

SYSTEMS AND METHODS FOR DETECTING TUMOR DNA IN MAMMALIAN BLOOD

CROSS REFERENCE TO RELATED PATENT APPLICATION

[0001] The present application claims the benefit of U.S. Patent Application No. 17/930,705 filed September 8, 2022, entitled “SYSTEMS AND METHODS FOR DETECTING TUMOR DNA IN MAMMALIAN BLOOD,” which claims the benefit of Vietnam Patent Application No.: 1-2022-00556 SC, filed January 25, 2022, entitled “BIOPSY PROCEDURE FOR DETECTING TUMOR DNA IN MAMMALIAN BLOOD,” and of U.S. Provisional Patent Application No. 63/373,012, filed August 19, 2022, entitled “SYSTEMS AND METHODS FOR DETECTING TUMOR DNA IN MAMMALIAN BLOOD,” which are incorporated herein by reference in their entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing that has been submitted electronically in XML file format and is hereby incorporated by reference in its entirety. The Sequence Listing for this application is labeled “132753-5001-US-Sequence Listing_XML”, which was created on September 8, 2022, and is 3,474 kilobytes in size

TECHNICAL FIELD

[0003] The present disclosure relates to the field of detecting cancer by screening for methylation patterns and size of cell-free DNA (cfDNA), also known as SPOT-MAS (Screening for Presence of Tumor by Methylation and Size of cfDNA) in biological samples.

BACKGROUND

[0004] In 2020, there was 19.2 million new cancer cases worldwide and 9.9 million cancer deaths in 2020. Among the most common types of cancer are liver cancer, lung cancer, breast cancer, stomach cancer, and colorectal cancer.

[0005] Patients with cancer found at an early stage have an increased chance of successful treatment. For post-treatment cancer patients, the early detection of cancer recurrence will also help promptly introduce new treatment regimens and increase survival time for patients.

[0006] Conventional cancer screening tests, such as endoscopic ultrasound, positron emission tomography and computed tomography (PET/CT), and biochemical tests based on marker proteins have many limitations in terms of sensitivity, specificity, invasiveness, and patient accessibility.

[0007] Recently, non-invasive testing (also known as liquid biopsy) has been proven to have potential applications in cancer diagnosis based on specific genetic variation (mutation carrier, variation in the number of genes, methylation, and size variation) of cell-free DNA (cfDNA) molecule of tumor in blood. However, many publications show that the sensitivity and specificity of cancer detection of these methods is limited by the quantity and individualization of these genetic variations. Most of the published tests used only one variable characteristic of the cfDNA molecule, so the sensitivity and specificity of detection is low and inconsistent in different types of cancer.

[0008] There are various known methods of early cancer screening based on the liquid biopsy technology such as CancerSEEK, PanSeer, Delfi and GRAIL which are detailed below herein.

[0009] *CANCERSEEK METHOD*

[0010] The CancerSEEK method, developed by the Ludwig Cancer Research at Johns Hopkins University (Cohen JD, et al., Science. 2018 Feb 23;359(6378):926-930), can detect 8 different types of cancer (including ovarian cancer, liver cancer, stomach cancer, pancreatic cancer, esophageal cancer, colon cancer, lung cancer and breast cancer). The CancerSEEK test method relied on detecting mutations of 16 specific cancer genes and combined with 8 biochemical markers to give conclusions on cancer risk.

[0011] 16 cancer-related genes were selected based on the somatic mutation dataset in cancer (Catalogue of Somatic Mutations in Cancer – COSMIC). These genes include: *TP53*, *GNAS*, *PPP2R1A*, *HRAS*, *KRAS*, *AKT1*, *PTEN*, *FGFR2*, *CDKN2A*, *BRAF*, *EGFR*, *APC*, *FBXW7*, *PIK3CA*, *CTNNB1* and *NRAS*. The presence of the mutation-carrying cfDNA molecule in the blood and combined with information from biochemical markers (CEA, CA-125, CA19-9, PRL, HGF, OPN, MPO and TIMP-1) was used to assess cancer risk.

[0012] The CancerSEEK test was performed sequentially in the following main steps:

[0013] Step 1: Collect samples, extract genetic material, prepare library and do sequencing.

[0014] Collect 10 ml of blood from patients with ovarian, liver, bronchial, pancreatic, stomach, colorectal, lung or breast cancers that are considered at stage I to III before surgery. The blood sample was then processed to obtain plasma. cfDNA was extracted from plasma using the commercial QIAAsymphony DSP Circulating DNA Kit (937556).

[0015] DNA from samples of leukemic cells and tissue embedded in paraffin from cancer patients was extracted using the commercial QIAAsymphony DSP DNA Midi Kit (937255).

[0016] Sequencing library was prepared by amplification of DNA obtained from plasma using 61 primer pairs designed to amplify the regions of interest in 16 genes of 66 to 80 base pairs in length. This library containing DNA regions (16 genes) of interest that have been purified and passed through the second amplification step to include indexing and compatible sequences for Illumina sequencing technology. Library samples were sequenced using an Illumina MiSeq or HiSeq4000 system.

[0017] *Step 2: Detect gene mutations from cfDNA.*

[0018] Gene mutations must meet one of the following two conditions: (i) being recognized in the COSMIC oncogenic somatic mutation database, or (ii) being predicted to cause inactivation of tumor suppressor genes (including nonsense mutations, addition or deletion of out-of-region fragments, classic splice site mutations). Synonymous mutations except for terminal exon and intron mutations excluding splice area were removed. The highlight of this procedure is the use of readings with unique molecular identifier (UMI) to identify each DNA fragment so that mutations with low variant allele frequency (VAF) can be detected.

[0019] *Step 3: Evaluate cancer marker protein in plasma.*

[0020] The concentration of biochemical markers in plasma samples (CEA, CA-125, CA19-9, PRL, HGF, OPN, MPO and TIMP-1) were measured using the Bioplex 200 platform system (Biorad, Hercules CA). The method was based on immunological principles using Luminex magnetic beads (Millipore, Bilerica NY) to help quantify the concentration indirectly through the calibration curve built (with Bioplex Manager 6.0 software) from standard samples and control samples available.

[0021] *Step 4: Combine gene and protein mutation analysis to detect tumor DNA.*

[0022] The VAF values of mutations detected in the DNA sample of cancer tissue and white blood cells will be used to build a probabilistic model that predicts the likelihood of mutations coming from tumor DNA. The model for the probability value of a mutation coming from the tumor is called Omega. This Omega value will be combined with the concentration of 8 biochemical markers in plasma to evaluate the probability of a diagnostic blood sample (diagnostic value of CancerSEEK) coming from 1 of 8 types of cancer surveyed. The average sensitivity of the CancerSEEK test for 8 published cancer types ranged from 33% to 98% and the specificity was 99%. In which, the detection sensitivity is less than 70% for 6/8 types of cancer surveyed, the sensitivity of the procedure to detect breast cancer is the lowest, reaching only 33%.

[0023] The CancerSEEK test for cancer detection was based on the detection of cfDNA carrying oncogenic mutations. Therefore, in the case of cancer at a very early stage, the amount of cfDNA carrying mutations existing in the blood is too small to be detected. For detection, it is necessary to increase the sequencing capacity many times over, but this significantly increases the cost of implementation. In addition, the majority of detected gene mutations can be benign mutations from white blood cells, mutations caused by cancer cells account for a small part and have individual characteristics. In order to eliminate benign mutations from white blood cells, sequencing is required twice, one for cfDNA and one for DNA from white blood cells. Combined sequencing with biochemical markers requires patients to have two tests simultaneously (with different natures in methodology) to have a basis for concluding cancer condition.

[0024] *PANSEER METHOD*

[0025] The PanSeer method relied on methylation variations of the cfDNA molecule for predictive cancer detection (Chen X, et al., Nat Commun. 2020 Jul 21;11(1):3475). The PanSeer test was implemented in the Taizhou Longitudinal (TZL) study, where collecting blood samples started from 2007 to 2016 in Taixing, Gaogang and Hailin counties. A total of 123,115 individuals aged 30-75 participated in the study, with an average condition monitoring of 8.1 years, focusing on researching 5 types of cancer, including stomach, esophagus, colorectal, lung and liver cancer.

[0026] DNA regions in the genome with different methylation states among cancer groups and normal people were selected through biological database banks such as: whole genome bisulfite sequencing (WGBS) data, methylation data from a variety of cancer tissues based on RRBS (Reduced Representation Bisulfite Sequencing) data of the research team and data from other scientific publications. From the above resources, a total of 595 DNA regions were selected to investigate the methylation states between cancer patients and healthy people.

[0027] The PanSeer test was performed sequentially in the following main steps:

[0028] *Step 1: Collect samples and extract genetic material.*

[0029] 10 ml of blood from study subjects was collected and processed for plasma collection. cfDNA was extracted from plasma using the commercial QIAamp Circulating Nucleic Acid Kit (Qiagen, 55114).

[0030] DNA from cancer tissue samples and normal human tissue samples were used from the Biochain biobank, DNA sample from the tissue was fragmented into DNA pieces with the size of about 150 nucleotides to simulate the size of cfDNA molecules using the Covaris system (which used physical force to fragment DNA).

[0031] *Step 2: Bisulfite processing, library preparation and sequencing.*

[0032] The cfDNA samples and DNA of tissue samples were treated with bisulfite using the Methylcode Bisulfite Conversion Kit (provided by ThermoFisher, MECOV50). After bisulfite processing, cfDNA molecules will be assigned sequences carrying a unique molecular identifier (UMI). The DNA sequence region of interest (595 regions of the genome containing 11,787 CpG points) was amplified using PCR (Polymerase Chain Reaction) with a specific primer set. The library containing the DNA sequence regions of interest was purified and passed through the second amplification step to include indexing and compatible sequences for Illumina sequencing technology. Library samples were sequenced on the Illumina NextSeq 500 system, paired-end sequencing mode with 300 cycles.

[0033] *Step 3: Evaluate the methylation fraction and select the DNA sequence region of interest.*

[0034] The average methylation fraction (AMF) for each sequence region was calculated as the total number of C nucleotides at all CpG sites in the sequence region of interest divided by

the total number of C nucleotides and T nucleotides at all CpG sites in this sequence region of interest. This fraction was calculated using the following formula:

$$\frac{\sum_i^M N_{C,i}}{\sum_i^M (N_{C,i} + N_{T,i})}$$

where

i: The i^{th} CpG site in the region of interest;

M: Total number of CpG in the sequence region of interest;

$N_{T,i}$: Number of T nucleotides observed at the i^{th} CpG site; and

$N_{C,i}$: Number of C nucleotides observed at the i^{th} CpG site.

[0035] AMF fractions in each sequence region of interest were compared between cancerous and healthy tissue samples. The dataset of 160 cancer tissue samples and 40 healthy tissue samples from Biochain was used to select DNA regions with different AMF values between these 2 groups of samples. The difference of AMF was tested using *t-test* (with Benjamini-Hochberg correction). Statistical test results showed that a total of 477 DNA regions (containing 10,613 CpG points) had clearly different AMF between the two groups of samples.

[0036] *Step 4: Build an algorithm model to predict cancer detection.*

[0037] To distinguish incoming plasma samples of cancer patients from the ones of healthy individuals, the PanSeer test used a logistic regression (LR) classification model that was built on the training dataset of average methylation fraction (AMF) of 477 regions of samples known as cancerous or non-cancerous samples, accompanied by a cross validation model to avoid overfitting during algorithm training. This classification model was then evaluated on the model evaluation dataset.

[0038] The limitation of the PanSeer method is that it can only distinguish between cancerous or healthy samples, in case of positive samples (classified as cancerous), the patient needs to have other blood tests and tumor monitoring with imaging tests to determine the tissue of origin.

[0039] *DELFI METHOD*

[0040] The analytical DELFI test evaluated the length of cfDNA molecules obtained from blood, to predict whether the analyzed blood sample contains the cfDNA molecule of cancer

cells (Cristiano S, et al., Nature. 2019 Jun;570(7761):385-389; Mathios D, et al., Nat Commun. 2021 Aug 20;12(1):5060). Because size-specific variations of DNA occur across the entire chromosome of cancer cells, this procedure can overcome sensitivity limitations compared with mutational markers that occur at individual sites. The DELFI procedure was implemented on 215 healthy volunteers and 208 patients in 7 cancer groups including breast cancer, colorectal cancer, lung cancer, ovarian cancer, prostate cancer, stomach cancer and gallbladder cancer.

[0041] The DELFI procedure was performed sequentially in the following main steps:

[0042] *Step 1: Collect samples and extract genetic material.*

[0043] 10 ml of blood from study subjects was collected and processed for plasma collection and monocyte subclass. cfDNA was extracted from plasma using the commercial QIAamp Circulating Nucleic Acid Kit (Qiagen, 55114). The quality of cfDNA was assessed using the Bioanalyzer 2100 electrophoresis system (Agilent Technologies).

[0044] *Step 2: Create sequencing library.*

[0045] The cfDNA sample was carried out to prepare the sequencing library using commercially available kits (NEBNext DNA library Prep kit) suitable for the Illumina sequencing technology. The cfDNA library was sequenced on Hiseq 2000/2500 system (Illumina), set to paired-end sequencing mode with 100 cycles. The DELFI test used genome-wide sequencing and DNA region-sequencing technology to evaluate abnormalities in the length of cfDNA molecules.

[0046] *Step 3: Evaluate variation in length of cfDNA.*

[0047] Sequencing data includes reads of paired-end sequences of cfDNA molecule. Typically, a cfDNA fragment will range from 50 bp to 200 bp in length. For cost savings, only sequencing about 50 bp in length was performed at each end of the cfDNA fragment. The sequencing results are put through a processing procedure to locate 2 ends of the cfDNA fragment on the original genome, thereby determining the length of that cfDNA fragment. The length of this cfDNA fragment will be used to distinguish between cancer and healthy samples. In addition, the sequencing results also give indication of mutations appearing on cfDNA and DNA from leukocytes, aiding to perform the following steps in building the predictive model.

[0048] *Step 4: Build a predictive model to detect cancer samples in two groups of people.*

[0049] The predictive model was built based on the anomalous attributes in the length of the tumor-derived cfDNA molecule. These attributes used to train the algorithm include:

[0050] The length difference between cfDNA fragments carrying mutations from the tumor and those without mutations was evaluated using Welch's two-sample t-test on 100 mutation-carrying fragments.

- The length difference of cfDNA between cancer patients and healthy subjects showed that, on average, samples from healthy subjects had longer cfDNA fragments than cancer samples (Wilcoxon rank sum test).
- The length difference of cfDNA among samples at different cancer stages and after cancer treatment.

[0051] The “Gradient tree boosting model” machine learning algorithm model was applied on 208 patients (54 breast cancer patients, 27 colorectal cancer patients, 12 lung cancer patients, 28 uterine cancer patients, 34 pancreatic cancer patients, 27 stomach cancer patients and 26 bile duct cancer patients) and 215 healthy subjects. To build a machine learning model, the algorithm divided the data into ten parts, and the algorithm used 9 parts in turn to find the differences between two groups of samples in the above 504 regions, selected those regions as characteristics to identify groups of sick and healthy people, and then rechecked the rest of samples. Since there are ten parts, the algorithm performed this calculation 10 times and found the best characteristics to help predict the two groups of samples. The DELFI model achieved a sensitivity of 80% and a specificity of 95%. This model also identified the location of cancer and achieved an accuracy of 61%. When combined with mutations detected on cell-free DNA, the model achieved a sensitivity of 91% and a specificity of 98%.

[0052] The DELFI procedure achieved a high specificity-sensitivity in patients with stage III (91%) and stage IV (82%) cancer but a lower sensitivity in patients with stage I (73%) and stage II (78%) cancer with a specificity of 95%. In addition, the procedure achieved different sensitivities, depending on the type of cancer, the highest is 100% in lung cancer, and the lowest is 70% in breast cancer and 71% in pancreatic cancer. The effectiveness of the DELFI model has not been proven through clinical trials with large samples.

[0053] *GALLERI[®] METHOD*

[0054] GALLERI (Grail) is a test to screen for >50 types of early-stage cancers based on specific methylation variation of tumor DNA released into the bloodstream (Liu MC, et al., Ann Oncol. 2020 Jun;31(6):745-759; Liu L, et al., Ann Oncol. 2018 Jun 1;29(6):1445-1453). These variations are often related to mechanisms that control the expression of many oncogenes and occur at an early stage in tumor formation and development. Using data of potential methylation markers from the whole genome sequencing and the human genome data system associated with all common cancers (The Cancer Genome Atlas - TCGA), a hybrid capture detector was designed that covers more than 100,000 target sequence regions and over 1,000,000 CpG.

[0055] The GALLERI procedure comprises the following main steps:

[0056] *Step 1: Collect samples and extract genetic material.*

[0057] cfDNA was obtained from 10 ml of blood in cancer patients and healthy subjects in the same way as the above procedures.

[0058] *Step 2: Create sequencing library.*

[0059] The sequencing library was prepared by performing bisulfite transformation of cfDNA fragments extracted from plasma. The cfDNA was then tagged with the reads needed for sequencing by the Illumina system and identifiers before being hybrid captured by the probes designed for 100,000 targets mentioned above. The entire cfDNA library was 150 bp sequenced from 2 ends of an Illumina's NovaSeq system. Target sequence fragments were aligned with the standard genome to determine the methylation status of known CpGs. Then, based on data on methylation levels at target regions in healthy people and cancer patients, models were built to assess the probability of this sequence from cancer patients.

[0060] *Step 3: Build a model to distinguish cancer samples and tumor tissue origins.*

[0061] The data was randomly divided into 2 sets including training set and control set so that the proportion of cancer samples and control samples was equivalent. In order to find the origin of sequence fragments, a model was built to detect methylation markers in each target sequence region, comparing them with the markers specific to each cancer type. Finally, a set of 2 machine learning models based on logistic regression algorithms are applied for 2 purposes: i) to distinguish the cancer group and the control group; ii) to determine the origin of tumor DNA. The effectiveness of this model combination has been verified in clinical trials. Specifically, a

recent study applying this method of the author group with the participation of about 4,000 volunteers (including 2800 cancer patients and 1200 healthy people) achieved an average sensitivity of 51.5% at a specificity of 99.5%. For some common cancers, sensitivity was improved at 67.6%.

[0062] The GALLERI test is a non-invasive method to detect cancer at early stages (I-III A). Moreover, this method can also distinguish tumor origin with high accuracy. However, due to the requirements of the analytical method, rather large sequencing capacity (30,000X) increases testing costs and reduces patient accessibility. Considering the current situation, when the cost of next-generation sequencing is still high for developing countries, reducing requirements for the depth of the sequencing method will contribute to making this research direction easier to access and soon achieve practical results.

[0063] Despite the recent development of non-invasive testing for early detection of cancer, there remains a need in the art for systems and methods to overcome the limitations of existing testing procedures. The present disclosure addresses this need.

SUMMARY OF THE INVENTION

[0064] Disclosed herein are systems and methods for detecting tumor DNA in mammalian blood cells by screening for methylation patterns and size of cell-free DNA (cfDNA).

[0065] In one aspect, the present disclosure provides methods for detecting the presence of a cancer and for identifying the cancer origin in a test subject.

[0066] The disclosed methods comprise the steps of: (a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject; (b) using the bisulfite treated cfDNA to prepare (i) a first sequencing library for a plurality of specific target genomic regions and (ii) a second sequencing library for a genome of the species of the test subject from a flow through of the first sequencing library; (c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results; (d) analyzing the corresponding first and second plurality of sequencing results by measuring:

- i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a

- plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
- ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for the plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of the plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from a cohort of a healthy subject; and
- (e) responsive to inputting into a combination model of each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model:
- i. a categorical indication of a presence or absence of the cancer in the test subject, and in the case where the model determines presence of the cancer in the test subject, an origin of the cancer.

[0067] In some embodiments, the plurality of specific target genomic regions comprises at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500 or more cancer specific regions. In some embodiments, the plurality of specific target genomic regions comprises between 400 and 500 cancer specific gene regions. In some embodiments, wherein the plurality of specific target genomic regions consists of between 17,500 and 18,500 CpG sites. In some embodiments, the plurality of specific target genomic regions comprises at least five nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments, the plurality of specific target genomic regions comprises at least 50 nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments, the plurality of specific target genomic regions comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments,

the plurality of specific target genomic regions comprises at least 300 nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments, each respective target genomic region in the plurality of specific target genomic regions encompasses a sequence selected from SEQ ID NOs: 1-450.

[0068] In some embodiments, at least 20 respective cancer specific genomic regions in the plurality of cancer specific genomic regions encompass an oncogene and/or a tumor suppressor gene listed in Table 23. In some embodiments, the plurality of cancer specific genomic regions, their respective chromosomal locations and their sequences (SEQ ID Nos: 1-450) are listed in Table 23.

[0069] In some embodiments, the plurality of specific target genomic regions is captured by a set of DNA probes. In some embodiments, the set of DNA probes comprises DNA fragments with a size ranging between 40 base-pair (bp) and 50 bp, between 51 bp and 60 bp, between 61 bp and 70 bp, between 71 bp and 80 bp, between 81 bp and 90 bp, between 91 bp and 100 bp, between 101 bp and 110 bp, between 111 bp and 120 bp, between 121 bp and 130 bp, between 131 bp and 140 bp, between 141 bp and 150 bp, between 151 bp and 160 bp, between 161 bp and 170 bp, between 171 bp and 180 bp, between 181 bp and 190 bp, or between 191 bp and 200 bp. In some embodiments, the set DNA probes comprises DNA fragments with a size ranging between 111 bp and 120 pb or between 121 bp and 130 bp. In some embodiments, the set of DNA probes consists of between 400 DNA probes and 500 DNA probes, between 501 DNA probes and 1000 DNA probes, between 1001 DNA probes and 1500 DNA probes, between 1501 DNA probes and 2000 DNA probes, between 2001 DNA probes and 2100 DNA probes, between 2101 DNA probes and 2150 DNA probes, between 2151 DNA probes and 2200 DNA probes, between 2201 DNA probes and 2250 DNA probes, between 2251 DNA probes and 2300 DNA probes, between 2301 DNA probes and 2350 DNA probes, between 2351 DNA probes and 2400 DNA probes, between 2401 DNA probes and 2450 DNA probes, between 2451 DNA probes and 2500 DNA probes, between 2501 DNA probes and 3000 DNA probes, between 3001 DNA probes and 3500 DNA probes, or between 3501 DNA probes and 4000 DNA probes. In some embodiments, the set DNA probes consists of between 2201 DNA probes and 2250 DNA probes or between 2251 DNA probes and 2300 DNA probes. In some embodiments, the set of DNA probes comprises at least 10 nucleic acid sequences selected from SEQ ID NOs: 451-2700. In some embodiments, the set of DNA probes comprises at least 100 nucleic acid sequences

selected from SEQ ID NOs: 451-2700. In some embodiments, the set of DNA probes comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 451-2700. In some embodiments, the set of DNA probes, their respective chromosomal locations, their sequences (SEQ ID NOs: 451-2700) and size (120 pb) are listed in Table 25 of U.S. Patent Application No. 17/930,705, which is hereby incorporated in its entirety by reference herein.

[0070] In some embodiments, the first sequencing library is prepared for paired-end sequencing.

[0071] In some embodiments, the plurality of specific target genomic regions have a different methylation percentage between the test subject and the cohort of healthy subjects. In some embodiments, the plurality of specific target genomic regions have a methylation percentage higher in the test subject as compared to the cohort of healthy subjects.

[0072] In some embodiments, the methylation in the test subject is about two-fold higher than the methylation in the cohort of healthy subjects.

[0073] In some embodiments, the second sequencing library comprises universal adapter sequences. In some embodiments, the genomic sequencing comprises rolling circle sequencing or MGI-DNBseq sequencing.

[0074] In some embodiments, the analysis of the sequencing results from (d)(ii)-(d)(iv) is performed by measuring non-duplicating fragments in the genome. In some embodiments, the genome comprises 22 chromosomes.

[0075] In some embodiments, the methylation density for the genome in (d)(ii) is determined for each respective second bin, in a plurality of second bins, wherein the plurality of second bins consists of between 2500 second bin regions and 3000 second bin regions. In some embodiments, each respective second bin region consists of between 800,000 nucleotides and 1,200,000 nucleotides of the genome. In some embodiments, the measuring of the methylation density identifies respective second bin regions in the plurality of second bin regions that are differentially methylated between the test subject and the cohort of healthy subjects. In some embodiments, the methylation density in each respective second bin region is evaluated based on a Z score value.

[0076] In some embodiments, the plurality of first bins is between 2500 first bins and 3000 first bins. In some embodiments, each first bin in the plurality of first bins represents a different between 800,000 nucleotides and 1,200,000 nucleotides of the genome.

[0077] In some embodiments, the measuring of respective copy number of cfDNA identifies a subset of first bins in the plurality of first bins with variation in the number of copies of DNA per bin between the test subject and the cohort of healthy subjects. In some embodiments, the variation in the number of copies of DNA between the test subject and the cohort of healthy subjects in each first bin is evaluated based on a Z score value. In some embodiments, the Z score identifies regions of instability in the genome.

[0078] In some embodiments, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third bins, wherein the plurality of third bins consists of between 500 third bins and 600 third bins. In some embodiments, each respective third bin in the plurality of third bins represents a different between 4.5 million nucleotides (4.5 megabases) and 5.5 million nucleotides (5.5 megabases) of the genome.

[0079] In some embodiments, the measuring of the fragment size pattern distribution of cfDNA identifies a subset of third bins in the plurality of third bins with a variation in the fragment size pattern distribution of cfDNA per bin between the test subject and the cohort of healthy subjects. In some embodiments, the variation in the fragment size pattern distribution of the cfDNA in each third bin in the plurality of third bins is evaluated based on cfDNA fragment length ratio (RF) value. In some embodiments, the RF value identifies presence of cancer, wherein cfDNA fragment length released from tumor cells from the test subject is shorter than cfDNA fragment length released by cells of the cohort of healthy subjects. In some embodiments, the cohort of healthy subjects consists of between 5 and 50 healthy subjects, between 5 and 100 healthy subjects, between 5 and 1000 healthy subjects, between 5 and 5000 healthy subjects, between 50 and 500 healthy subjects, between 50 and 1000 healthy subjects, between 50 and 5000 healthy subjects, between 100 and 500 healthy subjects, between 100 and 1000 healthy subjects, between 100 and 5000 healthy subjects, between 500 and 1000 healthy subjects, or between 500 and 5000 healthy subjects, or more.

[0080] In some embodiments, the liquid biopsy sample comprises a body fluid, blood, or plasma. In some embodiments, the origin of the cancer comprises colorectal cancer (CRC), liver cancer, lung cancer, breast cancer, or gastric cancer. In some embodiments, the subject is a human.

[0081] In some embodiments, the model is a composite model comprising four attribute models and a combination model, wherein each respective attribute model in the four attribute models produces an initial categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and wherein the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the four attribute models. In some embodiments, the combination model is a logistic regression combined linear model of the four attribute models, in which each of the four attribute models is independently assigned a different probability weight. In some embodiments, the model comprises at least 100 parameters. In some embodiments, the model comprises a logistic regression, a deep neural network, a fully connected neural network, a convolutional neural network, a graph based neural network, or a support vector machine. In some embodiments, the deep neural network specifies a tissue for cancer origin. In some embodiments, the deep neural network specifies a tissue for cancer origin.

[0082] In one aspect, the present disclosure provides methods for monitoring likelihood of cancer recurrence in a subject previously treated for cancer. The disclosed methods comprise the steps (a)-(e) as described above herein, wherein the detection of a cancer is indicative of cancer recurrence and need of resuming treatment to the subject.

[0083] In another aspect, the present disclosure provides methods for assessing the efficacy of a cancer treatment in a subject suffering from cancer. The disclosed methods comprise the steps (a)-(e) as described above herein, wherein the detection of a cancer is indicative of efficacy of treatment and need of continuing, modifying or discontinuing treatment of the subject.

[0084] In a further aspect, the present disclosure provides methods treating cancer in a subject in need thereof. The disclosed methods comprise the steps (a)-(e) as described above herein, wherein the detection of a cancer and the identification of the cancer origin are indicative

of the need to treat the subject and the type of treatment that is the most efficacious given the cancer origin.

[0085] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure.

[0086] Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0087] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference in their entireties to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0088] The implementations disclosed herein are illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings. Like reference numerals refer to corresponding parts throughout the several views of the drawings.

[0089] Figures 1A, 1B, and 1C collectively illustrate a computer system for detecting tumor DNA in mammalian blood, in accordance with an embodiment of the present disclosure.

[0090] Figures 2A, 2B, and 2C, collectively provide a flow chart illustrating exemplary methods for detecting tumor DNA in mammalian blood, in which dashed boxes indicate optional features, in accordance with some embodiments of the present disclosure.

[0091] Figure 3 shows a schematic diagram of the protocol for detecting tumor DNA in peripheral blood using the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0092] Figure 4 illustrates 353 sequence regions out of 450 target sequence regions to be surveyed with statistically significant differences in methyl density (p -value < 0.05) between a liver cancer group and a healthy group specified when performing the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0093] Figure 5 is a heatmap illustrating the clustering of target sequence regions between liver cancer patients and healthy subjects obtained after performing the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0094] Figure 6 illustrates the results of analysis of mean values of methylation density on all survey bins belonging to 22 chromosomes of patients with colorectal cancer (CRC) and a group of healthy people who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0095] Figure 7 shows a graph illustrating the hypomethylation change (decreased methyl ratio) on all the 'bin' regions of 22 chromosomes of the CRC group compared with the healthy group who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0096] Figure 8 shows a graph illustrating the percentage of bins that are determined to be hypomethylated between the group of colorectal cancer patients and the group of healthy people who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0097] Figure 9 is a chart illustrating the variation of DNA copy number on all 22 chromosomes of the group of colorectal cancer patients and the group of healthy people who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0098] Figure 10 is a chart comparing the percentage (%) of CNA bins in the total number of surveyed bins between the CRC group and the healthy group who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[0099] Figure 11 is a histogram showing the size distribution of cfDNA fragments in colorectal cancer samples and healthy subjects who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[00100] Figure 12 is a chart showing comparison of the ratio of small size (≤ 150) cfDNA fragments to large size (> 150 bp) ones between CRC patients and healthy people who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[00101] Figure 13 is a chart illustrating the results of evaluating the effectiveness of blood sample classification of four groups of patients with liver cancer, lung cancer, colorectal cancer, and breast cancer with blood samples of healthy people who underwent the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[00102] Figure 14 is a diagram showing the test results of blood samples from patients with liver cancer, lung cancer, colorectal cancer, and breast cancer using the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[00103] Figure 15 is a diagram depicting a Deep Neural Network (DNN) model for determining the tissue of origin for cancer. The model is built from epigenetic signatures including GC methylation, fragment length and motif end.

[00104] Figure 16 is a table depicting the tissue of origin for cancer classification performance of DNN model. The model provided probability scores of 5 cancer types (breast cancer, gastric cancer, colorectal cancer, liver cancer and lung cancer) and probability scores of unknown cancer.

DETAILED DESCRIPTION OF THE PRESENT DISCLOSURE

[00105] The present disclosure relates to the medical field, specifically relating to a liquid biopsy procedure based on screening for the presence of tumor(s) by methylation and size of cell-free DNA (cfDNA), also known as SPOT-MAS (Screening for Presence of Tumor by Methylation and Size of cfDNA) test procedure to detect tumor DNA in blood for application in screening and early detection of cancer and monitor the likelihood of post-treatment recurrence in mammals.

[00106] Reference will now be made in detail to embodiments, examples of which are illustrated in the accompanying drawings. In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the present disclosure. However, it will be apparent to one of ordinary skill in the art that the present disclosure may be practiced without these specific details. In other instances, well-known methods, procedures, components, circuits, and networks have not been described in detail so as not to unnecessarily obscure aspects of the embodiments.

[00107] The implementations described herein provide various technical solutions for screening liquid biopsy samples for detecting cancer based on the methylation and size of cfDNA, also known as SPOT-MAS (Screening for Presence Of Tumor by Methylation and Size of cfDNA) test procedure.

[00108] *Definitions.*

[00109] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described.

[00110] As used herein, each of the following terms has the meaning associated with it in this section.

[00111] As used herein, the term “about” or “approximately” mean within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which depends in part on how the value is measured or determined, e.g., the limitations of the measurement system. For example, in some embodiments, the term “about” refers to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of $\pm 20\%$ or $\pm 10\%$, more preferably $\pm 5\%$, even more preferably $\pm 1\%$, and still more preferably $\pm 0.1\%$ from the specified value, as such variations are appropriate to perform the disclosed methods. In some embodiments “about” mean within 1 or more than 1 standard deviation, per the practice in the art. In some embodiments, “about” means a range of $\pm 20\%$, $\pm 10\%$, $\pm 5\%$, or $\pm 1\%$ of a given value. In some embodiments, the term “about” or “approximately” means within an order of magnitude, within 5-fold, or within 2-fold, of a value. Where particular values are described in

the application and claims, unless otherwise stated the term “about” meaning within an acceptable error range for the particular value should be assumed. The term “about” can have the meaning as commonly understood by one of ordinary skill in the art. In some embodiments, the term “about” refers to $\pm 10\%$. In some embodiments, the term “about” refers to $\pm 5\%$.

[00112] As used herein, the terms “control,” “control sample,” “reference,” “reference sample,” “normal,” and “normal sample” describe a sample from a non-diseased tissue. In some embodiments, such a sample is from a subject that does not have a particular condition (e.g., cancer). In other embodiments, such a sample is an internal control from a subject, e.g., who may or may not have the particular disease (e.g., cancer), but is from a healthy tissue of the subject. For example, where a liquid or solid tumor sample is obtained from a subject with cancer, an internal control sample may be obtained from a healthy tissue of the subject, e.g., a white blood cell sample from a subject without a blood cancer or a solid germline tissue sample from the subject. Accordingly, a reference sample can be obtained from the subject or from a database, e.g., from a second subject who does not have the particular disease (e.g., cancer).

[00113] As used herein the term “cancer,” “cancerous tissue,” or “tumor” refers to an abnormal mass of tissue in which the growth of the mass surpasses, and is not coordinated with, the growth of normal tissue, including both solid masses (e.g., as in a solid tumor) or fluid masses (e.g., as in a hematological cancer). A cancer or tumor can be defined as “benign” or “malignant” depending on the following characteristics: degree of cellular differentiation including morphology and functionality, rate of growth, local invasion and metastasis. A “benign” tumor can be well differentiated, have characteristically slower growth than a malignant tumor and remain localized to the site of origin. In addition, in some cases a benign tumor does not have the capacity to infiltrate, invade or metastasize to distant sites. A “malignant” tumor can be a poorly differentiated (anaplasia), have characteristically rapid growth accompanied by progressive infiltration, invasion, and destruction of the surrounding tissue. Furthermore, a malignant tumor can have the capacity to metastasize to distant sites. Accordingly, a cancer cell is a cell found within the abnormal mass of tissue whose growth is not coordinated with the growth of normal tissue. Accordingly, a “tumor sample” refers to a biological sample obtained or derived from a tumor of a subject, as described herein.

[00114] Non-limiting examples of cancer types include ovarian cancer, cervical cancer, uveal melanoma, colorectal cancer, chromophobe renal cell carcinoma, liver cancer, endocrine tumor, oropharyngeal cancer, retinoblastoma, biliary cancer, adrenal cancer, neural cancer, neuroblastoma, basal cell carcinoma, brain cancer, breast cancer, non-clear cell renal cell carcinoma, glioblastoma, glioma, kidney cancer, gastrointestinal stromal tumor, medulloblastoma, bladder cancer, gastric cancer, bone cancer, non-small cell lung cancer, thymoma, prostate cancer, clear cell renal cell carcinoma, skin cancer, thyroid cancer, sarcoma, testicular cancer, head and neck cancer (e.g., head and neck squamous cell carcinoma), meningioma, peritoneal cancer, endometrial cancer, pancreatic cancer, mesothelioma, esophageal cancer, small cell lung cancer, Her2 negative breast cancer, ovarian serous carcinoma, HR+ breast cancer, uterine serous carcinoma, uterine corpus endometrial carcinoma, gastroesophageal junction adenocarcinoma, gallbladder cancer, chordoma, and papillary renal cell carcinoma.

[00115] A “disease” is a state of health of an animal where the animal cannot maintain homeostasis, and where if the disease is not ameliorated, then the animal's health continues to deteriorate. In contrast, a “disorder” in an animal is a state of health in which the animal is able to maintain homeostasis, but in which the animal's state of health is less favorable than it would be in the absence of the disorder. Left untreated, a disorder does not necessarily cause a further decrease in the animal's state of health.

[00116] As used herein, “isolated” means altered or removed from the natural state through the actions, directly or indirectly, of a human being. For example, a nucleic acid or a peptide naturally present in a living animal is not “isolated,” but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is “isolated.” An isolated nucleic acid or protein can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

[00117] As used herein, the terms “biological sample,” “patient sample,” and “sample” are interchangeably used and refer to any sample taken from a subject, which can reflect a biological state associated with the subject. In some embodiments such samples contain cell-free nucleic acids such as cell-free DNA. In some embodiments, such samples include nucleic acids other than or in addition to cell-free nucleic acids. Examples of biological samples include, but are not

limited to, blood, whole blood, plasma, serum, urine, cerebrospinal fluid, fecal, saliva, sweat, tears, pleural fluid, pericardial fluid, or peritoneal fluid of the subject. In some embodiments, the biological sample consists of blood, whole blood, plasma, serum, urine, cerebrospinal fluid, fecal, saliva, sweat, tears, pleural fluid, pericardial fluid, or peritoneal fluid of the subject. In such embodiments, the biological sample is limited to blood, whole blood, plasma, serum, urine, cerebrospinal fluid, fecal, saliva, sweat, tears, pleural fluid, pericardial fluid, or peritoneal fluid of the subject and does not contain other components (e.g., solid tissues, etc.) of the subject. A biological sample can include any tissue or material derived from a living or dead subject. A biological sample can be a cell-free sample. A biological sample can comprise a nucleic acid (e.g., DNA or RNA) or a fragment thereof. A sample can be a liquid sample or a solid sample (e.g., a cell or tissue sample). A biological sample can be a bodily fluid, such as blood, plasma, serum, urine, vaginal fluid, fluid from a hydrocele (e.g., of the testis), vaginal flushing fluids, pleural fluid, ascitic fluid, cerebrospinal fluid, saliva, sweat, tears, sputum, bronchoalveolar lavage fluid, discharge fluid from the nipple, aspiration fluid from different parts of the body (e.g., thyroid, breast), etc. A biological sample can be a stool sample. In various embodiments, the majority of DNA in a biological sample that has been enriched for cell-free DNA (e.g., a plasma sample obtained via a centrifugation protocol) can be cell-free (e.g., greater than 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the DNA can be cell-free). A biological sample can be treated to physically disrupt tissue or cell structure (e.g., centrifugation and/or cell lysis), thus releasing intracellular components into a solution which can further contain enzymes, buffers, salts, detergents, and the like which can be used to prepare the sample for analysis. A biological sample can be obtained from a subject invasively (e.g., surgical means) or non-invasively (e.g., a blood draw, a swab, or collection of a discharged sample). In some embodiments, a biological sample is derived from one tissue type (e.g., from a single organ such as breast, lung, prostate, colorectal, renal, uterine, pancreatic, esophageal, lymph, ovarian, cervical, epidermal, thyroid, bladder, or gastric). In some embodiments, a biological sample is derived from a two or more tissue types (e.g., a combination of tissue from two or more organs). In some embodiments, a biological sample is derived from one or more cell types (e.g., cells originating from a single organ or from a predetermined set of organs).

[00118] As used herein, the term “tissue” corresponds to a group of cells that group together as a functional unit. More than one type of cell can be found in a single tissue. Different types of

tissue may consist of different types of cells (e.g., hepatocytes, alveolar cells or blood cells), but also can correspond to tissue from different organisms (mother vs. fetus) or to healthy cells vs. tumor cells. The term “tissue” can generally refer to any group of cells found in the human body (e.g., heart tissue, lung tissue, kidney tissue, nasopharyngeal tissue, oropharyngeal tissue). In some aspects, the term “tissue” or “tissue type” can be used to refer to a tissue from which a cell-free nucleic acid originates. In one example, viral nucleic acid fragments can be derived from blood tissue. In another example, viral nucleic acid fragments can be derived from tumor tissue.

[00119] As used herein, the term “liquid biopsy” refers to a technique performed on non-solid biological tissue by detecting cells and cell-free DNA that have entered body fluids, primarily blood. Liquid biopsy refers to real-time monitoring of dynamic changes of the disease by detecting free tumor cells, cfDNA, exosomes, etc. This technique has great application value as a tool for early diagnosis of diseases, monitoring of progression in real time, observation and evaluation of treatment effect, prognosis assessment and metastasis risk analysis with the added benefit of being non-invasive and flexible for repeated tumor sampling.

[00120] As used herein, the term “liquid biopsy sample” refers to a liquid sample obtained from a subject that includes cell-free DNA. Examples of liquid biopsy samples include, but are not limited to, blood, whole blood, plasma, serum, urine, cerebrospinal fluid, fecal material, saliva, sweat, tears, pleural fluid, pericardial fluid, or peritoneal fluid of the subject. In some embodiments, a liquid biopsy sample is a cell-free sample, e.g., a cell free blood sample. In some embodiments, a liquid biopsy sample is obtained from a subject with cancer. In some embodiments, a liquid biopsy sample is collected from a subject with an unknown cancer status, e.g., for use in determining a cancer status of the subject. Likewise, in some embodiments, a liquid biopsy is collected from a subject with a non-cancerous disorder, e.g., a cardiovascular disease. In some embodiments, a liquid biopsy is collected from a subject with an unknown status for a non-cancerous disorder, e.g., for use in determining a non-cancerous disorder status of the subject.

[00121] As used herein, the term “cell-free DNA” and “cfDNA” interchangeably refer to DNA fragments that circulate in a subject's body (e.g., bloodstream) and originate from one or more healthy cells and/or from one or more cancer cells. These DNA molecules are found outside cells, in bodily fluids such as blood, whole blood, plasma, serum, urine, cerebrospinal

fluid, fecal material, saliva, sweat, tears, pleural fluid, pericardial fluid, or peritoneal fluid of a subject, and are believed to be fragments of genomic DNA expelled from healthy and/or cancerous cells, e.g., upon apoptosis and lysis of the cellular envelope. In some embodiments cell-free DNA (cfDNA) refers to degraded DNA fragments ranging from 50 bp to 200 bp in size that can be derived from both normal and diseased cells. cfDNA can be used to describe various forms of DNA that circulate freely in body fluids including, but not limited to, blood, sputum, urine, cerebrospinal fluid, or ascites from dead and necrosis cells. These different forms of DNA include circulating tumor DNA (ctDNA), circulating cell-free mitochondrial DNA (ccf mtDNA) and cell-free fetal DNA (cffDNA). Variations in concentrations, integrity, genetics, and epigenetics in cfDNA can suggest pathological conditions of the body, such as inflammatory diseases, autoimmune diseases, stress or even malignancies. High levels of cfDNA are commonly observed in many types of cancer, especially in advanced cancers. Clinical detection of cfDNA is a major application of liquid biopsy and is used for early diagnosis of clinical tumors, real-time monitoring of progression, observation and assessment of treatment efficacy, and prognosis assessment and metastatic risk analysis of cancer.

[00122] As used herein, the term “fragment” is used interchangeably with “nucleic acid fragment” (e.g., a DNA fragment), and refers to a portion of a polynucleotide or polypeptide sequence that comprises at least three consecutive nucleotides. In the context of sequencing of cell-free nucleic acid molecules found in a biological sample, the terms “fragment” and “nucleic acid fragment” interchangeably refer to a cell-free nucleic acid molecule that is found in the biological sample or a representation thereof. In such a context, sequencing data (e.g., sequence reads from whole genome sequencing, targeted sequencing, etc.) are used to derive one or more copies of all or a portion of such a nucleic acid fragment. Such sequence reads, which in fact may be obtained from sequencing of PCR duplicates of the original nucleic acid fragment, therefore “represent” or “support” the nucleic acid fragment. There may be a plurality of sequence reads that each represent or support a particular nucleic acid fragment in the biological sample (e.g., PCR duplicates). In some embodiments, nucleic acid fragments can be considered cell-free nucleic acids. In some embodiments, sequence reads from PCR duplicates can be misleading; for example, when the abundance level of a particular cell-free nucleic acid molecule needs to be determined. In such embodiments, only one copy of a nucleic acid fragment is used to represent the original cell-free nucleic acid molecule (e.g., duplicates are removed through

molecular identifiers that are attached to the cell-free nucleic acid molecule during the library preparation process). In some embodiments, methylation sequencing data can be used to further distinguish these nucleic acid fragments. For example, two nucleic acid fragments that share identical or near identical sequences may still correspond to different original cell-free nucleic acid molecules if they each harbor a different methylation pattern.

[00123] By “nucleic acid” is meant any nucleic acid, whether composed of deoxyribonucleosides or ribonucleosides, and whether composed of phosphodiester linkages or modified linkages such as phosphotriester, phosphoramidate, siloxane, carbonate, carboxymethylester, acetamidate, carbamate, thioether, bridged phosphoramidate, bridged methylene phosphonate, phosphorothioate, methylphosphonate, phosphorodithioate, bridged phosphorothioate or sulfone linkages, and combinations of such linkages. The term nucleic acid also specifically includes nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine and uracil).

[00124] As used herein, the term “parameter” refers to any coefficient or, similarly, any value of an internal or external element (e.g., a weight and/or a hyperparameter) in an algorithm, model, regressor, and/or classifier that can affect (e.g., modify, tailor, and/or adjust) one or more inputs, outputs, and/or functions in the algorithm, model, regressor and/or classifier. For example, in some embodiments, a parameter refers to any coefficient, weight, and/or hyperparameter that can be used to control, modify, tailor, and/or adjust the behavior, learning, and/or performance of an algorithm, model, regressor, and/or classifier. In some instances, a parameter is used to increase or decrease the influence of an input (e.g., a feature) to an algorithm, model, regressor, and/or classifier. As a nonlimiting example, in some embodiments, a parameter is used to increase or decrease the influence of a node (e.g., of a neural network), where the node includes one or more activation functions. Assignment of parameters to specific inputs, outputs, and/or functions is not limited to any one paradigm for a given algorithm, model, regressor, and/or classifier but can be used in any suitable algorithm, model, regressor, and/or classifier architecture for a desired performance. In some embodiments, a parameter has a fixed value. In some embodiments, a value of a parameter is manually and/or automatically adjustable. In some embodiments, a value of a parameter is modified by a validation and/or training process for an algorithm, model, regressor, and/or classifier (e.g., by error minimization and/or backpropagation methods). In some embodiments, an algorithm, model, regressor, and/or

classifier of the present disclosure includes a plurality of parameters. In some embodiments, the plurality of parameters is n parameters, where: $n \geq 2$; $n \geq 5$; $n \geq 10$; $n \geq 25$; $n \geq 40$; $n \geq 50$; $n \geq 75$; $n \geq 100$; $n \geq 125$; $n \geq 150$; $n \geq 200$; $n \geq 225$; $n \geq 250$; $n \geq 350$; $n \geq 500$; $n \geq 600$; $n \geq 750$; $n \geq 1,000$; $n \geq 2,000$; $n \geq 4,000$; $n \geq 5,000$; $n \geq 7,500$; $n \geq 10,000$; $n \geq 20,000$; $n \geq 40,000$; $n \geq 75,000$; $n \geq 100,000$; $n \geq 200,000$; $n \geq 500,000$, $n \geq 1 \times 10^6$, $n \geq 5 \times 10^6$, or $n \geq 1 \times 10^7$. As such, the algorithms, models, regressors, and/or classifiers of the present disclosure cannot be mentally performed. In some embodiments, n is between 10,000 and 1×10^7 , between 100,000 and 5×10^6 , or between 500,000 and 1×10^6 . In some embodiments, the algorithms, models, regressors, and/or classifier of the present disclosure operate in a k -dimensional space, where k is a positive integer of 5 or greater (e.g., 5, 6, 7, 8, 9, 10, etc.). As such, the algorithms, models, regressors, and/or classifiers of the present disclosure cannot be mentally performed.

[00125] The term, “polynucleotide” includes cDNA, RNA, DNA/RNA hybrid, anti-sense RNA, siRNA, miRNA, snoRNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified to contain non-natural or derivatized, synthetic, or semisynthetic nucleotide bases. Also, included within the scope of the invention are alterations of a wild type or synthetic gene, including but not limited to deletion, insertion, substitution of one or more nucleotides, or fusion to other polynucleotide sequences.

[00126] Conventional notation is used herein to describe polynucleotide sequences: the left-hand end of a single-stranded polynucleotide sequence is the 5'-end; the left-hand direction of a double-stranded polynucleotide sequence is referred to as the 5'-direction.

[00127] The term “oligonucleotide” typically refers to short polynucleotides, generally no greater than about 60 nucleotides. It will be understood that when a nucleotide sequence is represented by a DNA sequence (i.e., A, T, G, C), this also includes an RNA sequence (i.e., A, U, G, C) in which “U” replaces “T”.

[00128] As used herein, the terms “peptide,” “polypeptide,” or “protein” are used interchangeably, and refer to a compound comprised of amino acid residues covalently linked by peptide bonds. A protein or peptide must contain at least two amino acids, and no limitation is placed on the maximum number of amino acids that may comprise the sequence of a protein or peptide. Polypeptides include any peptide or protein comprising two or more amino acids joined

to each other by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. "Polypeptides" include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs and fusion proteins, among others. The polypeptides include natural peptides, recombinant peptides, synthetic peptides or a combination thereof. A peptide that is not cyclic will have a N-terminal and a C-terminal. The N-terminal will have an amino group, which may be free (i.e., as a NH₂ group) or appropriately protected (for example, with a BOC or a Fmoc group). The C-terminal will have a carboxylic group, which may be free (i.e., as a COOH group) or appropriately protected (for example, as a benzyl or a methyl ester). A cyclic peptide does not have free N- or C-terminal, since they are covalently bonded through an amide bond to form the cyclic structure. Amino acids may be represented by their full names (for example, leucine), 3-letter abbreviations (for example, Leu) and 1-letter abbreviations (for example, L). The structure of amino acids and their abbreviations may be found in the chemical literature, such as in Stryer, "Biochemistry", 3rd Ed., W. H. Freeman and Co., New York, 1988. tLeu represents tert-leucine. neo-Trp represents 2-amino-3-(1H-indol-4-yl)-propanoic acid. DAB is 2,4-diaminobutyric acid. Orn is ornithine. N-Me-Arg or N-methyl-Arg is 5-guanidino-2-(methylamino) pentanoic acid.

[00129] The terms "subject", "patient", "individual", and the like are used interchangeably herein, and refer to any animal, or cells thereof whether in vitro or in situ, amenable to the methods described herein. In certain non-limiting embodiments, the patient, subject or individual is a human. Non-human mammals include, for example, livestock and pets, such as ovine, bovine, porcine, canine, feline and murine mammals. Preferably, the subject is human. The term "subject" does not denote a particular age or sex. In some embodiments, the subject from whom a sample is taken, or is treated by any of the methods or compositions described herein can be of any age and can be an adult, infant or child. In some cases, the subject, e.g., patient is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 years old, or within a range therein

(e.g., between about 2 and about 20 years old, between about 20 and about 40 years old, or between about 40 and about 90 years old). A particular class of subjects, e.g., patients that can benefit from a method of the present disclosure is subjects, e.g., patients over the age of 40.

[00130] Another particular class of subjects, e.g., patients that can benefit from a method of the present disclosure is pediatric patients, who can be at higher risk of chronic heart symptoms. Furthermore, a subject, e.g., patient from whom a sample is taken, or is treated by any of the methods or compositions described herein, can be male or female.

[00131] The term "measuring" according to the present invention relates to determining the amount or concentration, preferably semi-quantitatively or quantitatively. Measuring can be done directly.

[00132] As used herein the term "amount" refers to the abundance or quantity of a constituent in a mixture.

[00133] The term "concentration" refers to the abundance of a constituent divided by the total volume of a mixture. The term concentration can be applied to any kind of chemical mixture, but most frequently it refers to solutes and solvents in solutions.

[00134] As used herein, the term "primers" or "probes" refers to DNA strands which can prime the synthesis of DNA. DNA polymerase cannot synthesize DNA de novo without primers: it can only extend an existing DNA strand in a reaction in which the complementary strand is used as a template to direct the order of nucleotides to be assembled. The synthetic oligonucleotide molecules which are used in a polymerase chain reaction (PCR) as primers are referred to as "primers".

[00135] As used herein, the term "methylation status" (also called methylation profile) can include information related to DNA methylation for a region. Information related to DNA methylation can include a methylation index of a CpG site, a methylation density of CpG sites in a region, a distribution of CpG sites over a contiguous region, a pattern or level of methylation for each individual CpG site within a region that contains more than one CpG site, and non-CpG methylation. A methylation profile of a substantial part of the genome can be considered equivalent to the methylome. "DNA methylation" in mammalian genomes can refer to the addition of a methyl group to position 5 of the heterocyclic ring of cytosine (e.g., to produce 5-

methylcytosine) among CpG dinucleotides. Methylation of cytosine can occur in cytosines in other sequence contexts, for example 5'-CHG-3' and 5'-CHH-3', where H is adenine, cytosine or thymine. Cytosine methylation can also be in the form of 5-hydroxymethylcytosine. Methylation of DNA can include methylation of non-cytosine nucleotides, such as N6-methyladenine.

[00136] As used herein, the term “methylation” refers to a modification of deoxyribonucleic acid (DNA) where a hydrogen atom on the pyrimidine ring of a cytosine base is converted to a methyl group, forming 5-methylcytosine. In particular, methylation tends to occur at dinucleotides of cytosine and guanine referred to herein as “CpG sites”. In other instances, methylation may occur at a cytosine not part of a CpG site or at another nucleotide other than cytosine; however, these are rarer occurrences. In this present disclosure, methylation is discussed in reference to CpG sites for the sake of clarity. Anomalous cfDNA methylation can be identified as hypermethylation or hypomethylation, both of which may be indicative of cancer status. As is well known in the art, DNA methylation anomalies (compared to healthy controls) can cause different effects, which may contribute to cancer. Various challenges arise in the identification of anomalously methylated cfDNA fragments. First, determining a subject's cfDNA to be anomalously methylated only holds weight in comparison with a group of control subjects, such that if the control group is small in number, the determination loses confidence with the small control group. Additionally, among a group of control subjects' methylation status can vary which can be difficult to account for when determining a subject's cfDNA to be anomalously methylated. On another note, methylation of a cytosine at a CpG site causally influences methylation at a subsequent CpG site. Those of skill in the art will appreciate that the principles described herein are equally applicable for the detection of methylation in a non-CpG context, including non-cytosine methylation.

[00137] As used herein, the terms “cut-off” or “threshold” or “reference” are used interchangeably, and refer to a value that is used as a constant and unchanging standard of comparison. In some embodiments, the terms “cutoff” and “threshold” refer to predetermined numbers used in an operation. In one example, a cutoff size refers to a size above which fragments are excluded. In some embodiments, a threshold value is a value above or below which a particular classification applies. Either of these terms can be used in either of these contexts.

[00138] As used herein, the term “ratio” refers to any comparison of a first metric X, or a first mathematical transformation thereof X' (e.g., measurement of a number of units of a genomic sequence in a first one or more biological samples or a first mathematical transformation thereof) to another metric Y or a second mathematical transformation thereof Y' (e.g., the number of units of a respective genomic sequence in a second one or more biological samples or a second mathematical transformation thereof) expressed as X/Y , Y/X , $\log N(X/Y)$, $\log N(Y/X)$, X'/Y , Y/X' , $\log N(X'/Y)$, or $\log N(Y/X')$, X/Y' , Y'/X , $\log N(X/Y')$, $\log N(Y'/X)$, X'/Y' , Y'/X' , $\log N(X'/Y')$, or $\log N(Y'/X')$, where N is any real number greater than 1 and where example mathematical transformations of X and Y include, but are limited to, raising X or Y to a power Z, multiplying X or Y by a constant Q, where Z and Q are any real numbers, and/or taking an M based logarithm of X and/or Y, where M is a real number greater than 1. In one non-limiting example, X is transformed to X' prior to ratio calculation by raising X by the power of two (X^2) and Y is transformed to Y' prior to ratio calculation by raising Y by the power of 3.2 ($Y^{3.2}$) and the ratio of X and Y is computed as $\log_2(X'/Y')$.

[00139] As used herein, the terms “sequencing,” “sequence determination,” and the like refer to any biochemical processes that may be used to determine the order of biological macromolecules such as nucleic acids or proteins. For example, sequencing data can include all or a portion of the nucleotide bases in a nucleic acid molecule such as an mRNA transcript or a genomic locus. Many sequencing techniques are available and known in the art such as but not limited to, Sanger sequencing, paired-end sequencing, pyrosequencing, and SMRT sequencing and DNB generation (e.g., Rolling circle and MGI-DNBseq G-400 sequencing).

[00140] As used herein, the term “DNA amplification” will be typically used to denote the in vitro synthesis of double-stranded DNA molecules using PCR. It is noted that other amplification methods exist and they may be used in the present invention without departing from the gist.

[00141] The term “genome”, as used herein, relates to a material or mixture of materials, containing genetic material from an organism. The term “genomic DNA” as used herein refers to deoxyribonucleic acids that are obtained from an organism. The terms “genome” and “genomic DNA” encompass genetic material that may have undergone amplification, purification, or fragmentation.

[00142] The term “sequence variation”, as used herein, refers to a difference in nucleic acid sequence between a test sample and a reference sample that may vary over a range of 1 to 10 bases, 10 to 100 bases, 100 to 100 kb, or 100 kb to 10 MB. Sequence variation may include single nucleotide polymorphism and genetic mutations relative to wild-type. In certain embodiments, sequence variation results from one or more parts of a chromosome being rearranged within a single chromosome or between chromosomes relative to a reference. In certain cases, a sequence variation may reflect a difference, e.g. abnormality, in chromosome structure, such as an inversion, a deletion, an insertion or a translocation relative to a reference chromosome, for example.

[00143] As used herein, the term “sequence reads” or “reads” refers to nucleotide sequences produced by any nucleic acid sequencing process described herein or known in the art. Reads can be generated from one end of nucleic acid fragments (“single-end reads”) or from both ends of nucleic acid fragments (e.g., paired-end reads, double-end reads). The length of the sequence read is often associated with the particular sequencing technology. High-throughput methods, for example, provide sequence reads that can vary in size from tens to hundreds of base pairs (bp). In some embodiments, the sequence reads are of a mean, median or average length of about 15 bp to 900 bp long (e.g., about 20 bp, about 25 bp, about 30 bp, about 35 bp, about 40 bp, about 45 bp, about 50 bp, about 55 bp, about 60 bp, about 65 bp, about 70 bp, about 75 bp, about 80 bp, about 85 bp, about 90 bp, about 95 bp, about 100 bp, about 110 bp, about 120 bp, about 130, about 140 bp, about 150 bp, about 200 bp, about 250 bp, about 300 bp, about 350 bp, about 400 bp, about 450 bp, or about 500 bp). In some embodiments, the sequence reads are of a mean, median or average length of about 1000 bp, 2000 bp, 5000 bp, 10,000 bp, or 50,000 bp or more. Nanopore® sequencing, for example, can provide sequence reads that can vary in size from tens to hundreds to thousands of base pairs. Illumina® parallel sequencing, for example, can provide sequence reads that do not vary as much, for example, most of the sequence reads can be smaller than 200 bp. A sequence read (or sequencing read) can refer to sequence information corresponding to a nucleic acid molecule (e.g., a string of nucleotides). For example, a sequence read can correspond to a string of nucleotides (e.g., about 20 to about 150) from part of a nucleic acid fragment, can correspond to a string of nucleotides at one or both ends of a nucleic acid fragment, or can correspond to nucleotides of the entire nucleic acid fragment. A sequence read can be obtained in a variety of ways, e.g., using sequencing techniques or using probes, e.g., in

hybridization arrays or capture probes, or amplification techniques, such as the polymerase chain reaction (PCR) or linear amplification using a single primer or isothermal amplification.

[00144] As used herein, the term “read count” refers to the total number of nucleic acid reads generated, which may or may not be equivalent to the number of nucleic acid molecules generated, during a nucleic acid sequencing reaction.

[00145] As used herein, the term “read-depth,” “sequencing depth,” or “depth” can refer to a total number of unique nucleic acid fragments encompassing a particular locus or region of the genome of a subject that are sequenced in a particular sequencing reaction. Sequencing depth can be expressed as “Y×”, e.g., 50×, 100×, etc., where “Y” refers to the number of unique nucleic acid fragments encompassing a particular locus that are sequenced in a sequencing reaction. In such a case, Y is necessarily an integer, because it represents the actual sequencing depth for a particular locus. Alternatively, read-depth, sequencing depth, or depth can refer to a measure of central tendency (e.g., a mean or mode) of the number of unique nucleic acid fragments that encompass one of a plurality of loci or regions of the genome of a subject that are sequenced in a particular sequencing reaction. For example, in some embodiments, sequencing depth refers to the average depth of every locus across an arm of a chromosome, a targeted sequencing panel, an exome, or an entire genome. In such case, Y may be expressed as a fraction or a decimal, because it refers to an average coverage across a plurality of loci. When a mean depth is recited, the actual depth for any particular locus may be different than the overall recited depth. Metrics can be determined that provide a range of sequencing depths in which a defined percentage of the total number of loci fall. For instance, a range of sequencing depths within which 90% or 95%, or 99% of the loci fall. As understood by the skilled artisan, different sequencing technologies provide different sequencing depths. For instance, low-pass whole genome sequencing can refer to technologies that provide a sequencing depth of less than 5×, less than 4×, less than 3×, or less than 2×, e.g., from about 0.5× to about 3×.

[00146] As used herein, the term “reference genome” refers to any sequenced or otherwise characterized genome, whether partial or complete, of any organism or pathogen that may be used to reference identified sequences from a subject. Typically, a reference genome will be derived from a subject of the same species as the subject whose sequences are being evaluated. Exemplary reference genomes used for human subjects as well as many other organisms are

provided in the on-line genome browser hosted by the National Center for Biotechnology Information (“NCBI”) or the University of California, Santa Cruz (UCSC). A “genome” refers to the complete genetic information of an organism or pathogen, expressed in nucleic acid sequences. As used herein, a reference sequence or reference genome often is an assembled or partially assembled genomic sequence from an individual or multiple individuals. In some embodiments, a reference genome is an assembled or partially assembled genomic sequence from one or more human individuals. The reference genome can be viewed as a representative example of a species' set of genes. In some embodiments, a reference genome comprises sequences assigned to chromosomes. Exemplary human reference genomes include but are not limited to NCBI build 34 (UCSC equivalent: hg16), NCBI build 35 (UCSC equivalent: hg17), NCBI build 36.1 (UCSC equivalent: hg18), GRCh37 (UCSC equivalent: hg19), and GRCh38 (UCSC equivalent: hg38). For a haploid genome, there can be only one nucleotide at each locus. For a diploid genome, heterozygous loci can be identified; each heterozygous locus can have two alleles, where either allele can allow a match for alignment to the locus.

[00147] As disclosed herein, the term “regions of a reference genome,” “genomic region,” or “chromosomal region” refers to any portion of a reference genome, contiguous or non-contiguous. It can also be referred to, for example, as a bin, a partition, a genomic portion, a portion of a reference genome, a portion of a chromosome and the like. In some embodiments, a genomic section is based on a particular length of genomic sequence. In some embodiments, a method can include analysis of multiple mapped nucleic acid fragments to a plurality of genomic regions. Genomic regions can be approximately the same length or the genomic sections can be different lengths. In some embodiments, genomic regions are of about equal length. In some embodiments genomic regions of different lengths are adjusted or weighted. In some embodiments, a genomic region is about 10 kilobases (kb) to about 500 kb, about 20 kb to about 400 kb, about 30 kb to about 300 kb, about 40 kb to about 200 kb, and sometimes about 50 kb to about 100 kb. In some embodiments, a genomic region is about 100 kb to about 200 kb. A genomic region is not limited to contiguous runs of sequence. Thus, genomic regions can be made up of contiguous and/or non-contiguous sequences. A genomic region is not limited to a single chromosome. In some embodiments, a genomic region includes all or part of one chromosome or all or part of two or more chromosomes. In some embodiments, genomic regions

may span one, two, or more entire chromosomes. In addition, the genomic regions may span joint or disjointed portions of multiple chromosomes.

