A CLINICAL METHOD FOR GENOTYPING LARGE GENES FOR MUTATIONS THAT POTENTIALLY CAUSE DISEASE

Field of the Invention

[0001]. The present invention relates to screening and prescreening methods for genotyping large genes for mutations that potentially cause disease such as muscle disease and kits for carrying out the same.

Background Art

[0002]. There are more than 40 primary congenital muscle disorders known and the most common is a genetic disease called Duchenne muscular dystrophy (DMD) followed by Becker muscular dystrophy (BMD). Both are allelic disorders caused by mutations in the same gene, the DMD gene of the X chromosome, with the latter having a milder phenotype and slower disease progression. Together, DMD and BMD occur with an incidence of approximately 1 in 3,500 newborn males, accounting for 80% of all new cases of dystrophy and 56% of all new cases of congenital myopathy of all types. DMD is a severe disease resulting in severe and progressive muscle wasting. The patient becomes wheel-chair bound by about 12 years of age, and usually do not survive beyond the second decade of life due to pulmonary or cardiac failure. BMD patients may be ambulant with milder symptoms and survive longer.

[0003]. The DMD gene responsible for both disorders is the largest known in the human genome spanning 2.3 Megabases (or 2,300 kB) which is encoded by 79 exons. This gene is localized to the X chromosome, hence, the disease is recessive x-lined and affects only males while females can be silent carriers. The gene codes for a muscle protein called dytsrophin which is absent or semi-functional in patients. About 60% of DMD and BMD cases are caused by large deletions or duplications, but the remaining mutations are mostly small and are known as point mutations. They involve single or multiple base changes,

insertions, deletions or duplications. To detect these unknown mutations, the sequences in all the coding exonic regions as well as the intronic regions (flanking sequences, branch-points, promoter regions, etc.) have to be interrogated. This is akin to looking for a needle in a hay-stack and every single base in the large gene sequences has to be checked. There are no known mutation hot-spots for point mutations in the DMD gene. All the reported mutations so far have been heterogeneous and can occur anywhere in the gene, both in exonic or intronic regions which affect protein functioning.

[0004]. Mutation screening is important to help confirm the clinical diagnosis, detect carriers, perform genetic counselling and allow for prenatal diagnosis or diagnosis of pre-implantation of embryos. In addition, mutation screening in patients is also useful as there are now various novel molecular therapeutic treatments which are currently in various phases of clinical trials. These treatment approaches aim to restore the function of the defective dystrophin muscle protein by targeting specific classes of mutations in the DMD gene. Hence, it is important to identify the specific underlying mutation in each patient in order to select the best possible treatment approaches in future.

[0005]. The most widely offered test in diagnostic labs for molecular diagnosis in DMD involves detection of common large deletion/duplication mutations costing about a few hundred US dollars or British sterling pounds. Sequencing analysis is offered only at a few university labs and takes a few months for turn-around of results. Sequencing is expensive and depends on the size of the gene and number of exons to be analysed, costs escalating with larger sequences. As the DMD gene is very large with a large number of exons sequencing of the gene costs more than 10 times that of testing for common deletion/duplication mutations. Hence, to currently test for uncommon point mutations it is expensive, time consuming and limited to very few locations.

[0006]. In mutation screening or genotyping, large genes are challenging to interrogate when a large region has to be analysed quickly in a short time. For

typical diagnostic labs offering genetic service, results for patients have to be reported as soon as possible. Prenatal diagnosis and embryo pre-implantation diagnosis require urgent testing within a limited time-frame. While direct DNA sequencing is one of the most common and accurate methods for detecting unknown mutations, this process involves designing individual assays based on location and primer sequences, and subsequently iteratively optimizing each individual assay, and analysing the individual sequences of each region covered by the assays. Thus, it is laborious, costly and time-consuming when a large number of assays and large regions of sequences have to be analysed. Consequently, a large number of mutation screening methods have been introduced to reduce the number of such analyses required by narrowing the region which needs to be sequenced or by using an approach which can easily detect a variant from wild-type sequence. These include denaturing high performance liquid chromatography (DHPLC), single strand conformational polymorphism (SSCP), single condition amplification/internal primer (SACIP), etc. which employs multiple protocol steps and procedures. However, these current methods are not sensitive or efficient enough or suited for high throughput analysis of genes while other chip-based technology are currently too expensive to be affordable for routine diagnosis of heritable genetic disorders in clinics.

[0007]. Generally detection of mutations in this disease may be required from biological samples from infants or young children. These samples include blood samples, fetuses or embryos. Only very small amount of samples are able to be collected from such sources. All the current detection methods require a significant volume of DNA sample.

[0008]. High resolution melting (HRM) is a simple, rapid method for mutation scanning which requires template DNA of good quality and quantity. The method involves the melting of double stranded DNA into single stranded DNA with increasing temperature such that amplicons with sequence differences will generate different melt profiles due to the difference in thermal stability. These melt profiles are detected due to the use of saturating intercalating DNA dyes which are released as the DNA strands melt and are captured by new

generation melt instruments which are capable of much higher data density capture than conventional real-time PCR machines. The lowest volume sample that can be used to conduct HRM is 10ng. For small genes or when genotyping a few assays, the DNA amount required is small as compared to assays for DMD gene due to the large size of the gene and the number of assays required to interrogate the entire gene.

Summary of the Invention

[0009]. The present invention seeks to provide novel methods to ameliorate some of the difficulties with the current detection and pre-screening of disease caused by a genetic mutation in a large gene sequence.

[00010]. We have developed a method that can enrich the DNA such that only a small volume is required that may accurately and quickly detect mutations within a large gene sequence.

[00011]. Accordingly the present invention provides a method of determining polymorphisms within a large gene comprising the steps of: (a) making a Whole-Genome Amplification (WGA) to obtain sufficient amounts of genetic templates for DNA analysis; (b) enriching the WGA sample with nested primers designed for the large gene; (c) using the enriched WGA sample for high resolution melt (HRM); and (d) detecting differential melt profiles during the transition from double strand to single strand with an increase in temperature wherein sequence point mutations within the gene affects the thermal stability and gives a different melt profile from the normal non-mutated gene sequence.

[00012]. Preferably the large gene is the DMD gene that may potentially contribute to muscle disease or muscular dystrophy such as Duchenne muscular dystrophy or Becker muscular dystrophy.

[00013]. The method may further comprise a cleaning step prior to the HRM step such as washing with a high magnesium concentration, exo-nuclease digestion, dephosphorylation treatment or filtration.

[00014]. A further aspect of the present invention provides a kit comprising reagents for whole genome amplification, at least two nested primers specific to the gene of interest and reagents' for high resolution melt analysis.

[00015]. Preferably the kit is to rapidly detect mutations that may potentially contribute to muscle disease.

Brief Description of the Drawings

Figure 1: Depicts a symbolized flow chart of a preferred embodiment of the method of the invention

Figure 2: Melt-analysis for WGA enriched and purified products 2A depicts results from Exon 48 showing use of normal WGA products (without purification and/or enrichment) : the three different genotypes cannot be distinguished from each other. In comparison **2B** depicts results Exon 48 showing enriched WGA products: the three different genotypes (A/A, A/C and C/C) cluster in distinct groups. **2C** depicts results from Exon 37 showing use of normal WGA products (without purification and/or enrichment): the three different genotypes cannot be distinguished clearly from each other. In comparison **2D** depicts results Exon 37 showing enriched WGA products: where the three different genotypes (A/A, A/G and C/C) cluster in distinct group.

Figure 3. HRM analysis using probes and primers shows use of probes to differentiate two SNPs (A/G and A/C) at probe melt regions Rs1800274 (A/G) peak **A** and Rs1800275 (A/C) peak **B**

Figure 4. HRM analysis using primers shows use of different dyes for high resolution melt to differentiate between normals (N), patient (P) and carrier (C) samples using the DMD primers for Exon 46 where Differentiation of Genotypes are indicated as Mutant (P), Wild Type (N), and Heterozygote carriers (C).

Detailed Disclosure

[00016]. According to the invention there is provided a method of determining sequence variants, including mutations and polymorphisms within a large gene preferably within the DMD gene, contributing to disease, preferably muscle disease comprising the steps of: (a) making a Whole-Genome Amplification (WGA) step to obtain sufficient amounts of genetic templates for DNA analysis. (b) enriching the WGA sample with nested primers designed for the large gene; (c) using the enriched WGA sample for high resolution melt (HRM); and (d) detecting differential melt profiles during the transition from double strand to single strand with an increase in temperature wherein sequence point mutations within the gene affects the thermal stability and gives a different melt profile from the normal non-mutated gene sequence.

[00017]. The method involves the melting of double stranded DNA into single stranded DNA with increasing temperature such that amplicons with sequence differences will generate different melt profiles due to the difference in thermal stability. These melt profiles are detected due to the use of saturating intercalating DNA dyes which are released as the DNA strands melt and are captured by new generation melt instruments which are capable of much higher data density capture than conventional real-time PCR machines. This method is thus also called high resolution melting (HRM) analysis.

[00018]. The HRM step may include a cleaning step such as the use of high magnesium concentration, exo-nuclease digestion, dephosphorylation treatment or filtration steps or any other method known in the art that could adequately clean the enriched WGA sample. Four mutations in the DMD gene were amplified using multiple displacement amplification (MDA) and analyzed by HRM. Melting patterns of WGA and genomic DNA were concordant for 3 out of 4 mutations, and discordant results were eliminated after post-PCR dephosphorylation treatment was performed. This is the first known study showing the effect of magnesium salt concentrations and dephosphorylation treatment on melt-analysis using WGA samples. Additionally, use of an additional nested PCR step for enriching the WGA products has been shown to improve the

melt analysis and differentiate the different genotypes of mutation present in a sample. Use of WGA samples for melt-analysis is useful for genotyping/mutation screening in samples of limiting DNA templates such as in prenatal diagnosis, embryo pre-implantation diagnosis, and other such methods.

[00019]. This method allows for quick screening of both known and unknown mutations as primers can be designed to optimally amplify regions simultaneously at same or similar sets of conditions for the entire gene, and the amplified products melted immediately after the PCR process to generate the melt profiles within 1 or 2 minutes either in a single closed tube or plate-based approach without a need for further laboratory manipulation. This reduces external contamination as well as reduces the number of steps for analysis. Following this primary screen, only the candidate sequence region showing abnormal melt-profiles (as compared with controls) need to be further analysed by sequencing. This reduces the need to sequence all the 79 exons per patient (which actually involve sequencing of more than 100 sequence regions as some exons are too large or contain repeat sequences to allow amplification or sequencing in a single reaction assay per exon).

[00020]. For genotyping of known mutations, carrier screening by testing for absence or presence of the variant can similarly be quickly done on large number of samples by merely comparing the melt-profiles that are generated.

[00021]. In some regions of the gene, the presence of polymorphisms or repeat sequences may generate complex profiles which are difficult to analyse or interpret. For such regions, we have designed probes to hybridize to these complex regions such that a nested PCR is carried out with the second PCR generating an asymmetric product with more copies of the strand to which the probe will bind. The probe is designed to target the complex sequence but with a T_m that can generate a different temperature window upon melting. Hence, the probe-amplicon duplex and the whole amplicon duplex generates different melt-curves at two different distinct temperature windows allowing interrogation of

both the region of polymorphism as well as mutation screening in the larger amplicon. The probe can be detected without being labelled but is blocked at its 3'end to prevent extension of the oligonucleotides during PCR elongation step.

[00022]. This method is simple, rapid and cheap. It requires only a PCR reaction, dye and melting equipment capable of capturing and generating the different melt profiles. It does not need real-time amplification and is performed within 1-2 minutes after PCR in a closed tube homogeneous assay. It does not require expensive labelled probes or extensive technical time for analysis of the results. As a clinical assay, it is affordable to patients. It can be customized as a commercial kit based assay-system offering diagnosis for entire gene. It has practical applications in muscle diseases such as DMD and BMD due to the large size of the gene and the considerable cost involved if more than 100 PCR and sequencing reactions have to be carried out per patient It is also a rapid method with short turn-around time as only a PCR amplification and melt step is required. This method makes genetic testing more affordable for patients. It will contribute towards not only confirming molecular diagnosis but also for differential diagnosis.

Whole genome amplification

[00023]. Whole Genome Amplification (WGA) : may be a technique designed to amplify all or some of the DNA in a sample. Several whole-genome amplification (WGA) techniques have been presented as capable of amplifying DNA from trace quantities and with less error than traditional PCR. Some preferred methods to perform WGA may include multiple-displacement amplification (MDA), that uses the highly processive φ29 DNA polymerase and random exonuclease-resistant primers in an isothermal amplification reaction. This method is based on strand-displacement synthesis there are commercially available MDA kits such as GE MDA kit and the Sigma WGA kit. Some other methods may include PCR-based methods for WGA including degenerated oligonucleotide primed PCR (DOP-PCR) and primer extension PCR and ligation

mediated PCR (LM-PCR). WGA kit (Sigma): may be a commercially available kit for conducting whole genome amplification. A new WGA method known as OmniPlex, converts randomly fragmented genomic DNA into a library of inherently amplifiable DNA fragments of defined size. This library can be effectively amplified several thousand fold with the help of a high-fidelity DNA polymerase. The library can be re-amplified to achieve a final amplification of over a million fold without degradation of representation. Similarly any WGA known in the art may be suitable for the invention.

[00024]. Potential advantages of WGA include amplifying DNA from nanogram amounts, and large cost and time savings compared with alternatives such as generating cell lines from individuals. Typically between 0.1 to 10 ng of starting genomic DNA can be used for WGA, and after the WGA treatment, between 0.1 ng to 10 ng of the products can be used for the melt analysis.

PROBES AND PRIMERS FOR MUTATION SCREENING IN DMD/BMD

[00025]. The probes and primers used for mutation screening and genotyping in DMD/BMD are targeted to specific sequences and uniquely designed so that they can be amplified and melted at universal sets of conditions and temperatures to reduce the number of different PCR assays required to screen the entire gene simultaneously. Nested primers are designed to reduce the contamination in products due to the amplification of unexpected primer binding sites effectively enriching the samples for use in the HRM ensuring very clean melt profiles that allow a quick visual detection of any mutation in a particular amplicon.

[00026]. Polymerase chain reaction (PCR) itself is the process used to amplify nucleic acid samples. Conventional PCR requires primers complementary to the termini of the target nucleic acid. A problem with primers is that they often bind to incorrect regions of the nucleic acid, giving unexpected products. Nested primers comprise two sets of primers, used in two successive runs of polymerase chain reaction, the second set intended to amplify a

secondary target within the first run product or a sequence that binds to a segment of the first amplicon to ensure only the intended locus is examined in the HRM. It is very unlikely that any of the unwanted PCR products contain binding sites for both the new primers, ensuring the product from the second PCR has little contamination from unwanted products of primer dimers, hairpins, and alternative primer target sequences.

[00027]. The product from the first reaction undergoes a second run with the second set of primers. This requires unique knowledge of various factors such as melt-domains, melting temperatures, sequence specific properties and experimental parameters affecting melt-profiles. This methodology can be applied for interrogation of any sets of genes which can be carried out under similar assay conditions.

[00028]. The outer primers are designed based on coverage of the sequences amplified by the inner primers. All the inner primers are similar to those for the normal melt analysis, and for the WGA enrichment method, outer primers are designed such that primer melting temperatures (Tm) range from 56-64°C and the maximum difference between the Tm of each primer in a pair should not exceed 1°C. Primer GC Content range from 10-90%. Primer Sizes range from 17-30 bases. Maximum Self-Complementarity is set at 6, while Maximum 3' Stability set to 8.

[00029]. "Probes". Polynucleotide polymorphisms associated with DMD alleles are detected by hybridisation with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridisation and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridisation stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these

indications need further analysis to demonstrate detection of a DMD in muscle disease such as DMD and BMD.

[00030]. Probes for DMD nucleic acid may be derived from the sequences of the DMD gene or its cDNAs. The probes may be of any suitable length, which span all or a portion of an exon or intron of the DMD gene and which allow specific hybridisation to the DMD gene. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8-30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridises to the target sequence with the requisite specificity.

[00031]. The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g Sambrook et al., 1989: "Molecular Cloning: a laboratory manual. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Coldspring Harbour Laboratory Press, Coldspring Harbour, NY Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

[00032]. Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically

synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

[00033]. Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding dystrophin protein are preferred as probes. The probes may also be used to determine whether mRNA encoding dystrophin protein is present in a cell or tissue.

[00034]. The present invention provides one or more DMD polynucleotides or fragments thereof comprising mutations with respect to the wild type sequence, such as the sequence shown in SEQ ID No. 1. In a further embodiment, the present invention provides a plurality of DMD polynucleotides or fragments thereof for use in screening the DNA of an individual for the presence of one or more mutations/polymorphisms. The plurality of sequences is conveniently provided immobilized to a solid substrate as is described below.

High Resolution Melt

[00035]. Method includes purification and enrichment of products for melt analysis. It employs technique to amplify single product at high efficiency using optimized conditions with Mg salt concentrations; and/or treatment with exonuclease I amd shrimp alkaline phosphatase enzymes to remove primers and nucleotides from post-PCR products prior to melt; and/or ethanol precipitation of WGA products; and/or enrichment of the primary WGA product by a nested amplification step using outer primer pairs in first round PCR followed by inner primer pairs in second round PCR as a step before the final melt analysis. High Resolution Melt or HRM analysis as it will be referred to herein is a technique for the detection of mutations, polymorphisms and epigenetic differences in double stranded DNA samples. HRM analysis is performed on double stranded DNA samples. Typically the user will use polymerase chain reation (PCR) prior to HRM analysis to amplify the DNA region in which their mutation of interest lies.

Essentially the PCR process turns a tiny amount of the region of interest in the DNA into a large amount, so you have quantities large enough for better analysis. In the tube there are now many of copies of the region of DNA of interest. This region that is amplified is known as the amplicon. After the PCR process the HRM analysis begins. The process is simply a precise heating of the amplicon DNA from around 50˚C up to around 95˚C. At some point during this process, the melting temperature of the amplicon is reached and the two strands of DNA separate or "melt" apart. HRM is monitored by using a fluorescent dye. The dyes that are used for HRM are known as intercalating dyes and have a unique property. They bind specifically to double-stranded DNA and when they are bound they fluoresce brightly. As the type of dyes used are fully saturating, they will bind at all possible positions on the double stranded DNA such that the strands are fully statured with the dye. In the absence of double stranded DNA they have nothing to bind to and they only fluoresce at a low level. At the beginning of the HRM analysis there is a high level of fluorescence in the sample because of the billions of copies of the amplicon. But as the sample is heated up and the two strands of the DNA melt apart, presence of double stranded DNA decreases and thus fluorescence is reduced. The HRM machine has a camera that monitors and captures this process by measuring the fluorescence. The machine then simply plots this data as a graph known as a melt curve, showing the level of fluorescence vs the temperature. The melting temperature of the amplicon at which the two DNA strands come apart is entirely predictable. It is dependant on the sequence of the DNA bases. If two samples from two different people are compared, they should give exactly the same shaped melt curve or profile. However if one of the individual has a mutation in the DNA region that was amplified, then this will alter the temperature at which the DNA strands melt apart. So now the two melt curves appear different. The difference may only be subtle, perhaps a fraction of a degree, but because the HRM machine has the ability to monitor this process in "high resolution", it is possible to accurately document these changes and therefore identify if a mutation is present or not.

