcagttggt tttgcgatta
agatcagggg
      601 agccacagac atttacatta aaattcaaga gagctgaaga ctatcccatt
gacctctact
      661 accttatgga cctgtcttac tcaatgaaag acgatttgga gaatgtaaaa
agtcttggaa
      721 cagatctgat gaatgaaatg aggaggatta cttcggactt cagaattgga
tttggctcat
      781 ttgtggaaaa gactgtgatg ccttacatta gcacaacacc agctaagctc
aggaaccctt
      841 gcacaagtga acagaactgc accagcccat ttagctacaa aaatgtgctc
agtcttacta
      901 ataaaggaga agtatttaat gaacttgttg gaaaacagcg catatctgga
aatttggatt
      961 ctccagaagg tggtttcgat gccatcatgc aagttgcagt ttgtggatca
ctgattggct
     1021 ggaggaatgt tacacggctg ctggtgtttt ccacagatgc cgggtttcac
tttgctggag
     1081 atgggaaact tggtggcatt gttttaccaa atgatggaca atgtcacctg
gaaaataata
     1141 tgtacacaat gagccattat tatgattatc cttctattgc tcaccttgtc
cagaaactga
     1201 gtgaaaataa tattcagaca atttttgcag ttactgaaga atttcagcct
gtttacaagg
     1261 agctgaaaaa cttgatccct aagtcagcag taggaacatt atctgcaaat
tctagcaatg
```

1321 taattcagtt gatcattgat gcatacaatt ccctttcctc agaagtcatt
ttggaaaacg
1381 gcaaattgtc agaaggcgta acaataagtt acaaatctta ctgcaagaac
ggggtgaatg
1441 gaacagggga aaatggaaga aaatgttcca atatttccat tggagatgag
gttcaatttg
1501 aaattagcat aacttcaaat aagtgtccaa aaaaggattc tgacagcttt
aaaattaggc
1561 ctctgggctt tacggaggaa gtagaggtta ttcttcagta catctgtgaa
tgtgaatgcc
1621 aaagcgaagg catccctgaa agtcccaagt gtcatgaagg aaatgggaca
tttgagtgtg
1681 gcgcgtgcag gtgcaatgaa gggcgtgttg gtagacattg tgaatgcagc
acagatgaag
1741 ttaacagtga agacatggat gcttactgca ggaaagaaaa cagttcagaa
atctgcagta
1801 acaatggaga gtgcgtctgc ggacagtgtg tttgtaggaa gagggataat
acaaatgaaa
1861 tttattctgg caaattctgc gagtgtgata atttcaactg tgatagatcc
aatggcttaa
1921 tttgtggagg aaatggtgtt tgcaagtgtc gtgtgtgtga gtgcaacccc
aactacactg
1981 gcagtgcatg tgactgttct ttggatacta gtacttgtga agccagcaac
ggacagatct
2041 gcaatggccg gggcatctgc gagtgtggtg tctgtaagtg tacagatccg
aagtttcaag
2101 ggcaaacgtg tgagatgtgt cagacctgcc ttggtgtctg tgctgagcat
aaagaatgtg
2161 ttcagtgcag agccttcaat aaaggagaaa agaaagacac atgcacacag
gaatgttcct
2221 attttaacat taccaaggta gaaagtcggg acaaattacc ccagccggtc
caacctgatc
2281 ctgtgtccca ttgtaaggag aaggatgttg acgactgttg gttctatttt
acgtattcag
2341 tgaatgggaa caacgaggtc atggttcatg ttgtggagaa tccagagtgt
cccactggtc
2401 cagacatcat tccaattgta gctggtgtgg ttgctggaat tgttcttatt
ggccttgcat
2461 tactgctgat atggaagctt ttaatgataa ttcatgacag aagggagttt
gctaaatttg
2521 aaaaggagaa aatgaatgcc aaatgggaca cgggtgaaaa tcctatttat
aagagtgccg
2581 taacaactgt ggtcaatccg aagtatgagg gaaaatgagt actgcccgtg
caaatcccac
2641 aacactgaat gcaaagtagc aatttccata gtcacagtta ggtagcttta
gggcaatatt
2701 gccatggttt tactcatgtg caggttttga aaatgtacaa tatgtataat
ttttaaaatg
2761 ttttattatt ttgaaaataa tgttgtaatt catgccaggg actgacaaaa
gacttgagac
2821 aggatggtta ctcttgtcag ctaaggtcac attgtgcctt tttgaccttt
tcttcctgga
2881 ctattgaaat caagcttatt ggattaagtg atatttctat agcgattgaa
agggcaatag
2941 ttaaagtaat gagcatgatg agagtttctg ttaatcatgt attaaaactg
atttttagct
3001 ttacaaatat gtcagtttgc agttatgcag aatccaaagt aaatgtcctg
ctagctagtt
3061 aaggattgtt ttaaatctgt tattttgcta tttgcctgtt agacatgact
gatgacatat

```
     3121 ctgaaagaca agtatgttga gagttgctgg tgtaaaatac gtttgaaata
gttgatctac
     3181 aaaggccatg ggaaaaattc agagagttag gaaggaaaaa ccaatagctt
taaaacctgt
     3241 gtgccatttt aagagttact taatgtttgg taacttttat gccttcactt
tacaaattca
     3301 agccttagat aaaagaaccg agcaattttc tgctaaaaag tccttgattt
agcactattt
     3361 acatacaggc catactttac aaagtatttg ctgaatgggg acctttttgag
ttgaatttat
     3421 tttattattt ttattttgtt taatgtctgg tgctttctgt cacctcttct
aatcttttaa
     3481 tgtatttgtt tgcaattttg gggtaagact ttttttatga gtacttttttc
tttgaagttt
     3541 tagcggtcaa tttgcctttt taatgaacat gtgaagttat actgtggcta
tgcaacagct
     3601 ctcacctacg cgagtcttac tttgagttag tgccataaca gaccactgta
tgtttacttc
     3661 tcaccatttg agttgcccat cttgtttcac actagtcaca ttcttgtttt
aagtgccttt
     3721 agttttaaca gttcactttt tacagtgcta tttactgaag ttatttatta
aatatgccta
     3781 aaatacttaa atcggatgtc ttgactctga tgtattttat caggttgtgt
gcatgaaatt
     3841 tttatagatt aaagaagttg aggaaaagca aaaaaaaaa
            //
```

Protein sequence:

NCBI Reference Sequence: NP_002202.2

LOCUS: NP_002202

ACCESSION: NP_002202

```
       1 mnlqpifwig lissvccvfa qtdenrclka nakscgeciq agpncgwctn
stflqegmpt
      61 sarcddleal kkkgcppddi enprgskdik knknvtnrsk gtaeklkped
itqiqpqqlv
     121 lrlrsgepqt ftlkfkraed ypidlyylmd lsysmkddle nvkslgtdlm
nemrritsdf
     181 rigfgsfvek tvmpyisttp aklrnpctse qnctspfsyk nvlsltnkge
vfnelvgkqr
     241 isgnldspeg gfdaimqvav cgsligwrnv trllvfstda gfhfagdgkl
ggivlpndgq
     301 chlennmytm shyydypsia hlvqklsenn iqtifavtee fqpvykelkn
lipksavgtl
     361 sanssnviql iidaynslss evilengkls egvtisyksy ckngvngtge
ngrkcsnisi
     421 gdevqfeisi tsnkcpkkds dsfkirplgf teevevilqy icececqseg
ipespkcheg
     481 ngtfecgacr cnegrvgrhc ecstdevnse dmdaycrken sseicsnnge
cvcgqcvcrk
     541 rdntneiysg kfcecdnfnc drsnglicgg ngvckcrvce cnpnytgsac
dcsldtstce
     601 asngqicngr gicecgvckc tdpkfqgqtc emcqtclgvc aehkecvqcr
afnkgekkdt
     661 ctqecsyfni tkvesrdklp qpvqpdpvsh ckekdvddcw fyftysvngn
nevmvhvven
     721 pecptgpdii pivagvvagi vliglallli wkllmiihdr refakfekek
mnakwdtgen
     781 piyksavttv vnpkyegk
```

MYH10

Official Symbol: **MYH10**

Official Name: myosin, heavy chain 10, non-muscle [Homo sapiens]

Gene ID: 4628

Organism: Homo sapiens

Other Aliases: NMMHC-IIB, NMMHCB

Other Designations: cellular myosin heavy chain, type B; myosin heavy chain, nonmuscle type B; myosin, heavy polypeptide 10, non-muscle; myosin-10; nonmuscle myosin II heavy chain-B; nonmuscle myosin heavy chain IIB

Nucleotide sequence:

NCBI Reference Sequence: NM_001256012.1

LOCUS: NM_001256012

ACCESSION : NM_001256012