[00148] As used herein, the term “specificity” or “true negative” or “true negative rate” refers to the number of true negatives divided by the sum of the number of true negatives and false positives. Specificity can characterize the ability of an assay or method to correctly identify a proportion of the population that truly does not have a condition. For example, specificity can characterize the ability of a method to correctly identify the number of subjects within a population not having cancer. In another example, specificity characterizes the ability of a method to correctly identify one or more markers indicative of cancer.

[00149] As used herein, an “effective amount” or “therapeutically effective amount” is an amount sufficient to affect a beneficial or desired clinical result upon treatment. An effective amount can be administered to a subject in one or more doses. In terms of treatment, an effective amount is an amount that is sufficient to palliate, ameliorate, stabilize, reverse or slow the progression of the disease, or otherwise reduce the pathological consequences of the disease. The effective amount is generally determined by the physician on a case-by-case basis and is within the skill of one in the art. Several factors are typically taken into account when determining an appropriate dosage to achieve an effective amount. These factors include age, sex and weight of the subject, the condition being treated, the severity of the condition and the form and effective concentration of the therapeutic agent being administered.

[00150] The terminology used in the present disclosure is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. As used in the description of the invention and the appended claims, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will also be understood that the term “and/or” as used herein refers to and encompasses any and all possible combinations of one or more of the associated listed items. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. Furthermore, to the extent that the terms “including,” “includes,” “having,” “has,” “with,” or variants thereof are used in either the

detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising.”

[00151] As used herein, the term “if” may be construed to mean “when” or “upon” or “in response to determining” or “in response to detecting,” depending on the context. Similarly, the phrase “if it is determined” or “if [a stated condition or event] is detected” may be construed to mean “upon determining” or “in response to determining” or “upon detecting [the stated condition or event]” or “in response to detecting [the stated condition or event],” depending on the context.

[00152] It will also be understood that, although the terms first, second, etc. may be used herein to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another. For example, a first subject could be termed a second subject, and, similarly, a second subject could be termed a first subject, without departing from the scope of the present disclosure. The first subject and the second subject are both subjects, but they are not the same subject. Furthermore, the terms “subject,” “user,” and “patient” are used interchangeably herein.

[00153] Reference will now be made in detail to embodiments, examples of which are illustrated in the accompanying drawings. In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the present disclosure, including example systems, methods, techniques, instruction sequences, and computing machine program products that embody illustrative implementations. However, the illustrative discussions below are not intended to be exhaustive or to limit the implementations to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The features described herein are not limited by the illustrated ordering of acts or events, as some acts can occur in different orders and/or concurrently with other acts or events.

[00154] The implementations provided herein are chosen and described in order to best explain the principles and their practical applications, to thereby enable others skilled in the art to best utilize the various embodiments with various modifications as are suited to the particular use contemplated. In some instances, well-known methods, procedures, components, circuits, and networks have not been described in detail so as not to unnecessarily obscure aspects of the

embodiments. In other instances, it will be apparent to one of ordinary skill in the art that the present disclosure may be practiced without one or more of the specific details.

[00155] It will be appreciated that, in the development of any such actual implementation, numerous implementation-specific decisions are made in order to achieve the designer's specific goals, such as compliance with use case- and business-related constraints, and that these specific goals will vary from one implementation to another and from one designer to another. Moreover, it will be appreciated that though such a design effort might be complex and time-consuming, it will nevertheless be a routine undertaking of engineering for those of ordinary skill in the art having the benefit of the present disclosure.

[00156] *Description.*

[00157] To overcome the limitations of existing test methods for early detection of cancer, the systems and method of the present disclosure provide a novel liquid biopsy test procedure based on the screening of cancer cells for presence of tumor by methylation and size of cfDNA, also known as SPOT-MAS (Screening for Presence Of Tumor by Methylation and Size of cfDNA) test procedure. This SPOT-MAS test procedure allows simultaneous detection of four patterns of characteristic variations of tumor DNA including: i) methylation at specific sites of genes related to tumor growth; ii) genome-wide methylation of tumor DNA; iii) genome-wide copy number abnormalities of tumor DNA; and iv) the typical size of the DNA released by the tumor into the bloodstream.

[00158] The present disclosure provides simultaneous combination of four patterns of characteristic variations of tumor DNA in the SPOT-MAS liquid biopsy test procedure helps to improve the detection efficiency of early-stage cancers, differentiate benign from malignant tumor, monitor post-treatment recurrence of tumor and locate tumor. Moreover, different types of cancer carry different characteristic variations, therefore the investigation of many attributes helps to pinpoint the exact origin of the cancer. Simultaneous analysis of many different attributes of tumor DNA is the basis for the SPOT-MAS test procedure to increase the sensitivity of cancer detection compared with procedures that rely solely on one type of attribute such as gene mutations or methyl changes in certain regions.

[00159] In the present disclosure, unless expressly stated otherwise, descriptions of devices and systems will include implementations of one or more computers. For instance, and for

purposes of illustration in Figures 1A, 1B, and 1C, a computer system 100 is represented as a single device that includes all the functionality of the computer system 100. However, the present disclosure is not limited thereto. For instance, in some embodiments, the functionality of the computer system 100 is spread across any number of networked computers and/or reside on each of several networked computers and/or by hosted on one or more virtual machines and/or containers at a remote location accessible across a communications network (*e.g.*, communications network 186 of Figure 1A). One of skill in the art will appreciate that a wide array of different computer topologies is possible for the computer system 100, and other devices and systems of the present disclosure, and that all such topologies are within the scope of the present disclosure. Moreover, rather than relying on a physical communications network 186, the illustrated devices and systems may wirelessly transmit information between each other. As such, the exemplary topology shown in Figures 1A, 1B, and 1C merely serves to describe the features of an embodiment of the present disclosure in a manner that will be readily understood to one of skill in the art.

[00160] Figures 1A, 1B, and 1C collectively depicts a block diagram of a distributed computer system (*e.g.*, computer system 100) according to some embodiments of the present disclosure. The computer system 100 at least facilitates detecting the presence of a cancer and cancer origin in a test subject.

[00161] In some embodiments, the communication network 186 optionally includes the Internet, one or more local area networks (LANs), one or more wide area networks (WANs), other types of networks, or a combination of such networks.

[00162] Examples of communication networks 186 include the World Wide Web (WWW), an intranet and/or a wireless network, such as a cellular telephone network, a wireless local area network (LAN) and/or a metropolitan area network (MAN), and other devices by wireless communication. The wireless communication optionally uses any of a plurality of communications standards, protocols and technologies, including Global System for Mobile Communications (GSM), Enhanced Data GSM Environment (EDGE), high-speed downlink packet access (HSDPA), high-speed uplink packet access (HSUPA), Evolution, Data-Only (EV-DO), HSPA, HSPA+, Dual-Cell HSPA (DC-HSPDA), long term evolution (LTE), near field communication (NFC), wideband code division multiple access (W-CDMA), code division

multiple access (CDMA), time division multiple access (TDMA), Bluetooth, Wireless Fidelity (Wi-Fi) (*e.g.*, IEEE 802.11a, IEEE 802.11ac, IEEE 802.11ax, IEEE 802.11b, IEEE 802.11g and/or IEEE 802.11n), voice over Internet Protocol (VoIP), Wi-MAX, a protocol for e-mail (*e.g.*, Internet message access protocol (IMAP) and/or post office protocol (POP)), instant messaging (*e.g.*, extensible messaging and presence protocol (XMPP), Session Initiation Protocol for Instant Messaging and Presence Leveraging Extensions (SIMPLE), Instant Messaging and Presence Service (IMPS)), and/or Short Message Service (SMS), or any other suitable communication protocol, including communication protocols not yet developed as of the filing date of this document.

[00163] In various embodiments, the computer system 100 includes one or more processing units (CPUs) 172, a network or other communications interface 174, and memory 192.

[00164] In some embodiments, the computer system 100 includes a user interface 176. The user interface 176 typically includes a display 178 for presenting media, such as a result by a respective model (*e.g.*, first model 122-1, second model 122-2, . . . , model Y 120-Y of Figure 1C). In some embodiments, the display 178 is integrated within the computer systems (*e.g.*, housed in the same chassis as the CPU 172 and memory 192). In some embodiments, the computer system 100 includes one or more input device(s) 180, which allow a subject to interact with the computer system 100. In some embodiments, input devices 180 include a keyboard, a mouse, and/or other input mechanisms. Alternatively, or in addition, in some embodiments, the display 178 includes a touch-sensitive surface (*e.g.*, where display 178 is a touch-sensitive display or computer system 100 includes a touch pad).

[00165] In some embodiments, the computer system 100 presents media to a user through the display 178. Examples of media presented by the display 178 include one or more images, a video, audio (*e.g.*, waveforms of an audio sample), or a combination thereof. In typical embodiments, the one or more images, the video, the audio, or the combination thereof is presented by the display 178 through a client application 120. In some embodiments, the audio is presented through an external device (*e.g.*, speakers, headphones, input/output (I/O) subsystem, *etc.*) that receives audio information from the computer system 100 and presents audio data based on this audio information. In some embodiments, the user interface 176 also includes an

audio output device, such as speakers or an audio output for connecting with speakers, earphones, or headphones.

[00166] Memory 192 includes high-speed random access memory, such as DRAM, SRAM, DDR RAM, or other random access solid state memory devices, and optionally also includes non-volatile memory, such as one or more magnetic disk storage devices, optical disk storage devices, flash memory devices, or other non-volatile solid state storage devices. Memory 192 may optionally include one or more storage devices remotely located from the CPU(s) 172. Memory 192, or alternatively the non-volatile memory device(s) within memory 192, includes a non-transitory computer readable storage medium. Access to memory 192 by other components of the computer system 100, such as the CPU(s) 172, is, optionally, controlled by a controller. In some embodiments, memory 192 can include mass storage that is remotely located with respect to the CPU(s) 172. In other words, some data stored in memory 192 may in fact be hosted on devices that are external to the computer system 100, but that can be electronically accessed by the computer system 100 over an Internet, intranet, or other form of network 186 or electronic cable using communication interface 184.

[00167] In some embodiments, the memory 192 of the computer system 100 for detecting the presence of a cancer and for identifying the cancer origin in a test subject stores:

- an operating system 102 (*e.g.*, ANDROID, iOS, DARWIN, RTXC, LINUX, UNIX, OS X, WINDOWS, or an embedded operating system such as VxWorks) that includes procedures for handling various basic system services;
- optionally, an electronic address 104 associated with the computer system 100 that identifies the computer system 100 (*e.g.*, within the communication network 186);
- a sequencing library store 106 that retains a record of a plurality of sequencing libraries (*e.g.*, first sequence library 108-1, second sequence library 108-2, . . . , sequence library T 108-T of Figure 1C), each sequence library prepared for a plurality of specific target genomic regions (*e.g.*, first plurality of genomic regions 110 of Figure 1B), whereby one or more sequencing libraries 108 includes a corresponding plurality of sequencing results produced therefrom that is utilized by one or more models 122 for detecting tumor DNA in mammalian blood; and
- a model library 118 that retains a plurality of models (*e.g.*, first model 120-1, second model 120-2, . . . , model Y 122-X of Figure 1C), each respective model 120 utilized for

providing, at least in part, for detecting tumor DNA in mammalian blood based on one or more parameters of a corresponding model 120 (*e.g.*, first parameter 122-1, second parameter 122-2, . . . , parameter W 122-W of first model 120-1 of Figure 1C); and

- a client application 124 for presenting information (*e.g.*, media) using a display 178 of the computer system 100.

[00168] As indicated above, an optional electronic address 104 is associated with the computer system 100. The optional electronic address 204 is utilized to at least uniquely identify the computer system 100 from other devices and components of the distributed system 100, such as other devices having access to the communications network 186. For instance, in some embodiments, the electronic address 104 is utilized to receive a request from a remote device to detect tumor DNA in mammalian blood.

[00169] Referring to Figure 1B, the sequence library 106 stores a record of a plurality of sequence libraries 108. In some embodiments, each sequencing library 108 includes data associated with a plurality of specific target genomic regions including reads of paired-end sequences of cfDNA molecule. In some such embodiments, each sequencing library 108 includes a plurality of sequencing results, such as a first plurality of sequencing results that are utilized to locate two ends of a cfDNA fragment on an original genome, thereby determining a length of that cfDNA fragment as a respective result 116.

[00170] Referring to Figure 1C, the computer system includes a model library 118 that stores a plurality of models 120 (*e.g.*, classifiers, regressors, clustering, *etc.*). In some embodiments, the model library 118 stores two more models 120 (*e.g.*, a first model 120-1 and a second model 120-2), three or more models 120, four or more models 120, ten or more models 120, 50 or more models 120, or 100 or more models 120.

[00171] In some embodiments, a model 120 in the plurality of models is implemented as an artificial intelligence engine for the subject question and answering system (QAS). For instance, in some embodiments, the model 120 includes one or more gradient boosting models 120, one or more random forest models 120, one or more neural network (NN) models 120, one or more regression models, one or more Naïve Bayes models 120, one or more machine learning algorithms (MLA) 116, or a combination thereof. In some embodiments, an MLA or a NN is trained from a training data set that includes one or more features identified from a data set.

MLAs include supervised algorithms (such as algorithms where the features/classifications in the data set are annotated) using linear regression, logistic regression, decision trees, classification and regression trees, Naïve Bayes, nearest neighbor clustering; unsupervised algorithms (such as algorithms where no features/classification in the data set are annotated a priori), such as means clustering, principal component analysis, random forest, adaptive boosting; and semi-supervised algorithms (such as algorithms where an incomplete number of features/classifications in the data set are annotated) using generative approach (such as a mixture of Gaussian distributions, mixture of multinomial distributions, hidden Markov models), low density separation, graph-based approaches (such as minimum cut, harmonic function, manifold regularization, *etc.*), heuristic approaches, or support vector machines.

[00172] In some embodiments, a model 120 is in the form of a hybrid deep learning (DL) model such as a Long Short Term Memory (LSTM) model, or a bidirectional LSTM (BiLSTM) model with an attention layer based on a neural network (NN). In some embodiments a model 120 is a deep learning model in the context of a network topology and word embedding technique customized for QAS. In some embodiments, a model 120 is a conditional random fields model 120, a convolutional neural network (CNN) model 120, an attention based neural network model 120, a deep learning model 120, a long short term memory network model 120, or another form of neural network model 120.

[00173] While MLA and neural networks identify distinct approaches to machine learning, the terms may be used interchangeably herein. Thus, a reference to MLA may include a corresponding NN or a reference to NN may include a corresponding MLA unless explicitly stated otherwise. In some embodiments, the training of a respective model 120 includes providing one or more optimized datasets, labeling these features as they occur (*e.g.*, in sequence results), and training the MLA to predict or classify based on new inputs. Artificial NNs are efficient computing models which have shown their strengths in solving hard problems in artificial intelligence. For instance, artificial NNs have also been shown to be universal approximators, that is, they can represent a wide variety of functions when given appropriate parameters.

[00174] One of skill in the art will readily appreciate other models 120 that are applicable to the systems and methods of the present disclosure. In some embodiments, the systems and

methods of the present disclosure utilize more than one model 120 to provide an evaluation (*e.g.*, arrive at an evaluation given one or more inputs), such as detecting tumor DNA in mammalian blood with an increased accuracy. For instance, in some embodiments, each respective model 120 arrives at a corresponding evaluation when provided a respective data set. Accordingly, in some embodiments, each respective model 120 independently arrives at a result and then the result of each respective model 120 is collectively verified through a comparison or amalgamation of the models 120. From this, a cumulative result is provided by the models 120. However, the present disclosure is not limited thereto.

[00175] In some embodiments, a respective model 120 is tasked with performing a corresponding activity. As a non-limiting example, in some embodiments, the task performed by the respective model 120 includes, but is not limited to, detecting a presence of a cancer and identifying a cancer origin in a test subject (*e.g.*, block 202 of Figure 2A, block 230 of Figure 2C), preparing a first sequence library 108-1 and/or a second sequencing library 108-2 (*e.g.*, block 208 of Figure 2A), sequencing the prepared first and/or second sequencing libraries (*e.g.*, block 220 of Figure 2B), producing a corresponding first and/or second plurality of sequencing results 114 (*e.g.*, block 220 of Figure 2B), analyzing the corresponding first and second plurality of sequencing results (*e.g.*, block 222 of Figure 2B), determining a categorical indication of a presence or absence of the cancer in the test subject (*e.g.*, block 230 of Figure 1C), converting the second sequencing library into cfDNA sequencing library spheres for genomic sequencing (*e.g.*, block 234 of Figure 2C) or any combination thereof.

[00176] In some embodiments, each respective model 120 of the present disclosure makes use of 10 or more parameters, 100 or more parameters, 1000 or more parameters, 10,000 or more parameters, or 100,000 or more parameters. In some embodiments, each respective model of the present disclosure cannot be mentally performed.

[00177] In some embodiments, a client application 124 is a group of instructions that, when executed by the processor 174, generates content for presentation to the user, such as a result provided by one or more models 120. In some embodiments, the client application 124 generates content in response to one or more inputs received from the user through the computer system 100, such as the inputs 180 of the computer system 100.

[00178] Each of the above identified modules and applications correspond to a set of executable instructions for performing one or more functions described above and the methods described in the present disclosure (*e.g.*, the computer-implemented methods and other information processing methods described herein; method 200 of Figures 2A through 2C; *etc.*). These modules (*e.g.*, sets of instructions) need not be implemented as separate software programs, procedures or modules, and thus various subsets of these modules are, optionally, combined or otherwise re-arranged in various embodiments of the present disclosure. In some embodiments, the memory 192 optionally stores a subset of the modules and data structures identified above. Furthermore, in some embodiments, the memory 192 stores additional modules and data structures not described above.

[00179] It should be appreciated that the computer system 100 of Figures 1A, 1B, and 1C is only one example of a computer system 100, and that the computer system 100 optionally has more or fewer components than shown, optionally combines two or more components, or optionally has a different configuration or arrangement of the components. The various components shown in Figures 1A, 1B, and 1C are implemented in hardware, software, firmware, or a combination thereof, including one or more signal processing and/or application specific integrated circuits.

[00180] Now that a general topology of the distributed system 100 has been described in accordance with various embodiments of the present disclosures, details regarding some processes in accordance with Figures 2A through 2C will be described.

[00181] Figures 2A through 2C illustrate a flow chart of methods (*e.g.*, method 200) for detecting a presence of a cancer and identifying a cancer origin in a test subject, in accordance with embodiments of the present disclosure. Specifically, an exemplary method 200 for detecting a presence of a cancer and identifying a cancer origin in a test subject is provided, in accordance with some embodiments of the present disclosure. In the flow charts, the preferred parts of the methods are shown in solid line boxes, whereas optional variants of the methods, or optional equipment used by the methods, are shown in dashed line boxes.

[00182] Various modules in the memory 192 of the computer system 100 (*e.g.*, sequence library 106, model library 118, client application 124, or a combination thereof of Figures 1A, 1B, and 1C), the memory 192 of the computer system 100, or both perform certain processes of

the methods 200 described in Figures 2A through 2C, unless expressly stated otherwise. Furthermore, it will be appreciated that the processes in Figures 2A through 2C can be encoded in a single module or any combination of modules.

[00183] *Block 202.* Referring to block 202 of Figure 2A, a method 200 detecting the presence of a cancer and for identifying the cancer origin in a test subject is provided.

[00184] In some embodiments, the method 200 is implemented at a computer system (*e.g.*, computer system 100 of Figures 1A, 1B, and 1C). The computer system includes one or more processors (*e.g.*, CPU 174 of Figure 1A) and a memory (*e.g.*, memory 192 of Figures 1A, 1B, and 1C) coupled to the one or more processors 174. The memory 192 includes one or more programs (*e.g.*, sequence library 106, model library 118, client application 124, or a combination thereof of Figures 1A, 1B, and 1C) configured to be executed by the one or more processors 174. Accordingly, in such embodiments, the one or more programs, when executed by the one or more processors, perform the method 200. As such, portions of the method 200 require a computer (*e.g.*, computer system 100 of Figures 1A, 1B, and 1C) to be used because the considerations used by the systems and methods of the present disclosure, on the scale performed by the systems and methods of the present disclosure, cannot be mentally performed. In other words, given an input to a model 120 to collectively consider each respective result, the model 120 output needs to be determined using the computer rather than mentally in such embodiments.

[00185] In one aspect, provided herein is a method for detecting the presence of a cancer and for identifying the cancer origin in a test subject. In one aspect, disclosed herein is a method for monitoring likelihood of cancer recurrence in a subject previously treated for cancer. In another aspect, provided herein is a method for assessing the efficacy of a cancer treatment in a subject suffering from cancer. In yet another aspect the present disclosure provides a method for treating cancer in a subject in need thereof.

[00186] The various disclosed methods comprise the following: (a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject (*e.g.*, block 204 of Figure 2A); (b) using the bisulfite treated cfDNA to prepare a first sequencing library for (i) a plurality of specific target genomic regions and (ii) a second sequencing library for a genome from a flow through of the first sequencing library (*e.g.*, block 208 of Figure 2A); (c) sequencing the

prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results (*e.g.*, block 220 of Figure 2B); (d) analyzing the corresponding first and second plurality of sequencing results by measuring:

[00187] i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;

[00188] ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from a cohort of healthy subjects;

[00189] iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from a cohort of healthy subjects, and

[00190] iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from a cohort of a healthy subject (*e.g.*, block 222 of Figure 2B); and

[00191] (e) responsive to inputting into a model each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model:

[00192] i. a categorical indication of a presence or absence of the cancer in the test subject, and

[00193] in the case where the model determines presence of the cancer in the test subject, an origin of the cancer (*e.g.*, block 230 of Figure 2C).

[00194] In some embodiments, the plurality of specific target genomic regions comprises at least 2550, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least

225, at least 250, at least 275, at least 300, at least 325, at least 350, at least 375, at least 400, at least 425, at least 450, at least 475, at least 500, at least 525, at least 550, at least 575, at least 600, at least 625, at least 650, at least 775, at least 800, at least 825, at least 850, at least 875, at least 900, at least 925, at least 950, at least 975, at least 1000, or more cancer specific regions.

[00195] In some embodiments, the plurality of specific target genomic regions comprises at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500 or more cancer specific regions. In some embodiments, the plurality of specific target genomic regions comprises at least 400, at least 410, at least 420, at least 430, at least 440, at least 450, at least 460, at least 470, at least 480, at least 500 or more cancer specific regions (*e.g.*, block 210 of Figure 2A). In some embodiments, the plurality of specific target genomic regions comprises at least 440, at least 441, at least 442, at least 443, at least 444, at least 445, at least 446, at least 447, at least 448, at least 449, at least 450, at least 451, at least 452, at least 453, at least 454, at least 455, at least 456, at least 457, at least 458, at least 459, at least 460 or more cancer specific regions. In some embodiments, the plurality of specific target genomic regions comprises 450 cancer specific regions. In some embodiments the 450 cancer specific regions are disclosed in Table 23 as provided elsewhere herein (SEQ ID NOs: 1-450).

[00196] In some embodiments, the methylation status comprises a methylation state of each respective CpG site in a corresponding plurality of CpG sites. In some embodiments, the plurality of specific target genomic regions consists of between 10,000 and 11,000 CpG sites, between 11,000 and 12,000 CpG sites, between 12,000 and 13,000 CpG sites, between 14,000 and 15,000 CpG sites, between 15,000 and 16,000 CpG sites, between 16,000 and 17,000 CpG sites, between 17,000 and 18,000 CpG sites, between 18,000 and 19,000 CpG sites, between 19,000 and 20,000 CpG sites, between 20,000 and 21,000 CpG sites, between 21,000 and 22,000 CpG sites, between 22,000 and 23,000 CpG sites, between 23,000 and 24,000 CpG sites, between 24,000 and 25,000 CpG sites, or more. In some embodiments, the plurality of specific target genomic regions consists of between 17,500 and 18,500 CpG sites, between 17,600 and 18,400 CpG sites, between 17,700 and 18,300 CpG sites, between 17,800 and 18,200 CpG sites, or between 17,900 and 18,100 CpG sites. In some embodiments, the plurality of specific target genomic regions consists of 18,000 CpG sites.

[00197] In some embodiments, the plurality of specific target genomic regions comprises at least 3, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 105, at least 110, at least 115, at least 120, at least 125, at least 130, at least 135, at least 140, at least 145, at least 150, at least 155, at least 160, at least 165, at least 170, at least 175, at least 180, at least 185, at least 190, at least 195, at least 200, at least 205, at least 210, at least 215, at least 220, at least 225, at least 230, at least 235, at least 240, at least 245, at least 250, at least 255, at least 260, at least 265, at least 270, at least 275, at least 280, at least 285, at least 290, at least 295, at least 300, at least 305, at least 310, at least 315, at least 320, at least 325, at least 330, at least 335, at least 340, at least 345, at least 350, at least 355, at least 360, at least 365, at least 370, at least 375, at least 380, at least 385, at least 390, at least 395, at least 400, at least 405, at least 410, at least 415, at least 420, at least 425, at least 430, at least 435, at least 440, at least 441, at least 442, at least 443, at least 444, at least 445, at least 446, at least 447, at least 448, at least 449 nucleic acid sequences selected from SEQ ID NOs: 1-450 (e.g., block 212 of Figure 2A).

[00198] In some embodiments, the plurality of specific target genomic regions comprises at least 50 nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments, the plurality of specific target genomic regions comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments, the plurality of specific target genomic regions comprises at least 300 nucleic acid sequences selected from SEQ ID NOs: 1-450. In some embodiments, each respective target genomic region in the plurality of specific target genomic regions encompasses a sequence selected from SEQ ID NOs: 1-450.

[00199] In some embodiments, at least 5, at least 10, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 105, at least 110, at least 115, at least 120, at least 125, at least 130, at least 135, at least 140, at least 145, at least 150, at least 155, at least 160, at least 165, at least 170, at least 175, at least 180, at least 185, at least 190, at least 195, at least 200, at least 205, at least 210, at least 215, at least 220, at least 225, at least 230, at least 235, at least 240, at least 245, at least 250, at least 255, at least 260, at least 265, at least 270, at least 275, at least 280, at least 285, at least 290, at least 295, at least 300, at least

305, at least 310, at least 315, at least 320, at least 325, at least 330, at least 335, at least 340, at least 345, at least 350, least 355, at least 360, at least 365, at least 370, at least 375, at least 380, at least 385, at least 390, at least 395, at least 400, at least 405, at least 410, at least 415, at least 420, at least 425, at least 430, at least 435, at least 440, at least 441, at least 442, at least 443, at least 444, at least 445, at least 446, at least 447, at least 448, at least 449 respective cancer specific genomic regions in the plurality of cancer specific genomic regions encompass an oncogene and/or a tumor suppressor gene listed in Table 23.

[00200] In some embodiments, the plurality of specific target genomic regions is captured by a set of DNA probes (*e.g.*, block 214 of Figure 2A). In some embodiments, the set of DNA probes comprises DNA fragments with a size ranging between 2 base-pair (bp) and 9 bp, between 10 bp and 19 bp, between 20 bp and 39 bp, between 40 bp and 50 bp, between 51 bp and 60 between 40 bp and 50 bp, between 51 bp and 60 bp, between 61 bp and 70 bp, between 71 bp and 80 bp, between 81 bp and 90 bp, between 91 bp and 100 bp, between 101 bp and 110 bp, between 111 bp and 120 bp, between 121 bp and 130 bp, between 131 bp and 140 bp, between 141 bp and 150 bp, between 151 bp and 160 bp, between 161 bp and 170 bp, between 171 bp and 180 bp, between 181 bp and 190 bp, between 191 bp and 200 bp or more. In some embodiments, the set DNA probes comprises DNA fragments with a size ranging between 111 bp and 120 pb or between 121 bp and 130 bp. In some embodiments, the set DNA probes comprises DNA fragments having a size of 111 bp, 112 bp, 113 bp, 114 bp, 115 bp, 116 bp, 117 bp, 118 bp, 119 bp, 120 bp, 121 bp, 122 bp, 123 bp, 124 bp, 125 bp, 126 bp, 127 bp, 128 bp, 129 bp, 130 bp. In some embodiments, the set DNA probes comprises DNA fragments having a size of 120 bp.

[00201] In some embodiments, the set of DNA probes consists of between 50 DNA probes and 99 DNA probes, between 100 DNA probes and 199 DNA probes, between 200 DNA probes and 299 DNA probes, between 300 DNA probes and 399 DNA probes, between 400 DNA probes and 500 DNA probes, between 501 DNA probes and 1000 DNA probes, between 1001 DNA probes and 1500 DNA probes, between 1501 DNA probes and 2000 DNA probes, between 2001 DNA probes and 2100 DNA probes, between 2101 DNA probes and 2150 DNA probes, between 2151 DNA probes and 2200 DNA probes, between 2201 DNA probes and 2250 DNA probes, between 2251 DNA probes and 2300 DNA probes, between 2301 DNA probes and 2350

DNA probes, between 2351 DNA probes and 2400 DNA probes, between 2401 DNA probes and 2450 DNA probes, between 2451 DNA probes and 2500 DNA probes, between 2501 DNA probes and 3000 DNA probes, between 3001 DNA probes and 3500 DNA probes, or between 3501 DNA probes and 4000 DNA probes, or more. In some embodiments, the set DNA probes consists of between 2201 DNA probes and 2250 DNA probes or between 2251 DNA probes and 2300 DNA probes.

[00202] In some embodiments, the set DNA probes consists of 2240 DNA probes, 2241 DNA probes, 2242 DNA probes, 2243 DNA probes, 2244 DNA probes, 2245 DNA probes, 2246 DNA probes, 2247 DNA, 2248 DNA probes, 2249 DNA probes, 2250 DNA probes, 2251 DNA probes, 2252 DNA probes, 2253 DNA probes, 2254 DNA probes, 2255 DNA probes, 2256 DNA probes, 2257 DNA probes and 2258 DNA probes, 2259 DNA probes or 2260 DNA probes. In some embodiments, the set DNA probes consists of 2250 DNA probes selected from Table 25 of U.S. Patent Application No. 17/930,705, which is hereby incorporated in its entirety by reference herein).

[00203] In some embodiments, the set of DNA probes comprises at least 5, at least 10, at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, least 400, at least 450, at least 500, at least 550, at least 600, at least 650, at least 700, at least 750, at least 800, at least 900, at least 1000, least 1100, at least 1150, at least 1200, at least 1250, at least 1300, at least 1350, least 1400, at least 1450, at least 1500, at least 1550, at least 1600, at least 1650, at least 1700, at least 1750, at least 1800, at least 1900, at least 2000, at least 2100, at least 2150, at least 2200, at least 2210, at least 2220, at least 2230, least 2240, at least 2249 nucleic acid sequence selected from SEQ ID NOs: 451-2700.

[00204] In some embodiments, the set of DNA probes comprises at least 10 nucleic acid sequence selected from SEQ ID NOs: 451-2700. In some embodiments, the set of DNA probes comprises at least 100 nucleic acid sequences selected from SEQ ID NOs: 451-2700. In some embodiments, the set of DNA probes comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 451-2700. In some embodiments, the set of DNA probes comprises 2250 nucleic acid sequences selected from SEQ ID NOs: 451-2700. Further information regarding SEQ ID Nos is found in Table 25 of U.S. Patent Application No. 17/930,705, which is hereby incorporated in its entirety by reference herein.

[00205] In some embodiments, the first sequencing library is prepared for paired-end sequencing. Details of exemplary sequencing library preparation are provided elsewhere herein. In some embodiments, the sequencing library allows proceeding with genomic sequencing, such as but not limited to Illumina sequencing technology (e.g., ILLUMINA MISEQ[®] or HISEQ4000[®] system).

[00206] In some embodiments, the genome comprises 22 chromosomes.

[00207] In some embodiments, the plurality of specific target genomic regions have a different methylation percentage between the test subject and a cohort of healthy subjects (*e.g.*, block 216 of Figure 2A).

[00208] In some embodiments, the methylation in the test subject is about one fold, about two fold, about three fold, about four fold, or about five fold higher or more than the methylation in the cohort of healthy subjects.

[00209] In some embodiments, the second sequencing library comprises universal adapter sequences. Usage of universal adapter and their sequences are well known in the art. In some embodiments, the universal adapters comprise a biotin-bound probes such as but not limited to, biotin-bound P5/P7 probes (Integrated DNA Technologies- IDT, USA). In some embodiments, the second sequencing library is converted into cfDNA sequencing library spheres for genomic sequencing. In some embodiments, the genomic sequencing comprises, but is not limited to, rolling circle sequencing or MGI-DNBseq G-400 sequencing.

[00210] In some embodiments, the analysis of the sequencing results from the presently disclosed methods (e.g., (d)(ii)-(d)(iv)) is performed by measuring non-duplicating fragments in the genome (*e.g.*, block 224 of Figure 2B).

[00211] In some embodiments, the methylation density for the genome in (d)(ii) of the disclosed methods is determined for each respective second bin region in between 1500 second bin regions and 2000 second bin regions, in between 200 second bin regions and 2500 second bin regions, in between 2500 second bin regions and 3000 second bin regions, or in between 3000 second bin regions and 3500 second bin regions. In some embodiments, the methylation density for the genome in (d)(ii) of the disclosed methods is determined for each respective second bin region in between 2500 second bin regions and 3000 second bin regions. In some

embodiments, the methylation density for the genome in (d)(ii) of the disclosed methods is determined for each respective second bin region of about 2730, about 2731, about 2732, about 2733, about 2734, about 2735, about 2736, about 2737, about 2738, about 2739, or about 2740 second bin regions.

[00212] In some embodiments, each respective second bin region consists of between 500,000 nucleotides and 600,000 nucleotides, between 600,000 nucleotides and 700,000 nucleotides, between 700,000 nucleotides and 800,000 nucleotides, between 800,000 nucleotides and 900,000 nucleotides, between 900,000 nucleotides and 1,000,000 nucleotides, between 1,000,000 nucleotides and 1,100,000 nucleotides, between 1,100,000 nucleotides and 1,200,000 nucleotides, between 1,200,000 nucleotides and 1,300,000 nucleotides, between 1,300,000 nucleotides and 1,400,000 nucleotides, or between 1,400,000 nucleotides and 1,500,000 nucleotides. In some embodiments, each respective second bin region consists of between 600,000 nucleotides and 1,000,000 nucleotides, between 700,000 nucleotides and 1,100,000 nucleotides, between 800,000 nucleotides and 1,300,000 nucleotides, between 900,000 nucleotides and 1,400,000 nucleotides, or between 1,000,000 nucleotides and 1,500,000 nucleotides. In some embodiments, each respective second bin region consists of between 1,000,000 nucleotides (1 megabase).

[00213] In some embodiment, the measuring of the methylation density identifies second bin regions in the between 2500 second bin regions and 3000 second bin regions that are differentially methylated between the test subject suffering and a cohort of healthy subjects. In some embodiment, the measuring of the methylation density identifies second bin regions of about 2730, about 2731, about 2732, about 2733, about 2734, about 2735, about 2736, about 2737, about 2738, about 2739, or about 2740 second bin regions that are differentially methylated between the test subject suffering and a cohort of healthy subjects.

[00214] In some embodiments, the methylation density in each respective second bin region is evaluated based on a Z score value. In some embodiments, as provided in details elsewhere herein, variation in values of methylation density in each bin is evaluated based on the “Z score” value as computed based the following formula:

$$Zscore = \frac{MD \text{ in surveyed bin} - \text{Mean MD in corresponding bin of the reference group}}{\text{Standard deviation MD in corresponding bin in the reference group}}$$

[00215] In some embodiments, the plurality of first bins is between 1500 first bin regions and 2000 first bin regions, between 200 first bin regions and 2500 first bin regions, between 2500 first bin regions and 3000 first bin regions, or between 3000 first bin regions and 3500 first bin regions. In some embodiments, the plurality of first bins is between 2500 first bin regions and 3000 first bin regions. In some embodiments, the plurality of first bins is about 2730, about 2731, about 2732, about 2733, about 2734, about 2735, about 2736, about 2737, about 2738, about 2739, or about 2740 first bin regions.

[00216] In some embodiments, each first bin consists of between 500,000 nucleotides and 600,000 nucleotides, between 600,000 nucleotides and 700,000 nucleotides, between 700,000 nucleotides and 800,000 nucleotides, between 900,000 nucleotides and 1,000,000 nucleotides, between 1,000,000 nucleotides and 1,100,000 nucleotides, between 1,200,000 nucleotides and 1,300,000 nucleotides, between 1,300,000 nucleotides and 1,400,000 nucleotides, or between 1,400,000 nucleotides and 1,500,000 nucleotides. In some embodiments, each first bin consists of between 600,000 nucleotides and 1,000,000 nucleotides, between 700,000 nucleotides and 1,100,000 nucleotides, between 800,000 nucleotides and 1,300,000 nucleotides, between 900,000 nucleotides and 1,400,000 nucleotides, or between 1,000,000 nucleotides and 1,500,000 nucleotides. In some embodiments, each first bin consists of about 1,000,000 nucleotides (1 megabase).

[00217] In some embodiment, the measuring of respective copy number of cfDNA identifies a subset of first bins in the plurality of first bins with variation in the number of copies of DNA per bin between the test subject and a cohort of healthy subjects. In some embodiments, the variation in the number of copies of DNA between the test subject and a cohort of healthy subjects in each first bin is evaluated based on a Z score value.

[00218] In some embodiment, as provided in details elsewhere herein, variation of gene copy number in each bin is evaluated based on the “Z score” value as computed in the following formula:

$$Zscore = \frac{\text{number of reads in surveyed bin} - \text{Average number of reads in corresponding bin of the reference group}}{\text{Standard deviation of the number of reads in the corresponding bin in the reference group}}$$

[00219] In some embodiments, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third binds, where the plurality of third bins consists of between 500 third bins and 600 third bins (*e.g.*, block 228 of Figure 2B).

[00220] In some embodiment, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third binds, where the plurality of third bins consists of between 100 third bins and 200 third bins, between 200 third bins and 300 third bins, between 300 third bins and 400 third bins, between 400 third bins and 500 third bins, between 500 third bins and 600 third bins, between 600 third bins and 700 third bins, between 800 third bins and 900 third bins, or between 900 third bins and 1,000 third bins. In some embodiment, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third binds, where the plurality of third bins consists of between 500 third bins and 600 third bins. In some embodiment, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third binds, where the plurality of third bins consists of between 550 third bins and 600 third bins. In some embodiment, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third binds, where the plurality of third bins consists of about 550, about 570, about 580, about 590, or about 600 third bins. In some embodiment, the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third binds, where the plurality of third bins consists of 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, or 600 third bins.

[00221] In some embodiments, each respective third bin consists of between 1 million (1 megabase) nucleotides and 1.5 million nucleotides, between 1.5 million nucleotides and 2 million nucleotides, between 2 million nucleotides and 2.5 million nucleotides, between 2.5 million nucleotides and 3 million nucleotides, between 3.5 million nucleotides and 4 million nucleotides, between 4 million nucleotides and 4.5 million nucleotides, between 5 million nucleotides and 5.5 million nucleotides, between 5.5 million nucleotides and 6 million nucleotides, between 6.5 million nucleotides and 7 million nucleotides, between 7 million

nucleotides and 7.5 million nucleotides, or between 7.5 million nucleotides and 8 million nucleotides. In some embodiments, each respective third bin consists of between 4.5 million nucleotides (4.5 megabases) and 5.5 million nucleotides (5.5 megabases). In some embodiments, each respective third bin consists of 5 million nucleotides (5 megabases).

[00222] In some embodiments, the measuring of the fragment size pattern distribution of cfDNA identifies a subset of third bins with a variation in the fragment size pattern distribution of cfDNA per bin between the test subject and a cohort of healthy subjects (*e.g.*, block 226 of Figure 2B). In some embodiments, the variation in the fragment size pattern distribution of the cfDNA in each third bin in the plurality of third bins is evaluated based on cfDNA fragment length ratio (RF) value. In some embodiments, the RF value identifies presence of cancer, where cfDNA fragment length released from tumor cells from the test subject is shorter than cfDNA fragment length released by cells of a cohort of healthy subjects.

[00223] In some embodiments, the plurality of specific target genomic regions have a methylation percentage higher in the test subject as compared to a cohort of healthy subjects. In some embodiments, the cohort of healthy subjects consists of between 5 and 50 healthy subjects, between 5 and 100 healthy subjects, between 5 and 1000 healthy subjects, between 5 and 5000 healthy subjects, between 50 and 500 healthy subjects, between 50 and 1000 healthy subjects, between 50 and 5000 healthy subjects, between 100 and 500 healthy subjects, between 100 and 1000 healthy subjects, between 100 and 5000 healthy subjects, between 500 and 1000 healthy subjects, or between 500 and 5000 healthy subjects, or more. In some embodiments, healthy subjects include for instance subjects that are not diagnosed with any disease and/or are not diagnosed with cancer. In some embodiments, the healthy subjects have the same sex and/or age range as the test subject.

[00224] In some embodiments, the liquid biopsy sample comprises a body fluid, blood, or plasma.

[00225] In some embodiments, the origin of the cancer comprises but is not limited to colorectal cancer (CRC), liver cancer, lung cancer, breast cancer (*e.g.*, block 232 of Figure 2C), or gastric cancer.

[00226] In some embodiments, the subject is a mammal. In some embodiments, the subject is a non-human mammal, such as but not limited to a livestock or a pet (e.g. ovine, bovine, porcine, canine, feline and marine mammals). In some embodiments, the subject is human.

[00227] In some embodiments, the disclosed machine learning model is a composite model comprising four attribute models and a combination model, where each respective attribute model in the four attribute models produces an initial categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and where the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the four attribute models.

[00228] In some embodiments, the combination model is a logistic regression combined linear model of the four attribute models, in which each of the four attribute models is independently assigned a different probability weight.

[00229] In some embodiments, the disclosed model (e.g., machine learning model) comprises at least 5, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least 170, at least 180, at least 190, at least 200 or more parameters. In some embodiments, the disclosed machine learning model comprises at least 100 parameters.

[00230] In some embodiments, the disclosed machine learning model comprises a logistic regression, a deep neural network, a fully connected neural network, a convolutional neural network, a graph based neural network, or a support vector machine. In some embodiments, the deep neural network specifies a tissue for cancer origin. In some embodiments, the disclosed model comprises machine learning models known in the art including but not limited to supervised algorithms (such as algorithms where the features/classifications in the data set are annotated) using linear regression, logistic regression, decision trees, classification and regression trees, naïve Bayes, nearest neighbour clustering; unsupervised algorithms (such as algorithms where no features/classification in the data set are annotated) using Apriori, means clustering, principal component analysis, random forest, adaptive boosting; and semi-supervised algorithms (such as algorithms where an incomplete number of features/classifications in the data set are annotated) using generative approach (such as a mixture of Gaussian distributions,

mixture of multinomial distributions, hidden Markov models), low density separation, graph-based approaches (such as mincut, harmonic function, manifold regularization), heuristic approaches, or support vector machines.

[00231] In one aspect, the disclosure provides a method for detecting the presence of a cancer and for identifying the cancer origin in a test subject. The disclosed method comprises a computer system having one or more processors, and memory storing one or more programs for execution by the one or more processors: obtaining, in electronic form, a sequencing data generated from a first sequencing library for (i) a plurality of specific target genomic regions and (ii) a second sequencing library for a genome from a flow through of the first sequencing library; determining a methylation pattern based on the sequencing data from the first sequencing library from the test subject relative to a cohort of healthy subjects, where the methylation pattern comprises a methylation state of each CpG site in a corresponding plurality of CpG sites in 450 cancer specific gene regions; determining a methylation pattern based on the sequencing data from the second sequencing library from the test to a cohort of healthy subjects, where the methylation pattern comprises a methylation state of each CpG site in a corresponding plurality of CpG sites in 2734 bin regions, where each bin region comprises one million nucleotides (one megabase); determining number of copies of cfDNA based on the sequencing data from the second sequencing library from the test subject suffering from cancer relative to a cohort of healthy subjects, where the number of copies of cfDNA comprises measuring of the number of copies of cfDNA in 2734 bin regions, where each bin region comprises one million nucleotides (one megabase), further where the measuring of number of copies of cfDNA identifies bin regions with variation in the number of copies of cfDNA per bin between the test subject and a cohort of healthy subjects; determining size patterns of cfDNA based on the sequencing data from the second sequencing library from the test subject relative to a cohort of healthy subjects, where the size patterns of cfDNA comprises measuring of the number of copies of cfDNA in 588 bin regions, where each bin region comprises five million nucleotides (five megabases), further where the measuring of number of copies of DNA identifies bin regions with variation in the number of copies of DNA per bin between the test subject and a cohort of healthy subjects; and applying a machine learning model for the data set for each of the (b) –(e) to indicate presence or absence of the cancer in the test subject, and in the case where the model determines presence of the cancer in the test subject, identify an origin of the cancer.

[00232] Details of an exemplary system for providing clinical support detecting cancer using a liquid biopsy assay are described in conjunction with **Figure 3** which illustrates the protocol for detecting tumor DNA in peripheral blood using the SPOT-MAS test procedure according to an embodiment of the present disclosure.

[00233] Specifically, the present disclosure provides a SPOT-MAS test procedure for detection of tumor DNA in the blood of mammals, comprising:

[00234] Element 1: Create a sequencing library of bisulfite-treated cell-free DNA (cfDNA)

[00235] *Block 204.* Referring to block 204 of Figure 2A, in some embodiments, the first element comprises collecting blood samples and processing blood sample to collect plasma and stratify monocytes. In some embodiments, the cfDNA is extracted from plasma. To perform this extraction of cfDNA, any known commercially available kit can be used, such as but not limited to the MagMAX cell-free DNA extraction kit (supplied by Thermo Fisher, USA) on KingFisher Flex Magnetic 96DW automatic system (supplied by Thermo Fisher, USA).

[00236] *Block 208.* Referring to block 208, in further embodiments, the obtained cfDNA is treated with bisulfite (BS) to convert C nucleotides without methyl moiety (-CH₃) into T nucleotides, while the C nucleotides with methyl moiety are preserved (*e.g.*, block 234 of Figure 2C). In other embodiments, purification, desulfurization and resolution are carried out to recover the bisulfite-treated cfDNA. In some embodiments, the processing of the cfDNAs can use the bisulfite conversion kit EZ_DNA methylation Gold Kit (supplied by Zymo) with the advantages of being able to convert DNA at with low cfDNA input (minimum 500 pg), achieving a conversion efficiency of over 99% and a recovery efficiency of over 75%.

[00237] In some embodiments, the cfDNAs, after being treated with bisulfite, is used to create a sequencing library. The process of preparing a sequencing library is known in the art and involves attaching fragments of nucleotide sequences (also known as adapters and indexes that contain sequences that help distinguish different library samples and sequences that pair with primers that help attach to the expository substrate) to the 2 ends of the cfDNA. In some embodiments, the procedure for attaching adapters and indexes to bisulfite-converted cfDNAs can be performed using the Accel-NGS™ Methyl-Seq DNA library kit (supplied by Swift Bioscience, USA). In some embodiments, the generated cfDNA library will be used for 2

purposes: (i) to analyze characteristic variations at 450 target sequence regions (see details in Table 23 provided elsewhere herein) and (ii) across the entire genome.

[00238] *Start here Fragmentation of the cfDNA library for variation analysis at 450 target sequence regions:*

[00239] In some embodiments, the disclosed cfDNA library relates to 450 regions (e.g., containing 18,000 CpG sites) carrying methylation characteristic variations of many recorded types of cancer (See Table 23), hybrid captured by a probe set consisting of 2250 probes with the size of 120 bp specifically designed to capture these target sequence fragments through the principle of complementary pairing (See Table 25 of U.S. Patent Application No. 17/930,705, which is hereby incorporated in its entirety by reference herein). In some embodiments, the disclosed hybrid capture procedure is performed using the xGEN® Lockdown Reagent kit (supplied by Integrated DNA Technologies-IDT, USA). To reduce the rate of nonspecific capture (including adapter fragments and high repeat sequence regions in the genome), locking and preventing probes from binding can be implemented, for example, Human Cot1 DNA (provided by Invitrogen, USA) and xGen Universal Blockers (provided by IDT, USA) can be used. After locking nonspecific sequences, this cfDNA library is hybridized with a probe set to capture target sequence regions. Next, magnetic beads are used to retain the probes bound to target sequence regions, for example, Dynabead™ streptavidin (provided by Invitrogen, USA). Meanwhile, the remaining sequences that are not captured by magnetic beads (called the “flow through” fragment) are recovered to analyze other markers. In some embodiments, the target sequence regions that have been retained by magnetic beads are then PCR amplified by, for instance, KaPa Hifi hotstart Polymerase enzyme (provided by Roche, Switzerland) with specific primers for 2 adapter fragments at 2 ends of each cfDNA fragment.

[00240] *Library fragment for analysis of genome-wide variations (“flow through” fragment):*

[00241] In some embodiments, the other cfDNA library fragment (“flow through” fragment) is recovered by hybridization with biotin-bound probes (e.g. a biotin-bound P5/P7 probe assembly provided by Integrated DNA Technologies- IDT, USA). In some embodiments, the cfDNA library fragment is obtained by streptavidin-bound magnetic beads (Dynabeads® M-270 Streptavidin beads - Invitrogen) via this bead's biotin-streptavidin binding. In some embodiments, the cfDNA library fragment is then PCR amplified and purified. PCR

amplification can be performed using various suitable polymerases enzymes such as but not limited to KaPa Hifi hotstart Polymerase enzyme (provided by Roche, Switzerland). Purification can be performed using for instance, Kapa Pure Beads (provided by Roche, Switzerland). In some embodiments, the disclosed cfDNA library fragments are further sequenced. Sequencing can be performed via various suitable sequencing techniques known in the art, such as the MGI DNB -G400 system (provided by BGI, China). In some embodiments, after sequencing, the cfDNA library for such fragment (after hybrid capture) can be used to analyze methylation density, copy number abnormalities, and typical size of cfDNA across the whole genome including 22 autosomes.

[00242] *Element 2: Analyze different variation patterns of cfDNA.*

[00243] Methylation density analysis at 450 target sequence regions:

[00244] In some embodiments, the sequencing data from the disclosed cfDNA library fragment comprises the promoter, the exons, the introns, and specific regions in the whole genome. In some embodiments, the disclosed SPOT-MAS test procedure comprises sequencing at a higher depth which increases the resolution to identify differences of methylation at the threshold level of at least 1%. Thus, the SPOT-MAS test procedure as provided herein improves sensitivity in detecting methyl changes that occur at early stages of cancer cell development.

[00245] *Genome-wide methylation density analysis:*

[00246] In some embodiments, the standard human genome is uniformly subdivided into non-duplicating fragments (bin) of 1 megabase (one million nucleotides) length (e.g., block 224 of Figure 2B). In some embodiments, the methylation density (MD) per bin is calculated using the following formula:

$$MD = \frac{\sum mC}{(\sum mC + \sum T)} \times 100$$

where $\sum mC$ is the total number of methylated C nucleotides and $\sum T$ is the total number of nucleotides.

[00247] In some embodiments, the methylation trend is evaluated based on the Z-score of each bin using the following formula:

$$\text{Zscore} = \frac{\text{MD in survey bin} - \text{Mean MD in corresponding bin of the reference group}}{\text{Standard deviation MD in corresponding bin in the reference group}}$$

[00248] In some embodiments, if the Zscore of the tested bin region is less than -3 (Zscore < -3), that bin region is less methylated than the bin in the reference group.

[00249] In some embodiments, if the Zscore of the tested bin region is between -3 and 3 (-3 < Zscore < 3), methylation in that bin region is equivalent to the bin in the reference group.

[00250] In some embodiments, if the Zscore of the test bin region is more than 3 (Zscore > 3), that bin region is more methylated than the bin in the reference group.

[00251] The analysis element as disclosed herein, helps selecting bin regions with different methyl variation levels between cancer patients and healthy people.

[00252] *Analysis of genome-wide copy number abnormalities:*

[00253] In some embodiments, the standard human genome is uniformly subdivided into non-duplicating fragments (bin) of 1 megabase (one million nucleotides) length. In some embodiments, the copy number abnormalities are evaluated using the Zscore value using the formula:

$$\text{Zscore} = \frac{\text{number of reads in survey bin} - \text{Average number of reads in the corresponding bin of the standard reference group}}{\text{Standard deviation of the number of reads in the corresponding bin in the reference group}}$$

[00254] In some embodiments, if the Zscore of the tested bin region is less than -3 (Zscore < -3), that bin region has fewer copies than the bin in the standard reference group.

[00255] In some embodiments, if the Zscore of the tested bin region is between -3 and 3 (-3 < Zscore < 3), the number of copies that bin region has is equivalent to the bin in the standard reference group.

[00256] In some embodiments, if the Zscore of the tested bin region is more than 3 (Zscore > 3), that bin region has more copies than the bin in the standard reference group.

[00257] In some embodiments, the Zscore value for variation in methyl density and DNA copy number as determined by the SPOT-MAS test helps identifying regions of genetic instability in the tumor genome. This is a prominent advantage of the SPOT-MAS test procedure because these markers contribute to accurate determination of the presence of cancer cells as well as their tissue origin based on the regions carrying these characteristic variations.

[00258] *Analysis of variation in cfDNA size:*

[00259] In some embodiments, the standard human genome is uniformly subdivided into non-duplicating fragments (bin) of 5 megabase (five million nucleotides) length. In some embodiments, within each of these bins, the ratio of the number of DNA fragments with size ≤ 150 bp to those with size > 150 bp is determined and used as a characteristic attribute of cfDNA size. It is known in the art that cancer cells tend to release more cfDNA fragments that are less than 150 bp in size. Thus determining the size difference of DNA fragments via the disclosed SPOT-MAS test procedure allows increasing the chances of tumor DNA being detected.

[00260] In one aspect, the disclosed SPOT-MAS test procedure provides generating data on different patterns of variation across the entire cell's DNA and identifying which variations are characteristic of tumor DNA. It is known in the art that methyl or size changes in tumor DNA are also markers to determine the origin of tumor DNA. Thus, incorporating the simultaneous analysis of these features by the disclosed SPOT-MAS test procedure addresses the need of increasing the chance of detecting tumor DNA and identifying its origin.

[00261] *Element 3: Build a machine learning model that predicts samples carrying cancer and tumor origin*

[00262] In some embodiments, the machine learning model distinguishes samples with/without cancer.

[00263] *Build a machine learning model for each attribute.*

[00264] In some embodiments, the process of building a machine learning model for each attribute comprises the following:

[00265] *Divide dataset:* In some embodiments, the dataset is divided into two sets, the training set and the leave-out test set using the 7:3 ratio. For the model training set, the data is

further randomly divided several times (with cross-validation) into model training and validation sets.

[00266] **Model training:** In some embodiments, the algorithm model is trained in turn with the models using the training data sets and evaluates the effectiveness of the model after training with the model validation sets using the algorithm combining 1000 basic classification models of the same type called Bagging Ensemble. This model is trained based on classification algorithms including Extreme Gradient Boosting (XGBoost), logistic regression (LR) and support vector machine (SVM) models. Nowadays, LR and SVM classification algorithms are widely applied to perform binary classification. XGBoost is a recently developed boosting algorithm and has been shown to have good speed and performance on many large datasets. For each algorithm, the parameters are adjusted to optimize for the performance (e.g., sensitivity, specificity, accuracy, etc.) of the model using the GridsearchCV algorithm.

[00267] **Set the cut-off threshold:** To set a suitable cut-off threshold for the model, it is necessary to determine the sensitivity, specificity, and accuracy of the model. In some embodiments, sensitivity, specificity and accuracy are calculated using the formula:

$$Accuracy = \frac{(a + d)}{(a + b + c + d)}$$

$$Sensitivity = \frac{(a)}{(a + c)}$$

$$Specificity = \frac{(d)}{(b + d)}$$

where:

a (true positive) is a cancer sample and is classified as cancer by the algorithm.

b (false positive) is a healthy sample and is classified as cancer by the algorithm.

c (false negative) is a cancer sample and is classified as a healthy sample by the algorithm.

d (true negative) is a healthy sample and is classified as a healthy sample by the algorithm.

[00268] In some embodiments, the cut-off threshold value is set based on the value of specificity and is surveyed to range from 0 to 1. In some embodiments, for each specificity value, a different set of sensitivity and accuracy values is obtained. From there, the ROC (receiver operating curve) model is built. In some embodiments, based on the ROC curve, a cut-off threshold is selected so that the specificity is at least 80%, at least 85%, at least 90%, at least

91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%. In some embodiments, based on the ROC curve, a cut-off threshold is selected so that the specificity is at least 95%. The area under the ROC curve is then calculated, often called AUC (area under the ROC curve). It is known in the art that the larger the area, the higher the accuracy of the model.

[00269] In some embodiments, the weight and number of occurrences of gene or bin regions in each attribute in 1000 times when training the model will be recorded and rated. The larger the weighted bin or gene regions and the higher the frequency of occurrence, the greater the significance of contributing to the model's performance.

[00270] In some embodiments, the effectiveness of the model on the leave-out test set is evaluated based on the following: After selecting a model with the best performance, the effectiveness of the selected model will be evaluated on the model evaluation dataset. Like the model training element, the indicators of specificity, sensitivity, accuracy, and AUC values of the model are determined on the model evaluation dataset. The model achieves the best performance when these values are highest and are equivalent to the values obtained in the model training element.

[00271] *Build a model that combines different attributes.*

[00272] In some embodiments, after evaluating the effectiveness of the models built on each attribute, the multi-attribute combination model is built with a strategy of linearly combining the categorical prediction results of each individual attribute.

[00273] The prediction result of individual models built on each attribute group of cfDNA is the probability value corresponding to that attribute for each sample. In some embodiments, a new dataset is formed, consisting of four categorical prediction values corresponding to four attribute groups. In some embodiments, the newly built logistic regression combined linear model as disclosed herein allows combining these attributes and determining the weight of each attribute's contribution to the final categorical prediction result. In some embodiments, the final model applied in the disclosed SPOT-MAS test procedure is a stacking model of individual attributes for the first layer and a logistic regression model for the second layer.

[00274] *Determining the origin of the tumor*

[00275] In some embodiments, after classifying cfDNA as being of tumor origin, the SPOT-MAS test procedure as provided herein further analyzes the source (from which organ in the body) of cfDNA release. The analytical procedure is based on the principle that cfDNA released from which organ will have variations in the methylation level, the size of DNA fragments that is characteristic of that organ. Specifically, the classification of tumor origin is built based on machine learning classification algorithms. In some embodiments, the attributes initially included in the analysis comprise variation in genome-wide methylation density, target methylation density, and size of cfDNA fragments (long fragment, short fragment, size ratio). In some embodiments, for each attribute type, machine learning algorithms are used to classify the tumor origin from different organ types (e.g., liver, lung, colorectal, stomach, and breast) by default to find the most suitable algorithm and attribute for the highest classification efficiency. In some embodiments, the machine learning algorithms to be surveyed include a deep neural network, logistic regression, random forest, and support vector machine. In some embodiments, the machine learning algorithm is a deep neural network.

[00276] In some embodiments, four patterns of characteristic variations in tumor DNA include:

[00277] *Methylation at specified sites of genes involved in tumor growth*

[00278] Methylation is a epigenetic mechanism known in the art that indicates when cytosine sites (C sites) in CpG islands are linked with CH₃ group. In some embodiments, to detect C sites that are linked with CH₃ group, the DNA is treated with bisulfite chemicals. Under the influence of chemicals, which C sites do not have “protection” of CH₃ group will be converted to T nucleotides while C sites that are linked with CH₃ group will be preserved. In some embodiments, sequencing methods allow determining which C sites are or are not methylated. Based on such determination, the methylation density at these sites can be calculated.

[00279] In some embodiments, the relevant genomic regions selected for investigation in the SPOT-MAS procedure are a list of 450 target gene regions containing 18,000 CpG sites that control the expression of tumor suppressor genes (Table 23). In the early stages of cancer, these regions are highly methylated to inhibit the expression of tumor suppressor genes that promote tumor proliferation and transformation. Therefore, based on this feature, it is possible to distinguish the DNA released by cancer cells into sample from the DNA of normal cells.

[00280] *Genome-wide methylation of tumors*

[00281] The methylation and determination of genome-wide methylation status of tumor are similar to the methylation at specific sites of genes associated with tumor growth. However, when investigating genome-wide methylation characteristics, many studies demonstrated that the methylation status tends to decrease in many different cancers. This tendency of methylation decrease facilitates the activation of oncogenes, especially in the early stages of tumorigenesis. Thus, when comparing the trend of genome-wide methylation in cancer patients with healthy people, the trend of methylation decrease in cancer patients has been observed. Harnessing this feature allows cancer to be identified at a very early stage.

[00282] *Genome-wide copy number abnormalities of tumor DNA.*

[00283] The presence of structural abnormalities of the chromosome is a common characteristic found in all types of cancer. These abnormalities often occur very early and accumulate gradually during the formation and growth of the tumor. Abnormalities range from fragment deletions, duplications, and inversions on whole branches of chromosomes to fragment amplifications or deletions located at different sites in the genome. The consequence of these abnormalities is structural rearrangement of genes and instability of the genome, and the resulting proteins are structurally and functionally defective.

[00284] Often, the genome in cancer patients will have regions that are amplified many times or lost some regions. By sequencing the whole genome, the number of cfDNA molecules on each bin region of the chromosome will be counted, thereby determining which bin regions increase or decrease the copy number of the entire tumor genome. When comparing the copy number of each bin region of the genome in cancer patients and healthy people, copy number abnormalities were noted. Based on the abnormality of the copy number on the whole genome, it is possible to identify the presence of cancer cells.

[00285] *Characteristic size of DNA released by the tumor into the bloodstream*

[00286] The cfDNA molecules present in the blood are released from cells undergoing the apoptosis. This apoptosis of cancer cells and normal cells is different, resulting in cfDNA released from these two cell types with different lengths. Specifically, the size of cfDNA released from tumors is usually shorter than that of cfDNA released normal cells.

[00287] To determine the size of cfDNA, whole-genome sequencing is performed to “measure” the length of the cfDNA fragments. Count the number of cfDNA molecules of the same size and use them to calculate the distribution density on a scale from 0 to 250 nucleotides. The density of cfDNA fragments smaller than 150 nucleotides is usually higher in the blood of cancer patients than in the blood of healthy individuals. Based on the size characteristics of cfDNA, it is possible to identify the presence of cancer cells.

[00288] **EXAMPLES**

[00289] The present disclosure is now described with reference to the following Examples. These Examples are provided for the purpose of illustration only and this disclosure should in no way be construed as being limited to these Examples, but rather should be construed to encompass any and all variations which become evident as a result of the teaching provided herein.

[00290] Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the compounds of the present disclosure and practice the claimed systems and methods. The following working examples, therefore, specifically point out the preferred embodiments of the present disclosure, and are not to be construed as limiting in any way the remainder of the disclosure.

[00291] In the examples disclosed herein blood tests of a group of patients with colorectal cancer (CRC), liver cancer, lung cancer, breast cancer, gastric cancer and a group of healthy people were conducted using a liquid biopsy procedure (SPOT-MAS test procedure) to detect tumor DNA.

[00292] As shown in **Figure 3**, the disclosed liquid biopsy procedure (SPOT-MAS test procedure) allows simultaneous detection of four patterns of characteristic variations of tumor DNA including: i) methylation at specific sites of genes related to tumor growth; ii) genome-wide methylation of tumor; iii) genome-wide copy number abnormalities of tumor DNA; and iv) the typical size of DNA released by the tumor into the bloodstream.

[00293] The materials and methods employed in the experiments disclosed herein are now described.

[00294] Materials and Methods

[00295] Element 1: Prepare a sequencing library of bisulfite-treated cell-free DNA (cfDNA)

[00296] *1.1 Preparing cfDNA library*

[00297] Cell-free DNA (cfDNA) is DNA that can be released from cancer cells and normal cells (leukemic cells) into the bloodstream when undergoing the apoptosis or necrosis. For cfDNA collection, blood samples can be collected and stored in a Streck cell-free DNA BCT (218997) anticoagulant test tube. First, plasma and cellular components were separated twice by centrifugation. Then, extract cfDNA from the plasma using extraction kits, for example, the MagMAX cell-free DNA extraction kit (supplied by Thermo Fisher, USA) on the KingFisher Flex Magnetic 96DW automated system (provided by Thermo Fisher, USA) following the manufacturer's instructions. At the end of the program, the resulting cfDNA was recovered and stored in a Lobind tube (Eppendorf AG), kept at -20°C if not used immediately and the concentration was evaluated using the QuantiFluor dsDNA system (provided by Promega, USA).