[00036]. The present invention relates to a fluorescence-based method which is able to determine whether a DNA sample is identical to another DNA sample in DNA sequence. This method is sufficiently sensitive to detect a difference of a single nucleotide base pair in DNA fragments of up to 250 bases in length. This method utilizes fluorescence-monitoring of the gradual and progressive thermal denaturation of double-stranded DNA (dsDNA) of various origins and configurations. Fluorescence is provided by fluorescent dsDNA-binding dyes which fluoresce while incorporated into dsDNA. As the dsDNA denatures, dye molecules are released into the solution and cease to fluoresce. Sufficient amount of the sample DNA for screening may be produced by various Polymerase Chain Reaction (PCR) methods as well as various Whole Genome Amplification (WGA) methods. A wild-type DNA fragment may be used as a control for comparison in mutation scanning applications.

[00037]. One preferred embodiment of the invention is for the detection of variants (including polymorphisms and mutations) in the human DMD gene, though the method can be applied to any gene in any organism of known genome.

[00038]. An isolated DMD nucleic acid molecule is disclosed which molecule typically encodes a dystrophin polypeptide, allelic variant, or analog, including fragments, thereof. Specifically provided are DNA molecules selected from the group consisting of: (a) DNA molecules set out in SEQ ID NO: 1, or fragments thereof; (b) DNA molecules that hybridize to the DNA molecules defined in (a) or hybridisable fragments thereof; and (c) DNA molecules that code an expression for the amino acid sequence encoded by any of the foregoing DNA molecules.

[00039]. Preferred DNA molecules according to the invention include DNA molecules comprising the sequence set out in SEQ ID NO: 1 or fragments thereof.

[00040]. A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the

polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

[00041]. An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein that has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

[00042]. "DMD Allele" refers to normal alleles of the DMD gene sequence as well as alleles carrying variations.

[00043]. "DMD gene sequence," "DMD gene," "DMD nucleic acids" or " DMD polynucleotide" each refer to polynucleotides that are likely to be expressed in muscle tissue. Mutations at the DMD gene sequence may be involved in muscle disease.

[00044]. The DMD gene sequence is intended to include coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The DMD gene sequence is intended to include all allelic variations of the DNA sequence.

[00045]. These terms, when applied to a nucleic acid, refer to a nucleic acid that encodes a DMD polypeptide, fragment, homologue or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence that is either derived from, or substantially similar to a natural DMD encoding gene or one having substantial homology with a natural DMD encoding gene or a portion thereof. The coding sequence for human DMD polynucleotide is shown in SEQ ID NO: 1, with the amino acid sequence shown in SEQ ID NO: 2.

[00046]. A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

[00047]. Alternatively, substantial homology or (identity) exists when a nucleic acid or fragment thereof will hybridise to another nucleic acid (or a complementary strand thereof) under selective hybridisation conditions, to a strand, or to its complement. Selectivity of hybridisation exists when hybridisation that is substantially more selective than total lack of specificity occurs. Typically, selective hybridisation will occur when there is at least about 55% identity over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

[00048]. Thus, polynucleotides of the invention preferably have at least 75%, more preferably at least 85%, more preferably at least 90% homology to the sequences shown in the sequence listings herein. More preferably there is at least 95%, more preferably at least 98%, homology. Nucleotide homology comparisons may be conducted as described below for polypeptides. A preferred sequence comparison program is the GCG Wisconsin Bestfit program described below. The default scoring matrix has a match value of 10 for each identical nucleotide and -9 for each mismatch. The default gap creation penalty is -50 and the default gap extension penalty is -3 for each nucleotide.

[00049]. In the context of the present invention, a homologous sequence is taken to include a nucleotide sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level over at least 20, 50, 100, 200, 300, 500 or 1000 nucleotides with the nucleotides sequences set out in SEQ ID. No 1. In particular, homology should typically be considered with respect to those regions of the sequence that encode contiguous amino acid sequences known to be essential for the function of the protein rather than non-essential neighbouring sequences. Preferred polypeptides of the invention comprise a contiguous sequence having greater than 50, 60 or 70% homology, more preferably greater than 80, 90, 95 or 97% homology, to one or more of the nucleotides sequences of SEQ ID NO: 1 from 1 to **[how long is the sequence]**, which encode amino acids 1 to **[how long is the sequence]** of SEQ ID NO:2. Preferred polynucleotides may alternatively or in addition comprise a contiguous sequence having greater than 80, 90, 95 or 97% homology to the sequence of SEQ ID NO: 1 that encodes amino acids 1 to 3685 of SEQ ID NO:2.

[00050]. Other preferred polynucleotides comprise a contiguous sequence having greater than 40, 50, 60, or 70% homology, more preferably greater than 80, 90, 95 or 97% homology to the sequence of SEQ ID NO: 1 that encodes amino acids 1 to 3685.

[00051]. Nucleotide sequences are preferably at least 15 nucleotides in length, more preferably at least 20, 30, 40, 50, 100 or 200 nucleotides in length.

[00052]. Generally, the shorter the length of the polynucleotide, the greater the homology required to obtain selective hybridization. Consequently, where a polynucleotide of the invention consists of less than about 30 nucleotides, it is preferred that the % identity is greater than 75%, preferably greater than 90% or 95% compared with the DMD nucleotide sequences set out in the sequence listings herein. Conversely, where a polynucleotide of the invention consists of, for example, greater than 50 or 100 nucleotides, the % identity compared with

the DMD nucleotide sequences set out in the sequence listings herein may be lower, for example greater than 50%, preferably greater than 60 or 75%.

[00053]. The main approach for this set of assays is the use of primers and probes to encompass coding exons, flanking intronic splice sites and branch-points, upstream promoter regionsand downstream regulatory regions at the 3' untranslated region of the DMD gene. A pair of primers were used for the amplification and melt assay of the regions with no known SNPs while additionally a set of probes were used individually for regions with known presence of SNPs (as provided by dbSNP in NCBI database). Primers were designed based on the following features :

[00054]. Primer Melting Temperatures (T_m) range from 56-64°C in 2°C increments. Maximum Difference between the T_m of each primer in a pair should not exceed 1°C. Primer GC Content range from 10-90%. Primer Sizes range from 17-30 bases. Maximum Self-Complementarity set at 6, while Maximum 3' Stability set to 8. Product Sizes range from 100-250 bases in length, though whenever possible the amplicon size range should be 150-200 bases long. Whenever possible, primers for the individual exons encompass the predicted branching points for the exon and at least 5 bases of the following intron at the 3' end. Due to the relatively small amplicon sizes, many exons require 2 or more overlapping amplicons for complete coverage of the branch points, exon, and 3' flank. Primer sequences require checking by BLAST or other similar programmes to rule out the possibility of multiple binding sites. Where possible the primers were designed to amplify amplicons with only one melting domain. The designing of primers for this invention can be performed using Primer3 or other primer design software.

[00055]. Probes were designed based on the following features: A list of known Single Nucleotide Polymorphisms (SNPs) in the exonic regions can be obtained from online databases such as dbSNP. SNPs are avoided by placement of primers directly on the SNP or avoiding the SNP region entirely. Coverage of

the SNP regions can be obtained by the use of unlabelled hybridization probes which are modified at the 3' end to prevent PCR extension. Hybridization probes are short fragments of DNA complementary to one strand of the DNA duplex. The probe works by effectively shortening the region of interrogation to the length of the probe, thereby magnifying any existing sequence difference. An important criteria in probe design is to have the SNP fall roughly in the middle of the probe as the further from the centre of the probe a SNP is located, the higher the probability the probe will fail to work as intended. The probes can be designed by Primer3 or any other primer design or probe selection software. Parameters for probe selection in Primer3 can be set. Hybridization Oligo Size range from 20-40 bases. Hybridization Oligo T_m range from 50-65°C, as it is vital that there be at least 10°C between the denaturation temperature of the probe with that of the full-length PCR product. Fluorescent-labeled probes can also be designed using the same parameters and software, provided the excitation wavelength of the fluorescent label is close to the emission wavelength of the dsDNA-binding dye and the detector is capable of reading the excitation wavelength of the fluorescent label.

[00056]. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, natural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring.

[00057]. In the context of the present invention, a homologous sequence is taken to include an amino acid sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level over at least 20, 50, 100, 200, 300 or 400 amino acids with the amino acid sequences

set out in SEQ ID. Nos 2. In particular, homology should typically be considered with respect to those regions of the sequence known to be essential for the function of the protein rather than non-essential neighbouring sequences. Preferred polypeptides of the invention comprise a contiguous sequence having greater than 50, 60 or 70% homology, more preferably greater than 80 or 90% homology, to one or more of amino acids of SEQ ID NO: 2.

[00058]. Other preferred polypeptides comprise a contiguous sequence having greater than 40, 50, 60, or 70% homology, of SEQ ID No: 2. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity. The terms "substantial homology" or "substantial identity", when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 70% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 80% identity, and preferably at least about 90 or 95% identity.

[00059]. Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

[00060]. Percentage (%) homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an “ungapped” alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues (for example less than 50 contiguous amino acids).

[00061]. Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences,

one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

[00062]. However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package (see below) the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

[00063]. Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux *et al.,* 1984, Nucleic Acids Research 12:387). Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel *et al.*, 1999 *ibid* – Chapter 18), FASTA (Atschul *et al.,* 1990, J. Mol. Biol., 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see

Ausubel *et al.*, 1999 *ibid*, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program.

[00064]. Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

[00065]. Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

[00066]. "dystrophin protein" or "DMD polypeptide" refers to a protein or polypeptide encoded by the DMD gene sequence, variants or fragments thereof. Also included are proteins encoded by DNA that hybridize under high or low stringency conditions, to DMD encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the dystrophin protein(s).

[00067]. A polypeptide "fragment," "portion" or "segment" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

[00068]. Any DMD nucleic acid specimen, in purified or non-purified form, can be utilised as the starting nucleic acid or acids.

[00069]. PCR is one such process that may be used to amplify DMD gene sequences. This technique may amplify, for example, DNA or RNA, including messenger RNA, wherein DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes, and/or conditions optimal for reverse transcribing the template to DNA would be utilized. In addition, a DNA-RNA hybrid that contains one strand of each may be utilized. A mixture of nucleic acids may also be employed, or the nucleic acids produced in a previous amplification reaction described herein, using the same or different primers may be so utilised.

[00070]. The specific nucleic acid sequence to be amplified, i.e., the polymorphic gene sequence, may be a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid. It is not necessary that the sequence to be amplified is present initially in a pure form; it may be a minor fraction of a complex mixture, such as contained in whole human DNA.

[00071]. DNA utilized herein may be extracted from a body sample, such as blood, buccal cells, sera, tissue material, muscle tissue and the like by a variety of techniques. If the extracted sample has not been purified, it may be treated before amplification with an amount of a reagent effective to open the cells, or animal cell membranes of the sample, and to expose and/or separate the strand(s) of the nucleic acid(s). This lysing and nucleic acid denaturing step to expose and separate the strands will allow amplification to occur much more readily.

[00072]. The deoxyribonucleotide triphosphates dATP, dCTP, dGTP and dTTP are added to the synthesis mixture, either separately or together with the primers, in adequate amounts and the resulting solution is heated to about 90 degrees – 100 degrees C from about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period, the solution is allowed to cool, which is preferable for the primer hybridization. To the cooled mixture is added an

appropriate agent for effecting the primer extension reaction (called herein "agent for polymerization"), and the reaction is allowed to occur under conditions known in the art. The agent for polymerization may also be added together with the other reagents if it is heat stable. This synthesis (or amplification) reaction may occur at room temperature up to a temperature above which the agent for polymerization no longer functions. Thus, for example, if DNA polymerase is used as the agent, the temperature is generally no greater than about 40 degree C. Most conveniently the reaction occurs at room temperature.

[00073]. Specific oligonucleotide primers derived from DMD gene sequence may be useful in determining whether a subject is at risk of suffering from the ailments described herein. Primers direct amplification of a target polynucleotide (eg DMD or any of other muscle gene like FKRP or DYSF or CAPN3 or LAMA2 or LMNA or SG or SGCB) prior to sequencing. Primers used in any diagnostic assays derived from the present invention should be of sufficient length and appropriate sequence to provide initiation of polyrmerisation. Environmental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerisation, such as DNA polymerase, and a suitable temperature and pH.

[00074]. Primers of the invention are preferably single stranded for maximum efficiency in amplification, but may be double stranded. If double stranded, primers may be first treated to separate the strands before being used to prepare extension products. Primers should be sufficiently long to prime the synthesis of DMD gene extension products in the presence of the inducing agent for polymerization. The exact length of a primer will depend on many factors, including temperature, buffer, and nucleotide composition. Oligonucleotide primers will typically contain 12-20 or more nucleotides, although they may contain fewer nucleotides. Primers were designed to amplify amplicons covering coding exons, flanking intronic splice sites and branch-points, upstream promoter regions and downstream regulatory regions at the 3' untranslated region of the gene. A pair of primers were used for the amplification and melt

assay of the regions with no known SNPs while additional a set of probes were used individually for regions with known presence of SNPs (as provided by dbSNP in NCBI database).

[00075]. For regions with SNPs, a set of probes were designed based on the following features: A list of known Single Nucleotide Polymorphisms (SNPs) in the exonic regions can be obtained from online databases such as dbSNP. SNPs are avoided by placement of primers directly on the SNP or avoiding the SNP region entirely. Coverage of the SNP regions can be obtained by the use of unlabelled hybridization probes which are modified at the 3' end to prevent PCR extension. Hybridization probes are short fragments of DNA complementary to one strand of the DNA duplex. The probe works by effectively shortening the region of interrogation to the length of the probe, thereby magnifying any existing sequence difference. An important criteria in probe design is to have the SNP fall roughly in the middle of the probe as the further from the centre of the probe a SNP is located, the higher the probability the probe will fail to work as intended. The probes can be designed by Primer3 or any other primer design or probe selection software. Parameters for probe selection in Primer3 can be set. Hybridization Oligo Size range from 20-40 bases. Hybridization Oligo Tm range from 50-65°C, as it is vital that there be at least 10°C between the denaturation temperature of the probe with that of the full-length PCR product. Fluorescent-labeled probes can also be designed using the same parameters and software, provided the excitation wavelength of the fluorescent label is close to the emission wavelength of the dsDNA-binding dye and the detector is capable of reading the excitation wavelength of the fluorescent label.

[00076]. Nested primers are designed using an outer and inner pair of primers. The outer primers are designed based on coverage of the sequences amplified by the inner primers. All the inner primers are similar to those for the

normal melt analysis, and for the WGA enrichment method, outer primers are designed such that primer melting temperatures (Tm) range from 56-64°C and the maximum difference between the Tm of each primer in a pair should not exceed 1°C. Primer GC Content range from 10-90%. Primer Sizes range from 17-30 bases. Maximum Self-Complementarity is set at 6, while Maximum 3' Stability set to 8.

[00077]. Primers that may be used in diagnostic assays derived from the present invention should be designed to be substantially complementary to each strand of the DMD genomic gene sequence. This means that the primers must be sufficiently complementary to hybridise with their respective strands under conditions that allow the agent for polymerisation to perform. In other words, the primers should have sufficient complementarity with the 5' and 3' sequences flanking the mutation to hybridise therewith and permit amplification of the DMD genomic gene sequence.

[00078]. Oligonucleotide primers of the invention employed in the PCR amplification process that is an enzymatic chain reaction that produces exponential quantities of DMD gene sequence relative to the number of reaction steps involved. Typically, one primer will be complementary to the negative (-) strand of the DMd gene sequence and the other is complementary to the positive (+) strand. Annealing the primers to denatured nucleic acid followed by extension with an enzyme, such as the large fragment of DNA polymerase I (Klenow) and nucleotides, results in newly synthesised + and - strands containing the target a DMD gene sequence. Because these newly synthesized sequences are also templates, repeated cycles of denaturing, primer annealing, and extension results in exponential production of a region of the DMD gene sequence defined by the primers. The product of the chain reaction is a discreet nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

[00079]. Oligonucleotide primers may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized.

[00080]. The agent for polymerisation may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase, polymerase muteins, reverse transcriptase, other enzymes, including heat-stable enzymes (ie, those enzymes which perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation), such as Taq polymerase. Suitable enzyme will facilitate combination of the nucleotides in the proper manner to form the primer extension products that are complementary to each DMD gene sequence nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths.

[00081]. The newly synthesised DMD strand and its complementary nucleic acid strand will form a double-stranded molecule under hybridizing conditions described above and this hybrid is used in subsequent steps of the process. In the next step, the newly synthesized double-stranded molecule is subjected to HRM denaturing conditions using any of the procedures described above to provide heat melt profiles.

[00082]. The steps of denaturing, annealing, and extension product synthesis can be repeated as often as needed to amplify the target polymorphic gene sequence nucleic acid sequence to the extent necessary for detection. The amount of the specific nucleic acid sequence produced will accumulate in an exponential fashion. Amplification may also be achieved via real time PCR as known in the art.