```
        1 agcagtgcta aaggagcccg gcggaggcag cggtgggttt gggactgagg
cgctggatct
       61 gtggtcgcgg ctggggacgt gcgcccgcgc caccatcttc ggctgaagag
gcaattgctt
      121 ttggatcgtt ccatttacaa tggcgcagag aactggactc gaggatccag
agaggtatct
      181 ctttgtggac agggctgtca tctacaaccc tgccactcaa gctgattgga
cagctaaaaa
      241 gctagtgtgg attccatcag aacgccatgg ttttgaggca gctagtatca
aagaagaacg
      301 gggagatgaa gttatggtgg agttggcaga gaatggaaag aaagcaatgg
tcaacaaaga
      361 tgatattcag aagatgaacc cacctaagtt ttccaaggtg gaggatatgg
cagaattgac
      421 atgcttgaat gaagcttccg ttttacataa tctgaaggat cgctactatt
caggactaat
      481 ctatacttat tctggactct tctgtgtagt tataaaccct tacaagaatc
ttccaattta
      541 ctctgagaat attattgaaa tgtacagagg gaagaagcgt catgagatgc
ctccacacat
      601 ctatgctata tctgaatctg cttacagatg catgcttcaa gatcgtgagg
accagtcaat
      661 tctttgcacg ggtgagtcag gtgctgggaa gacagaaaat acaaagaaag
ttattcagta
      721 ccttgcccat gttgcttctt cacataaagg aagaaaggac cataatattc
ctcaggaatc
      781 gcctaaacca gtgaaacacc agggggaact tgaacggcag cttttgcaag
caaatccaat
      841 tctggaatca tttggaaatg cgaagactgt gaaaaatgat aactcatctc
gttttggcaa
      901 atttattcgg atcaactttg atgtaactgg ctatatcgtt ggggccaaca
ttgaaacata
```

961 ccttctggaa aagtctcgtg ctgttcgtca agcaaaagat gaacgtactt
ttcatatctt
1021 ttaccagttg ttatctggag caggagaaca cctaaagtct gatttgcttc
ttgaaggatt
1081 taataactac aggtttctct ccaatggcta tattcctatt ccgggacagc
aagacaaaga
1141 taatttccag gagaccatgg aagcaatgca cataatgggc ttctcccatg
aagagattct
1201 gtcaatgctt aaagtagtat cttcagtgct acagtttgga aatatttctt
tcaaaaagga
1261 gagaaatact gatcaagctt ccatgccaga aaatacagtt gcgcagaagc
tctgccatct
1321 tcttgggatg aatgtgatgg agtttactcg ggccatcctg actccccgga
tcaaggtcgg
1381 ccgagactat gtgcaaaaag cccagaccaa agaacaggca gattttgcag
tagaagcatt
1441 ggcaaaagct acctatgagc ggctctttcg ctggctcgtt catcgcatca
ataaagctct
1501 ggataggacc aaacgtcagg gagcatcttt cattggaatc ctggatattg
ctggatttga
1561 aatttttgag ctgaactcct ttgaacaact ttgcatcaac tacaccaatg
agaagctgca
1621 gcagctgttc aaccacacca tgtttatcct agaacaagag gaataccagc
gcgaaggcat
1681 cgagtggaac ttcatcgatt tcgggctgga tctgcagcca tgcatcgacc
taatagagag
1741 acctgcgaac cctcctggtg tactggccct tttggatgaa gaatgctggt
tccctaaagc
1801 cacagataaa acctttgttg aaaaactggt tcaagagcaa ggttcccact
ccaagtttca
1861 gaaacctcga caattaaaag acaaagctga tttttgcatt atacattatg
cagggaaggt
1921 ggactataag gcagatgagt ggctgatgaa gaatatggac cccctgaatg
acaacgtggc
1981 cacccttttg caccagtcat cagacagatt tgtggcagag ctttggaaag
atgagattca
2041 gaatattcag agagcttctt tctatgacag tgtttctggt cttcatgagc
caccagtgga
2101 ccgtatcgtg ggtctggatc aagtcactgg tatgactgag acagcttttg
gctccgcata
2161 taaaaccaag aagggcatgt ttcgtaccgt tgggcaactc tacaaagaat
ctctcaccaa
2221 gctgatggca actctccgaa acaccaaccc taactttgtt cgttgtatca
ttccaaatca
2281 cgagaagagg gctggaaaat tggatccaca cctagtccta gatcagcttc
gctgtaatgg
2341 tgtcctggaa gggatccgaa tctgtcgcca gggcttccct aaccgaatag
ttttccagga
2401 attcagacag agatatgaga tcctaactcc aaatgctatt cctaaaggtt
ttatggatgg
2461 taaacaggcc tgtgaacgaa tgatccgggc tttagaattg gacccaaact
tgtacagaat
2521 tggacagagc aagatatttt tcagagctgg agttctggca cacttagagg
aagaaagaga
2581 tttaaaaatc accgatatca ttatcttctt ccaggccgtt tgcagaggtt
acctggccag
2641 aaaggccttt gccaagaagc agcagcaact aagtgcctta aaggtcttgc
agcggaactg
2701 tgccgcgtac ctgaaattac ggcactggca gtggtggcga gtcttcacaa
aggtgaagcc

2761 gcttctacaa gtgactcgcc aggaggaaga acttcaggcc aaagatgaag
agctgttgaa
2821 ggtgaaggag aagcagacga aggtggaagg agagctggag gagatggagc
ggaagcacca
2881 gcagctttta gaagagaaga atatccttgc agaacaacta caagcagaga
ctgagctctt
2941 tgctgaagca gaagagatga gggcaagact tgctgctaaa aagcaggaat
tagaagagat
3001 tctacatgac ttggagtcta gggttgaaga agaagaagaa agaaaccaaa
tcctccaaaa
3061 tgaaaagaaa aaaatgcaag cacatattca ggacctggaa gaacagctag
acgaggagga
3121 aggggctcgg caaaagctgc agctggaaaa ggtgacagca gaggccaaga
tcaagaagat
3181 ggaagaggag attctgcttc tcgaggacca aaattccaag ttcatcaaag
aaaagaaact
3241 catggaagat cgcattgctg agtgttcctc tcagctggct gaagaggaag
aaaaggcgaa
3301 aaacttggcc aaaatcagga ataagcaaga agtgatgatc tcagatttag
aagaacgctt
3361 aaagaaggaa gaaaagactc gtcaggaact ggaaaaggcc aaaagaaaac
tcgacgggga
3421 gacgaccgac ctgcaggacc agatcgcaga gctgcaggcg cagattgatg
agctcaagct
3481 gcagctggcc aagaaggagg aggagctgca gggcgcactg gccagaggtg
atgatgaaac
3541 actccataag aacaatgccc ttaaagttgt gcgagagcta caagcccaaa
ttgctgaact
3601 tcaggaagac tttgaatccg agaaggcttc acggaacaag gccgaaaagc
agaaaaggga
3661 cttgagtgag gaactggaag ctctgaaaac agagctggag gacacgctgg
acaccacggc
3721 agcccagcag gaactacgta caaaacgtga acaagaagtg gcagagctga
agaaagctct
3781 tgaggaggaa actaagaacc atgaagctca aatccaggac atgagacaaa
gacacgcaac
3841 agccctggag gagctctcag agcagctgga acaggccaag cggttcaaag
caaatctaga
3901 gaagaacaag cagggcctgg agacagataa caaggagctg gcgtgtgagg
tgaaggtcct
3961 gcagcaggtc aaggctgagt ctgagcacaa gaggaagaag ctcgacgcgc
aggtccagga
4021 gctccatgcc aaggtctctg aaggcgacag gctcagggtg gagctggcgg
agaaagcaag
4081 taagctgcag aatgagctag ataatgtctc cacccttctg gaagaagcag
agaagaaggg
4141 tattaaattt gctaaggatg cagctagtct tgagtctcaa ctacaggata
cacaggagct
4201 tcttcaggag gagacacgcc agaaactaaa cctgagcagt cggatccggc
agctggaaga
4261 ggagaagaac agtcttcagg agcagcagga ggaggaggag gaggccagga
agaacctgga
4321 gaagcaagtg ctggccctgc agtcccagtt ggctgatacc aagaagaaag
tagatgacga
4381 cctgggaaca attgaaagtc tggaagaagc caagaagaag cttctgaagg
acgcggaggc
4441 cctgagccag cgcctggagg agaaggcact ggcgtatgac aaactggaga
agaccaagaa
4501 ccgcctgcag caggagctgg acgacctcac ggtggacctg gaccaccagc
gccaggtcgc

4561 ctccaacttg gagaagaagc agaagaagtt tgaccagctg ttagcagaag
agaagagcat
4621 ctctgctcgc tatgccgaag agcgggaccg ggccgaagcc gaggccagag
agaaagaaac
4681 caaagccctg tcactggccc gggccctcga ggaagccctg gaggccaagg
aggagtttga
4741 gaggcagaac aagcagctcc gagcagacat ggaagacctc atgagctcca
aagatgatgt
4801 gggaaaaaac gttcacgaac ttgaaaaatc caaacgggcc ctagagcagc
aggtggagga
4861 aatgaggacc cagctggagg agctggaaga cgaactccag gccacggaag
atgccaagct
4921 tcgtctggag gtcaacatgc aggccatgaa ggcgcagttc gagagagacc
tgcaaaccag
4981 ggatgagcag aatgaagaga agaagcggct gctgatcaaa caggtgcggg
agctcgaggc
5041 ggagctggag gatgagagga aacagcgggc gcttgctgta gcttcaaaga
aaaagatgga
5101 gatagacctg aaggacctcg aagcccaaat cgaggctgcg aacaaagctc
gggatgaggt
5161 gattaagcag ctccgcaagc tccaggctca gatgaaggat taccaacgtg
aattagaaga
5221 agctcgtgca tccagagatg agatttttgc tcaatccaaa gagagtgaaa
agaaattgaa
5281 gagtctggaa gcagaaatcc ttcaattgca ggaggaactt gcctcatctg
agcgagcccg
5341 ccgacacgcc gagcaggaga gagatgagct ggcggacgag atcaccaaca
gcgcctctgg
5401 caagtccgcg ctgctggatg agaagcggcg tctggaagct cggatcgcac
agctggagga
5461 ggagctggaa gaggagcaga gcaacatgga gctgctcaac gaccgcttcc
gcaagaccac
5521 tctacaggtg gacacactga acgccgagct agcagccgag cgcagcgccg
cccagaagag
5581 tgacaatgca cgccagcaac tggagcggca gaacaaggag ctgaaggcca
agctgcagga
5641 actcgagggt gctgtcaagt ctaagttcaa ggccaccatc tcagccctgg
aggccaagat
5701 tgggcagctg gaggagcagc ttgagcagga agccaaggaa cgagcagccg
ccaacaaatt
5761 agtccgtcgc actgagaaga agctgaaaga aatcttcatg caggttgagg
atgagcgtcg
5821 acacgcggac cagtataaag agcagatgga gaaggccaac gctcggatga
agcagcttaa
5881 acgccagctg gaggaagcag aagaagaagc gacgcgtgcc aacgcatctc
ggcgtaaact
5941 ccagcgggaa ctggatgatg ccaccgaggc caacgagggc ctgagccgcg
aggtcagcac
6001 cctgaagaac cggctgaggc ggggtggccc catcagcttc tcttccagcc
gatctggccg
6061 gcgccagctg caccttgaag gagcttccct ggagctctcc gacgatgaca
cagaaagtaa
6121 gaccagtgat gtcaacgaga cgcagccacc ccagtcagag taaagttgca
ggaagccaga
6181 ggaggcaata cagtgggaca gttaggaatg cacccggggc ctcctgcaga
tttcggaaat
6241 tggcaagcta cgggattcct tcctgaaaga tcaactgtgt cttaaggctc
tccagcctat
6301 gcatactgta tcctgcttca gacttaggta caattgctcc ccttttttata
tatagacaca