[00298] *1.2 Bisulfite treatment*

[00299] The treatment of cfDNA with bisulfite was carried out to convert cytosine (C)-type nucleotides with a methyl moiety (-CH₃) to uracil-type (U) nucleotides, while C-type nucleotides without methyl moiety are not converted. Thus, the treatment of cfDNA with bisulfite (BS) helps detecting methylation on cfDNA. Bisulfite conversion was performed on cfDNA using the EZ DNA Methylation-Gold Kit (provided by Zymo Research, USA) following the manufacturer's instructions. The product was then purified and desulfurized on Zymo-Spin™ IC Column. The resulting cfDNA was resolved in 7.5 µL of M-elution buffer.

[00300] *1.3 Creating cfDNA sequencing library*

[00301] After processing with BS, cfDNA was attached with adapters and indexes. An adapter is a nucleotide sequence attached to two ends of a DNA fragment that enables the DNA to attach to a rack on the surface of a flow cell in a sequencing system and be recognized by primer sequences to be amplified. An index is a nucleotide sequence that is specific to each sample and helps to distinguish different samples when performing simultaneous sequencing of multiple samples. The procedure for attaching adapters and indexes to bisulfite-converted cfDNA is known in the art and can be performed for instance by using the Accel-NGS™

Methyl-Seq DNA library kit (supplied by Swift Bioscience, USA) following the manufacturer's instructions. After attaching adapters and indexes, the cfDNA fragments were called cfDNA library and used for the portions of the pipeline.

[00302] Tumor formation and growth is the result of expression changes of many oncogenes and tumor suppressor genes. The expression of these genes is closely controlled through a methylation mechanism that occurs at regulatory regions such as promoters and enhancers regions. These regions often contain CpG islands which are CG sequences that appear with high frequency and the addition of CH₃ group (referred to as methylation) at C sites of CpG islands inhibits gene expression. Methylation at regulatory regions of tumor suppressor genes often occurs during tumor initiation. Therefore, methylation variation in these regions can be used as tumor markers. Based on previous publications and knowledge in the art, a list of 450 target genomic regions containing 18,000 CpG sites carrying characteristic methylation variation of many types of cancer has been established. To investigate the methylation density at 450 target genomic regions (Table 23), a probe set consisting of 2250 DNA fragments with the size of 120 bp was specifically designed to capture these target sequences through the principle of complementary pairing (see Table 25 of in U.S. Patent Application No. 17/930,705, which is hereby incorporated in its entirety by reference herein).

[00303] The hybrid capture procedure was performed with the xGEN® Lockdown Reagent kit (provided by Integrated DNA Technologies-IDT, USA) following the manufacturer's instructions. To reduce the rate of nonspecific capture (including adapter fragments and high repeat sequence regions in the genome), locking and preventing probes from binding was implemented, for example by using Human Cot1 DNA (provided by Invitrogen, USA) and xGen Universal Blockers (provided by IDT, USA). After locking the nonspecific sequences, the disclosed cfDNA library was hybridized with a probe set to capture target sequence regions. Next, Dynabead™ streptavidin magnetic beads (supplied by Invitrogen, USA) were used to retain the probes bound to target sequence regions. Meanwhile, the remaining sequences that were not captured by magnetic beads (called the "flow through" fragment) were recovered for other markers analysis. The target sequence regions that was retained by magnetic beads was subsequently used for PCR amplification by KAPA Hifi hotstart Polymerase enzyme (provided by Roche, Switzerland) with specific primers for 2 adapter fragments at 2 ends of each cfDNA fragment. After PCR, the concentration of cfDNA library product after hybrid capture was

quantified using the Quantus system. After the amplification reaction, the cfDNA library fragments was sequenced using paired-end sequencing mode at 100-bp on the MGI DNB -G400 system (provided by BGI, China) with a depth of 20 million reads for 1 sample.

[00304] *1.4 Collecting and processing “flow through” fragments*

[00305] After hybrid capture, the remaining cfDNA library fragments (“flow through” fragments) was recovered by hybridization with a P5/P7 probe assembly (provided by Integrated DNA Technologies- IDT, USA). These probes are nucleotide sequences with biotin molecules attached and additionally paired with adapter sequences P5 and P7 at both ends of the cfDNA library. cfDNA in this flow-through fragment, after being specifically attached to the P5/P7 probe, were collected using magnetic beads (Dynabeads® M-270 Streptavidin beads - Invitrogen) through the magnetic beads’ biotin-streptavidin binding. Then, the cfDNA library in this flow-through fragment was PCR amplified using the KaPa Hifi hotstart Polymerase enzyme (provided by Roche, Switzerland). After amplification, the product was purified using Kapa Pure Beads (provided by Roche, Switzerland). Amplified product concentration was quantified using the Quantus system. cfDNA sequencing was performed on this flow-through fragment using the MGI DNB G400 system with a depth of 20 million reads per sample as described above.

[00306] *Element 2: Analyze different variation patterns of cfDNA.*

[00307] 2.1 Analysis of methylation variation at 450 target gene regions (containing 18,000 CpG sites)

[00308] Sequencing data from cfDNA sequencing library fragments was particularly focused on promoters, exon, intron, and intergenic regions of cancer-related genes. The quality of the raw data was checked using FastQC tool (Babraham Institute, version 0.11.9). Poor quality data and adapter sequences were removed using a trimmomatic tool (USADL lab, version 0.39).

[00309] Read sequences were aligned with the standard genome and analyzed to determine methylation percentage using the Bismark aligner tool (Babraham Institute, version 16.0.2). Regions with different methylation percentages between cancer and healthy groups (called DMR: Differentially Methylated Regions) were determined by the methylation percentage per CpG determined using the following formula:

$$\text{Methylation percentage} = \frac{N_{C,i}}{N_{C,i} + N_{T,i}} \times 100\%$$

where:

i: The i^{th} CpG site in the region of interest;

$N_{T,i}$: Number of T nucleotides observed at the i^{th} CpG site; and

$N_{C,i}$: Number of C nucleotides observed at the i^{th} CpG site.

[00310] The regions with different methylation percentage between the cancer group and the healthy group were determined accordingly. Specifically, the percentage of methylation of the healthy group and the cancer group on each corresponding CpG site were compared by the Wilcoxon ranked sum test (Mann Whitney U test), in order to identify regions with (statistically significant) differences on the methylation density of CpG. The Wilcoxon ranked sum test is suitable when comparing multiple variables simultaneously between 2 groups of independent samples and variables that are not normally distributed (non-parametric test). In addition, the p-value of the statistical test was corrected using the Benjamini Hochberg method to avoid the false-positive situation encountered when the number of variables to be compared is much larger than the number of analyzed samples. The regions with different percentages of methylation between cancer and healthy groups were identified when p-value was less than 0.05 (p-value < 0.05).

[00311] The methylation fold change between the cancer group and the healthy group was determined. Specifically, the percentage of methylation (between cancer and healthy groups) on each respective CpG site is used to determine how many times the methylation fold change has changed. The methylation fold change was corrected by taking the log to base 2 ($|\log_2|$) of the absolute value of the above percentage. If this value was greater than 1, the methylation fold change has changed more than 2 times between the cancer group and the healthy group.

[00312] *2.2 Genome-wide methylation density change analysis*

[00313] The quality of the sequencing data of the flow-through library fragments was checked by using FastQC software. Poor quality data and adapter sequences were removed using a trimmomatic tool. Read sequences were aligned against the human reference genome sequence (version hg19) using the BSAAligner software in the Methyl pipe analysis package (DOI: 10.1371/journal.pone.0100360). The following parameters were checked: (1) proportion of reads is aligned against the reference genomic sequence in total mappability, (2) depth of sequencing, (3) sequencing coverage of all samples.

[00314] Genome-wide methylation variation consisting of 22 chromosomes was determined as follows. The standard human genome was uniformly subdivided into non-duplicating fragments (bin) of 1 megabase (one million nucleotides) length. Analysis of methylation variation was performed on each bin. The methylation density (MD) per bin was calculated using the following formula:

$$MD = \frac{\sum mC}{(\sum mC + \sum T)} \times 100$$

where: $\sum mC$ is the total number of methylated C nucleotides; and $\sum T$ is the total number of T nucleotides. Bins with variation in methylation state were identified. Sequencing data from 19 healthy subjects were randomly selected to determine the reference MD value for each bin. Variation in values of methylation density in each bin was evaluated based on the “Z score” value using the following formula:

$$Zscore = \frac{MD \text{ in surveyed bin} - \text{Mean MD in corresponding bin of the reference group}}{\text{Standard deviation MD in corresponding bin in the reference group}}$$

[00315] If $Zscore < -3$, that bin region was less methylated than the bin in the reference group.

[00316] If $-3 < Zscore < 3$, methylation in that bin region was equivalent to the bin in the reference group.

[00317] If $Zscore > 3$, that bin region was more methylated than the bin in the reference group.

[00318] 2.3 Genome-wide DNA copy number abnormalities analysis

[00319] Sequencing data of the flow through library fragments was used for genome-wide DNA copy number abnormalities analysis. Data quality was checked using FastQC software. Poor quality data and adapter sequences were removed using a trimmomatic tool. Read sequences were aligned against the human reference genome sequence (version hg19) using the BSAAligner software in the Methyl pipe analysis package (DOI: 10.1371/journal.pone.0100360).

[00320] The following parameters were checked: (1) proportion of reads was aligned against the reference genomic sequence in total mappability, (2) depth of sequencing, (3) sequencing

coverage of all samples. DNA copy number abnormalities analysis on 22 chromosomes was performed on each bin.

[00321] The number of copies of DNA in the bins were determined: Differences in the number of reads between bins can occur due to the influence of the bin region containing many G and C nucleotides (GC-bias) or the presence of repeat sequence regions (tandem repeat). Therefore, after alignment, the number of reads in each bin were corrected using the QDNASeq tool (DOI: 10.1101/gr.175141.114). The median copy number of all bins after correction were calculated. The degree of variation in the number of copies per bin was determined by taking the log to base 2 ($|\log_2|$) of the absolute value of the ratio of the number of reads in that bin to the median of the reads of all bins. If this value was greater than 1, then the degree of variation was more than 2 times between the investigated bin and the whole genome.

[00322] The proportion of bins with DNA copy number abnormalities between the cancer group and healthy people was determined.

[00323] Sequencing data from 19 healthy subjects were randomly selected to determine the average number of reads for each bin. Variation of gene copy number in each bin was evaluated based on the “Z score” value using the following formula:

$$Zscore = \frac{\text{number of reads in surveyed bin} - \text{Average number of reads in corresponding bin of the reference group}}{\text{Standard deviation of the number of reads in the corresponding bin in the reference group}}$$

[00324] If $Zscore < -3$, that bin region had fewer copies than the bin in the reference group

[00325] If $-3 < Zscore < 3$, the number of copies that bin region had was equivalent to the bin in the reference group

[00326] If $Zscore > 3$, that bin region had more copies than the bin in the reference group

[00327] 2.4 Analysis of variation in cfDNA size.

[00328] The sequencing data of the flow through library fragments was used to analyze variation in cfDNA size. Data quality was checked using FastQC software. Poor quality data and adapter sequences were removed using a trimmomatic tool.

[00329] Read sequences were aligned against the human reference genome sequence (version hg19) using the BSAAligner software in the Methyl pipe analysis package (DOI: 10.1371/journal.pone.0100360). Check parameters: (1) proportion of reads is aligned against the reference genomic sequence in total mappability, (2) depth of sequencing, (3) sequencing coverage of all samples.

[00330] Variation in cfDNA size was determined as follows. The standard human genome was uniformly subdivided into non-duplicating fragments (bin) of 5 megabase (5 million nucleotides) length. Size variation analysis was performed on each bin. After alignment, the length of each cfDNA fragment was calculated using software (bsalign). The size of cfDNA fragment was calculated based on the distance between the starting point of the Watson reading in the standard genome and the end point of the reading in the opposite direction (Crick). The size distribution ratio of cfDNA fragments of cancer and healthy samples in the range of 0 to 250 nucleotides was determined. Fragment ratio (RF) per bin was calculated using the following formula:

$$RF = \frac{(P \leq 150bp)}{(P > 150bp)} \times 100$$

where: $P \leq 150bp$ means length of reads is 150 nucleotides or less and $P > 150bp$ means length of reads is over 150 nucleotides.

[00331] RF variation on all 22 chromosomes was determined.

[00332] *Element 3: Build a machine learning model that predicts samples carrying cancer and tumor origin.*

[00333] Resulting analytical data in sections 2.1, 2.2, 2.3 and 2.4 as provided above herein was converted to quantitative data of 4 different attributes for each cfDNA sample including: methylation density attribute of 450 target regions (2.1); methylation density attribute of genome-wide bins (22 chromosomes) (2.2); DNA copy number attribute of genome-wide bins (22 chromosomes) (2.3); cfDNA size-specific ratio attribute of genome-wide bins (22 chromosomes) (2.4). The machine learning model was built for each individual group of attributes and combination of all attribute groups. The effectiveness of this model was evaluated based on its ability to classify 2 groups of samples as cancer and healthy people or between malignant and benign tumors.

[00334] *3.1 Machine learning model can distinguish samples with and without cancer.*

[00335] *Build a machine learning model for each attribute.*

[00336] The process of building a machine learning model for each attribute comprised the following:

[00337] *Dividing dataset:* The dataset was divided into two sets, the training set and the leave-out test set using 7:3 ratio. For the model training set, the data was further randomly divided several times (with cross-validation) into model training and validation sets.

[00338] *Model training:* The algorithm model was trained in turn with the models using the training data sets and evaluated the effectiveness of this model after training with the model validation sets using the algorithm combining 1000 basic classification models of the same type called Bagging Ensemble. This model was trained based on classification algorithms including Extreme Gradient Boosting (XGBoost), logistic regression (LR) and support vector machine (SVM) models. Nowadays, LR and SVM classification algorithms are widely used in the art to perform binary classification. XGBoost is a recently developed boosting algorithm and was shown to have good speed and performance on many large-sized datasets. For each algorithm, the parameters used in this disclosure were adjusted to optimize the efficiency of the model using the GridsearchCV algorithm.

[00339] *Set the cut-off threshold:* To set a suitable cut-off threshold for the model, it is necessary to determine sensitivity, specificity and accuracy of the model. In the present disclosure the sensitivity, specificity and accuracy were calculated using the formula:

$$Accuracy = \frac{(a + d)}{(a + b + c + d)}$$

$$Sensitivity = \frac{(a)}{(a + c)}$$

$$Specificity = \frac{(d)}{(b + d)}$$

where:

a (true positive) is a cancer sample and is classified as cancer by the algorithm,

b (false positive) is a healthy sample and is classified as cancer by the algorithm,

c (false negative) is a cancer sample and is classified as a healthy sample by the algorithm,

and

d (true negative) is a healthy sample and is classified as a healthy sample by the algorithm.

[00340] The cut-off threshold value was set based on the value of specificity and it was surveyed to range from 0 to 1. For each specificity value, a different set of sensitivity and accuracy values were obtained. From there, the ROC (receiver operating curve) model was built. From the ROC curve, the cut-off threshold was selected so that the specificity was at least 95%. The area under the ROC curve, often called AUC (area under the ROC curve), was calculated. The larger the area, the higher the accuracy of the model.

[00341] The weight and number of occurrences of the gene or bin regions in each attribute in 1000 times when training the model was recorded and rated. The larger the weighted bin or gene regions and the higher the frequency of occurrence, the greater the significance of contributing to the model's performance.

[00342] The effectiveness of the model was evaluated on the leave-out test set: After selecting the model with the best performance, the effectiveness of the selected model was evaluated on the model evaluation dataset. Similar to the model training element, the indicators of specificity, sensitivity, accuracy and AUC values of the model were determined on the model evaluation dataset. The model had the best performance when these values were the highest and were equivalent to the values obtained in the model training element.

[00343] *Build a model that combines different attributes.*

[00344] After evaluating the effectiveness of the models built on each attribute, the multi-attribute combination model was built with a strategy of linearly combining the categorical prediction results based on each individual attribute.

[00345] The prediction result of individual models built on each attribute group of cfDNA corresponded to the probability value corresponding to that attribute for each sample. Thus, a new dataset was formed, consisting of 4 categorical prediction values corresponding to 4 attribute groups. The newly built logistic regression combined linear model allowed combining these attributes and determining the weight of each attribute's contribution to the final categorical prediction result. The final model applied in the SPOT-MAS test procedure was a stacking

model of individual attributes for the first layer and a logistic regression model for the second layer.

[00346] *3.2 Determining the origin of the tumor.*

[00347] The sequence for building a model to determine the tumor origin included the following selected attributes: methyl region or bin region with methylation, the size of DNA fragments that was characteristically different between five (5) types of cancer:

- Each sample had fragment size data of 588 bins, methylation of 2734 bins and 450 regions.
- All data from samples in the cancer (5 types) group and healthy group were divided into algorithm training set (7 parts) and algorithm test set (3 parts).
- In the algorithm training sample group, the Least Absolute Shrinkage and Selection Operator (LASSO) was used to find bins with characteristically different DNA methylation or fragment sizes between 4 types of cancer.

[00348] After selecting useful attributes, a logistic regression machine learning algorithm was used to build a model using a training sample group to help determine the probability value of 5 cancer types of that sample. From there, the organ origin of ctDNA was determined based on the highest probability value of that organ.

[00349] After training, the classification algorithm was tested on a test sample set, and for each true or false classification result, the sensitivity, specificity and accuracy of the model were calculated to evaluate the classification effectiveness of the model.

[00350] *Example 1: Element 1 - Create a sequencing library of bisulfite-treated cell-free DNA (cfDNA).*

[00351] *1.1 Process blood samples to collect plasma*

[00352] A 10ml BD Vacutainer blood collection tube, USA (368589) with anticoagulant (K2- EDTA) was used to collect blood samples from the patients. Process the collected blood samples within no longer than 6 hours at a temperature of about 4°C. Separate the plasma twice by centrifugation as follows:

[00353] First centrifugation: Blood tubes were centrifuged at 1,600g for 10 min at 4°C. The upper plasma layer was gently aspirated into a 2ml Eppendorf tube without touching the

mononuclear cell layer. Then the mononuclear cells were aspirated into a 2ml Eppendorf tube and freeze at -80°C.

[00354] Second centrifugation: The above-mentioned plasma layer was centrifuged at the speed of 16,000g for 10 minutes, at 4°C. The supernatant was collected into 1.5 ml Eppendorf tubes and the residue at the bottom of the tubes was discarded. The obtained plasma sample was either used immediately for cfDNA extraction or frozen at -80°C.

[00355] 1.2 Extraction of cfDNA:

[00356] cfDNA extraction was performed on KingFisher Flex Magnetic 96DW automated system using the commercial MagMAX cell-free DNA Isolation kit (supplied by ThermoFisher Scientific, USA).

[00357] 880 μ L of plasma was used for cfDNA extraction. The plasma was divided equally between the 2 sample plates. Table 1 below lists the chemicals used for cfDNA extraction corresponding to the elements to perform the cfDNA extraction in the KingFisher Flex Magnetic 96DW with 96 deep well plate process. Be sure to use the standard plate for the 6th position and deep well plates for all other positions.

Table 1

Purpose	Plate position on the extractor	Chemicals used	Volume of each well
Lysing and mixing sample with magnetic beads	1	MagMAX™ Cell Free DNA Lysis/Binding Solution	550 μ L
		MagMAX™ Cell Free DNA Magnetic Beads	8 μ L
		Plasma blood sample	440 μ L
Lysing and mixing sample with magnetic beads	2	MagMAX™ Free DNA Cell Lysis/Binding Solution	550 μ L
		MagMAX™ Cell Free DNA Magnetic Beads	8 μ L
		Blood sample plasma	440 μ L

Purpose	Plate position on the extractor	Chemicals used	Volume of each well
1 st wash	3	MagMAX™ Cell Free DNA Wash Solution	1 mL
2 nd wash	4	80% alcohol	1 mL
3 rd wash	5	80% alcohol	500 mL
Recover cfDNA	6	MagMAX™ Cell Free DNA Elution Solution	30 µL
	7	The tip-comb was placed in deep well plate for lysis	

[00358] The attachment, washing and elution of the obtained cfDNA were performed as follows: setting parameter, selecting function for suitable plate position on KingFisher Flex Magnetic 96DW extractor. The chemical plates and samples were paced in suitable positions on the extractor and the extraction was carried out. At the end of the cycle (approximately 47 minutes), the cfDNA recovery plate located at the 6th position on the extractor was removed from the extractor. The cfDNA sample was either used immediately for the next element or transferred to a Lobind tube (Eppendorf AG) for storage at -20°C for a long-term use.

[00359] *1.3 Measure cfDNA concentration using QuantiFluor dsDNA system.*

[00360] The concentration of cfDNA was measured with Quantus™ Fluorometer (E6150) measuring system, using QuantiFlour dsDNA system (E2670). This was as follows: Dilute 20X TE buffer 20 times with distilled water to obtain 1X TE buffer. Dilute QuantiFlour dsDNA dye 400 times with 1X TE buffer to obtain a measuring buffer. Aspirate 198 µL of measuring buffer into a 0.5 ml thin-walled PCR tube (Cat.# E4941). Add 2 µL of cfDNA sample to be measured into the PCR tube and incubate at room temperature for 5 minutes, avoiding direct sunlight. Measure sample with Quantus™ Fluorometer meter system and record the obtained cfDNA concentration.

[00361] *1.4 Bisulfite treatment (BS).*

[00362] Bisulfite treatment of cfDNA was performed with 2ng cfDNA using Zymo EZ DNA Gold methylation reagent kit (D5006), including the following:

[00363] *CT conversion reaction.*

[00364] CT conversion reagent tube was dissolved with 900 μL of H_2O , 300 μL of M-Dilution buffer and 50 μL of M-Dissolving buffer. The tube was placed on a shaker for 10 minutes or until completely dissolved. 20 μL of cfDNA were aspirated into 0.2 mL PCR tube. The amount of H_2O was adjusted so that the volume of cfDNA in the tube reached 2ng. 130 μL of CT conversion reagent were added and mixed by suction and release 10 times. The mixture was placed in a heat cycler and the thermal process followed the settings shown in the Table 2 below.

Table 2

Element	Temperature	Time
1	98°C	10 minutes
2	64°C	2.5 hour
Kept at 4°C		

[00365] Purifying the product after bisulfite modification.

[00366] The purification element involved the following: Prepare an M-wash buffer by adding 24 ml of 100% alcohol to 6 ml of concentrated M-wash buffer. Prepare the Zymo-Spin™ IC membrane kit and collection column. Add 600 μL of M-binding buffer into the membrane kit. Aspirate all 150 μL of the CT conversion product mixture in the PCR tube into the collection column and mix well by manually inverting several times. Centrifuge the collection column at 11,000g for 30 seconds and then discard the solution in the collection column. Add 100 μL of M-wash buffer to the collection column and centrifuge the second time at 11,000g for 30 seconds. Add 200 μL of M-Desulphonation buffer to the collection column and incubate at room temperature for 15 minutes. Then centrifuge the column for the third time at 11,000g for 30 seconds. Add another 200 μL of M-wash buffer to the collection column and centrifuge the fourth time at 11,000g for 30 seconds. Discard the solution in the collection column and continue adding 200 μL of M-wash buffer. Then centrifuge the column for the fifth time at 11,000g for 30

seconds. Empty the collection column and transfer Zymo-Spin™ IC membrane to a new 1.5 ml Eppendorf tube. Add 7.5 µL of M-elution buffer to the center of the membrane and incubate for 5 minutes at room temperature, centrifuge at maximum speed for 1 minutes to obtain cfDNA sample. This cfDNA sample can be used immediately or stored at -20°C.

[00367] *1.5 Generating a sequencing library for bisulfite treated cfDNA.*

[00368] *Attaching adapters and indexes.*

[00369] Denaturation-separation of cfDNA: After bisulfite treatment, cfDNA product was denatured to separate single-stranded cfDNA by incubation at 95°C for 2 minutes in a heat cycler. The sample was immediately removed and placed on cold ice for 2 minutes to prevent regurgitation. A reaction mixture was prepared for attaching the adapter 1 to the components as shown in the Table 3 below.

Table 3

Chemicals	Volume (µL)
Low TE buffer	6.75
G1 buffer	2
G2 chemicals	2
G3 chemicals	1.25
G4 yeast	0.5
G5 yeast	0.5
G6 yeast	0.5
Total volume	13.5

[00370] 13.5 µL of the above reaction mixture was added into 7.5 µL of cfDNA sample after the denaturation-separation element. The reaction mixture was mixed well by suction-release 10 times and incubated in a heat cycler with the program set at the temperature and time shown in the Table 4 below.

Table 4

Element	Temperature	Time
---------	-------------	------

1	37°C	15 minutes
2	95°C	2 minutes
Kept at 4°C		

[00371] Extend strands to create non-Uracil library: The chemical mixture was prepared for strand extension reaction with the components and volumes shown in the Table 5 below.

Table 5

Chemicals	Volume (μL)
Y1 chemicals	1
Y2 yeast	21
Total volume	22

[00372] Right at the end of attaching adapter 1 process, 22 μL of the extension chemical mixture was added. This mixture was mixed well by suction-release 10 times and incubated in a heat cycler with the program parameters as shown in the Table 6 below.

Table 6

Element	Temperature	Time
1	98°C	1 minute
2	62°C	2 minutes
3	65°C	5 minutes
Kept at 4°C		

[00373] Purifying the product after strand extension: 50.4 μL of KAPA magnetic beads were added into the tube containing the strand extended product, mixed well by suction-release 10 times and incubated at room temperature for 5 minutes. The sample tube was placed on a magnetic tray to capture magnetic beads until the solution cleared, and then the supernatant was discarded. 200 μL of 80% alcohol solution was added, incubated for 30 seconds and the supernatant was discarded. Add 200 μL of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. The magnetic beads were left to dry naturally for 1 to 3 minutes but without letting them dry too much. The tube from the magnetic tray was removed and 7.5 μL

were added of low TE. A magnetic bead suspension was created by suction-release 10 times and incubated at room temperature for 5 minutes. The tube containing the amplified product was placed on the magnetic tray to capture the magnetic beads, until the solution became clear, then the supernatant was transferred into a new 0.2 ml tube to prepare for the next element.

[00374] Connecting and attaching the 2nd adapter: The chemical mixture for the coupling reaction and attaching the 2nd adapter with the components and volumes are shown in the Table 7 below.

Table 7

Chemicals	Volume (μL)
B1 buffer	1.5
B2 chemicals	5
B3 yeast	1
Total volume	7.5

[00375] The connection of the 2nd adapter involved the following: Add 7.5 μL of the above chemical mixture to 7.5 μL of the cfDNA product purified in the previous element. Mix this mixture well by suction-release 10 times. Incubate this mixture in a heat cycler at 25°C for 15 minutes. To purify the product after connecting and attaching the 2nd adapter, add 18 μL of KAPA magnetic beads into the tube containing the amplified product. Mix well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Add another 200 μL of 80% alcohol solution into the sample tube, incubate for 30 seconds and discard the supernatant. Add another 200 μL of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. Let the magnetic beads dry naturally for 1 to 3 minutes and avoid letting them too dry. Remove the tube from the magnetic tray, add another 10 μL of low TE. Create magnetic bead suspension by suction-release 10 times and incubate at room temperature for 5 minutes. Place the tube containing the amplified product on the magnetic tray to capture the magnetic beads, wait for the solution to clear and transfer the supernatant into a new 0.2 ml tube to prepare for the next element.

[00376] Amplify and attach indexes: The chemical mixture for amplification reaction was prepared and the index attachment including the components and volumes are shown in the Table 8 below.

Table 8

Chemicals	Volume (μL)
Low TE buffer	5
R1 buffer	5
R2 chemicals	2
R3 yeast	0.5
Total volume	12.5

[00377] The amplification and attachment of the indexes involved the following: Add 12.5 μL of the above chemical mixture into a sample tube containing 10 μL of the cfDNA product purified in the previous element. Add another 2.5 μL of different index primer pairs specified for each sample. Mix the mixture well by suction-release 10 times and place the sample tube containing the mixture in the heat cycler. The amplification program followed the parameters shown in Table 9 below.

Table 9

Element	Temperature ($^{\circ}\text{C}$)	Time (seconds)
1	98	30
2	98	10
3	60	30
4	68	60
Repeat 2 – 4 for 15 cycles		
Kept at 4°C		

[00378] After amplification, the purification of the product involved the following: add 20 μL of KAPA magnetic beads into the sample tube containing the above amplified product. Mix well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the

sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear, and discard the supernatant. Add another 200 μ L of 80% alcohol solution and incubate for 30 seconds, then discard the supernatant. Add another 200 μ L of 80% alcohol solution and incubate for 30 seconds, then discard the supernatant. Let the magnetic beads dry naturally for 1 to 3 minutes and avoid letting them too dry. Remove the tube from the magnetic tray and add 20 μ L of TE with less EDTA. Create magnetic bead suspension by suction-release 10 times and incubate at room temperature for 5 minutes. Place the tube containing the amplified product on the magnetic tray to capture the magnetic beads, wait for the solution to clear, and transfer the supernatant into a new 1.5 ml Eppendorf tube. Check concentration of cfDNA library after amplification using Quantus™ Fluorometer meter system.

[00379] *Fragmentation of the cfDNA library for variation analysis at 450 target sequence regions*

[00380] Hybrid capture was performed using xGEN® Lockdown reagent kit (1080584) combined with human DNA Cot reagents (1080769) and xGen Universal Blocker-TS key mixture (1075474) to increase the specificity of hybrid capture process. The process of hybrid capture included the following:

[00381] *Hybrid reaction:* 16 libraries of different samples were pooled together in 1 hybrid reaction with an input of 50ng for each sample. A chemical mixture was prepared for nonspecific site-locking reaction including the components shown in the Table 10 below.

Table 10

Component	Volume (μ L)
Human DNA Cot	5
xGen Universal Blocker-TS key mixture	2
Total	7

[00382] 7 μ L of the above key mixture were added into the sample tube containing the pooled libraries. The mixture was mixed and concentrated the sample on a concentrator at 1700 rpm, 65°C until the solution turns colloidal. The hybrid buffer mixture included the components shown in the Table 11 below.

Table 11

Component	Volume (μ L)
xGen 2X hybrid buffer	8.5
xGen hybrid enhancer	2.7
Target probe	4
Water	1.8
Total	17

[00383] The sample suspension was reconstituted with 17 μ L of the above hybrid buffer mixture. The solution was mixed and incubated at room temperature for 5 to 10 minutes. The entire sample was transferred into a 0.2 ml PCR tube, then placed it in a heat cycler and run the thermal process with the settings shown in the Table 12 below.

Table 12

Element	Temperature	Time
1	95°C	30 seconds
2	65°C	4 hours
Kept at 65°C		

[00384] The wash buffers were diluted and the probe capture reagent were prepared onto magnetic beads. The high-concentration stock buffers were defrosted and if the buffers have crystallized, incubated at 65°C until completely dissolved. The components were diluted according to the Table 13 below.

Table 13

Component	Water (μ L)	Buffer (μ L)	Total (μ L)	Storage
xGen 2X magnetic beads	250	250	500	Room temperature

wash buffer				
I xGen 10X wash buffer	270	30	300	Divide into 2 parts: at 65°C and room temperature
II xGen 10X wash buffer	180	20	200	Room temperature
III xGen 10X wash buffer	180	20	200	Room temperature
xGen 10X strong wash buffer	360	40	400	At 65°C

[00385] The reaction mixture was prepared for probe hybrid capture onto magnetic beads and included the components shown in the Table 14 below.

Table 14

Component	Volume (μL)
xGen 2X Hybridization Buffer	8.5
xGen Hybridization Buffer Enhancer	2.7
Nuclease-Free Water	5.8
Total	17

[00386] The washing of the streptavidin magnetic beads included the following: Bring Dynabeads M-270 Streptavidin magnetic beads from 4°C to room temperature at least 30 minutes before use. Create magnetic bead suspension using a shaker for 15 seconds. Aspirate 100 μL of magnetic beads into each 1.5ml non-stick tube. Add 100 μL of magnetic beads wash buffer into each tube. Create suspension by suction-release 10 times. Place the tube in a magnetic tray, wait until the magnetic beads separate from the supernatant (about 1 minute) and discard the supernatant, making sure that the magnetic beads remain in the tube. Remove the tube from the magnetic tray and perform the washing again with 100 μL of magnetic bead wash buffer. Reconstitute the magnetic bead suspension in 17 μL of the above capture reaction mixture solution. Mix well to ensure that the magnetic beads do not dry on the wall of the tube. Magnetic beads are ready for capture reaction.

[00387] *After hybridization the library capture followed the protocol as detailed herein:*

After incubation for 4 hours, end the hybridization program, remove the sample from the PCR machine. Transfer 17 μL of the above-suspended magnetic bead mixture into the tube containing the hybrid sample. Mix well by suction-release 10 times and incubate the sample tube in a heat cycler at 65°C for 45 minutes. Make sure the cap of the heat cycler is at 70°C. Every 15 minutes, gently create suspension to mix well the magnetic beads. After 45 minutes, remove the sample from the PCR machine and immediately proceed to the washing with annealing.

[00388] *The 65°C hot washing involved the following:* Use wash buffer I and strong wash solution that has been incubated at 65°C. Transfer 100 μL of wash buffer I into the sample tube and do suction-release 10 times without forming air bubbles. Place the tube on a magnetic tray for 1 minute. Collect the supernatant into a 1.5 ml non-stick tube, used for the flow through the library fragment collection. Remove the tube from the magnetic tray and add 200 μL of strong wash solution to the sample. Suction and release 10 times using a pipet without air bubbles and incubate the sample at 65°C for 5 minutes. Place the tube on a magnetic tray for 1 minute and discard the supernatant. Remove the tube from the magnetic tray and add 200 μL of strong wash solution to the sample tube. Suction and release 10 times using a pipet without air bubbles and incubate the sample at 65°C for 5 minutes. Place the tube on a magnetic tray for 1 minute.

[00389] *The room temperature washing involved the following:* Wash buffers I, II and III are placed at room temperature. Discard the supernatant and add another 200 μL of wash buffer I. Create suspension to mix the sample well and incubate for 2 minutes (alternately shake for 30 seconds, rest for 30 seconds). After incubation, quickly centrifuge the sample tube and place it on a magnetic tray for 1 minute. Discard the supernatant and add another 200 μL of wash buffer II. Create suspension to mix the sample well and incubate for 2 minutes (alternately shake for 30 seconds, rest for 30 seconds). After incubation, quickly centrifuge the sample tube and place it on a magnetic tray for 1 minute. Discard the supernatant and add 200 μL of wash buffer III. Create suspension to mix the sample well and incubate for 2 minutes (alternately shake for 30 seconds, rest for 30 seconds). After incubation, quickly centrifuge the sample tube and place it on a magnetic tray for 1 minute. Discard the supernatant and use a suitable aspirator to remove all residual solution, then remove the tube from the magnetic tray. Add another 20 μL of H_2O , magnetic bead suspension by suction-release 10 times. Magnetic beads in the form of suspension are used directly for the next element of the method.

[00390] *The Post-capture library amplification involved the following:* Prepare chemical mixture for amplification reaction (after capture) including the components shown in the Table 15 below.

Table 15

Component	Volume (μL)
KAPA HiFi HotStart 2X mixture	25
P5/P7 primer mixture	5
Total	30

[00391] Add 30 μL of chemical mixture to 20 μL of magnetic beads in the form of suspension in the previous element of the method. Mix the mixture well by suction-release 10 times. Place mixture tube in a heat cycler and run amplification program with the parameters shown in Table 16 below.

Table 16

Element	Temperature	Time
1	98°C	45 seconds
2	98°C	15 seconds
3	60°C	30 seconds
4	72°C	30 seconds
Repeat 2 – 4 for 14 cycles (*)		
5	72°C	60 seconds
Kept at 4°C		

[00392] Purifying the product after amplification: Place the tube containing the amplified product on the magnetic tray to capture the magnetic beads, wait for the solution to clear and transfer the supernatant into a tube containing 45 μL of KAPA magnetic beads. Mix the sample well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Add 200 μL of 80% alcohol solution and incubate for 30 seconds, then

discard the supernatant. Add another 200 μ L of 80% alcohol solution and incubate for 30 seconds, then discard the supernatant. Let the magnetic beads dry naturally for 1 to 3 minutes, avoid letting them too dry. Remove the tube from the magnetic tray and add 22 μ L of TE 0,1X. Create magnetic bead suspension by suction-release 10 times and incubate at room temperature for 5 minutes. Place the tube containing the amplified product on the magnetic tray to capture the magnetic beads, wait for the solution to clear and transfer the supernatant into a new 1.5 ml tube. Check concentration of cfDNA library after the amplification using Quantus™ Fluorometers meter system.

[00393] The collection of library fragments for analysis of genome-wide variation (“flow through” fragment) involved the following:

- Prepare chemicals, tools and equipment:
- Wash solution I (high salt concentration): NaCl 1M, Tris-HCl 10mM, Tween-20 0,05%.
- Wash solution II (low salt concentration): NaCl 15mM, Tris-HCl 10mM.
- Dynabeads® M-270 Streptavidin magnetic beads (Cat No. 11205D)
- Biotin-bound P5 Probe (12.5 μ M) (Integrated DNA Technologies-IDT)
- Biotin-bound P7 Probe (12.5 μ M) (Integrated DNA Technologies-IDT)
- Hybridization buffer
- Hybridization enhancer.
- KaPa Hifi HotStart Ready mixture (Cat No. KK2601)
- P5, P7 Primer mixture (Integrated DNA Technologies-IDT)
- Kapa Pure Beads magnetic beads (Cat No. KK8002)
- Sample concentrator (Thermo Fisher Scientific SpeedVac system)
- Magnetic 1.5 ml and 0.2 ml tube trays (magnetic trays)
- Vortexer.
- PCR heat cycler

[00394] *The concentration of library fragments involved the following:* Wash solution I sample containing the remaining cfDNA library fragments is evaporated on the sample concentrator system at 1700 rpm at 65°C. Attach P5/P7 probe to Dynabeads® M-270

Streptavidin magnetic beads. Add another 100 μ L of magnetic beads to a 1.5 ml Eppendorf tube. Place the tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Remove the tube from the magnetic tray, add 100 μ L of wash solution I into the tube. Mix well the mixture for 5 seconds on a vortexer. Place the tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Wash the magnetic beads again with wash solution I for 2 more times. Place the tube on a magnetic tray to capture magnetic beads, wait for the solution to clear, discard the supernatant. Add 16 μ L of H₂O into the tube containing washed magnetic beads, mix well and transfer to a 0.2 ml tube. Add 2 μ L of P5 probe and 2 μ L of P7 probe and mix well, incubate at room temperature for 15 minutes. Place the tube containing the mixture of magnetic beads fitted with P5/P7 probe on a magnetic tray to collect magnetic beads, wait for the solution to clear and discard the supernatant. Add 100 μ L of wash solution I and mix well the mixture for 5 seconds. Place the tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Wash the magnetic beads again with wash solution I for 2 more times. Place the tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Add the following components into the library tube (concentrate): 1.8 μ L of H₂O; 8.5 μ L of hybrid buffer and 2.7 μ L of hybrid enhancer. Incubate this mixture at room temperature for 10 minutes. Mix well by suction-release 10 times and transfer the entire mixture to a 0.2 ml tube. Place the tube in a heat cycler and incubate at 95°C for 10 minutes. Transfer the entire mixture to a tube containing the magnetic bead mixture fitted with P5/P7 probe. Mix well by suction-release 10 times and incubate at room temperature for 30 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Remove the sample tube from the magnetic tray, add 100 μ L of wash solution I into the tube. Mix the mixture well by suction-release 10-20 times. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Wash again with wash solution I for one more time. Then, add 100 μ L of wash solution II to the tube and mix the mixture well by suction-release 10-20 times. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Add 20 μ L of H₂O into the tube, suspend the magnetic bead evenly by suction-release 10 times. Magnetic beads in the form of suspension are used for the next element of the method.

[00395] *The amplification of DNA with KAPA HiFi DNA Polymerase yeast involved the following:* Transfer 3 μL of the mixture of magnetic beads in form of suspension to a 0.2 ml tube. Place the tube in a heat cycler and incubate at 65°C for 10 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Measure the concentration of cfDNA in the supernatant using Quantus™ Fluorometer meter system.

[00396] *The preparation of the library amplification reaction involved the following:* Add another 3 μL of H_2O ; 25 μL of KAPA HiFi HotStart Ready Mix and 5 μL of P5/P7 primer mixture into 17 μL of magnetic beads in the form of suspension. Mix the mixture well by suction-release 10-20 times. Place the sample in a heat cycler and run the heat program as shown in Table 17 below.

Table 17

Element	Temperature (°C)	Time (seconds)
1	98	45
2	98	15
3	60	30
4	72	30
Repeat 2 – 4 for 10 cycles (*)		
5	72	60
Kept at 4°C		

(*) number of cycles is adjusted depending on the library concentration before amplification and the amount of library required after the amplification.

[00397] *The purification of the product after amplified involved the following:* Place the tube containing the amplified product on the magnetic tray to capture magnetic beads, wait for the solution to clear, transfer the supernatant into a tube containing 45 μL of KAPA magnetic beads. Mix well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear and discard the supernatant. Add another 200 μL of 80% alcohol solution, incubate for 30 seconds

and discard the supernatant. Add another 200 μ L of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. Let the magnetic beads dry naturally for 1-3 minutes and avoid letting them too dry. Remove the tube from the magnetic tray and add 20 μ L of TE 0,1X. Mix well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the tube containing the amplified product on the magnetic tray to capture magnetic beads, wait for the solution to clear, transfer the supernatant into a new 1.5 ml Eppendorf tube. Check concentration of cfDNA library after the amplification using Quantus™ Fluorometer meter system.

[00398] *The procedure for library transformation and sequencing using MGI-DNBseq system involved the following:*

[00399] To be sequenced on a DNBseq system, the cfDNA library needed to be converted into DNA library spheres, the process is done with MGI Easy Universal library conversion reagent kit (1000004155). The specific protocol was as follows:

[00400] *Adapter conversion:* The libraries of each sample were mixed with equal amounts of DNA to form a mixture of pooled library. The pooled library was fitted with a suitable adapter for the MGI-DNBseq sequencing system through the AC-PCR reaction amplification. The reaction components included 25 μ L of AC-PCR amplification chemical mixture and 3 μ L of AC-PCR primer mixture. The PCR reaction was done in a heat cycler with thermal cycling as shown in the Table 18 below.

Table 18

Element	Temperature	Time
1	98°C	3 minutes
2	98°C	30 seconds
3	62°C	15 seconds
4	72°C	30 seconds
Repeat 2 – 4 for 5 cycles		
5	72°C	5 minutes
Kept at 4°C		

[00401] *After amplification, the purification of the product involved the following:* Add 60 μL of KAPA magnetic beads into the tube containing the amplified product. Mix well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear, and discard the supernatant. Add another 200 μL of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. Add another 200 μL of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. Let the magnetic beads dry naturally for 1-3 minutes, avoid letting them too dry. Remove the tube from the magnetic tray and add 30 μL of TE 0,1X. Create magnetic bead suspension by suction-release 10 times and incubate at room temperature for 5 minutes. Place the tube containing the amplified product on the magnetic tray to capture magnetic beads, wait for the solution to clear and transfer the supernatant into a new 1.5 ml Eppendorf tube. Check concentration of cfDNA library after the amplification using Quantus™ Fluorometer meter system.

[00402] Denaturation – separation: The library were denatured to separate into a single strand. Specifically, after AC-PCR, 1 pmol of product was denatured in a heat cycler at 95°C for 3 minutes and then placed on cold ice immediately to prevent regurgitation of single-stranded DNAs.

[00403] Cyclization reaction: The straight single-stranded DNA library was converted to cyclic form by a cyclization reaction. The reaction used 1 short single-stranded DNA fragment (splint Oligo) capable of complementary pairing with 2 adapters attached in the AC-PCR. This splint Oligo fragment acted as a splint to connect 2 ends of single-stranded DNA fragments. The reaction components included: 11.6 μL of splint buffer and 0.5 μL of ligation enzyme, done in a heat cycler at 37°C for 30 minutes and then immediately place the product on cold ice.

[00404] Reaction of cleavage of non-cyclic DNA library fragments: Non-cyclic single-stranded DNA library fragments were enzymatically chopped. The reaction used 4 μL of a mixture of cutting enzymes (including 1.4 μL of cutting buffer and 2.6 μL of cutting yeast). The reaction was incubated at 37°C for 30 minutes using a heat cycler. After being chopped, DNA fragments were removed using the purification process.

[00405] After fragmentation, the purification of DNA product involved the following: Add 170 μL of KAPA magnetic beads into the tube containing chopped product. Mix well by suction-

release 10 times and incubate at room temperature for 5 minutes. Place the sample tube on a magnetic tray to capture magnetic beads, wait for the solution to clear, discard the supernatant. Add another 500 μL of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. Add another 500 μL of 80% alcohol solution, incubate for 30 seconds and discard the supernatant. Let the magnetic beads dry naturally for 1-3 minutes, avoid letting them too dry. Remove the tube from the magnetic tray and add 27 μL of TE 0,1X. Mix well by suction-release 10 times and incubate at room temperature for 5 minutes. Place the tube on the magnetic tray to capture magnetic beads, wait for the solution to clear, transfer the supernatant into a new 1.5 ml Eppendorf tube. Check the concentration of cfDNA library after fragmentation using Quantus Fluorometer meter system.

[00406] DNA sphere (DNB) generation – circle amplification reaction: A mixture of 20 μL of App-A buffer produced DNB and 60fmol (equivalent to 9.9ng) of cyclic DNA library. The mixture was placed in a heat cycler using program parameters as shown in Table 19 below.

Table 19

Element	Temperature ($^{\circ}\text{C}$)	Time (minutes)
1	95	1
2	65	1
3	40	1
Kept at 4°C		

[00407] 44 μL of mixture for generation of DNB 2 were added to the element 1 product (kept on cold ice). The mixture was placed in a heat cycler using program parameters as shown in the Table 20 below.

Table 20

Element	Temperature ($^{\circ}\text{C}$)	Time (minutes)
1	30	25
2	Kept at 4°C	

[00408] As soon as the temperature reached 4°C , 20 μL of Stop DNB reaction buffer were added. The DNB library mixture was mixed well by suction-release gently with a wide-mouth

straw to avoid breaking DNBs. The amount of formed DNB was quantified using the QuBit system.

[00409] Load DNB onto a flowcell: The DNB mixture was mixed with 8 μ L of DNB II loading buffer and 0.25 μ L of DNB II LC yeast mixture. The mixture was mixed well by suction-release using a wide-mouth straw. The flowcell was fitted to the sample feeder. Using a wide-mouth straw, 30 μ L of the DNB library mixture was transferred to the sample loading position on the feeder. The DNB library solution automatically flew into the flowcell without being injected.

[00410] Preparation the sequencing reagent cartridge: After the sequencing reagent cartridge was defrosted, it was stirred well and wiped dry the outer shell. A pointed tip was used to puncture the membrane of the wells marked with 1, 2, 3, 4, 6, 7 and 8 on the sequencing reagent cartridge. The sample was loaded according to the Table 21 below.

Table 21

Well	Absorb the liquid that is already inside	Add to the solution mixture
1		1.8 ml of dNTPs mixture 1.8 ml of sequencing yeast mixture
2		1.8 ml of dNTPs mixture 1.8 ml of sequencing yeast mixture
3	App-A insertion primer 1	2.2 ml of App-A insertion primer 1 (1 μ M)
4		2.9 ml of App-A index primer 3 (1 μ M)
6	App-A insertion primer 2	2.9 ml of App-A index primer 2 (1 μ M)
7	App-A MDA primer	3.1 ml of App-A MDA primer (1 μ M)
8	App-A index primer 2	3.3 ml of App-A insertion primer 2 (1 μ M)

[00411] The sequencing reagent cartridge and flowcell were placed into MGISEQ-2000 sequencer, the required information was entered and the sequencing process was started.

[00412] *Example 2: Element 2 - Analyze different variation patterns of cfDNA.*

[00413] 2.1 Analysis of methylation variation at 450 target regions (containing 18,000 CpG sites)

[00414] Raw data was quality checked using FastQC tool (Babraham Institute, version 0.11.9). Poor quality data and adapter sequences were removed using a trimmomatic tool. Read sequences were aligned with the standard genome and analyzed to determine methylation percentage using the Bismark aligner tool (Babraham Institute, version 16.0.2).

[00415] Regions with different methylation percentages between cancer and healthy groups (called DMR - Differentially Methylated Regions) were determined by the methylation percentage per CpG determined using the following formula:

$$\text{Methylation percentage} = \frac{N_{C,i}}{N_{C,i} + N_{T,i}} \times 100\%$$

where:

i: The i^{th} CpG site in the sequence region of interest,

$N_{T,i}$: Number of T nucleotides observed at the i^{th} CpG site, and

$N_{C,i}$: Number of C nucleotides observed at the i^{th} CpG site.

[00416] The regions with different methylation percentage between the cancer group and the healthy group were determined. Specifically, the percentage of methylation of the healthy group and the cancer group were compared on each corresponding CpG site by the Wilcoxon rank sum test (Mann Whitney U test), in order to identify regions with differences (statistically significant) on the methylation density of CpG. The Wilcoxon rank sum test is suitable when comparing multiple variables simultaneously between 2 groups of independent samples and variables that are not normally distributed (non-parametric test). In addition, the p-value of the statistical test was corrected using the Benjamini Hochberg method to avoid the false-positive situation encountered when the number of variables to be compared was much larger than the number of analyzed samples. Regions identified with different percentages of methylation between cancer and healthy groups when p-value was less than 0.05 (p-value < 0.05).

[00417] The methylation fold change was determined between the cancer group and the healthy group. Specifically, the percentage of methylation (between cancer and healthy groups) on each respective CpG site was used to determine how many times the methylation fold change had changed. The methylation fold change was corrected by taking the log to base 2 ($|\log_2|$) of

the absolute value of the above percentage. If this value was greater than 1, the methylation fold change had changed more than 2 times between the cancer group and the healthy group. With some of the results depicted in the figures:

[00418] **Figure 4** illustrates 353 sequence regions out of 450 target sequence regions surveyed with statistically significant differences in methylation density (p -value < 0.05) between the liver cancer group and the healthy group specified when performing the SPOT-MAS test procedure according to the present invention (as described above). Specifically, in each survey region, the percentage of methylation was compared between the cancer group and the healthy people using the Wilcoxon rank sum test with correction using the Benjamini-hochberg method. It was noted that 353 out of 450 target sequence regions had differences in methylation density (p -value less than 0.05) (including dots above the solid line with value $-\log_{10}(p\text{-value}) > 1.30$). In these 353 regions, there were 154 regions with methylation density in liver cancer patients being 2 times that of healthy people (including large dots, located to the right of the dashed line with $\log_2$ value (fold ratio) > 1).

[00419] **Figure 5** is a heatmap illustrating the clustering according to the methylation density at target sequence regions between liver cancer patients and healthy subjects obtained after performing the SPOT-MAS test procedure according to the present invention. The lightness on the heatmap represented the degree of change in methylation density (with a scale of 0 to 100, the darker the color indicates the higher the methylation density). Specifically as shown in **Figure 5**, from top to bottom, the regions of DNA sequences were grouped according to the descending order of the methylation density. From left to right was the list of analyzed samples, with the left side being the group of liver cancer patients, the right side being the group of healthy people. The results from the heatmap showed that the samples in the liver cancer group with multiple target sequence regions had increased methylation density compared with the healthy control group.

[00420] *2.2 Methylation density change analysis on 22 chromosomes*

[00421] The quality of the sequencing data of the remaining flow through the library fragment was assessed using MultiQC software (<https://multiqc.info/>). Poor quality data and adapter sequences were removed using a trimmomatic tool. Read sequences were aligned against the human reference genome sequence (version hg19) using the BSAigner software in the

(Methyl pipe analysis package, DOI: 10.1371/journal.pone.0100360). Check parameters: (1) proportion of reads was aligned against the reference genomic sequence in total mappability, (2) depth of sequencing, (3) sequencing coverage of all samples.

[00422] Genome-wide methylation variation was determined as follows. The standard human genome was uniformly subdivided into non-duplicating fragments (bin) of 1 megabase (one million nucleotides) long. Analysis for methylation variation was performed on each bin. The methylation density (MD) per bin was calculated using the following formula:

$$MD = \frac{\sum mC}{(\sum mC + \sum T)} \times 100$$

where $\sum mC$ is the total number of methylated C nucleotides and $\sum T$ is the total number of nucleotides.

[00423] Bins with variation in methylation state were identified. Sequencing data from 19 healthy subjects were randomly selected to determine the reference MD value for each bin. Variation in values of methylation density in each bin was evaluated based on the “Z score” value using the following formula:

$$\text{Zscore} = \frac{\text{MD in survey bin} - \text{Mean MD in corresponding survey bin of the reference group}}{\text{Standard deviation MD in corresponding bin in the reference group.}}$$

- If Zscore < -3, that bin region was less methylated than the bin in the reference group.
- If -3 < Zscore < 3, methylation in that bin region was equivalent to the bin in the reference group.
- If Zscore > 3, that bin region was more methylated than the bin in the reference group.

[00424] **Figure 6** illustrates the results of analysis of mean values of methylation density on all survey bins belonging to 22 chromosomes of patients with colorectal cancer (CRC) and a group of healthy people who underwent SPOT-MAS test procedure according to the present disclosure (as described above). Specifically, the solid curve represents the distribution of methylation density values of all the survey bins belonging to 22 chromosomes of the group of patients with colorectal cancer. The dotted curve depicts the distribution of methylation density values of all the survey bins of the 22 chromosomes of the healthy group. It can be seen that the distribution of methylation density values in the cancer group was skewed to the left (the tendency to decrease methylation) compared with the healthy group.

[00425] **Figure 7** shows a graph illustrating the decrease in methylation on all the 'bin' regions of the 22 chromosomes of the CRC group compared with the healthy group who underwent the SPOT-MAS test according to the invention (as described above). Specifically, the vertical axis represents the values of methylation density and the median represents the list of 22 chromosomes examined in healthy people (top chart) and CRC patients (bottom chart). The methylation density values of each bin are indicated by dots. When setting the benchmark (dotted line at the values of methylation density reaching 60%), it can be seen that the methylation density on some bins in the group of people with colorectal cancer was lower than in the healthy group.

[00426] **Figure 8** shows a graph illustrating the percentage of bins that are determined to be less methylated ($Z\text{score} < 3$ according to the analysis described above) between the group of colorectal cancer patients and the group of healthy people who underwent the SPOT-MAS test procedure according to the invention. Accordingly, the vertical axis represents the percentage of bins that were less methylated, and the horizontal axis is the list of analyzed samples (with cancer samples being bars with slashes, and healthy samples being bars without slashes). The percentage of bins less methylated in the total number of bins surveyed was calculated for each sample. The results showed that, 5/15 (ZL10071, ZL10335, ZL10516, ZL0819, ZL12643) colorectal cancer samples had a higher percentage of less methylated bins than the healthy group.

[00427] *2.3 DNA copy number abnormalities analysis on 22 chromosomes*

[00428] Sequencing data of the remaining flow through library fragments was used for genome-wide DNA copy number abnormalities analysis. Data quality was checked using FastQC software. Poor quality data and adapter sequences were removed using a trimmomatic tool. Read sequences were aligned against the human reference genome sequence (version hg19) using the BSAigner software in the (Methyl pipe analysis package, DOI: 10.1371/journal.pone.0100360).

[00429] Check parameters: (1) proportion of reads was aligned against the reference genomic sequence in total mappability, (2) depth of sequencing, (3) sequencing coverage of all samples.

[00430] Identifying DNA copy number abnormalities on 22 chromosomes

[00431] The standard human genome was uniformly subdivided into non-duplicating fragments (bin) of 1 megabase (one million nucleotides) long. Copy number abnormalities analysis was performed on each bin.

[00432] The number of copies of DNA in the bins was determined. Differences in the number of reads between bins can occur due to the influence of the bin region containing many G and C nucleotides (GC-bias) or the presence of repeat sequence regions (tandem repeat). Therefore, after alignment, the number of reads in each bin was corrected using the QDNaseq tool (DOI: 10.1101/gr.175141.114). The median copy number of all bins was calculated after correction. The degree of variation in the number of copies per bin was determined by taking the log to base 2 ($\log_2$) of the absolute value of the ratio of the number of reads in that bin to the median of the reads of all bins. If this value was greater than 1, then the degree of variation was more than 2 times between the investigated bin and the whole genome.

[00433] The proportion of bins with DNA copy number abnormalities between the cancer group and healthy people was determined. Sequencing data from 19 healthy subjects were randomly selected to determine the average number of reads for each bin. Variation of gene copy number in each bin was evaluated based on the “Z score” value using the following formula:

$$\text{Zscore} = \frac{\text{Number of reads in survey bin} - \text{Average number of reads in corresponding bin of standard reference group}}{\text{Standard deviation of the number of reads in the corresponding bin in the reference group.}}$$

[00434] - If $\text{Zscore} < -3$, that bin region had fewer copies than the bin in reference group.

[00435] - If $-3 < \text{Zscore} < 3$, the number of copies that bin region had is equivalent to the bin in the references group.

[00436] - If $\text{Zscore} > 3$, that bin region had more copies than the bin in reference group.

[00437] The obtained test results are shown in **Figures 9 and 10**.

[00438] **Figure 9** is a chart illustrating DNA copy number variations on all 22 chromosomes of the group of colorectal cancer patients and the group of healthy people who underwent the SPOT-MAS test procedure according to the disclosure, as described above. Specifically, the vertical axis represents the log to base 2 value of the number of DNA copies and the horizontal axis represents the list of chromosomes examined in healthy people (top chart) and CRC patients

(bottom chart). The chromosome outlined by the dashed line is the chromosomes with DNA copy number abnormality. This result showed that colorectal cancer patients had copy number abnormalities in peripheral blood compared with the group of people with colorectal cancer.

[00439] **Figure 10** is a chart illustrating the percentage of the bins with gene copy number abnormalities in the total number of surveyed bins between the CRC group and the healthy group who underwent the SPOT-MAS test procedure according to the disclosure, as described above. Accordingly, the vertical axis represents the percentage of the bins with copy number abnormalities, and the horizontal axis is the list of analyzed samples (with cancer samples being spotted bars, and healthy samples being non-spotted bars). The percentage of bins with abnormalities (when absolute value of Zscore ($|Zscore|$) > 3) in the surveyed bins was calculated for each sample. The results show that, 6/15 colorectal cancer samples (ZL10071, ZL10516, ZL10335, ZL10672, ZL0819 and ZL12643) that were surveyed had a higher percentage of bins with abnormalities than that of the healthy group. This result demonstrated instability in the DNA copy number in peripheral blood of the colorectal cancer group.

[00440] *2.4 Analysis of variation in cfDNA size*

[00441] Sequencing data of the remaining flow through library fragments was used to analyze variation in cfDNA size. Data quality was checked using MultiQC software (<https://multiqc.info/>). Poor quality data and adapter sequences were removed using a trimmomatic tool.

[00442] Read sequences were aligned against the human reference genome sequence (version hg19) using the BSAigner software in the (Methyl pipe analysis package, DOI: 10.1371/journal.pone.0100360). The parameters: (1) proportion of reads was aligned against the reference genomic sequence in total mappability, (2) depth of sequencing, and (3) sequencing coverage were checked for all samples.

[00443] Variation in cfDNA size was determined as follows.

[00444] The standard human genome was uniformly subdivided into non-duplicating fragments (bin) of 5 megabase (5 million nucleotides) long. Size variation analysis was performed on each bin.

[00445] After alignment, the length of each cfDNA fragment was calculated using software (bsalign). The size of cfDNA fragment was calculated based on the distance between the starting point of the Watson reading in the standard genome and the end point of the reading in the opposite direction (Crick).

[00446] The size distribution ratio of cfDNA fragments of cancer and healthy samples in the range of 0 to 250 nucleotides was determined.

[00447] **Figure 11** is a histogram showing the size distribution of cfDNA fragments in colorectal cancer samples and healthy subjects who underwent the SPOT-MAS test procedure according to the disclosure, as described above. Specifically, the horizontal axis of the graph represents the scale of cfDNA size (from 0 to 250 nucleotides) and the vertical axis represents the density of cfDNA fragmentation in the blood. The black dashed line represents the cfDNA size distribution in the blood of CRC patients, while the gray solid line represents the cfDNA size distribution in the blood of the healthy people. The results showed that the density of cfDNA in colorectal cancer samples with cfDNA size < 150bp was higher than in healthy samples. This result suggested that a person's condition can be represented by the distribution of cfDNA lengths found in that person's plasma.

[00448] Fragment ratio (RF) per bin was calculated using the following formula:

$$RF = \frac{(P \leq 150bp)}{(P > 150bp)} \times 100$$

where $P \leq 150bp$ means length of reads is 150 nucleotides or less and $P > 150bp$ means length of reads is over 150 nucleotides.

[00449] **Figure 12** is a histogram showing the RF ratio variation across all 22 chromosomes as determined by the SPOT-MAS test procedure according to the disclosure, as described above. Specifically, the vertical axis represents the RF ratio and the median represents the list of surveyed chromosomes. Within each region (bin) on the chromosome, the RF ratio is represented as a dot. When comparing patients with colorectal cancer (left graph) and healthy people (right graph), the RF ratio was higher in the colorectal cancer group than in healthy people on the entire surveyed chromosome. This result established that there was a difference in cfDNA size fluctuations in peripheral blood that can help distinguish between cancer and healthy people.

[00450] *Example 3: Element 3 - Building a machine learning model that predicts samples carrying cancer and tumor origins.*

[00451] The analytical data as provided above in Example 2, sections 2.1, 2.2, 2.3 and 2.4, established the basis of quantitative data of four different attributes for each cfDNA sample: methylation density attribute of 450 target regions (2.1); methylation density attribute of bins in 22 chromosomes (2.2); DNA copy number attribute of bins in 22 chromosomes (2.3); cfDNA size-specific ratio attribute of bins in 22 chromosomes (2.4). The machine learning model was built for each individual group of attributes as well as the combination of all four attribute groups. The effectiveness of this model was evaluated based on its ability to classify 2 groups of samples as cancer and healthy people or between malignant and benign tumors.

[00452] The model applied in the SPOT-MAS test procedure was a stacking model of individual attributes analyzed in element 2. The results of building the accuracy of the model are depicted in **Figure 13**.

[00453] **Figure 13** is a chart illustrating the results of evaluating the effectiveness of blood sample classification of 4 groups of patients with liver cancer, lung cancer, colorectal cancer, and breast cancer with blood samples of healthy people who underwent SPOT-MAS test procedure according to the invention. Specifically, in the graph, the vertical axis represents the test's sensitivity and the horizontal axis represents the [1-specificity] value (or false-positive rate) of the test. Corresponding to a pair of sensitivity and [1-specificity] values, a point will be plotted on the graph. The changes in value of [1-specificity] from 0 to 1 will create a receiver operating curve (ROC). The area bounded by the ROC curve and the right and bottom sides of the graph is called the area under the ROC curve (or AUC). The larger the area, the higher the accuracy of the model. **Figure 13** showed that the AUC area is 0.94 (with confidence intervals ranging from 0.92 to 0.95), which means that the model's accuracy was up to 94% when classifying cancer samples and healthy samples.

[00454] After selecting the model with the best performance, the effectiveness of the selected model was evaluated on the model evaluation dataset. Similar to the model training, the specificity, sensitivity, accuracy and AUC values of the model were determined on the model evaluation dataset. The model has the best performance when these values were the highest and

were equivalent to the values obtained in the model training. The model's evaluation results are described in Table 22 and **Figure 14**.

Table 22

	Average	Confidence interval
Sensitivity (%)	70.00	66.90 – 73.10
Specificity (%)	89.67	87.18 – 92.16

[00455] The results when applying the model on the leave-out test set show that the sensitivity of the test reaches 70% (with confidence intervals ranging from 66.90% – 73.10%) and the specificity reaches 89.67% (with confidence intervals ranging from 87.18% to 92.16%).