[00083]. Preferably, the method of amplifying DMD is by PCR, as described herein or real time PCR and as is commonly used by those of ordinary skill in the art. Alternative methods of amplification have been described and can also be employed as long as the DMD gene sequence amplified by PCR using primers of the invention is similarly amplified by the alternative means. Such alternative amplification systems include but are not limited to self-sustained sequence replication, which begins with a short sequence of RNA of interest and a T7 promoter. Reverse transcriptase copies the RNA into cDNA and degrades the RNA, followed by reverse transcriptase polymerizing a second strand of DNA. Another nucleic acid amplification technique is nucleic acid sequence-based amplification (NASBA) which uses reverse transcription and T7 RNA polymerase and incorporates two primers to target its cycling scheme. NASBA can begin with either DNA or RNA and finish with either, and amplifies to 10^8 copies within 60 to 90 minutes. Alternatively, nucleic acid can be amplified by ligation activated transcription (LAT). LAT works from a single-stranded template with a single primer that is partially single-stranded and partially double-stranded. Amplification is initiated by ligating a cDNA to the promoter oligonucleotide and within a few hours, amplification is 10^8 to 10^9 fold. The QB replicase system can be utilized by attaching an RNA sequence called MDV-1 to RNA complementary to a DNA sequence of interest. Upon mixing with a sample, the hybrid RNA finds its complement among the specimen's mRNAs and binds, activating the replicase to copy the tag-along sequence of interest. Another nucleic acid amplification technique, ligase chain reaction (LCR), works by using two differently labeled halves of a sequence of interest that are covalently bonded by ligase in the presence of the contiguous sequence in a sample, forming a new target. The repair chain reaction (RCR) nucleic acid amplification technique uses two complementary and target-specific oligonucleotide probe pairs, thermostable polymerase and ligase, and DNA nucleotides to geometrically amplify targeted sequences. A 2-base gap separates the oligonucleotide probe pairs, and the RCR fills and joins the gap, mimicking normal DNA repair. Nucleic acid amplification by strand displacement activation (SDA) utilizes a short primer

containing a recognition site for hincII with short overhang on the 5' end that binds to target DNA. A DNA polymerase fills in the part of the primer opposite the overhang with sulfur-containing adenine analogs. HincII is added but only cuts the unmodified DNA strand. A DNA polymerase that lacks 5' exonuclease activity enters at the site of the nick and begins to polymerize, displacing the initial primer strand downstream and building a new one which serves as more primer. SDA produces greater than 10^7 -fold amplification in 2 hours at 37 degrees C. Unlike PCR and LCR, SDA does not require instrumented temperature cycling. Another amplification system useful in the method of the invention is the QB Replicase System. Although PCR is the preferred method of amplification if the invention, these other methods can also be used to amplify the JMJD6 or JMJD6 and estrogen receptor gene sequence as described in the method of the invention.

[00084]. A "sample", as used herein, refers to a biological sample obtained from body fluid or a tissue in the body, for example a biopsy. In a preferred embodiment the sample is will be a "clinical sample," which is a sample derived from a patient such as a fine needle biopsy sample or blood extraction. A "sample" may also include a section of tissue such as a section taken from a frozen or fixed tissue. Tissue samples can be obtained from muscle cells.

Diagnosis

[00085]. Diagnostic and prognostic methods will generally be conducted using a biological sample obtained from either a pre-natal or pre embryo implantation source or from any individual undergoing screening of muscle disease such as DMD or BMD or the like. A "sample" refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, organs, tissue including muscle tissue and samples of *in vitro* cell culture constituents.

[00086]. According to the diagnostic and prognostic methods of the present invention, alteration of the DMD gene sequence when compared to the DMD gene sequence taken from a normal, sample may be detected using anyone of the methods described herein. In addition, the diagnostic and prognostic methods can be performed to detect the DMD gene sequence expression and confirm the presence or absence of a functional dystrophin protein.

Detection kits

[00087]. Detection kits may comprise reagents for whole genome amplification, at least two nested primers specific to the gene of interest and reagents' for high resolution melt analysis. The kit may further contain one or more of the primers and probes of the invention, reagents for whole genome amplification, intercalating dyes, PCR reagents, cleaning or filtering reagents such as magnesium salts preferably $MgCl_2$ or physical filters as known in the art or any combination thereof. Kit components may be packaged for either manual or partially or wholly automated practice of the foregoing methods. In other embodiments involving kits, this invention contemplates a kit including compositions of the present invention, and optionally instructions for their use. Such kits may have a variety of uses, including, for example, screening for genetic mutations in large genes, Specifically in screening for uncommon mutations. In one embodiment the screening is for and other applications.

Examples of preferred embodiments

Protocol for Whole Genome Amplification of Genomic DNA followed by Product Purification and nested primer amplification for High Resolution Melting Applications

[00088]. This protocol is for the whole genome amplification (WGA) of extracted genomic DNA using Multiple Displacement Amplification method, for example, with use of Repli-g Midi Kit (Qiagen) followed by WGA product purification using Exonuclease I – Shrimp Alkaline Phosphatase (ExoSAP) digestion, PureLink™ PCR Purification Kit (Invitrogen), or Nested Polymerase

Chain Reaction (nPCR) prior to High Resolution Melting (HRM) with any of the methods incorporated as well.

For each reaction, 5 µl of 20 ng/µl DNA template used. This protocol is for a single reaction. Proportional scale-ups should be conducted for higher sample numbers.

Prepare Buffer D1 by diluting 0.7 µl of reconstituted Buffer DLB with 5 µl of sterile water.

Prepare Buffer N1 by diluting 1.1 µl of Stop Solution with 10.3 µl of sterile water.

Prepare Buffer DLB by adding 500 µl of sterile water to the tube.

Place the 5 µl of sample in a 1.5 ml Eppendorf tube and add 5 µl of Buffer D1.

Vortex briefly and pulse spin. Incubate at room temperature for 3 minutes.

Add 10 µl Buffer N1 to the sample. Vortex briefly and pulse spin.

Thaw Repli-g Polymerase on ice and all other components at room temperature. Vortex briefly and pulse spin.

Prepare Master Mix on ice by adding 1 µl of Repli-g Polymerase to 29 µl of Repli-g Reaction Buffer. Mix and pulse spin.

Add the 30 µl of Master Mix to the 20 µl of denatured DNA template.

Incubate at 30°C for 10-16 hours.

Heat the sample to 65°C for 3 minutes to denature the polymerase. WGA products can be stored frozen or refrigerated.

[00089]. WGA products can be quantitated spectrometrically and diluted to suitable working concentrations. However, measured quantities may not be

accurate due to difficulties in establishing a calibrating blank and the presence of nonspecific DNA products and various proteins.

PureLink™ PCR Purification Kit

[00090]. This section of the protocol deals with using the PureLink™ PCR Purification Kit to remove truncated and nonspecific WGA products and primer dimers by spin column techniques. Binding Buffer HC should be used as it will retain only DNA fragments larger than 600 bp after purification.

[00091]. The recommended volume of product to be purified is 50-100 µl. WGA products should be diluted at least 2 times before being applied to the spin column to reduce the viscosity and to prevent overloading the column filter. This protocol is for a single reaction, based on the 250 reaction PureLink™ PCR Purification Kit. Proportional scale-ups should be conducted for higher sample numbers.

Prepare Binding Buffer HC by adding 11 ml isopropanol to 109 ml undiluted Binding Buffer HC.

Prepare Wash Buffer by adding 160 ml 96-100% ethanol to 40 ml undiluted Wash Buffer.

Mix 4 volumes of prepared Binding Buffer HC with 1 volume of WGA product in an appropriate sterile microfuge tube.

Pipette the mixture into the provided PureLink™ Spin Column in Collection Tube without touching the filter. Centrifuge at 10,000g at room temperature for 60 seconds.

Discard the flowthrough liquid in the collection tube and briefly dry the collection tube by tapping it on a piece of clean absorbent paper. Reuse the collection tube for further steps.

Add 650 µl of prepared Wash Buffer to the spin column and centrifuge at 10,000g at room temperature for 60 seconds.

Discard flowthrough and dry the collection tube. OPTIONAL: Repeat Steps 8 & 9.

Return collection tube to spin column and centrifuge the spin column at maximum speed for 3 minutes.

Discard collection tube and place spin column in provided sterile 1.7 ml PureLink™ Elution Tube.

Add 30 µl Elution Buffer or sterile water (pH 7-8) to the spin column and incubate at room temperature for 2 minutes.

Centrifuge for at maximum speed for 2 minutes at room temperature.

Repeat Step 12 using 20 µl of Elution Buffer or sterile water. The combined DNA-containing elute volume should be around 48 µl.

Purified WGA products can be quantitated spectrometrically and diluted to suitable working concentrations.

ExoSAP

[00092]. This section of the protocol deals with using the Exonuclease I – Shrimp Alkaline Phosphatase enzymatic digestion to remove truncated and nonspecific WGA products and primer dimers. The protocol for 10 reactions is given for ease of calculation. For more or less reactions, appropriate scale-ups or scale-downs should be performed.

Prepare 6 µl of each WGA product to be purified in sterile microfuge tubes or PCR plates.

Prepare a master mix on ice in the following order:

Sterile water	38.1 µl
Exonuclease I (20,000U/ml)	0.7 µl
Shrimp Alkaline Phosphatase (1U/µl)	3.2 µl
Total	**42.0 µl**

Aliquot 4 µl of the master mix into each well or tube containing 6 µl of WGA product.

Incubate the reactions at 37°C for 40 minutes, followed by an enzyme denaturation step at 85°C for 10 minutes.

Purified WGA products can be quantitated spectrometrically and diluted to suitable working concentrations. However, measured quantities may not be accurate due to difficulties in establishing a calibrating blank and the presence of various proteins.

Nested PCR

[00093]. This section of the protocol deals with using nested PCR (nPCR) for the dilution of nonspecific WGA products and primer dimers during the PCR stage. For larger sample sizes, appropriate scale-ups should be performed. This protocol utilizes batches of 8 individual DNA samples (henceforth referred to as a 'set') per pair of primers. Each set serves as 1`HRM analyses unit.

An example is shown as follows:

[00094]. NESTED PRIMER SEQUENCES

Exon 2 :

Outer Primers:
F: 5' – gcctggccatttttcaga – 3'
R: 5'- atatttccagatttgcacagctaa – 3'

Inner Primers:

F: 5' – catcataatggaaagttactttg – 3'
R: 5' – agatcttaaaagtaaagtaacaaacc– 3'

Exon 30 :

Outer Primers:
F: 5' – ggaagctgcgaaatctgtct – 3'
R: 5'- atcaaaacaaccccatggaa – 3'

Inner Primers:
F: 5' – acatttaactgatacactcttattcc – 3'
R: 5' – ACTGCTTGTCATGAATGTG – 3'

Exon 37 :

Outer Primers:
F: 5' – gcatgtgcttgctctcattt – 3'
R: 5'- aaaccttgctgtggggtcta – 3'

Inner Primers:
F: 5' – GCAAACTTGATGGCAAA – 3'
R: 5' – aagtttccaccttggagtag – 3'

Exon 48

Outer Primers:
F: 5' – tgctgctaaaataacacaaatcagt – 3'
R: 5'- gtccctgtgcctattgtggt – 3'

Inner Primers:
F: 5' – AAGCTTGAAGACCTTGAAGAGC – 3'
R: 5' – cctgaataaagtcttccttaccacac – 3'

[00095]. The nPCR method involves performing an initial round of PCR using primers which form a large amplicon that encompasses the smaller amplicon formed by the primers used for the latent round of PCR and HRM. The PCR product of the first round is then used as the template for the second round of PCR. This effectively increases the concentration of template DNA (large amplicon) while simultaneously decreasing the amount of nonspecific DNA fragments and primer dimers from the WGA step.

Prepare 1 master mix for each set by mixing the following in order in a sterile microfuge tube:

Sterile autoclaved water	To top up to total of 76.0 µl
10x PCR Buffer	8.4 µl
$MgCl_2$ solution	As needed for optimal efficiency
10 mM dNTPs mix	1.7 µl
Forward Primer 10 µM	2.5 µl
Reverse Primer 10 µM	2.5 µl
Polymerase (5 U/ µl)	0.8 µl
Total	**76.0 µl**

Deposit 1 µl of requisite DNA template into PCR plate well or microfuge tube and add 9 µl aliquots of the master mix to the template. Repeat this for all samples in the same set.

Pulse spin the PCR plate or tubes and PCR performed on the sample using the following cycling conditions:

Temperature (°C)	Time (s)	Number of cycles
94	120	1
94	30	35
Optimal Annealing Temperature	30	
72	60	
72	120	1
12	As required	1

*Note that the cycling conditions may change depending on the requirements of the polymerase and thermal cycler being used

[00096]. 1 µl of the PCR products will be used as template for the HRM PCR. Alternatively, PCR products can be quantitated spectrometrically and diluted to suitable working concentrations. However, measured quantities may not be accurate due to difficulties in establishing a calibrating blank and the presence of various proteins.

Protocol for Symmetrical PCR Followed by HRM (Glass capillaries)

[00097]. This protocol is for the amplification of multiple 10 µl Polymerase Chain Reaction (PCR) in LightCycler-format glass capillaries for High Resolution Melting (HRM) simultaneous mutation scanning and Single Nucleotide Polymorphism (SNP) genotyping of DMD gene amplicons. DNA template for use should be of as high purity as possible and at a quantity of at least 10 ng per reaction.

[00098]. DNA probes blocked with phosphate groups at the 3' end to prevent PCR extension are used in excess of the primers to induce the formation of probe-amplicon duplexes during and after asymmetric PCR. The probe can be designed to be complementary to either strand of the amplicon but the primer which primes the complementary strand must be present in 10-20 times higher amount than its partner. For best effect the SNP base should be located near the middle of the probe.

[00099]. List of saturating fluorescent HRM dyes and their concentrations used are given below:

Dye	Source	10x Concentration	1x Final Concentration for HRM
LC Green+	Idaho Technology	Proprietary information	Proprietary information
EvaGreen	Biotium Inc.	20 µM	2 µM
BEBO	TATAA Biocentre	50 µM	5 µM
SYTO 9	Invitrogen	15 µM	1.5 µM
ResoLight	Roche	Proprietary information	Proprietary information

*Optimal concentrations for EvaGreen, BEBO, and SYTO 9 were empirically determined while LC Green+ and ResoLight are sold in fixed concentrations

[000100]. Most polymerases can be used for HRM PCR, but reaction inhibition by the HRM dye may occur. Platinum Taq (Invitrogen) has been tested to work with all 5 dyes while Amplitaq Gold (Applied Biosystems) works with all dyes except LC Green+.

[000101]. HRM-capable platforms currently on the market which use LightCycler-format glass capillaries are the LightScanner32 (IdahoTechnology)

and Rotor-Gene Q (Qiagen), but this protocol can be applied to fit any capillary-based HRM platform.

[000102]. This protocol utilizes batches of 8 individual DNA samples (henceforth referred to as a 'set') per pair of primers. Each set serves as 1`HRM analyses unit. LightScanner32 can accommodate 32 capillaries (4 sets) and Rotor-Gene Q can accommodate either 72 (9 sets) or 100 (12 sets) capillaries per run depending on configuration. In other words, the LightScanner32 can screen 4 primer pairs of 8 DNA samples each per run. Primer pairs to be screened together in one run must possess identical PCR amplification and HRM conditions.

[000103]. Number of DNA samples screened can also be increased in multiples of 8 by correspondingly reducing the number of sets screened.

Prepare 1 master mix for each set by mixing the following in order in a sterile microfuge tube:

Sterile autoclaved water	To top up to total of 76.0 µl
10x PCR Buffer	8.4 µl
$MgCl_2$ solution	As needed for optimal efficiency
10 mM dNTPs mix	1.7 µl
Forward Primer 0.5 µM - 1 µM	2.5 µl
Reverse Primer 10 µM	2.5 µl
Probe 10 µM	3.4 µl
10x Fluorescent HRM dye	8.4 µl
Polymerase (5 U/ µl)	0.8 µl
Total	**76.0 µl**

Deposit 1 µl of requisite DNA template into a glass capillary and add 9 µl aliquots of the master mix to the template. Repeat this for all samples in the same set.

8 µl of molecular biology-grade mineral oil added to the capillaries before sealing them with the capillary caps.

PCR performed on the sample using the following cycling conditions:

Temperature (°C)	Time (s)	Number of cycles
94	90	1
94	10	60
Optimal Annealing Temperature	10	
72	20	
72	30	1
12	As required	1

*Note that the cycling conditions may change depending on the requirements of the polymerase and thermal cycler being used
**When using LC Green+ as the fluorescent dye, it is necessary to raise the annealing temperature by around 6°C as the dye increases the stability of double-stranded DNA. For best results it may be necessary to re-optimise the PCR conditions in the presence of the dye

Transfer the PCR products to a HRM-capable platform if necessary and perform an initial denaturation at 95°C for 45 s. Cool the PCR products down to 40-45°C to allow renaturation before performing HRM from 50°C to 95°C. This melt will contain both the probe melt and full length amplicon melt data.

A 2nd melt of the temperature region of the probes is then conducted to obtain only the probe melt peaks data. PCR products can be stored refrigerated or frozen and remelted when necessary.

Protocol for Symmetrical PCR Followed by HRM (Glass capillaries)

[000104]. This protocol is for the amplification of multiple 10 µl Polymerase Chain Reaction (PCR) in LightCycler-format glass capillaries for High Resolution Melting (HRM) mutation scanning of DMD gene amplicons. DNA template for use should be of as high purity as possible and at a quantity of at least 10 ng per reaction.

[000105]. List of saturating fluorescent HRM dyes and their concentrations used are given below:

Dye	Source	10x Concentration	1x Final Concentration for HRM
LC Green+	Idaho Technology	Proprietary information	Proprietary information
EvaGreen	Biotium Inc.	20 µM	2 µM
BEBO	TATAA Biocentre	50 µM	5 µM

SYTO 9	Invitrogen	15 µM	1.5 µM
ResoLight	Roche	Proprietary information	Proprietary information

*Optimal concentrations for EvaGreen, BEBO, and SYTO 9 were empirically determined while LC Green+ and ResoLight are sold in fixed concentrations

[000106]. Most polymerases can be used for HRM PCR, but reaction inhibition by the HRM dye may occur. Platinum Taq (Invitrogen) has been tested to work with all 5 dyes while Amplitaq Gold (Applied Biosystems) works with all dyes except LC Green+.

[000107]. HRM-capable platforms currently on the market which use LightCycler-format glass capillaries are the LightScanner32 (IdahoTechnology) and Rotor-Gene Q (Qiagen), but this protocol can be applied to fit any capillary-based HRM platform.

[000108]. This protocol utilizes batches of 8 individual DNA samples (henceforth referred to as a 'set') per pair of primers. Each set serves as 1`HRM analyses unit. LightScanner32 can accommodate 32 capillaries (4 sets) and Rotor-Gene Q can accommodate either 72 (9 sets) or 100 (12 sets) capillaries per run depending on configuration. In other words, the LightScanner32 can screen 4 primer pairs of 8 DNA samples each per run. Primer pairs to be screened together in one run must possess identical PCR amplification and HRM conditions.

[000109]. Number of DNA samples screened can also be increased in multiples of 8 by correspondingly reducing the number of sets screened.