```
     6361 cacaggacac atatattaaa cagattgttt catcattgca tctattttcc
atatagtcat
     6421 caagagacca ttttataaaa catggtaaga ccctttttaa aacaaactcc
aggcccttgg
     6481 ttgcgggtcg ctgggttatt ggggcagcgc cgtggtcgtc actcagtcgc
tctgcatgct
     6541 ctctgtcata cagacaggta acctagttct gtgttcacgt ggcccccgac
tcctcagcca
     6601 catcaagtct cctagaccac tgtggactct aaactgcact tgtctctctc
atttccttca
     6661 aataatgatc aatgctattt cagtgagcaa actgtgaaag gggctttgga
aagagtagga
     6721 ggggtgggct ggatcggaag caacacccat ttggggttac catgtccatc
ccccaagggg
     6781 ggccctgccc ctcgagtcga tggtgtcccg catctactca tgtgaactgg
ccttggcgag
     6841 ggctggtctg tgcatagaag ggatagtggc cacactgcag ctgaggcccc
aggtggcagc
     6901 catggatcat gtagacttcc agatggtctc ccgaaccgcc tggctctgcc
ggcgccctcc
     6961 tcacgtcagg agcaagcagc cgtggacccc taagccgagc tggtggaagg
ccctccccg
     7021 tcgccagccg ggccctcatg ctgaccttgc aaattcagcc gctgctttga
gcccaaaatg
     7081 ggaatattgg ttttgtgtcc gaggcttgtt ccaagtttgt caatgaggtt
tatggagcct
     7141 ccagaacaga tgccatcttc ctgaatgttg acatgccagt gggtgtgact
ccttcatttt
     7201 tccttctccc ttccctttgg acagtgttac agtgaacact tagcatcctg
tttttggttg
     7261 gtagttaagc aaactgacat tacggaaagt gccttagaca ctacagtact
aagacaatgt
     7321 tgaatatatc attcgcctct ataacaattt aatgtattca gttttgactg
tgcttcatat
     7381 catgtacctc tctagtcaaa gtggtattac agacattcag tgacaatgaa
tcagtgttaa
     7441 ttctaaatcc ttgatcctct gcaatgtgct tgaaaacaca aacctttttgg
gttaaaagct
     7501 ttaacatcta ttaggaagaa tttgtcctgt gggtttggaa tcttggattt
tccccttta
     7561 tgaactgtac tggctgttga ccaccagaca cctgaccgca aatatctttt
cttgtattcc
     7621 catatttcta gacaatgatt tttgtaagac aataaattta ttcattatag
atatttgcgc
     7681 ctgctctgtt tacttgaaga aaaaagcacc cgtggagaat aaagagacct
caataaacaa
     7741 gaataatcat gtgaacgtgg aaaaaaaaaa aaaaaaaa
            //
```

Protein sequence:

NCBI Reference Sequence: NP_001242941.1

LOCUS: NP_001242941

ACCESSION: NP_001242941