[00456] **Figure 14** is a diagram showing the test results of blood samples from patients with liver cancer, lung cancer, colorectal cancer, and breast cancer using the SPOT-MAS test procedure according to the invention. Specifically, the vertical axis represents the probability (likelihood) of cancer prediction of the analyzed sample, and the horizontal axis is the list of surveyed cancers. The classification threshold value from the algorithm was 0.5 (solid line). The samples above the classification line were predicted by the model as cancerous and below this line were considered noncancerous. The results showed that the model was able to correctly predict 13/16 liver cancer samples, 9/21 colorectal cancer samples, 6/8 lung cancer samples and 3/22 breast cancer samples. In the group of healthy people, the model only wrongly predicted 1 case of cancer in a total of 36 surveyed samples. This result demonstrated that the disclosed SPOT-MAS classification model achieved different detection efficiency for different cancer groups. The model delivers good results for the group of healthy, liver cancer and lung cancer samples while the effectiveness is lower for the group of colorectal cancer and especially breast cancer samples.

[00457] cfDNA released from different organs have variations in epigenetic marks including the methylation, fragment length and motif-end profiles that can differentiate one cancer type from other cancer types. To determine the tumor tissue origin, a deep neural networks (DNN) model was built from such epigenetic signatures (**Figure 15**) as inputs. Structural for deep neural networks model was based on the multi-layer feedforward artificial neural network that was

trained with stochastic gradient descent using back-propagation. A random grid search in H2O platform was used to select the hyperparameter for of the deep neural networks. The model was built from epigenetic signatures such as GC methylation, fragment length and motif end. The hyperparameters included for instance (1) three hidden layer with 60 nodes in a layer; (2) activation function: Rectifier With Dropout; (3) Input layer dropout ratio: 0.01; (4) Loss function: Cross Entropy; (5) Rate annealing: 1e-06; (6) L1 regularization: 0; (7) L2 regularization: 0.

[00458] The disclosed DNN model returned probability scores of five (5) cancer types (breast cancer, gastric cancer, colorectal cancer, liver cancer and lung cancer) and probability scores of unknown cancer. The DNN model had 3 hidden layers and 60 nodes in each layer.

[00459] The performance of deep neural networks with hyperparameter was tested using leave-one-out cross validation (train in (n-1) sample of data, leave one sample to test the model). The result for the leave-one-out cross validation was shown in **Figure 16**. The model achieved a mean accuracy for five (5) cancer types of 0.69 (95% CI: 0.68-0.76). Of the five cancer types, liver cancer can be effectively differentiated from others with the highest accuracy of 0.93 while breast cancer showed lowest accuracy of 0.57. The accuracy for identifying colorectal, gastric and lung cancer were of 0.66, 0.66 and 0.65, respectively.

[00460] Example 4: Effectiveness of the systems and methods of the present disclosure.

[00461] Due to the combination of simultaneously identifying four attributes carrying characteristic variations occurring in the entire tumor genome, the SPOT-MAS test procedure according to the systems and methods of the present disclosure provides higher accuracy (sensitivity and specificity) than published tests that rely solely on one or two attributes. Therefore, the SPOT-MAS test is effective in detecting benign tumor DNA in the following cases:

- Early stage cancer with low tumor cfDNA level in the blood.
- Certain types of cancer tend to release less tumor cfDNA.
- Tumor recurrence after treatment.

[00462] Using a single cfDNA library preparation procedure (bisulfite treatment) for simultaneous analysis of four tumor DNA markers also helped reducing the cost of the disclosed SPOT-MAS test as compared with similar tests that need to take blood samples and multiple

independent cfDNA processing procedures. Therefore, the SPOT-MAS test allow increasing the patient's chance of accessing a cancer screening test.

[00463] The disclosures of each and every patent, patent application, and publication cited herein are hereby incorporated herein by reference in their entirety.

[00464] While this disclosure was provided with reference to specific embodiments, it is apparent that other embodiments and variations of this disclosure may be devised by others skilled in the art without departing from the true spirit and scope of the disclosure. The appended claims are intended to be construed to include all such embodiments and equivalent variations.

Table 23: List of target sequence regions of interest

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1	1163599	1163832	C1orf15 9_TTL1 0	CCGAGAGGGGTACGTTCTTGCCGCTACCTGACAGCAGGCCTTCT AGAAAGTTCTCTCCAGAAGCAGCCACCGCCGTCTGAGGCACTTTG TGCGGAGACGGGAAGCTGTCGCCTCAGAGGTGGGTGCGTAGAAG GGTTTGCCGGGTGCGAGGATGACCGCGTCTCCCTGGGCTCTGG AGTCTGCGGTGGGAAGGGCTTGTTTCAGCACCCCTCTGGTCAGAG GCCGGCCG	1
chr1	2443865	2444245	PEX10_ PLCH2	CCGCTGACTGCGCCTCCCGGCCGACGCCCCGCCCGCCGCCCT CGCTGCCCTCCTGTCAGCCGCCACGGAGACAATGGACGCGGGAGC CGCCCCGAGAACACAGTAGGTGCCGCTCTGCCGCTGCGCCGCT GCCAACCGGGATGCGCGGTGGACGCGCGGGGCGCCGACAGCCC TGGTGC GGCTCGGGGCTGAGCCGCTGGGCTTCAGACTCGGGAGC GGAGGCTCGGATCGCGGTGGCACGGGCAGGGGTGCGGGCGCGG GACTGTGGGCGGGACGGGCGGAGCGGTCTTGAGCTCTCCGATG GCCTCAGGTGCGGGGTGAGGGATCTGGGGGCCGCCCTCGGCAA ACTTTCCTTCCCCGGGCTTCTGCG	2
chr1	2789325	2789609	ACTRT2 _MMEL 1	CCGACTGCGGCCCGGTGATGGAAGCAGCGGAGCTAGACCTGCC TCGGGTGCTTTGGGAAGTCACCGCCACTGCCTCCGTTTATTCTTT GTAAAATAGGAGGAAACACATCCGTCGTACCTCGAAGGAGACCC GCAGGAAGCAGCGGCCCCAGCGTGCCCGGGCGGGTCTCACCCCT CTGCGTGGTGGGGCCGCCCGTCTCTGCGGCTCCCTCCGGCCCTG CGCTCTGGACGGCCCGCGCGTGGAGATCGCTGCAGCATCCACG GGCCTCCTCCCG	3
chr1	2789723	2790126	ACTRT2 _MMEL 1	ACGCGTGCCCCGCCAGCACCCGACAGTCCCCGGCCGCACAGCG CGCGCACACAGCCCCCGGGTGCGGCGCCCCCTGCTCCACTACCGT CTGGAAGTCTCTCATGGCGCGCCGCGAGTCCCCGGCGTGACGCG GCAGAAGCCGCGCAGAGCCAGGAGAGGTGCGCGGTCCCTGCAC GGTCCCCCGCAGGGTCTCAGGGCGTGCGGGCCGACAGAGGCT CGCACATGGCCCCGGGCGCCCGCGTCCCTGCCAGCAGCAGGC ACTCGGCGAGGCGGGCTCCAGCGCGGGTGCGCCCGGCGAGGC GCCAGGCGAAGCAGGACGCGGAGCGCGGGTTACGGGGGCCAG CAGCTCGCGGCCCGCATCTCGCGGAGCTCAGGGTGGTCCGCAC G	4
chr1	6064999	6065249	KCNAB2 _CHD5	CCGCGTTTCCTTCTTGGGCCATCTGTGCATAACCATCAAGACGCA GTGGCTTCTTACATTTCTGGTGATGTTGCTTCTCCATGTGCCAATC CCCCAGCGGATACCCCACTCTCCGGAGGGAGAACCCCAAGCAGGT GCCGCTGGGCATGCGCCAGGGAGGCTGTGACCGGAGCAAGCACT GCCTTGCTTGGAGCTGGCTGCCTACAAGCTCAGACATCCAGCCCGC AGAGTCCACCCTGGCTGCAGCG	5
chr1	7668718	7668819	VAMP3 _CAMT A1	GCGGCGGTTTCCATGGAGAAGGTCTGATGTTCTCCAGTAATTTCT GCAGTTCTTTGTTCCCGGCAGCAGCCCCAGCCTCATGCTAGCAGCT GTTGATTGCG	6
chr1	7668966	7668971	VAMP3 _CAMT A1	CCGCCG	7
chr1	9317873	9318231	SLC25A 33_SPS B1	GCGGGTCGCTTTGGTGGGAGTTTCTTGCTTCTTGGCACACCATTC GCTCCGCGAGTTTGTTAAGGGCCCTGTGTGCCAGGCTCGGCCGA GCATCTGTGGAACAGAGGAAGCTGGGTGGACAGTCGAGGTTTG	8

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1	15345942	1534624 6	EFHD2_ TMEM5 1	GTGACGTGCCAGGTGGGGAGAGGAAGCAGCTGCACTCATTCCCCT TTCCGGGCAGGTTGGGGAAACGCAGCGATTGTTCTGGGAAGCTGC AGCTTAGGGAGAGATGACGTTCCCTGTGCCCCAGTGAGGGTGGGG CCCTGGGGTCTGGGCTGACAGCAGGCAGTGGGGGAAGGTGGGTG TGGGCACCCGGAGGCCCATGATGCCCCAGATCCTCCACCACG CCGTCCAGCATGGATGCCTCAGGCCGACAGAAAGTTTTCCCTTTA GGTTGAGTTGTGTCAAACCTTACGCCCCGAGGAGTATCAGTCCT CCGCCCTCCCTGCGCTCCACAAGATACTACTTCTCTTCCACA TGATGACTCAGATGTGTGAAAACAGGGGCGCCGACCCCTGTGTCT GCTCTCCCCGGGCCAAGCGCCCTTGTCTCAGGTCCCTCACAG GACTAGAGCCTGGCCCTGGCTGCCTCCTGTGGCCTGTGCTGCTCTC CAGAAGTCACAGACTGGTAGCTCAGCG	9
chr1	29259675	2926013 4	PTPRU	GCGCACAGCGTCCCGGCCCTCCCTAGCTCTGCTCTGCGCTTCTTG GGTCCCCATTCCCCAGGTTAGAGCGCGGCTCCAGGAACCTATGT CCGCGCGGTGTAGTAGGGACGGCTAAATGGGGCCCGGTGAGAG CGAGATCGGGACCCCTCGCTCGAGGCGCCCTGACCCCTCACTC TCTTCCCTGCAGCGGCAGAGCGGGGCGTGGTGCCGCGGCGGGG GTGCGGCACATCAGCCACCGCGCTTCTGGCCACTTTCGCGTGG CTGCCGTGAGCCGCGCCGAGCAGGACCTGTACCGCTGTGTGCCA GGCCCCGCGCGGCGCGGGCTCTAACTTCGCGGAGCTCATGCTC AAGGGTCAGCTGGTGACGCCGGGGAGCGCCGGGACCTCACCTC GAGGGGCGGGGCCGCGACGGGGGCGGGCTGCCCCGGGGCG TGGCCG	10
chr1	37475577	3747579 6	ZC3H12 A_MEA F6	CCGTACAGGGCACCCAAAGCCGGGTCAAGAGCTGGCCGCTGAGGA GGCCTCGGCCCTGGAAGTGCAGATGAAGGTGGACTTCTCCGGAA GCTGGGCTATTCATCCACGGAGATCCACAGCGTCTGCAGAAGCTG GGCGTCCAGGCAGACACCAACACGGTGCTGGGTGAGCTGGTGAAA CACGGGACAGCCACCGAGCGGGAGCGCCAGACCTCACCG	11
chr1	43550045	4355025 5	KDM4A _PTPRF	GCGGGTGGAGGTGGATTGGAGGGAAAGCGGAGGGCGAGGCCCTGG TTGAGGGGCGGGGCTGCTGTCTGGTCCCCGGGCTGCCTTGGG CCAGCTTGGCCTAGTCTGTTGGGTGGGCGGCGAGGGTGACAGCTC CTCTCCAGCTCCAAGGGAGGGGAGTTGTTCTGCCTCTCGATAGC CCCAGGCCTTGGGCACAGCCAGCCTCCACG	12
chr1	47708297	4770848 4	FOXO2_ TRABD2 B	TCGCGGGCAAAGATCCGATGAGAGAGAGGCAGAGAAAATGAGAG GCAGAGACAGAGGCAAAGGCACAGCGAGACACCGGGGAAACGG GGAAGCAGGTGAGAGAGGAAGAGAGAGACAGGCCGGAAGAGAC TGTGCCAGGAGCCTGGACAAGGGATGCCGTGCCAGCAGCCTGG ACAAGGGATGCCG	13
chr1	50416318	5041648 3	DMRTA 2_ELAV L4	GCGGTGGGGCAGAGGACGGGGATGAGGCGGCGGGAACCGCCCTA CGAGGAGAGGCTGGGAGGCTCCGAAAACCTGGGGTAGGGGAGC GCACCGGGGCTTTAGAGGGCGCAGCGGCCAAGGGCAAGAAAGTT TACACTCCAGAAAGCTTCCGCACGCTTCTCCCG	14
chr1	50418877	5041947 8	DMRTA 2_ELAV L4	CCGCGGAGTAGGCCAGGCGCAGGGGGCTGAGGCCGAGCGGCGCG CCCAGCGGGTAGGCGCCCGCTCGGCACCGAAGTGAAGTGGCGTTG GGCTGCAGCGGCGAGAAGGCCGAGCGGCTGCTCAGCGAGCCAG CGCCCCAGGCGCATGGCGCCGCGCAGCAAGGGTCTGTGGTGCGG AGGTGCGGCGGGCCCCGCTGACGCGCGCAGGCGAGCCAGGCC CCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGACGCGGCTG GGCCACGCGTCTGCTGAGCTGCTGACGACCCACGGCGGCTTAT CTGGGGGCGCCGAGGGGCCAGGCCGCGCGCAGGCCCCACAGG	15

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1	50419666	50419798	DMRTA2_ELAVL4	TGGTGGTTCAGCACCTGCTCGATGGCCTGCACACGTCGCCGCCGC AGCCCTGCAACACCAGCTCCAGGACGCCTCGCCGGTGGCTGGGA ACACGCGTGTCAAGATATCCAGCGGCGTCCGCTGCCGTGGACCCG AGCCTCCGCCCAGCCCTGGCGCCGGCGCGGCCTCACCTCTTCTTT GTCAGCCTCTGAACCGATTAGAGCCAGAGGGCTAGCGGAGCC CGGGCTGTCTCTCGCCG	16
chr1	63320169	63320413	FOXO3	GCGGGCTGCCCCGGCGGCCTGCCTGCAGCAGCGTCTTAGGAAACA GGTCAAACCTTCTGCAACTTGGCCTCTGGGAGGGGAGAAAACGTGT CGTGAGGAGCGGTTAGCTAGAAGACAGCAGTCAACAGCACCTCG CCGGGGAATGGACGGATCAGGCTGGGCCGTGGCAGAGGGAGGGT AGGAGGCAGCGACACGAGCAGCGTGGAGGGAGTCCAGAGAGCTAGC CTCTCGGACGCGGGAATCGAAATTAGGCTCATTGGAGACTACTT CGAGACCGGTGAGGGGAGCCCTGTAGCCACCATCTCCGGCGCGC ATCCACACATACTAGTCCACGCGGGCCAGCCACCAAGGCCGCGGC AGGGCCAGCGTGCGCCCG	17
chr1	67559431	67559622	SERBP1_GADD45A	TCGCTGCTTGTAGGCTTTTGTGCTTTGATGCCAAGAGCCTCAGTC TCACACGCCCCCTCTGGCCGTCCTGCCTGGGACACCGAGTTGAATT TCCCCACCTGCGTCTGGGTCTCACTCCCGCGTCCGGGCGTCCA GCTCACGCCTGTCTGGTGGATCTTCTAGTCTCTGCGTTGGCTCTCTC TGACCG	18
chr1	90717886	90718030	BARHL2	GCGGCGGAGACGCGATGCCGGGCGACTCCGGCCGCTGCCGGGCG CGTTCGCTTGAATCCGGCTGCTGGCGGGCGGCCGACCCCTCC CGTGACGTCACGGCCACTACCGCCGCTCCCCGCGCCGCGCCGCGCC GGCCCCGCG	19
chr1	109792189	109792254	CSF1_EPS8L3	ACGAGAGAAGCATGTTGCTGCCGCGAGAGGCTGCTGAGAGACCAG CCTGTTTGCATGGCTGGAGCG	20
chr1	109792367	109792397	CSF1_EPS8L3	GCGTCTGGCCCCACAGGACAACCTGGAGCCG	21
chr1	110069838	110070192	ALX3	CCGCAGTCCCCAGCCGACCCCGATTGACCACTCTAGGTTGAGGCC CAGCCTCAGGGCCCTCAAAGGGCGCCAGACACAAAAGCCGCGCTT CTTCGTCAGGTCTCAGTGTGGCTCCACAGCCCTCGGCCGGTCTGG GCTTCAGGGTAGGTGGCAGTTCAGTCCAACCTCGGCAGAGCATG CTCTCTCTTCCAGGTCCAACCTGCTTTCGGGCCCGACTGGACTCC GGGCCGTCGCCACTGCACCTTCCCTCGACCTCCCGCCTTCCATTCCC GCCGCCGAGGAACGGTGGTTCACCTCCCGCCACACTGGCCTTT GCCTGGCCCGGGCCAGCGCCAACCCGGCTTCCG	22
chr1	110083552	110083775	UBL4B_ALX3	CCGCTTGGGGAGGATCTGGCTGGTTAATGGTGATTGATGCAAA AACCCTTGATTCCATTCTGATGTACTCAAGAACAGAGATGGCTGGA GACAGAGACAAGGAGAGTCAGAAAGCGACAGAAAGTAAGTCTCTC CGGGCCTCTCCACCCAGCCAATGACAGTATCACTTCAGGAAGAGAC ACTCCCTGTCCCCAATTCGGTTCCTCCCTCCGCCAAAACCG	23
chr1	111270865	111271077	CHIA_CHI3L2	GCGGTGACCCACCGGTGAGTCCCGGGTGGCCTAGGGTAAGGCGG ACCGGGAGCCACCTCACACCCACACAGCCTGCGGGAAGGATCCGA CAAGGTGAGGGTAGCCCCGCGCGGGGCGCAACAGCCTATTCTC CCGTGTGTGACGACCCAGCCAGAGAGAACCAACCTGAGTGCCA GCGAGAGCCTGTCCTTGGTCGCTCCGACCCCTCG	24
chr1	111271216	111271381	CHIA_CHI3L2	GCGGGGCAAGGAAGCGGATCTTCATCCATGTCCTGGATGGAGTA AGGCACACTCTGGAGGTAGCAGCGAGTTTGAAGTGTCTAAGAAAA AGGCCTTCTGCAATTCACAATTCTATGGCTACCTGCACCTTTCATT ACCACTCAAAGCTAAAGGTAGCCGACG	25

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1	118979744	118979926	SPAG17_TBX15	CCGCCATCCCTCAGGGTTCCGGGTCCCGGGTTTCCAGGGTCCCGGGTTTCCAAGGCCCGCGATAACCCGGGCGCACGCGCGCGATGCGGCGAGGCGAGGCGAGGCGGTGGGGCCAGCGCGGAGCCCCAGGC GCGAGAACAGGAACTCGGGCTGGCACACCGAGGCCTCGCAGCCAA GCCG	26
chr1	118984362	118984632	SPAG17_TBX15	TCGCTCCGCGGGAGACCCGGCTTCGGCAGCACTTAGCAGAAGATTT TGGCGGGAAAGGCCCAAGCCCTAGCTGAGGACTCCGGGTGGAGC AGGGGCTGAGGTCCGAGCGCAGATGGCGCCGCGAGCGCTGAA ATATACTTGCAAGGCCGAGCAATATACTTGCAAGGCCGAGCCG GAGCAGCTGTTCCAGCCGATCCTAGCTCGAAAGTTCCTCTGTTGCTC TGGGAGAGGGCGGGGAGAGCAGGCTCGAGAGCCAGGCTCCTCC G	27
chr1	118990070	118990360	TBX15	ACGAACATGAACTCTGGGGAGCTGGAAGCAGGGTACTGGTCCCCG CCTCTGCACTCTGCCAGAGGACTTGGGGAGCCCGGATGGAGA GGCGCAGGATCTCCCACTTCAGTCAGCATTTGGCGTTGCTTCCAGG AGTCGTCGCTGAAAGTCAGCGCGCATTCACTGCTACCGGGCTTCAG CAGAGAAGCTGGAGACAAGGCAGACGGGAACCCGCAATTTCTTTC CCCAGCGGCTGGGGCCTCTCTCTACCTCCCAACTCTGGTGTGCGCC GCGTTTTCCGCTGCG	28
chr1	118993059	118993286	TBX15	TCGCTTCGCGCCGCCGCGGTGTGGCCGCGAGAGCCGGGGCCGGCG GGCCGCAAAATTGCGCGATTGTTGCTGACTTCGGTCTGCGCAGGA GCAGGGCCCTCCACAAAGGGAGCCTTGTGTGGCCAGGCCGAGC GGCCGCGCCCAAGAGGTGAGGAAATCCTGTTCCCCAGGCCACG TTCTCTTTCCCCACGGCGTTTCGTGCAACGCCGAGCCGACCTTCG GCGCCCTGGCAGTCCCGGAAAACACCAGGAAAACAAGCAGGAACC	29
chr1	148808424	148808612	ENSG00000255168_PD E4DIP	GTAGCTAGGACTGGGGTGGCCAGGCCAGGAAATCCATGAAGGG CACAGACAGCGGGTCCTGCTGCCGCCCGGATGCGACTTTGGCTGC TGCTGTGCGCGTCCCGCCGGGCTCACTACAGCCTTACCGGTCCG GGGACGCG	30
chr1	145871997	145872290	PIAS3_I TGA10	TCGGCAAGCCCCAATGAGATGCTCCATCTTCTTTTCAGCAGCTCTG CCGTTTTCTCAAAGTCTCGGAGCGCCCCGCATCTCGCTGGCCTG GCGCTCTTGTTCGGTGCCTGAGCGGCCAGGTCCCGCTCCGGCGG CGCTCCTCGGCCACCGCCTCCCGCAGCCGACCCACTTCCCGGCAGG CCGTATTATACTTTTCCAACAGAGCCTTAGCTCCTGGCCCCAAGCC TGAGCCTGGGCCACAGCGAACCCCGAGTTTCTCGTGTGGCGTA GGCTGGCTGCCTCCCG	31
chr1	155295211	155295696	PKLR_H CN3	ACGGTGTTGCGGTTCCCCCGCGTCCGGAACGCGGGGTCCACAGTCA CCAGCACCTGGGAGCCCTTACCAGCTCCACTTCCGACTCTGGACC CTAAGGAGGGAGCCAGAGGAGATGTGAGTTCTGAGCCCCGAGT CCGGGACCCGCCCTGCCACGCTGGGCCAACCTACAGGCGCC GCCTTTCGGCCCTGGCCAGCGAGTCCAGCCCCACTGCTCACCC CCTGCAGGATCCCACTGCGGATCTCCGGTCCCTTGGTGTCAGGGC GATGGCCACGGGCCGGTAGCTGAGTGGGGAACCTGCAAAGCTCTC CACCGCTCCCGGACGTTGGCGATGGACTCAGCATGGTACTGGGG GAGGGAGCGGAGCGAGGGTTTCAGGGGAAGGTGGCCAGGACCTC GAGGCATCCTCTGCCCCACCACTGCCCCGCGGCCGTCCCGCAC CTCGTGGGAGCCGTGGGAGAAGTTGAGTCGCG	32
chr1	156122596	156122849	SEMA4 A_LMN A	TCGTTTTGATGCCTCTCCCTTCTGGACGGTGGAAAGGGCTGTGTC ATAGAGTAGGAACGGGAGATGCGGCACAGGAATGGCTCCATTGA CCCGGGTTGGGGGCTAGGGCGAAGGCCTAGGAGAGGCAGAACTG	33

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1	156436615	156436840	C1orf61 _MEF2 D	TTACCTTAGAGCTGGCCAGGATTAGAGAACAGTGCCTGGAACCGG GGGGAGGGGCACGGTGACCTTGGGCTGCCACCTTCTACCTTCCA GCACCCATACTGGCTCCCCAACCTGCG CCGGGGAGAGCGGGGAAGCCTGGCAAGCCAGGGAAAGGGAAGAT GAGACAGAGAGACATAGAGAGACAGGGACAGAGGGAGACAGAG AGGGGGCTAAGAGCGACGCGGGCGAGAGAGGAAGAAAGGCTGG GGAGAAGGAAAAATGAGATAAATAAGGAAAAAAGAGAAGCGAA GGCGGTGGGAGAGGCAGCCGGGCCTCTCTGGGAGCTTAGCCAG AGGCGCCCCG	34
chr1	156642049	156642306	BCAN	CCGGGGAGGGCGGGGCAGGGGCGGGGGGAAGAAAGGGGGTTTT GTGCTGCGCCGGGAGGGCCGGCGCCCTCTTCCGAATGCTCTCGG CCCCAGCCTCTCTCACGCTCGCGCAGTCTCCGCCGAGTCTCAGCT GCAGCTGCAGGACTGAGCCGTGACCCCGAGGAGACCCCGGAG GAGGCGACAACTTCGCAGTGCCGCGACCCAACCCAGCCCTGGG TAGGTGAGTGCCTCCGACGCCCCGCCGCCGCCG	35
chr1	161069183	161069244	ARHGA P30	TCGCCTCACCTCCCTCTCCTGTTCCAGTCACCTGCCCGCTGTTTCA TCCACTCCTCCTCG	36
chr1	166921155	166921374	TADA1_ ILDR2	CCGTAGTACTCTCCAAGGAGTCGTCTGGTAGAAGCCGCTGTGCG CCCGGACTCCGAGCGCTCGAAGCGGCTCCCGCCCCGCGCCTCGTG ACTGTTGCCGTCTGCCCGGCGGGGCGGCTGGCCGTAGGAGTCAGC GAAGGCCGCCAGCTCGTCCATGGAACGGCCGCGCACCCCCGTGGC GAAGTTCTTCCGCGACAGCATCTCCGACTTGAGAGCGCG	37
chr1	220876745	220876918	HLX	GCGGATTTGCGTCACCCGAGCAACTTGCCGGTGAGATAAAGTTG CACAAATATTGAAAGGGGAAGTGCTAGGAGTCATTATAGAGTTTTT CTCCGGAAGAAATAAGGATTTCTGCAGTATCCTAAAATACTAAGGC CGCTTCTATTTTGAGACCAATCTCGCAGGCACATCCG	38
chr1	220877030	220877266	HLX	GCGGGAGTCTGCGGGCTCAGAACTCGGCGAGGGGCTGCAGGGG CCAGGCTTGGGCTGGGGAGGGGTAGAGGGGGCGCGGGGGT CGCTCCAAAGACTTGATTTGCGGTTTGCTCCGGGAGCTGGGAGT AAGGCCTTGGATGGCGCCGACGCGGTTGCGAGGAAGCTGAGGCC GGGAGAGCAAGGGGCGCGCAGGCGAAGTTGCAACTTGCACTCCA GCCGCGGGCCTGGCG	39
chr1	237042230	237042430	RYR2	GCGAGCGCGGCTGGGCTGCGGGGCTGCTTCCCCGCGTCTCCGGG CCCGGGCCGCCCTCTCCCGACAGTGCGGAGCAGGGAGGCCCG CGCCTCGACCAACCGCGCCCGAGCGTCCGCGCCTCCTCCGCTCT GCAGGCGGGGACCGCCCGGCGCTCGGCACCCGGCAGCGCGGCC CCTCCAGCCCCCGGCTCCCCG	40
chr1	237042609	237042936	RYR2	GCGTCAGGGCATCCACTAGCGGGGTCCGGGCAGAGTGACAGCGG GCAGCGGGGACTCGCGGGCGGGGCGAGGGGTGCCCTTGAGGA TGCGGGAGGAGCGGGCATCACCAAGTGTGTGCAGGTGTGCGTGT GGGGCGAGGGAAGGCAAGGGCGCGTGTCTGTGCGCGCTGTGGA AAGCTAGAGGATGGAGCGCGGCTAGCCGGCGGAGGCGCCCGGG CTCGGACCCGGGGACCGGGGACAGGAGCGTCCGAGCTGCGGGA ACCGGGAGAGGAGGGGACGGCCGGTCCGGCTGCCTGGTGGCAC GGCTGGGACCTCCCGGGCG	41
chr1	239997847	239998011	FMN2_ CHRM3	GCGCGCCCCGTGCGGGACCGGGCGGGGACGGGAGAAGGAAAAG GGCCCTGGCTCCGGGACAGGGCTCCGAGGGTGCCGGGCGGG GAGCGGAACAGGGAACGGGCTGGTGGCGGCCCAAGCGGGAGG GACGGACCGACACGCGGCCCTGGCGGCCTTGCG	42

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1	239998124	239998259	FMN2_CHRM3	ACGGTCGCCGCGGGCAAGGACCGCGAGGTTGCGGCCCTGCTCCGA ATCCCCGGCTGCGCTGGCCACGCTCCTCCACGCGGGGGCGGCCGC TCCGCCACCCGCACGGCGCCCCGAGCTGCTCCGGTGGGGATTCTG GCGCCGCCCGGGGAGCGGCTGCGCGAGGATGCGCGGTGCCCGGT GTGCTTGATTCTCTGAGGAGCCGGTCAGCGTGGACTGCGGCCA CAGCTTCTGCCTCAGGTGCATCTCCGAGTTCTGCGAGAAGTCGGAC GGCGCGCAGGGCGGCGTCTACGCTGTCCGAGTGCCGGGGCCCC TTCCGGCCCTCGGGCTTTCGCCCCAACGGCAGCTGGCGGGCCTGG TGGAGAGCGTGCGGCGGCTGGGGTTGGGCGCGGGGGCCGGGGC GCGGCGATGCGCGCGGCACGGCGAGGACCTGAGCCGCTTCTGCGA GGAGGACGAGGCGGCGCTGTGCTGGGTGTGCGACGCCGGCCCCG AGCACAGGACGCACCGCACGGCGCCGCTGCAAGGAGGCCGCCGC AGCTACCAGGTGAGGCGCCCCCGCGGGGGCTGCG	43
chr1	247857256	247857693	TRIM58	CCGCGCTGCTCCCCCTCCACCCCGAGGCAGCTCCAGATGGACACA GCAGGTCGGAACATCCACACCCAAAGACAGACTACGGAGCAGA GCCGGCTTCCGCAGCG	44
chr1 0	509413	509519	DIP2C_Z MYND1 1	CCGGCAGGTTCTGGGAAGTCTCCGTATTTCGAGGTACAGGAGCC ATAAATCCATATTTAATTAGCTTTGAACG	45
chr1 0	3472562	3472635	PITRM1 _KLF6	GCGTCGTCCCGGATTCTCGGACACCACAAACGCCATCAACCACGA GCACCGGTGTCCGTGGCTATTGCCCCGAATGTCCCCATCCGCGTC CCCGGGAATCCCTCGGCTTTTCGCGCATCCAGGTCCCCAGCCCCA GCTACTGGTGCGCCCCGAGCCCCCTAGGTGCCAGAGCGGTGGTCGG CCGGGCTCCTGCCAGTCTCG	46
chr1 0	7407762	7407965	PRKCQ_ SFMBT2	CCGCGCTGCGCCTACCCAGTGCCCTGGCCCCGAGGGCGACAGC GGCTGCTCCCTCCCATTTGCGTCCCAGACCGCGCGCCTCGCTTAG CTCCCGGGAGCCGACAGGCGCTTGCCCTGGTGCCAGCGCAGGGCT TCCCG	47
chr1 0	7413140	7413280	SFMBT2	TCGAGATCTTTATTTTCTAAAGGTGGGGTTGCCCTTCTCCATCC CCGGCCAGTCCGACTTGGTGCTCGCGATTGAATTTAAACGAATAAT CCCTACTTCCCCATCCAAAATTAGCGGATAGGCGCCCTTGACCCG CCGGATCACCTTCCAATGACACCCGCATATACTTGCAAACTGTGCA AAAGCCCTTGAAAAGTCCAGAGATGGGACAGAAGCCCCCAGCAGA ACCCAGGCCGAGCCCCGCGCACCTCGGATAAGGGGGTGCGGGA ATGCACCCACCTGGTCCCTGAGGGCAGCACCTTAGATTGCCCAGG CTGCCGCGGAGGAGGACGATCGCCGCGGGGCTCCGCTCTCGCCG TCTGGGCCACCGCGCG	48
chr1 0	8052248	8052385	GATA3	CCGCGCGCGGCCACCCGCGCCTCTTCTCAAATCACTTACCCGATT ACTCCAGACTGTGGCCGGGAGGTCACTCCCTGCAGAAGTGTCCTCC CTCCCCAACGCCGCGCAATAATTTTAAAGCAAAGGAGCGCGCGC CAGGTGGGCTCCCAAGCTCCGCGCAGACCTTGGGCCAGCCTTGG CCGCTACCCGAGCG	49
chr1 0	23191669	23191912	PTF1A	GCGGGGCCGACGGCCGGCTGCAGGGCGGCTGGCTCTCCGCGCTCG AGACTAGGCGCACTCCCATCCCCGCCGATGTTCTCCACGCGGGCT CCAGCGGCTCACACCGCCACCGCGCTCGTCTCGGCTTTATTTACC CAGCCCCGCGCGCCGCCCGGGGAACAGGAATAGCGAGGCCCTTCT CATGTTCTGACTGCCGGTCCCAGCCGGCG	50
chr1 0	27745381	27745577	MXK	ACGCGCTCGGCCCGCAGGTGGCACTCAGTAGACCTGACGCACGT GTTCTGCTGTGTGGTAGCCTGGGGAGGCTCCCCAGCCTGCCTCA GTGGGCTCTCCCTGGTGGCCCGGCAAGAGCAGAGCTTCATGAG AGCCCCGTGCTGCACTGCTGGGCTGCCTCGATGCCAGCCAGGCCG	51
chr1 0	27745715	27745928	MXK	CCGCGCTCGGCCCGCAGGTGGCACTCAGTAGACCTGACGCACGT GTTCTGCTGTGTGGTAGCCTGGGGAGGCTCCCCAGCCTGCCTCA GTGGGCTCTCCCTGGTGGCCCGGCAAGAGCAGAGCTTCATGAG AGCCCCGTGCTGCACTGCTGGGCTGCCTCGATGCCAGCCAGGCCG	52
chr1 0	70560556	70560867	PRF1_P ALD1	ACGCGCTCGGCCCGCAGGTGGCACTCAGTAGACCTGACGCACGT GTTCTGCTGTGTGGTAGCCTGGGGAGGCTCCCCAGCCTGCCTCA GTGGGCTCTCCCTGGTGGCCCGGCAAGAGCAGAGCTTCATGAG AGCCCCGTGCTGCACTGCTGGGCTGCCTCGATGCCAGCCAGGCCG	53

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 0	72318365	72318398	DNAJB1 2_DDIT 4	GAGGCTTGAGATGCCCGAAGTACCCAGTGCCCCGGCCACCTCTCCT GGCCCTCTTCTATTTTAGGGCTCAGTCCAATGGATGAGGAAGCCTT GTCCGGCTCCACCACAGCTAATGACAGCCTGGCAGGCCG TCGACCTTTCAGCCCGGTGGAGAAAGCAACTTCG	54
chr1 0	72450476	72450539	DNAJB1 2_MICU 1	GCGAATGGAGGTGACTGAAGGTATCAGTGCCAAACAGGTTCTTTTC TGCTTCATACACATTCCG	55
chr1 0	92692207	92692364	EXOC6_ HHEX	TCGGTGGGAACGTGTTAGGTCCACGTGCCGGTGGGTGTATGTGAA TGTGTCTGGTTGGGTGGCCTCCTGGCCTACCTTTGTATCCCTGGG GCCCGACAGCTCTGGGGTCTGGCCAGGCCGCTCCAGGGCAGTGGG TGAGCGCCGCTCTTCCCGCTCG	56
chr1 0	93062579	93063009	CYP26A 1_CYP2 6C1	GCGAAAGCAAAAGCCAGGAAGTTTAGGTCTGGGCCGCTTGGAAGA GGGAGAAAGGACCGGAAGTGGCCTTCTGGTACTCCGGAATCGCC AAGCAGATGAGGCCAGACCGCCGCCAGCGTGATCACGCGCGCTC CCACAGGTCCTGGCGCGCGTGTTAGCCGCGCCGCGCTGGAGCGC TACGTGCCGCGCTGCAGGGGGCGCTGCGGCATGAGGTGCGCTCC TGGTGCGCGCGGGCGGGCCGGTCTCAGTCTACGACGCTCCAAA GCGCTACCTTCCGCATGGCCGCGCGCATCTGTGGGTTGCGGC TGGACGAGGCGCAGTGCGCCACGCTGGCCCGGACCTTCGAGCAGC TCGTGGAGAACCTCTTCTACTGCCTCTGGACGTTCCCTTCAGTGGC CTACGCAAGGTACGGCCGCCCG	57
chr1 0	93068345	93068620	CYP26A 1_CYP2 6C1	GCGTGATGTATAGCATCCGGGACACGCACGAGACGGCTGCGGTGT ACCGCAGCCCTCCCGAAGGCTTCGATCCAGAGCGCTTCGGCGCAGC GCGCGAAGATTCCCGGGGCGCCTCCAGCCGCTTCCATTACATCCCG TTCGGCGGCGGTGCGCGCAGCTGCCTCGGCCAGGAGCTGGCGCAA GCCGTGCTCCAGCTGTAGCTGTGGAGCTAGTGCGCACCGCGCGCT GGGAAGTGGCCACACCCGCTTCCCCGCCATGCAGACGGTGCCCAT CG	58
chr1 0	97321084	97321318	FRAT1_ FRAT2	ACGCACTGGGTTGCGGGACAGAGTAGCCAGGTTCTGCCGGTGCTC GGAGAAGAGCGCAGTGTTTTGCAAGTGCTGGAGTCTCCTGAGGAC ACGCGCGTCGCCGCCACCGCGGGTGTGGGAAAGCGCGGACGTGCT GGGCGGCTGTGCTTCGGTAGGCGACCACCGCCCTGGCCGCGCTC CGGGCTTTCACGAAACTCCCGAGACCGGGCCCTGGGTTCTCTCTC TCCTACTCG	59
chr1 0	101140065	101140385	TLX1_LB X1	CCGCGGAGAGCACATGCAGGCCGGAGCCCTCAGCCCGGCAGCTCT CGGACCCTGCCAGCTCGACGCGGACTCATGCAGAAGAGGACATT CCGCAAGTAGGTACAATCCAGCGCTGGGGCCTGGGGCGTCCGGG GGGCGGCCTTTGAGCTTCCCGGATACCGCTCGCTGCTCCCGGAGC TGTTCCGGCCGCCGGCTGCCGGGTCTGTGCACTTTCAGTAGGGCCCC GCTGACTCTCTGCCCTTGGGCTAGGCCTCCCGGGGATGCCAGACT CCTGGGGACGCTGGGACCCGCGGCGCGGGGACACGCAGGACT CCCG	60
chr1 0	101284337	101284644	BTRC_L BX1	CCGCGCGCAGCTGGAGCCCGGCGAGAGGGCCGCGGAAGGGGGGT GCGAACCAGGGGCCGACCCGGGGAGGAGCCGGGAGGCGAGCG GCGAGGGGCACTGCGCGGCTGGGTCTGCCCCGGGGTTTCGCACTG CGCCGCGGGTCAAGTACCGCGAGTTGGCCCTGACTGTCTGCAGG ATGAGGGTGTGAGGAGGGTTCCAGGCCAGCGTGCCTGCCTCGCC	61

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 0	107164022	107164546	SORCS1	TCCAGCCCGGGGTAAGGAGATCCACGGAGGCTCTGCGCCTAAAC TCAGGTGGCCAGACAGAGTTGGGGCGGGAGGCGGGTATACG CCGTCAGCGCAAACGTGGTGCTGGTCAGTCTCAGCTCCTCCATCCG GAAGCGGGTGGCTTTGTCCGGGTCCCGTCCCGAGTCCCAGGCTCC TGCTGCCCTCCATCTCTTAGCACTCCCCGGGGGTCCGACTCGCGCC CTCTCCCCGTTCTGCCTTCTCCTGATCCGCTCCGCTCCGTCTCCTCCG GCCGGAGCGTGACGAACCGCCATGGATGCCCCAGTGCCCCGAGC CCGCTCCAGGGATAGCGCTCGGTCCCCGGGGGCCACTGAGAACAG GGGACGCACTACGAGGGGCAGGGGCGTGGCAGGAGCCCTGCCTG GCCGCCCTGGTGGGAAAAGCCCCAGGGGTCGAGGCCGAGCGT GGAGCGGAGCTGGGGTGCGGCGAGGGGCAGCAGGAGCCGCCGC CGCAGACCCCCGGGCGCAGAGGATCAAGAGCCCCGCGCCGGCG AGGAGCGCGCTCAGCCGGGCTTGGGAGCCGCCGCCGCGCCAAT TTCCCATCGCGGGAGCGAAGAGCAGCG	62
chr1 0	108912398	108912568	NONE	GCGGGCTGGCTGCCTGGGCAGCACAGGACTTGAGGGAGCTGCGG GGACTCCTGGAGTCTCATCAGGCCTTCCAGTCGCTGTGGGGACCCC GGCTGCGCGCGGATCGCCTGCGCCACTGTCCCACTGACCCGCCCG CCGGGTTTGCCAATTACCAGCGCCACCTGGTCCCG	63
chr1 0	122163615	122163857	PLEKHA 1_TACC 2	TCGGACCACACCGGCGCTCACGCTCATACCCGCACGCCCGGGCAG AGCCGCGCACGCCGCCCACTCGGGCGCGCGCCGCCCACTCG CGCGCACACATACGCGGCGCTCGCCCCCGCCCCGGCTCGGGC CGCGAGTCGAGCTCCCTGCCCGCTCCCGGCCCACGGATGCC GCAGTGCTCCCTCTGCAGTGACGAACCCCGCCGCCCGCGCGC TCGCCCCGGTCCCG	64
chr1 0	122164217	122164660	PLEKHA 1_TACC 2	TCGCTCAGCAGTGGGTGCATGGCTGGGGGGCTTCTCCTGCCGTCA GCATCTTCTCTGCACCCCCGGCACAGTGGTATTTCTGCAAGGG AACAGCCAGGCATCAGCGACTGCCTCCTCCTAGGAAGAACCCATGA GCGTGGCAGCTCCGTGCCCGGGGCGACAGCCAGTTTCCGGGCAG CTGCGCTTGTGGCTGGGCAGATGGCGTGGTGCGCTCTGGTGGACG TTCCGTCTAGTTAGCCTAAGCATCATCCATACTCTGGTGAACACT CGAGGACAAGGCCGCTTGCTATTATTAGTAAAGGGCCGAACCGTC CTGTCAATTGGTGGAGGCAGTGCTTGACTGTGCATCGATCCAGGAAT CCGATCTTTTCTCTCAACCACAGAGCTAACGTGCTCAGAAGTGCCCT TTATCTGGCCGAGTGTTATTAGAATTACG	65
chr1 0	123145980	123146613	HMX2	CCGCACGACATATTTACAGTTCAGGAAGGTTGACCAACTTTCCCT GCCTGCCCCAGCTTTCTTCCCAGCGGGGTGGCTGGCACTGCTCC CCGAGTTAGCTGGCCAGTTCCCTCGGGGCTGCCTTGACCTGGCT CCGAGGCAGCGCCTAGCTCAGGATGTCTGCGAGAAGCGGATGGT TAGTGAGAATCCGACGATTCTTTCGCTGAACCTCCCGCTACCCCC AACAGCGCGGGAGCACGCGGGACCCGCTGCGACGTGGCCAGGA GCCTGCGCCGCCGCGCGCAGAGGAGAACGCACAAATTGTATTT AGCGCCAGGTCCTTCCGGGTTAATGAGCTGACACCATGATTAAGC TGACCATTTGTAATGTGTCTCGACCCTGCCGCTGAGCCCTGAAGAG GTTAATGCGGTGACGAGGCCGGCACCTGCCCCCTGCTGGCCTCCC GGGCCGCTGCGCGCACCCCTGGCCCCCGCCCCCTCGCTGCCCT GCCCGGCTGCGCGCGCACTCCTAATCAATTAGCCATTAAACGAG CCCCTCGAGGAGTTAAGTAGGGAAGAGTTCTGCCACGGGCAGGGC CGCAGTCGGTAACTACCGCGGCTAATGATATTATAAGCG	66
chr1 0	123151236	123151480	BUB3	GCGCGGAGAGGGAAGTGGGCGCGGTGAGGCAGTTCTGCGGCTCA GGAGAGATCCGAGGCCCGGGACCAGGCAAAGAAGGTGAGGGAG	67

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 0	124447401	124447594	OAT_NK X1-2	GCAAAGGCGCTTCCCTACACTCTTTGTTGTTAATAGTTTGCATTGG TTCAGCGTGTGGCTGGATCACCGGCTAGCACGCGGCGCTTGTCT GAATGGAACCTTGACGCGCGGCGGGGGCGCCACGGACTTCTCTG CCCTGACACCTGCGGCGCG GCGAAAGAGGGGCCAGGGGGCTCCGGATTCATAGACGCGGGGCG TAGAAGGGGGTCAAGTAGGAAGGCCAAGGAACGGCGCGAAAGG GCTCCCGGGGGCGGCGAGCCGTCAGCGGGAAGGAGGCGGCGGACG GGAAGAGGACATTGGCCGCGGAGTAGGAGGGGAAAGTCTGGAAG TGCAGAGCGCCGTGCCG GCGCGGAGAAACCTGGCGGGGGCCCCGGACTCCCCGGCTTGGGAA AAGCGATGACTGCCCTGAACTGCTGGGGCGTTCGAAATTTCCAGG GTCCCGACCCCTCCGTGGGGTACGCGCGACTTCGGCGCAGATGTCA GTCCGCTGCCCTCCGGGTTGAGGGAGCGAGGACTCCAGACGACCC CAGGGCCGCTGTCCAGGCCAGCCCCGCG CCGGCAGTGGGGAGCACCAGCTGGAGAGTGGGTGTGAGGCCACC ACATCCCCCTGCAGCTCCAGCGCCATTGAACTTTGAGGAAA GATCTCAGCTCCTGCCGGAAGGCCCTGCACAGGCTGATGACCT GCTCTCCTGACTCTTTCTGACTCTTTTCCGGCGAACCTGCCACCTC CTCCTTCAGGCTGGGCGG ACGGATGATTCCGTAAGCAACTGGAAACCCAGTACAAATAGTCC AACTTAGACAGTAGGACGGAGTAGAAGACAGGGTTCTGTGAAA AAAAATAAATGCTTTTCTAAGGTTAACGCCGGGAAAGTCCGGG GCCTCCGAATTCCACTCCAGTGTCTTTAGTCACCGGGCCACTTGC CTTGTCAAATGTGCGGCTGGGTTTCATCTGCACTGATGACAACG AAGGCCGTGGCAGCTATTAATCTTCACTATGGTCTCATGAACTAG TTAAGCATGAAGGGTGACAGCCCTGAGCCCCAGGGGCTTGACAA CTGCG CCGAAGAGCTGGCGGAGAGAAGCGGCTCCAGTGCTTAGCCGGCC TGTCGGAGCTTCTCTGCCTGTACAGCGCCCTCGCTCTTAGCACATG TTTTCAAGGTATCTCTAACACCGGCTGCCAGTTGCCCAATCGATA GAAGCAACATCACACTCTTCTTAAAAAGGGAACAAAGTGC TTTCGATAAAGCCTCATCATCTATAGCTTCTCGG TCGCTCCGAGGTCCCCGCGCCAGAGACGAGCCGCGCTCCACCAC CCACACCCACCGCGCCCTCGTTCCCTCTTCCGGGAGCCAGTCCG CGCCACCGCCCGCCAGGCCATCGCCACCCTCCGAGCCATGTC CACCAGGTCCGTGTCTCTGCTCCTACCGCAGGATGTTCCGGCGGC CCGGGCACCGCGAGCCGGCCGAGTCCAGCCG TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	68
chr1 0	127736442	127736649	FOXI2	GCGCGGAGAAACCTGGCGGGGGCCCCGGACTCCCCGGCTTGGGAA AAGCGATGACTGCCCTGAACTGCTGGGGCGTTCGAAATTTCCAGG GTCCCGACCCCTCCGTGGGGTACGCGCGACTTCGGCGCAGATGTCA GTCCGCTGCCCTCCGGGTTGAGGGAGCGAGGACTCCAGACGACCC CAGGGCCGCTGTCCAGGCCAGCCCCGCG CCGGCAGTGGGGAGCACCAGCTGGAGAGTGGGTGTGAGGCCACC ACATCCCCCTGCAGCTCCAGCGCCATTGAACTTTGAGGAAA GATCTCAGCTCCTGCCGGAAGGCCCTGCACAGGCTGATGACCT GCTCTCCTGACTCTTTCTGACTCTTTTCCGGCGAACCTGCCACCTC CTCCTTCAGGCTGGGCGG ACGGATGATTCCGTAAGCAACTGGAAACCCAGTACAAATAGTCC AACTTAGACAGTAGGACGGAGTAGAAGACAGGGTTCTGTGAAA AAAAATAAATGCTTTTCTAAGGTTAACGCCGGGAAAGTCCGGG GCCTCCGAATTCCACTCCAGTGTCTTTAGTCACCGGGCCACTTGC CTTGTCAAATGTGCGGCTGGGTTTCATCTGCACTGATGACAACG AAGGCCGTGGCAGCTATTAATCTTCACTATGGTCTCATGAACTAG TTAAGCATGAAGGGTGACAGCCCTGAGCCCCAGGGGCTTGACAA CTGCG CCGAAGAGCTGGCGGAGAGAAGCGGCTCCAGTGCTTAGCCGGCC TGTCGGAGCTTCTCTGCCTGTACAGCGCCCTCGCTCTTAGCACATG TTTTCAAGGTATCTCTAACACCGGCTGCCAGTTGCCCAATCGATA GAAGCAACATCACACTCTTCTTAAAAAGGGAACAAAGTGC TTTCGATAAAGCCTCATCATCTATAGCTTCTCGG TCGCTCCGAGGTCCCCGCGCCAGAGACGAGCCGCGCTCCACCAC CCACACCCACCGCGCCCTCGTTCCCTCTTCCGGGAGCCAGTCCG CGCCACCGCCCGCCAGGCCATCGCCACCCTCCGAGCCATGTC CACCAGGTCCGTGTCTCTGCTCCTACCGCAGGATGTTCCGGCGGC CCGGGCACCGCGAGCCGGCCGAGTCCAGCCG TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	69
chr1 0	128934041	128934243	MKI67_ MGMT	CCGGCAGTGGGGAGCACCAGCTGGAGAGTGGGTGTGAGGCCACC ACATCCCCCTGCAGCTCCAGCGCCATTGAACTTTGAGGAAA GATCTCAGCTCCTGCCGGAAGGCCCTGCACAGGCTGATGACCT GCTCTCCTGACTCTTTCTGACTCTTTTCCGGCGAACCTGCCACCTC CTCCTTCAGGCTGGGCGG ACGGATGATTCCGTAAGCAACTGGAAACCCAGTACAAATAGTCC AACTTAGACAGTAGGACGGAGTAGAAGACAGGGTTCTGTGAAA AAAAATAAATGCTTTTCTAAGGTTAACGCCGGGAAAGTCCGGG GCCTCCGAATTCCACTCCAGTGTCTTTAGTCACCGGGCCACTTGC CTTGTCAAATGTGCGGCTGGGTTTCATCTGCACTGATGACAACG AAGGCCGTGGCAGCTATTAATCTTCACTATGGTCTCATGAACTAG TTAAGCATGAAGGGTGACAGCCCTGAGCCCCAGGGGCTTGACAA CTGCG CCGAAGAGCTGGCGGAGAGAAGCGGCTCCAGTGCTTAGCCGGCC TGTCGGAGCTTCTCTGCCTGTACAGCGCCCTCGCTCTTAGCACATG TTTTCAAGGTATCTCTAACACCGGCTGCCAGTTGCCCAATCGATA GAAGCAACATCACACTCTTCTTAAAAAGGGAACAAAGTGC TTTCGATAAAGCCTCATCATCTATAGCTTCTCGG TCGCTCCGAGGTCCCCGCGCCAGAGACGAGCCGCGCTCCACCAC CCACACCCACCGCGCCCTCGTTCCCTCTTCCGGGAGCCAGTCCG CGCCACCGCCCGCCAGGCCATCGCCACCCTCCGAGCCATGTC CACCAGGTCCGTGTCTCTGCTCCTACCGCAGGATGTTCCGGCGGC CCGGGCACCGCGAGCCGGCCGAGTCCAGCCG TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	70
chr1 0	129415181	129415505	MGMT	ACGGATGATTCCGTAAGCAACTGGAAACCCAGTACAAATAGTCC AACTTAGACAGTAGGACGGAGTAGAAGACAGGGTTCTGTGAAA AAAAATAAATGCTTTTCTAAGGTTAACGCCGGGAAAGTCCGGG GCCTCCGAATTCCACTCCAGTGTCTTTAGTCACCGGGCCACTTGC CTTGTCAAATGTGCGGCTGGGTTTCATCTGCACTGATGACAACG AAGGCCGTGGCAGCTATTAATCTTCACTATGGTCTCATGAACTAG TTAAGCATGAAGGGTGACAGCCCTGAGCCCCAGGGGCTTGACAA CTGCG CCGAAGAGCTGGCGGAGAGAAGCGGCTCCAGTGCTTAGCCGGCC TGTCGGAGCTTCTCTGCCTGTACAGCGCCCTCGCTCTTAGCACATG TTTTCAAGGTATCTCTAACACCGGCTGCCAGTTGCCCAATCGATA GAAGCAACATCACACTCTTCTTAAAAAGGGAACAAAGTGC TTTCGATAAAGCCTCATCATCTATAGCTTCTCGG TCGCTCCGAGGTCCCCGCGCCAGAGACGAGCCGCGCTCCACCAC CCACACCCACCGCGCCCTCGTTCCCTCTTCCGGGAGCCAGTCCG CGCCACCGCCCGCCAGGCCATCGCCACCCTCCGAGCCATGTC CACCAGGTCCGTGTCTCTGCTCCTACCGCAGGATGTTCCGGCGGC CCGGGCACCGCGAGCCGGCCGAGTCCAGCCG TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	71
chr1 0	129425147	129425366	MGMT	CCGAAGAGCTGGCGGAGAGAAGCGGCTCCAGTGCTTAGCCGGCC TGTCGGAGCTTCTCTGCCTGTACAGCGCCCTCGCTCTTAGCACATG TTTTCAAGGTATCTCTAACACCGGCTGCCAGTTGCCCAATCGATA GAAGCAACATCACACTCTTCTTAAAAAGGGAACAAAGTGC TTTCGATAAAGCCTCATCATCTATAGCTTCTCGG TCGCTCCGAGGTCCCCGCGCCAGAGACGAGCCGCGCTCCACCAC CCACACCCACCGCGCCCTCGTTCCCTCTTCCGGGAGCCAGTCCG CGCCACCGCCCGCCAGGCCATCGCCACCCTCCGAGCCATGTC CACCAGGTCCGTGTCTCTGCTCCTACCGCAGGATGTTCCGGCGGC CCGGGCACCGCGAGCCGGCCGAGTCCAGCCG TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	72
chr1 0	17229289	17229505	VIM	TCGCTCCGAGGTCCCCGCGCCAGAGACGAGCCGCGCTCCACCAC CCACACCCACCGCGCCCTCGTTCCCTCTTCCGGGAGCCAGTCCG CGCCACCGCCCGCCAGGCCATCGCCACCCTCCGAGCCATGTC CACCAGGTCCGTGTCTCTGCTCCTACCGCAGGATGTTCCGGCGGC CCGGGCACCGCGAGCCGGCCGAGTCCAGCCG TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	73
chr1 0	129467252	129467332	MGMT	TCCCGACGCCCGCAGGTCTCGCGGTGCGCACCGTTTGCAGCTTGG TGAGTGTCTGGTCTGCTCGCTCCCGAAGAGTGC CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	74
chr1 0	91633125	91633224	PPP1R3 C	CCTGGGACCAATCGCCGGGCTCGAGCCCCAGGGCGCGACCAACC AGCGCCAGCTGGGGCGCCAGCCCTCGCCCCGGCAACGTGATCGC CCCGGGGCGA TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	75
chr1 0	86924203	86924324	BMPR1 A	TTTATGATAGTTTGTCTGTCTTAGTGATGTGTGTGTGTCTCCA TGCACATGCACGCCGGGATTCCTCTGCTGCCATTTGAATTAGAAGA AAATAATTTATATGCATGCACAGGAAGAT TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	76
chr1 0	17453751	17453872	ST8SIA6	TCTCGACTCCCCGGCTCCAGGCCAGGTCCCCAGCCCCAGAGTTG GAAGAGCCTTAGGGCGGGAAGGAAGAGACAGCAAGGACCAAGAT GGGGAGCATGAGATCCTGATGCGGAACCCGAC TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	77
chr1 0	17454616	17454782	ST8SIA6	TCACCTGAAGGTTGGGGCGCGGAAGCTCAACTCCGTGCTGATTGG GCTCCAAGTTTCTGCGCCCTCGCTCGTCCGAGTGCCCGCAATC	78

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 1	280945	281044	ATHL1_ NLRP6	CCCCGGACGCCACGCAGACCAACCAGCCACACCACTCTGCCT GCGGAGAGAGGAGAGGAGAAAAAGGGGCC TCGGTCGGGACCTGTGCGGCACGTCCAAGACCACACGTCAGTGTA CCTGCTTTTCATCACCAGCGTTCTGAGCTCGGCTCCGGTAGCCGAC GGGCCCCG	79
chr1 1	281278	281496	ATHL1_ NLRP6	ACGGCGGGGTGCCAGGACCGCGGCTGGCGGCGTTGGGACACTC CTGCGTGGGGACGCCAGCCGCACAGCCACTTGGTGCTACCACG CGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGGATGCGCGACATC GAGCGCCACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGGAG GCCCTGCGGTGGGTGCAGGGACAGGGACAGGGCTGCCCG	80
chr1 1	636846	637203	DRD4	GCGTCTGGCGGAACGGGCTGGGAGGGAGTTTTGCCAGATACCA GGTGACTAGGGTGAGCGCCGAGGGCCGGGACGCACGCACGGG CCGGGTAGGATGCGCTGCGCTCGATGCCGCGCGCTTCAGGGCC TGGTCTGGCGCCCCCTCCATCCTTGTCGGTTTCTCGGGTCGCGGAC CCGCGCGGCGCGGGGCGATGCTGGCCTGCCGTGGCCACCACT CGCTTCATTCCCGTCTCTTGGGCGCGCCGATTGTCACGTGCCCG TCTCTCCCTGCGCAAAATTCAGATGAGCAAACTGCGGCTCACG GTGGAGCGCGCGGGGGCCCCCTGAGCCGGGCGGGTCTG	81
chr1 1	1310494	1310628	TOLLIP	GCGGAGGACAGGCGTTATGCAAAGATTGGCAATCCTTTGACGAGC CCAGGTAGTACAGCACGTCTCCCCGTGATGTTTTTGGCTTTTATC TTACATATAAAACAGCGTACCCAGGTGGACGCCTTCTCCTCG	82
chr1 1	1310748	1310787	TOLLIP	ACGAATCCTCTTTTGGGGTCTGGATCAGGACCCTTTTCCG	83
chr1 1	1661927	1662031	KRTAP5 - 6_KRTA P5-5	GCGCCCGTGGCTTCTGTCATCTGCCGACACCACCCGAGGCTGCCAG GCCACAACATGAAGTCAGCTGTGCCAGGAAATCCCAAGCTCGCCC ACACCTGGCCCCG	84
chr1 1	31815810	3181625 9	PAX6_E LP4	TCGGCGCTTTTCGTCACTTCTAACCAGTCTCACAGAGGGTGACTT CCAAACCTGGCTAGCGGGGAAAACCGCTGCCCGGGGGACAGAGG GGCTGACAGGAACTGCGGGTTGGCTCAGCCGAATGCGGCCGGGG AGAATTTAAGAATTCTCAGCCGCGCGGCCGATGCCTTGATTCC TCACGAGAGGAAAGGGAATGAAAAATGAAGCAACAATGACACC ACCCAGGCTGGCAGCCCTCGTTCCCGGCCAGACCCCGCTCCTCAGG CCCGGCTCTGGCGCCGGGTGGCGTCCAGCCCTGCACGCGCGGCG CGGCCCCGCGGAAAGTTTGTGACGCGAGAGTACTGTCCTCCGC CTCGCGCGCGCTGCCCTTCTGCCCGGAGGGGCGTTGGGTCC TTGGTTTTCTTTCCAATTCTAAATAAATAAAATAAACTCCG	85
chr1 1	58905378	5890557 6	GLYATL 2_GLYA TL1	CCGCTGGATCCCGCTGGATGCACGTCCCGCCACCGCCGACCC ATCAGCGCGCAGAAGGGCAGCAATGGCCACACACCGAAGCACCTTG GCGGGCTATTCCCTTGACGCTCTCCTCAGCGCGTGCTCCCACTCG CAATCAAAAGGCGGAAAAAGCGGAAACCGCCAGGCATCTCCCAT ACCCACCCGGCTGCCG	86
chr1 1	61769415	6176972 3	MYRF_T MEM25 8	GCGGCTGCCAACGGGCTGAGATTATCGCTGGTCAAATACTCCCTG GCGCTTGGCTATTGTTTTCCACGGGCGGGTGGGGAGCCTGGCCCT GCCTCTGAGCAAGTATCCCCGCGGTGATGCCACCCGCTGCCCGCC TGCGCCATCATGGACGCACCTTCGGCGGTAAGTGGGTGGCTGGG GAAGGCCGTGGGTGCAGCCTGGGTGCAGGCTTCCAGGCCGGGCC CACCTCACCTTAGAGGGTGCTCAGGGGTGCCCTGGCCCCAGGTG GCCAAGAGCAGAACCACCGCGGGAGCAGGCTCCCCG	87

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 1	62444331	62444643	SCGB1A1_AHN AK	CCGGCCTCTGCCACAGCTGGGTGGGTGCCAGCCAAGGAAGCTTG TGCCCCATCATTACAGGCATTGTTCTCCCTTAGAAGAGGATCTCGA AAGCAGAAGGAAATTAGAAACAACCGCACAAATGAATACCAGATTC TGCTTTCTCTCAGCTCTGTCTGCCAGGAGATTAGGCAGGGTTGGCT GACAGCGTGCCCCGCCCGGCAGCTGCTGCCCTCCAGGATGTCCGC GCCGTGGGGAAGCGGGGGTCCCGCTGGCCTTCTAGCTCTCTATTTA TCTCCAAAGTGTCGGTTTTCTTTCTCTGCTAGATGCG	88
chr1 1	63919655	63919894	RCOR2	GCGGAAGGGGCCAAGGAAGCTGGGCAGCGCGGCCGAGAACCCGG GGCCTCACCTACCCGAGCTACCTCCGAGCTTGCGCGGAGCCGGA GGGCTCCCGGGAATGCCCTCCCCGCCATTTTCGCCGATGAGCTCGG GCTCACCTTCCACTGGAAGCGACAGCGCCTTCTTTTCGAGGGCTG CAGGCCAGGACGACAGGCCGCTGGAAGCAAGTGTGATCAGGGCA CATTTATTTCTACG	89
chr1 1	76322032	76322348	WNT11 _PRKRIR	TCGGGAATATTTGTGGGCTGCCGGCGGGGCAGGCGGGGTGGGGG AGGCTGCCCGCGGGCGGGAAGCCCCGCGCACTCGGGTCCCTTCG GGTCCCGCGGGGGTTCGGCGCGTGCGGAAAGCGGCCGAGCCG CCAACCTCGGCCCGTCCGCAACCGAAGAGGAGCGACCGCAGCCT GGAAAAGAAGAGCCCCAGCTGTTTCTTCCACCGGGCGGGCGG GACGGAGAAGGGAGGGAGCCTGGGAGAGACGCAGGTGTGGCGC TCGCCTGTGCTGGCGGGGTGGCAGCCGGGGCGTGGCACCTCGGA GTCTCG	90
chr1 1	77039610	77039900	CAPN5_ B3GNT6	GCGTGAGTTTCTTAGCACTGCAGCAGTGGTTCTCCAGGCGCCAAG GTCCCCGCGGGAGGAGAGGTCCCCGAGGAGGAGACGCCAGAGG GTCCACCGACGCTCCCGCGGCTGACGAGCCGCCCTCGGAGCTCGT CCCCGGCCCCCGTGCCTGGCGAACGCCTCGGCGAACGCCACGGC CGACTTCGAGCAGCTGCCCGCGCGCATCCAGGACTTCTGCGGTAC CGCCACTGCCGCCACTTCCGCTGCTTTGGGACGCACCGGCCAAGT GCGCCGGCGGCCGAGGCG	91
chr1 1	94867437	94867593	AMOTL1_CWC 15	ACGTGCAGCCAGGCAGGCATCTCTGGTGTCTGTGCCCGTATGCCCC AGGACCTGGCATGTCTAAACCAGGCCTGGGAGCCGGAGGACTTGT TTGAAGGAAGAGCTGCTGTGTTCCCTGCACTGATATTCTCTCATT GTTGTATTGGTGTCCACG	92
chr1 1	125166486	125166741	PKNOX2 _FEZ1	TCGCGGGGCTGGGAGTGATCTGAGGTCCCGACCCAGGCGGCTCG GAGTGCTCCAGGAGCCACCTGGGTCTGCGGGCGCAGCGCGGCGG GGCGGGAGCGGTGGCCGCGAGGGCGCGGCTGCGATGAAGGC CGGGGGGCGAGCGCTAGCAGCGAGGTGCCACAGTGGGCCGAGGAG TCTGGGCTGTGGCCAGGGTAGGACCGGCTCAAACCTCAGTGCCCT GATTGGAGCCGCTTCTGTGCTTACCCGCGCCG	93
chr1 1	30585461	30585582	MPPED2 2Ã~Â¿Â½	GGCTCGGGCCGCGCGGGAGCCCGGGGATCGGGCCAACACAAT GCACCCAGGCCTAGGCCGGGGCGGCTCGAACACATCACCCGGGA CTTTCTAGTAAACAGCTCGCTGAGCCCTCGTCC	94
chr1 1	131910529	131910600	OPCML	CGCTCCGAGGGCGCACCCGGAGAAAGTGGCGGTCCAGGGATGGAG CTGCTGCCATGACAACCCCGCGGTCGG	95
chr1 2	6075235	6075514	ANO2_ VWF	CCGCACATACGTGACACAGCCCCGAAGCACCTAAGGGACACCACC CAGGACAGACCGTTATCCCCGGCAGGGCAGGACGGGGCAGGGG GCCGACTTACTGCACGCGCTGTGGTCGGTCCAGCCGTACAGCACCA TTCCCTCCTGGGCACAGTCCGGGCGTACTCCAGGAGGGCAGGGC AGGCGCACTCCAGCCCCCAGCACACTCACACAAAGTCTTCTACA CAGGGCCACAAAAGGCTCGGGGTCCACCAGAGGGTGGCAGCGGG CAAACACCG	96