Prepare 1 master mix for each set by mixing the following in order in a sterile microfuge tube:

Sterile autoclaved water	To top up to total of 76.0 µl
10x PCR Buffer	8.4 µl
$MgCl_2$ solution	As needed for optimal efficiency
10 mM dNTPs mix	1.7 µl
Forward Primer 10 µM	2.5 µl
Reverse Primer 10 µM	2.5 µl
10x Fluorescent HRM dye	84 µl
Polymerase (5 U/ µl)	0.8 µl

Total	76.0 µl

Deposit 1 µl of requisite DNA template into a glass capillary and add 9 µl aliquots of the master mix to the template. Repeat this for all samples in the same set.

8 µl of molecular biology-grade mineral oil added to the capillaries before sealing them with the capillary caps.

PCR performed on the sample using the following cycling conditions:

Temperature (°C)	Time (s)	Number of cycles
94	90	1
94	10	35
Optimal Annealing Temperature	10	
72	20	
72	30	1
12	As required	1

*Note that the cycling conditions may change depending on the requirements of the polymerase and thermal cycler being used
**When using LC Green+ as the fluorescent dye, it is necessary to raise the annealing temperature by around 6°C as the dye increases the stability of double-stranded DNA. For best results it may be necessary to re-optimise the PCR conditions in the presence of the dye

[000110]. Transfer the PCR products to a HRM-capable platform if necessary and perform an initial denaturation at 95°C for 45 s. Cool the PCR products down to 40-50°C to allow renaturation before performing HRM from 65°C to 95°C. PCR products can be stored refrigerated or frozen and remelted when necessary.

Protocol for Asymmetrical PCR Followed by HRM (PCR plate)

[000111]. This protocol is for the amplification of multiple 10 µl Polymerase Chain Reaction (PCR) in a 96- (or 384-) well PCR plate for High Resolution Melting (HRM) simultaneous mutation scanning and Single Nucleotide Polymorphism (SNP) genotyping of DMD gene amplicons. DNA template for use should be of as high purity as possible and at a quantity of at least 10 ng per reaction.

[000112]. DNA probes blocked with phosphate groups at the 3' end to prevent PCR extension are used in excess of the primers to induce the formation of probe-amplicon duplexes during and after asymmetric PCR. The probe can be designed to be complementary to either strand of the amplicon but the primer which primes the complementary strand must be present in 10-20 times higher amount than its partner. For best effect the SNP base should be located near the middle of the probe.

[000113]. List of saturating fluorescent HRM dyes and their concentrations used are given below:

Dye	Source	10x Concentration	1x Final Concentration for HRM
LC Green+	Idaho Technology	Proprietary information	Proprietary information
EvaGreen	Biotium Inc.	20 μM	2 μM
BEBO	TATAA Biocentre	50 μM	5 μM
SYTO 9	Invitrogen	15 μM	1.5 μM
ResoLight	Roche	Proprietary information	Proprietary information

*Optimal concentrations for EvaGreen, BEBO, and SYTO 9 were empirically determined while LC Green+ and ResoLight are sold in fixed concentrations

[000114]. Most polymerases can be used for HRM PCR, but reaction inhibition by the HRM dye may occur. Platinum Taq (Invitrogen) has been tested to work with all 5 dyes while Amplitaq Gold (Applied Biosystems) works with all dyes except LC Green+.

[000115]. HRM platforms capable of probe melt peak analyses currently on the market which use 96- or 384-well plate formats are the LightScanner (Idaho Technology) and LightCycler 480 (Roche), but this protocol can be applied to fit any PCR plate-based HRM platform.

[000116]. Roche and Bio-Rad manufacture white-welled plates optimized for use in their own platform, and it is recommended that their respective, recommended plates be used. For the LightScanner, Bio-Rad's hard-skirted

white plates(Catalogue code: HSP 9655) is recommended for 96-well format, though other compatible plates can also be utilized.

[000117]. This protocol utilizes batches of 8 individual DNA samples (henceforth referred to as a 'set') per pair of primers. Each set serves as 1`HRM analyses unit. 96-well plates can accommodate 12 sets and 384-well plates can accommodate 48 sets. In other words, 96-well plates can be used to screen 12 primer pairs of 8 DNA samples each per run. Primer pairs to be screened together in one run must possess identical PCR amplification and HRM conditions.

[000118]. Number of DNA samples screened can also be increased in multiples of 8 by correspondingly reducing the number of sets screened.

Prepare 1 master mix for each set by mixing the following in order in a sterile microfuge tube:

Sterile autoclaved water	To top up to total of 76.0 µl
10x PCR Buffer	8.4 µl
$MgCl_2$ solution	As needed for optimal efficiency
10 mM dNTPs mix	1.7 µl
Forward Primer 0.5 µM - 1 µM	2.5 µl
Reverse Primer 10 µM	2.5 µl
Probe 10 µM	3.4 µl
10x Fluorescent HRM dye	8.4 µl
Polymerase (5 U/ µl)	0.8 µl
Total	**76.0 µl**

Deposit 1 µl of requisite DNA template into PCR plate well and add 9 µl aliquots of the master mix to the template. Repeat this for all samples in the same set.

12 µl of molecular biology-grade mineral oil added to the wells before the PCR plate is sealed with optically-clear adhesive seal.

Subject the plate to centrifugation of 2000-3000 rpm for 60-120 seconds to separate the oil and aqueous phases.

PCR performed on the sample using the following cycling conditions:

Temperature (°C)	Time (s)	Number of cycles
94	120	1
94	30	60
Optimal Annealing Temperature	30	
72	30	
72	60	1
12	As required	1

*Note that the cycling conditions may change depending on the requirements of the polymerase and thermal cycler being used
**When using LC Green+ as the fluorescent dye, it is necessary to raise the annealing temperature by around 6°C as the dye increases the stability of double-stranded DNA. For best results it may be necessary to re-optimise the PCR conditions in the presence of the dye

[000119]. Transfer the PCR products to a HRM-capable platform if necessary and perform an initial denaturation at 95°C for 45 s. Cool the PCR products down to 40-45°C to allow renaturation before performing HRM from 50°C to 95°C. This melt will contain both the probe melt and full length amplicon melt data.

[000120]. A 2nd melt of the temperature region of the probes is then conducted to obtain only the probe melt peaks data. PCR products can be stored refrigerated or frozen and remelted when necessary.

Protocol for Symmetrical PCR Followed by HRM (PCR plate)

[000121]. This protocol is for the amplification of multiple 10 µl Polymerase Chain Reaction (PCR) in a 96- (or 384-) well PCR plate for High Resolution Melting (HRM) mutation scanning of DMD gene amplicons. DNA template for use should be of as high purity as possible and at a quantity of at least 10 ng per reaction.

[000122]. List of saturating fluorescent HRM dyes and their concentrations used are given below:

Dye	Source	10x Concentration	1x Final Concentration for HRM
LC Green+	Idaho Technology	Proprietary information	Proprietary information

EvaGreen	Biotium Inc.	20 μM	2 μM
BEBO	TATAA Biocentre	50 μM	5 μM
SYTO 9	Invitrogen	15 μM	1.5 μM
ResoLight	Roche	Proprietary information	Proprietary information

*Optimal concentrations for EvaGreen, BEBO, and SYTO 9 were empirically determined while LC Green+ and ResoLight are sold in fixed concentrations

[000123]. Most polymerases can be used for HRM PCR, but reaction inhibition by the HRM dye may occur. Platinum Taq (Invitrogen) has been tested to work with all 5 dyes while Amplitaq Gold (Applied Biosystems) works with all dyes except LC Green+.

[000124]. HRM-capable platforms currently on the market which use 96- or 384-well plate formats are the LightScanner (Idaho Technology), LightCycler 480 (Roche), and CFX96/CFX384 (Bio-Rad), but this protocol can be applied to fit any PCR plate-based HRM platform.

[000125]. Roche and Bio-Rad manufacture white-welled plates optimized for use in their own platform, and it is recommended that their respective, recommended plates be used. For the LightScanner, Bio-Rad's hard-skirted white plates(Catalogue code: HSP 9655) is recommended for 96-well format, though other compatible plates can also be utilized.

[000126]. This protocol utilizes batches of 8 individual DNA samples (henceforth referred to as a 'set') per pair of primers. Each set serves as 1`HRM analyses unit. 96-well plates can accommodate 12 sets and 384-well plates can accommodate 48 sets. In other words, 96-well plates can be used to screen 12 primer pairs of 8 DNA samples each per run. Primer pairs to be screened together in one run must possess identical PCR amplification and HRM conditions.

[000127]. Number of DNA samples screened can also be increased in multiples of 8 by correspondingly reducing the number of sets screened.

Prepare 1 master mix for each set by mixing the following in order in a sterile microfuge tube:

Sterile autoclaved water	To top up to total of 76.0 µl
10x PCR Buffer	8.4 µl
$MgCl_2$ solution	As needed for optimal efficiency
10 mM dNTPs mix	1.7 µl
Forward Primer 10 µM	2.5 µl
Reverse Primer 10 µM	2.5 µl
10x Fluorescent HRM dye	84 µl
Polymerase (5 U/ µl)	0.8 µl
Total	**76.0 µl**

Deposit 1 µl of requisite DNA template into PCR plate well and add 9 µl aliquots of the master mix to the template. Repeat this for all samples in the same set.

12 µl of molecular biology-grade mineral oil added to the wells before the PCR plate is sealed with optically-clear adhesive seal.

Subject the plate to centrifugation of 2000-3000 rpm for 60-120 seconds to separate the oil and aqueous phases.

PCR performed on the sample using the following cycling conditions:

Temperature (°C)	Time (s)	Number of cycles
94	120	1
94	30	35
Optimal Annealing Temperature	30	
72	30	
72	60	1
12	As required	1

*Note that the cycling conditions may change depending on the requirements of the polymerase and thermal cycler being used
**When using LC Green+ as the fluorescent dye, it is necessary to raise the annealing temperature by around 6°C as the dye increases the stability of double-stranded DNA. For best results it may be necessary to re-optimise the PCR conditions in the presence of the dye

[000128]. Transfer the PCR products to a HRM-capable platform if necessary and perform an initial denaturation at 95°C for 45 s. Cool the PCR products down to 40-50°C to allow renaturation before performing HRM from 65°C to 95°C. PCR products can be stored refrigerated or frozen and remelted when necessary.

Primer design

[000129]. The primers are listed in Table 1.

Exon	F Primer (5'-3')	Primer size (bp)	R Primer (5'-3')	Primer size (bp)
1P	GCCAAGCACAGATAAAGCAG	20	GTAGATTGAGCCGGTGAGG	19
1C	CATGATTCTTTCAGTCATTTATTATGC	27	AAGAGATTTGCAGATCCACTGTT	23
1L	AAAACGGATTTTTAAGATACACAGG	25	ATGAATTATATTTAAAGTTGCTTCCTAACT	30
1Ma	tccactgtgctattctggtttg	22	gtagaggcccccggatatt	19
1Mb	agagagaaggcgggtcactt	20	AAAGTAACACTTCAGTTTTTCCTATTCG	28
1Mc	GGCAATTACCTTCGGAGAAA	20	CTACTTCTTCCCACCAAAGCA	29
1Md	CGCTGCCTTGATATACACTTTTC	23	tgcttctttgcaaactactgtgat	24
2	catcataatggaaagttactttg	23	agatcttaaaagtaaagtaacaaacc	26
3a	tttttcatccgtcatcttcg	20	CCTCCCATCCTGTAGGTCAC	20
3b	ggaagtgtgctttgttaaattgag	24	cagtttctggtctgaaattctactaagtt	29
4A	atgcctcacaggctctgttc	20	GCCTTGTTGACATTGTTCAGG	21
4b	gtagattgtcggtctctctgctg	23	acagcatccagaccttgtcc	20
5	ttattgcaactaggcatttggtc	23	cccaaaaggaaaccattcatc	21
6a	tctatttatcactgaagatcaaggaca	27	TTTTCACTGTTGGTTTGTTGC	21
6b	ttgcaacaaaccaacagtga	20	tcagagtctaaatcaccacttttaca	26
7a	aggactatgggcattggttg	20	TGGCAAACCACACTATTCCA	20
7b	tttgtctttgtgtatgtgtgtatgtgt	27	gcagtggtagtccagaaatttacc	24
8A	tcgtcttcctttaactttgatttgt	25	TGGCTTCAATGCTCACTTGT	20
8B	cagATGTTGATACCACCTATCCAG	24	TGCATTTGATGATGTAGCTGAA	22
9A	ccttcattctgggagtataccaa	23	GTGCTAGACTGACCGTGATctg	22
9B	tctctgcagATCACGGTCAG	20	ttatcttggaagcagttctctgg	23
10A	ctggaacaatctgcaaagacat	27	AAATCTCTCCGTGTGCTTGC	20
10B	AGCACACGGAGAGATTTCTAATG	23	gttggaatcccaagcacatc	20
11A	aatctaaatgggccacaagttt	22	TTTTCCTGTTCCAATCAGCTT	21
11B	TGATTGGAACAGGAAAATTATCAG	24	ttaaccatcaaaccacatcaaaa	23
12A	aatagatgcccccaaatgc	19	TGTTCTTTCTTCTGTTTTTGTTGC	24
12B	TTAATGGATCTCCAGAATCAGAAAC	25	ctagtagaaagcacgcaacataaga	25
13A	tcaccatttgagagcaaatca	21	TTGTTCTTCCAAAGCAGCAG	20
13B	TGATGAATCTAGTGGAGATCACG	23	cagctgattatgagtgtgtgtatattg	27
17A	tggtgctattttgatctgaagg	22	CAGCCTGTGAAATctgtgaga	21
17B	GATTTCACAGGCTGTCACCA	20	tCCACAGTAATCTGCCTCTTCTT	23
18A	aaagaatgataaatgtggataaattgc	27	TGAGAAGTTGCCTTCCTTCC	20
18B	CTCGCTCAGAAGCTGTGTTG	20	caaagcacggagtttacaagc	21
19A	ttaaggcttgaaagggcaag	20	GAGCTGATCTGCTGGCATC	19
20A	ttttataattaggatgtgttggctttc	27	ACTGGCAGAATTCGATCCAC	20

20B	GCCTCAGAACAACTGAACAGC	21	GGTGGTGGGTTGGATTTTC	19
20C	TCAGAACAACATCATCGCTTTC	22	attatgctccaaatggaaggag	22
21A	ggctggtgatagaggcttgt	20	GAAGATCTGATAGCCGGTTGA	21
21B	TCAACCGGCTATCAGATCTTC	21	TCTTTCTCTCTGACCTGCACA	21
22A	taatataattgagtttgctgacaatttagg	30	CCATGTCCTGATGGCACTC	19
22B	ATGCGCTATCAGGAGACCAT	20	gcttgataagcgtgctttattg	22
23A	gatcatctactttgtttacatgtttgaa	29	GCGCTTTCTTCGACATCTCT	20
23B	TGAAAGAGATGTCGAAGAAAGC	22	GCTTTTGACAATGCTCAACC	20
24	tgttaaaagtaatcagcacaccagt	25	tcatacaaaattattcatattaaaggcatc	30
25A	caattttacattttctagctatgtttca	28	ATCTTCTGCCCACCTTCATT	20
25B	tagCTTTTAGTCAGTGATATTCAGACAAT	29	CATGTGATCCCACTGAGAGTTAAG	24
26A	tctgatccccatgagttattttct	22	AAGGCCTCCTTTCTGGCATA	20
26B	gtttgtttgttttgtggaagGTC	23	TTCATCTCTTCAACTGCTTTCTGT	24
27A	tgcattttggatgtaaagttattttc	26	CTGGTAGTTGATGGTTAGAGTTTCAA	26
27B	GAAACTCTAACCATCAACTACCAGTG	26	tgcctgaatgaggaattcaa	20
28A	cctgctacaaagtaaaggtgaaaa	24	CCACTTGTTTGCTTTCTCCA	20
28B	AAGCAAACAAGTGGCTAAATGAA	23	tcttgggttgttttctttgga	21
29A	tgctccttggtgtttttaatcc	22	TCCTCTGAATGTCGCATCAA	20
29B	TTTGATGCGACATTCAGAGG	20	ttagattaaagagatttttcacttatcttca	31
30A	caaaaaggtgattgtggaagag	22	TTAAGTGTAAGGAATTTTCAGTCTCC	26
30B	TCCAGGAGTCCCTCACAGTC	20	aatggaagctgattcccaga	20
31A	tggtagaggtggttgaggagag	22	GCCTCCTTCCCCTGATTATG	20
31B	gactaataatgctatcctcccaaca	25	ccaacgaaaacacgttcctt	20
32A	ctgctttattttgtttttattttttctgt	28	TAGACGCAGCTCAAAATTGG	20
32B	TTGAGCTGCGTCTACAAGAAAG	22	gcgtatttgccaccagaaat	20
33A	aggggatctctatttatttctgttcat	27	GGATTTTCCGTCTGCTTTTTC	21
33B	GAAGTGAAGTCTGAAGTGGAAATG	20	ttgtggtctcagcatgcac	19
34A	ggttataacgaaatttgaattaaagagta	29	CAACTGCTGATCTCTTTGTCAAT	23
34B	ATTGACAAAGAGATCAGCAGTTG	23	ttcacgtatgttcaaaataaccttc	25
35A	ccatacagaaagccgtttca	20	TCTCCTTCTTGCCCAAAACT	20
35B	CAGAGGTAGGAGAGGCCTTG	20	GGAGGTGACAGCTATCCAGTTAC	23
36A	aataccacttaaaactaatctcaatgaaac	30	GGTCCACATTCTGGTCAAAAG	21
36B	tgccatggtatgtctctgtacaat	24	tggacattacttttcatattttatttgc	28
37A	gcatgtgcttgctctcattt	20	AGTCCACCTTTGGGCGTAT	19
37B	AACCACGGTGACCACTGC	18	agaccatttagcacaagtttcca	23
38A	tttagGCCTCCATTCCTTTG	20	tttccactcctagttcattcacac	24
38B	CCTTTGAAGGAATTGGAGCA	20	tcattcacacttttatcacaacca	24
40A	aaaagatgagggacgcaaatta	22	GTCATCCAAGCATTTCAGGAG	21
40B	TACAAGAGGCAGGCTGATGA	20	caacattttaaatatgatcttcacagg	27
41A	cttgcaagtcggttgatgtg	20	GCACTGCATTCAGCTCCTCT	20
41B	AAAGAGGAGCTGAATGCAGTG	21	acatacgtgggtttgccagt	20