```
      1 maqrtgledp erylfvdrav iynpatqadw takklvwips erhgfeaasi
keergdevmv
     61 elaengkkam vnkddiqkmn ppkfskvedm aeltclneas vlhnlkdryy
sgliytysgl
```

```
      121 fcvvinpykn lpiyseniie myrgkkrhem pphiyaises ayrcmlqdre
dqsilctges
      181 gagktentkk viqylahvas shkgrkdhni pqespkpvkh qgelerqllq
anpilesfgn
      241 aktvkndnss rfgkfirinf dvtgyivgan ietylleksr avrqakdert
fhifyqllsg
      301 agehlksdll legfnnyrfl sngyipipgq qdkdnfqetm eamhimgfsh
eeilsmlkvv
      361 ssvlqfgnis fkkerntdqa smpentvaqk lchllgmnvm eftrailtpr
ikvgrdyvqk
      421 aqtkeqadfa vealakatye rlfrwlvhri nkaldrtkrq gasfigildi
agfeifelns
      481 feqlcinytn eklqqlfnht mfileqeeyq regiewnfid fgldlqpcid
lierpanppg
      541 vlalldeecw fpkatdktfv eklvqeqgsh skfqkprqlk dkadfciihy
agkvdykade
      601 wlmknmdpln dnvatllhqs sdrfvaelwk deiqniqras fydsvsglhe
ppvdrivgld
      661 qvtgmtetaf gsayktkkgm frtvgqlyke sltklmatlr ntnpnfvrci
ipnhekragk
      721 ldphlvldql rcngvlegir icrqgfpnri vfqefrqrye iltpnaipkg
fmdgkqacer
      781 miraleldpn lyrigqskif fragvlahle eerdlkitdi iiffqavcrg
ylarkafakk
      841 qqqlsalkvl qrncaaylkl rhwqwwrvft kvkpllqvtr qeeelqakde
ellkvkekqt
      901 kvegeleeme rkhqqlleek nilaeqlqae telfaeaeem rarlaakkqe
leeilhdles
      961 rveeeeernq ilqnekkkmq ahiqdleeql deeegarqkl qlekvtaeak
ikkmeeeill
     1021 ledqnskfik ekklmedria ecssqlaeee ekaknlakir nkqevmisdl
eerlkkeekt
     1081 rqelekakrk ldgettdlqd qiaelqaqid elklqlakke eelqgalarg
ddetlhknna
     1141 lkvvrelqaq iaelqedfes ekasrnkaek qkrdlseele alkteledtl
dttaaqqelr
     1201 tkreqevael kkaleeetkn heaqiqdmrq rhataleels eqleqakrfk
anleknkqgl
     1261 etdnkelace vkvlqqvkae sehkrkklda qvqelhakvs egdrlrvela
ekasklqnel
     1321 dnvstlleea ekkgikfakd aaslesqlqd tqellqeetr qklnlssrir
qleeeknslq
     1381 eqqeeeeear knlekqvlal qsqladtkkk vdddlgties leeakkkllk
daealsqrle
     1441 ekalaydkle ktknrlqqel ddltvdldhq rqvasnlekk qkkfdqllae
eksisaryae
     1501 erdraeaear eketkalsla raleealeak eeferqnkql radmedlmss
kddvgknvhe
     1561 lekskraleq qveemrtqle eledelqate daklrlevnm qamkaqferd
lqtrdeqnee
     1621 kkrllikqvr eleaeleder kqralavask kkmeidlkdl eaqieaanka
rdevikqlrk
     1681 lqaqmkdyqr eleearasrd eifaqskese kklksleaei lqlqeelass
erarrhaeqe
     1741 rdeladeitn sasgksalld ekrrlearia qleeeleeeq snmellndrf
rkttlqvdtl
     1801 naelaaersa aqksdnarqq lerqnkelka klqelegavk skfkatisal
eakigqleeq
     1861 leqeakeraa anklvrrtek klkeifmqve derrhadqyk eqmekanarm
kqlkrqleea
```

```
     1921 eeeatranas rrklqreldd ateaneglsr evstlknrlr rggpisfsss
rsgrrqlhle
     1981 gaslelsddd tesktsdvne tqppqse
      //
```

NCL

Official Symbol: **NCL**

Official Name: nucleolin [Homo sapiens]

Gene ID: 4691

Organism: Homo sapiens

Other Aliases: C23

Nucleotide sequence:

NCBI Reference Sequence: NM_005381.2

LOCUS: NM_005381

ACCESSION : NM_005381 XM_002342275

```
        1 ctttcgcctc agtctcgagc tctcgctggc cttcgggtgt acgtgctccg
ggatcttcag
       61 cacccgcggc cgccatcgcc gtcgcttggc ttcttctgga ctcatctgcg
ccacttgtcc
      121 gcttcacact ccgccgccat catggtgaag ctcgcgaagg caggtaaaaa
tcaaggtgac
      181 cccaagaaaa tggctcctcc tccaaaggag gtagaagaag atagtgaaga
tgaggaaatg
      241 tcagaagatg aagaagatga tagcagtgga gaagaggtcg tcatacctca
gaagaaaggc
      301 aagaaggctg ctgcaacctc agcaaagaag gtggtcgttt ccccaacaaa
aaaggttgca
      361 gttgccacac cagccaagaa agcagctgtc actccaggca aaaaggcagc
agcaacacct
      421 gccaagaaga cagttacacc agccaaagca gttaccacac ctggcaagaa
gggagccaca
      481 ccaggcaaag cattggtagc aactcctggt aagaagggtg ctgccatccc
agccaagggg
      541 gcaaagaatg gcaagaatgc caagaaggaa gacagtgatg aagaggagga
tgatgacagt
      601 gaggaggatg aggaggatga cgaggacgag gatgaggatg aagatgaaat
tgaaccagca
      661 gcgatgaaag cagcagctgc tgcccctgcc tcagaggatg aggacgatga
ggatgacgaa
      721 gatgatgagg atgacgatga cgatgaggaa gatgactctg aagaagaagc
tatggagact
      781 acaccagcca aaggaaagaa agctgcaaaa gttgttcctg tgaaagccaa
gaacgtggct
      841 gaggatgaag atgaagaaga ggatgatgag gacgaggatg acgacgacga
cgaagatgat
```