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 2	6555180	6555371	IFFO1	TCGGAACCCACACCAACTCGCGGCCCGTTGTGAGTGGTATGACACA GAGAGACCTGTCCCCCTTTCCCAATCCCTACCTCCGCTTGTACTCGT CCCGCTCCCGCTTCACTTTGGCCAGCACGTTGTAGAGAGCGCGGAT CTCGGGCGTGATGGTGTGATCTGGACGCCACCCATCCGGGTGC ACCCACG	97
chr1 2	6555638	6555744	IFFO1	ACGAAGCCGGTCTGCACTGCCTGGTCGCGACGACCCAGGCCCGC CGGCCCTGCTTACCCTCCTCCAGCGCTTGCTGCACTTGCTTCTCAA CAGCCGGTTCCGGCG	98
chr1 2	6555890	6556364	IFFO1	CCGGCCGCGAGAGAGGCGCGGGGGCAAGTCTCTCCCCCGGC GAAGTGGTCGCTCCAGTGAGTCCCCAGTGGCCCGCCAGGCC CTGCTGCTCCTGCTGCAGGAGGAAGAGGTTGGGGCCGAATAACGG ATTCATGGCTGCGCCTTCTGCTGGGAGATGCAGACCGTGCAGGA GCAGGGATGGAAGGCGAGCCAGAAGAGCCAATGCGGCGCCGCG GGACAGAGCCGACCAATCAGGCGGCTCGGCAGCGGGGACAGGT CAGGGGGCGGGCCGAGGGGAAGCCAATGACAGGCTCAAATTGGA GGCCGGACCTGGACCTTTCGGGTCTGAGGCCGAGCCCTGTGAT GAGGGGAGCCACCGCTGACTCCAGCCGGGGTGGCGTAAAGCC CAGGACCTCCAGTACCCATGGGTTCTGGTGGAAGCCCATCTCCC CTACACGACTTTTTTTTTTTTGTAGACCG	99
chr1 2	6963339	6963534	PHB2_P TPN6	CCGGTGACAGGTAAGGCCACCAGGGGAGAGGTCTGGGCTGAG CTTGGGACTGCAGAGGGGGGATGAGGGTGGGTAATCGGTGTGT GTCGCGGGTGGGAAAGGCTGCCGGGGGTAGGGGAAGGTGGCTC AGAGGCGGCGGGCCGACGGTCGAGGGGCTTCGGAGGGCCTGCTT GGACTGCAACCTGGGCCTCG	100
chr1 2	24902884	2490345 7	BCAT1	GCGAGCTACCGAGACCCGGGTTCCAATCTCCCCCTTCCGCAAC GCCCCGGTTTCAGGTACCTGGCGGGCAAGGGCCGACGCGAGCG AAGCGGGTGGCCATGGGGAGGCTGCGGGGACGCGGGGCTGCAG AGAGCGGCAGTGGCACGAGCGCGCGCTGGAAGCGAAAGCAGG CGGTGTGGCCAAGCCCCGGCGCACGGCCCATAGGGCGCTGGGTAC CACGACCTGGGGCCGCGCCAGGGCCAGGCGCAGGGTACGACG CAACCCCTCCAGCATCCCTTGGGGAGGAGCCTCCAACCGTCTGTC CCAGTCTGTCTGCAGTCGCTAAAACCGAAGCGGTTGTCCCTGTCAC CGGGGTGCTTGCAGAGGCCGAGAATGCGCGCCACGAACGAGC GCCTTTCAAGCGCAGATATTTGCGGAGCATCCTGTTTATTAAACA ACCTCTAGGTGAATGGCCGGGAAGCGCCCTCGGTCAAGGCTAAG GAAACCTCGGAGAACTACATTAGGGCAGCTTTTCCACCGACTCCA AATCCAAGTACAAAAAGCAGTTTCTGCCCTCG	101
chr1 2	33439383	3343960 2	SYT10	CCGCCCTGGCTGCCCTGTCCCGAGGGAAGATGCCCGAGCACTTCT CCCACTCCACCTGGCCGGCGAAGCACAGCTCGGTGACGATGTGCA GAGCCTTCTGGCACAGACTGTTCACTCCGTCCTCCTTGTGGAACTC ATCGTTTGGCTTTTCTTCGTTTTCTTTTTTTTCCAGTTAGCCGTCT TTTCTCTTCCCGTACCTCTAACCCCTCTGGCG	102
chr1 2	33439838	3344001 9	SYT10	CCGTAAAAAGCCAAAGCAAGCCCTCGACTCGCAAGCACGCCCCC TCCTCTCCCGACGCACTGGTGTTTCTGGCGGGTGCCTGGCGGCGA CGCGTCCAATCGCAGCCCGCGCGGGCGCTAGGTGACAGGCGGC GGAGCGCGCAGACCCGGCTCCCCGCTCCTCTGAAGAAGGGAATC G	103
chr1 2	54047491	5404771 1	HOXC4_ HOXC5	CCGCCGGGAGGACTCGGAAATACAAAAAGGAGCCGAAAGATTTA AACAGTCGGAGGCAGAGGCGTCCCGAGGCGGCCAAAGCGGAAAT CAATCACGTAATTAACAGGGAGGGGACGAAGCCCAAGGCTGG	104

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 2	57619711	57619959	SLC26A10	GGGTCCCGGGTTTCGGAGGAGGCGGCCAAGGTGCAGGCCGAGGCT GGCGAGCGGCTTAGGGACGTGGCTCGCCGCCAGGACCAGAGCG TCGGGCTGTGGAGGCTCGGGGCTCGCGCTTGTTCGGGACAGGGG CGTGGCGCCTGCTGCTGGCTCGGCTGCCCGCGTGCCTGGCTGCC CCATTACCGCTGGCGGGCCTGGCTGCTCGGAGATGCGGTGGCCGG AGTGACCGTGGGCATCGTGACGTGCCCCAGGGTGAGAGGCCCTA ACAGCAGCCTGTCCGGAGCACAAGCTCTAGAGGGCTTCCGGGAGG AGGCTTAGGGAGCTGGGAATCCG	105
chr1 2	63150667	63151102	AVPR1A	ACGGCGATCTCCAGTTTGCCAGCTCCTCGTTGCGCACGTCCCTCG GTGGGCGGTTGCCCTCCCGAGGGCTTCGGCTCCCGGCTTGTGTT GCCAGCGCCGGTGGCCAGAGGCCACCATGGGCTGGAGTTGCCCGA GGGCCCCGCGTCGGGACCGCGGAGAGACGCATGCTGTCCATGCA GCTCCTACTCGGCCCTCTTCGGAGCTCCAGCCCTCGCGGGCCGCTC CCTCCCCGTCTCGGAGGACTTGGGCTCCTCGTCCGAAGCGCAGGGT CTTTGGCGGCTCGCAGCTTGCCGGGCTCTGCGATCCCTCCAGTGG GCGTCTCCCGAGCAGCGTCCCGCTGCCACTGAGCAGCTCTCAG CAGGGTGAGCTGGCCCTCTCCTGCTCTGCCTTTTTCAACTTCGG CGAGGTCGGGAAGGTGAGTCCG	106
chr1 2	65882147	65882322	HMGA2 _ENSG0000228144	GCGGCGAGGTCTTGGGGCTGGCCTTTCTGCTGCTGGTAGGAGGA TCATGTGCTGCTATTTTCGGAGGCTCTGCCAGTTGGCCCTGCCAC CTTTCTGTTCTACTGAAAGCAGCCAGGAAGTGAAGAAAGAGGA AGCCTCGGCTGTGCTCCGGGCTGCGCTGCCAGGGTTGCG	107
chr1 2	69657752	69657934	LRRC10 _BEST3	ACGTGCTTCTCATGTTTATGAATAAACATGGATGACTGAGATGA TTAACTGGCTGAATGTCCTGGGACGGCGTTTCGATTACAACCTTGTG CTGTTTTCTAAAGCCTCAGCAGCGCCCTTGGCTACCAGATAGCCTT CTGACCCACCTCCACTGTGTGAGGGTCAGATTCTATTACATCG	108
chr1 2	80919125	80919187	LIN7A MYF5	TCGTTAAGGAATGCATGCCGGTAGTTGCTGAGATGTACAAATAAGC ACCAAAAATTAACCACG	109
chr1 2	103799993	10380029	NT5DC3 _STAB2	ACGAAACAACAGACTGAATAGTACAGGAAATGTCACG	110
chr1 2	103800133	103800197	NT5DC3 _STAB2	ACGATATCATTATGTTTTGATATGTAACGTTAACAAAAAGATCACT TCAACCTCTTCTCCCG	111
chr1 2	113463473	113463622	LHX5_S DSL	TCGGCCGGGACTGCGCCTGCGAAGGCGGGCCGTGCGCGAAGAA GTCGTAGTTGCTTCCCGCGCGTAGTAGTCGCTTGGTAGTCTGCG GAGGGGGAGCGGGAAGGAGACAGGGCGCGGTGAGAGAAGGCGA AGTAGGCGGGGGACCCG	112
chr1 2	113464137	113464171	LHX5_S DSL	GCGGTTAGAGACACGCGTGGAACCCCGGGGGCG	113
chr1 2	113464275	113464499	LHX5_S DSL	GCGGAGGCTGACAGGCCCGGGAGAGGAACCGGGCAGGGACAA ACCAGCGGACAGAGCAGAGCGGAAATGTTGAGACCGGGAAGC GACCTGGCCGGGGAACTGGATCCGGGCGCGGCAGGAGCGAC TGGTGGGTGGGGCGGGCGGGCGGCTTGGCGCCCTAAACTCGG TCCCTGCGCCCTACCAACCCAGTCCAAGTCTTCGCTCGCCAAGTA CG	114
chr1 2	113479599	113479898	LHX5_R BM19	ACGCTTTTTCTGGCGAAACGGAGAAAAAACGCCGCGGAAACGGTG CGCAGGGTTGGGGAGTATAGTTCTGATTGCAACATAATTCCGCAA GCTTTTTTATTTTTATTTTCCCGGACGCGTTGCGTCGGAAGAA ACGCTTCTAATCTTTCTAGCTCCCTGGATTGAAGTTGCGGGTCTT GGGGCGAGGCTTAGCTGGTCTGGGGTCTTGGTGTCCACAGCC	115

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 2	114402994	1144035 28	RBM19 _TBX5	CCGGATACGCACCCGCGAAACGTTTCGACATCGCCGCTTTTTGTTTT GTTTTGCTTTGTTTTTTAGTCG CCGTTTCACCCCATGTGACACCTTATTTAAAAATTACCAGGATCTAC TGAGGGGGCCGACTTGAGCGCCCACTGCGTCTGGGTTTTGGGCGC AGAGCGCAAGGTGAGGCTCCTCCCTCTGCTGGGCCAGGTTGTA GCCTGGCGAACCCGAGGCTCCTGGTGCCCTCCGGGCAGAGCTCTG TGCGCTCCCAGCGGCCGGTGATGGCGCGCCAGCCAGCCAGGCCCC GACCGCAAGACAAATGGTGC GGCGCGCGGGTCTAGTCGGCGCGC CGGAGGAGGCAGGAGGAGGCAGGAGGAGGCGGGAGGAGGCGGA AGGCTACGGAAGATCAGAAGAGGGGTCAAGCCATCGCTCATGCCG GCCTGAATCGGCCGCTGACCTGGCCCTTATTAAGATGCTGGGGGCC GATTCTACACATAGTGCAGAGGGAAGGAATTATCTAGGCCATTGT TAGCTGACCCCAAACGCGCGATAATTGAGATTCTCGAACATTT AAATAGATTTCAAAATCCTTTGGCCGTAAGATAACCG	116
chr1 2	114448549	1144488 14	TBX5_T BX3	GCGCGCGCGCACCCAGCGCGCAACTGCTCCATCAAGCATCCACTG GCCTCCAGCCCGGTTTTCCGGTTGTAGCACTGGCGGCCCCAGAGTG GACCCGATAAGCTATCGGCGCGGCCAGGAGGGGCGGTGAGCGG CGAGTCAGGGCACCTCGGACCGGCTCCCGGCTCCCGGTCCGGCTG CCTGCCAGCGGCCGCTCAGGACAGAAGCGAGATGCCTGCCTAGGC GTTTCTGGTTACAATCACCTCACACACCGCCTGCATTCCG	117
chr1 2	122035274	1220355 58	MLXIP_ BCL7A	ACGTGTGCGCACACACATGATCTGGTGACTTGGTTTCTGCTCCATTT TCCCTGCAGAAAAACAAGAATAAGAAAAAAGGCAAGGACGAGA AGTGTGGCTCAGAGGTGACCACTCCGGAGAACAGTTCTCCCCAG GGATGATGGACATGCATGTTGAGTGCCCATGGCTGCCAGCCTCTC CTGCCAGCCCGGGCCTTGCCCAAGCACTCGGTGATTTTTGTTT CTCCAGCAGGTTTGTTCATTCCAGGCAAGGGGTAGGAGGGCTG GGCAGGGCCCCG	118
chr1 2	124457139	1244574 99	NCOR2_ ZNF664	CCGATGTGCAGCTTCAGCCTTCTTCTGCAGGGTGATGGCGAAGAG GAGGAATTTTTTAAAAAACAAAAAACACAGATTATAAATAGAGG CTTCCCGGAGCAGCGGGCACCTGCCAGCCAGTCCAGCATGCTGA TCCTCAGCACGGGGGAGGGAGGCCCGGGGCCCTGCAGGCCCTC CCCACGCTGAAAAAAACACAGAGGAGCCTCAATACCCCAACAGC GGCCCCAGCAAGCCAGCCAAGTTTCGATTTTAGCAATGCGCCGG GTCCACTGAAGCCTGCTCCCGGCAGGCGCGCAGGCCTCGTCCCC CAGGGCCCAGCGACGTGGGCACCGTCCCAACAGCCAGCCG	119
chr1 2	131773905	1317741 04	MMP17 _SFSWA P	GCGAGGCCCTTGAGGAGCTTACCAGAATAGTGAGGGCCACAGAGG GCCAAAGACCCACAAGTGGTAAAGGACAGGTGGCCCCACTCAGGA AGACACTTTCTCAGGCAGAACCGGAATGACAATGGGAGGCCAGTT GTGGAGAGCCTGGGACGCCAGAATAAGTGAGCACGAGAGACCGA CAGGATGAGAGCCGATTTCG	120
chr1 2	132904700	1329050 56	CHFR_Z NF605	GCGAGAGCCACCGCGCCCGCCTATAAAAAATTTTTAAAAAAGG ACAATGACTCTAGAGATTCCCGGCAGAGTTCTCTGGGAAGCTTT TCCTACCGAAGACGCGGCCTCAAGTCATCCCAAGCCGGGGCTCC TGGGTGGCTTCTCAGGAAGCCAAGCTCCCTACCCCTGTGGCGACGC CGCGGGCGGAATGCGCATGCGCGCCACGAGCCACAATCGTAGGGT TGGGCGCGCCCTGCCGGCCACGAGGGGAGCGCAGGAGCTGAGC GCACCCATCAGCGAAAAGAGCGCGCCTCCCGCTTTTTCTGAAC CGTATCTCTAAACTATAATTTTGAGATCAAAAGTGCG	121

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 2	24949058	2494915 9	BCAT1	CAGTGCCCGAGGCGGCGGCGAGTACACGTGGCGGGCTGGATTGC AGACCGGCCCTCTCGCGGCGGAGACTCGCGACCTAGCGGATTGCA TCAGCAGGAAGAC	122
chr1 2	65121476	6512167 8	WIF1	CTGGCGAGGCCAGCAGTCAGCGGGGCAAATAGAGCGAGAACAGA AGAGCGGGAAGGGCTGGCGCGAGCGAGGTGCGAGCGAGGAGTG GGGCCCGCGAGGCCTGGGCGGCCGCCACTTGGGGGCGCTGTGGG GCCCCCGGGGGCGGGGCCGCGAGGGACCCCCGAGGCTGCATTC ACAGTGGGTGCGCCCACTGGAGCGCC	123
chr1 3	20946079	2094627 4	XPO4_L ATS2	ACGGTGGAGAGACGCGGGAGGGTCCGGAAAACTGCGTTCTCACA AGACCAAAGGGAGGGGAGGGAGGGGAGATGTGGCTGCAAGTG CAGTTGGAGAGGGTGTGAAGAGATCGGGAGTCTCTGCGAGGCTC TGGAGCACCCGGCGCCTAAGAGGCTAGTGCGCCCGTGCCGCTGC GGTAGGACCTGGCGGTCCG	124
chr1 3	24745936	2474615 7	RNF17_ ATP12A	GCGCCCGCAGGGCCCGCCACCGCTTTGCTTACGCCGCTGCCCCGTG GGCCACCCCGCGCGCAGGGTCCCCAGCCCGCCTCCGCCACAG CCGGCTTTCCCGCGCAGCCACGGAAGTCACTGCCCGCAGCCGGCA AGGGCTCCAGCTGGACGAGGGGGCCTTCCTCGCTCCGGGATCCC TGTCCCACTGTGTGGCTTCCCGAGGCCTCCCTTCTCTGCG	125
chr1 3	24746281	2474656 0	RNF17_ ATP12A	CCGGCTGAGATTAGAGAGCCTGGCGAGGTGTGGGGTGCAGCAG GGAGAATGGGTGTGGTGCCTATGTGCGTGTGGTCTTGTGGAG ATGGATGCTCTCCGGGTCAATCTCTGCCTTCTCGGGGTGCCCCCTCA GTGTCGTGCTGAAAAGGCCTCCGTCTCTCTGGTCTTGTGTGCG CTCCCCACGTACCCGCTTCTCTTGAAGGGCCGCGGGCGTTGGC GAAGGTGGTGGGACTGTCTGAGGATCATCATGGGCAGGGAAG GGCGCGCG	126
chr1 3	50843351	5084355 1	RNASEH 2B_DLE U1	TCGGCGCCCCCTCAGCGCCTCGCACTACCTCCTCTGGGGAGTT CGCCCCGCGCGCGGTCCGCCGACTCCTGGTCCCCACGCCCCGCCC CGCTCTCGCGCCCGGGCCCCGGCCGGGCCCGCGCGGGCGCTGAG CGACGGGTGGAGCGGTGGACACGTGGTCTGGGTCCCGCGGGTTCC CCGGGGGCGACTGGACCG	127
chr1 3	52739274	5273948 7	LECT1	TCGGGCGGGAAACAGCTCGCCCGGGCTCCTACGGGTGCCCTTTC GCCGCGTCCCTCCCGAGGGTCTTTGCACTCGGGCGTGGAAGTG GGATGAGCAAACCCCGCAGCACAGGGCCTTCGCCCCAGGACCTGC ACCTCTACCGGCCACGGGACGTCCCTCCGACCCGCTGTGGATG CCGTGACCCCTGCACACTCATACGCTGGGGCG	128
chr1 3	99955868	9995613 4	ZIC5_CL YBL	GCGGAAATCGGGGCCGGGGCAAGGACGCAGGGGCGTGTGCCCCA CGTTTCTGGCCCGGTAGCCGCAACTCCTTGATGTAAACGAGATT TGGCCGCGCTGCGGCGTGTGGGAAAGATGATTACACTCGAAAG GAATCAGACTCCTTGCGGAGCCATTACTGTGCCGTCCGCACGC GCAGGTTCTGGCCCGGCTTTCAGCAACTCCCCGCTCCTCGCTAACCA CTCGCTCGTAATTTGTGGGCCGAGTGAGCTGCGCCCG	129
chr1 3	99997312	9999759 4	PCCA_ZI C2	CCGCGAGGTCCCGGGTTTCGCCATCCTGAGACCCCCGCGGGATG GCCCAGGAGGGGCGCGGCGGCCCTGAGTCAAGGTGGGCGGGGGC AGGTGCTTCCCTCCACCGCTTGTCTATGCCGGCGCGGTCCCCAC CGCCCGACCTAGCCCGCGCGGGCCGAGCACGGCGGCCGCGCTTC GCACTCCTTCTCCACCGGGTCCGACGGCCCGGCTTACGATTCCC GGGCCCTCGGGCATGTGAGGGACTTGAGTGAATGCAGCTCCCTCA ACTACTCCCG	130

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 3	108495649	108495841	MYO16_TNFSF13B	GCGCGCGGGGAGGGGAGAGGCGGGGCCGCGGGGACTGTGTCG CCGCCGACGCCGCGGCTGCGGGTCGAGAGCGGGCAGAGAGAG CCGCCCGGAGCGGGTGCGGAGCAGTCCCAGCCTCCAGCCGGC CTGGCTGCGCGCAACCGCGCCGGCCCCGGGCACAGGGGCAACTGC CGACCCCTCTCACCCG	131
chr1 3	108495969	108496120	MYO16_TNFSF13B	ACGCGGCGGGGAGCCTCTCCGAGTCTGGAGGTACGCGGGGCGC AGAGGCTGTTCTGCACCGCCGGGCTGGGGACGCCGGGAGGGTGC CCCGGGTCGGACTTGCGGCGCTGGGTCCCCACCCAGAGTTCGCCA CGGTGAGGGTTGGACGCG	132
chr1 3	110525911	110526052	RAB20_COL4A2	ACGCTCCTGGTGATGCATTGTTTCAATCACCACAAAGCAAACCCCA AGTGAGATCTTCCAACCACAAAGCACCTGCTCCCAACCACACCTGC CGGGGGCACGCTTTCGAAGAGGAATGAGACTGAGACCTGTGCTCA GACG	133
chr1 3	111893313	111893725	SOX1_T EX29	GCGTCCGGGAGGGGATCACATTCCTGCGCAGTTGCGTGTGCGG GAAGTGACTTGTTTTCTAACGACCCTCGTGACAGCCAGAGAATGTC CGTTTCTCGGAGCGCAGCACAGCCTGTCCCATCGAGAAGCCTCGGG TGAGGGGCCCGGTGGGCGCCCGAGGCCGCTGGAGGGCTGTGGG AGGGACGGTGGCTCCCCACTCCGTGGCGAAGGGCAGGCAAACCA GAAGCCTCTTTGAGAGCCGTTGGGATTGAGACGAGTAAGCCAC AGCGAGTGGTTAGAAGTAGGTTAGGAAGAAGGGGAGGTAAGAAA GCCGAGTAGGGTTCTGGGCCGAGCCGTTCACTGAGACAGGAACC CTGGGGGAGATGCGTGTCTCCCTGGCGTCTCGGTGCAATGCCCA GAGAGCG	134
chr1 3	112062727	112063125	SOX1	CCGGGCCAGGGCGCAGATGATGGACTCAGAGCGCCAGGGACCCT AGAGAGAGGAGCACTCCTCAAGAGCCCCCTGGCCATCACCCGAGC GCCCTGGAGCGCCATCACCCGAACGCGCGCTCCAGGCCCTCGAAC AAGGCCTCTGGCTGCCAGAGCGAGTGAGGGGCGCAGAGCGGCA GAGAGCGGAGAGCCCCGGTGTCTCCGCGAGGGCGCGCGCGGCCA GCAGACGGCGATCGAGGCGCGCGCCACGGCACGGCCAGCGCAGA CACGCCGCGGGGTCTCGGGCCGAGCCGTGCAGCCGGGCCCGCTG CCTCTTTCCCTCATGGCTCCGCGCGGGAGGAAACCGGCCCTTCT CCGCCCCCCTCTCTCGCTGCGGTGTCCCCAGCACCCCCG	135
chr1 3	112782153	112782378	MCF2L_ATP11A	ACGTCTGCTCGCCGTGTGAGACTTTGGAGTGGGCTTCATCCATT CATCCTGATCGTTCCTCCATGAGACAGGGTCCCTTTGTTGCTGGCTG GAAGCGGCCGGAAGCGTGGGCTCGCTGTGGCATGGGCAATGCC ACACGGCTCCAGGGAAGCGTTCAGCTTTCAAACCAAGTGTCTGGGC TCGTGGCCACTCCTGAAATTCAGTTGCCGTCTTTGAAGCTTCG	136
chr1 3	27968877	27968960	CDX2	GGTAACCGCCGTAGTCCGGGTAAGTGGGGGGCTGACGAAGTTCT GCGGCGCCAGGTTGAGGCCGCCAGAGTGCGGCACGGAGC	137
chr1 3	36346455	36346548	SPG20	GCCTCGCTCCCGCCACAGAGCCGCGAGCACGCCGCGCGCAGCCT AGGTACGTGAGTACCACGCGCGCTCTGCCAGCGGATTTCATCA CC	138
chr1 4	20625539	20625748	RNASE1 2_ORF5 1	CCGATTACACAGCATCAGTTCCTCTGAATTCTGCATTCTGAATTAA AATCCTGATTTCGAATTGGCATTCTTTTCGGTTAGGCAGGGAGGCCT TCTCGCTCGGGTCTCCTACTTTATCCGTTGTACTGACTCTCTGGACC CCAGTTTTTGCACTGCACCATTTGGGTTCCCGCAATCAGGAAAGCTC AGTTCTCATCTAAAATACACG	139
chr1 4	54776350	54776707	GCH1_S AMD4A	GCGGCTCTGCTCTCCACCCAGTGGGGCTGAACTAACAAAGTTCCCC TTTTGCTTTTCTACCAGAACCTGTGGTTTGCAACCCCGGGGGCAG CAATAGCATGCCAAGCCGCACCCACAGCTCAGTCCAGAGGACCCGC	140

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 4	56238737	56238785	TMEM260_PELI2	TCGCTGCCCCGTGCACACTTCCCCACAGAACATGCTGATGTTCCAGC AGCCAGGTAGGGCCCGGCGCTTCATGTCCCCTTGACACAGAGGGG AGGCCAAAATAGATGCCCTAGCAAACCCAGCCAGAAAGTGCTTAG CCTCGACTGTCACCGTGCATTCTTTGGAGCTTATAGAAGCCTTTCCT TTTTTAAACTGTGCCTTGCCAGCATGAATAGCGGCG ACGAAGCTTGATCTAAAAGCCAGGTGAGTGGCAGATTCCGGGCC CACG	141
chr1 4	56798239	56798843	OTX2_T MEM260	CCGAGGCCGACCCGACCCCTGCACTCCGCCAGGCCGCGAGGTTTCC CAGCGACCGGCGCCCCGGCCCGCGCCGACCTGGAGGCCTGACTG CAGGGCTCGGGCGGGGCCCTCTCTCGGCTCTGGCTGGCGGCCAC TCCCGCGGGCGTACAGGCCTCGCCACCGGGCCTCGGCCTTGCCGC GGCCACAGCGCCCTGGGACCGGCGCCCCGAGGCCTGAGAACTA CGCCCCGGGGGCGCGGGCTGAGGCTCAAGAGAGGTCCTAGGTGC GGGCCAGGGATGGAGCCAGCCAGAGAGAAAGGGGAAAAACCCGG CAAGGCAAGAGCCTCAGTCTCGCCCTGCCTGGCCCGCCAGGCTGT GAGTGGGGCCATTGGGCAGCGCCAACCTGGGGAGTCCGGCGTCT GCCCCAGCTGGGGGCCCTCGGGGCAGAGATGTGAGTGCTGTTCCC AGGTAACCTCGACTGGGCACTGGGGAGTTAGAAAAGCCAGCTCTT TAGCCAGAGCGCCTAGGGCGCGGCGAGAGCGGGCCGCCGGCA CCACGTTCTTCTGGCAGTTCGCCCCAGCCTCCAGCGTCTTGCGC CTGTGGCGGCGGCAGTACG	142
chr1 4	56809049	56809322	OTX2_T MEM260	CCGCCCCGCCCCGAGCCACGACACCTCATTGTCCTGGAGCCTGGGA AGGGGGTGC GCGAGCGCGCGGGCGAGCCCTGCCTCTCCCCGCCAG AGAACAGCTGAGGGGCCGCGGTCCAGCGGGAGGATTCCGGTCCC TGGCCCCGCCGCGGCCTTGGGCGGAGCAGGGGCCACTAGTGCCA CTTCTGCCCCCCCCAGGTGCGCGCGGAGGGCTACGTGGGGCGGGC CGCGACCCGGCAAAGTCATGTTGAAAAAACACTCTTACGTTCCGT CG	143
chr1 4	59741061	59741184	RTN1_J KAMP	ACGCTGCTATTAGGACTCCCTTTGTTCTGGCTCATTCTCCATGCAG CATCCGGAAGGATCAATTGAAAACCCAAAGTCTGATCCTGCCACACT TCCCTTCTTCCCCCTCTCCAGCCAAGCCG	144
chr1 4	77025610	77025968	IRF2BPL _VASH1	ACGAAATGCGTATCCTCAAACGTTCTGGCAAATGGTGCAGCAG AGGGGTCCGCTGTTGGCCATGGGGGAATCCGGAATGTTTTGGGGG TGCACTTGGTCCATGCCCGGTGGGCGCTAGGCGCGGGGGCGCC ACCTGTAAATTCAGGTCCCCGTTACGTGATGCCAAGCGGCGTGCC CCGGCACGGAGGCCGCGGAGACTGGGCTGCTGCTGTTTCGCCGCG CCGACGCAGTGGTAGAGTGACGGAAGTCCCATCTTGGGCGAGT GCGCTGTGCCAGAGTATCTGCCACCGACATGAGAGCGGCCATAG GGGACGGACCGTTCTGGGGGGCTGACTCAGGTGGGGTGGTCCG CCGGGGCTCGACGAATCAAGGCCACACAGGCAGTGGGAGCAAAG GCAAAGCCCCGGCAGGTGTGGGGCTGGGTCCCTAGGGGTGGAGGA CGGCGGGCGGGCGCCCTGCTCGTGCTGCGAGTGCCAGCCCCAGC CCGAGGCGTCCCTCGCCTTGCCCGCCCTGCTCATGCCGGGCCTT CCCCACCCGACTGCGCCAGCCTCTTACCCGGTCCCTCCCGCTT TACCAATCCCTGCCCCACGCAGTCTCTCAGAGCCCCAGGGCTCTTG CAGCCCTAAGGGGCTGGACTGTGTTGCCCGCCCGCACTATGGGAT GCCCCG	145
chr1 4	92687197	92687518	LGMN_ RIN3	TCGGCCCAGCTGGGAAGCCAGGCAGGGAGGGGGACGGGCCCCC GCAGGCTCGCGGCAGAGACGGGAAAGGCGCAGGTGCCGGACTCG	146
chr1 4	92687724	92687988	LGMN_ RIN3	TCGGCCCAGCTGGGAAGCCAGGCAGGGAGGGGGACGGGCCCCC GCAGGCTCGCGGCAGAGACGGGAAAGGCGCAGGTGCCGGACTCG	147

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 4	93056040	93056042	ITPK1_C HGA	CAGACAGCTTGCGCGCCGCCACCCGCTATCCATCCAGGGAGGGGC CTGGGCCGGGAGAGGGCGCCTGAGGAGACAGGGCCCCGCGTGAC CCACAGGCCCTCGCGTCTCCGAGGACTTCATCTGCGTGTCTGTAC CTGGAGCCCCGAGCAGCAGGCGCGGACGCTGGCGTCTCGGGGCG CCG	148
chr1 4	97033237	97033740	VRK1	CCGAGCAGCAGCCACCTCAGGGCCAGGGAGCCCGAGCTGCGGGA TCCGCCGCCCCGGGGCCGAGCAGCTTCAGCTCCTTGCGCTCTGCG CCGGGGTCTCGCGGCCGCCGGAACCGCTCCTTCAGTTTCGCTAT GCGGAGCGGGCGCGGGACCCAGCAGGTGAGGGCCAGGGCAG GTGCTTCCCTCGCCCCGGCTCCCGCCCCAGCTCTGGCCGGCCCA GCGCGTCTGCTCCCGTCTCGCCGTGCTCTCGGCGCTGCATGTCCC CGGGGCGCGGCGCAGCAGCTGGTGCCGCGGTGGGCATCTGTTG GCCTCTCTGTCCCAACGCGTGACCTGATCGCTGCGACAGCGGAAT CCCACGGTGACAGGCCAGAGCTGCGCCGAGAGCCGCGCTCCAGC TCCTCCCGGGCTGGGTTTAGGGTCCACAGCTCTTGCCAAATCCA GAGGCTGGAAGGGACGCGAAGTTCTTCGTGACCCAGCTTCTCAG GCAGCG	149
chr1 4	100390711	100390944	WARS_ BEGAIN	CCGGCACATCTTTTCCACCAAGTGTGCAGATCTGTGCCGCTCTTTT GGGGGCTGTGTAGCGCTCAGTGTCTGACACACACCATATTATG GTAAAAGTGGCTGCGTCTTCTCACCTCCACAGCGGCAGAAATTAT ACCTTTGAAAATGTCTGCTAATTTAATGGTGTCTTGTGAGAAC AAGTGAGTTTACATTTACGTACAGCTCTTTAGAACGGGCCGCACT CG	150
chr1 4	105284109	105284340	PACS2_ BTBD6	GCGCTGCACCCGCTTCTGCAGGAAACGCATTCAAGCGCCCAACAC ACATGCACGTCCACAAAAGTGGCCTTCCACCCGGCCACGGCTCGAA GCATTTCCGAAGACTGAAATCACACAGAGGGTGTCTCTACTGACAG AAGAATCACACCGGCAGTCAGGAAGAAAGGCGCTGACTATACTCC TCTACTAGTAAGTCCACAGCAGGACAAGGAAAAAGCACAAAGGGA AGCG	151
chr1 4	105864223	105864329	TMEM1 21	ACGGTGACCAGGGTTCCTGGCCCCAGTAGTCAAAGTAGTCACATT GTGGGAGGCCCCATTAAAGGGGTGCACAAAACCTGACTCTCCGAC TGTCGCCGGGCCGCGCG	152
chr1 4	93788894	93788893	PRIMA1	CGGCTGCCCCGGGGCACTGGGGTGCCCCAGCTCTCTACTACCTCAC GCTGGCCCGCGAGAGGCAGCGCGGGAGGCGCCGCGAGGGAGCT CCCCTGGGG	153
chr1 5	22964409	22964554	CYFIP1_ NIPA2	CCGGCAGCCCTGCCAGCAGACTCCGACGCTGGAAGGCAGGAAGC AGCCTCCAGCCCCAGCAAGAAGGCAGGTCTTGGCCTTTGGCTGACC TCGGCCACGGTGCCCCAGGCCAGCAGGGCAGTTTCCCCTGCCCGG CAGTCCCCG	154
chr1 5	29103277	29103526	NDNL2_ APBA2	CCGAGGGTGGTCTGCCAGCAACAGCAGCTCCTCTTCCCCACCT CTCCAGCGCTGCAGGCTCTGCCCACAGCCCACTTGAGGAGGCGG CTTGAGCCCTGAGGTGGGGCTGGGCTGGGCTCCTGGACTCACAG CAGTGAACGCCACAGGCTTGGTGTGCGAGTTGGGGCCGCGAGGGC GACCCCTTCTCTGAAGCGCCAGCCGAGAGAGAGCCCCCTGAACCC CACACCTCCAGGAGGCAGCCG	155
chr1 5	41495639	41496110	ITPKA_L TK	CCGACCCAGGATCGTTTCTGGGGTAACCTTGCTAGGTGCGGG GGCGGATGCCGGGGCTTCCAGGATGTGGAGTGTGGGCGAGTGA GAGGCCCCCGCCCCGCTTTCGGAAGCCTGAGCAGCAGCTCC CGGGGCGCGGAAGCTCTGACACCTGAGAGCCGGTGCAGGCGAAA	156

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 5	41501068	4150130 2	ITPKA_L TK	GGGCGCGAAGCGCGGGCGCGTCCCGCTTCCCTCTTCCGCCCGCAG GGACTCGGCGAAGTGCCTGGGAGAGGGAGTGCCTAGGAGGAGG TCCTGCGGCCCAAGCCTGGGTGTAGAGACCGCCCCGGCTAAGGTC AAGCCTCGGGGACCTGGGCGACCCCGCCCTCCGAGCCGTCGG GAGCCGGTGCAAATCGCCGCTGAGGGCCCTTCCAGCTCCAAGGCT GCGGCTTCAGGCCCTTCCCCACCCCAAGGCCCGCGGGGCTCCCC GAAGTCAAACAGCCACAGCGGCG	157
chr1 5	41502802	4150305 3	ITPKA_L TK	GCGGTAGGCCTGATAATCTGCAATTTTAAACAGGGTGACCATCAG GTAATTCGATGTTTACAACAGTTCAAAAACCTCGACAGAGCATTTT CGTAACCTGCCACGCGTTCTTCACTGCGCAGAGCTGGAGCGCAAAAC CGGGGGCTTCAGATGCTAAGTCCAGGCTCTTGACAGCTCACTGGA GACGCTGGAATCACCTTCACTGCGCCTGTATCAGCACCCGCCACAC AGGCG	158
chr1 5	45135110	4513528 4	DUOX1 _SHF	CCGCTGCAGCAGATCCGGGACACCTGGAGGTATCCGAGTTCTTC AGGAGGCACGAGGTAAGCGGCGGCTGCCCGGGTGCCCGGGCCGC GAGGGCTAGGGCGGGAACCCGGCAAGGGCGTCTCTGGGCAGGGC CGCGGCCTGACGGTGCGGGGCTCGCAGGTATCGGCAGCTCGCTC CTTTTGTGCACGATCACTGCCATCGCGCCGGCGTGTGGCTCATCG ACTTCGGCAAGACCACGCCCCCTCCCCG	159
chr1 5	52791169	5279140 8	ONECU T1	ACGTGCTTTCAGACCTGGTGAGCGTGGAACCTCCCGGCTGCCCGGC CGAGTTCCTCAACATTCGCATCCCGCCCGGAGACCCCATGTTGCAC CCCACCAGCGCGGGGACGTGGTGCTGCCCTTCAGAGAAGCCGC TGGGACCCCGAGACCGGACGAGTCCACAGCAATCCCCG	160
chr1 5	67822046	6782221 6	PIAS1_S KOR1	CCGGGTGCTGGTGCGGGCCGTGGAGGCTCGGGCCGTCCCTGCGGTT ACTCCCAAGGCCCTCTGCTAAAGCACCCGGAGGCGGTTGCTTTCC AGAAGTACTGACGCAGACAGGGTGGACGCCGCGCGCGGGTCTC CGCTTGCCCCCTAGGGACGCCCTTTTCCCGCGTCCCCGAGAGACG CCTCCAGATTGAAAATCAATTACGCTTCGGGAGTAATTCGCCCTT CCCACAGTCACG	161
chr1 5	76347350	7634748 2	ISL2_SC APER	CCGGTAGCCCCGAGGGAACGAGGCGAGAGGGGAGAAGGCGA CCCCGCGCTGCTACCCGCGGAAGATTATGGCGCTCCCGGGTTCC AAGGACAGGCTGCGTTCTGTCGCTGCTGCCACCGCCGTTAGTCGCC GTGGCCGCTGCGCCCCCTGCCAGGCGGCGCCGTCGCG	162
chr1 6	11076366	1107654 6	SOCS1_ CIITA	CCGTCCCTCTGGCTTGGAGCTGCGGGTCCCCGCCCTCGAGCCGGAG CGCCGCGCTGGACACCCGCGGGGTGGGGGCTCGGCTGGGCTGAG CCACGGAGACGCCAGGTTCCCGCGGTGGCGGGGGCGCCGATCG ACGGGGAGGGGAGGGCAGTAAGAGCCGCCACAGAAAACAGGAAT TCATGGGGGGAGTGGGGTTGAGGATTAACGTTGAGTTTCAAGACA TCCCTCGTCCAGCCCACTCTGTGAGCTGTCTGGGGCTCCGCCTACA CACAGCTCCTCACCTGAAGCTGCTGGGTTCCCTGCATCACACG	163
chr1 6	11233151	1123337 5	SOCS1_ CIITA	GCGGCTGCCGGGTGCGAGCGGGCTCAGGCCTGTGGCCCTGCCTGA CGTTGGTCCCCATCAAGCCATGTGACGAGACCAGGCCACAAGAAA GAGGTTTCAACAAGCGTTATCGTTTCTGGAACCTCAACTCGGCGA CTTCCCCGAAGACCGGCTGTGCTGGCGGGCGGGGTGCGCACAGC GGGGACAAGGCTGCCCCCTTCTCTCCGCTGCCTCCGCGGCCG	164
chr1 6	22814288	2281488 1	HS3ST2	TCGGGCGCTGGGCGCGCTCCGAACCGGCGCACGTAAGAGCCTGG GAGCGCCCGAGCCGCCGCTGCCCGGAGCCCCATCGCCTAGGAC CGGGAGATGCTGGAATGCAACCGCTGTTCCCCGAGGAGCCGCT GCCCCCGGACCCCTGGCACTGTGCGCACCTGGTCAGCAGCCCC CGGAGAAGACGGCGCCCCAACGCCGACCCGCTGGCCGTGGCA	165

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
				GCGCCACGCGAGCCCTCTAGGCGACCGCAGGGCCACAGCAGCTCA GCCGCCGGTGCCCCCTCGGAAACCATGACCCCCGCGCGGGGCCCA TGAGGCCATGGCCTATAGGGTCTGGCCGCGCGGGGCCACCTCA GCCGCGGAGGGCGCGCAGGCTGCTCTTCGCCTTCACGCTCTCGCTC TCCTGCACTTACCTGTGTACAGTTCTGTGCTGCTGCGACGACCT GGGTGCGAGCCGCTCTCGGCGCGCTCGCTGCCCTCCGCGGCCCC AGCGCGGGCGGCCAGAACTTCTCCAGAAGTCCCGCCCTGTGATC CTCCGGGCGGACGCCAGCGAGCCAGCGCTCCAGCGCGCCCG CCG	
chr1 6	27228988	2722919 4	KDM8_ NSMCE 1	ACGCACTCGCTACCGAACAAGCCTGGCCCTGTCACTCCCAACTCAC CCCCACCCAGGGCTTCCACCACCTTAGGTCCAAGAGCCAAGCC CCTAATACGCGTATCTCCCGGGCTGCCCTCCGTCTGCTCGCCTCGCA ATCTTTGTGCTCAGATGGCCCTGGCCTTAGCTTCTTGAGTGACCTG CTGGCCACAGGGCCACTGCCG	166
chr1 6	51156004	5115634 9	SALL1	ACGCAGTTTTTTGGGGGAACTCCCGCCGCCGCCACCAAGGGCTAT CTCCAGACGGGCGCCGGGTGCAGCGCCGTGACCGGGCGCCCTGGC GCCGGCTCGGGCGCGAAATTCAGCGGTGGCAAGCGGAGGGTGGG CTTGTAACCAACCGCGCGCGCCGAGCCAAGAGTCGCGTACTGTC TGCCCGCGGCAAAGTTCTGTTTCTCCGCTTGGAGGGCTGTTCTTAC ACCGGTATTAAGAAACCGACTTCGCTAGCGACTGCAAGTGCTTGCG ATTTTGACTTTCGTCACAGTTGAGCGTCTTGCACTTAAATTA GCGCCCCGATGCAACAGTGCCCTCG	167
chr1 6	57620289	5762062 6	GPR56_ GPR114	GCGTCTCTCAGTGGAGGCCCTGGCTGTTCTGGGGTTACCCCTTGCA GTGCACAGCATGGCCGGGATGCTGGCATGGTGCTATCCTAGCA CCGGAAGCTGGCAGGTGTGAGGTGTGTTCCCGGTGTCAACGGA CACTGCAGGACGCAGGGCAAGGGTGACGCCGCGGAGCCTGAGCA TGGACGGGAGGCAGGCGGCAGGACCTGAAGTCTCCTGCTGCTTT CCGAGCGCCCTGAGCAGCTTCTCCTGGGATCCACGAAACCGG TTTGGGAGCAGGTTGGCCAGGTCGTTTGACTTTTGACTGGGGAG GAGAAGGCAGCCTCCCTTAGCG	168
chr1 6	70737654	7073791 1	MTSS1L _VAC14	GCGGCCGGGGAGCCAGCCCTGCAGATGTTACTAAGTGAAACCTGA TGTGGTGACATGAGAATCCACAGAACGTCTCACAACACCTGCC CGGGATGTTTTGGATTGAGTTTTGTGGTTATGACGTGAAGAAACCT CACATGTCAGGATAAAAATAACCTGGCTTCAGTACATAACGCGAG TTACAGTTCAACAGAACAGATGTGAAAACGTACGCCACCCAGTTC AGGCCACGAGGGTCCCTGCTCCACTCCG	169
chr1 6	86287871	8628821 2	FOXF1_ RF8	ACGCTGAAGATCACCTTGTAAGGTGGAGTTTCTCAGGCTTTACTC CGGGAGCCCTCCCTGGGGAGCAAGAGAAGGCAGGGTCAGTGCTG AGCCATCCCGGTGTGTGGACCTGCTACGCTAGGTCTGGTCTGGAC GGTGCTGATGGGACCGGGATGACAGAGCCAGGAGGGGCCAGAA TGAAAGTCGCAGAAAACAGAAACAGGCTACAACTTCTCCAGTCT GCCACCCCTCCCTTCCGTTTGTTCATGAAAACCCATTTCCAATCAG AGGACCACAGGCCAGGGAACATGGTGAGCCAGCCAAAGACACTT TCAGGACAGATGGTATAGAAACG	170
chr1 6	86578761	8657899 2	FOXL1	CCGCTCGCCCATGCTGTATCTGTACGGTCCCGAGAGACCCGGCCT CCCTCTGGCCTTCGCCCCGCGGCTGCTCTAGCTGCCTCGGGCCGG GCCGAGACCCCGCAGAAGCCTCCCTACAGCTACATCGCGCTCATCG CCATGGCGATCCAGGACGCGCCGAGCAGAGGGTCACGCTCAACG GCATCTACCAGTTCATCATGGACCGTCCCTTCTACCACGACAAC CG	171

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 6	86579316	86579457	FOXL1	ACGGCCCTCTCCGCCGGCGCCCTCCACTGGCCGGGGACCGCGTC CCCGAACGAGGACGCTGGTGACGCTGCCAGGGCGCAGCGGCCGT GGCGGTTCGGCCAGGCAGCGCGCACAGGGGACGGCCCGGGTCCC CTCTGCG	172
chr1 6	86579589	86579743	FOXL1_ FBXO31	CCGGCGGCGCTGTGGGTGCCTCGCTCTGGCCGCTCTCCAGCCT CCGTCCGCCTTTAACGCTTCCTGATGCTCGACCCGCATGTCCAGG GCGGCTTTTACCAGCTCGGGATCCCCCTCTCTTATTTCCCCCTGC AGGTTCCCGACACG	173
chr1 6	88703445	88703727	CTU2_R NF166	TCGTCTCCCGCGAAGGACTCAGGAAAGACACAAGAGGGAACCCA GCCCCACTGGCAGGGCGGCTGGGCCCGAGGAGCAGGAGGCAGAA CGAGGCACCCACAGGGTGGGTGCTCTATCGGCCTAGTTTCCAGTGA CTGCCAGCCTGGTGTTCAGAGAGCCAGCAGCCGGGAGTAGTGCCC GCTTCCCCACAGGAAGTTCTGTCTGCGCCACCCAGGGGCTGGT GCTGAGCAGCTTCTCAGCTGAAGGAAGTGGCTGAGGGCGATGGGT GTGGGGGCGTCG	174
chr1 6	58463491	58463554	NDRG4	CGGTCCCGCTCGCCCTCCCGCCCGCCACCGGGCACCCACGCCGC GCAGAAGCGGAAGCCAC	175
chr1 7	27540520	27540613	LGALS9 _KSR1	GCGGGGAGGTTGTCTCTACACAAATGTAAGCCTGGCAGCTTCCC CAGGAGAGTGCGGGTATGGGCCGGCCGGGAGAGGGCTGGCTGT TGCG	176
chr1 7	27540719	27540727	LGALS9 _KSR1	CCGTGGGCG	177
chr1 7	36808307	36808664	LHX1_ MRM1	CCGCGCGGGTGCTCCAGAGCATCCAATTCAATTTCCACTTCAATTT ATCAGCGGCGGGGAGCCGGGCGGGAGATAGGAGGCGGCCCTG ACACGAATTAGCCCGGAGATTGTCCGATACGCCTTGGCCAGGGCG CCGGGCGCGCGCTCGCCTCCCTCGCTCTCTTTGTGTCCGCTC GCCTCGCTCTCGGCTCGCCGCGCTCCATCCCGCGGCGCTGGCC CGGGCCGAGCGAACTGCTTTGCCTTTGCCACGTTGAGCGCGCCG AGGCAGCCGGGGCGCGGGGCTCCAGGACCCGTCTGCTCCTGGTG CCCCAGCTCCTCAGGGTCCGGCCGGGTACCTGGGCCG	178
chr1 7	36942478	36942694	AATF_L HX1	GCGAGTAGGGAGAAGGCTGGGAGTAAATCAAGGGGAGGCGGCG AGACCGAGGACCCAATTACGGCCCTGAATAACGGGGTAGCTGG TAAGGGGAGCTCCCGGGCTTGCGCCAGCCTCCTCCTGCACCCA GGCCCCGAGGGGCTCCCGCGATCCGCGAGTTCCTCCGCGGCGCT TCCTCAGCCCGCCGAGTCTCGCTCTTCCCTCCCTTTG	179
chr1 7	39165372	39165565	PLXDC1 _ARL5C	CCGGGGCGCTTCGGGGCTTGCCAAGAGACGGTGTTAGAGAAAGA GCATAACGCGAAGTCACAATCGCAGGAACTCGCAGCAGCCCCC ATCCCCGCGCTGGCTCCGTTAGCGGGGAGAAAGGAGGGTCGCC CAGCTTTCGCTCCTGGGGCGCACCGAAGCGCCGGACCCAAGAGG AGCAGGCAGGGACG	180
chr1 7	48544346	48544518	HOXB1_ HOXB2	ACGCTGTTAGCGGCCAGGCCTGAACCCAGTGGGATATTCTACTTC CCCATCCAGGAATGGAGGGGGTAAGGAACCCCAACAGGCTCGCC ACCATTTTTTTAAACCTCCTTCCACTGCTTTTCTCCCCCTCTCTAG CTGCCCTCACCCACCCACACGCTTACCG	181
chr1 7	48544736	48544865	HOXB1_ HOXB2	CCGGGCTGGAGGCTGGGGAAGGTTTGCTCGAAAGGAGGAGGAGG AGGAATTAATGTCGACTCCTTGATTGATGAAGTTTGAATGTCTCC AAGACAGCGGGGAAGGAAGTCAGACACTCGGCGAGCGACG	182
chr1 7	48754927	48755218	HOXB13 _TTLL6	CCGGTCTGCTTCTCCAGCCTCTGCTGGATTCTCTCCGACCCCTCT GGAGCGAAGCCCTTGGCCCTGCGTTGCATGCGGCACGGTGCGGG TTCGGGCTCTGCGCTGGAGCCGGGATGCCCTCCGGCGGAGGGTGC	183

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 7	50469276	5046952 4	CHAD	GCGTAGGCGGGCGCCTGGGCGTGAGCCCCGCTGCAAGGCTCAGCG TCGGGGAAGCACTTTTCTCGTCGACCCGGGGTCTTTTCCGCCAAG GAGCTCGGGGCTCAAGAACTCGGGACTGGGCTGTGGGCGGGGCA TGGTTTTCTCTCTGCGGCG ACGCTCGGCCGGGTGCCCTGGATGCGAGGCGGGAGGAAGCGGGG CCGACAGCTGGATGCGTCTCCCTGCGGTGGGCCAGCTGCTGCG CTTTAAAGGGGCGCTTGTGCGGCGCCTGCCGAGCGTGAGAGCCGC CCCGGCGTCGGTCTCCCACTTCAGACTCGACGCGCCGAAGCTGGCC CTGGGTAGACCCGAGCTCCTTCCCCACCCTCGGGCGCGCCCCACC CCTCTCTTCAACCCCGCTTGC TCGGCCTTGGGTAAAGGGAGTGGGGGGCCATGTGTGGAGCCCTCT GGAAGGTCTGGACTCCTGCTTTTCTTGGCTCTTCTCGTTCTCCAAC CACCCCAAGGTTACGACAGAGTCTTGGGCGCGTCTCCTCCGTTGT GCCGCGTGTGTGGCAGCAGCTGTTGGTGCTGACTAATAGGACTT CCTGGCAGCTGTGCCGGGCACACGTGGCACCGGCAGGAAGTGCCT CTCCTCG	184
chr1 7	57443192	5744342 7	MSI2_E NSG000 001663 29	CCGTCTGTGCAACCCCTCAGCCCCGCCAGAGGCTTCAGGAGCTGC TGGGGGTGATCCCCAGTGGTCCGCTGTGGTCTTTATCTCCGGCTC TGCTCTGCTGCTCTTTTCGCTTGTGGTGCTGGGCTGGTCC TAAGGAGGCTGGGTCTG GCGGCGGGGGGCTCTCAGGTCGCTGGGCTGGTCTTTTGTGAGCC ACCCGCTAACCTGAAAGGCCAGGAAGGAACGTCGGCGAGTGTCT GGGATGGGGTTTCCGTCCCGGGACTCCCTACGAGGGCGGTCCCC GGTAGCCAGAAGATCCGGCCGACTCCGAGCCTGGCCCCCTTGGGC GCCG GCGGGTGAGCAGAAGGGCCGTGCCAGGGCCTGGAAGTGCAAGG CCGCGTGGTGGGCATGGTAGGGAAGCGGAGCGTGGGCCTGTGAG GCGCGTGTGCGCCTGCGACCTCGGGACCGGGGCTCCCAATGAAC AGCGCGCACAGCTGGGAGCAGGGCTTGGGGAGCGGGGCTCTGCG GCCGGGGATCCGTAGAAGCCG GCGCGTAGGACTGAGAGCGCAGGGCGCGAGCCGAGGGCTCCGC TGACGGCTCCGGGTGTGACAAGAGCCAGCAGAGGACCCCATGG CCATGCGGGCCAAGCGCAGACGGCCCTCCTTGCGACCCCGCAG GCCGCCACATCTGGGACCAGCGGATCGTTGGTCTGCTGGAGCCGA TCCCGCCG CCGCTCCCCGGGCCCTGTCCCGCCTGGACGCTCCCTCCAGGAGCC TGCGCCCCGGCCCCGGGGTCAGGGTTGGGATGCGGGCTCTGCAGG CGCCCCGGCGAACAGCTCTACCTGGAGGCTGTCCCTGCCCGCTTA GTCCAAGGGCCTTGGTGTGGGGGCTCCGCTGTCAAGGCGGGGA ACCGGTTCTCTCGTTTCTCTCCCTTCCCCAGCGGCTTCAACG CCG	185
chr1 7	58328520	5832867 5	BZRAP1	CCGTCTGTGCAACCCCTCAGCCCCGCCAGAGGCTTCAGGAGCTGC TGGGGGTGATCCCCAGTGGTCCGCTGTGGTCTTTATCTCCGGCTC TGCTCTGCTGCTCTTTTCGCTTGTGGTGCTGGGCTGGTCC TAAGGAGGCTGGGTCTG GCGGCGGGGGGCTCTCAGGTCGCTGGGCTGGTCTTTTGTGAGCC ACCCGCTAACCTGAAAGGCCAGGAAGGAACGTCGGCGAGTGTCT GGGATGGGGTTTCCGTCCCGGGACTCCCTACGAGGGCGGTCCCC GGTAGCCAGAAGATCCGGCCGACTCCGAGCCTGGCCCCCTTGGGC GCCG GCGGGTGAGCAGAAGGGCCGTGCCAGGGCCTGGAAGTGCAAGG CCGCGTGGTGGGCATGGTAGGGAAGCGGAGCGTGGGCCTGTGAG GCGCGTGTGCGCCTGCGACCTCGGGACCGGGGCTCCCAATGAAC AGCGCGCACAGCTGGGAGCAGGGCTTGGGGAGCGGGGCTCTGCG GCCGGGGATCCGTAGAAGCCG GCGCGTAGGACTGAGAGCGCAGGGCGCGAGCCGAGGGCTCCGC TGACGGCTCCGGGTGTGACAAGAGCCAGCAGAGGACCCCATGG CCATGCGGGCCAAGCGCAGACGGCCCTCCTTGCGACCCCGCAG GCCGCCACATCTGGGACCAGCGGATCGTTGGTCTGCTGGAGCCGA TCCCGCCG CCGCTCCCCGGGCCCTGTCCCGCCTGGACGCTCCCTCCAGGAGCC TGCGCCCCGGCCCCGGGGTCAGGGTTGGGATGCGGGCTCTGCAGG CGCCCCGGCGAACAGCTCTACCTGGAGGCTGTCCCTGCCCGCTTA GTCCAAGGGCCTTGGTGTGGGGGCTCCGCTGTCAAGGCGGGGA ACCGGTTCTCTCGTTTCTCTCCCTTCCCCAGCGGCTTCAACG CCG	186
chr1 7	61404804	6140498 7	TBX4_T BX2	GCGGCGGGGGGCTCTCAGGTCGCTGGGCTGGTCTTTTGTGAGCC ACCCGCTAACCTGAAAGGCCAGGAAGGAACGTCGGCGAGTGTCT GGGATGGGGTTTCCGTCCCGGGACTCCCTACGAGGGCGGTCCCC GGTAGCCAGAAGATCCGGCCGACTCCGAGCCTGGCCCCCTTGGGC GCCG GCGGGTGAGCAGAAGGGCCGTGCCAGGGCCTGGAAGTGCAAGG CCGCGTGGTGGGCATGGTAGGGAAGCGGAGCGTGGGCCTGTGAG GCGCGTGTGCGCCTGCGACCTCGGGACCGGGGCTCCCAATGAAC AGCGCGCACAGCTGGGAGCAGGGCTTGGGGAGCGGGGCTCTGCG GCCGGGGATCCGTAGAAGCCG GCGCGTAGGACTGAGAGCGCAGGGCGCGAGCCGAGGGCTCCGC TGACGGCTCCGGGTGTGACAAGAGCCAGCAGAGGACCCCATGG CCATGCGGGCCAAGCGCAGACGGCCCTCCTTGCGACCCCGCAG GCCGCCACATCTGGGACCAGCGGATCGTTGGTCTGCTGGAGCCGA TCCCGCCG CCGCTCCCCGGGCCCTGTCCCGCCTGGACGCTCCCTCCAGGAGCC TGCGCCCCGGCCCCGGGGTCAGGGTTGGGATGCGGGCTCTGCAGG CGCCCCGGCGAACAGCTCTACCTGGAGGCTGTCCCTGCCCGCTTA GTCCAAGGGCCTTGGTGTGGGGGCTCCGCTGTCAAGGCGGGGA ACCGGTTCTCTCGTTTCTCTCCCTTCCCCAGCGGCTTCAACG CCG	187
chr1 7	61451790	6145198 7	TBX4	GCGGGTGAGCAGAAGGGCCGTGCCAGGGCCTGGAAGTGCAAGG CCGCGTGGTGGGCATGGTAGGGAAGCGGAGCGTGGGCCTGTGAG GCGCGTGTGCGCCTGCGACCTCGGGACCGGGGCTCCCAATGAAC AGCGCGCACAGCTGGGAGCAGGGCTTGGGGAGCGGGGCTCTGCG GCCGGGGATCCGTAGAAGCCG GCGCGTAGGACTGAGAGCGCAGGGCGCGAGCCGAGGGCTCCGC TGACGGCTCCGGGTGTGACAAGAGCCAGCAGAGGACCCCATGG CCATGCGGGCCAAGCGCAGACGGCCCTCCTTGCGACCCCGCAG GCCGCCACATCTGGGACCAGCGGATCGTTGGTCTGCTGGAGCCGA TCCCGCCG CCGCTCCCCGGGCCCTGTCCCGCCTGGACGCTCCCTCCAGGAGCC TGCGCCCCGGCCCCGGGGTCAGGGTTGGGATGCGGGCTCTGCAGG CGCCCCGGCGAACAGCTCTACCTGGAGGCTGTCCCTGCCCGCTTA GTCCAAGGGCCTTGGTGTGGGGGCTCCGCTGTCAAGGCGGGGA ACCGGTTCTCTCGTTTCTCTCCCTTCCCCAGCGGCTTCAACG CCG	188
chr1 7	61454733	6145491 9	TBX4	GCGGGTGAGCAGAAGGGCCGTGCCAGGGCCTGGAAGTGCAAGG CCGCGTGGTGGGCATGGTAGGGAAGCGGAGCGTGGGCCTGTGAG GCGCGTGTGCGCCTGCGACCTCGGGACCGGGGCTCCCAATGAAC AGCGCGCACAGCTGGGAGCAGGGCTTGGGGAGCGGGGCTCTGCG GCCGGGGATCCGTAGAAGCCG GCGCGTAGGACTGAGAGCGCAGGGCGCGAGCCGAGGGCTCCGC TGACGGCTCCGGGTGTGACAAGAGCCAGCAGAGGACCCCATGG CCATGCGGGCCAAGCGCAGACGGCCCTCCTTGCGACCCCGCAG GCCGCCACATCTGGGACCAGCGGATCGTTGGTCTGCTGGAGCCGA TCCCGCCG CCGCTCCCCGGGCCCTGTCCCGCCTGGACGCTCCCTCCAGGAGCC TGCGCCCCGGCCCCGGGGTCAGGGTTGGGATGCGGGCTCTGCAGG CGCCCCGGCGAACAGCTCTACCTGGAGGCTGTCCCTGCCCGCTTA GTCCAAGGGCCTTGGTGTGGGGGCTCCGCTGTCAAGGCGGGGA ACCGGTTCTCTCGTTTCTCTCCCTTCCCCAGCGGCTTCAACG CCG	189
chr1 7	64779504	6477972 9	SMURF 2_LRRC 37A3	CCGCTCCCCGGGCCCTGTCCCGCCTGGACGCTCCCTCCAGGAGCC TGCGCCCCGGCCCCGGGGTCAGGGTTGGGATGCGGGCTCTGCAGG CGCCCCGGCGAACAGCTCTACCTGGAGGCTGTCCCTGCCCGCTTA GTCCAAGGGCCTTGGTGTGGGGGCTCCGCTGTCAAGGCGGGGA ACCGGTTCTCTCGTTTCTCTCCCTTCCCCAGCGGCTTCAACG CCG	190
chr1 7	75487890	7548789 2	CASKIN 2_KIAA 0195	CCG	191
chr1 7	75488000	7548811 0	CASKIN 2_KIAA 0195	GCGCCATCCTGGTCTTGCACTGGGCCTACAGAGACGGACACCTGG TCAACCTGCCAGTCAGCCTGCTGGTTGAAGGAGACATCATAGCTTT GAGGCCTGGCCAGGAATCG	192
chr1 7	75640072	7564024 4	SMIM6 _SMIM 5	GCGGGCTGCGGATGGGTGCGAGGGTGGAATCTCGGTGCTGCGAC GAGTGTGGGGCCAGCCGTGGAGGCTCCAGGTGTTCTCTGCCCC AGCAGAGCCCGGAGGAGCCCCAACAGGAAGCCAGCGCGGCATG GCTGCCACCGACTTCGTGAGGAGATGCGCGCCGTGGGCG	193