42A	gaaatgcttttaacactttctgga	24	TCAAATAAGTAGAAGGCACATAAGAAA	27
42B	GCACACTGTCCGTGAAGAAA	20	TTCAGAGACTCATCTTGCTTAAAGA	25
43A	ccatttgctacctttgggatt	21	CACAGGCGTTGCACTTTG	18
43B	GCAAGAAGACAGCAGCATTG	20	tgaaaacaaatcatttctgcaag	23
44A	tgcaaatgcaggaaactatca	21	AAACGCCGCCATTTCTC	17
44B	ATGGCGGCGTTTTCATT	17	aagagtccagatgtgctgaaga	22
45A	ttcacatggagcttttgtatttct	24	AGTTTGCCACTGCCCAAT	18
45B	ATTGGGCAGTGGCAAACT	18	aaatgttttcattcctattagatctgtc	28
46a	ttaaattgccatgtttgtgtcc	22	TTGCTGCTCTTTTCCAGGTT	20
46b	ttctccaggCTAGAAGAACAAAA	23	cctgggggatttgagaaaata	21
47A	tttcagtcaatcagctctgtgc	22	GGGCAACTCTTCCACCAGTA	20
47B	cgttgttgcatttgtctgtttc	22	gacggaagagatggttaatgtctaa	26
48A	tgctaaaataacacaaatcagtaagattc	29	TTTCAAGCTGCCCAAGGT	18
48B	cagGTTTCCAGAGCTTTACCTG	22	CGTCAAATGGTCCTTCTTGG	20
49A	tctgtttcttttctctgcacca	22	CCGGTTGTTTAGCTTGAACTG	21
49B	ctgcactatatgggttcttttcc	23	gcaaatgtacaacaggggaag	21
50a	tcctttaaaagaaattctacccactaa	27	GTTTACCGCCTTCCACTCAG	20
50b	cgaataagtaatgtgtatgcttttctg	27	ccaaagagaatgggatccag	20
51A	tgaaattggctctttagcttgtg	23	TGAAATCTGCCAGAGCAGGTA	21
51B	TGCCATCTTCCTTGATGTTG	20	aaacttctgccaacttttatcatttt	26
52A	ttctaaaagtgttttggctggtc	23	GGCAGCGGTAATGAGTTCTT	20
52B	GGCAACAATGCAGGATTTG	19	ttgtgtgtcccatgcttgtt	20
53A	aaatgtgagataacgtttggaagt	24	TCCTTTAACATTTCGTTCAACTG	23
53B	ACAGTTGAAcGAAATGTTAAAGGA	24	ccagtattttatttaaacaggtatctttg	30
54A	gacctgaggattcagaagctg	21	GTTTCAGGGCCAAGTCATTT	20
54B	GCCAGTGGCAGACAAATGTA	20	tcaccaccccattattacagc	21
55A	caactcaccccattgttggta	21	TTCTTCCAAAGCAGCCTCTC	20
55B	GGCTGCTTTGGAAGAAACTC	20	agcggaaatgcctgacttac	20
56A	atattttgcaattctccaaattcac	25	TTTTTGGCTGTTTTCATCCAG	21
56B	TCCAAGGTGAAATTGAAGCTC	21	tttttgctccacatcttttcc	21
57A	attcaagactgctttgtagttcaca	25	GCCACACCAGCAAGTTCC	18
57B	GGAACTTGCTGGTGTGGCTA	20	aaggcacgaggcttaaaaatg	21
58A	gacctgggagtttcataaacaag	21	GCTGCTCTGTCAGAAATATTCG	22
58B	cacttcttttcatctcatttcacag	25	ccgtcaccactgatccttc	19
60A	actggcactgcaccctaaag	20	AGAGCTGAATGCCCAAAGTG	20
60B	CCTTGCTCGCCAGCTTAC	18	actttcattgttctttacaaattttatctg	30
61	gaaaggatacattgttttaattgttcc	27	tgcaataaagttaagtgataaaagctg	27
62	tgtctttcctgtttgcgatg	20	aatgctcttttaaaaatgtgatacttcc	28
63	catgttgttgttattgttgttttcttt	27	catttaacttggaggaaacatgg	23
64	gggtgggttgtctgttatttct	22	ccaacctacttttattctaagcaaag	27
65A	ttattaaagggtatgagagagtcctagc	28	AGGATATCCATGGGCTGGT	19

65B	GCACAACCTCAAGCAAAATG	20	agagcccgggaaataaaaac	20
66A	tttttccttttcaaggctttattc	24	ACACGGATCCTCCCTGTTC	19
66B	aaagggtcagtaattgttttctgc	24	aaaagaatacagcattagtatacacgactt	30
67A	ctgtggaaatactggctactcttg	24	TGGATAGAATCATGCAGAAGGA	22
67B	GCAAGTGGCAAGTTCAACAG	20	aacttacAAACTGGAAGCAGCTC	23
68A	gctttgcaaccattgttctttc	20	GGGGTTCCAGTCTCATCCA	19
68B	AATAATAAGCCAGAGATCGAAGC	19	aaaggtgaaaaatgatgagaaaaa	26
69A	ttgtgggactaatgaacattgc	22	CCTTTTGCAACTCGACCAG	19
69B	ATTATGACATCTGCCAAAGCTG	22	actgaaatttatcccaggtgaact	24
70A	agtgtcatggggcagaagac	20	AGTCTGCACTGGCAGGTAGC	20
70B	AAGGTATTTTGCGAAGCATCC	21	gggtgttcagctgagaggag	20
71	ggtttggctattgctttcca	20	cagcacccttcagcaaaaa	19
72	aaaagcattctaggccatgtg	21	gcctggcatacaactagtctca	22
73	tttcaggaatgttcgattaggtc	23	agcaatttcattgtcaggaaca	22
74A	tctattttcaaatacactcctgagtcc	27	TTTCCTCACTCTCTAAGGAAATCAAG	26
74B	AATCCAGCATTACTGCCAAAG	21	ggcacttttctatgtgtgcaag	22
75A	ccatggtatataaaatttggtgatga	26	AGAGGTGGGCATCATTTCAG	20
75B	CATAAAGGCCTGTCCCCACT	20	TCCAGCAGCTGCCTTAGC	18
76A	agcgtcaaaatttatgagtcctg	23	ACACCGTTGTGCCATTCAC	19
76B	aattctgttttcttttggatgacttag	27	cctttcttcagacaacaaaatctg	24
77	catggccctttaatatctgttttc	24	agcaaatctgagtcccttctagg	23
79A	tgcagcttctgtgttgtcttc	21	ACTAAGGACTCCATCGCTCTG	21
79B	CCACATGGCAGATGA	15	TTTATTCTGCTCCTTCTTC	19
79C	GCAGAGCGATGGAGTCCTTA	20	AATACCACTACCCTTCACAAAAATATAGA	29
79D	GAGGATTAGACAGTAAGAGTTTACAAGAA	29	AAATGGCAAGTTATTTAGCTATCAAG	26
79E	CAGGTTTTACACGTCTATGCAAT	23	AAAGCACACTTTAGTTTACAATCTTTCT	28
79F	TCTTTATATGGAACGCATTTTGG	20	TGGTAGAGGAAGTCTTATCTTTAATATGC	29
79G	GCATTGTTTTGCATCCTTTTG	21	ACTCAAGCCTGCCCCACT	18
79H	ACCACCACACCAAATGACTACT	22	TTCTTTCTATTAGGATGTGACATGAA	26
79I	GCAGGCTTGAGTTTTCATTTC	21	TCCGGGAAAAATCCATTCTA	20
79J	CAGGCTTACCTGCTTGGTC	19	TCTGATATAATAAGTCCTGTGTATTCATTC	30
79K	TTTCCACTGACAACGAAAGTAAA	23	CCTTCTGATTGATTTCCACTGA	22
79M	GCTAGCAATGCCACGATTTAG	21	GTGGCCTACTCCTTCACAGG	20
79N	CTCCCAAGCAGTAGCAGGAC	20	CAAAAACCAAACCACCCAGT	20
79O	TGCCCTCTTCTCACAGTCAA	20	GGCTTCTGGGTTGATACCTG	20
79P	TAAGACAGTAGCCCCATCACA	21	CTAATCCACCAAGAAGGGTTTT	22
79Q	TCCACACAGGTTTGTAAGTAAGTAAGA	27	AAAAGAAATAAAATGGCATGAAAGA	25
79R	GGAGAAAAGCTCAAGAGGAAAA	22	CACTTGATGTCAGCCCACTC	20
79S	TTGCAAATCTGTTACCTCTGACA	23	ACACCCCAAAACCAAAGTGA	20
79T	ACATCAAACACGGCTTCTCA	20	AGAGCTTTGGGTTTTCTTTTGA	22
79U	CACAGCTTCACCACTTGTCC	20	AATCCTTGGGTAAAGAAAAGGTC	23

79V	TTGGTTTTTGTCTTGCATTTTT	22	AATTATTTATGCACTCTATTTACCTCTGA	29
79W	AAAACAAGGGGTTACTTTACATCCT	25	TTAAAATGCAGCAATAAAGCTCTC	24
79X	GACACATTAGCTCTGGAGTGAGTC	24	TCAGTACCTGAAAACAATGACAAAA	25
79Y	ACCACCGAGTATTAAACTGTAAATCA	26	TGGTGAAGTCTGATATGTTGTGAA	24
79Z	CACCAACACTGTAACATTTACGAA	24	acaaagtactgcagtccagtcaa	23

Legend

Redesigned due to similarities to other primer sets; Tested

Redesigned twice due to similarities to other primer sets; Tested

Redesigned due to similarities to more than one other primer set; Untested

Untested

Tested but not fully optimised due to various issues

For Exon 79, all primers except 79A cover 3' UTR. UTR sequences are also given in capital letters

[000130]. The probes are listed in table 2.

Exon	Probe Sequence	Probe size (bp)	SNP
6P	GATTGCAACaAACCAACAGTGAAAAG	26	Rs1800256 (A/C)
8P1	CATTTTCAGcTACATCATCAAATG	24	Rs1800264 (C/T)
8P2	GCACTATTCTCAACAGgtaaag	22	
9P	ttctctgcagATCACgGTCAGTCTAGCACA	30	Rs1800265 (A/G)
10P	CATTGCAAGCACAcGGAGAGATTTC	25	Rs1800266 (A/C)
11P	TAAGCTGATTGGAaCAGGAAAATTATCAGA	30	rs34155804 (67.6% T/A)
17P1	cttctcacagATTTcACAGGCTGTCACCAC	30	rs34563188 (A/G)
17P2	CAAAAGAAGAGGCAGATTaCTGTGG	25	rs16998350 (A/66.7%T)
17P3	GGATTCTGAAATTAGGAAAAGgtgag	26	
19P1	GATGCCAGCAGATCAGcTCAGGCCCTGG	28	rs1800260 (C/G)
19P2	GGTGGAACAGATGGTGAATGgtaa	24	
21P1	TCAACCGGCTATCAGaTCTTCAAC	24	rs41312094 (A/G)
21P2	GATGTGCAGGtCAGAGAGAAAG	22	rs228406 (86.8% C/13.2% T)
21P3	GAAAGAGCTACAGACAAgtaagt	23	
23P1	GTGAAAGAGATGTCgAAGAAAGCGCC	26	rs1800268 (A/G)
23P2	TGGTTGAGCATTGtCAAAAGCTAGAGGAGC	30	rs1800261 (C/T)
23P3	CAAATGAATAAACTCCGAAAAATTCAGgta	30	
25P1	GAACTTAACtCTCAGTGGGATC	22	rs3827462 (94.1%T, 5.9%A)
25P2	CACATGTGCCAACAGgtatag	21	
26P	TACAGAAAGCAgTTGAAGAGATGAAGgta	29	rs1800262 (G/T)
27P	CTTGAAACTCTAACCAtCAACTACCAGTGG	30	rs1800269 (C/T)
28P	TGGAGAAAgCAAACAAGTGGCTAAATGAAG	30	rs1800270 (C/G)

29P	ATTTGATGcGACATTCAGAGGATA	24	rs5030730 (C/T)
30P1	AGGAGACTGAAAAtTCCTTACACTTAATC	29	rs1800263 (A/T)
30P2	CCAGGAGTCCCTCACAgTCATTGAC	25	rs28715870 (C/100%A)
32P	GCCAATTTTGAGCtGCGTCTACAAGAAAGT	30	rs1057872 (T/100%A)
34P	GAATTGACAAAGAgATCAGCAGTTGA	26	rs36072930 (98.7%C/1.3%T)
35P1	AGTAAcTGGATAGCTGTCACCTCCCGAGCA	30	rs1801185 (C/T)
35P2	AGAAGAGTGGTTAAAtCTTTTGTTGgtaag	30	rs16990264 (66.7%A/T)
37P1	GAATGACATAcGCCCAAAGGTGGACTCTACACGT GAC	37	rs34102501 (1.4%A/98.6%G*)
37P2	CAAGCAGCAAACTTGATGGCAAACCaCGGTGAC CACTGC	39	rs1801187 (76.1%A/23.9%G)
41P	GAAAGAGGAgCTGAATGCAGTGCG	24	rs1800272 (A/G)
42P	CTCTTTAAGCAAGAtGAGTCTCTGAAGgta	30	rs331312 (A/100%C)
44P	TGAGAAATGGCGGcGTTTTCATTATG	26	rs16990169 (A/G)
45P	CATTGGGCAGtGGCAAACTG	20	rs1800273 (T/100%C)
48P1	TAACCAACCAAACCAAGAgGGACCAT	26	rs1800274 (A/G)
48P2	TGACGTTcAGgtagggaacttttgc	26	rs1800275 (88.1%A/11.9%C)
53P	GCAACAGTTGAAcGAAATGTTAAAGGATTC	30	rs1801188 (35.4%C/64.6%T)
55P	CGAGAGGCTGCTTTGGAaGAAACTCAT	27	rs1801189 (A/G)
57P	CAGGAACTTgCTGGTGTGGCTAC	23	rs35527662 (-/C)
67P	GGAGCTGCTTCCAgTTTgtaag	22	rs12690372 (C/100%T)
75P	GCTAAGGCAGCTGcTGGAGCAAgtgagg	28	rs1800281 (C/T)
78P1	gatatctctgcctcttcctctctctat	27	
78P2	ttattaaaggGAAgAAATACCCCTG	25	rs1795743 (A/100%G)
78P3	GGAAAGCCAATGAGAGAGttagtg	24	
79P	GGGCAGAGCgATGGAGTCCTTAGTATC	27	rs16989352 (100%C/T)

Legend

Untested

[000131]. While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

[000132]. Each document, reference, patent application or patent cited in this text is expressly incorporated herein in their entirety by reference, which means that it should be read and considered by the reader as part of this text. That the

document, reference, patent application or patent cited in this text is not repeated in this text is merely for reasons of conciseness.

[000133]. Any manufacturer's instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention.

[000134]. The present invention is not to be limited in scope by any of the specific embodiments described herein. These embodiments are intended for the purpose of exemplification only. Functionally equivalent products, formulations and methods are clearly within the scope of the invention as described herein.

[000135]. The invention described herein may include one or more range of values (eg size, concentration etc). A range of values will be understood to include all values within the range, including the values defining the range, and values adjacent to the range which lead to the same or substantially the same outcome as the values immediately adjacent to that value which defines the boundary to the range.

[000136]. Throughout this specification, unless the context requires otherwise, the word "comprise" or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers. It is also noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[000137]. Other definitions for selected terms used herein may be found within the detailed description of the invention and apply throughout. Unless otherwise defined, all other scientific and technical terms used herein have the

same meaning as commonly understood to one of ordinary skill in the art to which the invention belongs.

SEQUENCE LISTING

<110> DSO National Laboratories
National University of Singapore
Lai, Poh-San
Yap, Eric
Yim, Onn Siong