```
      901 gaagatgatg atgatgaaga tgatgaggag gaggaagaag aggaggagga
agagcctgtc
      961 aaagaagcac ctggaaaacg aaagaaggaa atggccaaac agaaagcagc
tcctgaagcc
     1021 aagaaacaga aagtggaagg cacagaaccg actacggctt tcaatctctt
tgttggaaac
     1081 ctaaacttta acaaatctgc tcctgaatta aaaactggta tcagcgatgt
ttttgctaaa
     1141 aatgatcttg ctgttgtgga tgtcagaatt ggtatgacta ggaaatttgg
ttatgtggat
     1201 tttgaatctg ctgaagacct ggagaaagcg ttggaactca ctggtttgaa
agtctttggc
     1261 aatgaaatta aactagagaa accaaaagga aaagacagta agaaagagcg
agatgcgaga
     1321 acacttttgg ctaaaaatct cccttacaaa gtcactcagg atgaattgaa
agaagtgttt
     1381 gaagatgctg cggagatcag attagtcagc aaggatggga aaagtaaagg
gattgcttat
     1441 attgaattta agacagaagc tgatgcagag aaaacctttg aagaaaagca
gggaacagag
     1501 atcgatgggc gatctatttc cctgtactat actggagaga aaggtcaaaa
tcaagactat
     1561 agaggtggaa agaatagcac ttggagtggt gaatcaaaaa ctctggtttt
aagcaacctc
     1621 tcctacagtg caacagaaga aactcttcag gaagtatttg agaaagcaac
ttttatcaaa
     1681 gtaccccaga accaaaatgg caaatctaaa gggtatgcat ttatagagtt
tgcttcattc
     1741 gaagacgcta aagaagcttt aaattcctgt aataaaaggg aaattgaggg
cagagcaatc
     1801 aggctggagt tgcaaggacc caggggatca cctaatgcca gaagccagcc
atccaaaact
     1861 ctgtttgtca aaggcctgtc tgaggatacc actgaagaga cattaaagga
gtcatttgac
     1921 ggctccgttc gggcaaggat agttactgac cgggaaactg ggtcctccaa
agggtttggt
     1981 tttgtagact tcaacagtga ggaggatgcc aaagctgcca aggaggccat
ggaagacggt
     2041 gaaattgatg gaaataaagt taccttggac tgggccaaac ctaagggtga
aggtggcttc
     2101 gggggtcgtg gtggaggcag aggcggcttt ggaggacgag gtggtggtag
aggaggccga
     2161 ggaggatttg gtggcagagg ccggggaggc tttggagggc gaggaggctt
ccgaggaggc
     2221 agaggaggag gaggtgacca caagccacaa ggaaagaaga cgaagtttga
atagcttctg
     2281 tccctctgct ttccctttc catttgaaag aaaggactct ggggttttta
ctgttacctg
     2341 atcaatgaca gagccttctg aggacattcc aagacagtat acagtcctgt
ggtctccttg
     2401 gaaatccgtc tagttaacat ttcaagggca ataccgtgtt ggttttgact
ggatattcat
     2461 ataaactttt taaagagttg agtgatagag ctaaccctta tctgtaagtt
ttgaatttat
     2521 attgtttcat cccatgtaca aaaccatttt ttcctacaaa tagtttgggt
tttgttgttg
     2581 tttctttttt ttgttttgtt tttgtttttt ttttttttgc gttcgtgggg
ttgtaaaaga
     2641 aaagaaagca gaatgtttta tcatggtttt tgcttcagcg gctttaggac
aaattaaaag
     2701 tcaactctgg tgccagaaaa aaaaaaaaaa aa
```

//

Protein sequence:

NCBI Reference Sequence: NP_005372.2

LOCUS: NP_005372

ACCESSION: NP_005372 XP_002342316

```
  1 mvklakagkn qgdpkkmapp pkeveedsed eemsedeedd ssgeevvipq
kkgkkaaats
 61 akkvvvsptk kvavatpakk aavtpgkkaa atpakktvtp akavttpgkk
gatpgkalva
121 tpgkkgaaip akgakngkna kkedsdeeed ddseedeedd ededededei
epaamkaaaa
181 apasededde ddeddedddd deeddseeea mettpakgkk aakvvpvkak
nvaededeee
241 ddededdddd eddeddded deeeeeeeee epvkeapgkr kkemakqkaa
peakkqkveg
301 tepttafnlf vgnlnfnksa pelktgisdv fakndlavvd vrigmtrkfg
yvdfesaedl
361 ekaleltglk vfgneiklek pkgkdskker dartllaknl pykvtqdelk
evfedaaeir
421 lvskdgkskg iayiefktea daektfeekq gteidgrsis lyytgekgqn
qdyrggknst
481 wsgesktlvl snlsysatee tlqevfekat fikvpqnqng kskgyafief
asfedakeal
541 nscnkreieg rairlelqgp rgspnarsqp sktlfvkgls edtteetlke
sfdgsvrari
601 vtdretgssk gfgfvdfnse edakaakeam edgeidgnkv tldwakpkge
ggfggrgggr
661 ggfggrgggr ggrggfggrg rggfggrggf rggrggggdh kpqgkktkfe
```

//

SEC61A1

Official Symbol: **SEC61A1**

Official Name: Sec61 alpha 1 subunit (S. cerevisiae) [Homo sapiens]

Gene ID: 29927

Organism: Homo sapiens

Other Aliases: HSEC61, SEC61, SEC61A

Other Designations: Sec61 alpha-1; protein transport protein SEC61 alpha subunit; protein transport protein Sec61 subunit alpha; protein transport protein Sec61 subunit alpha isoform 1; sec61

Nucleotide sequence:

NCBI Reference Sequence: NM_013336.3

LOCUS: NM_013336

ACCESSION : NM_013336 NM_015968

```
        1 agcgatccga ggcccggccc cggccccgcc ccgcgccgcg ccgcgccgct
tgccgccggg
       61 ctagcactga cgtgtctctc ggcggagctg ctgtgcagtg gaacgcgctg
ggccgcgggc
      121 agcgtcgcct cacgcggagc agagctgagc tgaagcggga cccggagccc
gagcagccgc
      181 cgccatggca atcaaatttc tggaagtcat caagcccttc tgtgtcatcc
tgccggaaat
      241 tcagaagcca gagaggaaga ttcagtttaa ggagaaagtg ctgtggaccg
ctatcaccct
      301 ctttatcttc ttagtgtgct gccagattcc cctgtttggg atcatgtctt
cagattcagc
      361 tgacccttтc tattggatga gagtgattct agcctctaac agaggcacat
tgatggagct
      421 agggatctct cctattgtca cgtctggcct tataatgcaa ctcttggctg
gcgccaagat
      481 aattgaagtt ggtgacaccc caaaagaccg agctctcttc aacggagccc
aaaagttatt
      541 tggcatgatc attactatcg gccagtctat cgtgtatgtg atgaccggga
tgtatgggga
      601 cccttctgaa atgggtgctg gaatttgcct gctaatcacc attcagctct
ttgttgctgg
      661 cttaattgtc ctacttttgg atgaactcct gcaaaaagga tatggccttg
gctctggtat
      721 ttctctcttc attgcaacta acatctgtga aaccatcgta tggaaggcat
tcagccccac
      781 tactgtcaac actggccgag gaatggaatt tgaaggtgct atcatcgcac
ttttccatct
      841 gctggccaca cgcacagaca aggtccgagc ccttcgggag gcgttctacc
gccagaatct
      901 tcccaacctc atgaatctca tcgccaccat ctttgtcttt gcagtggtca
tctatttcca
      961 gggcttccga gtggacctgc caatcaagtc ggcccgctac cgtggccagt
acaacaccta
     1021 tcccatcaag ctcttctata cgtccaacat ccccatcatc ctgcagtctg
ccctggtgtc
     1081 caacctttat gtcatctccc aaatgctctc agctcgcttc agtggcaact
tgctggtcag
     1141 cctgctgggc acctggtcgg acacgtcttc tgggggccca gcacgtgctt
atccagttgg
     1201 tggcctttgc tattacctgt cccctccaga atcttttggc tccgtgttag
aagacccggt
     1261 ccatgcagtt gtatacatag tgttcatgct gggctcctgt gcattcttct
ccaaaacgtg
     1321 gattgaggtc tcaggttcct ctgccaaaga tgttgcaaag cagctgaagg
agcagcagat
     1381 ggtgatgaga ggccaccgag agacctccat ggtccatgaa ctcaaccggt
acatccccac
     1441 agccgcggcc tttggtgggc tgtgcatcgg ggccctctcg gtcctggctg
acttcctagg
     1501 cgccattggg tctggaaccg ggatcctgct cgcagtcaca atcatctacc
agtactttga
     1561 gatcttcgtt aaggagcaaa gcgaggttgg cagcatgggg gccctgctct
tctgagcccg
     1621 tctcccggac aggttgagga agctgctcca gaagcgcctc ggaaggggag
ctctcatcat
     1681 ggcgcgtgct gctgcggcat atggactttt aataatgttt ttgaatttcg
tattctttca
     1741 ttccactgtg taaagtgcta gacattttcc aatttaaaat tttgcttttt
atcctggcac
```

1801 tggcaaaaag aactgtgaaa gtgaaatttt attcagccga ctgccagaga
agtgggaatg
1861 gtataggatt gtccccaagt gtccatgtaa cttttgtttt aacctttgca
ccttctcagt
1921 gctgtatgcg gctgcagccg tctcacctgt ttccccacaa agggaatttc
tcactctggt
1981 tggaagcaca aacactgaaa tgtctacgtt tcattttggc agtagggtgt
gaagctggga
2041 gcagatcatg tatttcccgg agacgtggga ccttgctggc atgtctcctt
cacaatcagg
2101 cgtgggaata tctggcttag gactgtttct ctctaagaca ccattgtttt
cccttatttt
2161 aaaagtgatt tttttaagga cagaacttct tccaaaagag agggatggct
ttcccagaag
2221 acactcctgg ccatctgtgg atttgtctgt gcacctattg gctcttctag
ctgactcttc
2281 tggttgggct tagagtctgc ctgtttctgc tagctccgtg tttagtccac
ttgggtcatc
2341 agctctgcca agctgagcct ggccaagcta ggtggacaga cccttgcagt
gatgtccgtt
2401 tgtccagatt ctgccagtca tcactggaca cgtctcctcg cagctgccct
agcaagggga
2461 gacattgtgg tagctatcag acatggacag aaactgactt agtgctcaca
agcccctaca
2521 ccttctgggc tgaagatcac ccagctgtgt tcagaatttt cttactgtgc
ttaggactgc
2581 acgcaagtga gcagacacca ccgacttcct ttctgcgtca ccagtgtcgt
cagcagagag
2641 aggacagcac aggctcaagg ttggtagtga agtcaggttc ggggtgcatg
ggctgtggtg
2701 gtgttgatca gttgctccag tgtttgaaat aagaagactc atgtttatgt
ctggaataag
2761 ttctgtttgt gctgacaggt ggcctaggtc ctggagatga gcaccctctc
tctggccttt
2821 agggagtccc ctcttaggac aggcactgcc cagcagcaag ggcagcagag
ttgggtgcta
2881 agatcctgag gagctcgagg tttcgagctg gctttagaca ttggtgggac
caaggatgtt
2941 ttgcaggatg ccctgatcct aagaaggggg cctgggggtg cgtgcagcct
gtcggggaga
3001 ccccactctg acagtgggca cacggcagcc tgcaaagcac agggccaccg
ccacagcccg
3061 gcagaggggc acactctgga gaccttgctg gcagtgctag ccaggaaaca
gagtgaccaa
3121 gggacaagaa gggacttgcc taaagccacc cagcaactca gcagcagaac
caagatgggc
3181 cccaggctcc tccatatggc ccagggctta ccaccctatc acacgtggcc
ttgtctagac
3241 ccagtcctga gcaggggaga ggctcttgag acctgatgcc ctcctaccca
catggttctc
3301 ccactgccct gtctgctctg ctgctacaga ggggcagggc ctcccccagc
ccacgcttag
3361 gaatgcttgg cctctggcag gcaggcagct gtacccaagc tggtgggcag
ggggctggaa
3421 ggcaccaggc ctcaggagga gccccatagt cccgcctgca gcctgtaacc
atcggctggg
3481 ccctgcaagg cccacactca cgccctgtgg gtgatggtca cggtgggtgg
gtgggggctg
3541 accccagctt ccaggggact gtcactgtgg acgccaaaat ggcataactg
agataaggtg
3601 aataagtgac aaataaagcc agtttttac aaggtaaaaa aaaaaaaaaa aaaaaa

//

Protein sequence:

NCBI Reference Sequence: NP_037468.1

LOCUS: NP_037468

ACCESSION: NP_037468 NP_057052