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 7	75753505	75753724	GALK1_I TGB4	GCGGGTCCGGGGGTCTCTCTCCCCAGCTGTGCCGAGGCTGCACTC GCTCATCTGGAAGGCTTCAGCCGCGCAAGGGTTTCACCTGCCGCG GCCTTCCCGCTCCGGCCGTGCGCATCTACCCCCCCCCAACACACA CCCCGGGATCCCGGGAGCTGGAGACGGGCTCCCCTCGCAGAGCCT ACGGCCTTCCCCCGCTGGCCCTGCTCGGCCCGGCG	194
chr1 7	77373433	77373778	TNRC6C _SEPT9	CCGGGCCCCCGGGGGGCGCTTCTCGCCGCTGCCCTCCGCGCGAC CCGCTGCCCCACCAGCCATCATGTCCGACCCCGCGGTCAACGCGCAG CTGGATGGGATCATTTCCGACTTCGAAGGTGGGTGCTGGGCTGGC TGCTGCGGCCGCGGACGTGCTGGAGAGGACCTGCGGGTGGGCT GGCGCGGGACGGGGGTGCGCTGAGGGGAGACGGGAGTGCCTG AGGGGAGACGGGACCCCTAATCCAGGCGCCCTCCGCTGAGAGCG CCGCGCGCCCCGCGCCCGTGCCGCGCGCCTACGTGGGGGACC CTGTTAGGGGGACCCGCTAGACCCTGCGCG	195
chr1 7	79390056	79390201	RBF0X3 _ENGAS E	CCGCCGGGTCTCCGCAGCCTCCGGGTCTCCGCAGCCTCCGGGTCTC CGTAGCCAGCCACCCGCGGAGGGGCTGGGTCCACAGAGGAGGA CCAGCAGCAGTGAAGGGCAAGTCCACAGAGTTCTGAGGTGTCCAA CCTCCGGGACG	196
chr1 7	79815775	79815944	CBX8_C BX4	TCGTGCGTGGCCGCGGGGCTGCCGTCTCGGCCCTGTGCGGGTCT GCGCTTTGGCGGCCGCCGAGCCGAGGGGAGAAAATGGCCGTGG CGCGGGGCGCCGCGAGGGTGC GGGAGGGCTGGCAGGCGCGG CCGCTGGAGGGGCGCCGCTCTCAGGGCTCGGTGAGGCG	197
chr1 7	81025756	81025942	BAIAP2 _CHMP 6	TCGAGCTTAACACTCAAATCATGTTTTCTGAAATCATGTTACTTTCT GGCAAAGTATGCCGGCGAAGCCACTGAGACACGCTCCGCACATCTT TAGAACATAAAGGCCCTGGCAGTAGCTTGCGGGCTCTTTGGAAA ACTGCTTGGCTCTCACTGGAACACAGCCACGCCTCCTCTGGGCC CG	198
chr1 7	82888917	82889090	ZNF750 _B3GNT L1	CCGTGGGTGCACTTTGCTGGGTCTTCTGGGACACTGAAGTCTCCT GTGTCTCCAGCCCTGAGAACTCGGAGCCCGGGTCTTTGGGAAG GACGGGGCACCAGCTGGTGACACATGGGAAGGGAGGTGTGGTTG TCACCTTGCCAGGTAACCTGCTCTGCCTGGTCCGTGCG	199
chr1 7	82889202	82889334	ZNF750 _B3GNT L1	CCGGCCCTGGGACTCGGCCTGGAGAGCCTATTGACACCGTGCCAT GGGTGCGGGCAGGGCGCCCTCCCTGGAGGGCGGCACGTGGTGCC AGTTGGTGACCATGAGCTGCCTCACTCCTGAGGAAGAGTGTTGCG	200
chr1 8	909016	909218	ADCYAP 1	TCGATGCAAATCCAGGGCAGCAGCCAGACTGGCATATGTAGGGC TCTCCGGTTACTTTCTGTATGTCGCGGGTGAGAGGAACAGCGAG GACAATTTAGCGAAACACAGGAAGGGTCGGATCTCAAGGGGGCA GCGCTGGGAGAAAGGTTAGGCTTGAAGCGCGCTCGCTGCCCGG ATCTTATCCCGGGCCCCCTCCG	201
chr1 8	50268570	50268724	CCDC11	CCGGTGGGTGACTGTGGCTGGGAACACTACGGGCTTTCTCGCCCCGG CGCCCCCTGGCGGACCCACCAGCAGGTTGAAGGTGTCCGCGCAGT GCTGAGCACCAAGAGCCTCAGCCTTCAGCCAACCCCCCGCCCCGC GGCCTAGGTAAGTGAATCG	202
chr1 8	78949408	78949588	SALL3	ACGCGAGGACACAACCCGGAAGAGTCTCCCCGGAGCGGCACTGT GCCGGCCCCGGTCTCGGACCTCCAGCCCCAGAGTGCTGGAGAAT AAAGGCCCGTTGCTCATGAGCCACTGCGCTATGCATTTTGTTACAA CAGCCTCACCAGGAGTCCAACACCAACATCCAGGTGAAACTGACG	203
chr1 9	641770	641986	FGF22_ RNF126	TCGGGCTGGGAGGCTGCCCCGAGGAGCTTCACTTTGACAGGGAG CTGGCCGGGCACGCAGGGAAGTGTACACCCAGCTGACAAAGCGGC AGACACCCAGGCCGGGTGAGCGAGTGTGGGTGAGGAGTGGCGG	204

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 9	1210277	1210602	STK11_ C19orf2 6	CTGGCCCCAGGGTCCTTGCTGGACAAGACACTTCAGCTCAGGGTG GGGCAGGGGCTACCCAGGGGTACCCACAGACGATGGCG GCGCTGCAGGGAAAAAGCCTCCTTGTGTGTGGGAAGTTAATAA ACTCCGCTCAGATTGTGTCTCGCAGCGAGTGTCTGGAACCTTCCAG ACAAGCCTCAGGCGTCCGGTCTCCAGTTGGTGTGGAAGCGTGG GCGATACCAAGGGGGTGGGTTGGGGCAGATGGAGCCGGCGTG AGTCCCGTCTCTTCCCTTCTTCCAGAAAGGCAGCCCTGGAGTCCA TGCCTTGTCCCGTCTCACCAGGCAAAAAGTATAATCTTATTAGAAAT AGGAAAGTTCCAAAAAGCATCAATGAGTTAAAAAGAGGGCTGGGC ATGTTTCG	205
chr1 9	1467885	1468077	C19orf2 5_APC2	GCGCACATCGGCCATCCCTCGCGCTTTACGCGGGAGCGTCCGCAG GGCCGGAAGGAGGCCCTGCCCGTCCAAGGCTGCACCAGCTGCC CCGCCGCCCGCCGACCCAGCCAGCCTCATTGCTGACGAGACCC CGCCCTGCTACTCCCTGAGCTCCTCCGCCAGCTCCCTCAGCGAGCCC GAGCCCTCG	206
chr1 9	3688011	3688251	CACTIN _PIPK1 C	GCGTGGCCAGCCCGCAGGTGGCGGGGCCGACGGGATGGGTGAGG GTGCACAGAGCACACGCCAGCCCTGGGGGAAGCCCGGCCGTGC GGGCTGCGGGAGATCCTGATGGGCCCCGAGCTGAGGCTCCCGCAG CCAGGGTCTGCGCGTGGTCCCCACCTCCTTGC GCGCTCCGTCTCCA GCACAGCAGAGGTGGACGCCCTCGCGGCTGGCTCCCCAGCGTCC CTGTCTCCAGGGGCG	207
chr1 9	5229634	5229887	PTPRS_ KDM4B	CCGTGGCGTTGAGCGCCTCCGCCTCCACCTCCGCGGCGGCGCGCT GGGCACTGGCGGGCGGGAGGGGAGGGGAGGGGCGGGCGGAGC CGTTACCAAGGGCGCCCGCCCTGCCCGGGCAGTGCCACTGTCCG ATTCCAGGATGCCGAGTGGCTGCCGTTGAATAACTGGGCGCTCTTA GCGCTCACCACCGGGCGGGAGGACATGGCCTCCTGCACACCCCC ACAGCCCTGGGAGGGGCCCTGAAGGTGCG	208
chr1 9	10891471	1089153	CARM1 _YIPF2	CCGTGGGGTGGGTGTCAGGGCTTGTCTGGGAGATTCCAAGCTGAG GAAAGCAGGGCTGTCCG	209
chr1 9	10891633	1089170	CARM1 _YIPF2	CCGGCTGCCCACTCTAGGGAGGGGCCAGATAACTTGCCTAGAC GCCGGCCTCCCGCCCCAGCCTTCG	210
chr1 9	15181568	1518182	ILVBL_N OTCH3	CCGCCCACTGGGGCTGCAGTGGGCGAGGTCTGTTCGAGTGGA AGCCTCCGTAGCTGGCGGGCAGGTGCAGGTGAAGGAGGCCACG TGGTCGGTACAGGTGCCCGGGCCGAGGGGTTGCTCAGGCACTCA TCCACATCGCGGGCGCATCGTGGCCGGCGAAACCAGGGAGGCA GGAGCAGGAAAAGAGCCACGCCGTCTTGACACGAGCCACCGTT CAGGCATGGGTCTGCGGACAGGAGGAAGGCG	211
chr1 9	39264155	3926429	IFNL2	CCGACGCCCCCCAGGGGACAGTGGCCGGCAGCACCTGCTGCAGC ACGAGGCACAGAGGTGCACTGCAGGAGAAAGTGAAGGCAGAG GCCAAGGCGAGGAGGGGGCCGGCTCCCGCTCTCTCCCTCTGTGT GTGCTGCG	212
chr1 9	41499327	4149958	CEACA M21_A TP5SL	GCGTGGGGGAAGGAAGAGGGTATGAGGCTGGCATGAAGTGGGG ACTAGAGAAAGGGTGAGTAGTTTTTCAGAGAAAAGGCCAGTGCCA GGGCTGTCCAGGAGCGAATCTGGTCACTTGTCTGAAACAGGGGT CCGGGTCTGGCAGTGGCAGCATGGTGGGGTGGGTGAGTGGCACT ATGGAAGAGCCAAATCTCCACCTCTATCTCAAAGCCTTTCTCCAC ACAGCTTTCGGTTAGCAAGGCTCCATGAGAATG	213
chr1 9	46493754	4649396	CCDC8_ PPP5D1	ACGCCCCGGCCTCGGCCTCGGCCGCCGCGGGGTTTTGCGGGCC CCGGAAGCGGTGGGAGGCGCGCCGGCCGAGTCAGGCCCTGGG GGCCGTGCGCGCCCTCTTGGCCGGGGCTTCTGGATGCCCTGTCC	214

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr1 9	48735170	48735556	MAMSTR_RASIP1	TCCGGCTCCGACGCCTCGCTCTCGGTGTCCTCCGACTCCTCCTCGGACTGTTCTCCGAAGCCTCCTCCGACCCCTCG GCGTGCGGGGCTGGGGCGGCGTTACCTGGGCGTCTGGTAGCCC TGGAGCAGCAGGAAGTAGGGGCGGTTGCTGGGGGCTGGATGAG GCACTGAGTCAACTGATCGAAGTCCCCGGGGTCTGCAGTTCGATT TGGGCGTCGGGCTGCCCCTGGGGCCATGCTAAGTGCCTGCTGTCTC GCTCCTGCTGCCGCCGCCGCCCTGAAGGCTAAGCTCCGACAC GCTGCGCCGCAAAGACAAGTTTTCTGAGCGCTCCTTGCTCCAGAC CCAGCTGGGGCCCTGATCCGGTCCCCGGGCCAGGACTGGCCAGC GCTCCCCACCCGACGCCGCCGGGAGCGGTTCTTCTGTGGCCGCC ACGAAGGGGCGCCGGTGCCTGCG	215
chr1 9	57076511	57076538	ZIM2_USP29	TCGGGGCCGAGAAGCATTAAAATGACG	216
chr2 468115	468115	468367	FAM150B_TME18	CCGCGAGGGGCGAGGACGAGGCTGCATGGGCCAGCGAGGGGGTCG ACACCGAGCCAGAGTGAGCGCGGGGCTGGGGCGCAGAGCCCGC CCAGGGAGCCGGGAGACGCCGCGCAAGCTCCCCGACAAACGCA ATGACCGAGGACGCGCGGGCGAGGCCGTCCAGGGAGCCCTGGTC CCTCAGCTGCACCGGACTGAGCCGCGACCGCTCAGCACGCGCTGCT TATAAATCAGGGGTGCGCTTCCCAAGCCCCG	217
chr2 1040869	1040869	1040871	TPO_SNTG2	CCG	218
chr2 1040974	1040974	1041007	TPO_SNTG2	ACGGCTTTTGGTGGAGGCTAATGTTAAATCCG	219
chr2 2116598	2116598	2116856	PXDN_MYT1L	CCGTCCTATGACTCTCTTTTGATCAACGCAATGCAGTGCAATTGATG CCATCTGACTTGACGAGTGGGTTAGAAGATGCCTCTCAGATTCCA TATAGGTCTCTTGGAAGATCCGCCCCGGGAAAGCCAGGCCATGTA AGACCATTGACCACCTTAGGACCACCAGGCTTGGAGGAAGCCAAG ACACCCACGTGGAGAGGCTGTGACGGGAGTGAGGGAGGTGCAGC CAACCTCACCTGGCTCCACTTCAAGGCCCG	220
chr2 5692734	5692734	5692953	SOX11	GCGGAGAGCTTGGAAGCGGAGAGCAACCTGCCCGGGAGGCGCT GGACACGGAGGAGGGCGAATTATGGCTTGACGCCGTTGGCCCT GGACGAGAGCGACCCAGACTGGTGCAAGACGGCGTCGGGCCACA TCAAGCGGCCGATGAACGCGTTCATGGTATGGTCCAAGATCGAAC GCAGGAAGATCATGGAGCAGTCTCCGACATGCACAACGCCG CCGTTTCTGAACCCAGGAGACACTCAGGAAACCTTGCTGGTGGAAC GGATGCAGCAGCGAGGTTTTCCGGGGCAGGAACACCTCCAGGA GCTTTTCCACGGCCAAGCGCTGGCTGGTGGTGGAGCTGCGCTGAA GTCAGTGTGTGCTTTGGGCCAGCTGCACTGTGCCGGGGTCCAG GGATGGGTGTGAGGCTGTCTGCCCCCACTGCACGCCGGCTGTCA GAGGCATCTGTCTTCCCCCGCATGCATCTTCTCCCCGTCTGGCA TGGTGTCTAGTCTTTGTGGATGGGGACATAAACAAGCCGCCAT CAACTGCTTGGTGACATTGGCCAATCCTGTGGTGGCCCCAGCTGGG CTTGCTGCCTGTGTGTGGTGAGGGTGCCCTTCTGTCAACCG	221
chr2 10331450	10331450	10331857	HPCAL1_ODC1	TCGCGAGGTTGCGGGCAAGACCCCTTGAGGTGCCAAGTCTGGGC CGCCCCCAGGGGCTGGCCAGCAGGGGGCAGCGTGGCTCTGAGCG TGGAGGCCAGGGCTGGTCCGCGCCGGCAGGGGCCAGCCTCCAGTGC CCAGTTGGGTTCCCGGGCCTCGAAGTTCTAGCCCGCAGGACTCA GGAGCGTTCCCGGAGGAGGTGGGGATGGGGTGGTGAAGCCAG AGCGTTTTAACTTCTGCATCCCCTGCCGCTTCTCAGCCAGCAGGGC CCGGCTTGAGGCTGGGATTTTTGGTGCCTGCAGCAGGGAAGCTTAT	222
chr2 19350218	19350218	19350866	NT5C1B-RDH14_OSR1	TCGCGAGGTTGCGGGCAAGACCCCTTGAGGTGCCAAGTCTGGGC CGCCCCCAGGGGCTGGCCAGCAGGGGGCAGCGTGGCTCTGAGCG TGGAGGCCAGGGCTGGTCCGCGCCGGCAGGGGCCAGCCTCCAGTGC CCAGTTGGGTTCCCGGGCCTCGAAGTTCTAGCCCGCAGGACTCA GGAGCGTTCCCGGAGGAGGTGGGGATGGGGTGGTGAAGCCAG AGCGTTTTAACTTCTGCATCCCCTGCCGCTTCTCAGCCAGCAGGGC CCGGCTTGAGGCTGGGATTTTTGGTGCCTGCAGCAGGGAAGCTTAT	223

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
				AGTCCAGTTGTCATCCGCGGCCGCCGCGCTCCGGGCGCTGAAGCT GGAGAGGCCATCCTGCGCTTGGGAAAGGCCGCGGGCGCCACCGCC TGCGCGGTCCC GCGGT CAGGGCGCTGGAGCTGGGGGAGCCCCG CCTTGCCCCAAGGAGAAGAGCCCCGGCGGCTGGCTTCTAACTGTG GGAAAAGTAGACACCCAGGGAAGTTTCTAGCTTATGGAAGCGGA CTCGAATTTTCTCTAAGCGTCCCGGGCCTCCAGGGCGCCCGCC CCCACCATTCCTGACAAGGCTTTAAATTTGTAGGGAATCTTCGCGG GTGCAGAGCCTCG	
chr2	30230708	3023092 8	LBH	TCGGAGAAAGACGTGGGAGTCAAGGATGGGGGGCGCGTGCACAC CGCCCCGCCACACCTTCTGCCCCGCTGCAGACCGGGCGTATGTGT GTCTCCAATGGAAAAATCCTACCCAGGACGACACCACATCCTTGCT CCCACAAATAAAACCTTCCACGGAAGTCTAGGGCTGCAGACAGCCC TTCGCAAGCCAACGCGCCCCGTGGGCACTCGTCCCCCG	224
chr2	31580228	3158047 7	XDH_M EMO1	GCGGGGCGCGATATGCCACAGGTAACCGCCGCTGCGCGCAGTTA AGGAACAGTCTGTCCAATAGGTCTCCCCAACCTGAGCTTTCCAGG TCGCCTCCCCCGCCGAGGACCTCTTCTCTCGAGCAGCCAGAGGAT TTGGAGCTGCTGAGAGCGGATGAGGTCCTGGGGGAGTGAAGGCG GCGTCTGTGCCGAGCCGCTTGTCAACTCTCTAGCGTCCAAGCCCC GGCCCCGGCCCCGCCAGGTGCG	225
chr2	31580671	3158092 6	XDH_M EMO1	CCGAAGAGGGAGAGGGGCTGCCGGGCGAGGATCCCCGCGGGCAC CGCAAGGAAGGCAGCTCCTGCAGGAACAGGCGGCGCGGGCTG GCAGGCGGGTAGCCGCGGGCTTCAGGCTCTCCGTGTGCTTCCCGTA GCCGAGGGCTTCGCGACGTACAAGGCCAGTGCCCAAGGGCGAC CAAAGTGCGCTGCTGCCAGCACTGGGCTCTGCTGGCACTGAACC TGCATCGCGCCGTGTTCTCGCCGGTGGCCG	226
chr2	44801816	4480209 7	SIX3_CA MKMT	ACGGTGCGGCCGCTTGGGCGTGATCCCTTGGCTGGGGCTGCAGGG GGCCCGTCTCCAGGGGCGCAGAGGGAAGGACCAGCGTTTCCAAG CCGGGCTCTGGCCGCGCGCGAGAGCGAGGCCAAGGTCTGGGG GCAGTTCAGGGGGACCCCGAAGTCGGGACGGCCAGAAACGCTTT GCCCCAGCCACCGCCCTTTCTTTGTGAGTTTCCCCAAGCCGTGCG GTGCGACCCGGCGCCGACTCTCTCTCTTCCCTGCGAGGGCCC GCGCCGCCCCG	227
chr2	45000612	4500089 1	SIX2_SI X3	ACGCTCCCCTGACCTCAGGGCCAGAGCCTCGCATTACCCCGAGCA GTGCGTTGGTTACTCTCCCTGGAAAGCCGCCCGCGGGGCAAGT GGGAGTTGCTGCACTGCGGTCTTTGGAGGCCTAGGTGCGCCAGAG TAGGCGGAGCCCTGTATCCCTCCTGGAGCCGGCCTGCGGTGAGGT CGGTACCCAGTACTTAGGGAGGGAGGACGCGCTTGGTGCTCAGGG TAGGCTGGGCGCTGCTAGCTCTTGATTTAGTCTCATGTCCGCTTT GTGCCG	228
chr2	45004557	4500469 3	SIX2_SI X3	GCGGCCGCGCGCCGCGCCCTGAGTCCGATTTCCTCCTTCCCT GACCTTCAGTTTCACTGCAATCCACAGAAGCAGGTTTGCGAGCT CGAATACCTTTGCTCCACTGCCACACGACGACCCGGGACTGGGCG TCGGGTTGAGAAAATCCG	229
chr2	47043762	4704377 9	TTC7A_ CALM2		230
chr2	47043900	4704399 3	TTC7A_ CALM2	TCGGGTCTGCCCTAGACCCATTCCGGCCCTCAAAGATGAAGAAAAT GAGAAGGGGGCTCTGGCAGAGAGAAGTGTGATGCCTGCAGAGGG CCCG	231
chr2	66439219	6643973 3	ETAA1_ MEIS1	GCGGTGGGGGCTATCAGCGAAGGGAGGGGAATGTGCGTGGAGCT GAGGAGGAGCCTCCCGGCTCTCCGAGGGCCTTGGGGTTGGGATCC CTAGGTGCAGCCCGTTGACAGTCGGCCCCACGGCCATGGACGTCCT	232

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2	68799906	6880009 1	ARHGA P25_B MP10	TTCCCCAAGTTAGCTGAGCGCCTGCCACCGAGATCCCCGAGCCTG GGCTTCGCGCGGGCCCTAGGAGGAACCCGAGGAACCAGCCCTC CCCAACTCTCCGCCCGCGCCTTTCTCCTCCACCGGATCCTGGATGT GCAGTGGAGGGGACGAGGGCTTGTGCGGTGGGAACTTAATTCA AAATGGCTGCTGAAACGCTTGGGTTTTATTCTAGCAAATGTTGC CAATTTCTCCGGCCAGATACGCTAAACCGATCCTCAGATACCGTCCA TGGCTCAGGGCCTCCGACTTCAGGGCTCCAGGAGGAAGGGGAGGT GAGCGGTACCTGGGTCTGGGGGAGGGGAGGAAAAAGGAAAAA AGTAGATGACACAATCG TCGGAGGCGTGAGTCTTCGGCCCTGCCATGCCTCACATCCCCAGGA TGCCCGGTGGGAACTGGGCTGTGGCTTTCCTGCCCTGGCACTGCT TGTTTGCTGGGATTTTCAGGAGGAAAACCCCAAGCTCCGAAAGAA AGGTATTTCTTTTTATTTGTAGTTCACTTCTTCCACTAGAAGACTC G	233
chr2	72920275	7292088 5	EMX1_5 FXN5	CCGCCGCTTCTGAGCCATCAGTCCCAGCGGGTACGTTATCGAGTA GCACAAACAGTTGGATTTTCCCTCAAGAACCGAGCTGGACGCGG AGATGGAGCCAAGTGTGGCTGCATTTTCGGACCCGGAATCCGTT GGGCACTGAAGGACTTTTCGAACCTGTAGCGCTGTTGCTTCGCGG TCCATCGTCGCCGCTGCAGACGGATGCGCTCCCCGGCGGCTCTACG CCCTCCAGTCCCGGCCAGGCCTCTGGGCTGGGAGCCGAGCCGTCTC GGGCCCTCCGGCGCCGCGTTTTCTAGAGAACCAGGGTCTCAGCGAT GCTCATTTTCAGCCCCGTCTTAATGCAACAAACGAAACCCACACGA ACGAAAAGGAACATGTCTGCGCTCTCTGCGCAGCGCTTGGGCGGC GCGGTCCCGCGCGCGGGGAAGCGGCGTCTCCGCTAACCGAGGC GCTGGAAGGGGAAAAGCGAATGCGGAATCGTCCAGGACTCCGAA GGTGCGGGCCGCTCGCGAGCACCGAAGGGGAGGAGCCGACGAAG ACCAGGAGTGGGCCGATTTCCGGTACTGTTTCCCCGAGATCAGGAA CTTTCCGGGTCTAGGAGCAACG	234
chr2	74499264	7449968 3	MRPL53 _LBX2	ACGGGGAACCAGGAGGAGAGAGGTGAGGAAAAGGCTAAGTCAG AGTCCGCGACCTTGCCGGCTCTATACCTTCAGAGGGCTGCAGAGCG CGCGCGTCAAGTCCGCGGAAAGTTTTACTAGTCAGCTCCTCCAGCG CGCACAGCGGCGACGTTGGACCCGACCCGACTCTGGAAGCTGCG GCGCAGAGGGTGCTCGGGGACCATGCGCGGGGCTAGGATGTCT GCGATGCTTAAGAGTGTCCGGGGTGTTCGGGGCTCGCGTCCCGAG TTCATGGTTCGCGCGGGCTGGGGCGGTCCGGCTGTCCGTTGCGCTA GGCTCCGCAAACGCCTGGGCCCAAGTGCTCGGCTCCCAATCCGGGC CCCCAGCCTCGGACCCGCCCCGGCTCTGGGCCCCGAGTCCCGTGTG CCCCTCCTCCTGCG	235
chr2	85584347	8558473 2	VAMP5	TCGCCACTCGCGGAAGGCGCGCCCCCGCCCTCGCTCGCGGGCCC GCCCCGCCCCGCCCCCTGCTTTCCTCCGGGGCCGCTGGCACTGCGG CCGCTCCGCAGGCAGAGAAGCCGGGAGCGGGCGAGGCGGCGGCG GCAGCAGCGATGGTGAGGGCCAGGCGGGGCCGCGCCAGCCCTGC GACGGGCAGAGGGCGAGTGCGGAGGGTGGGAGAGAGAGTCCA AAGTCCGCGGGCTGGGGCCTCCCTGGGGCCACGAGGGCCAGAC CTGAGGCGGTGACCACTGCTGGAGCAGGACGGGGCGGACCTCCA CTCCCTGCGCGCCGCATGGGAGAGAAATGCGTGAGCCCCGTCTG GCTGCACCGCGCAGAGCGAGCGGGACTCG	236
chr2	85936659	8593691 9	ST3GAL 5_POLR 1A	GCGGGAAGGGGACAGAGTGGGAGGTCCCTCCTCGGTGCCGGCT GCGCCAGCTGCTGCCGTGTTCTGGTGTACCAGGCCGGACCTTGCGC AATGCCTTTGGGGTAATCTTCAAACCTATGTCTGCTGATCACTCTCT	237

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2	98823029	98823181	MGAT4A_TSGA10	TTAGCTGCCTGGCAGTACCGCAAACCCAGTTGTGGAAGTCCCACC ACAAGGACCTTGACAGAGGTGGAGGCCCTCCCATGCAGAAGCCA GAGAACTGCGCCATTCTCCCGGTATCCTTCCG TCGGGGGGAGTCTGTCCCCCTCAGGGATGGCGGTGGGAAACGG GCTCGCGACGTCTTCGGGAGCACAGACCACCTCCTCCGCCTTGTC GTGGCCGGGGCACACGGGCTGCGGGGGGCGCTCCCATCTCTGC TTCCGCCGTCTGGGACCG	238
chr2	104842514	104842734	POU3F3	GCGAAAGAGGGAGATGCCCGTGTAGAGAACCAGAGAGGGGGGCT GGGGTAGAATAATCAGCTCTAAGGTTGCAGATTAGATCTCAAGGC TGAAAAGGATAAGCTTCCACCAGAGCATCTGTAGCGCTCTGTCTG CTGCCCTGCCCTGCCCTGCGCGCGCACCGCACTCACACGTACACCC GGTCTCTCGACGCGCACACGCACTGTTCCCCGCCG	239
chr2	104842858	104843001	POU3F3	GCGCGGCCTTCGGGGCTCCAGAGCGCGCGGGCCCGAACGAGGC GCGCGGCCCTGGCACATGCGGGGACTGCCAGCGCGGACTGGA GAAGGGGAGCGAAGGGGTGGGGAGGGGGTGACGCCGGCTGCC ACCCCGCTCCCG	240
chr2	104844478	104844803	POU3F3	ACGTTACACACCGCTTGCTAAATGCAGTGGCGAGAGGAGGGAGC AGCGTCTACATGAAGCGAACTTTCAAGCGCAGAGCCCTGACTCCC AGGCGCGGGGGCTCACCGGGAGGGGCCGGGCGAGAGAGCGCG TGGGTGCGTGAGTGCCTGTGTGCGCCCGCCCTTGCTTGCTCGGG TGTCGCCCTTTGTCCCCCGCCGCGGGCTCCACGGTGGGATCTGCG CGCGGCCGGTGGGCAAGCCCTCGACCCGGGCGCGTCCACAGCGCC CACCCGCGGCCCCCAACACCTCGAGAGCAGATCTTAGGGTTAAC CAGGCACCG	241
chr2	106065435	106065782	C2orf40	CCGTTTCGCTGCGGGCAGCGCTGGCCACGCGGCCCGCCCGCCG GCGGTTCTCCGTGGCCAAGCATCCTTGGCCTTGAGAGCCAGGGGCT GCGTTCCCTTGGGGCCGGGCGGGAGAGAGGACCTCGGTGGTA CTCGCCGTGCGCTGGGCGCAGCCGCTTGCCCTCAGCCCTCTGGC GCGGCGCCACCCGCTGGGTCCCGCCCGGCAGCGACGACGAGGAT AACCCGCGGCCGCGCTGCCGCTCGCACCCCTCTCCGCGCCCGG TTCTCCCTCGCAGACCTCGAAGTGCGCCCTCGCCCTCTGCTCGC GCCCGCCGCGCATGGCTGCCTCCCCCGCG	242
chr2	113173940	113173989	PSD4	GCGGAAGTCGGAAGCTCCAGCCGTCACAGCCACATTCACTGGGCA AGCCG	243
chr2	113276786	113276958	PAX8_PSD4	CCGCCGGAAGGGTCAGGGGAAGGTTAGGAGGAAAGATGGACCTC CAGAGCCGAGCAGAAGTGCCATTGCACCACTTGCGCAGAAAGTG CCATTGCACCAGCTTGGCATGGGCACCGGGCACTGCACATTAGGCC TCAGGGATGGTCTGGCATGTCTGGTATCGTACCACG	244
chr2	130963976	130964318	ARHGEF4_FAM168B	GCGGCCGCGCACCGCCGCCCCGCCCCAGCCTTCCCCGAGCCTGT GGCTGGAGCTCGGGCCCGCTGCGTGCGGGCGCAGCAATGCCCA GCGAGTCAAGCGGGCAGACGAGTGGCGATCTCGGCACTAGCAGC AGCAGCAGCGCCGGGCTGTCCCCGGGCTCCGACTCGGACAGCAGC GGCGTGGTGTGTGCGGCCGCGGAGGCAACGGGGGCATGCGCGG CGCCGTGTCCCGCTCTGGAGCTGGAGAGCTGCGCTCGGCCACC GCCGGTAAGGACGCCGCCATCCCCGCGCCGACGCGCCCTCCGCG CCCGGGTCTGTCTTGGGACCCCCG	245
chr2	131034614	131034830	FAM168B_ARHG EF4	GCGGCCAGTCCTTGTAAGGAATCAGAGTCCCTGGCCCATCCCTCCC CAAAGCGCCGGTGCCAGGCGTTTTGGCCTCTGTATCTCTGAAACGA GGAGGTCCCGGGGCATCCCCGAGCGCCCCCGTGGCCATCTGTGCC	246

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2	161423935	161424183	SLC4A1 O_TBR1	ACTGGCCAGCCCAGGGCCAGGACTGCTGTGCCGGCGTGAGATTCCGACCTTTCCAAGGAGGTGCCAAGGGCGCAGCG CCGAGGGCCTGGCCGCCGAGCGCTCGCCGCTGCCGCCCGGCGCCG CCGAGGACGCCAAGCCCAAGGACCTGTCCGATTCCAGCTGGATCG AGACGCCCTCCTCGATCAAGTCCATCGACTCCAGCGACTCGGGGAT TTACGAGCAGGCCAAGCGGAGGGCGGATCTCGCCGGCCGACACGCC CGTGTCGAGAGTTCTGTCCCCGCTCAAGAGCGAGGTGCTGGCCCA GCGGGACTGCGAGAAGAAGAACTGCG	247
chr2	165793968	165794218	GALNT3	ACGCAGCCCAGGGGTACCGCGTCTCCCTCCGCTGCCGCCGGCTTACCTGGCGGGTGGGCAGGGCAGGGTGGCGGGAAGCGCGCGCCGG GCAGGCGCTGGACGTGGGCTAGGCGCCAGGTGCAGGTGGCGGCG GCTGCGACTCCGTTGCTGTGCCACAGTTGCGGCTCAGTAGAGCT CCTCTCCGCCGCCCTCCTGCCTTCCGCTGGGCCTCCCGCTTG CCTGGAGAGGCAGAACCGAGGCTCG	248
chr2	170823080	170823362	GORASP 2_GAD1	CCGACTAAAATTCTAGCCTTATCGGGCCAGAAAATACGGATGTC CCCGGGCAGAGGTTGGAGAGGCGGGGGAAGATTACCGGCCGCT TATTAAGAGCCATCCGTGAGCTCCTGCGCGCGGAGATAGCGGC AGAGCAGGCACGGGACACGCCGCCGCCCTAGCCCCGAGCGCC GAGAGCCGCCCGCGCTGGGTGCTCTGTGCACCTGATCTTCCAG CCTCCTGGGTCCCGGGGCGAGGGCGGTGGCAGTTTGCACTCAGA GCAGAGTGGCCG	249
chr2	172100815	172100923	DLX2_D LX1	CCGGCGCTGAGACTGGCGGCGAAGCACAAGGTGGAGAAGCGCTG GCCCCAGGGTGCTGCTCCGAGGGGATCTCACCCTTTTCCACATCTT CTTGAACCTTGGACCGGCG	250
chr2	174337530	174337828	SP9_CIR 1	ACGACTCTTAGAGGCCGGGCGAGAGGCGCGAGCACACAAGCGAG TAGAGACACCGAGAACGAACGAGAGGTTGCGAGGGCGAGCGAGC GGGAGGCGGGAGGGCAGGGGCTTCAGTGACGCCCCAGGGCCCCG GGCTGGGCGCGAGGTGGAGCCGCTCAGGGCTCCCGGGCTGCGGT TCGCCCCGTGTGCGAGGAGCTCCCTCTGCCTTCCGCGCCCGGATA AGAATCGAACGCGTGGTCCGGAACAAAAGCGAACCATCTCCGA CACAAACACTTTAAAACTGTACTCCAGACG	251
chr2	176082294	176082535	KIAA17 15_HOX D10	ACGCGGTACGGTAGCGCCGCACTTGATCCGCGCCAGAGCCGAGC CACCCAGCGCCGCGCTCCCGCCGCTGCCTCCGCTGCCTCCATGCAG GCTTCCGAGGCTGAGCCCGACGCCGACGTGCTGGTGCCGGCAGC CGAGCCGCTCTCTGCGTACCCTGGCAAACAAACGACCAACAGCGCA TGAGTGCTGTAGGACCAACAGCCCGGCGCTGGCGCTGCGCGCGG ATCGGGGAAGCCCCG	252
chr2	176091806	176091979	HOXD1 O_HOX D11_	GCGAGGCCGGTCCGGCTGCTGGAGAGACACAGAAGTTTACGGTG GGAGGCTGAGTGGCTTTCTCCCCGGCGCCGTTCTCAGGGTCTTTC TGCGGGTCGAAGAAGGACCCGCGGGAGCTGAGAGGCCAAGGTCG GAAGCACTCCCGCTGGCCCAAGAGTAGAGGCGAAGAGCG	253
chr2	176100105	176100308	HOXD1 O_HOX D11_HO XD12	GCGCCCCAAGCGGCCGCTGGGCCAGAGGAGCGCGGTCTGATCCCG GCCGTCTTCGCCCCCGAGTCTAGCCTGGCTCCTGCAGTGCTGCT CTCAAAGCGGCCAAGTATGACTACGCTGGTGTGGGTGCTGCCACG CCGGGCTCCACGACCTGTCCAGGGGGCTCCTGCGCCCCCTGGCT TCAAGGACGACACCAAGGGCCCCG	254
chr2	176104613	176104885	HOXD1 O_HOX D11	TCGGGGTCTTACGGTAGGTTCTCGAGCGGGACGCGCGGGTCCGG AGGCTGCGGTTTTCCCTGGGTTTGGGGAATGGGGGTAGGAACTAG GAGGGAGCTGGGGCCAAAGAGCCAAGCGGGCTGGGACTGGAATG AAAGCGCTCTGGGTTGTGGAGTGGGTGCGGGGGCAAGGTTCCGC	255

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2	176122574	176123118	HOXD9	GCTAAGGAGCCGAAAGGGGCCGCGCCCCCTTCCCCTATGCACC GGCGCGCCACTGCAGATGGCTCACCCCTCCCCGCCAAATCGTGCT CCCG GCGGGCTCTAATTGCGGCGCTTATGTTGATGATTTTTTTTTTAATCA CAGCAGCCCCAGTTTAGCGGACTGATTTACTCCCGGTATTGGTAA ATATGATCACGTGGGCCGCGCGACCAATGGTGGAGGCTGCAGCCT GCGAACTAGTCGGTGGCTCGGGCGCCGCGGGGAGCTGCTCGGC GGCGGACAGTGTAATGTTGGGTGGGAGTGCGGGACGCTCAAAAT GTCTTCCAGTGGCACCCCTCAGCAACTACTACGTGGACTCGCTTATA GGCCATGAGGGCGACGAGGTGTTGCGGCGCGCTTCGGGCGCGCC GGGGCCAGGCGCGCAGGGCCGGCTGCAGGTGTGGCTGATGGCC CGGCCGCCACCGCCGCGGAGTTCGCCTCGTGTAGTTTTGCCCCAG ATCGGCCGTGTTCTCTGCCTCGTGGTCCGCGGTGCCCTCCAGCCCC CGGCAGCGGCGGCGATGAGCGGCCTTACCACCCGTACGTTCCCCC GCGCCCTGCGCCCTCTGCCTCCGAGCCCGGCCGCTACGTGCG ACGGAAGTGAAGTCTCCGTGGCCGGGAGTCCAGGGGAGCAGCG GCCCCGAGTTCTCGTCAACTCGTTCCTGCAGGAGAAGGCGGCAG CGGCGACGGGGGAACCGGGCCTGGGGCAGGGATCGGGGCCGC GACTGGGACGGGCGGCTCGTCGGAGCCCTCAGCTTGACGACCA CCCG	256
chr2	176123335	176123515	HOXD8 _HOXD 9	GCGGGGAGGTCGCCTGGGGCGTCGGCGATTATATTGCGGCCGAG CCGGGGCGCGCCGGGAAAGGCCGGGAGGGCGGCGCGCGGG GGCTGGGCGAGGCCCGCGACCCGCGAGGGAGGCGGCGCAAGC CGAGGCGGCGGGCGCAAGAGCCGGGCATGAGCGCCAGTAGCTG AGCGCCCGCGGCTGCCTGGCCTCAGAAGCGACGCGCAGCGCGG GCGGGCGGCAGCAGCGACGTAGCCCGCGGTCCCGGCGCGGAGA GCAGCCGCCCCACAGGCCCGCGGCGAGTGCAGGCGAGTCGAGGC TCGCTCTCTGGCTGCTTAGCGCCGCCCG	257
chr2	176129999	176130335	HOXD8	GCGGGGAGGTCGCCTGGGGCGTCGGCGATTATATTGCGGCCGAG CCGGGGCGCGCCGGGAAAGGCCGGGAGGGCGGCGCGCGGG GGCTGGGCGAGGCCCGCGACCCGCGAGGGAGGCGGCGCAAGC CGAGGCGGCGGGCGCAAGAGCCGGGCATGAGCGCCAGTAGCTG AGCGCCCGCGGCTGCCTGGCCTCAGAAGCGACGCGCAGCGCGG GCGGGCGGCAGCAGCGACGTAGCCCGCGGTCCCGGCGCGGAGA GCAGCCGCCCCACAGGCCCGCGGCGAGTGCAGGCGAGTCGAGGC TCGCTCTCTGGCTGCTTAGCGCCGCCCG	258
chr2	176152444	176152662	HOXD1 _HOXD 4	GCGTGTGCGCCGGGGAGAGGGCGGGAGGGAGGAAGCAAGCGAG CTTGGGAGCGCGCGGGGAGGGCCGCGGGCCTCGGGGCGCGCCAG GAAGTGAGCGGCGGAGGCGAGGGGCCCTAACTAGTGCGCGGCGC TGACCTGCCTGTCTGTCTGTTTTGTCTCGCAGTGAACCCCACTAC ACCGGTGGGGAACCAAGCGGTCCCGAACGGCCTACACCCG CCGTGGTGCGGGATTCCCGAGTGTGGCCCCGCTGGGGGAGGGTC TTGGGCGCTCATTACAGGCCAGGAGGTCCGCTGCTGGCGCTGGCA CGCTTAATCTTTTTTCCACATTGCAGAATCATTCCACACAGCCACT CG	259
chr2	176159549	176159688	HOXD1 _HOXD 4	CCGTGGTGCGGGATTCCCGAGTGTGGCCCCGCTGGGGGAGGGTC TTGGGCGCTCATTACAGGCCAGGAGGTCCGCTGCTGGCGCTGGCA CGCTTAATCTTTTTTCCACATTGCAGAATCATTCCACACAGCCACT CG	260
chr2	197786184	197786353	BOLL	GCGGGTGGGGAGAAGCGGACTGCGTCGCCTCGGGTGGCAGGTGG CGGTGCGGGCGGGCGCTGCAAGCCGGAGAGGGGCGCGGAGGG CGAGTTTCGGCTGTGGCCCTGGGACTCCGAGCCGGGGCGTCTCAG GGGCAGAGCGCACGGCACAGCGGGGCGGGCGTGGGGCG CCGGGACAGAGTGGAGGGAAGCAGAAACATTGCAATCGGGGT GGCGGCAGCAGCGACATGAGATCCTTTGCCCTCCGCCCCCTGGGCT GCGGGACCCAGTGACTTCGAGGAGGAGCGCGAGCGCAGCCGCGC GGGGCGCACCCGGATCCGCTGGGGCGGGAGCCGCCCTTCCCG CCGAGGCGGCGGGGCTGCGAGTCAAGTCCAGGACTCGGGCC AGTCTCTCCG	261
chr2	208406581	208406813	PTH2R	CCGGGACAGAGTGGAGGGAAGCAGAAACATTGCAATCGGGGT GGCGGCAGCAGCGACATGAGATCCTTTGCCCTCCGCCCCCTGGGCT GCGGGACCCAGTGACTTCGAGGAGGAGCGCGAGCGCAGCCGCGC GGGGCGCACCCGGATCCGCTGGGGCGGGAGCCGCCCTTCCCG CCGAGGCGGCGGGGCTGCGAGTCAAGTCCAGGACTCGGGCC AGTCTCTCCG	262
chr2	219448479	219448881	GMPPA _SPEG	TCGAGCGCGGCGCACCGTGGGGCACCCCCGGGGCCTCGCAGGA AGAACTGCGGGCGCCAGGACGCGTGGCCGAGCGGCGCCGCTGTT CCAGCAGAAAGCGGCCTCGCTGGACGAGCGCACGCTCAGCGCAG	263

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2	222298537	222298790	PAX3	CCCGGCCTCAGACCTCGAGCTGCGCTTCGCCACGAGCTGGGCCG CATCCGCCGCTCCACGTGCGGGAGGAGCTGGTGCCTCGCACGA GTCCCTGCGCGCCACGCTGCAGCGTGCCCATCCCCCTCGAGAGCCC GGCGAGCCCCGCTCTTCTCTCGGCCCTCCACCCCAAGACATCGC GGGCCGTGAGCCCCGCCGCCGCCAGCCGCCCTCTCCGAGCAGCG CGGAGAAGCCGGGGACGAGCCTGGGAGGCCAGGAGCCGCG GCGGGAACCCGCTACGCGGGTAGTTCTGCCCCGGGCCCGGCCGA TCATCCTGGGCACAGCGCCGCCAGCGTGGTCATCCTGGGGGACG CTTCGCTCGGAAATTATATCCAGGTGAAGGCGAAACGGAAAGGCG AGTGCGGCGCGGATGACCCTCGGGAACTATCCGGAGCGTGGAGA GCCCCCCCCAAAACGGCTGGAGAGAGAGGGAGGGACGCGGGGA GGGGGGCTGTCGGTTCCTAGTCCAGAGGCCG	264
chr2	222298991	222299175	PAX3	CCGAGTGCGGGGATCCGGGCTCGGGAGCATTATTAGTTCTTTAC CCAAAGCTTGTCAGGAGCCCTGAGCTGCGATTGGCCGACGGGTA GACCGTCCCGGTGGCGGAGACACGCGTGATTGGGCAACAGCG ACCACTTTCTCTCCCATCTCTGGTGGTGCCGAGGCCTCTGCTGGCC CCG	265
chr2	233060290	233060637	INPP5D	CCGCAGCTCAGTTTCCTTCCCTCACTGAGCGCCTGAAACAGGAAG TCAGTCAGTTAAGCTGGTGGCAGCAGCCGAGGCCACCAAGAGGCA ACGGGCGGCAGGTTGCAGTGGAGGGGCCCTCCGCTCCCTCGGTGG TGTGTGGGTCTGGGGGTGCCTGCCGGCCCGGCCGAGGAGGCCCA CGCCACCATGGTCCCTGCTGGAACCATGGCAACATACCCCGCTC CAAGGCGGAGGAGCTGCTTCCAGGACAGGCAAGGACGGGAGCT TCCTCGTGCGTGCCAGCGAGTCCATCTCCCGGGCATACGCGCTCTG CGTGCTGTGAGTACAACCTGCTCCCTCCCCG	266
chr2	241882165	241882427	CXXC11	GCGTGGGTGGCTCCTGGTGGGGAAGTGAGAAGCCCTCCGTGCGG TGTCTCTGAAGCAGCCCCAGGCCAAGGCTGTGGCGTGTGTTGGTGG TGCTGTAGGCCCAAGATGTTTATGGGTGAGGGTCCCCGGGGCCG GGATTCTGATCCCTGGTGAGAGGTGGCTGGGAGGAAGTCCAGACG TGTCTGAGTGGCCATTCTCACACTGAGGTGACACCGCCTCTCCA AACACGTGACGTGGCTGGAAGCAGATGCTGTGTCGG CCTCAGCCTGCGAGGAGCGCGCCGCCGCCATCCCTGCGTCCC GTCCCGCGTCCCCGCGTTCCCGCGTCTGCGATCCGCCGCCATG	267
chr2	232633618	232633707	EFHD1	GAGGGCCAACGGCCCCCGCGCACCTGCGCCCCTCTGAAGCGCGC CGCCTCCCCGCGCCGGGACTGGGACCTGCCTCTGGGGAATCCGC CTAGAAGACGCGCGCGGAC	268
chr2 0	4822740	4822848	RASSF2 A	GCGATGGTCTGTGACCCTGCGCGGCTCAGAGCCTAGGGGACAGG GGCAGGAGCGGAAAGCGCGGCCTGATTACCGACGTGGAGACG CTGTCTGCTGCGCTTCTGGCGGCCGAGCGCAGGCGCGGACGCTG GGAGCCAGCGGGGACGCGGGCTCGGGGCCCTGGGCGGCAGGAA CGGCACGTCCGCTAGGAGCAGGCAGGGTGCTCGAGCGCCGCCG GCGGCTGCGTGCCG	269
chr2 0	25081598	25081833	VSX1	GCGTGGCTGTGTGCCGAATTGGTGGGTTCTTGGTCTCACTGACT TCAAGAATGAAGCCGCGGACCCTTGCGGTGAGTGTTACAGTTCTTA AAGGCGCGGTGTCGGGAGTTTGTCTTCTGATGTTTCGGATGTGTT TGGAGTTTCTTCTTCTGGTGGGTTCTGTTGCTGCTGGCTCAGGA AAGAAGCTGCAGACCTTCGCG	270
chr2 0	40969181	40969385	MAFB_ TOP1	GCGCGTCGCCAGGCTAACCTGCGTGGAATAATCGAGGTGGAAG GCGAGGCGCCTTATTGAGGGGGCCGGCAGCGCGCGCGCGCGG	271
chr2 0	50010033	50010192	SNAI1_ UBE2V1		272

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2 0	56627071	56627303	TFAP2C	CGAGGGGGCGCGGGGGCTGTGCGGCCCGGGCCGGAACGTGA GCCGGGCTGGGGGCGGCGACACCCCCG CCGTACAGAGGGCGCGGAGGTTGCGCTCCAGTTCGAACGCTTACC CATTGGAAAGAGGGCAGCGCCGGGTCCAGGGAAGCTCCTTGGG AATGAATGGCCTTTGCCAAGCGTTCCGGATCCTCTGGGTCTTTG GGCCCACGGCACGGTGCTGCGCGAGCCCTCAGTGCCCATCGGCTC CCTTCGCTCCTGCGTAGACGCTCCCAGGCGGGGAGGCATATCGGT TCCTCCG	273
chr2 0	57390020	57390254	RBM38	GCGGGAGCTGGGGGAGGGAGAGGTGAGAGGTCAAGGCTGCCGC GTGGAGCGTGGGCCGTGGAGTGGGGGAGGGGCGGGCAGACTC CTCCCCCGGCAGCCAGGGCAGAGGGCTGGAGGAAACGCGGAG AACTCCTCGGTGCTGGAGGAAACGAGGGGAACCTCCTCGCCGGCT TGCGGTCCCCACAGCCACGGAGTGCCACTCCAGTCCCCACAGA CCCCACCTGCGTCG	274
chr2 0	62603877	62604085	GATA5_ SLCO4A 1	CCGCTGTCAGTAACTGACAGGAAGGGGCGGGAGGCGGATGGGCC GTGACAGCTTAATGGCTTCGGTTAAAGCATCCTCTGATCGTGCTGG CGCTGGGAGAGGCTCTGAGCTCGGGTGGCACTGCGGGCACTCTGG ACACTGTCTCCGGCTGCCGCTGAGCTGGGAGGCTCCTTCCAGCAG GCCACGCGGTGAGGGGCACCTCCTGCCG	275
chr2 1	36696422	36696602	SIM2	CCGCGCCAGAAGGGAAAGACATAGGAGGTGTCCCAATCTGCGGTC ACCGCCGATGCTCCTGACCACTCTAGTGAGCACCTGCCCGTACTTT TCCATTCCAACAGAGCTTCCAGCTTCATACTAACTATCCACATACG GCCTGTGGGTATTAGCTCTAAGTGTCTTTTCCGAGGGCCCCG	276
chr2 1	36697196	36697567	SIM2	ACGCATTAAATCCTCCCGAAGCCAGGAGGTGCCAGAGCGGGCTC AGGGGGCCGCTGCGGAAGCTGCGGCAGGGGCTGGGTCCGTAGC CTCTAACCCCTTGAGCTCCTTCTCCAGAGGCCCGGAGCCGGCAG CTGTCAGCGCAGCCAGGAGCGGGATCCTGGGCGCGGAGGTGGGT CCGACTCGCCAGGCTTGCGCATTTGAGACCCGCGCCGCTAGCCCAT GGCCCTCTGCTCAAGCCGCTGCAACAGGAAAGCGCTCTGGATCCG AAACCCCAAAGGAAAGCGCTGTTACTCTGTGCGTCCGGCTCGCGTG GCGTCGCGGTTTTCGGAGCACCAAGCCTGCGAGCCCTGGCCACGAT GTGGACTCCG	277
chr2 1	36704462	36704653	SIM2_H LCS	CCGAGGGCGCAGAGGGGACAATCCGGGAAGTGGTAAAGGGGACA CCCGGGCAGAGGGCTGTGCTTTCGTTGAGGCGAGGAAGTGGAG CGCGCGCTGCAGATTCAGCGCGGGGCTAGAGGAGGGGACCTGGA TCCCTGAACCCCGGGGCGGAAAGGGAGCCTCCGGGCGGCTGTGG GTGCCGCGCTCCTCG	278
chr2 1	44185013	44185201	C21orf3 3_ICOSL G	CCGCTGTGGTTGAACTCCTACTTACTTTTCGGCAGATGGTGTTTGC CAAGTTAGTTTTGCAGCTGCCTGGGGTACTGGGTGGAAAGCAGC CCCGGGAACCCCATGGGGGACTTTGTGTCTTTTACTCCATCACAG CGAAGCCACGGGGCTGGGCCAGGCCCTGCCCTTTGGGAACGGGCT CCTCCG	279
chr2 2	19765578	19765956	TBX1_C 22orf29	CCGCCCCCTGCAGGAGGGAGCACCAGCTCCGTAGAGGAGGGGC AGACGTGGACTGGTTCTTGTGTCAGGGCAGCAGAAAGGCCCTTGGTG CGCTTCTCTAACACTCCCTATCCTCCGCCGAGGTGCGGTGGCCCA GGCTGCAGGGCTCCAGCGGCTTGCTCACACCCACCTCCCTGCAGAT CACGCAGCTCAAGATTGCCAGCAATCCCTTCGCGAAAGGCTTCCGG GACTGTGACCCTGAGGACTGGTGAGTGTCTCCCCGAGAGAGTG AGCGCCGGGCGCTGGCGCAGGCGCCGCCCTGATCCGCCTCCCGC	280

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2 2	19767183	19767448	TBX1_C 22orf29	CCGCAGGCCCCGGAACCAACCGGCCCGGCGCACTGCCGCTCATGAG CGCCTTCGCGCGCTCG CCGAGACCGCGCTCGCCCCGCGGCCCGGCCGCGAGTTGCAGTGTAGA CAGCCCCGAGAGCCCCGCCTGCAGGCGGTGTAGATACATGTAGATA CTGTAGATACTGTAGATACCGCCCCGCGGCCGACTTGATAAACGGT TTCGCCTCTTTTGAAGCCGCCTGCGTGTCCATTTATTTGTGCCAG TTAGATCGCGTTGGGAATCTTCGGGACAGCGAGCCCCGGGGTAGCT CAGGGCCCTCAGGGCCTCCCCAGCCCCAATCCCTGCCG ACGGAGAGAGAGGAGGCGACCACTGGGGCTCGGGCAACCATCCC GGTACCCCCGCCCCGCGCCAGGAGAGGAGGGAAGCCTTGAA GTGCCAGGCTTTGAATCGCCCATCTCCATGGCAACGCGTGGGCAC AAAGGGCCCGGCCGCGAGCAGCGCGCGCTGCG CCGGACCAAGAGGCTGGGGGAGGGGTCTCCGTAGGGACGGATG GGAGAGATACAGAGGAAGTAGAATGGCCAGGCTGTGGACTGCGG TAGGAAGTAGAGGTAAAGACAGAAGGAGACCCCCGGGATGGAAA CCCTGCAGTCTAGTTGAGGAGTGAAGGGGGCTGGGGGAGCCTG GGCGGTGGATTCTGCTGGCTGTCG CCGGGCCCTGGGCGGAAGGGATGTCTGCGTGAGTCAGCTGTGTCT GAGGAGGGGATCCTGGGCTGGGCTGGGCGGCCCTACTCGGCGGG TCAGGCGGAGGGGCGCGGCCGGATCCCG CCGGAAGTATGTCGTGCAGGGTTTCAGTGTTCAGTCAAAGCCCTGTC ATCATGGGGACAAGGTAGTTTCTTGGGAATTCATATACAGACT GTAGACCAGAAGTGTCTCAGAGTTCAGCCATGCCTCTGTACCG GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	281
chr2 2	20280279	20280448	RTN4R_ DGCR6L	ACGGAGAGAGAGGAGGCGACCACTGGGGCTCGGGCAACCATCCC GGTACCCCCGCCCCGCGCCAGGAGAGGAGGGAAGCCTTGAA GTGCCAGGCTTTGAATCGCCCATCTCCATGGCAACGCGTGGGCAC AAAGGGCCCGGCCGCGAGCAGCGCGCGCTGCG CCGGACCAAGAGGCTGGGGGAGGGGTCTCCGTAGGGACGGATG GGAGAGATACAGAGGAAGTAGAATGGCCAGGCTGTGGACTGCGG TAGGAAGTAGAGGTAAAGACAGAAGGAGACCCCCGGGATGGAAA CCCTGCAGTCTAGTTGAGGAGTGAAGGGGGCTGGGGGAGCCTG GGCGGTGGATTCTGCTGGCTGTCG CCGGGCCCTGGGCGGAAGGGATGTCTGCGTGAGTCAGCTGTGTCT GAGGAGGGGATCCTGGGCTGGGCTGGGCGGCCCTACTCGGCGGG TCAGGCGGAGGGGCGCGGCCGGATCCCG CCGGAAGTATGTCGTGCAGGGTTTCAGTGTTCAGTCAAAGCCCTGTC ATCATGGGGACAAGGTAGTTTCTTGGGAATTCATATACAGACT GTAGACCAGAAGTGTCTCAGAGTTCAGCCATGCCTCTGTACCG GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	282
chr2 2	20438412	20438611	SCARF2	CCGGACCAAGAGGCTGGGGGAGGGGTCTCCGTAGGGACGGATG GGAGAGATACAGAGGAAGTAGAATGGCCAGGCTGTGGACTGCGG TAGGAAGTAGAGGTAAAGACAGAAGGAGACCCCCGGGATGGAAA CCCTGCAGTCTAGTTGAGGAGTGAAGGGGGCTGGGGGAGCCTG GGCGGTGGATTCTGCTGGCTGTCG CCGGGCCCTGGGCGGAAGGGATGTCTGCGTGAGTCAGCTGTGTCT GAGGAGGGGATCCTGGGCTGGGCTGGGCGGCCCTACTCGGCGGG TCAGGCGGAGGGGCGCGGCCGGATCCCG CCGGAAGTATGTCGTGCAGGGTTTCAGTGTTCAGTCAAAGCCCTGTC ATCATGGGGACAAGGTAGTTTCTTGGGAATTCATATACAGACT GTAGACCAGAAGTGTCTCAGAGTTCAGCCATGCCTCTGTACCG GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	283
chr2 2	21652304	21652421	PPIL2_S DF2L1	CCGGGCCCTGGGCGGAAGGGATGTCTGCGTGAGTCAGCTGTGTCT GAGGAGGGGATCCTGGGCTGGGCTGGGCGGCCCTACTCGGCGGG TCAGGCGGAGGGGCGCGGCCGGATCCCG CCGGAAGTATGTCGTGCAGGGTTTCAGTGTTCAGTCAAAGCCCTGTC ATCATGGGGACAAGGTAGTTTCTTGGGAATTCATATACAGACT GTAGACCAGAAGTGTCTCAGAGTTCAGCCATGCCTCTGTACCG GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	284
chr2 2	26050281	26050416	SEZ6L_ MYO18 B	CCGGAAGTATGTCGTGCAGGGTTTCAGTGTTCAGTCAAAGCCCTGTC ATCATGGGGACAAGGTAGTTTCTTGGGAATTCATATACAGACT GTAGACCAGAAGTGTCTCAGAGTTCAGCCATGCCTCTGTACCG GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	285
chr2 2	27799853	27800150	MN1	GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	286
chr2 2	46007845	46008261	PPARA_ WNT7B	CCGGGCCCTGGGCGGAAGGGATGTCTGCGTGAGTCAGCTGTGTCT GAGGAGGGGATCCTGGGCTGGGCTGGGCGGCCCTACTCGGCGGG TCAGGCGGAGGGGCGCGGCCGGATCCCG CCGGAAGTATGTCGTGCAGGGTTTCAGTGTTCAGTCAAAGCCCTGTC ATCATGGGGACAAGGTAGTTTCTTGGGAATTCATATACAGACT GTAGACCAGAAGTGTCTCAGAGTTCAGCCATGCCTCTGTACCG GCGAGTCGACGGCTCCTTGTTTCGTACCCCTCCGTGGCTCCAGACT GTGGGAATCGAGCCGCTGGAGGACGGCAGGCGGTGGAAGGAG GCGGCTCGGTTAGGGCTCTGGTCCAGCGGCAGGCATGGGGCCGGC ACGGCGTGGCTGGAGGCACCTGAACTGTGGAAGTCCGGGAGGTTT CCCGGTGCTGCGGGCCGAAGCTCTAGGCCCTGGCTCTCCGCCA TGTGCTCATAGCCCTCGGCGAAGGGCGGCTGGCTGCCAGGCCTC CGGCTGCGCCCGCTAGCCGAGCAGGCG CCGGCTCCTGTTCTCAGCCTGCCAGGCCCTTGTGCGGTGGCGTCGG GCAGGCAGGGCAGGGAGGCCACGGCAGCCATCTTCCGCGGGAGC TGGGGCTGGCCAGCAGCGTTTCCAGTGGCCTCCTCTGTGCTCC GAGCTGCATTACCTCATCGGGAAGCCATTCAGAAAGGAGCTGCG GAGCCCCGAGTGGGAGTGGGAGTGGGAGAGCTGCGTCAGCGCCCTC CTGGCAGCCTCGGTGCCAGACGAGGGCAGGCGTCACGCCTCCGGG TGTCTGCCTGCCGAGCGACTGCTGGGAGAGCAGTGGCTTTTGTCA GCGTTTCGGGGTGACCGGGCTGGGCTGCAGCAGGCAGGTGGCGT GGCACGGCCCATGGCCGGCCAGCTACCAGGTGGGCAGAGGATCTA TTTCAAGAGCCG CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	287
chr2 2	46525740	46525932	CELSR1_ TRMU	CCGCTGCCAGGGGCCCCAGACCCCATCTACCCACCTATCCCCTTCCT CAACAGGTTCTGCTATCGGGTTTCAAAATGTGCAGGCACAGGCACC AGCCCCAACCAAGGGGACCCCTCAGCAGCGACACTGGGGCCAGC GTGGGGTTCTGGCACGCCAGCAACATGGGCCCGCTCCAGGCGTG GCCAGCACCG ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	288
chr2 2	50548840	50549077	TYMP_S YCE3	ACGTGCTGGCCTTTGCCAGCAGCACGGAGAGCCCGGCTGGCGC AGGAGACCTACGCGCTGATGAGCGACAACCTGCTGCGAGTGCTGG GAGACCGTGCTCTACCGCCGGCTGAGCGCGGCCGACCGCGAGC GCATCCTCAGCCTGCGGACCGGCCGGGGCCGGGCGGTGCTGGGC	289