<120> A CLINICAL METHOD FOR GENOTYPING LARGE GENES FOR MUTATIONS THAT POTENTIALLY CAUSE DISEASE

<130> 2009.4507

<160> 2

<170> PatentIn version 3.5

<210> 1
<211> 13993
<212> DNA
<213> Homo sapiens

<400> 1
tcctggcatc agttactgtg ttgactcact cagtgttggg atcactcact ttccccctac 60

aggactcaga tctgggaggc aattaccttc ggagaaaaac gaataggaaa aactgaagtg 120

ttacttttt taaagctgct gaagtttgtt ggtttctcat tgtttttaag cctactggag 180

caataaagtt tgaagaactt ttaccaggtt ttttttatcg ctgccttgat atacactttt 240

caaaatgctt tggtgggaag aagtagagga ctgttatgaa agagaagatg ttcaaaagaa 300

aacattcaca aaatgggtaa atgcacaatt ttctaagttt gggaagcagc atattgagaa 360

cctcttcagt gacctacagg atgggaggcg cctcctagac ctcctcgaag gcctgacagg 420

gcaaaaactg ccaaaagaaa aaggatccac aagagttcat gccctgaaca atgtcaacaa 480

ggcactgcgg gttttgcaga acaataatgt tgatttagtg aatattggaa gtactgacat 540

cgtagatgga aatcataaac tgactcttgg tttgatttgg aatataatcc tccactggca 600

ggtcaaaaat gtaatgaaaa atatcatggc tggattgcaa caaaccaaca gtgaaaagat 660

tctcctgagc tgggtccgac aatcaactcg taattatcca caggttaatg taatcaactt 720

caccaccagc tggtctgatg gcctggcttt gaatgctctc atccatagtc ataggccaga 780

cctatttgac tggaatagtg tggtttgcca gcagtcagcc acacaacgac tggaacatgc 840

attcaacatc gccagatatc aattaggcat agagaaacta ctcgatcctg aagatgttga 900

taccacctat ccagataaga agtccatctt aatgtacatc acatcactct tccaagtttt 960

gcctcaacaa gtgagcattg aagccatcca ggaagtggaa atgttgccaa ggccacctaa 1020

agtgactaaa gaagaacatt ttcagttaca tcatcaaatg cactattctc aacagatcac 1080

ggtcagtcta gcacagggat atgagagaac ttcttccccc aagcctcgat tcaagagcta 1140

tgcctacaca caggctgctt atgtcaccac ctctgaccct acacggagcc catttccttc 1200

acagcatttg gaagctcctg aagacaagtc atttggcagt tcattgatgg agagtgaagt 1260

aaacctggac cgttatcaaa cagctttaga agaagtatta tcgtggcttc tttctgctga 1320

ggacacattg caagcacaag gagagatttc taatgatgtg gaagtggtga aagaccagtt 1380

tcatactcat gaggggtaca tgatggattt gacagcccat cagggccggg ttggtaatat 1440
tctacaattg ggaagtaagc tgattggaac aggaaaatta tcagaagatg aagaaactga 1500
agtacaagag cagatgaatc tcctaaattc aagatgggaa tgcctcaggg tagctagcat 1560
ggaaaaacaa agcaatttac atagagtttt aatggatctc cagaatcaga aactgaaaga 1620
gttgaatgac tggctaacaa aaacagaaga aagaacaagg aaaatggagg aagagcctct 1680
tggacctgat cttgaagacc taaaacgcca agtacaacaa cataaggtgc ttcaagaaga 1740
tctagaacaa gaacaagtca gggtcaattc tctcactcac atggtggtgg tagttgatga 1800
atctagtgga gatcacgcaa ctgctgcttt ggaagaacaa cttaaggtat tgggagatcg 1860
atgggcaaac atctgtagat ggacagaaga ccgctgggtt cttttacaag acatccttct 1920
caaatggcaa cgtcttactg aagaacagtg ccttttttagt gcatggcttt cagaaaaaga 1980
agatgcagtg aacaagattc acacaactgg ctttaaagat caaaatgaaa tgttatcaag 2040
tcttcaaaaa ctggccgttt taaaagcgga tctagaaaag aaaaagcaat ccatgggcaa 2100
actgtattca ctcaaacaag atcttctttc aacactgaag aataagtcag tgacccagaa 2160
gacggaagca tggctggata actttgcccg gtgttgggat aatttagtcc aaaaacttga 2220
aaagagtaca gcacagattt cacaggctgt caccaccact cagccatcac taacacagac 2280
aactgtaatg gaaacagtaa ctacggtgac cacaagggaa cagatcctgg taaagcatgc 2340
tcaagaggaa cttccaccac cacctcccca aaagaagagg cagattactg tggattctga 2400
aattaggaaa aggttggatg ttgatataac tgaacttcac agctggatta ctcgctcaga 2460
agctgtgttg cagagtcctg aatttgcaat ctttcggaag gaaggcaact tctcagactt 2520
aaaagaaaaa gtcaatgcca tagagcgaga aaaagctgag aagttcagaa aactgcaaga 2580
tgccagcaga tcagctcagg ccctggtgga acagatggtg aatgagggtg ttaatgcaga 2640
tagcatcaaa caagcctcag aacaactgaa cagccggtgg atcgaattct gccagttgct 2700
aagtgagaga cttaactggc tggagtatca gaacaacatc atcgctttct ataatcagct 2760
acaacaattg gagcagatga caactactgc tgaaaactgg ttgaaaatcc aacccaccac 2820
cccatcagag ccaacagcaa ttaaaagtca gttaaaaatt tgtaaggatg aagtcaaccg 2880
gctatcaggt cttcaacctc aaattgaacg attaaaaatt caaagcatag ccctgaaaga 2940
gaaaggacaa ggacccatgt tcctggatgc agactttgtg gcctttacaa atcattttaa 3000
gcaagtcttt tctgatgtgc aggccagaga gaaagagcta cagacaattt ttgacacttt 3060
gccaccaatg cgctatcagg agaccatgag tgccatcagg acatgggtcc agcagtcaga 3120
aaccaaactc tccatacctc aacttagtgt caccgactat gaaatcatgg agcagagact 3180
cggggaattg caggctttac aaagttctct gcaagagcaa caaagtggcc tatactatct 3240
cagcaccact gtgaaagaga tgtcgaagaa agcgccctct gaaattagcc ggaaatatca 3300
atcagaattt gaagaaattg agggacgctg gaagaagctc tcctcccagc tggttgagca 3360
ttgtcaaaag ctagaggagc aaatgaataa actccgaaaa attcagaatc acatacaaac 3420

cctgaagaaa tggatggctg aagttgatgt ttttctgaag gaggaatggc ctgcccttgg 3480
ggattcagaa attctaaaaa agcagctgaa acagtgcaga cttttagtca gtgatattca 3540
gacaattcag cccagtctaa acagtgtcaa tgaaggtggg cagaagataa agaatgaagc 3600
agagccagag tttgcttcga gacttgagac agaactcaaa gaacttaaca ctcagtggga 3660
tcacatgtgc caacaggtct atgccagaaa ggaggccttg aagggaggtt tggagaaaac 3720
tgtaagcctc cagaaagatc tatcagagat gcacgaatgg atgacacaag ctgaagaaga 3780
gtatcttgag agagattttg aatataaaac tccagatgaa ttacagaaag cagttgaaga 3840
gatgaagaga gctaaagaag aggcccaaca aaaagaagcg aaagtgaaac tccttactga 3900
gtctgtaaat agtgtcatag ctcaagctcc acctgtagca caagaggcct taaaaaagga 3960
acttgaaact ctaaccacca actaccagtg gctctgcact aggctgaatg ggaaatgcaa 4020
gactttggaa gaagtttggg catgttggca tgagttattg tcatacttgg agaaagcaaa 4080
caagtggcta aatgaagtag aatttaaact taaaaccact gaaaacattc ctggcggagc 4140
tgaggaaatc tctgaggtgc tagattcact tgaaaatttg atgcgacatt cagaggataa 4200
cccaaatcag attcgcatat tggcacagac cctaacagat ggcggagtca tggatgagct 4260
aatcaatgag gaacttgaga catttaattc tcgttggagg gaactacatg aagaggctgt 4320
aaggaggcaa aagttgcttg aacagagcat ccagtctgcc caggagactg aaaaatcctt 4380
acacttaatc caggagtccc tcacattcat tgacaagcag ttggcagctt atattgcaga 4440
caaggtggac gcagctcaaa tgcctcagga agcccagaaa atccaatctg atttgacaag 4500
tcatgagatc agtttagaag aaatgaagaa acataatcag gggaaggagg ctgcccaaag 4560
agtcctgtct cagattgatg ttgcacagaa aaaattacaa gatgtctcca tgaagtttcg 4620
attattccag aaaccagcca attttgagca gcgtctacaa gaaagtaaga tgattttaga 4680
tgaagtgaag atgcacttgc ctgcattgga aacaaagagt gtggaacagg aagtagtaca 4740
gtcacagcta aatcattgtg tgaacttgta taaaagtctg agtgaagtga agtctgaagt 4800
ggaaatggtg ataaagactg gacgtcagat tgtacagaaa aagcagacgg aaaatcccaa 4860
agaacttgat gaaagagtaa cagctttgaa attgcattat aatgagctgg gagcaaaggt 4920
aacagaaaga aagcaacagt tggagaaatg cttgaaattg tcccgtaaga tgcgaaagga 4980
aatgaatgtc ttgacagaat ggctggcagc tacagatatg gaattgacaa agagatcagc 5040
agttgaagga atgcctagta atttggattc tgaagttgcc tggggaaagg ctactcaaaa 5100
agagattgag aaacagaagg tgcacctgaa gagtatcaca gaggtaggag aggccttgaa 5160
aacagttttg ggcaagaagg agacgttggt ggaagataaa ctcagtcttc tgaatagtaa 5220
ctggatagct gtcacctccc gagcagaaga gtggttaaat cttttgttgg aataccagaa 5280
acacatggaa acttttgacc agaatgtgga ccacatcaca aagtggatca ttcaggctga 5340
cacacttttg gatgaatcag agaaaaagaa accccagcaa aaagaagacg tgcttaagcg 5400
tttaaaggca gaactgaatg acatacgccc aaaggtggac tctacacgtg accaagcagc 5460

aaacttgatg gcaaaccgcg gtgaccactg caggaaatta gtagagcccc aaatctcaga 5520
gctcaaccat cgatttgcag ccatttcaca cagaattaag actggaaagg cctccattcc 5580
tttgaaggaa ttggagcagt ttaactcaga tatacaaaaa ttgcttgaac cactggaggc 5640
tgaaattcag caggggggtga atctgaaaga ggaagacttc aataaagata tgaatgaaga 5700
caatgagggt actgtaaaag aattgttgca aagaggagac aacttacaac aaagaatcac 5760
agatgagaga aagcgagagg aaataaagat aaaacagcag ctgttacaga caaaacataa 5820
tgctctcaag gatttgaggt ctcaaagaag aaaaaaggct ctagaaattt ctcatcagtg 5880
gtatcagtac aagaggcagg ctgatgatct cctgaaatgc ttggatgaca ttgaaaaaaa 5940
attagccagc ctacctgagc ccagagatga aaggaaaata aaggaaattg atcgggaatt 6000
gcagaagaag aaagaggagc tgaatgcagt gcgtaggcaa gctgagggct tgtctgagga 6060
tggggccgca atggcagtgg agccaactca gatccagctc agcaagcgct ggcgggaaat 6120
tgagagcaaa tttgctcagt ttcgaagact caactttgca caaattcaca ctgtccgtga 6180
agaaacgatg atggtgatga ctgaagacat gcctttggaa atttcttatg tgccttctac 6240
ttatttgact gaaatcactc atgtctcaca agccctatta gaagtggaac aacttctcaa 6300
tgctcctgac ctctgtgcta aggactttga agatctcttt aagcaagagg agtctctgaa 6360
gaatataaaa gatagtctac aacaaagctc aggtcggatt gacattattc atagcaagaa 6420
gacagcagca ttgcaaagtg caacgcctgt ggaaagggtg aagctacagg aagctctctc 6480
ccagcttgat ttccaatggg aaaaagttaa caaaatgtac aaggaccgac aagggcgatt 6540
tgacagatct gttgagaaat ggcggcgttt tcattatgat ataaagatat ttaatcagtg 6600
gctaacagaa gctgaacagt ttctcagaaa gacacaaatt cctgagaatt gggaacatgc 6660
taaatacaaa tggtatctta aggaactcca ggatggcatt gggcagcggc aaactgttgt 6720
cagaacattg aatgcaactg gggaagaaat aattcagcaa tcctcaaaaa cagatgccag 6780
tattctacag gaaaaattgg gaagcctgaa tctgcggtgg caggaggtct gcaaacagct 6840
gtcagacaga aaaaagaggc tagaagaaca aaagaatatc ttgtcagaat ttcaaagaga 6900
tttaaatgaa tttgttttat ggttggagga agcagataac attgctagta tcccacttga 6960
acctggaaaa gagcagcaac taaaagaaaa gcttgagcaa gtcaagttac tggtggaaga 7020
gttgcccctg cgccagggaa ttctcaaaca attaaatgaa actggaggac ccgtgcttgt 7080
aagtgctccc ataagcccag aagagcaaga taaacttgaa aataagctca agcagacaaa 7140
tctccagtgg ataaaggttt ccagagcttt acctgagaaa caaggagaaa ttgaagctca 7200
aataaaagac cttgggcagc ttgaaaaaaa gcttgaagac cttgaagagc agttaaatca 7260
tctgctgctg tggttatctc ctattaggaa tcagttggaa atttataacc aaccaaacca 7320
agaaggacca tttgacgttc aggaaactga aatagcagtt caagctaaac aaccggatgt 7380
ggaagagatt ttgtctaaag ggcagcattt gtacaaggaa aaaccagcca ctcagccagt 7440
gaagaggaag ttagaagatc tgagctctga gtggaaggcg gtaaaccgtt tacttcaaga 7500

gctgagggca aagcagcctg acctagctcc tggactgacc actattggag cctctcctac 7560
tcagactgtt actctggtga cacaacctgt ggttactaag gaaactgcca tctccaaact 7620
agaaatgcca tcttccttga tgttggaggt acctgctctg gcagatttca accgggcttg 7680
gacagaactt accgactggc tttctctgct tgatcaagtt ataaaatcac agagggtgat 7740
ggtgggtgac cttgaggata tcaacgagat gatcatcaag cagaaggcaa caatgcagga 7800
tttggaacag aggcgtcccc agttggaaga actcattacc gctgcccaaa atttgaaaaa 7860
caagaccagc aatcaagagg ctagaacaat cattacggat cgaattgaaa gaattcagaa 7920
tcagtgggat gaagtacaag aacaccttca gaaccggagg caacagttga atgaaatgtt 7980
aaaggattca acacaatggc tggaagctaa ggaagaagct gagcaggtct taggacaggc 8040
cagagccaag cttgagtcat ggaaggaggg tccctataca gtagatgcaa tccaaaagaa 8100
aatcacagaa accaagcagt tggccaaaga cctccgccag tggcagacaa atgtagatgt 8160
ggcaaatgac ttggccctga aacttctccg ggattattct gcagatgata ccagaaaagt 8220
ccacatgata acagagaata tcaatgcctc ttggagaagc attcataaaa gggtgagtga 8280
gcgagaggct gctttggaag aaactcatag attactgcaa cagttccccc tggacctgga 8340
aaagtttctt gcctggctta cagaagctga aacaactgcc aatgtcctac aggatgctac 8400
ccgtaaggaa aggctcctag aagactccaa gggagtaaaa gagctgatga aacaatggca 8460
agacctccaa ggtgaaattg aagctcacac agatgtttat cacaacctgg atgaaaacag 8520
ccaaaaaatc ctgagatccc tggaaggttc cgatgatgca gtcctgttac aaagacgttt 8580
ggataacatg aacttcaagt ggagtgaact tcggaaaaag tctctcaaca ttaggtccca 8640
tttggaagcc agttctgacc agtggaagcg tctgcacctt tctctgcagg aacttctggt 8700
gtggctacag ctgaaagatg atgaattaag ccggcaggca cctattggag gcgactttcc 8760
agcagttcag aagcagaacg atgtacatag ggccttcaag agggaattga aaactaaaga 8820
acctgtaatc atgagtactc ttgagactgt acgaatattt ctgacagagc agcctttgga 8880
aggactagag aaactctacc aggagcccag agagctgcct cctgaggaga gagcccagaa 8940
tgtcactcgg cttctacgaa agcaggctga ggaggtcaat actgagtggg aaaaattgaa 9000
cctgcactcc gctgactggc agagaaaaat agatgagacc cttgaaagac tccaggaact 9060
tcaagaggcc acggatgagc tggacctcaa gctgcgccaa gctgaggtga tcaagggatc 9120
ctggcagccc gtgggcgatc tcctcattga ctctctccaa gatcacctcg agaaagtcaa 9180
ggcacttcga ggagaaattg cgcctctgaa agagaacgtg agccacgtca atgaccttgc 9240
tcgccagctt accactttgg gcattcagct ctcaccgtat aacctcagca ctctggaaga 9300
cctgaacacc agatggaagc ttctgcaggt ggccgtcgag gaccgagtca ggcagctgca 9360
tgaagcccac agggactttg gtccagcatc tcagcacttt ctttccacgt ctgtccaggg 9420
tccctgggag agagccatct cgccaaacaa agtgccctac tatatcaacc acgagactca 9480
aacaacttgc tgggaccatc ccaaaatgac agagctctac cagtctttag ctgacctgaa 9540

taatgtcaga ttctcagctt ataggactgc catgaaactc cgaagactgc agaaggccct 9600
ttgcttggat ctcttgagcc tgtcagctgc atgtgatgcc ttggaccagc acaacctcaa 9660
gcaaaatgac cagcccatgg atatcctgca gattattaat tgtttgacca ctatttatga 9720
ccgcctggag caagagcaca acaatttggt caacgtccct ctctgcgtgg atatgtgtct 9780
gaactggctg ctgaatgttt atgatacggg acgaacaggg aggatccgtg tcctgtcttt 9840
taaaactggc atcatttccc tgtgtaaagc acatttggaa gacaagtaca gataccttt 9900
caagcaagtg gcaagttcaa caggattttg tgaccagcgc aggctgggcc tccttctgca 9960
tgattctatc caaattccaa gacagttggg tgaagttgca tcctttgggg gcagtaacat 10020
tgagccaagt gtccggagct gcttccaatt tgctaataat aagccagaga tcgaagcggc 10080
cctcttccta gactggatga gactggaacc ccagtccatg gtgtggctgc ccgtcctgca 10140
cagagtggct gctgcagaaa ctgccaagca tcaggccaaa tgtaacatct gcaaagagtg 10200
tccaatcatt ggattcaggt acaggagtct aaagcacttt aattatgaca tctgccaaag 10260
ctgctttttt tctggtcgag ttgcaaaagg ccataaaatg cactatccca tggtggaata 10320
ttgcactccg actacatcag gagaagatgt tcgagacttt gccaaggtac taaaaaacaa 10380
atttcgaacc aaaaggtatt ttgcgaagca tccccgaatg ggctacctgc cagtgcagac 10440
tgtcttagag ggggacaaca tggaaactcc cgttactctg atcaacttct ggccagtaga 10500
ttctgcgcct gcctcgtccc ctcagctttc acacgatgat actcattcac gcattgaaca 10560
ttatgctagc aggctagcag aaatggaaaa cagcaatgga tcttatctaa atgatagcat 10620
ctctcctaat gagagcatag atgatgaaca tttgttaatc cagcattact gccaaagttt 10680
gaaccaggac tccccctga gccagcctcg tagtcctgcc cagatcttga tttccttaga 10740
gagtgaggaa agaggggagc tagagagaat cctagcagat cttgaggaag aaaacaggaa 10800
tctgcaagca gaatatgacc gtctaaagca gcagcacgaa cataaaggcc tgtccccact 10860
gccgtcccct cctgaaatga tgcccacctc tccccagagt ccccgggatg ctgagctcat 10920
tgctgaggcc aagctactgc gtcaacacaa aggccgcctg gaagccagga tgcaaatcct 10980
ggaagaccac aataaacagc tggagtcaca gttacacagg ctaaggcagc tgctggagca 11040
accccaggca gaggccaaag tgaatggcac aacggtgtcc tctccttcta cctctctaca 11100
gaggtccgac agcagtcagc ctatgctgct ccgagtggtt ggcagtcaaa cttcggactc 11160
catgggtgag gaagatcttc tcagtcctcc ccaggacaca agcacagggt tagaggaggt 11220
gatggagcaa ctcaacaact ccttccctag ttcaagagga agaaataccc ctggaaagcc 11280
aatgagagag gacacaatgt aggaagtctt ttccacatgg cagatgattt gggcagagcg 11340
atggagtcct tagtatcagt catgacagat gaagaaggag cagaataaat gttttacaac 11400
tcctgattcc cgcatggttt ttataatatt catacaacaa agaggattag acagtaagag 11460
tttacaagaa ataaatctat atttttgtga agggtagtgg tattatactg tagatttcag 11520
tagtttctaa gtctgttatt gttttgttaa caatggcagg ttttacacgt ctatgcaatt 11580