```
          1 maikflevik pfcvilpeiq kperkiqfke kvlwtaitlf iflvccqipl
fgimssdsad
       61 pfywmrvila snrgtlmelg ispivtsgli mqllagakii evgdtpkdra
lfngaqklfg
      121 miitigqsiv yvmtgmygdp semgagicll itiqlfvagl ivllldellq
kgyglgsgis
      181 lfiatnicet ivwkafsptt vntgrgmefe gaiialfhll atrtdkvral
reafyrqnlp
      241 nlmnliatif vfavviyfqg frvdlpiksa ryrgqyntyp iklfytsnip
iilqsalvsn
      301 lyvisqmlsa rfsgnllvsl lgtwsdtssg gparaypvgg lcyylsppes
fgsvledpvh
      361 avvyivfmlg scaffsktwi evsgssakdv akqlkeqqmv mrghretsmv
helnryipta
      421 aafgglciga lsvladflga igsgtgilla vtiiyqyfei fvkeqsevgs mgallf
//
```

PAPSS2

Official Symbol: **PAPSS2**

Official Name: 3'-phosphoadenosine 5'-phosphosulfate synthase 2 [Homo sapiens]

Gene ID: 9060

Organism: Homo sapiens

Other Aliases: RP11-77F13.2, ATPSK2, SK2

Other Designations: 3-prime-phosphoadenosine 5-prime-phosphosulfate synthase 2; ATP sulfurylase/APS kinase 2; ATP sulfurylase/adenosine 5'-phosphosulfate kinase; PAPS synthase 2; PAPS synthetase 2; PAPSS 2; SK 2; bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2; bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthethase 2

Nucleotide sequence:

NCBI Reference Sequence: NM_001015880.1

LOCUS: NM_001015880

ACCESSION : NM_001015880