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr2 2	50577928	5057817 9	CPT1B	GTCCTCGTACTGCCAGCCTCTACCAGGGGGGCGCTCAGGGCTCC CCAGGGGGCCCTCG TCGGCACCTAGGACGGGGGAGATGGGTGCGCGGGCGCGCTTAG GCCGGCCCCGCCGAGCCGCGCCGAGACGCCCCAGCCAGTCCG CGACCCCTCGCGCCCCACCCCGGACTAGCGGCTGCCCCGGCC CGCGCCCCCGCCAGGCCAACCCCGCCAAATCCTCGCGCCAGCCT TCCGGGTGGGCACAGCCACTGTGGTGACAGGGGATTGGGCCTGA AAGCTCCAGGAGCCCCAAGGACGGCG CCGGGGTGGCTGAAAGCGGGCTCCTAAGCCATCTCTCGGATTCTT TCTTCGAGACGCGAGCAAGCTCCTGGCACCCTGTAGTCTCTCCCTC TCCCTTCTGTATTTCGGCCAACGACCGACATCAGGCCATTCTTTAT TAACCTTTATCAAGCCAGGCCGGTCAGCG	290
chr3	9136446	9136614	RAD18_ SRGAP3	CCGGGGTGGCTGAAAGCGGGCTCCTAAGCCATCTCTCGGATTCTT TCTTCGAGACGCGAGCAAGCTCCTGGCACCCTGTAGTCTCTCCCTC TCCCTTCTGTATTTCGGCCAACGACCGACATCAGGCCATTCTTTAT TAACCTTTATCAAGCCAGGCCGGTCAGCG	291
chr3	52794515	5279469 1	ITIH3	ACGCCAGGGAGTCCCAGGGTCCATTGTGTTGGCCACAGCTTGTCT TTCTGCGGGCCTCTTCCGAGTCCCCGGTCTCTCAGGAATGAAGC CCCAACGCTGTTGGCAAGACCCACCAGGCCTTCTTCAGACAGACAC ATCGAGGGACCCGTCATTCCACCCATGCCAGCTTCCCG GCGACGCTTGGCTAGGCGGGCGCGACCTCTTCGAGTGAAGAAGTT GTCAAACCTCGTAAGCGTCAAGCCGGGTGCTCTCCGACAAGACCG AGACTGAGTCCCGCGGAGCCGCTCTGCGCTCTGCTCTGCCGCCA CAGAGGCTGGTGCAGCTTCCCTCCCGCCGCGCTCCGCGGGCCGGG AAACCTTTGCGTAGCCAGAGACGCACCGAGTCTTCTCTGGCTG ATGCCTCGTAGAAGAAATTCGCACG	292
chr3	62370444	6237069 7	FEZF2_P TPRG	GCGACGCTTGGCTAGGCGGGCGCGACCTCTTCGAGTGAAGAAGTT GTCAAACCTCGTAAGCGTCAAGCCGGGTGCTCTCCGACAAGACCG AGACTGAGTCCCGCGGAGCCGCTCTGCGCTCTGCTCTGCCGCCA CAGAGGCTGGTGCAGCTTCCCTCCCGCCGCGCTCCGCGGGCCGGG AAACCTTTGCGTAGCCAGAGACGCACCGAGTCTTCTCTGGCTG ATGCCTCGTAGAAGAAATTCGCACG	293
chr3	125141549	1251421 16	HEG1_S LC12A8	ACGCCGCTGGGGGCTGCTGAAATTAGAAGAGGGAGTCGGGAAGT CATACCCCTCCCTGTGGGCGTCGGGTTGCACTGTTGACTAACTTAG AAAGCGAGATTCTAAAAATGATGCTGGGGCTGCAGGCTGCGGGC TGCGGGCTGCGGGCTGCGGGCTGCTGCCGCGGCGGGGGCTTCCG GCGGGCTCTTCTGCGGTCACCCACCCCTGGACCAGCGACCGACG ACCAGCCAGACAGCCCTTCTGCGAATGGACAATGGAGAGAGGCT GGCGCAACCGAGAATAGCCAGCGCGGAGGAAGGGCTCCGGACGG AGCTAGGAGGGTGGGGCTCGGAGGGCGCAGGAAGAGCGGCTCTG CGAGGAAAGGAAAGGAGAGGCCGCTTCTGGGAAGGGACCCGCA CGACGACGCCCGAAGGGCGTCGGGGGAAGTGGTAGGCCCGGAG ACTGCGGAGGCTCCTCAGCAAAGGAAGTGGGCGCGGCGCGCAC GCAAGACCTCGCACCCGGCCTCGCGCGCCGCTCTGGACAGCCCA GCGCCTCTCAGCACCTGTACCTCGCCAGACGCG GCGGGGGCGGCTGCGCTCAGCGCCCTTCCCGGCGGCCGCGACCC CTCCCCGCTGACCTCACTCGAGCCGCCGCTGGCGCAGATATAAGC GGCGGCCCATCTGAAGAGGGCTCGGCAGGCG	294
chr3	129974607	1299747 28	TRH	GCGCAAGCCCAAGAACCTGCTCAAGAAGGACAGGTATGTCTTCCC TTGCCCTACCTGGGCGACACGACCCGCTCAAGGCGGCTGGCCTG CCCGTGGGGGCTCCGACGGCTCCTGAGCGCGCCGAGAAAAGCC CGGGCCTTCTTGCCCGGCTCGGGCCCTACTCCCTGCTGGACC CCGCGAGTTTAGCTCGAGCGCCATCCAGAAGATGGGCGAAGTGC CCCACACCTTGGCTACCGGCGCTCTGCCCTACGCGTCCACCCTGGG CTACCAGAACGGCGCCTTCG	295
chr3	137765024	1377653 16	SOX14	TCGCGCCCCAAGACCTGGGCTTGACGCGCCGCCAACAGGCCCGGG GACACGAGGCGCTCCAGGCCGGGGTCTCCCGGCTGCTGGCCCT CTCGCTCCCCACCCGCTGGCGGCGCTCGGTGCGCCGCAATTGACC CAACCCGCTTCTGCGTTTCCCCCTCAGGTTTCCCGTTTCTCCACAA AGGCCTAGGGGAGCCTCG	296
chr3	138939663	1389398 63	PIK3CB_ FOX12	TCGCGCCCCAAGACCTGGGCTTGACGCGCCGCCAACAGGCCCGGG GACACGAGGCGCTCCAGGCCGGGGTCTCCCGGCTGCTGGCCCT CTCGCTCCCCACCCGCTGGCGGCGCTCGGTGCGCCGCAATTGACC CAACCCGCTTCTGCGTTTCCCCCTCAGGTTTCCCGTTTCTCCACAA AGGCCTAGGGGAGCCTCG	297

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr3	138940084	138940265	PIK3CB_FOXL2	CCGCTTTGGGGGAAGCGAGAGGGAGGTTGAGAGACCCCGGGCG GGGTCTCAGCGCCACCAGCTGTGCCTTCAGGGCTTGGGTGTTGCG TGCAACGGCAACCGCGTGAGCCTCACTCCACGGCCAAGGGGCTA GGGCAGGGTGGATGCAATCGCGTGCCTGGCCCCGAAGGTGCT CG	298
chr3	147392077	147392581	PLSCR1_ZIC4	CCGCACTGACTTGCGATGTCGACCGGTCTGCCAGACCACCCAC CTGGCTGTGCGGCCTCTCGGTCTTAAGACGAGGGGTGGCGCGGT AGGGTCCGCACAGGCCAAATGGGATCCGAGGTGTCTACCGCAACC ACGCCCTTGAGCGCTGCGGCTTCGGGAAGAAAACAGCTGCTGCTG TCAGGCCAGGCCTGGCTCCGCAGCCCGAGGGGCCACGAGCGGCT GGCATAGCCGGGGAGGGGCTGGGATCGGTGGCTGCGATGCCCT GTAGAGCCGAGGGAAGGCGCGAGTGCACGTTAGAGTGACAATATT GGCCGGACCGAGCCCAATCGGGGAGCTCACGGCCAGCTGAATTC GCTGACGTGTAGGAGAGGAAAGGACCCCGAGAACCCGGAAGCCT AGATTCTGCCGAGCTGCAAGTGTGCGGAAATGGGGGAAGAA GGTTTCTGGGCGCTTTAAACAAATGGCTGCCTCCCAGCGCTCTGAG TTAAGGGACCG	299
chr3	158094327	158094687	VEPH1_SHOX2	GCGGCCTCTGTCTCCGTTAGTCTTGGGGGAGCAGACGCAAGAGG AGGCAAGGGCGCCGCGAGCTCCCCGGATGCACTGGTCCCACAGGC CGTGCCCGAGTGGAGCACTGCGAATGGGGCCAAGAAATTTGGCC TTTCTCGCCGACCTGGCTGCCTCCGCGGGCTCTCCGCTACCGC GCTCCCGCCGCGGCCGACTCCCGCGGGTCTCCGCGCCGAACCCAC CTGGCTCCTATCGCACGGGACATTCGACCCACCCACGCCGCGTC ACTGAGCCTCTGTACCGATACCCGGCGCTCCGCCAGCAGGGCCTG GACGCACCGCCTCCTTTGACCTCGGGCTTCCCCCGCGCTCCG CCGACCTCCGACCGATTGCGAGCACCCACCCAGTCGGGGCCAT CCATCCACCTGATTAACCTCGCCGGCAGCAACTCCAGCGTAGAAAG TAGGGCAATGAACACACACAGTCGGTAAAGCAGGAAGCCACAGA CCTGGCCAATGCACCCACCCTGTTACCAACCCACCCCGCTGCGCA GGGGGCGAGCCG	300
chr3	171458160	171458353	SLC2A2_TNII	CCGACCTCCGACCGATTGCGAGCACCCACCCAGTCGGGGCCAT CCATCCACCTGATTAACCTCGCCGGCAGCAACTCCAGCGTAGAAAG TAGGGCAATGAACACACACAGTCGGTAAAGCAGGAAGCCACAGA CCTGGCCAATGCACCCACCCTGTTACCAACCCACCCCGCTGCGCA GGGGGCGAGCCG	301
chr3	171458454	171458456	SLC2A2_TNII	ACG	302
chr3	188295081	188295168	TPRG1_LPP	ACGTGTGTAGAGGCTGAAGGAGAGCTGTGTTGCTAGCTTTGTATTT GAACGGTTCGTACACAAACAGTTCTTTTGATTAAGTATCCG	303
chr3	192408009	192408504	FGF12	CCGGGCTTCTACTGACCTGGTCTCCGCCTCACC GGCTCTTGCGGCC GCTGCAGAAGCGCACTTTGCTGAACACCCCGAGGACGTGCCTCTCG CACAGGGAGCGCCCGTCTTTGCTGGGGCTGGAGCGGCGCTTGGAG GCCGACACTCGGTCGCTGTTGGACTCCCTCGCTGCCGCTTCTGCC GGATCAAGGAGCTGGCTATCGCCGACGCCATAGCTGCTCAGCGAG GGCTCAGGCCCCAGCCTCTACTGCGCCCTCCGGCTTGCCTCCGC CGGGGCGAGGGCAGGACCTGGGCGGCCAGGGAAGGGCAGTCG CGGGGAGGCAGTGCTAAAATTTGAGGAGGCTGCAGTATCGAAAAC CCGGCGCTCACAAGGTTAGTCAAAGTCTGGGCAGTGGCGACAAAA TGTGTGAAAATCCAGATGTAACTTCCCCAACCTCTGGCGGCCGGG GGGCGGGGCGGGGCGGTCCAGGCCCTTTCGCAAGTAGACG ACGTGCCAGTTGGTGGCTGCGACTGGAGGAGGCCGGATCGGGGG TCCTAGGAATGGAGCCTCTCCGGACAGGGCTGCTCGGGGCTGCTG TGCTTCCCTAGGGGCTGAGGGGACCCACCGAGGCTTCTCATGA TGGGCACAGCCCGTTAGGAGTCTGGGTGCTAGAAACATTACGCGT CTGTGGCCCTCCATGCTTCTGTGTGCTCCTCACCTGCCG	304
chr3	196640742	196640962	NRROS_CEP19	ACGTGCCAGTTGGTGGCTGCGACTGGAGGAGGCCGGATCGGGGG TCCTAGGAATGGAGCCTCTCCGGACAGGGCTGCTCGGGGCTGCTG TGCTTCCCTAGGGGCTGAGGGGACCCACCGAGGCTTCTCATGA TGGGCACAGCCCGTTAGGAGTCTGGGTGCTAGAAACATTACGCGT CTGTGGCCCTCCATGCTTCTGTGTGCTCCTCACCTGCCG	305

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr3	196660739	196660989	NRROS_CEP19	ACGCTTCACATTTCGGGAGCACGAGCCCCCGGAGCGCTCACCGAGCTGGACCTGAGCCACAACCAGCTGTGCGAGCTGCACCTGGCTCCGGGGCTGGCCAGCTGCCTGGGCAGCCTGCGCTTGTTCAACCTGAGCTCCAACCAGCTCCTGGGCGTCCCCCTGGCCTCTTCGCCAATGCTAGGAACATCACTACACTTGACATGAGCCACAATCAGATCTCACTTTGTC	306
chr3	50340611	50340770	RASSF1A	CCCTGCCAGCTGCCTCGGACCGGCACACAGTGTGCGTGGCGGGCCCCGCGGGCTGGAAGCGGTGGCCACGGCCAGGGACCAGCTGCCGTGTGGGGTTGCACGCGGTGCCCCGCGCGATGCGCAGCGCGTTGGCACGCTCCAGCCGGGTGCGGGCCCTTCCCAGCGCGCCAGCGGGTGCCAGC	307
chr4	3310274	3310386	RGS12	CCGTGTGCGGGAGGAGCTGGGACCCGGGAAATGGCAGGTGTCTCTGAGGGGAACCGGGCGGGAGAGAGCTGGGGCCTGGAAGGCCAAGGCAAGGGCTGTCTCCAGTCCACG	308
chr4	8580403	8580763	GPR78	TCGCTCCAGTTTGGTGCCAGCGCCTGGAGGGAGAGGCGTGCGGAGGGCTGTGCTGCTAGGATCCACTGAGTGGCTCTTGCTGGCGTGTGACGCTGCGCGCAACCAGGGCTGGGAGGCTCGGCTGGAGGTGTGACCAGGGCAGGGACTGACCTGGCCCCGGAACAGAAGCGCGCAGAGTCCATCCTGCCACGCCACGAGGAGAGAAGAAGGAAAGATACAGTGTAGGAAAGAGACCTCCCTCGCCCCTACGCCCCGCGCCCCTGCGCCTCGCTTCAGCTCAGGACAGTCTGCCGGGACGGTGAGCGCATTACGCACCTGGACAGCACCGCGTTGCGCTGCCTCCAGGGCGGCCCCG	309
chr4	8858107	8858297	HMX1_CPZ	CCGACCGCCCCAAGCCGGTCGAGGCCCCCGTCCATTGCGGGGAAATGGATTTTCGCGATTAAAGAAACAAACCCAAATCAAATGAGCGAGGCCCGGATGTGCTGACGCTGCGGTTACGCGCGCGGAGCTGGAGCCCGGAGAGCGCTCTAGGAAAGGCGCAGCGGCGACCGCGGGAGGGGGTGAGAAGCCG	310
chr4	8861458	8861732	HMX1_CPZ	TCGGGAAAGGGGGGTAGGGAACGACGGGGGAGCCTCGGTGACCAAGGAGATGCACGCGCGCGGGATCTCGTGCGCCGCAAGAGGGACGAGCAGAGGAGCATCGGAAGAAGACAGGCGAAGGGGACCGGGAGCAGCGTAGGCGGAGCCCCGGGGGCACGGCCGAGGCTGCGCTTCAGGAGTGTCCGCCAGGCGCCTTCCCGGGCGGTTGGCGAAACCCGAGGAGGCCACAGCTCTGGCCTGGGGCGCCGTCTGTCCAGGGCCTCTGCG	311
chr4	13541900	13542257	RAB28_NKX3-2	GCGGGGCGCCCCGTGCAAGGTACAGCTACAGCTGTACAGCCCGGTCCGGAGCCGGAGCGCGGGAATCACTCGCTGCCTCAGCCCAAGCGGTTCACTGGGTGCCTGCGGCAGCTGCGCAGGTGGAGAGCGCCCAAGCTGGGAGGCAAGTACGGGTAATAGTAGGAGGGCTGCAGTGCAGAAAGCGAGGGTGGCCGACGCACTTCGCCGGCAGGTATTGTCTGTGTCGTGCGCACCAAGCACCTTACGGCCACCTTCTTGCGGCGGGCGCCGAGGCCAGAGTTCGGCTGCCATCTGCCGGCGCTTGTCTGTAGCGACGGTTCTGGAACAGATTTTCACTGCGTCTCG	312
chr4	24800039	24800445	SOD3_GI2	TCGTGGGCGGGGCGGTGCTGCTCCACGCTGGCGAGGACGACCTGGCCGCGCGGCAACCAGGCCAGCGTGGAGAACGGGAACGCGGGCCGGCGGCTGGCCTGCTGCGTGGTGGGCGTGTGCGGGCCCCGGGCTCTGGGAGCGCCAGGCGCGGAGCACTCAGAGCGCAAGAAGCGGCGGCGCGAGAGCGAGTGCAAGGCCGCTGAGCGCGGCCCCACCCGCGCGCGGCCAGGGACCCCCGAGGCCCCCTCTGCCTTTGAGCTTCTCTCTGCTCCAACAGACACCTCCACTCTGAGGTCTACCTTCGCT	313

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr4	38671431	3867152 4	KLF3_TL R10	TTGCTGAAGTCTCCCCGAGCCCTCTCCACCCAGAGGTCTCCCTATA CCGAGACCCACCATCCTTCCATCCTGAGGACCGCCCAACCTCG GCGTACTGAGACAGGGTGGGCAGCAGGGGCCAGTTGGAAGGAGT GGAAACTGTCTACTAATGTAAACAGACTGTCCCACGTTCTGTCTTCT CCG	314
chr4	38671667	3867166 9	KLF3_TL R10	ACG	315
chr4	44447507	4444778 0	KCTD8	GCGGCGGCTCAGCAGGGGGCGAGGGGTGCTGGGAAACGCCGGG GCTGCGAATTACGGAAGAAAATGTAAGTCTGAGTGTAGCTCTCAG ATCTTGTCGTCGCGGTACTGGTTGACGAAGCGGCGGTGCCCCGAG GAGTTACACGCCACCATGTGGAAGCCGGCCTCGGACAGGCGATCA AAGGCCTGCTCCAAGTAGGTGAACTTGAGGTAGAAGCGGGACGTG TACTTCTCCGGCTGCCGGTCGGGGTCGCGGCTCTCGTTGAGCGTGT CCCCG	316
chr4	56655130	5665585 3	HOPX_A RL9	TCGGCTGCCGCTGCCGTCAGCTGAAATGTTAGCTATCTACCGTCTTA TAAACGCCAGGAAAAACCTCTAAACCTTAGAGCCGGGGAATTTT TAAAAATCGGAACCAATCTCCGTGGCTTCGTGCAGCGTGAGTTC TGCAGCTCGGGGACGCTGCAGTGTGATGTGGTGGAGAGAGCAT GCTTCACCGCTCCTGCCATCCTGACAGCGCCCTCCCTCCCGGCTCA GCCTCCTGGTTCGCCAAACCGGAGGACTGAATTTATGGCTAGCTGG TCTCTGGGGCGCCTTCCAGCTCTGACATTCGCGCTAGAAATAGATCT TCCCGAAGGTTTCGCAGACAGACCAGAGGGGACCGAGCCGGGAA GGCGAGACAGGGACAGGCGAGAGACGCTGCTCCCAACTCGCAGA GGGAGAAAAGCGTGTATCCCGGGCTGCCGGGGAGAGTGGAAAAGA AAGGACTGGTGACCGAGGGGTTTCTGCGCAGCTCCCGGGGAACCA CGGCTGGATGGGGGTGGCGGGGAGACCGGGCGCCCATGGGAGC GGGGAAGCGGGGAGGCGGCGGGGAGCCATGCAGGGTCTGGG CCCCTGGGATGCGGGCAGAAGCGATGGGAGATCATGGGAGGGG AGCCCGGCGGGAGGCGCGGACGAACAGGACCGCCAGCCGCGAG AAGGCTCAGCCAGGCAGGGGTGCGGGCGCGCTGGGCGCGTGTG GGGACG	317
chr4	73998536	7399868 7	CXCL5	TCGAAGGACCGGGGACACGGGCCGCGCGGCTGGACAGGAGGCTC ATAGTGGTCAAGAGAGCGCTGCGAGCGGTGCGGGTTCCTGAACT GGGTGGAGGAGCGGAGATTGGAGGAGCGAAGATTGGAGGATCC GGAGCACTGTGGCTTCCTCG	318
chr4	93834642	9383485 6	SMARC AD1_AT OH1	GCGGAGCGTCTGGAGCGGAGCACGCGCTGTAGCTGGTGAGCGC ACTCTCCTTTCAGGCAGCTCCCCGGGGAGCTGTGCGGCCACATTTA ACACCATCATCACCCCTCCCGGCTCCTCAACCTCGGCCTCCTCTC GTCGACAGCCTTCTTGCCCCCACCAGCAGAGCTCACAGTAGCG AGCGTCTCTCGCCGTCTCCGCACTCGGCCG	319
chr4	110612859	1106131 56	PITX2_E NPEP	ACGCGCCCAAGAATTGGGCTGCCACTGGTATGGGTCCCAAGTCAC ATTCAATAAGCTGCCACCGCTTCTGGGGGACAGCAGTGGTGGTT CTAGGTCTCATCTTCCAGAGCGACGAGGATAAAAGTTCTGCCCA GGACTGTGTGCGAGGGGGTCCCGCACTGCTGCAAACCTCTCAGCGG AGGCAGAGAGGCTTTGCTGTTTCTGGAGAGAGGAAGCATTGGCAG AGGCAGTCTCCGGGCTGTGAGGAATCCACCCTCATGCCTTAGTGTG GGTACGTCAAGTCCCAGCATCAGCG	320
chr4	139735767	1397359 15	MGST2 _MAML 3	CCGTGCGTCCCCGGCAGGACCTAGACTGCCTCTCGGCGCAGGCGG CCCTAACAAAGAAGCCACGAGGCGGTCCCGGGCGCGGGCAGGG	321

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr4	146638834	1466393 36	POU4F2	GCGGTGCGGCGGCGCTCGGGAGACCCGCGAGGGGCCCTGGAGGT CCTCGGCCCCGCGCGG GCGCGGGGGTAGGCGCGGGGAGAGGGGAGTATAACTCGCCGGCC GCGAGGAGCGGGGGCAGTTTCGGGTGCCGAGGTCTGCAGCTAGC GGCAAGCGGAGTCAGGCATCCGTTCACTGACAGCAGAGGCGG CGAAGGAGCGCGTAGCCGAGATCAGGCGTACAGAGTCCGGAGGC GGCGGCGGGTGAGCTCAACTTCGCACAGCCCTCCAGCTCCAGCC CCGGCTGCCCCGCACTTCTCGGAGGGTCCCGGCAGCCGGGACCA GTGAGTGCTCTACGGACCAGCGCCCGCGGGCGGGAAGATGAT GATGATGTCCCTGAACAGCAAGCAGGCGTTTAGCATGCCGACCG CGGCAGCTGCACGTGGAGCCCAAGTACTCGGCACTGCACAGCAC CTCGCCGGGCTCCTCGGCTCCCATCGCGCCCTCGGCCAGCTCCCC AGCAGCTCGAGCAACGCTGGTGGTGGCGGCGGCGGCGGCGCGG CGGCGGCGGCGG	322
chr4	153788339	1537886 38	SFRP2	TCGGTGGCTGGCAGGAGGTGGTCTGCTGCTAGCGAGGGGGATGCA AAGGTCGTGTCTCGGGGAAACGCTCGCACTCAAGCATGTCGGG CCAGGGGAAGCCGAAGGCGGACATGACCGGGGCGCAGCGGTCT TCACCTGCACGCAGAGCGAGTGGCATGGCTGGATGGTCTCGTCTA GGTCATCGAGGCGAGCGGGGCGAAGAGCGAGCACAGGAATTC TTGGTGTCCGGGTGGCACTGCTTCATGACCAGCGGGATCCAAGCG CCGGCTGCTCCAGCACCTCCTTCATGGTCTCG	323
chr4	154742893	1547431 15	LRAT	GCGGACAAAGTTTCGGTGGGTGAACTGAAGCTGGGTCCATGTGAC CCTGAAGCCGGAGAAATAAACTTAACATGAATCTTGCTTTCCTGGC GGGCGTTGGGACCCCGCCGTTTTTCATGCCAACCGTTGGAAGCTTC GTACTCAACGGCCACAGGTGCCTAGGAGCGCAGAGAGGCCTCGGG TTCAAATCACCGGCGCGCAGGGACTGGACTCGCGGGTAGCG	324
chr4	157220224	1572205 43	GRIA2	ACGTAAGACAGCAGGGCCTGGTGAGAGGACGCTTCGCCGCCAACA ATTAGCAATTCGGCTTCTACACAGCAGCCGGAGATCAGCTTTGCTG CATTGGTCCAGGTTGGAGCATCTCCGCAGCAGCTGCAACAGCCGC ACGAAGGTAGCTCCGGGCGGGGAGCGAGGCGCTGTCTCGGTGCT GAAAGGCCGAGGCGCGGTTGGCGCGGACAGCCCGGAGACCCG AGGTCTCGCGGAGGGACAGCGGCTACGGGCCCCGAGCTGTGCTTT CTCAGCGCCGCGCACGCGACGCTCCACGGTGGTGGGGGTGCCG GGCG	325
chr4	190019091	1900192 95	FRG2_F RG1	TCGGGCGCCTCAGCGGTCTGCGCGTGGTCTGGCCGCCGCGGATA GCGGGACGCTCTGCGAGGCCGCGGAAAACGAGCGCGCGGCACT GGTGCTTGGGCGTATAGAGGGGGAGAGCAGCCCGGCCGCGGGCG AGCGGCTCCGGGGGTGCCTGATCCCAGCCTCGCGGCCCGGGTTG GTGGTGACGCCTGGAATCAGACGCGCG	326
chr4	81031194	8103125 8	BMP3	GTTCAACCCTCGGCTCCGCCCGGGCTCCTTGCGCCTTCGGAGTGT CCCGCAGCGACGCCGGGAG	327
chr4	153789264	1537893 86	SFRP2	GCCGCCGCTCGCCCGCCTAGGATTTCTTTAAACAACAAACAGAGA AGCCTGGCCGCTGCGCCCCACAGTGAGCGAGCAGGGCGCGGGCT GCGGGAGTGGGGGGCACGCAGGGCACCCCGCG	328
chr4	83114709	8311483 0	PLAC8	GGAGAGAATCTCACCACAAATGAAAACACGTGAAAGGGGAGAG GTAAGTGTGTTTCTATCGCAGGGCATAGTACATAGAACAGTTTCAG ACGCTCTTATTGGCCAGAGTAATCCAGCAGAA	329
chr4	80266629	8026675 8	FGF5	GCGTTATAAATATCCCGGTGCCAGCGCGGAGATCCGCTCGGGTGG CCTCTCTTCCCTCTCCCTTCTCTTCCCCGAGGCTATGTCCACCC GGTGCGGCGAGGCGGGCAGAGCCAGAGGCACGCAGCC	330

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr5	1876171	1876370	IRX4_N DUF56	TCGCTCGCCAGGCCGGGGGCTCCCGCCGCAGCCTTTTGACAGGCAC ATGAGCCGCGAGCTTCCGAACCTCGATAATATCATCTCGAGCGCGA AAGTCAATACGGTGACAGCGCGCGCCGGATACAATCCAATTACG CTCGGCTGCCCGGGGCTCCTGGGGCTCGGGGTCCGGCGGCCGAG GGTCCCCCTCAGGGCCCCG	331
chr5	1877759	1878263	IRX4_N DUF56	CCGGTCAGGCTCAGGCCCAGGCGGTGGAGGCCCGGCGTGCCAG CGCCGGGCTTGTCCATGTTCCAGGAGTCCAAGTTCAGAAGCCCC TCTCCGGTGGGTGGCGGCTTCGCGGTGGCCGCGCTAGTCTTCCTC TGGAACTCAGTGAAAAGAGTCGCGCGCCGTCCGCTGAGCGCGGG TTCCCTCTGGGCTCGGGACCCGCCGCTCAGGCGCAGAAGGGTT TGCCGCCGCTTGGGCAGGGCGAGCAGCTCCCTGGCGCGCCTG CAGCTGGGGCGTCTGGGGCACGGCAGGCGGAAAGGCGCGGGCC AGGGGTGAGTCAGCACGTTTCGCGCCCGCCCCAGCGAGCTCCC AGAGGCCCGGGTCCAGGAGGGCGCCCTTGGCGGTGGCCAGGC CTGGTTCAAAGTGCTGTGCCTGAGGATGGGGTCGTGGAAGACCCC GTCCACCCAGTTTCTGAGACTGGTTACCGGGGAGTCTGTGCTGTG TCCAGGGCG	332
chr5	2149349	2149539	IRX4_IR X2	CCGAGTGAGCAGCTGGAAGCCCGGGGTTAAGTGTTATTGACTTCA GAGCAGCAGCAGCGTGATCGGGTTTCAAGCTAGTTCCTTGAATT AATTTTGTGGATCGGTGTTTGAAGTTGGGTGGGAATAATTGGCC TGGGAGAGACTCCTTGCATCCTTGCCGGGTAATGAAGCTGGAGGC AGGCGTGCG	333
chr5	5140414	5140596	ADAMT S16	ACGCTGCCGGCCGGGACCCTCCGGTGGCCCTAGCCCCCTCGGAG CGTCTCTGGATGAAGCCCCGCGCGCGGATGGCGGGGCTTGGCG GCGCTGTGGATGCTGTTGGCGCAGGTGGCCGAGCAGGTGAGTCCC GGGCGCTCCCACCAGCGCGGAAACCGCGGGTCCGGACAGCTGGA GGCG	334
chr5	16179848	16180346	11-Mar	CCGAGAAAGCGAGGGGGCGGGAGGGAGGAGCGGCGCGCGGG GGTGACGGGGCGCGGGCGCGGGGTGGGCTGGGGGCGCGGATCA GTGGGACGGAGTTCGGGGTTCGGCTCCGAGCGGGCGGGCTGGAA GTGGGGGATCCCTCAGCCGCTCCACGGGCCGGCCCCGCGCTCAC GTCGGTTCCGGGGCGGATGACCCCTCTCAAACGCGCGCAGCGCTG CGGCTCTCGTGAGCTGGGAAGTAGGGGGCAGGGGAGAGGCCGCG GGTCCAGAAACCGTTACTGGATGGGCCGGTGGGATGTGGCGCGG GCCGGGTGGGGCGCGACAGTCTGAGCCGAGACCCGCTGGGCTT AAGGGTGCGCGAGGCGGGTGCCCTGGGCGCGCCGAAGTGGCTG AGCAGTGGAGCGGGAAGGGCGCGGGACCCGGGACTGTAACCGC CACTTCCAGGCCCTCGCTCCCCGCGCTTGGAGCCCTAAGGGCACT CTCAGGGATCCTCG	335
chr5	40681066	40681252	PTGER4_PRKAA 1	GCGGTGATGTTATCTTCGGGGTGTTGGGCAACCTGGTGCCATC GTGGTGTGTGCAAGTCGCGCAAGGAGCAGAAGGAGACGACCTTC TACACGCTGGTATGTGGGCTGGCTGTACCGACCTGTTGGGCACTT TGTTGGTGAGCCCGGTGACCATCGCCACGTACATGAAGGGCCAAT GGCCCCG	336
chr5	73298958	73299150	TMEM174_FOX D1	CCGGGCACGGAGGTTTAATGTGAAGCATGTGAGCGGGGCTCAGTT TACAGGTACGCGGGCCGATGGCGAAGAGCGCTGTCAAGCGGCCTC GAGGATTTCGGGGGTTTTCGCGCCCGCGAGGAAACCTACCCGAC GAGGCGAGCAGCTGGTGGCCCTGGCGGCCGCGAGCTCCCGGCTG CCACCGCTAGGCG	337

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr5	73419497	73419688	FOXD1_TMEM174	CCGGGCCGCGCGGGAGCGGCCGGGCGCAGCTGACCACGGGTACAGATAGGTTAATTTCCACATGGAGCTGCAGAAACCTATCCGCGGGTTGCGAAGCGTGGGTACGCCAAGGCATGTTAATCTGTTTAGCATGTGCGCCGCGCGGAGGAGCCAGACCACCGGGGCGCAGGAGGCGCGGCCGCAGCCGGCG	338
chr5	76953575	76953986	AGGF1_CRHBP	CCGGA CTGACCTATGTTTCTTGCCAGCTGAGGGAAGCGGCGGACTACGATCCTTTCTGCTCTTCAGCGCCAACCTGAAGCGGGAGCTGGCTGGGGAGCAGCCGTACCGCCGCGCTCTGCGTGAGTCGAGGCTGCCCGGCTCGCGGGCGCCCGGGACGCGGGGAAGGTGGGACTCTGTGCGGGGGCAGAGGGCTCGCGGACATCTCGGGGAAGGGGCTGGCCGGAACCGCCAGGGGCGCGGTCCCCTTAGCTAAGGATCGGTCCGCGGAGGCGCGCCAGGAGCGGGAGAGGGTGGCGCGCCCGGGGCGCAGGAACCCAGCGCAGCCTAGGCTGGAAGTCGGGGCGCTGGGCGACTACAGAGCCCGGGAATGGGGCGCGCGGAGAGCGGCCGCCCGAGGACGGCGCTGCGGCG	339
chr5	135038973	135039179	PITX1	TCGGGGTCCGGGGCGAAGAGAGCCAGGGCGCGGACCGACGTCTGCTGCTTTTCTGCGGCATTGCTGCCCGAACGAACGAACGAACGAACGAAGCGGTTTCGTTTAGGAAAAATACCCTCTTGACGCGAAGCCACGGCTGAAGTCCCGGGCCACGCAGAGGGGCCAGCAATTCATGGGTGGTGGGGCCCTCCATCCCTGGACG	340
chr5	141421294	141421524	PCDHGA11	ACGCTGCGGGGGTTCCGGGCCAGGCAGATCCGATATTCGGTGCCA GAAGAGACCGAAAAGGGCTCCTTCGTGGGCAATATCTCCAAGGACCTGGGGCTGGAGCCCCGGGAGCTGGCGAAGCGCGGAGTCCGCATCGTCTCCAGAGGGAAGACACAGCTTTTCGCTGTGAATCCGCGAAGCGGCAGCTTGATCACGGCAGGCAGGATAGACCGGGAGGAGCTCTGTGAGACG	341
chr5	141513254	141513476	PCDHGC5_DIA PH1	GCGGCGTGTCAGTGTGCAGTGGAGTGTGCAGTCTAAGCTTGCGGC TGCTCCAGGCAGAAAGAGGAGACCCCGGCGCGGGCGGGGCGGGTTGGCGCCGGGCAACGCGCTTGGGTAGAGGGGAGAGGACGTTTCGTTAGTTCCCGCCCTTCTGACTAAAATTGCCTACCCGAAGCGCCC CGGAGGGCTTACGGGAGGAGGGTAGACTCTCCTTTGCCCCCGCCGGGCAATGCGAAGGTCCCTCAAGCCTGGACGTTCTGCAGTGGTG GGGGTCTCGCTCTTGCCCTAGCCCTCTCCTACCCTACCCCTATCCGCGCCCCCGGACTGGCAGGCCTCTGGAAGCCCAGGCCGCGGCGCCTACCGCAAAACCTTCTCCGCGCGAGTCCCGTGACCTTGACGCCACGGGCAATCCCCGCACCGGACCCCTTATCTAAATAGGGCAGTAAATCAAGGACCTGTCAAGGGCCCGGTAATTACAGGAACCTCATAAAAAGGACCCGGCCGGCCGCTGTTTATATTAGCGCGGTGTAAATATTCTCGCTGTCTTGGGGAATCGCGTCGCG	342
chr5	154482492	154482837	HAND1	GCGTGAGAGAGAGATACGAGCCTAAACCCTCACATTGGACTACAGCCTCATCTCTGCCCGACCTTTCCTTCTGCCACCTCCTCTGCCCCGGTTCCTTCCAGAACAAATGTTTTACCGTGATCTGTCCAGGGCAAAAGCCATCCACATTCTCAGTGCCTACATCTAAAGCCCATGCTCCTCGCAGTCAAGGCTCTCCAGCAACCG	343
chr5	168880138	168880349	PANK3_SLIT3	GCGGGGGGCGCTAGAACCCGAGGCTGGTAGGAGAGCAAACTCTCA AACGCGCTGAAACCGGCCATCTGGGAGAAATATTAGGGCGCATGTCTCTCCCGAGGGCTTCTTTTTTTTTTTTTTCTTAACACG	344
chr5	173245910	173246041	NKX2-5_STC2	GCGCCGGCCGGGTTGAGCCGGGTTGGTTCCGACCCAAGAGAGCTC GTCCCACGACGAGCAGGTCCCTTGCATCCCGCGGGGCGGCCAGGTGCAATTTTCGCTGGGCCGACGGCGCGGAGATGGGCCAGAGTCC	345
chr5	177984393	177984830	PROP1_B4GALT7	GCGCCGGCCGGGTTGAGCCGGGTTGGTTCCGACCCAAGAGAGCTC GTCCCACGACGAGCAGGTCCCTTGCATCCCGCGGGGCGGCCAGGTGCAATTTTCGCTGGGCCGACGGCGCGGAGATGGGCCAGAGTCC	346

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr5	178576722	178577059	PHYKPL_COL23A1	GGCCATCCAGAAGTGCCTGGAGCGCACAGCAAGGCCCTGCCCTCG GCTCCGTGAAGGTGAGGGGGTAAAGTCGGCCCGGAGTCCCGGG GGTGCAGGAGGGGCCCCGCGGGTTCCAGCAGACCTCGACGGAA CGTTCCAGGCAGGCGAGATCTCGCACAGAATCTGCCCTTTAAAGG CTCGGCTTTGTCTCGTTAAACTTGCCTCTGGCAACGCGACCGCTGC GGCTCCCAGCAAGATTAGAGGGTTTCCGCTCGCAGGGGCGCGCC CGGGGACCGCGCTCCCCGCTGGTCTCGGCG CCGCGCGCCAGGCCCTGCGAAAAGCCCCAACGGGTCCCCGGCGA CCGCCGCGCCGGCTCTCGGTCTGTCTCCGAGGCGCCAGGCCTC CGCTCCAGCGCGGGCTCTCGGGCAGCGCCGCCCTCCCCCTGCG CGCACGGGAGGCCGCTGGGTTCGGCTTTGGACCAGGCGAGCAGC GCGGCGCTGGCCGCTCTGCCGGGTACGCCCGCGGAGACGTCTTC CCCGCTGCGCCCCGGCCCCAGCGCAGCGCCCGGGGAGCGGCCCT CCTCGGGCAGCGCCGCGCCTGTGTCCCTAGCGCGGTACTGCTTC TGCTTAGGACTCCCCCGCG	347
chr5	180467246	180467413	GFPT2_CNOT6	CCGGGGCGGAGTGGGTGTGTTCAAGAGCTTGTCTTGTCTCTTGCCC TGGCCACAGCCGGAAGCCCTGGGCAGGCGCCCGTGATAGCTG GCACGCTCAGCCTTTGGTGGAGAACTGAGGTGAGCTGGAAGGACT AATGGGAGGGAGGAGAGGTGTACTGGGGCCCCG GCGCAGTGGATGTGACGCTGGACCCGCGCTCGGCGCACCCAGCC TGGAGGTGTCGGAGGATGGCAAGAGCGTGTCTTCCCGCGGGGCG CCGCCAGGCCCCGCGCCTGGCCACCCGAGCGGTTCTCGGAGCAG ACGTGCGCGCTGAGCCTGGAGCGGTTCTCCGCCGGCCGCCACTACT GGGAGGTGCACGTGGGCCGCCGACGCGCTGGTTCTTGGGCGCT GCCTGGCCGCGGTGCCGCGCGGGGCTGCGCGCTGAGCCCTG CGGCCGGCTACTGGGTGCTGGGGCTGTGGAACGGCTGCGAGTACT TCGTCTGGCCCCGCACCGCTCGCGCTCACCTGCGCGTGCCCCC GCGGCGCTGGGCGTCTTCTGACTACGAGGCCGGAGAGCTGTC CTTCTTAACGTGTCCGACGGCTCCACATCTTACCTTCCACGACA CCTTCTCGGGCGCGCTCTGTGCGTACTTCAGGCCAGGGGCCACGA CGGCGGCGAATATCCGGATCCCTGACCATCTGCCCG CACTGCGGAGTGCGGGTGCGGAAGCGGAGAGAGAAGCAGCTGTG TAATCCGCTGGATGCGGACCAGGGCGCTCCCCATTCCCGTCGGGA GCCCCCGCA	348
chr5	181059231	181059766	BTNL9_OR2V1	GCGCAGTGGATGTGACGCTGGACCCGCGCTCGGCGCACCCAGCC TGGAGGTGTCGGAGGATGGCAAGAGCGTGTCTTCCCGCGGGGCG CCGCCAGGCCCCGCGCCTGGCCACCCGAGCGGTTCTCGGAGCAG ACGTGCGCGCTGAGCCTGGAGCGGTTCTCCGCCGGCCGCCACTACT GGGAGGTGCACGTGGGCCGCCGACGCGCTGGTTCTTGGGCGCT GCCTGGCCGCGGTGCCGCGCGGGGCTGCGCGCTGAGCCCTG CGGCCGGCTACTGGGTGCTGGGGCTGTGGAACGGCTGCGAGTACT TCGTCTGGCCCCGCACCGCTCGCGCTCACCTGCGCGTGCCCCC GCGGCGCTGGGCGTCTTCTGACTACGAGGCCGGAGAGCTGTC CTTCTTAACGTGTCCGACGGCTCCACATCTTACCTTCCACGACA CCTTCTCGGGCGCGCTCTGTGCGTACTTCAGGCCAGGGGCCACGA CGGCGGCGAATATCCGGATCCCTGACCATCTGCCCG CACTGCGGAGTGCGGGTGCGGAAGCGGAGAGAGAAGCAGCTGTG TAATCCGCTGGATGCGGACCAGGGCGCTCCCCATTCCCGTCGGGA GCCCCCGCA	349
chr5	112737724	112737821	APC	CGGAGGCGGGGAGACCTGCGGGCAGCGCTCACGCGCACATCCCC GGCTTCCCCGGGCTCCGCGCCTTCCCAAGAGCCCCGTGTCTCCGG CGTCCCAGGGATCGCGTGGGCTCCG CCGGCTCGAAGCAAAAGACGACCGCCGAAACGCGACCGTTTACC GCCTGCTTTTCCAAAGCAAAATTTGGAGACAAGTCCACCCGCGGGA AGAACCTGGCTAAGGGTCGGACATGGAAGAGAAGACGCTAAAC AGAAATTGCCTCCCTGCTTTCCACCTGCAGCTTCTAGACGCCGCCCT CGGTGCCACCCCTCGCGGAAGGCG CCGAGGCGCGGACTGGAAGGACAGGTACCAGGCTGCGGGCGCG CGGCTGTGGCCATCTCTTCCGCCCTGAGGCCGACGAACCCGGCTG GAAGCTGAGTGCTTAGCGGCCAAAGCAGCCCGGGCGCCGGGAG GGCGCCAGAGAAGCACAGCGTTAGGGCGGGGAAGAAAGGGTGAA TCTCAGAAATCGAAATCCGCACTGGCGCCACGACCTGGGCGCGCG GCCTGGTCTCGGCAGCTTCTGGCGGCTGCGCTTGTGTGAATG TGTCCCGGGAGGACCGGACACCTCAATCCCCGGCCCCCAACGCG	350
chr5	115816763	115816878	CDO1	CGGAGGCGGGGAGACCTGCGGGCAGCGCTCACGCGCACATCCCC GGCTTCCCCGGGCTCCGCGCCTTCCCAAGAGCCCCGTGTCTCCGG CGTCCCAGGGATCGCGTGGGCTCCG	351
chr6	1378416	1378621	FOX2_FOXQ1	CCGGCTCGAAGCAAAAGACGACCGCCGAAACGCGACCGTTTACC GCCTGCTTTTCCAAAGCAAAATTTGGAGACAAGTCCACCCGCGGGA AGAACCTGGCTAAGGGTCGGACATGGAAGAGAAGACGCTAAAC AGAAATTGCCTCCCTGCTTTCCACCTGCAGCTTCTAGACGCCGCCCT CGGTGCCACCCCTCGCGGAAGGCG	352
chr6	6003674	6004066	NRN1_FARS2	CCGAGGCGCGGACTGGAAGGACAGGTACCAGGCTGCGGGCGCG CGGCTGTGGCCATCTCTTCCGCCCTGAGGCCGACGAACCCGGCTG GAAGCTGAGTGCTTAGCGGCCAAAGCAGCCCGGGCGCCGGGAG GGCGCCAGAGAAGCACAGCGTTAGGGCGGGGAAGAAAGGGTGAA TCTCAGAAATCGAAATCCGCACTGGCGCCACGACCTGGGCGCGCG GCCTGGTCTCGGCAGCTTCTGGCGGCTGCGCTTGTGTGAATG TGTCCCGGGAGGACCGGACACCTCAATCCCCGGCCCCCAACGCG	353

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr6	10417340	10417545	TFAP2A	GGCGCCTGTCCGCGAGCGCCGGGCCAGACGCCGAAGAGGAAGGT GACCGAACCCGTAGCAGCTTCCGAGAGCGTACCCG CCGCCGAGGGCGCCATTGAGGTGCAGATTGGGACCTGCCGGCTCT GGACTGCCGCCCCGGTGTAGGCGCTGATGAAAGGCCCGGGCGA GCGCCAGGGTCGCCTCTGGAGCCAGCCGAGCTGCAATTTATGCCAG CGTCATTACCACGCTAAGTCGCTTCATTGCATGTCAATGCTCCGGCG GGGCCAGAACCCCGGGACAGCAGCG	354
chr6	10421138	10421571	GCNT2_ TFAP2A	ACGGTGGAAATAGGGCGGTGACTAATTTTCAGAGTGGAAGACAC GCACGAAGGGCGCACCTGCAGCTCTCCGGGATTAGGCGGGGGTCTC GCTGTGCTCTCTTAAAGTGAGCGGCGGTTTCAGCCTGCCACCGCT TCGCTCGCCAGCTCGGAGGAACTCTGGCTGGAGGCGACCTCGG GCCAGCCGGACGGCGCGGCCGAGCCTAGGAGGGGCTGCCAGA CGTGTCCAGGGCCAGGGTGGGGCGTAGGGAGCGCCGTCTCCACC CTCAGTACTTTGGGGTGGGGGACCTGAGCGTGCAGAGAGCGGG AGGCAGAGCTGAGAGCGGGTTAAGCGCGAAGCTAAGGCGCCGC ATAGGGTTGGGTGGGAATGGACAGGGTGAGCTGGAAGCGAAGCA CCCCAGCCAGGCCTTAGGAGAGAGGACCGTCG	355
chr6	19691498	19691942	ID4	CCGGGGCCTTGGAGCTTTCGGATCCTGCCCGCCTTTCATCATGTAA ACAAACGCATCAGATTTAAAGCTTTCCTAATTTGTTATGCTAACCT TGGAGCGCAACCTCTCCATTTGCATTTGAAGGAGCTAAATATTAGG CAGGAAAGAAAGTGCTCTTTTGAAGCCTGAGAAAATGTCCCCGC TCGGGGCTGTCCGCCATCTGGGCCGCGGGCTGGGCGCGCGGCTC CCGCCCCAGCTCCTTGGCAGAGGCGCCGGAGGAAGGGGCGCCGC GAAGGGCCGTATCTTGTGGAAAAGAATGCAGAAATGCCCCCT AAGGCTGAATGAGCACCACTTCCACACTCAGGGCGGGGAGGCCG GGGGACGTGGGAGCGGCGGCCAGGAGCGAGGCGTCCCTGGTGA CAGCGCGTCCCGAGGGCTCTCCCTTTTCCAGAGCG	356
chr6	30163583	30163837	TRIM10 _TRIM15	CCGTTTCCCTCTCGGATTCATGTAAGTGTGACTCGATTTAGGGAAA GGGAACCTCGCGTGGGCTGAGGAGACCGGAGTGGACGGGCTGGGG AAGGCACCGTGATGCCCGCAACCCGTCCTGAAGGTGGTCCATG AGCTGCCTGCCTGTACCTCTGTGCGGGGCGCTGGAGGATGCGG TGACCATTCCTGTGGACACACCTTCTGCCGGCTCTGCCTCCCCGCG CTCTCCAGATGGGGGCCAATCCTCG	357
chr6	32189648	32189910	PBX2	ACGGGGTTTGCTGGGTCTGTGTGGGGTCCCGGAGTGGGGGCACTC ACTTGGCCTGGGCCTCGTCCAGGCTCTGGTCGGTGATGGTCATTAT CTGCTGCAGAATGTCCCCGATGTCTTGCTTCCCTCGGCCTCCGGGA CCCCCCGCTACCCCCACCGGGGTCTCCGCCACCGGGAGGCTCGCC AGGGCCCCCAGGCTCCCCACTACCAATCCAGGCCCCCCCGGCC CCGCCTGGAGGGGGCGGCCCCAGTAGCCGTTCTG	358
chr6	36285248	36285484	PNPLA1 _ETV7	GCGCCCCCTGCTTCCGCGCGCCACACGACGCTGCTCTGGGAG CAGGGCCGGCGGCGCGCCGCTCGCAGCGATTGGTTGAACCGGA GGTTGTTGCTAGGCTACCACTGCGCCCTGAGCCTGGGGCCCCGCA GTCCCATCCTCTGTGGCAGATCCATCCCTCACTGCAGACCTAATTCC GGTACCCTGTGAACGGCATCCTCAGCAGCTTAAATTATCAGCCCCA ACTGCCCCG	359
chr6	38715300	38715525	GLO1_D NAH8	CCGTACGCCTCGTTCCGGGCCGCGGAGGCCGAGCAGCTCCCCCG GGGCAGCGCAACCGCTGGGGCCGGCCTCAGTGGGCTGAGTGGTC GGGGCATCGGGGCCAGAGAGCGGCTGGTGAGTACTTGGTCGGA GCGCGCTGTGAGCGCCCGGCCCTGTCCGGGAGGCCCTGATGCAG	360

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr6	41373668	41374013	FOXP4_NCR2	CCGGGTTCCCCGCCCACTTTCCTTCTTTTAGGGGACTGGAATCCAC G GCGCCACTGCGGAAGGCCTGACCTGATCCGGCACGGTGTGGCCAC CGTGGGCCCCACAGAGGGTGAAGGGTAGCTTATGCTGAGTGGGG GTGTCCACCTGGACAGACCAGGCGAGCCTCGCTCTGTGCGGGA GCTAGTTTTCCCTGGATCTTCCGCGGCAGAGAAGCCTGCGTCCGGG ACCAGCAGAGTGAGCCGACCGGCGGATGCAGTTGACCCATTTCG GTCCAACTTCACTTCGAGAAAACGCAGCCCTGCGCGCAGTCCACG CAGGACGCGACAGCGCCACCCTGTTTGTACGGCTGCGCGAATGA CTCGAGAGAGTCGCGGTGGCTGCACGTGCG ACGTCAATAAAAAATTAATTGATGAGTTGGCAGGGCGGGCGGTGCG GGTTCGCGGCGAGGCGCAGGGTGTCTATGGCAAATGTTACGGCTCA GATTAAGCGATTGTTAATTAAGGACGCGTAATTAATACTCGCT ACGCCATATGGGCCCCGTGAAAAGGCACAAAAGGTTTCTCCGCATGT GGGGTCCCCCTTCTTTTCTCTTCCACAAAAGCACCCAGCCCGT GGGTCCCCCTTTGGCCCCAAGGTAGGTGGAATCGTCACTTCCGG CCAGGGAGGGGATGGGGCGGTCTCCGGCGAGTTCCAAGGGCGTC CCTCGTTGCGCACTCGCCCGCCAGGTTCTTGAAGAGCCAGGAGC CTCCGGGGAAGTGGGAGCCCCAGCGGCCCGCAGACTGCCTCAGA GCGGAAGAGGACGCCGCGGCTTTGACCCAGCTTCTCCGACGGC ATCTGCAGGAGCCTCTAGGCCTGACATAGGCTCCGAGGTGCCCTG GCTCCCCCACG GCGCCAACAGCGCCCTCTCCCGTAAGTGGGCCTCCCTCCCGCGTT CTACCTGCAAGGCCGAAGGGAGAAAAACCAATGTTTCTCTTGACG GATGGCCGGGACTCCTTGGCCCTCGCCTGGCTTCCACCCCTCTG GCTTCCCGCACCAGCCGGGCCCGCAGCTCACCTGCCGCGAGCTGG GGCGAAGCCGTAGTCGGCGCTGCCGGGCGCTTTGTGCTTGGCCTC CGCGGCGCCCCGGGCGCGCCCTCAGGGACAGCCTCGGCGCGTG CAGGCCTCCGGGGGGCGCGCGACCCGCCGAGTTACGCGCCGCAT CTCGGGGCCTCCGGGTGCGGCCGAAGCAGTTGGGAGAGCTCAG GCTGCGGCCGGTGCCACCGTGGGGTAGCCCTGGGCCTCGGTGCGG CTCCCCGACGTACAGGCGCTTCTTATGAGCGAGCGCCCCCTCCC GAGAAGCGCTCCAGGCCCCAGCCCCGGCGTAGCGCGCGCCCGCG GGAAAGCGCGAGAAGCCGAGAGCCGGGGGCGCCCGGGGCCAG CGTTCGGGAGCTGCCTCAAGTCTGAGTAGTTGTTCCGGGGAGGGG AGCTCTGGCGGCCAGATACTGGAGGGCCG CCGACACCAGTTGGGAGACTGGGTAATAACACACGCTCCGGGCAC AGGGACCGCGGGCCAACGAACCGCGCGTGCGCCGCGCCAGCCTGC GTCGAGCCGTGCGACACGGCTCCGGGAGCCCGCTAGGCAACGC TCTCCAGGTTGCCAAGCAGGGTGTCAACAAGTGCACACGCGCGGA CGCCACGCGAGGCGACGCGCGTGGCGCCCCCGGGCG TCGATCTCTCATGTTTAGGCAAATCCAGGGTAAGGTGTCTCCCGG AGCTGGGGATGCGGAGCCAGATTTCTGGCTGAAATCATCCTCATCG GAAAAATCCGCAGAGGAAGACATAGAGCAGCGATAGGACGCGTTC CCGGAATCTACAGAGAATGACACAGAAAAAGCATTAAACAGCAAA ATACTCACATATGCTCAATGATTTAAACATCTCCCCACCAACCACC GCCGCCCTCCCTGCCCCAACTGGGTCTGGCATATCCTGCACCATC CTCG GCGACCGGTTTAGAGCTGTGTGGTCCCTAGTGGGTCTCCAAGCTCC GGGGTACCCTAGCGCGGTATTACATCATTAAGAAGCGCAATC	361
chr6	41560710	41561220	MDFI_F OXP4	ACGTCAATAAAAAATTAATTGATGAGTTGGCAGGGCGGGCGGTGCG GGTTCGCGGCGAGGCGCAGGGTGTCTATGGCAAATGTTACGGCTCA GATTAAGCGATTGTTAATTAAGGACGCGTAATTAATACTCGCT ACGCCATATGGGCCCCGTGAAAAGGCACAAAAGGTTTCTCCGCATGT GGGGTCCCCCTTCTTTTCTCTTCCACAAAAGCACCCAGCCCGT GGGTCCCCCTTTGGCCCCAAGGTAGGTGGAATCGTCACTTCCGG CCAGGGAGGGGATGGGGCGGTCTCCGGCGAGTTCCAAGGGCGTC CCTCGTTGCGCACTCGCCCGCCAGGTTCTTGAAGAGCCAGGAGC CTCCGGGGAAGTGGGAGCCCCAGCGGCCCGCAGACTGCCTCAGA GCGGAAGAGGACGCCGCGGCTTTGACCCAGCTTCTCCGACGGC ATCTGCAGGAGCCTCTAGGCCTGACATAGGCTCCGAGGTGCCCTG GCTCCCCCACG GCGCCAACAGCGCCCTCTCCCGTAAGTGGGCCTCCCTCCCGCGTT CTACCTGCAAGGCCGAAGGGAGAAAAACCAATGTTTCTCTTGACG GATGGCCGGGACTCCTTGGCCCTCGCCTGGCTTCCACCCCTCTG GCTTCCCGCACCAGCCGGGCCCGCAGCTCACCTGCCGCGAGCTGG GGCGAAGCCGTAGTCGGCGCTGCCGGGCGCTTTGTGCTTGGCCTC CGCGGCGCCCCGGGCGCGCCCTCAGGGACAGCCTCGGCGCGTG CAGGCCTCCGGGGGGCGCGCGACCCGCCGAGTTACGCGCCGCAT CTCGGGGCCTCCGGGTGCGGCCGAAGCAGTTGGGAGAGCTCAG GCTGCGGCCGGTGCCACCGTGGGGTAGCCCTGGGCCTCGGTGCGG CTCCCCGACGTACAGGCGCTTCTTATGAGCGAGCGCCCCCTCCC GAGAAGCGCTCCAGGCCCCAGCCCCGGCGTAGCGCGCGCCCGCG GGAAAGCGCGAGAAGCCGAGAGCCGGGGGCGCCCGGGGCCAG CGTTCGGGAGCTGCCTCAAGTCTGAGTAGTTGTTCCGGGGAGGGG AGCTCTGGCGGCCAGATACTGGAGGGCCG CCGACACCAGTTGGGAGACTGGGTAATAACACACGCTCCGGGCAC AGGGACCGCGGGCCAACGAACCGCGCGTGCGCCGCGCCAGCCTGC GTCGAGCCGTGCGACACGGCTCCGGGAGCCCGCTAGGCAACGC TCTCCAGGTTGCCAAGCAGGGTGTCAACAAGTGCACACGCGCGGA CGCCACGCGAGGCGACGCGCGTGGCGCCCCCGGGCG TCGATCTCTCATGTTTAGGCAAATCCAGGGTAAGGTGTCTCCCGG AGCTGGGGATGCGGAGCCAGATTTCTGGCTGAAATCATCCTCATCG GAAAAATCCGCAGAGGAAGACATAGAGCAGCGATAGGACGCGTTC CCGGAATCTACAGAGAATGACACAGAAAAAGCATTAAACAGCAAA ATACTCACATATGCTCAATGATTTAAACATCTCCCCACCAACCACC GCCGCCCTCCCTGCCCCAACTGGGTCTGGCATATCCTGCACCATC CTCG GCGACCGGTTTAGAGCTGTGTGGTCCCTAGTGGGTCTCCAAGCTCC GGGGTACCCTAGCGCGGTATTACATCATTAAGAAGCGCAATC	362
chr6	42104294	42104910	GUCA1 A_TAF8	GCGCCAACAGCGCCCTCTCCCGTAAGTGGGCCTCCCTCCCGCGTT CTACCTGCAAGGCCGAAGGGAGAAAAACCAATGTTTCTCTTGACG GATGGCCGGGACTCCTTGGCCCTCGCCTGGCTTCCACCCCTCTG GCTTCCCGCACCAGCCGGGCCCGCAGCTCACCTGCCGCGAGCTGG GGCGAAGCCGTAGTCGGCGCTGCCGGGCGCTTTGTGCTTGGCCTC CGCGGCGCCCCGGGCGCGCCCTCAGGGACAGCCTCGGCGCGTG CAGGCCTCCGGGGGGCGCGCGACCCGCCGAGTTACGCGCCGCAT CTCGGGGCCTCCGGGTGCGGCCGAAGCAGTTGGGAGAGCTCAG GCTGCGGCCGGTGCCACCGTGGGGTAGCCCTGGGCCTCGGTGCGG CTCCCCGACGTACAGGCGCTTCTTATGAGCGAGCGCCCCCTCCC GAGAAGCGCTCCAGGCCCCAGCCCCGGCGTAGCGCGCGCCCGCG GGAAAGCGCGAGAAGCCGAGAGCCGGGGGCGCCCGGGGCCAG CGTTCGGGAGCTGCCTCAAGTCTGAGTAGTTGTTCCGGGGAGGGG AGCTCTGGCGGCCAGATACTGGAGGGCCG CCGACACCAGTTGGGAGACTGGGTAATAACACACGCTCCGGGCAC AGGGACCGCGGGCCAACGAACCGCGCGTGCGCCGCGCCAGCCTGC GTCGAGCCGTGCGACACGGCTCCGGGAGCCCGCTAGGCAACGC TCTCCAGGTTGCCAAGCAGGGTGTCAACAAGTGCACACGCGCGGA CGCCACGCGAGGCGACGCGCGTGGCGCCCCCGGGCG TCGATCTCTCATGTTTAGGCAAATCCAGGGTAAGGTGTCTCCCGG AGCTGGGGATGCGGAGCCAGATTTCTGGCTGAAATCATCCTCATCG GAAAAATCCGCAGAGGAAGACATAGAGCAGCGATAGGACGCGTTC CCGGAATCTACAGAGAATGACACAGAAAAAGCATTAAACAGCAAA ATACTCACATATGCTCAATGATTTAAACATCTCCCCACCAACCACC GCCGCCCTCCCTGCCCCAACTGGGTCTGGCATATCCTGCACCATC CTCG GCGACCGGTTTAGAGCTGTGTGGTCCCTAGTGGGTCTCCAAGCTCC GGGGTACCCTAGCGCGGTATTACATCATTAAGAAGCGCAATC	363
chr6	50850435	50850652	TFAP2B	CCGACACCAGTTGGGAGACTGGGTAATAACACACGCTCCGGGCAC AGGGACCGCGGGCCAACGAACCGCGCGTGCGCCGCGCCAGCCTGC GTCGAGCCGTGCGACACGGCTCCGGGAGCCCGCTAGGCAACGC TCTCCAGGTTGCCAAGCAGGGTGTCAACAAGTGCACACGCGCGGA CGCCACGCGAGGCGACGCGCGTGGCGCCCCCGGGCG TCGATCTCTCATGTTTAGGCAAATCCAGGGTAAGGTGTCTCCCGG AGCTGGGGATGCGGAGCCAGATTTCTGGCTGAAATCATCCTCATCG GAAAAATCCGCAGAGGAAGACATAGAGCAGCGATAGGACGCGTTC CCGGAATCTACAGAGAATGACACAGAAAAAGCATTAAACAGCAAA ATACTCACATATGCTCAATGATTTAAACATCTCCCCACCAACCACC GCCGCCCTCCCTGCCCCAACTGGGTCTGGCATATCCTGCACCATC CTCG GCGACCGGTTTAGAGCTGTGTGGTCCCTAGTGGGTCTCCAAGCTCC GGGGTACCCTAGCGCGGTATTACATCATTAAGAAGCGCAATC	364
chr6	56851460	56851739	DST_KIA A1586	CCGACACCAGTTGGGAGACTGGGTAATAACACACGCTCCGGGCAC AGGGACCGCGGGCCAACGAACCGCGCGTGCGCCGCGCCAGCCTGC GTCGAGCCGTGCGACACGGCTCCGGGAGCCCGCTAGGCAACGC TCTCCAGGTTGCCAAGCAGGGTGTCAACAAGTGCACACGCGCGGA CGCCACGCGAGGCGACGCGCGTGGCGCCCCCGGGCG TCGATCTCTCATGTTTAGGCAAATCCAGGGTAAGGTGTCTCCCGG AGCTGGGGATGCGGAGCCAGATTTCTGGCTGAAATCATCCTCATCG GAAAAATCCGCAGAGGAAGACATAGAGCAGCGATAGGACGCGTTC CCGGAATCTACAGAGAATGACACAGAAAAAGCATTAAACAGCAAA ATACTCACATATGCTCAATGATTTAAACATCTCCCCACCAACCACC GCCGCCCTCCCTGCCCCAACTGGGTCTGGCATATCCTGCACCATC CTCG GCGACCGGTTTAGAGCTGTGTGGTCCCTAGTGGGTCTCCAAGCTCC GGGGTACCCTAGCGCGGTATTACATCATTAAGAAGCGCAATC	365
chr6	84766369	84766612	TBX18	GCGACCGGTTTAGAGCTGTGTGGTCCCTAGTGGGTCTCCAAGCTCC GGGGTACCCTAGCGCGGTATTACATCATTAAGAAGCGCAATC	366