gtacaaaaaa gttataagaa aactacatgt aaaatcttga tagctaaata acttgccatt 11640
tctttatatg gaacgcattt tgggttgttt aaaaatttat aacagttata aagaaagatt 11700
gtaaactaaa gtgtgcttta taaaaaaaag ttgtttataa aaacccctaa aaacaaaaca 11760
aacacacaca cacacacata cacacacaca cacaaaactt tgaggcagcg cattgttttg 11820
catccttttg gcgtgatatc catatgaaat tcatggcttt ttctttttt gcatattaaa 11880
gataagactt cctctaccac cacaccaaat gactactaca cactgctcat ttgagaactg 11940
tcagctgagt ggggcaggct tgagttttca tttcatatat ctatatgtct ataagtatat 12000
aaatactata gttatataga taaagagata cgaatttcta tagactgact ttttccattt 12060
tttaaatgtt catgtcacat cctaatagaa agaaattact tctagtcagt catccaggct 12120
tacctgcttg gtctagaatg gatttttccc ggagccggaa gccaggagga aactacacca 12180
cactaaaaca ttgtctacag ctccagatgt ttctcatttt aaacaacttt ccactgacaa 12240
cgaaagtaaa gtaaagtatt ggatttttt aaagggaaca tgtgaatgaa tacacaggac 12300
ttattatatc agagtgagta atcggttggt tggttgattg attgattgat tgatacattc 12360
agcttcctgc tgctagcaat gccacgattt agatttaatg atgcttcagt ggaaatcaat 12420
cagaaggtat tctgaccttg tgaacatcag aaggtatttt ttaactccca agcagtagca 12480
ggacgatgat agggctggag ggctatggat tcccagccca tccctgtgaa ggagtaggcc 12540
actctttaag tgaaggattg gatgattgtt cataatacat aaagttctct gtaattacaa 12600
ctaaattatt atgccctctt ctcacagtca aaaggaactg ggtggtttgg tttttgttgc 12660
ttttttagat ttattgtccc atgtgggatg agtttttaaa tgccacaaga cataatttaa 12720
aataaataaa ctttgggaaa aggtgtaaga cagtagcccc atcacatttg tgatactgac 12780
aggtatcaac ccagaagccc atgaactgtg tttccatcct ttgcatttct ctgcgagtag 12840
ttccacacag gtttgtaagt aagtaagaaa gaaggcaaat tgattcaaat gttacaaaaa 12900
aacccttctt ggtggattag acaggttaaa tatataaaca aacaaacaaa aattgctcaa 12960
aaaagaggag aaaagctcaa gaggaaaagc taaggactgg taggaaaaag ctttactctt 13020
tcatgccatt ttatttcttt ttgattttta aatcattcat tcaatagata ccaccgtgtg 13080
acctataatt ttgcaaatct gttacctctg acatcaagtg taattagctt ttggagagtg 13140
ggctgacatc aagtgtaatt agcttttgga gagtgggttt tgtccattat taataattaa 13200
ttaattaaca tcaaacacgg cttctcatgc tatttctacc tcactttggt tttggggtgt 13260
tcctgataat tgtgcacacc tgagttcaca gcttcaccac ttgtccattg cgttattttc 13320
tttttccttt ataattcttt ctttttcctt cataattttc aaaagaaaac ccaaagctct 13380
aaggtaacaa attaccaaat tacatgaaga tttggttttt gtcttgcatt tttttccttt 13440
atgtgacgct ggacctttтc tttacccaag gatttttaaa actcagattt aaaacaaggg 13500
gttactttac atcctactaa gaagtttaag taagtaagtt tcattctaaa atcagaggta 13560
aatagagtgc ataaataatt ttgttttaat ctttttgttt ttcttttaga cacattagct 13620

ctggagtgag tctgtcataa tatttgaaca aaaattgaga gctttattgc tgcattttaa 13680

gcataattaa tttggacatt atttcgtgtt gtgttcttta taaccaccga gtattaaact 13740

gtaaatcata atgtaactga agcataaaca tcacatggca tgttttgtca ttgttttcag 13800

gtactgagtt cttacttgag tatcataata tattgtgttt taacaccaac actgtaacat 13860

ttacgaatta ttttttaaa cttcagtttt actgcatttt cacaacatat cagacttcac 13920

caaatatatg ccttactatt gtattatagt actgctttac tgtgtatctc aataaagcac 13980

gcagttatgt tac 13993

<210> 2
<211> 3685
<212> PRT
<213> Homo sapiens

<400> 2

Met Leu Trp Trp Glu Glu Val Glu Asp Cys Tyr Glu Arg Glu Asp Val
1 5 10 15

Gln Lys Lys Thr Phe Thr Lys Trp Val Asn Ala Gln Phe Ser Lys Phe
20 25 30

Gly Lys Gln His Ile Glu Asn Leu Phe Ser Asp Leu Gln Asp Gly Arg
35 40 45

Arg Leu Leu Asp Leu Leu Glu Gly Leu Thr Gly Gln Lys Leu Pro Lys
50 55 60

Glu Lys Gly Ser Thr Arg Val His Ala Leu Asn Asn Val Asn Lys Ala
65 70 75 80

Leu Arg Val Leu Gln Asn Asn Asn Val Asp Leu Val Asn Ile Gly Ser
85 90 95

Thr Asp Ile Val Asp Gly Asn His Lys Leu Thr Leu Gly Leu Ile Trp
100 105 110

Asn Ile Ile Leu His Trp Gln Val Lys Asn Val Met Lys Asn Ile Met
115 120 125

Ala Gly Leu Gln Gln Thr Asn Ser Glu Lys Ile Leu Leu Ser Trp Val
130 135 140

Arg Gln Ser Thr Arg Asn Tyr Pro Gln Val Asn Val Ile Asn Phe Thr
145 150 155 160

Thr Ser Trp Ser Asp Gly Leu Ala Leu Asn Ala Leu Ile His Ser His
165 170 175

Arg Pro Asp Leu Phe Asp Trp Asn Ser Val Val Cys Gln Gln Ser Ala

```
            180                 185                 190

Thr Gln Arg Leu Glu His Ala Phe Asn Ile Ala Arg Tyr Gln Leu Gly
        195                 200                 205

Ile Glu Lys Leu Leu Asp Pro Glu Asp Val Asp Thr Thr Tyr Pro Asp
    210                 215                 220

Lys Lys Ser Ile Leu Met Tyr Ile Thr Ser Leu Phe Gln Val Leu Pro
225                 230                 235                 240

Gln Gln Val Ser Ile Glu Ala Ile Gln Glu Val Glu Met Leu Pro Arg
                245                 250                 255

Pro Pro Lys Val Thr Lys Glu Glu His Phe Gln Leu His His Gln Met
            260                 265                 270

His Tyr Ser Gln Gln Ile Thr Val Ser Leu Ala Gln Gly Tyr Glu Arg
        275                 280                 285

Thr Ser Ser Pro Lys Pro Arg Phe Lys Ser Tyr Ala Tyr Thr Gln Ala
    290                 295                 300

Ala Tyr Val Thr Thr Ser Asp Pro Thr Arg Ser Pro Phe Pro Ser Gln
305                 310                 315                 320

His Leu Glu Ala Pro Glu Asp Lys Ser Phe Gly Ser Ser Leu Met Glu
                325                 330                 335

Ser Glu Val Asn Leu Asp Arg Tyr Gln Thr Ala Leu Glu Glu Val Leu
            340                 345                 350

Ser Trp Leu Leu Ser Ala Glu Asp Thr Leu Gln Ala Gln Gly Glu Ile
        355                 360                 365

Ser Asn Asp Val Glu Val Val Lys Asp Gln Phe His Thr His Glu Gly
    370                 375                 380

Tyr Met Met Asp Leu Thr Ala His Gln Gly Arg Val Gly Asn Ile Leu
385                 390                 395                 400

Gln Leu Gly Ser Lys Leu Ile Gly Thr Gly Lys Leu Ser Glu Asp Glu
                405                 410                 415

Glu Thr Glu Val Gln Glu Gln Met Asn Leu Leu Asn Ser Arg Trp Glu
            420                 425                 430

Cys Leu Arg Val Ala Ser Met Glu Lys Gln Ser Asn Leu His Arg Val
        435                 440                 445

Leu Met Asp Leu Gln Asn Gln Lys Leu Lys Glu Leu Asn Asp Trp Leu
```

```
     450                 455                 460

Thr Lys Thr Glu Glu Arg Thr Arg Lys Met Glu Glu Glu Pro Leu Gly
465                 470                 475                 480

Pro Asp Leu Glu Asp Leu Lys Arg Gln Val Gln Gln His Lys Val Leu
                485                 490                 495

Gln Glu Asp Leu Glu Gln Glu Gln Val Arg Val Asn Ser Leu Thr His
            500                 505                 510

Met Val Val Val Val Asp Glu Ser Ser Gly Asp His Ala Thr Ala Ala
        515                 520                 525

Leu Glu Glu Gln Leu Lys Val Leu Gly Asp Arg Trp Ala Asn Ile Cys
    530                 535                 540

Arg Trp Thr Glu Asp Arg Trp Val Leu Leu Gln Asp Ile Leu Leu Lys
545                 550                 555                 560

Trp Gln Arg Leu Thr Glu Glu Gln Cys Leu Phe Ser Ala Trp Leu Ser
                565                 570                 575

Glu Lys Glu Asp Ala Val Asn Lys Ile His Thr Thr Gly Phe Lys Asp
            580                 585                 590

Gln Asn Glu Met Leu Ser Ser Leu Gln Lys Leu Ala Val Leu Lys Ala
        595                 600                 605

Asp Leu Glu Lys Lys Lys Gln Ser Met Gly Lys Leu Tyr Ser Leu Lys
    610                 615                 620

Gln Asp Leu Leu Ser Thr Leu Lys Asn Lys Ser Val Thr Gln Lys Thr
625                 630                 635                 640

Glu Ala Trp Leu Asp Asn Phe Ala Arg Cys Trp Asp Asn Leu Val Gln
                645                 650                 655

Lys Leu Glu Lys Ser Thr Ala Gln Ile Ser Gln Ala Val Thr Thr Thr
            660                 665                 670

Gln Pro Ser Leu Thr Gln Thr Thr Val Met Glu Thr Val Thr Thr Val
        675                 680                 685

Thr Thr Arg Glu Gln Ile Leu Val Lys His Ala Gln Glu Glu Leu Pro
    690                 695                 700

Pro Pro Pro Pro Gln Lys Lys Arg Gln Ile Thr Val Asp Ser Glu Ile
705                 710                 715                 720

Arg Lys Arg Leu Asp Val Asp Ile Thr Glu Leu His Ser Trp Ile Thr
```

```
                725                 730                 735

Arg Ser Glu Ala Val Leu Gln Ser Pro Glu Phe Ala Ile Phe Arg Lys
            740                 745                 750

Glu Gly Asn Phe Ser Asp Leu Lys Glu Lys Val Asn Ala Ile Glu Arg
        755                 760                 765

Glu Lys Ala Glu Lys Phe Arg Lys Leu Gln Asp Ala Ser Arg Ser Ala
    770                 775                 780

Gln Ala Leu Val Glu Gln Met Val Asn Glu Gly Val Asn Ala Asp Ser
785                 790                 795                 800

Ile Lys Gln Ala Ser Glu Gln Leu Asn Ser Arg Trp Ile Glu Phe Cys
                805                 810                 815

Gln Leu Leu Ser Glu Arg Leu Asn Trp Leu Glu Tyr Gln Asn Asn Ile
            820                 825                 830

Ile Ala Phe Tyr Asn Gln Leu Gln Gln Leu Glu Gln Met Thr Thr Thr
        835                 840                 845

Ala Glu Asn Trp Leu Lys Ile Gln Pro Thr Thr Pro Ser Glu Pro Thr
    850                 855                 860

Ala Ile Lys Ser Gln Leu Lys Ile Cys Lys Asp Glu Val Asn Arg Leu
865                 870                 875                 880

Ser Gly Leu Gln Pro Gln Ile Glu Arg Leu Lys Ile Gln Ser Ile Ala
                885                 890                 895

Leu Lys Glu Lys Gly Gln Gly Pro Met Phe Leu Asp Ala Asp Phe Val
            900                 905                 910

Ala Phe Thr Asn His Phe Lys Gln Val Phe Ser Asp Val Gln Ala Arg
        915                 920                 925

Glu Lys Glu Leu Gln Thr Ile Phe Asp Thr Leu Pro Pro Met Arg Tyr
    930                 935                 940

Gln Glu Thr Met Ser Ala Ile Arg Thr Trp Val Gln Gln Ser Glu Thr
945                 950                 955                 960

Lys Leu Ser Ile Pro Gln Leu Ser Val Thr Asp Tyr Glu Ile Met Glu
                965                 970                 975

Gln Arg Leu Gly Glu Leu Gln Ala Leu Gln Ser Ser Leu Gln Glu Gln
            980                 985                 990

Gln Ser Gly Leu Tyr Tyr Leu Ser  Thr Thr Val Lys Glu  Met Ser Lys
```