```
        1 ctaggcggcg gcggccgggt ccccaaggct gggcgctgct tgcggaaccg
acggggcgga
       61 gaggagcgtg gcgggaggag gagtaggaga agggggctgg tcaagggaag
tgcgacgtgt
```

121 ctgcggagcc tttttatacc tccttcccgg gagtccggca gccgctgctg ctgctgctgc
181 tgctgctgcc gccgccgccg ccgccgtccc tgcgtccttc ggtctctgct cccgggaccc
241 gggctccgcc gcagccagcc agcatgtcgg ggatcaagaa gcaaaagacg gagaaccagc
301 agaaatccac caatgtagtc tatcaggccc accatgtgag caggaataag agagggcaag
361 tggttggaac aaggggtggg ttccgaggat gtaccgtgtg gctaacaggt ctctctggtg
421 ctggaaaaac aacgataagt tttgccctgg aggagtacct tgtctcccat gccatccctt
481 gttactccct ggatggggac aatgtccgtc atggccttaa cagaaatctc ggattctctc
541 ctggggacag agaggaaaat atccgccgga ttgctgaggt ggctaagctg tttgctgatg
601 ctggtctggt ctgcattacc agctttattt ctccattcgc aaaggatcgt gagaatgccc
661 gcaaaataca tgaatcagca gggctgccat tctttgaaat atttgtagat gcacctctaa
721 atatttgtga aagcagagac gtaaaaggcc tctataaaag ggccagagct ggggagatta
781 aaggatttac aggtattgat tctgattatg agaaacctga aactcctgag cgtgtgctta
841 aaaccaattt gtccacagtg agtgactgtg tccaccaggt agtggaactt ctgcaagagc
901 agaacattgt accctatact ataatcaaag atatccacga actctttgtg ccggaaaaca
961 aacttgacca cgtccgagct gaggctgaaa ctctcccttc attatcaatt actaagctgg
1021 atctccagtg ggtccaggtt ttgagcgaag gctgggccac tcccctcaaa ggtttcatgc
1081 gggagaagga gtacttacag gttatgcact ttgacaccct gctagatggc atggcccttc
1141 ctgatggcgt gatcaacatg agcatcccca ttgtactgcc cgtctctgca gaggataaga
1201 cacggctgga agggtgcagc aagtttgtcc tggcacatgg tggacggagg gtagctatct
1261 tacgagacgc tgaattctat gaacacagaa aagaggaacg ctgttcccgt gtttgggggga
1321 caacatgtac aaaacacccc catatcaaaa tggtgatgga aagtggggac tggctggttg
1381 gtggagacct tcaggtgctg gagaaaataa gatggaatga tgggctggac caataccgtc
1441 tgacacctct ggagctcaaa cagaaatgta aagaaatgaa tgctgatgcg gtgtttgcat
1501 tccagttgcg caatcctgtc cacaatggcc atgccctgtt gatgcaggac actcgccgca
1561 ggctcctaga gaggggctac aagcacccgg tcctcctact acaccctctg ggcggctgga
1621 ccaaggatga cgatgtgcct ctagactggc ggatgaagca gcacgcggct gtgctcgagg
1681 aaggggtcct ggatcccaag tcaaccattg ttgccatctt tccgtctccc atgttatatg
1741 ctggccccac agaggtccag tggcactgca ggtcccggat gattgcgggt gccaatttct
1801 acattgtggg gagggaccct gcaggaatgc cccatcctga aaccaagaag gatctgtatg
1861 aacccactca tgggggcaag gtcttgagca tggcccctgg cctcacctct gtggaaatca

1921 ttccattccg agtggctgcc tacaacaaag ccaaaaaagc catggacttc
tatgatccag
1981 caaggcacaa tgagtttgac ttcatctcag gaactcgaat gaggaagctc
gcccgggaag
2041 gagagaatcc cccagatggc ttcatggccc ccaaagcatg gaaggtcctg
acagattatt
2101 acaggtccct ggagaagaac taagcctttg gctccagagt ttctttctga
agtgctcttt
2161 gattaccttt tctatttta tgattagatg ctttgtatta aattgcttct
caatgatgca
2221 ttttaatctt ttataatgaa gtaaaagttg tgtctataat taaaaaaaaa
tatatatata
2281 tacacacaca catatacata caaagtcaaa ctgaagacca aatcttagca
ggtaaaagca
2341 atattcttat acatttcata ataaaattag ctctatgtat tttctactgc
acctgagcag
2401 gcaggtccca gatttcttaa ggctttgttt gaccatgtgt ctagttactt
gctgaaaagt
2461 gaatatattt tccagcatgt cttgacaacc tgtactcttc caatgtcatt
tatcagttgt
2521 aaaatatatc agattgtgtc ctcttctgta caattgacaa aaaaaaaaat
ttttttttct
2581 cactctaaaa gaggtgtggc tcacatcaag attcttcctg atattttacc
tcatgctgta
2641 caaagcctta atgttgtaat catatcttac gtgttgaaga cctgactgga
gaaacaaaat
2701 gtgcaataac gtgaatttta tcttagagat ctgtgcagcc tatttctgtc
acaaaagtta
2761 tattgtctaa taagagaagt cttaatggcc tctgtgaata atgtaactcc
agttacacgg
2821 tgacttttaa tagcatacag tgatttgatg aaaggacgtc aaacaatgtg
gcgatgtcgt
2881 ggaaagttat ctttcccgct ctttgctgtg gtcattgtgt cttgcagaaa
ggatggccct
2941 gatgcagcag cagcgccagc tgtaataaaa aataattcac actatcagac
tagcaaggca
3001 ctagaactgg aaaagaccac agaaaacaaa gaatccaacc ctttcatctt
acaggtgaac
3061 aaactgtgat gatgcacatg tatgtgtttt gtaagctgtg agcaccgtaa
caaaatgtaa
3121 atttgccatt attaggaagt gctggtggca gtgaagaagc acccaggcca
cttgactccc
3181 agtctggtgc cctgtctaca ccagacaaca caggagctgg gtcagattcc
cctcagctgc
3241 ttaacaaagt tcctcgaaca gaaagtgctt acaaagctgc cttctcggat
actgaaaggt
3301 cgagttttct gaactgcact gattttattg cagttgaaaa aaaaaaaaag
ctattccaaa
3361 gatttcaagc tgttctgaga catcttctga tggctttact tcctgagagg
caatgttttt
3421 actttatgca taattcattg ttgccaagga ataaagtgaa gaaacagcac
cttttaatat
3481 ataggtctct ctggaagaga cctaaattag aaagagaaaa ctgtgacaat
tttcatattc
3541 tcattcttaa aaaacactaa tcttaactaa caaaagttct tttgagaata
agttacacac
3601 aatggccaca gcagtttgtc tttaatagta tagtgcctat actcatgtaa
tcggttactc
3661 actactgcct ttaaaaaaaa aaaccagcat atttattgaa aacatgagac
aggattatag

```
     3721 tgccttaacc gatatatttt gtgacttaaa aaatacattt aaaactgctc
ttctgctcta
     3781 gtaccatgct tagtgcaaat gattatttct atgtacaact gatgcttgtt
cttattttaa
     3841 taaatttatc agagtgaaaa aaaaaaaaaa aaaa
            //
```

Protein sequence:

NCBI Reference Sequence: NP_001015880.1

LOCUS: NP_001015880

ACCESSION: NP_001015880

```
        1 msgikkqkte nqqkstnvvy qahhvsrnkr gqvvgtrggf rgctvwltgl
sgagkttisf
       61 aleeylvsha ipcysldgdn vrhglnrnlg fspgdreeni rriaevaklf
adaglvcits
      121 fispfakdre narkihesag lpffeifvda plnicesrdv kglykrarag
eikgftgids
      181 dyekpetper vlktnlstvs dcvhqvvell qeqnivpyti ikdihelfvp
enkldhvrae
      241 aetlpslsit kldlqwvqvl segwatplkg fmrekeylqv mhfdtlldgm
alpdgvinms
      301 ipivlpvsae dktrlegcsk fvlahggrrv ailrdaefye hrkeercsrv
wgttctkhph
      361 ikmvmesgdw lvggdlqvle kirwndgldq yrltplelkq kckemnadav
fafqlrnpvh
      421 nghallmqdt rrrllergyk hpvlllhplg gwtkdddvpl dwrmkqhaav
leegvldpks
      481 tivaifpspm lyagptevqw hcrsrmiaga nfyivgrdpa gmphpetkkd
lyepthggkv
      541 lsmapgltsv eiipfrvaay nkakkamdfy dparhnefdf isgtrmrkla
regenppdgf
      601 mapkawkvlt dyyrslekn
            //
```

CLAIMS

We claim:

1. A method for identifying a drug that causes or is at risk for causing drug-induced cardiotoxicity, comprising: comparing (i) a level of expression of one or more biomarkers present in a first cell sample obtained prior to the treatment with the drug; with (ii) a level of expression of the one or more biomarkers present in a second cell sample obtained following the treatment with the drug; wherein the one or more biomarkers is selected from the markers listed in table 2; wherein a modulation in the level of expression of the one or more biomarkers in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced cardiotoxicity.

2. A method for identifying a rescue agent that can reduce or prevent drug-induced cardiotoxicity comprising: (i) determining a normal level of expression of one or more biomarkers present in a first cell sample obtained prior to the treatment with a cardiotoxicity inducing drug; (ii) determining a treated level of expression of the one or more biomarkers present in a second cell sample obtained following the treatment with the cardiotoxicity inducing drug to identify one or more biomarkers with a change of expression in the treated cell sample; (iii) determining the level of expression of the one or more biomarkers with a changed level of expression in the cardiotoxicity inducing drug treated sample present in a third cell sample obtained following the treatment with the cardiotoxicity inducing drug and the rescue agent; and (iv) comparing the level of expression of the one or more biomarkers determined in the third sample with the level of expression of the one or more biomarkers present in the first sample; wherein the one or more biomarkers is selected from the markers listed in table 2; and wherein a normalized level of expression of the one or more biomarkers in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced cardiotoxicity.

3. A method for alleviating, reducing or preventing drug-induced cardiotoxicity, comprising administering to a subject a rescue agent identified by the method of claim 2, thereby reducing or preventing drug-induced cardiotoxicity in the subject.

4. The method of any one of claims 1-3, wherein the one or more biomarkers is selected from the group consisting TIMP1, PTX3, HSP76, FINC, CYB5, PAI1, IBP7 (IGFBP7), 1C17, EDIL3, HMOX1, NUCB1, CS010, and HSPA4.

5. The method of claim 4, wherein the drug-induced cardiotoxicity is cardiomyopathy, heart failure, atrial fibrillation, cardiomyopathy and heart failure, heart failure and LV dysfunction, atrial flutter and fibrillation, or heart valve damage and heart failure.

6. The method of any one of claims 1-5, wherein the cell samples are cardiomyocytes or diabetic cardiomyocytes.

7. The method of any one of claims 1-3, wherein the drug is a cancer drug, diabetic drug, neurological drug, or anti-inflammatory drug.

8. The method of any one of claims 1-7, wherein the drug is Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab, Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, or TNF antagonists.

9. The method of claim 3, wherein the subject is a mammal, a human, or a non-human animal.

10. The method of claim 3, wherein the rescue agent is administered to a subject that has already been treated with a cardiotoxicity-inducing drug.

11. The method of claim 3, wherein the rescue agent is administered to a subject at the same time as treatment of the subject with a cardiotoxicity-inducing drug.

12. The method of claim 3, wherein the rescue agent is administered to a subject prior to treatment of the subject with a cardiotoxicity-inducing drug.

13. The method of claim 3, wherein the rescue agent is Coenzyme Q10.

14. The method of claim 3, wherein the rescue agent is not Coenzyme Q10.

15. The method of claim 3, further comprising monitoring the subject for drug induced cardiotoxicity.

16. A method for identifying a modulator of drug-induced toxicity, said method comprising:

(1) establishing a model for drug-induced toxicity, using cells associated with drug-induced toxicity, to represents a characteristic aspect of drug-induced toxicity;

(2) obtaining a first data set from the model for drug-induced toxicity, wherein the first data set represents one or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data characterizing the cells associated with drug-induced toxicity;

(3) obtaining a second data set from the model for drug-induced toxicity, wherein the second data set represents a functional activity or a cellular response of the cells associated with drug-induced toxicity;

(4) generating a consensus causal relationship network among the expression levels of the one or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data and the functional activity or cellular response based solely on the first data set and the second data set using a programmed computing device, wherein the generation of the consensus causal relationship network is not based on any known biological relationships other than the first data set and the second data set;

(5) identifying, from the consensus causal relationship network, a causal relationship unique in drug-induced toxicity, wherein a gene, lipid, protein, metabolite, transcript, or SNP associated with the unique causal relationship is identified as a modulator of drug-induced toxicity.

17. The method of claim 16, wherein second data set representing the functional activity or cellular response of the cells comprises one or more of bioenergetics, cell proliferation, apoptosis, organellar function, a genotype-phenotype association actualized by functional models selected from ATP, ROS, OXPHOS, and Seahorse assays, global enzyme activity, and an effect of global enzyme activity on the enzyme metabolic substrates of cells associated with drug-induced toxicity.

18. The method of claim 17, wherein the global enzyme activity is global kinase activity, and wherein the effect of global enzyme activity on the enzyme metabolic substrates is the phospho proteome.

19. The method of any one of claims 1 to 3 and 16, wherein the first data set comprises two or more of genomics, lipidomics, proteomics, metabolomics, transcriptomics, and single nucleotide polymorphism (SNP) data.

20. The method of claim 16, wherein step (4) is carried out by an artificial intelligence (AI) -based informatics platform.

21. The method of claim 20, wherein the AI-based informatics platform comprises REFS(TM).

22. The method of claim 21, wherein the AI-based informatics platform receives all data input from the first data set and the second data set without applying a statistical cut-off point.

23. The method of claim 16, wherein the consensus causal relationship network established in step (4) is further refined to a simulation causal relationship network, before step (5), by *in silico* simulation based on input data, to provide a confidence level of prediction for one or more causal relationships within the consensus causal relationship network.

24. The method of claim 16, wherein the unique causal relationship is identified as part of a differential causal relationship network that is uniquely present in cells, and absent in the matching control cells.

25. The method of claim 16, wherein the unique causal relationship identified is a relationship between at least one pair selected from the group consisting of expression of a gene and level of a lipid; expression of a gene and level of a transcript; expression of a gene and level of a metabolite; expression of a first gene and a second gene; expression of a gene and presence of a SNP; expression of a gene and a functional activity; level of a lipid and level of a transcript; level of a lipid and level of a metabolite; level of a first lipid and a second lipid; level of a lipid and presence of a SNP; level of a lipid and a functional activity; level of a first transcript and level of a second transcript; level of a

transcript and level of a metabolite; level of a transcript and presence of a SNP; level of a first transcript and a functional activity; level of a first metabolite and level of a second metabolite; level of a metabolite and presence of a SNP; level of a metabolite and a functional activity; level of a first SNP and presence of a second SNP; and presence of a SNP and a functional activity.

26. The method of claim 25, wherein the functional activity is selected from the group consisting of bioenergetics, cell proliferation, apoptosis, organellar function, kinase activity, protease activity, and a genotype-phenotype association actualized by functional models selected from ATP, ROS, OXPHOS, and Seahorse assays.

27. The method of claim 16, further comprising validating the identified unique causal relationship in drug-induced toxicity.

28. The method of claim 16, wherein the drug-induced toxicity is drug-induced cardiotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, renaltoxicity, or myotoxicity.

29. The method of claim 28, wherein the drug-induced cardiotoxicity is cardiomyopathy, heart failure, atrial fibrillation, cardiomyopathy and heart failure, heart failure and LV dysfunction, atrial flutter and fibrillation, or, heart valve damage and heart failure.

30. The method of claim 16, wherein the model for drug-induced toxicity comprises cell cardiomyocytes, diabetic cardiomyocytes, hepatocytes, kidney cells, neuronal cells, renal cells, or myoblasts.

31. The method of claim 16, wherein the model for drug-induced toxicity comprises a toxicity inducing drug, cancer drug, diabetic drug, neurological drug, or anti-inflammatory drug.

32. The method of claim 16, wherein the drug is Anthracyclines, 5-Fluorouracil, Cisplatin, Trastuzumab, Gemcitabine, Rosiglitazone, Pioglitazone, Troglitazone, Cabergoline, Pergolide, Sumatriptan, Bisphosphonates, or TNF antagonists.

33. A method for identifying a drug that causes or is at risk for causing drug-induced toxicity, comprising: comparing (i) a level of one or more biomarkers present in a first

cell sample obtained prior to the treatment with the drug; with (ii) a level of the one or more biomarkers present in a second cell sample obtained following the treatment with the drug; wherein the one or more biomarkers is selected from the modulators identified by the methods of any one of claims 16-32; wherein a modulation in the level of the one or more biomarkers in the second sample as compared to the first sample is an indication that the drug causes or is at risk for causing drug-induced toxicity.

34. A method for identifying a rescue agent that can reduce or prevent drug-induced toxicity comprising: (i) determining a normal level of one or more biomarkers present in a first cell sample obtained prior to the treatment with a toxicity inducing drug; (ii) determining a treated level of the one or more biomarkers present in a second cell sample obtained following the treatment with the toxicity inducing drug to identify one or more biomarkers with a change of level in the treated cell sample; (iii) determining the level of the one or more biomarkers with a changed level in the toxicity inducing drug treated sample present in a third cell sample obtained following the treatment with the toxicity inducing drug and the rescue agent; and (iv) comparing the level of the one or more biomarkers determined in the third sample with the level of the one or more biomarkers present in the first sample; wherein the one or more biomarkers is selected from the modulators identified by the methods of any one of claims 16-32 and wherein a normalized level of the one or more biomarkers in the third sample as compared to the first sample is an indication that the rescue agent can reduce or prevent drug-induced toxicity.

35. A method for alleviating, reducing or preventing drug-induced toxicity, comprising administering to a subject the rescue agent of claim 34, thereby reducing or preventing drug-induced toxicity in the subject.