995 1000 1005

Lys Ala Pro Ser Glu Ile Ser Arg Lys Tyr Gln Ser Glu Phe Glu
1010 1015 1020

Glu Ile Glu Gly Arg Trp Lys Lys Leu Ser Ser Gln Leu Val Glu
1025 1030 1035

His Cys Gln Lys Leu Glu Glu Gln Met Asn Lys Leu Arg Lys Ile
1040 1045 1050

Gln Asn His Ile Gln Thr Leu Lys Lys Trp Met Ala Glu Val Asp
1055 1060 1065

Val Phe Leu Lys Glu Glu Trp Pro Ala Leu Gly Asp Ser Glu Ile
1070 1075 1080

Leu Lys Lys Gln Leu Lys Gln Cys Arg Leu Leu Val Ser Asp Ile
1085 1090 1095

Gln Thr Ile Gln Pro Ser Leu Asn Ser Val Asn Glu Gly Gly Gln
1100 1105 1110

Lys Ile Lys Asn Glu Ala Glu Pro Glu Phe Ala Ser Arg Leu Glu
1115 1120 1125

Thr Glu Leu Lys Glu Leu Asn Thr Gln Trp Asp His Met Cys Gln
1130 1135 1140

Gln Val Tyr Ala Arg Lys Glu Ala Leu Lys Gly Gly Leu Glu Lys
1145 1150 1155

Thr Val Ser Leu Gln Lys Asp Leu Ser Glu Met His Glu Trp Met
1160 1165 1170

Thr Gln Ala Glu Glu Glu Tyr Leu Glu Arg Asp Phe Glu Tyr Lys
1175 1180 1185

Thr Pro Asp Glu Leu Gln Lys Ala Val Glu Glu Met Lys Arg Ala
1190 1195 1200

Lys Glu Glu Ala Gln Gln Lys Glu Ala Lys Val Lys Leu Leu Thr
1205 1210 1215

Glu Ser Val Asn Ser Val Ile Ala Gln Ala Pro Pro Val Ala Gln
1220 1225 1230

Glu Ala Leu Lys Lys Glu Leu Glu Thr Leu Thr Thr Asn Tyr Gln
1235 1240 1245

Trp Leu Cys Thr Arg Leu Asn Gly Lys Cys Lys Thr Leu Glu Glu

1250 1255 1260

Val Trp Ala Cys Trp His Glu Leu Leu Ser Tyr Leu Glu Lys Ala
1265 1270 1275

Asn Lys Trp Leu Asn Glu Val Glu Phe Lys Leu Lys Thr Thr Glu
1280 1285 1290

Asn Ile Pro Gly Gly Ala Glu Glu Ile Ser Glu Val Leu Asp Ser
1295 1300 1305

Leu Glu Asn Leu Met Arg His Ser Glu Asp Asn Pro Asn Gln Ile
1310 1315 1320

Arg Ile Leu Ala Gln Thr Leu Thr Asp Gly Gly Val Met Asp Glu
1325 1330 1335

Leu Ile Asn Glu Glu Leu Glu Thr Phe Asn Ser Arg Trp Arg Glu
1340 1345 1350

Leu His Glu Glu Ala Val Arg Arg Gln Lys Leu Leu Glu Gln Ser
1355 1360 1365

Ile Gln Ser Ala Gln Glu Thr Glu Lys Ser Leu His Leu Ile Gln
1370 1375 1380

Glu Ser Leu Thr Phe Ile Asp Lys Gln Leu Ala Ala Tyr Ile Ala
1385 1390 1395

Asp Lys Val Asp Ala Ala Gln Met Pro Gln Glu Ala Gln Lys Ile
1400 1405 1410

Gln Ser Asp Leu Thr Ser His Glu Ile Ser Leu Glu Glu Met Lys
1415 1420 1425

Lys His Asn Gln Gly Lys Glu Ala Ala Gln Arg Val Leu Ser Gln
1430 1435 1440

Ile Asp Val Ala Gln Lys Lys Leu Gln Asp Val Ser Met Lys Phe
1445 1450 1455

Arg Leu Phe Gln Lys Pro Ala Asn Phe Glu Gln Arg Leu Gln Glu
1460 1465 1470

Ser Lys Met Ile Leu Asp Glu Val Lys Met His Leu Pro Ala Leu
1475 1480 1485

Glu Thr Lys Ser Val Glu Gln Glu Val Val Gln Ser Gln Leu Asn
1490 1495 1500

His Cys Val Asn Leu Tyr Lys Ser Leu Ser Glu Val Lys Ser Glu

1505 1510 1515

Val Glu Met Val Ile Lys Thr Gly Arg Gln Ile Val Gln Lys Lys
1520 1525 1530

Gln Thr Glu Asn Pro Lys Glu Leu Asp Glu Arg Val Thr Ala Leu
1535 1540 1545

Lys Leu His Tyr Asn Glu Leu Gly Ala Lys Val Thr Glu Arg Lys
1550 1555 1560

Gln Gln Leu Glu Lys Cys Leu Lys Leu Ser Arg Lys Met Arg Lys
1565 1570 1575

Glu Met Asn Val Leu Thr Glu Trp Leu Ala Ala Thr Asp Met Glu
1580 1585 1590

Leu Thr Lys Arg Ser Ala Val Glu Gly Met Pro Ser Asn Leu Asp
1595 1600 1605

Ser Glu Val Ala Trp Gly Lys Ala Thr Gln Lys Glu Ile Glu Lys
1610 1615 1620

Gln Lys Val His Leu Lys Ser Ile Thr Glu Val Gly Glu Ala Leu
1625 1630 1635

Lys Thr Val Leu Gly Lys Lys Glu Thr Leu Val Glu Asp Lys Leu
1640 1645 1650

Ser Leu Leu Asn Ser Asn Trp Ile Ala Val Thr Ser Arg Ala Glu
1655 1660 1665

Glu Trp Leu Asn Leu Leu Leu Glu Tyr Gln Lys His Met Glu Thr
1670 1675 1680

Phe Asp Gln Asn Val Asp His Ile Thr Lys Trp Ile Ile Gln Ala
1685 1690 1695

Asp Thr Leu Leu Asp Glu Ser Glu Lys Lys Lys Pro Gln Gln Lys
1700 1705 1710

Glu Asp Val Leu Lys Arg Leu Lys Ala Glu Leu Asn Asp Ile Arg
1715 1720 1725

Pro Lys Val Asp Ser Thr Arg Asp Gln Ala Ala Asn Leu Met Ala
1730 1735 1740

Asn Arg Gly Asp His Cys Arg Lys Leu Val Glu Pro Gln Ile Ser
1745 1750 1755

Glu Leu Asn His Arg Phe Ala Ala Ile Ser His Arg Ile Lys Thr

1760 1765 1770

Gly Lys Ala Ser Ile Pro Leu Lys Glu Leu Glu Gln Phe Asn Ser
1775 1780 1785

Asp Ile Gln Lys Leu Leu Glu Pro Leu Glu Ala Glu Ile Gln Gln
1790 1795 1800

Gly Val Asn Leu Lys Glu Glu Asp Phe Asn Lys Asp Met Asn Glu
1805 1810 1815

Asp Asn Glu Gly Thr Val Lys Glu Leu Leu Gln Arg Gly Asp Asn
1820 1825 1830

Leu Gln Gln Arg Ile Thr Asp Glu Arg Lys Arg Glu Glu Ile Lys
1835 1840 1845

Ile Lys Gln Gln Leu Leu Gln Thr Lys His Asn Ala Leu Lys Asp
1850 1855 1860

Leu Arg Ser Gln Arg Arg Lys Lys Ala Leu Glu Ile Ser His Gln
1865 1870 1875

Trp Tyr Gln Tyr Lys Arg Gln Ala Asp Asp Leu Leu Lys Cys Leu
1880 1885 1890

Asp Asp Ile Glu Lys Lys Leu Ala Ser Leu Pro Glu Pro Arg Asp
1895 1900 1905

Glu Arg Lys Ile Lys Glu Ile Asp Arg Glu Leu Gln Lys Lys Lys
1910 1915 1920

Glu Glu Leu Asn Ala Val Arg Arg Gln Ala Glu Gly Leu Ser Glu
1925 1930 1935

Asp Gly Ala Ala Met Ala Val Glu Pro Thr Gln Ile Gln Leu Ser
1940 1945 1950

Lys Arg Trp Arg Glu Ile Glu Ser Lys Phe Ala Gln Phe Arg Arg
1955 1960 1965

Leu Asn Phe Ala Gln Ile His Thr Val Arg Glu Glu Thr Met Met
1970 1975 1980

Val Met Thr Glu Asp Met Pro Leu Glu Ile Ser Tyr Val Pro Ser
1985 1990 1995

Thr Tyr Leu Thr Glu Ile Thr His Val Ser Gln Ala Leu Leu Glu
2000 2005 2010

Val Glu Gln Leu Leu Asn Ala Pro Asp Leu Cys Ala Lys Asp Phe

2015 2020 2025

Glu Asp Leu Phe Lys Gln Glu Glu Ser Leu Lys Asn Ile Lys Asp
2030 2035 2040

Ser Leu Gln Gln Ser Ser Gly Arg Ile Asp Ile Ile His Ser Lys
2045 2050 2055

Lys Thr Ala Ala Leu Gln Ser Ala Thr Pro Val Glu Arg Val Lys
2060 2065 2070

Leu Gln Glu Ala Leu Ser Gln Leu Asp Phe Gln Trp Glu Lys Val
2075 2080 2085

Asn Lys Met Tyr Lys Asp Arg Gln Gly Arg Phe Asp Arg Ser Val
2090 2095 2100

Glu Lys Trp Arg Arg Phe His Tyr Asp Ile Lys Ile Phe Asn Gln
2105 2110 2115

Trp Leu Thr Glu Ala Glu Gln Phe Leu Arg Lys Thr Gln Ile Pro
2120 2125 2130

Glu Asn Trp Glu His Ala Lys Tyr Lys Trp Tyr Leu Lys Glu Leu
2135 2140 2145

Gln Asp Gly Ile Gly Gln Arg Gln Thr Val Val Arg Thr Leu Asn
2150 2155 2160

Ala Thr Gly Glu Glu Ile Ile Gln Gln Ser Ser Lys Thr Asp Ala
2165 2170 2175

Ser Ile Leu Gln Glu Lys Leu Gly Ser Leu Asn Leu Arg Trp Gln
2180 2185 2190

Glu Val Cys Lys Gln Leu Ser Asp Arg Lys Lys Arg Leu Glu Glu
2195 2200 2205

Gln Lys Asn Ile Leu Ser Glu Phe Gln Arg Asp Leu Asn Glu Phe
2210 2215 2220

Val Leu Trp Leu Glu Glu Ala Asp Asn Ile Ala Ser Ile Pro Leu
2225 2230 2235

Glu Pro Gly Lys Glu Gln Gln Leu Lys Glu Lys Leu Glu Gln Val
2240 2245 2250

Lys Leu Leu Val Glu Glu Leu Pro Leu Arg Gln Gly Ile Leu Lys
2255 2260 2265

Gln Leu Asn Glu Thr Gly Gly Pro Val Leu Val Ser Ala Pro Ile

2270 2275 2280

Ser Pro Glu Glu Gln Asp Lys Leu Glu Asn Lys Leu Lys Gln Thr
2285 2290 2295

Asn Leu Gln Trp Ile Lys Val Ser Arg Ala Leu Pro Glu Lys Gln
2300 2305 2310

Gly Glu Ile Glu Ala Gln Ile Lys Asp Leu Gly Gln Leu Glu Lys
2315 2320 2325

Lys Leu Glu Asp Leu Glu Glu Gln Leu Asn His Leu Leu Leu Trp
2330 2335 2340

Leu Ser Pro Ile Arg Asn Gln Leu Glu Ile Tyr Asn Gln Pro Asn
2345 2350 2355

Gln Glu Gly Pro Phe Asp Val Gln Glu Thr Glu Ile Ala Val Gln
2360 2365 2370

Ala Lys Gln Pro Asp Val Glu Glu Ile Leu Ser Lys Gly Gln His
2375 2380 2385

Leu Tyr Lys Glu Lys Pro Ala Thr Gln Pro Val Lys Arg Lys Leu
2390 2395 2400

Glu Asp Leu Ser Ser Glu Trp Lys Ala Val Asn Arg Leu Leu Gln
2405 2410 2415

Glu Leu Arg Ala Lys Gln Pro Asp Leu Ala Pro Gly Leu Thr Thr
2420 2425 2430

Ile Gly Ala Ser Pro Thr Gln Thr Val Thr Leu Val Thr Gln Pro
2435 2440 2445

Val Val Thr Lys Glu Thr Ala Ile Ser Lys Leu Glu Met Pro Ser
2450 2455 2460

Ser Leu Met Leu Glu Val Pro Ala Leu Ala Asp Phe Asn Arg Ala
2465 2470 2475

Trp Thr Glu Leu Thr Asp Trp Leu Ser Leu Leu Asp Gln Val Ile
2480 2485 2490

Lys Ser Gln Arg Val Met Val Gly Asp Leu Glu Asp Ile Asn Glu
2495 2500 2505

Met Ile Ile Lys Gln Lys Ala Thr Met Gln Asp Leu Glu Gln Arg
2510 2515 2520

Arg Pro Gln Leu Glu Glu Leu Ile Thr Ala Ala Gln Asn Leu Lys

2525 2530 2535

Asn Lys Thr Ser Asn Gln Glu Ala Arg Thr Ile Ile Thr Asp Arg
2540 2545 2550

Ile Glu Arg Ile Gln Asn Gln Trp Asp Glu Val Gln Glu His Leu
2555 2560 2565

Gln Asn Arg Arg Gln Gln Leu Asn Glu Met Leu Lys Asp Ser Thr
2570 2575 2580

Gln Trp Leu Glu Ala Lys Glu Glu Ala Glu Gln Val Leu Gly Gln
2585 2590 2595

Ala Arg Ala Lys Leu Glu Ser Trp Lys Glu Gly Pro Tyr Thr Val
2600 2605 2610

Asp Ala Ile Gln Lys Lys Ile Thr Glu Thr Lys Gln Leu Ala Lys
2615 2620 2625

Asp Leu Arg Gln Trp Gln Thr Asn Val Asp Val Ala Asn Asp Leu
2630 2635 2640

Ala Leu Lys Leu Leu Arg Asp Tyr Ser Ala Asp Asp Thr Arg Lys
2645 2650 2655

Val His Met Ile Thr Glu Asn Ile Asn Ala Ser Trp Arg Ser Ile
2660 2665 2670

His Lys Arg Val Ser Glu Arg Glu Ala Ala Leu Glu Glu Thr His
2675 2680 2685

Arg Leu Leu Gln Gln Phe Pro Leu Asp Leu Glu Lys Phe Leu Ala
2690 2695 2700

Trp Leu Thr Glu Ala Glu Thr Thr Ala Asn Val Leu Gln Asp Ala
2705 2710 2715

Thr Arg Lys Glu Arg Leu Leu Glu Asp Ser Lys Gly Val Lys Glu
2720 2725 2730

Leu Met Lys Gln Trp Gln Asp Leu Gln Gly Glu Ile Glu Ala His
2735 2740 2745

Thr Asp Val Tyr His Asn Leu Asp Glu Asn Ser Gln Lys Ile Leu
2750 2755 2760

Arg Ser Leu Glu Gly Ser Asp Asp Ala Val Leu Leu Gln Arg Arg
2765 2770 2775

Leu Asp Asn Met Asn Phe Lys Trp Ser Glu Leu Arg Lys Lys Ser

2780 2785 2790

Leu Asn Ile Arg Ser His Leu Glu Ala Ser Ser Asp Gln Trp Lys
2795 2800 2805

Arg Leu His Leu Ser Leu Gln Glu Leu Leu Val Trp Leu Gln Leu
2810 2815 2820

Lys Asp Asp Glu Leu Ser Arg Gln Ala Pro Ile Gly Gly Asp Phe
2825 2830 2835

Pro Ala Val Gln Lys Gln Asn Asp Val His Arg Ala Phe Lys Arg
2840 2845 2850

Glu Leu Lys Thr Lys Glu Pro Val Ile Met Ser Thr Leu Glu Thr
2855 2860 2865

Val Arg Ile Phe Leu Thr Glu Gln Pro Leu Glu Gly Leu Glu Lys
2870 2875 2880

Leu Tyr Gln Glu Pro Arg Glu Leu Pro Pro Glu Glu Arg Ala Gln
2885 2890 2895

Asn Val Thr Arg Leu Leu Arg Lys Gln Ala Glu Glu Val Asn Thr
2900 2905 2910

Glu Trp Glu Lys Leu Asn Leu His Ser Ala Asp Trp Gln Arg Lys
2915 2920 2925

Ile Asp Glu Thr Leu Glu Arg Leu Gln Glu Leu Gln Glu Ala Thr
2930 2935 2940

Asp Glu Leu Asp Leu Lys Leu Arg Gln Ala Glu Val Ile Lys Gly
2945 2950 2955

Ser Trp Gln Pro Val Gly Asp Leu Leu Ile Asp Ser Leu Gln Asp
2960 2965 2970

His Leu Glu Lys Val Lys Ala Leu Arg Gly Glu Ile Ala Pro Leu
2975 2980 2985

Lys Glu Asn Val Ser His Val Asn Asp Leu Ala Arg Gln Leu Thr
2990 2995 3000

Thr Leu Gly Ile Gln Leu Ser Pro Tyr Asn Leu Ser Thr Leu Glu
3005 3010 3015

Asp Leu Asn Thr Arg Trp Lys Leu Leu Gln Val Ala Val Glu Asp
3020 3025 3030

Arg Val Arg Gln Leu His Glu Ala His Arg Asp Phe Gly Pro Ala

3035 3040 3045

Ser Gln His Phe Leu Ser Thr Ser Val Gln Gly Pro Trp Glu Arg
3050 3055 3060

Ala Ile Ser Pro Asn Lys Val Pro Tyr Tyr Ile Asn His Glu Thr
3065 3070 3075

Gln Thr Thr Cys Trp Asp His Pro Lys Met Thr Glu Leu Tyr Gln
3080 3085 3090

Ser Leu Ala Asp Leu Asn Asn Val Arg Phe Ser Ala Tyr Arg Thr
3095 3100 3105

Ala Met Lys Leu Arg Arg Leu Gln Lys Ala Leu Cys Leu Asp Leu
3110 3115 3120

Leu Ser Leu Ser Ala Ala Cys Asp Ala Leu Asp Gln His Asn Leu
3125 3130 3135

Lys Gln Asn Asp Gln Pro Met Asp Ile Leu Gln Ile Ile Asn Cys
3140 3145 3150

Leu Thr Thr Ile Tyr Asp Arg Leu Glu Gln Glu His Asn Asn Leu
3155 3160 3165

Val Asn Val Pro Leu Cys Val Asp Met Cys Leu Asn Trp Leu Leu
3170 3175 3180

Asn Val Tyr Asp Thr Gly Arg Thr Gly Arg Ile Arg Val Leu Ser
3185 3190 3195

Phe Lys Thr Gly Ile Ile Ser Leu Cys Lys Ala His Leu Glu Asp
3200 3205 3210

Lys Tyr Arg Tyr Leu Phe Lys Gln Val Ala Ser Ser Thr Gly Phe
3215 3220 3225

Cys Asp Gln Arg Arg Leu Gly Leu Leu Leu His Asp Ser Ile Gln
3230 3235 3240

Ile Pro Arg Gln Leu Gly Glu Val Ala Ser Phe Gly Gly Ser Asn
3245 3250 3255

Ile Glu Pro Ser Val Arg Ser Cys Phe Gln Phe Ala Asn Asn Lys
3260 3265 3270

Pro Glu Ile Glu Ala Ala Leu Phe Leu Asp Trp Met Arg Leu Glu
3275 3280 3285

Pro Gln Ser Met Val Trp Leu Pro Val Leu His Arg Val Ala Ala

3290 3295 3300

Ala Glu Thr Ala Lys His Gln Ala Lys Cys Asn Ile Cys Lys Glu
3305 3310 3315

Cys Pro Ile Ile Gly Phe Arg Tyr Arg Ser Leu Lys His Phe Asn
3320 3325 3330

Tyr Asp Ile Cys Gln Ser Cys Phe Phe Ser Gly Arg Val Ala Lys
3335 3340 3345

Gly His Lys Met His Tyr Pro Met Val Glu Tyr Cys Thr Pro Thr
3350 3355 3360

Thr Ser Gly Glu Asp Val Arg Asp Phe Ala Lys Val Leu Lys Asn
3365 3370 3375

Lys Phe Arg Thr Lys Arg Tyr Phe Ala Lys His Pro Arg Met Gly
3380 3385 3390

Tyr Leu Pro Val Gln Thr Val Leu Glu Gly Asp Asn Met Glu Thr
3395 3400 3405

Pro Val Thr Leu Ile Asn Phe Trp Pro Val Asp Ser Ala Pro Ala
3410 3415 3420

Ser Ser Pro Gln Leu Ser His Asp Asp Thr His Ser Arg Ile Glu
3425 3430 3435

His Tyr Ala Ser Arg Leu Ala Glu Met Glu Asn Ser Asn Gly Ser
3440 3445 3450

Tyr Leu Asn Asp Ser Ile Ser Pro Asn Glu Ser Ile Asp Asp Glu
3455 3460 3465

His Leu Leu Ile Gln His Tyr Cys Gln Ser Leu Asn Gln Asp Ser
3470 3475 3480

Pro Leu Ser Gln Pro Arg Ser Pro Ala Gln Ile Leu Ile Ser Leu
3485 3490 3495

Glu Ser Glu Glu Arg Gly Glu Leu Glu Arg Ile Leu Ala Asp Leu
3500 3505 3510

Glu Glu Glu Asn Arg Asn Leu Gln Ala Glu Tyr Asp Arg Leu Lys
3515 3520 3525

Gln Gln His Glu His Lys Gly Leu Ser Pro Leu Pro Ser Pro Pro
3530 3535 3540

Glu Met Met Pro Thr Ser Pro Gln Ser Pro Arg Asp Ala Glu Leu

3545 3550 3555

Ile Ala Glu Ala Lys Leu Leu Arg Gln His Lys Gly Arg Leu Glu
3560 3565 3570

Ala Arg Met Gln Ile Leu Glu Asp His Asn Lys Gln Leu Glu Ser
3575 3580 3585

Gln Leu His Arg Leu Arg Gln Leu Leu Glu Gln Pro Gln Ala Glu
3590 3595 3600

Ala Lys Val Asn Gly Thr Thr Val Ser Ser Pro Ser Thr Ser Leu
3605 3610 3615

Gln Arg Ser Asp Ser Ser Gln Pro Met Leu Leu Arg Val Val Gly
3620 3625 3630

Ser Gln Thr Ser Asp Ser Met Gly Glu Glu Asp Leu Leu Ser Pro
3635 3640 3645

Pro Gln Asp Thr Ser Thr Gly Leu Glu Glu Val Met Glu Gln Leu
3650 3655 3660

Asn Asn Ser Phe Pro Ser Ser Arg Gly Arg Asn Thr Pro Gly Lys
3665 3670 3675

Pro Met Arg Glu Asp Thr Met
3680 3685

Claims:

1. A method of determining sequence variants within a large gene comprising the steps of:

(a) making a Whole-Genome Amplification (WGA) to obtain sufficient amounts of genetic templates for DNA analysis;

(b) enriching the WGA sample with nested primers designed for the large gene;

(c) using the enriched WGA sample for high resolution melt (HRM); and

(d) detecting differential melt profiles during the transition from double strand to single strand with an increase in temperature wherein sequence point mutations within the gene affects the thermal stability and gives a different melt profile from the normal non-mutated gene sequence.

2. The method of claim 1 wherein the large gene is the DMD gene that may potentially contribute to muscle disease.

3. The method of claim 2 wherein the muscle disease comprises muscular dystrophy.

4. The method of claim 3 wherein the muscular dystrophy comprises Duchenne muscular dystrophy.

5. The method of claim 3 wherein the muscular dystrophy comprises Becker muscular dystrophy.

6. The method of claim 1 further comprising a cleaning step prior to the HRM step.

7. The method of claim 6 wherein the cleaning step comprises washing with a high magnesium concentration.

8. The method of claim 6 wherein the cleaning step comprises exo-nuclease digestion, dephosphorylation treatment or filtration.

9. A kit comprising reagents for whole genome amplification, at least two nested primers specific to the gene of interest and reagents' for high resolution melt analysis.

10. The kit of claim 7 to rapidly detect mutations that may potentially contribute to muscle disease.