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr6	84767295	8476756 6	TBX18	CCATTTCTGAAGCTTAGCCGAAGGCAGGCGCCGGCAGGGAGAGCT AAGAGGCCCGCCTAGAGAGTTTGGGCCGGGAGTGGGAGTGGGACA AGGCGGGAGCTAACTTAGCTGGAGTAGACGCCAGAAGAAGTCCG TTCAGCTGAGGTGCCCG TCGGCTCCTGGAGAAGGGGCGTCAATCTCTCTTGGGCATGGGAG GGAAAGACATTCGAGTTGGCTGGGCGGAGTGGCAGCCTTGAGA GTGACGAGTGACAGCAAAGCCTCGTCCTAGCAAGGCCTTTACCAA CAGCGCGCATGCCCTTTCGAGGAGAGCGCCAGGCCCTCGCACTTT GCAAGTCAAGAGAGCAAAGAAAGCGGGGACAGGGCGCGTAATCG CAATGTCCGGTCGCGCGTGTGCACGTGTCTGTGTTGCATGTGTGC G	367
chr6	105986000	1059864 22	PREP_P RDM1	CCGGCCAGGAGTGAACGCTGTCAATTCATCTTGCCCTTAAGGGAGG GAAACCTCTACCGAATATAGTGCAGCCTCAATGGTGGGTCTGT CCTGGGGCCTGGGCGAGGGCGCCGGGTCTCCGGACTCAGGCAAGCA CCTTCTCTAACCAGCAAGCGAAGCGAGGAGGAGCGACAGAGCGC TTCCTCTCCCGCCGAGCTGAGTCCTCTGGGCCGACGTCTTCCTGG ACGAGCTCTGAGGCCGAAGATGCGTTGCGTGACTATGCTGCTGCC GGACGCGGGGTCTCTAGTCCGGAGGCACGGAAGGACCTGCCTGCC TGAAGTCTAGTCTGCAAGTCTCGGGCACACGCGCGGCTTCTGCCAC CCGCGTAAATGCCCTGGGGAAAGGCGCCCTTTCTTTATGATGTTTT TTAAGAGACG	368
chr6	137493335	1374937 24	OLIG3	CCGGGCCCCGCCGCTGCTCACTTGAGCAAGTCCTTGGACTCGGCCG ACAGCCGGGCCATGTTGGCTGTGGAGAGAGCGGACAGGTGCGGC GGCGCGGCATCTGGCAGATGGTGACGGGCGAGGGCAGACACAGC CCAGTGCTGGAAGCCGCTGCCAGCTGCAGCGCGGGCGCGTGGA GGGCGCCTTGAGTAGCGAGTGGGGAGGCCGGATGGTGCCGATGG CGGGAAGTGAGGCGGCGGACAGCGGTGACGAGGCGTTGCCAGAT GAGAGCGCGCCGCCAAGATGGGGTGACCGGGTGACCGGAGTT GGCCGCGTGCGCGGGGTGGCCGGCCGAGTGGCCACGGTCCCGC AGTGAAAGGCCGAGTGGTGGCCCCCATAGATCTCG ACGGGAAATGAAACCAAGTAACGTGGTGAGAGCACAACTGATGAC AATCACAGAGAGCACAGTCG	369
chr6	142913617	1429136 81	HIVEP2 _GPR12 6	ACGCCATCTCGTGCTCACCATTGTGGCATTCTTCATCGTCAACAT TCCAGATTGATAAAAAGTAGTAAATTAAGACTGGCCAGCAAAGT CCCTGATCAGCCGGATCACCAGCAGCAAGTTGCACGTTTGACAG CCGAGAGGAAATGACTTCATGGGCTCACTGTTGAGCTGCTCCCTT TGCATCTCGGGGAAGGTGTGTTTACCCGACGAGGTCCGGTGA AGGAAGCACGTGTGTGTGTGTGGAAGGGTGGCGCTGACCTCCAG ACAGGACATTACCCTTCTCCTCTTCTGACCACTGCTGTTCCACA GCAGTCACG	370
chr6	142913794	1429139 30	HIVEP2 _GPR12 6	ACGCCATCTCGTGCTCACCATTGTGGCATTCTTCATCGTCAACAT TCCAGATTGATAAAAAGTAGTAAATTAAGACTGGCCAGCAAAGT CCCTGATCAGCCGGATCACCAGCAGCAAGTTGCACGTTTGACAG CCGAGAGGAAATGACTTCATGGGCTCACTGTTGAGCTGCTCCCTT TGCATCTCGGGGAAGGTGTGTTTACCCGACGAGGTCCGGTGA AGGAAGCACGTGTGTGTGTGTGGAAGGGTGGCGCTGACCTCCAG ACAGGACATTACCCTTCTCCTCTTCTGACCACTGCTGTTCCACA GCAGTCACG	371
chr6	150683400	1506835 91	MTHFD 1L_PLEK HG1	CCGGAGGGAAATGACTTCATGGGCTCACTGTTGAGCTGCTCCCTT TGCATCTCGGGGAAGGTGTGTTTACCCGACGAGGTCCGGTGA AGGAAGCACGTGTGTGTGTGTGGAAGGGTGGCGCTGACCTCCAG ACAGGACATTACCCTTCTCCTCTTCTGACCACTGCTGTTCCACA GCAGTCACG	372
chr6	163396939	1633972 10	PARK2_ QKI	CCGGCGTGAAGAGATATTTAGAGGGGAGTGGTCTGGGCTAATC TGCATGTGAATCAGGGGGGTGGACAAAAGGATGAAAAGGTGGTG GAAACTCGAACACAAACCCTGCGGTCTCCAGGGGGTCATTATCTT GCCCCGGTCGACATCCTCGCGCCTGGCTTCTTCTGCGCATGAGC GAACAGAGCCTTTTCCAAAGACAGTTGGCAAAGGGTGCGTGTGC TTTGTCTGTGCGGCACTTTTTTAAGAAACAAAATTTCTTTACCCG GCGGGGTGGGGCAAAGGTGACCCAGCACGCAGCACGGTGCCAG GCATGGAAGTACACGTGATGCCCCGTCTGTTTAAACAGTGAACAAA GGCACCAGAGGCTTCTTCCCTTGAACACCAATCTTCAACCTAGAT	373
chr6	170178999	1701792 30	DLL1_C 6orf70	CCGGCGTGAAGAGATATTTAGAGGGGAGTGGTCTGGGCTAATC TGCATGTGAATCAGGGGGGTGGACAAAAGGATGAAAAGGTGGTG GAAACTCGAACACAAACCCTGCGGTCTCCAGGGGGTCATTATCTT GCCCCGGTCGACATCCTCGCGCCTGGCTTCTTCTGCGCATGAGC GAACAGAGCCTTTTCCAAAGACAGTTGGCAAAGGGTGCGTGTGC TTTGTCTGTGCGGCACTTTTTTAAGAAACAAAATTTCTTTACCCG GCGGGGTGGGGCAAAGGTGACCCAGCACGCAGCACGGTGCCAG GCATGGAAGTACACGTGATGCCCCGTCTGTTTAAACAGTGAACAAA GGCACCAGAGGCTTCTTCCCTTGAACACCAATCTTCAACCTAGAT	374

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr6	3247396	3247517	chr6:3	TAGCAGCCGAGCGAGAGGTGGCGTCTGAACAGCCTAGATTAGAGG CCGAGCGAGAGGGCGGCTCTGAACAGCACCTGGGATCAGGCAG CGCACG GTGTCGTATTTATGTGTGTCTGCCTCCCGTTCCAGCGGAGGGC GAGGCGGGGGTCATCGTTCTGAAGGGCATCTTTGTGTCTCCAGC ACTCAGGACAGTGCCTGGCACACAGATGCT	375
chr7	351427	351579	PDGFA_ FAM20 C	TCGGGCTGGTGGGGGCTGCAGAGGAAGCCGGCGGGGCCAAAGCG TTCTGTGATTGAAGGCGCTGACATCGGCTTCTGTTGTGACACGG CGCTCAGCTTTGCGAGATGGAACCATAGGGGACATTGCAAAAAGG GCACACAGAATCTCTCCG	376
chr7	601453	601766	PDGFA_ PRKAR1 B	CCGGGCTACCCAACATGCCACTTTTTATTCCAGATTCTTACTGAG CATCCTTTGATTCCCTTAAATGTGGCCTTACCCACACGGGCCCTGC GGATTTACCCTGCATGCGAAGGGCTCCACATCACAGGAGGGCC CCTGCAGGCAGTCTCTGCGCCCGGCCCGCCCGGCCCGCCGGGC ACTCCTGACGCCCCACCTGCCCCTGGCTGAAAAATCTGAAGTTGA TGGAGGTGCTTGGTGTTCGTGCACAGCCGCTGGGACTCACGGGA CAGCCCCATAAGTCACAGCCGTTCCCGAGGGGGCCCCG	377
chr7	1223944	1224204	ZFAND2 A_UNCX	CCGGGATTGTGGGTTTCTGCCCCAAGGGTTTCGCGGCGTGGGCAT GAGCGCTGGCATCTGCGCGCCCTGAGGTTGCGCGCTGCGTGCC TTCTCCGGGAGGTGGGGGGAATCCGAAGAGTCCACCCCAAGTT CGGTTCCCGGCTTCTGCTTTTGTTTACAGGCTCCGAGGAGGAC CTGCCTCTCTCTCCCGCAGCCCTGGGCCCCCACTCGACAGTTTCA CATCCAGGGAGGGACAAAGGGGACGCGGCCG	378
chr7	4819962	4820148	PAPOLB _AP5Z1	ACGGGGACCACATGGGACCCAGCTGCCTGCGGCCACCAAAACCCAG GCAGCCACGAAGCCACGTGGAAGTCAGCCGGGGACTCTCCAGGA ACACAGAGCCGAAAAATCACAGGTCCCTGAGCTGACTTCTCTGTG GGGGCCGGAACAAAGGGGCTCCTAAGCTGGCCCCGTCCTCTGT CACACG	379
chr7	27156798	2715700 6	HOXA7	CCGCCCGCGCCCGGCGGGCTGGCGCGTCCCGCGGAAAAAGACCT GGAGGCTCCGCGGGAGCGCCAGCTGGCGGCCAACCTCCGCACTG GGTCTGCGGACGCCAGGCGGCCCGGCCACGCAGCACCCCCCA CCCCGCCCCCGCGGACTCCTGCTAGTGAGCCCTGGACCAAGCTT GGGATCCTCCCATCCCTCTCTGTCCG	380
chr7	27166405	2716661 9	HOXA9	GCGGCCAGCGCCACCAGGGCGAAGTTTTGAGGGCTGTTGGT TGTGCGGCGCGCTCGGTCCCCGGCCCTCGACCCACGCACACGCG GCCAGCCCGCCTTCTCATCAGCTGGCAATCAGGATTCCAGGCG CAGGCGGCTGGCGACCCAGCCCTGTGCTCCAGCCTCAGAGGCTCTA ACCATGAGCGCTGCAAGCCTGGTTGCGTCCG	381
chr7	27220510	2722084 8	EVX1_H OXA13	CCGCCGCCAGACTGACCTGGTGTGGCGGTGCGGCGGGCCGGGC CAGGCCGCGACCGCGAGAAACACAGCCCCACGGAGGAGGCCGG GCCGCGGGGCTGGCGGGGACCTGCAGGCCGGGCCGAGGTGCGG TGAGGCCTCTCCGACCTGGCCGCTCCTCAGAGTTCGCTCGGGG CTTCGTGTTTGAGAGCAGCCTCCCGCTGCCCGCTTGCCCGGGG ATGTGGGTGGACCCGCCCGCGCGGCCGCGGCCAGTGCAAAACCG TGATCCACCCTCTCCGCTCGGTGGGAGGAACCCGGGGCTTTCGCG CCCTAACAGCAGCGTGACCCTCG	382
chr7	27243853	2724405 7	EVX1_H IBADH	GCGCGGAAGCCAGGAGTCCATAAAGGACCGTAAATTCGCGCCCA CTTGGGCGAGCCGGGTGCTGCAGCCCTCCGACCAAGTTGACGTCG GTCAGAGGTCCAAATTACCTTGCTACTTCCCGGGCTTCGCGGCGCC	383

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr7	29563592	2956379 2	PRR15	AGGTCGGAAATGGTCCCAATGGTCTAATTGCCTTTGGTCTCCGGTT GCATTTGAAAAGGCAGAGATCG TCGCGATGGGGCCAAGGGACAGCTGCTGCGGCAACTTTTACCCAG CGGAGCCCACCTACAGCCTCAGCCTCCGGGTCTCAGGTCTCCGCCG TTTCTTCTCAAGGAGTCGGTCGGGGGAGCGGCACTGCACAGCTTTT CTCCAATCAGACACCTCAAGGCTGGCGCCTGATCCAATCTCCTCCCC TGGAGGGTGGGAACGCG	384
chr7	29566025	2956667 4	WIPF3_ PRR15	GCGCAGTGGCGTCTAATGCTAATGTGGGCTACGTAGCTACGGGAT TGGGTCGCTCCGACCTGGCCGATCCGGTGCCAGACAGCATAAGG GAGGAAAGGGGACTGGGGGGGGCACGTGACTTCAACCAACCCAG TAACCAAGTTTTGTTTTCTTCCCCAGCACAGGCCGCTGCCTCAGCAT CCACCCCGCAGCCACGTGTGGCAAGCCGGGGAAGGGGTGGAGT GAACGGCCGGAGACCACGTGGAGAAAGGGGCCGCTTTGGCCCTTC CATCTGGGTGCCGGGAGCCCCCTAGGCCCTCCGGCCATGGCCGACA GCGGCGATGCTGGCAGCTCCGGCCCCCTGGTGAAATCGCTACCA ACAGCAGAAAGAAAAGCAAGGAAGCCGAGTGGGGGTGCCGCCT CCCGCCAGCCCGCTCCCGGGGAGCCACGCCACCTGCGCCGCCCA GCCCCGACTGGACCAGCAGCTCCCGGGAGAACCAGCACCCCAATC TCCTCGGGGGCGCCGGCAGCCCCCAACCAGACAAGTTATACG GGGACAAATCCGGCAGCAGCCGCCCAATTTGAAGATCTCGCGCT CCGGCCGCTTTAAGGAGAAGAGGAAAGTGCGGCCACGCTGCTCC CGGAGGCGGGCAGGTCCCCG	385
chr7	35254009	3525437 3	TBX20	CCGGGATGTCCAGGCTGAGGTGGCCACCAGCCGAGCGCGGCTGC TAGGACGCTGGCGTGGGGAGCGCGGCGCGAACTACGGACAGTG AGCCCTGGCGCTCGCTGCCCTGCGCCTTAATTTGCTGGCGGCGGCG ATCCCGGAGGCCCCGAGCCAGTCAGCGCCGTCTCACGTACCGCTT CCTGATTCCGCCGCCGGGGCGGGGCCGCGGGCCGGGCGCGGAG GGCGCGCCAGGGTGCGGGCGCCGCGTGCCCTGTGCCCCGCGCTG TTCGGTACCCAGCACAGGTTACAGGAAAAGGGTGCCACCACTAG GCTGACGCAGCAGCCATGGACATCCCCACCTGGTCTCACAGCCCCG GGCG	386
chr7	35257779	3525798 4	TBX20	CCGTGGGGAGCGCGCGCGCGCCTTGGAATTCACCGCAGTCGG GAGGGCGGGTCTGAGCCTTGCCCTCCAGGATCCTTCCGACGAACAC CCCGCGGGTTTTAGTTTATCGAGCCAAAGTGGTCCCGGAGAAGCG CTCCCTCGCAGCCAAGCTGCAAGAAGTGCCGGGAACCTACAGGC CTCGGGCCGACCCAGGAAGCCTCCG	387
chr7	55181179	5518139 5	LANCL2 _EGFR	ACGTATTTTGAAACTCAAGATCGCATTTCATGCGTCTTCACTGGAAG GGGTCCATGTGCCCCCTCCTTCTGGCCACCATGCGAAGCCACACTGA CGTGCCCTCTCCCTCCCTCCAGGAAGCCTACGTGATGGCCAGCGTGG ACAACCCCCACGTGTGCCGCCTGCTGGGCATCTGCCTCACCTCCACC GTGCAGCTCATCAGCAGCTCATGCCCTTCG	388
chr7	67551150	6755126 8	TYW1	ACGGCTGGCTTTGTTACAGCCGACCCGTGGCTTCCCGTGGCTGCA CTTGGA AAAAGCACTCGACGCTGCCCGGGCAGCTTTCATCTCAAG TGGGAACGCGGCTGCCGGCTGTCTCCG	389
chr7	70646560	7064677 2	WBSCR 17	CCGCTGGAGGGGAGCCACCGCCTCTGGCCCCCAAGGGGATTCT CTTTTTCTTTATGCCCAAGAAGCACTGCCCTGGAAGCATCCCCGGAAT GACTGAATCATTGCCATTTGTGCGGCATCGAACAGACTGTGCCGCT GACAGCTGTAGGCAAGATTGACTCCGATGCAGTGCCAGGAGATCT AGGCCATGCAAGGCGGCTGCTCAAGGCCCG	390

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr7	72337037	72337298	CALN1	CCGCGCGCTCCTCTACCCCTCCCGCTCCCGCTGGCCGCGCGGGTTC AGCCCCATGTGCGCGGCTGCCTCGCTGCGCCCGGAGCCCAAGTGGC CGAGGCCCCGCTGGAGTTGCGCGCCCTAGAACTCCATGCAGCTCC GGCCTCTCCCCAGCTCCTCCCCAGCGGATCCCCAGGGCCTTGCC GCCGACAGCACCACTCCTCGCTCTGCCGCGCGCCGCTTCAGGA GCCGGGCTTCTGGGCTCGCCTTGCCCGCTGCG	391
chr7	97732129	97732323	TAC1	GCGGAGCGACCAGCGTGCCTCGGAGGAACCAGAGAACTCAGC ACCCCGCGGACTGTCCGTGCGAGTAAGTGCCCGCGCGGTGCTGG CCGCGGTGCCCCGGGTACCCCGCCCGCATCTGTCCGAGGTGGCC GCGCTGGGGGCGCCGCTGCGGCGAGGGACAGTGGGGAGACTGGC TTCCCAAACGCCAACG	392
chr7	97732435	97732503	TAC1	ACGCGATTCTCTCGCCTAACCGGTACAGGTGAGACTTCAGTCTTAT GTTTTTGATCTTGTTTCATCCG	393
chr7	122310063	122310297	FEZF1_ RNF133	TCGATAATAGAAATTAACAACACAGAGCAAAGAACGAGCTTAG TGAAATGGAGAAGCAGTAGAGGTAATAAAATCCTCGAGCTAGA AAGCTCTAAGAACCGCTTATAAATTCAGTTACCTCTGAATCCGGC CGATGGCCACTCCGGCCCGGAGTGCCCGCGCCGACCCGCTGGC CTTGCCCGTCTCAGCCTTCATTATCGCCACGGCCTTGCGCCCCCTG CCCCCG	394
chr7	122310567	122310746	FEZF1_ RNF133	GCGGCTGGGAGTTGGGGCGCACTTCAGTGACCGGGCGCCGCTGC CGGGCTGGGGCTCCCAAGCGTCCGGCTCCCGGGGTGGTCGACGCG GCGCTGCCTTCGATCAGGTCCCGCCGACCTCGGGCCTCTGGACCAC CACCGCCCCAGCTGGTCTGGCAACCCATCCCGGGCGCAATCGCG CCGATTGGCCCGCTAGCCCAGGGCGTGACACCTCATAGCCCT	395
chr7	128351787	128351958	RBM28 _PRRT4	GCAGGGCTCCGCTCAGCAGCCGAAGGTGCCCGCCACCGGGGCCG TGCGCGCCGCGCGCCGCCAGGACTCCCGAGGGGCGAAGGGGCTG CGCCCCCTGCGGCAGGGGTGTGGCGCCCTTGAAGCCCG	396
chr7	128352148	128352385	RBM28 _PRRT4	GCAGCAGCAGGCCGCTGGGGCACCGCGCGGGCGCGCCGGGCC AGCAGCAGGCAGGCCAGCCCCAGGCCGGCAGCCAAGCAGGGCAG CGGAAGGTCCTGCAGCAGCAGCCAGGCGAGCGCGGGCAGTGCAT CCCTGTGCCATAGGCGTCGTAGAAGAGCGGGAAGGCCCGCTGG TCCCGGCCGACAGCAGCAGCAGGTCCAGCAGCGCCAGGCAGGGG GCGCCGGGCGGGCACCG	397
chr7	150958191	150958442	KCNH2_ AOC1	CCGAGGCGTCGGGGTTGAGGCTGTGCGCCCGGGGCGATGGGAGC TGGCCGGGCGCGCTGCGGGGCGGAGAGCCGGGACCCACCAGCGC ACGCCGCTCCTCCGCGGGCCCGAGCCCTGCCACGTGTTGTCCATG GCTGTCACTTCGTCCAGGGCCAGCGACTCGTGTGGGTGCCGCG GGCGTCAGGTCCACGTCCACCACCGCCCCCGGGGCGCCCGCG CCGCCCCGCGCCGCCGACCGCACCGACG	398
chr7	154617885	154618011	PAXIP1_ DPP6	GCGTCGTGCTTTTTTCATGGGAAAGAAAACCTTGACCCAGAGTGGC TTCATTAAAGAAGGGAGAGGGACTTCATAAGTGACCAGTCGAGAG CTAGGCCATAGGGGCTGCAGACCCGGGACTCAAACG	399
chr7	156608149	156608398	SHH_C7 orf13	CCGAGGGGTAGAAAGCGGATGCCTCCTAAACCTGCGTGCGATCTT CTGAGGATAGGAGGACACAGGCCAGCCCTGCAGCCCGGTGGG CTCCGCGGCGCCCCACCCGCTTCCCTCCAGGCCGTCTCTCCACT GCGGCCGACGCTCCAGCCAGGCTCCTTCTGGCCCTGAACACAG GTGACATTCTGCCACACGTCCACCCGAGGAGACTTTTCTCAAG CCCCTGCCTGGGACCCATCCG	400
chr7	157005707	157006028	MNX1_ NOM1	CCGGGCGCTGGCGGCCCCAGCAGCTCCTCGGCTCCCGGCTCCTCCG CGCCGCCCTTCCCCGCGCCCCCGCGCCGCCCTTCTGTTTCTCCGCT	401

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr7	157018159	157018379	UBE3C_MNX1	TCCTGCGCCGCTGCTCTTTGGCCTTTTGTGCTGCGTTTCCATTTTCATC CGCCGGTTCTGGAACCAAATCTTACCTGCGGGCACAAGCGGGCG TGAGAAACCGGCCACCGCCACCCAGGGCTTCTGTCCCCGGAGTC CCCCGGCCGCTGCGCCTGGGCCCCATTGGGTGCGCCCTGGAATG GCCTCAGGGTGAGACGACTTAGAAGCAGAAATGGGGAGGGGGCTC G CCGTGCGCTTCAGGCACAGGTAAGCGCAGCCCGCGCACCGCTTGG GACGCACCTGGCCACCTGCGCTGCCACCCAAGCTTGGGGTATGCG GGTGCCCGAGCAGAAACCCGAACTCGCACCGGGCTCCGAGGTTGG AGCAACTCCTAACACTGGGCTCGGAGCTAGGGGCTTGTGGAGGG GCGCTTGCCGCGCCGGCCCTCGGGGCTCACAGCCGGGCACG	402
chr7	157018557	157018747	UBE3C_MNX1	GCGGCCAGCCAGGCGCGGGGCCAAGCCTATTGCCAAAAACATAT TACCCTGCGACATTCTGTAATGAGATAATGATCCATAAACCCGGA TGATAGATGTGGCGTGCCTGCGATGTCTTCTCTAAATGAGCTGCTC GCATCGACTGCTAATAATGGTGAGTTATGGAAGCGATTTACGCGC AAACTGCG	403
chr7	157685586	157685799	DNAJB6_PTNRN2	GCGGCAGGAGGGACCCGGGGCCAGCCGAGGCTGTTCCAGGGAG GCAGACACCTGCTGTGCGCGGGACCTCGACACGCTCCGCACGCGC GGGAGCGGAACCGGGCTGCTTTGGAGGCTCCCTTGGCGCGCTT GGATTTACTCAAAGGTCAAAGAAAAATGTCAAGGAGAGCGATTGC CTGGAGAGCTCCTGGCTCTCTCCCGGGTCCCCG	404
chr7	97731982	97732134	TAC1	CGGCTAATTAATATTGAGCAGAAAGTCGCGTGGGGAGAATGTCA CGTGGGTCTGGAGGCTCAAGGAGGCTGGGATAAATACCGCAAGG CACTGAGCAGGCGAAAGAGCGCGCTCGGACCTCCTCCCGCGCGC AGCTACCGAGAGTGCGGAG	405
chr7	27096209	27096279	HOXA1	GCTGCTGCGGCGACTGCAAAGGCCGATTTGGAGTGCTGGAGCGAA GAAGAGCAAAAGCTGCGTTCTGCGCG	406
chr7	50304271	50304365	IKZF1	GACGACGCACCTCTCCGTGTCCCGCTCTGCGCCCTTCTGCGCGCCC CGCTCCCTGTACCGGAGCAGCGATCCGGGAGGCGGCCGAGAGGT GCGC	407
chr8	736980	737150	DLGAP2_TDRP	CCGATCGATTTTCCCTTTTCTCGGCTGTGCTGCCATACGCCACTC ACAGCAAACCCAGGCGGCGGGCCCCCTCCGAGGGCGCTCCTTGCG TCCGGACCCAGGTTCTCGGGGCGCCCCCGGTGGGTCCCCGCGAA GCCGCCGCCGCACACCTTCTCAGCGTAGCCCG	408
chr8	737364	737587	DLGAP2_TDRP	CCGGGGGCGACGGGTGTACCGGGTCCCCCGCTAATTTTCGGGCG CGGTGAGCGTGCCTGCGCGCGCCGCGGTGGAGGCCGCTGCTTTC CCGCCGGGAGCCCGGCACAGTCCCCGGGTGACCCGCGCGCCCCGC GCAACAGTTGGAGCCGGGCTGCCCGCGCTCCCCAAGCCGGGCC CTTCCCCAGATGCAGCCGCGCGCCGGCCGCCCCCACTGCGCCG	409
chr8	3691972	3692262	NONE	ACGGTCTTTGTCCAGCTCATGAGACAGGATGCTGGGCATCTGGTCT CATCATCAGCAGAGCCGCTCACTCAGCGATCTGCCTGCTCCGGGTGA GATCTCAGTCAACTTCGAATCATCTCTGACTCATCTGGAGAGGCC TGGGGAAGCCACTGCATCCGGGTCTCTATCCAGCCGCTAATGAC CATGGCCCTACAACATTGTTTCTCTGACTTTACGTTGTTATGCCCA TACACCTCAGTGCTCTGGGGGCAAAATCCTTACAGCCCCCTTAGTC GCTATCCTGCG	410
chr8	10731372	10731697	SOX7	GCGCTGCGACCTGCGAACTCCCCAGTTTCCCTCATCTGCACACCCT GGTGTAGACCGACCGTGCAGCGCCGGGCCACGTGCAGCCTGGGGA CTGCAGGCTGGGAGCTCACGGCCATCTCTGGCCGCGCTACCGCA GCTCCCTGTACCCGGCCCCCTGTGAGGAGCTCTGTTCCCGCGCTC	411

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr8	20518066	2051829 2	LZTS1	TCATATAAGCGCCGGCACACAGTAGGCGCTCAAGGCCTGCAGAAT GAGTGAGCAAATATAGCTCAGACACCTACTGAATGAAAGTCGGCA GGTTTGACTAGATCCTGGAATTTAAATTTACTGAGCGCCACCCAT GTGCG GCGGCACTTGCGGAGAGCTCGGAACACTCCGCCGAGAATGACTTT TGGAGCCATTTGGCAGAGATTAGGGAAAAGAATAAGTGGACACGC TCCAGTTATGAAGAAAAGACATATGGGGATTTAGATTATGAACAGA CGGAAGAGGAAGAATGAGGAATCATTCTTTGGAGATAAAGACTCT CCGGAACAGAAGCGATGCTGAAATGCGTAAGTCGACAGTAATGAC G	412
chr8	23018512	2301876 2	RHOBTB 2_TNFR SF10B	TCGACTCCAATGCCTTTCAGGAAAGGACTCGGCCTTCTCTGACTG CGGAGGCCCTGACCCTGCCAGCTGGCTCCGAGGGCAACACAGGGG CCTGGCCTCTAGAGGGCTGGTGATTGAGGGGCCCGGGCTGGCGGC AAAGAGGGGTTTGGTCTCGGGGCTTAAATGGCACCCAGACTCTTGT TTTGGCCATCTGGAGACTGCAGGCTCCCTTCTTACCCTCAGAGAGT GCTTATGGTGGGTGTTTTTGCG	413
chr8	23706484	2370681 3	NKX2-6	CCGGGCTCTTCCGCACCCGCGGATGTGGCGAAGCCGCGGGGCAGC TCCGCTCGCGCTCCAGTCGCAGGATGTCCTTGACCGAGAAGGGGG TGGAGGTGACGGGGCTCAGCAGCATCCCGAAGGCGGATGGGGCG GGCCCGAGGAGGTCCGGGTGAGGAGCGGCACCCTGAACCTCCCGT CTTGTCGCTGCAGGCCCCGCAGACAGACCCAAGCTCTGGGACAGA CGCCACGCGTCCCAGACAGCGCCTTCTCTGGGCCATGCTGGTAGG CCCGGGTCCAGGGCCGGGTGACGAGACCGTAGCCCCCATTGGTT CTCGCAGAAACCACG	414
chr8	38900207	3890059 4	PLEKHA 2	TCGGATGTTGTCCACCTGACTTGATGCATATTCAAATGTCTCTCTCC CGACGTGGGAGGCCGGAGTCAGAACCTGACAGACCTGCCGTTTAC TAAGTGGGTACCCAGGGCAAATTACTTCACAAGTCTGAGTCTCGGT TTCTCACCCTGAACCGGACTGGTACCCATAGGTTGCGGCGTGGAT CAAATGAGATAGCGCAGGGGCGGGACCCGCGCACAGCAGCTCTCT TAGTTCTCTTGGCGAGGTTTACGTAGTAACACATGCTTGCTGTTT CCCATTTTTCCAGAGCACCCCTCATGCTCTGGGGCAGGAAGGGA GTCTTCGCATCACACCGAAAAAGTCCCAACGGGCACGGTGTAGGC GCCTGTGGTCCCAGCTACTCG	415
chr8	54458148	5445830 3	SOX17	CCGGATGCGGGATACGCCAGTGACGACCAGAGCCAGACCCAGAGC GCGCTGCCCGCGGTGATGCGCGGGCTGGGCCCTGCCCTGGGCC GAGTCGCTGAGCCCCATCGGGGACATGAAGGTGAAGGGCGAGGC GCCGGCGAACAGCGGAGCACCG	416
chr8	54466959	5446739 8	RP1_SO X17	GCGGGAGCTTAGATTCTCTGTGGGCCACATGGTCTCAGAAGAGGC CCCGCGCCCGGGGGCGCCCGCAGTGTCGCTGGACCGCGCGCAG CGCTGGCCACGCCGTGGGCTGGGACTGGCCCGGAACGCGGGTGG CGGTTGCGCTCGGAGACCCGCGCAGCCGTCGGAGCATCTCCGTG CCTCGCTACCACTTCTTTTCTCCGCTCCGGCGAGGGTTTCGG CGCGCGGGGCAGGCCTGGAGCGCCGTGAGCAGGCCGGATGCGGG ATACGCCAGTGACTACCAGAGCCAGACCCGAGCGCGCTGCCGGC GGTGACGGCTAGGCTGGGCCCTGTCCTTGGGCCGAGTTGCCGAG CTCCCTCGGGGACTTGAAGGTGAAGGGCGAGGCGCCGGCCGGGG CCGCGGGCCGAGCCAAGGGCGAGTCTCGCATCCGGCG	417
chr8	54470079	5447031 7	RP1_SO X17	GCGAGGTGGGCGCAGGAGGAGGAGCTGCCTTCTCCGGGAGGCG GCGCAGCGCGGGGATCTTGGGGACCGAGCCAGAGACCAGGACC GTCCCCAACCGTTCGCGGCCGCGTAGCCCTGGGCGGCTGGGCT	418

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr8	55989071	55989233	RPS20_LYN	GCCCTTCCCGCGCAGGGCTTTCCCTCCTGCCGGTCGCTGCCCCGCA CATGGCTCTGGTCTGACTCCCGCTCCACTGCCACCACTGCCACGCGC CTGCGTCCCCG CCGGGTATGTGTGCTGAGCAAACAGTCCACAGGGCACATGCCCAG CAAGGCTGGTGTGCTGAGAGCCTGCGCCTCGGGTGGGAGAGA GCTTGCTGGAAGCCGGTTTACCCTGTGGGATGCTGGGGTTGACA GACTTCTACTGGGCCTTTGAGAAAAGCG	419
chr8	70034655	70034888	SLC05A1_PRDM14	GCGGCCCGAGTTGCAGGAAGGGCGCCGGCTCACTGCCCCAA GAGCTCGAACGCGCGCGCCGAGGAGTGCCGGCTGCGGGGTGCG GGTTGAGACTGGCGGGACCTCGGCCTGCGCGGGGTGCGGAAG GTGGATGCTACGGGCAAAGGGGCGGGGCTTGCGGTCCCAGATCC AGAGGCGGGTTGGGGACGTGAGCCGGCGTCCATGTGTTCTGCACC CCTTCTCGCCCCG	420
chr8	70072034	70072325	PRDM14	CCGGCCATTGAGGGAGAGAAAAGGAACGCTTAGTTCCATTACATT ACAGAAAGAAGCGCCGAGGGTGGGGGAAACGCAGTCTTGCCGGG TGAGCCGGGACAGGTTCTCGCTGCCCCCGCGCGCTGCTTCTC TTAGCTGAATGGGGAGCGACCCGCCCCGGGCGCGGCTTCGGGGC TGAAGACTGAGGTGCAGCCTACCCCCGGCCTGGCAGCGGCTTGG AAGAGAGAGGGAAGGAGGAACATCTACCCGGCTAAGAGACGCC GCCAGAGTCCCTAAAGCTGGCG	421
chr8	91594959	91595159	SLC26A7_RUNX1T1	GCGGCTGGATGTGAGGGCGATCTGGCTGCAACATGTGTACCCCCA TTGATTGCCAGGGTTGATTCTGATCCGGCTGACTAGGCGAGTG TCCCTTCTACTCTACTGCTCCATGTGTCTCCCTCTGAAGCTGCAC ACTTGCTCGAAGAGGACGACCATCCTGATAGAGGAGGACCGGTGT TCTGTCAAGGGTATACG	422
chr8	96159711	96159969	GDF6	CCGGCTGACCATCCCACCCAGCGCAGGGACCAACGAAAACCCGC GCGGCGCCAGGACCAGGGGGCTGCCGACGCCGCTCGCGGACTA GTTCTCAGACTGTGGGACTCCCTAGTGCCGCTTTGCCAGGGCT TTCCAAGGCTGTCTCATGCCCTAGATCTGCCCCAGCAGCTCAGGCCT TGGACTGCGAACCCAGTATCCCGAGACACCGATTCCATCAGTCCCC ATCCGACCCTCTCCAGCCGGGTTTCATCCG	423
chr8	98974317	98974773	VPS13B_OSIR2	TCGGTGAGGCGTTTCGGTATGGATTGGGTAGGAGCGGCCCTGGGCG ATGGGCTGACGTGCGTGGGCGCAGTTGAGGCCACTGCAAGGCCG CTGGATCCCGGATCCGCACCCGAGACGGAGCGGGGGCCACACGG GATAACCGAGGGGGCGAACGGGAGTTTCGGGCTCCGCTCCCTCT CCGGGTGGGGGACAGGTCGCCGAGTCCGAGGTGCGGCGCGAAGG CCACTCGCATTTTCCGCTTCCGCGAGCAACCCAGGGGCCCTGCG GGAGGAGGAGAGGGTCCCGGGAGTCCGCCCTTCCCTGCGCCTTCG GGACCGGCAGGAGGCGCTGCGCGGGCGAATTAAGAAAGAAAGGAA AAGCTCGTAGTGAGGTGTTACCGCATCCTGCCTTTGGACGTA CTTAGTTGAGTGACCCGATTTCGGACCTTAGGGGCGTTAGGGTCTCC TCCACCG	424
chr8	115667495	115667709	TRPS1	CCGCTGTCAGGCATTTAATCACCGGCCAGTGTCCTTGACCCGCGC GACACATGGCGCATCAACCGCATCGCAGAGGAAGTCTGCCCTTCC TCAGCCCTACGGAAGCGCCGGGCTGCAAGGCCCTGCCACATGG TACGGACAGGGCACAGACCGCTCGGCCAAGCTGTCTGAGCCGCT CTGAGGCGGGTGACCAAGGGATGCGACACCCG	425
chr8	142586176	142586403	ARC_BAI1	CCGGGTGCAGGTTGCGGGGAGGCATGAGGGGAGGCAATTCAGG CAGCAAAGCAGCAGGGTCAAAGGTCAGAGGACGTGGGCCCCGTA GCCTCGGAGGAACCGGAGGAGCAGAGCAGAGGCCAGAGGGCCAG	426

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr8	144050603	144051516	OPLAH_SPATC1	AGTGGGTGGCAGGGAGGCTGGCAAGGGAGTTGTGGCCATTGTC CCAGGACCAGGGGAGCCATCGTGAGCTCTGAACAGGGGAGTGGC ACAGCCCCG GCGCCAAAAGCAGCCCTGGGCCCTGGGTATCGCGCTTGGGGGGAG GGTACCCCCGCGGCTGGGCACGCGCCAAGAGCAGCCCTGGGCC TGGGTATCGTGCTTAGGGGGAGGGTATCGGAGCGGGAAGTGGAC CTGGGGAGCGCCGTCGGCTGAGGCTCTGGCTGATGCCGCCCTCCC CCGATCCCCAGGACCGCGCTGAGCACCTCCGTGCTCCACCAGT CCATGGCCTCTCCCCAAGATGCCGAGGCGGTGAGTTGCGACCTG GATGTAGGCACTGCCCGCCGAAGCGCGGAGGGGCCCTGGCCT TGATGACACCGCCCCCTACCAGGGCCCTGGAGCAGGAGAAAGGG CGCCACCTCTACCTGGCCGGCCTTCCCGGAGAAGCCGCCGAGCTA AGCCCTGGAGAGGTGCGGCGCTGGACTACATCACGTACCGCGGAG TTCCCGGGTGGCTGGGCCTGCGGCACTGGGACGACCTCAACCTG ACTCCCGCCCCAGGAGGTGGAGCAGGTGACGTTCACTACCGCCC TGGAGGGGCTCACGACCACCGGGCAGTGCGCTGCACTCCGAG TCTCAGTGCTCTTAAGGCAAGCACAGATGAGGGGCGCGCGCT GGCGCGCACAGACACGACTCGGAGCACGAAGTAGGCGCCGTAGCT GCGTCCCCAGAACCGGGAGACTTAAGGCATCTTTATTGCGGGATCC TCACACGGCCTCCTGGGCCGCGGATCTCATAGACGTGCGGTGC TCGGAAAGGCCAGTGCTTGCGGGGGCGACCCGCGCGTGGGGC GGGGTCTCCGGGTCCCCATAGCCACCGCCGCGCGGTGTGGAG ACAGAACACATCCTGTTGGCGCGGGGGGGGCGGGGAGGCGGGC TCAGTGCAGGCG TCGGGAGTGAGAAACCAACAAGTGAAGGGCGCCGCTTCCCG GGGCGCAGCTGCGGGCGGGGAGCAGGCGCAGGAGGAGGAAG CGAGCGCCCCGAGCCCCGAGCCGAGTCCCCGAGCCTGAGCCGC AATCGCTGCGGTACTCT	427
chr8	96494022	96494170	SDC2	GAAGCCGAAGAACTGCATGACCGGCTCGCACGAGTCGCGCACGGC CTCGAGAGCCAGCGACACG TTGGAAGGAGGAGGAGCGGCGGAGCAGGCGCAGGAGGAGGAAG AACGACGCGGGACCGGCCGCCCTC ATGACTGTGATACTCAAGTACAGAATTGTGTTGTCAGCCAGAAGTG GTTCAAGAGCCCTCCCGCAAATCATGACTTGCACTCTGGCTTTAAG TGAAGACGAGGGAATCTCAAGGCAGATGGG AAAGTATCAGCGTAGAAGGAATTGTGTCTGCTAGGAAAAGGGTG TGGCAAGAGGAGGAGCGGCACTTGCAGGAGAGCTCGGAACACTCC GCCGAGAATGACTTTTGGAGCCATTTGGCAGAG ACGGAATCTGACCAAGGCTGGACCCTCAATAATTGTGATTTCTTTTC CCCCTTTTCTTCTTGGTAAATCATCCACGAATCTACGCAAGTAG GGCCCTTCGTCATTCTTCGGAGTAGCCGCTTGAGGGCTGGAAGGA GCAGTGATAGAAACCCAGAGACGAGAGACCTCCGAACCTTCGA ACTCGATCACTGTCTCCCCGACCGCCGAACCCGCTGGAGAAGCG GGCGCGACAGGGCGATGAGTTAACGCGGAGGGAGCGCGGAGGCC GCGGAAGCCGGGGGCGCTGGGTCTCAGGCCCGGATGCTGAGCGC GGACCGGCGTGTCTCCCCACAGCGCCCCGCGCGGCTCTCCCCG CTGCGCCCCGACGGCGACCGCGCGGGTAGCCCTGGCGTTTGG CCACGCGCTCGGCTGAGGACCGCTAGAGCTGGGGGAGATCAAA GCATTCTATGGGGCCAAAGAGCCTGGGATTGCAGTGTTGTTAGC CTGGCCTCGCGCGTCAATAAATTTTCGGCG	428
chr8	41308713	41308777	SFRP1	GAAGCCGAAGAACTGCATGACCGGCTCGCACGAGTCGCGCACGGC CTCGAGAGCCAGCGACACG	429
chr8	54457811	54457880	SOX17	TTGGAAGGAGGAGGAGCGGCGGAGCAGGCGCAGGAGGAGGAAG AACGACGCGGGACCGGCCGCCCTC	430
chr8	123320561	123320682	ATAD2	ATGACTGTGATACTCAAGTACAGAATTGTGTTGTCAGCCAGAAGTG GTTCAAGAGCCCTCCCGCAAATCATGACTTGCACTCTGGCTTTAAG TGAAGACGAGGGAATCTCAAGGCAGATGGG	431
chr8	20518007	20518128	ch8:20	AAAGTATCAGCGTAGAAGGAATTGTGTCTGCTAGGAAAAGGGTG TGGCAAGAGGAGGAGCGGCACTTGCAGGAGAGCTCGGAACACTCC GCCGAGAATGACTTTTGGAGCCATTTGGCAGAG ACGGAATCTGACCAAGGCTGGACCCTCAATAATTGTGATTTCTTTTC CCCCTTTTCTTCTTGGTAAATCATCCACGAATCTACGCAAGTAG GGCCCTTCGTCATTCTTCGGAGTAGCCGCTTGAGGGCTGGAAGGA GCAGTGATAGAAACCCAGAGACGAGAGACCTCCGAACCTTCGA ACTCGATCACTGTCTCCCCGACCGCCGAACCCGCTGGAGAAGCG GGCGCGACAGGGCGATGAGTTAACGCGGAGGGAGCGCGGAGGCC GCGGAAGCCGGGGGCGCTGGGTCTCAGGCCCGGATGCTGAGCGC GGACCGGCGTGTCTCCCCACAGCGCCCCGCGCGGCTCTCCCCG CTGCGCCCCGACGGCGACCGCGCGGGTAGCCCTGGCGTTTGG CCACGCGCTCGGCTGAGGACCGCTAGAGCTGGGGGAGATCAAA GCATTCTATGGGGCCAAAGAGCCTGGGATTGCAGTGTTGTTAGC CTGGCCTCGCGCGTCAATAAATTTTCGGCG	432
chr9	1042341	1042870	DMRT2_DMRT3	ACGGAATCTGACCAAGGCTGGACCCTCAATAATTGTGATTTCTTTTC CCCCTTTTCTTCTTGGTAAATCATCCACGAATCTACGCAAGTAG GGCCCTTCGTCATTCTTCGGAGTAGCCGCTTGAGGGCTGGAAGGA GCAGTGATAGAAACCCAGAGACGAGAGACCTCCGAACCTTCGA ACTCGATCACTGTCTCCCCGACCGCCGAACCCGCTGGAGAAGCG GGCGCGACAGGGCGATGAGTTAACGCGGAGGGAGCGCGGAGGCC GCGGAAGCCGGGGGCGCTGGGTCTCAGGCCCGGATGCTGAGCGC GGACCGGCGTGTCTCCCCACAGCGCCCCGCGCGGCTCTCCCCG CTGCGCCCCGACGGCGACCGCGCGGGTAGCCCTGGCGTTTGG CCACGCGCTCGGCTGAGGACCGCTAGAGCTGGGGGAGATCAAA GCATTCTATGGGGCCAAAGAGCCTGGGATTGCAGTGTTGTTAGC CTGGCCTCGCGCGTCAATAAATTTTCGGCG	433

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr9	13323139	13323351	MPDZ_NFIB	TCGTGATCATTGGATGCATCCTCTCGATTCTCATCGTTGCACTGTGCGCGGAGAACACTTTGTTATCCGGCGTTTCTCCCTGCGTGATTATCATTCTTCCCCGCATTGTGGCGGGCTCTGCAGCTAGCAGGGAACCTGATCTCTGGCTGTGCCAAGGAGCTCGCGAGACCGCCCATCTGTCCGGTCTGTCTCCACCAGCTCCTTCGTGCG	434
chr9	14346834	14347128	NFIB_ZDHHC21	ACGTCAGAACAGGGTCTCCTATCAACTGCTACCTATTGCTGTCTCGCAAAACATCCCCCTAAACCCGCTGCATCGACAGCTTCGGGTGAGGGTGGGTAAGAGGCACTTACTGTGAGGCCGAGCTCCCGCACGAATTAGCCTACAACAGGACCTAGGTCTCCTAGGGAGACGAACTAGGCCAGCGAAATCGCGGCCAGGGAGCCCCCTGGCCCCACTCGGGAGACAAACCCGCCGCGCGAAGGGTGCGTCTCCTGAGCTCCACGCCGGGAGCTGGAAGGCAGGCAGACGCGCG	435
chr9	19788616	19788772	SLC24A2	GCGCCCCCTCTGCGCTCTCCCCGACGGCAGGCCCTGCCCCACGCCCCCCATCCCAAGCCAAAAGCAAGGGTAGGAGAGGCGGGGGCTCCAAATCCACGCCCCCGAGCACAGAGAGTTGGCTAACTCTAGCGGGGCCTGGGGGCGCCACATCCACG	436
chr9	19789007	19789198	SLC24A2	TCGCCAGCCGGGCTGGGTTCGGGAGGAGACTGAGCCGCTGTGAGCCGGCGCTCCGAGTCTGGCGCTGCCGGCCCCCGCGGCCCTCCCTCTGGGCTGTGCGCTGTGCGCTGGGAGCGGGGCCGACGCGCTCAGCTCCCGAGTCTTTGCTCCACGCCTCTGGGCGCAGAGCGCAGCTGGCAGCCG	437
chr9	27690518	27690700	C9orf72_LINGO2	CCGGGGAGGAGCCAAGATGGCCAAATAGGAACAGCTCCGGTCTACAGCTCCCAGCGTGAGCGACGCAGAAGACGCTGATTCTGCATTCCATCTGAGGTACCGGGTTCATCTCACTAGGGAGTGCCAGACAGTGGGCGCAGGTCAGTGGGTGCGTGCACCGTGCCTGAGCTGAAGCAGGCGG	438
chr9	36986086	36986754	PAX5_MELK	CCGGCGCCTCGCCCCGCGCGCATCATCTGCTCCGCTGCCAGCTCCCGGCTGCCGCCGCGCCCCGCGCCCCCGGGGCCCGGAAAGCTGGCATCCGTTGTTAGCATAACAACTCAATTGTTCTCAGCGGGGCCCGGCAAATAAAGTCATTATTACGGGCCTCTCCTGGCCGCCGCGGCGCGCGCAATCAGCGGGCCGAGCCACGCGCCAGCGCTGGGACCTGCAGGGCGCGCCGCCCTCCACGCTGCGCCCCGGGCCCGCGCGGCCCGCGCGCGGGGCGAGCGCCGGCCGCGGATTAGTTTTATCTCGAACGTCAATTGACTTAGACTGATTGGCTTCTGCCGCAATGTCAATTAAATTGCAAATGCTTGGCGAGCGCCGGCGCAGCGGGGCGCTCTTCCCGGGGCGCGCGCTCAGCCTTCTCTTTGCGCCACGTTCGGCCGCACTGAATTCATTTCTCCTTCCACGTGCGCGAGGAAATCAGGTGACCTCCTGGAAGTCGTCTGCCCTCCGCCCCCGGCCCTGGGGACTCTCCGTGCGAGCCCGAGCCCCGAGGACTCCCGGCCGGTGGCGGGGAGCTAGGCCACGGGGCGCCCGGACCGCGGGGCCGAGGAGGAAGGGACCGGCTCCCCGCAGGGACCTCG	439
chr9	36986901	36987144	PAX5_MELK	TCGAAGGAGATGGTGGCCGGGTCCCGTCCAGCCCATGCCAGTGCTGGGTGTCCAGAGGGAGGAAGGCCTGGCAGCATACACAGCGTTCACCTGGTGCTGACGCTGTGCCGAGCCACGGATGGGCACAGTCTATCTTCCCCACAGCCCTCCGAAGCAGATACTGTTACTGTCCGACTTCTACAGAGGAGCGAAGTGGGGTGACAGCCAGAGAGTGGCCAGTTGGGTTTCAAACGCTGCG	440
chr9	97853977	97854388	FOXE1	GCGCGGCGAGACGGCAGCAGGGGCCGGGGTCCAGGGGAGAGCCACGGGCCGCGGGGCGGGCGGGCGGCCGCAAGCGCCCCCTGCAGCGGGGAAGCCGCCCTACAGCTACATCGCGTCTATCGCCATGGCC	441

Chr. Num	Start	End	Gene	SEQUENCE (5'-3')	SEQ ID NO:
chr9	117744892	117745189	TLR4	ATCGCGCACGCGCCGAGCGCCGCTCACGCTGGGCGGCATCTAC AAGTTCATCACCGAGCGCTTCCCCTTCTACCGCGACAACCCCAAAAA GTGGCAGAACAGCATCCGCCACAACCTCACACTCAACGACTGCTTC CTCAAGATCCCGCGCGAGGCCGCCGCCGGGTAAGGGCAACTAC TGGGCGCTTGACCCCAACGCGGAGGACATGTTGAGAGCGGCAGC TTCTGCGCCGCCGAAGCGCTTCAAGCGCTCGGACCTCTCCACCT ACCCG CCGATGCCCGAAGTCCTGTGGGCGAGCCTAGCCACAGTAAGTGGT GGAACCTATTAGCGCAGGCCGTTCTCATCAGCGCCACGGAGGACG GAGACGCCGGGGTTCGCGGCTTTGAGCCTCTGGAGCGCCCGCGCC TTCGCGGGCTGCGCGGGGCTCAGGGAGCCGCGGCCACGGCTCCCG CGCGCTCGCTCGCCCGCAGGATCTGGGCGAGCCCGCGGGGACCCG GCTCTGCGCGCAGCCATTGTACAGCTGGCGCAGCCGCGCAAATG ACATCTGAGCCTCCTTTCAAGCCGCCG	442
chr9	124016065	124016323	NEK6_L HX2	GCGGTTCTTTTGTCTGCGCCGATCCTCCTTTAAAGACAGGTCTCAG TTTCCCGGACTTTTCTCCGAGTTTCTGCGCCTGCTGGGGTGA GGGCCGTGACCCTCGGAAGCGAGCCCCCGGGCGGGGACGAGAC CGGAGCAGGCCTGGCCTCGCGCCGGGGTGGGGTGGGGTGGGGT AGGTGGGGGGCTTGGTTCGGATTTCGGCATCTTTGAACCCAGGC CATTCCCGGAGAAGCTCTGCCCTCCCGCG	443
chr9	124503413	124503621	NR5A1_ GPR144	GCGGAGGGACAGCGGGTCAGGGAGGGCCGGCGGAGACCGGCAG CCTGGGGTCCCCGCGCGCCGCCAGCCGCTGTGCGCGGCCCGT CGCGTAATCCCCTCTCTGTGCCAGGCGCTGCCCGCGCACCCACC GAGCGCCCCGCGCAGCGTCCCGGGTGGGTCCGGTGAGTCCCCG CGCCCGGCTTCCCCTGCCAGGCCCCACG	444
chr9	129888368	129888598	USP20_ FNBP1	TCGTCCCCGTTGGCGGGGAGCCCATTGTGGAGCTGTGGGGACTG CCACACTACCATGACCTGTTGGTTTGCAGGGACAGAGGTGCGGC CCTGACTCTTCTACCCCTGTGTATCCGGGCTTGTCTTCTGTCTGCA AGTCAGTCCTCCTGCGTGAAGTGCACACGCTTAGGTACAC CCGTTGCAGGGACCGGAAGTCCATGGCTCTGCCCAACCTGAGC G	445
chr9	129888745	129888845	USP20_ FNBP1	CCGGAAGGGTGGTGTGTGGTCAACCTTGGTTGGTGAGAGGAGCA ATTTCTGGTTTCCACAAGTAAAGACAGCCCATCCCTTGGGACCTG TCCTTTCCG	446
chr9	130896111	130896371	QRFP	CCGGAGAGGACATGGGGTGGGTGGACATCTACCCGACACACCTAC TGCCCAGCTTGCAGGATGGCTTTCATGGGCAGGAAAGCCACAGAC ACCCATGAGGCCCGTGTTCACAGGCACCGGGCTGCGCGGCTAAG CCAGGTGCACCTCCCCGGCAGGTGGAGCCCTCAGCGGCCTGTTACC CAGGAACCAACCAAGGGGGCAGGCAGATGCCAGGACAGCAGT GGAGCATTTGCTGTGCCCCCAGCCCTCCACCG	447
chr9	132587027	132587206	GTF3C4_ BARHL 1	GCGCGGGCAGAGCGCCGAGCGCGCGCAGGGACTGGAGTTCTCG CCAGCTTCGGGTCTTTCTCCCCGAGCTGCCCGGGGGTCTCGGC CTCGGGCGCTCCCGCCCGCTCCTGTTCCCTCAGGGTTTCATGTCCT GTTCCCGGGGCCCCAGAGGTCCCGTCTGAGAGCGGCCCCCGCG	448
chr9	136534110	136534227	SEC16A_ NOTC H1	GCGGGAGACGGGGGAGTCCACTTCTCAAACCCGGTGATCCTGCA GGGCGCTGCACTCACAAAAAGGCTGACTCCACACAGGACCTGCCT CCTGGGCCTTGGCTCAGGCTGGGGCG	449
chr9	21974678	21974780	CDKN2A	CTGGATCGGCCTCCGACCCTAACTATTGGGTGCGTTGGGCAGCGCC CCCGCTCCAGCAGCGCCCGCACCTCCTCTACCCGACCCCGGGCCG CGGCCGTGGCC	450

What is claimed is:

1. A method for detecting the presence of a cancer and for identifying the cancer origin in a test subject, the method comprising:
 - a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
 - b) using the bisulfite treated cfDNA to prepare (i) a first sequencing library for a plurality of specific target genomic regions and (ii) a second sequencing library for a genome of the species of the test subject from a flow through of the first sequencing library;
 - c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
 - d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for the plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of the plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results

for a plurality of liquid biopsies obtained from a cohort of a healthy subject; and
e) responsive to inputting into a model of each of the analyzed sequencing results from
(d)(i)-(d)(iv), receiving as output from the model:

- i. a categorical indication of a presence or absence of the cancer in the test subject,
and
in the case where the model determines presence of the cancer in the test subject, an
origin of the cancer.
2. The method of claim 1, wherein the plurality of specific target genomic regions
comprises at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, or
at least 500 or more cancer specific regions.
3. The method of claim 1 or 2, wherein the plurality of specific target genomic regions
comprises between 400 and 500 cancer specific gene regions.
4. The method of claim 3, wherein the plurality of specific target genomic regions consists
of between 17,500 and 18,500 CpG sites.
5. The method of claim 3, wherein the plurality of specific target genomic regions
comprises at least five nucleic acid sequences selected from SEQ ID NOs: 1-450.
6. The method of claim 3, wherein the plurality of specific target genomic regions
comprises at least 50 nucleic acid sequences selected from SEQ ID NOs: 1-450.
7. The method of claim 3, wherein the plurality of specific target genomic regions
comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 1-450.
8. The method of claim 3, wherein the plurality of specific target genomic regions
comprises at least 300 nucleic acid sequences selected from SEQ ID NOs: 1-450.
9. The method of claim 3, wherein each respective target genomic region in the plurality of

specific target genomic regions encompasses a sequence selected from SEQ ID NOs: 1-450.

10. The method of any one of claims 2-5, wherein at least 20 respective cancer specific genomic regions in the plurality of cancer specific genomic regions encompass an oncogene and/or a tumor suppressor gene listed in Table 23.
11. The method of any one of claims 1-10, wherein the plurality of specific target genomic regions is captured by a set of DNA probes.
12. The method of claim 11, wherein the set of DNA probes comprises DNA fragments with a size ranging between 40 base-pair (bp) and 50 bp, between 51 bp and 60 bp, between 61 bp and 70 bp, between 71 bp and 80 bp, between 81 bp and 90 bp, between 91 bp and 100 bp, between 101 bp and 110 bp, between 111 bp and 120 bp, between 121 bp and 130 bp, between 131 bp and 140 bp, between 141 bp and 150 bp, between 151 bp and 160 bp, between 161 bp and 170 bp, between 171 bp and 180 bp, between 181 bp and 190 bp, or between 191 bp and 200 bp.
13. The method of claim 11 or 12, wherein the set DNA probes comprises DNA fragments with a size ranging between 111 bp and 120 pb or between 121 bp and 130 bp.
14. The method of claim 12 or 13, wherein the set of DNA probes consists of between 400 DNA probes and 500 DNA probes, between 501 DNA probes and 1000 DNA probes, between 1001 DNA probes and 1500 DNA probes, between 1501 DNA probes and 2000 DNA probes, between 2001 DNA probes and 2100 DNA probes, between 2101 DNA probes and 2150 DNA probes, between 2151 DNA probes and 2200 DNA probes, between 2201 DNA probes and 2250 DNA probes, between 2251 DNA probes and 2300 DNA probes, between 2301 DNA probes and 2350 DNA probes, between 2351 DNA probes and 2400 DNA probes, between 2401 DNA probes and 2450 DNA probes, between 2451 DNA probes and 2500 DNA probes, between 2501 DNA probes and 3000 DNA probes, between 3001 DNA probes and 3500 DNA probes, or between 3501 DNA

probes and 4000 DNA probes.

15. The method of any one of claims 11-14, wherein the set DNA probes consists of between 2201 DNA probes and 2250 DNA probes or between 2251 DNA probes and 2300 DNA probes.
16. The method of claim 15, wherein the set of DNA probes comprises at least 10 nucleic acid sequences selected from SEQ ID NOs: 451-2700.
17. The method of claim 15, wherein the set of DNA probes comprises at least 100 nucleic acid sequences selected from SEQ ID NOs: 451-2700.
18. The method of claim 15, wherein the set of DNA probes comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 451-2700.
19. The method of claim 1, wherein the first sequencing library is prepared for paired-end sequencing.
20. The method of any one of claims 1-19, wherein the plurality of specific target genomic regions have a different methylation percentage between the test subject and the cohort of healthy subjects.
21. The method of any one of claims 1-20, wherein the plurality of specific target genomic regions have a methylation percentage higher in the test subject as compared to the cohort of healthy subjects.
22. The method of claim 21, wherein the methylation in the test subject is about two fold higher than the methylation in the cohort of healthy subjects.
23. The method of claim 1, wherein the genome comprises 22 chromosomes.

24. The method of claim 1, wherein the second sequencing library comprises universal adapter sequences.
25. The method of claim 1, the method further comprising converting the second sequencing library into cfDNA sequencing library spheres for genomic sequencing.
26. The method of claim 25, wherein the genomic sequencing comprises rolling circle sequencing or MGI-DNBseq sequencing.
27. The method of claim 1, wherein the analysis of the sequencing results from (d)(ii)-(d)(iv) is performed by measuring non-duplicating fragments in the genome.
28. The method of claim 27, wherein the methylation density for the genome in (d)(ii) is determined for each respective second bin region, , in a plurality of second bins, wherein the plurality of second bins consists of between 2500 second bins and 3000 second bins.
29. The method of claim 28, wherein each respective second bin region consists of between 800,000 nucleotides and 1,200,000 nucleotides.
30. The method of claim 28 or 29, wherein the measuring of the methylation density identifies respective second bin regions in the plurality of second bin regions that are differentially methylated between the test subject and the cohort of healthy subjects.
31. The method of claim 30, wherein the methylation density in each respective second bin region is evaluated based on a Z score value.
32. The method of claim 1, wherein the plurality of first bins is between 2500 first bins and 3000 first bins.
33. The method of claim 32, wherein each first bin in the plurality of first bins represents a different between 800,000 nucleotides and 1,200,000 nucleotides of the genome.

34. The method of claim 1, wherein the measuring of respective copy number of cfDNA identifies a subset of first bins in the plurality of first bins with variation in the number of copies of DNA per bin between the test subject and the cohort of healthy subjects.
35. The method of claim 34, wherein the variation in the number of copies of DNA between the test subject and the cohort of healthy subjects in each first bin is evaluated based on a Z score value.
36. The method of claim 35, wherein the Z score identifies regions of instability in the genome.
37. The method of claim 1, wherein the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third bins, wherein the plurality of third bins consists of between 500 third bins and 600 third bins.
38. The method of claim 37, wherein each respective third bin in the plurality of third bins represents a different between 4.5 million nucleotides (4.5 megabases) and 5.5 million nucleotides (5.5 megabases) of the genome.
39. The method of claim 37 or 38, wherein the measuring of the fragment size pattern distribution of cfDNA identifies a subset of third bins in the plurality of third bins with a variation in the fragment size pattern distribution of cfDNA per bin between the test subject and the cohort of healthy subjects.
40. The method of claim 39, wherein the variation in the fragment size pattern distribution of the cfDNA in each third bin in the plurality of third bins is evaluated based on cfDNA fragment length ratio (RF) value.

41. The method of claim 40, wherein the RF value identifies presence of cancer, wherein cfDNA fragment length released from tumor cells from the test subject is shorter than cfDNA fragment length released by cells of the cohort of healthy subjects.
42. The method of any one of claims 1-41, wherein the cohort of healthy subjects consists of between 5 and 50 healthy subjects, between 5 and 100 healthy subjects, between 5 and 1000 healthy subjects, between 5 and 5000 healthy subjects, between 50 and 500 healthy subjects, between 50 and 1000 healthy subjects, between 50 and 5000 healthy subjects, between 100 and 500 healthy subjects, between 100 and 1000 healthy subjects, between 100 and 5000 healthy subjects, between 500 and 1000 healthy subjects, or between 500 and 5000 healthy subjects, or more.
43. The method of any one of claims 1-42, wherein the liquid biopsy sample comprises a body fluid, blood, or plasma.
44. The method of any one of claims 1-43, wherein the origin of the cancer comprises colorectal cancer (CRC), liver cancer, lung cancer, breast cancer, or gastric cancer.
45. The method of any one of claims 1-44, wherein the subject is a human.
46. The method of any one of claims 1-45, wherein the model is a composite model comprising four attribute models and a combination model, wherein each respective attribute model in the four attribute models produces an initial categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and wherein the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the four attribute models.
47. The method of claim 46, wherein the combination model is a logistic regression combined linear model of the four attribute models, in which each of the four attribute models is independently assigned a different probability weight.

48. The method of any one of claims 1-47, wherein the model comprises at least 100 parameters.
49. The method of any one of claims 1-48, wherein the model comprises a logistic regression, a deep neural network, a fully connected neural network, a convolutional neural network, a graph based neural network, or a support vector machine.
50. The method of claim 49, wherein the deep neural network specifies a tissue for cancer origin.
51. A method for monitoring likelihood of cancer recurrence in a subject previously treated for cancer, the method comprising:
- a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
 - b) using the bisulfite treated cfDNA to prepare (i) a first sequencing library for a plurality of specific target genomic regions and (ii) a second sequencing library for a genome of the species of the test subject from a flow through of the first sequencing library;
 - c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
 - d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from the cohort of healthy subjects;

- iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from the cohort of healthy subjects, and
- iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from the cohort of a healthy subject; and
- e) responsive to inputting into a model each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model:
 - i. a categorical indication of a presence or absence of the cancer in the test subject, and
 - in the case where the model determines presence of the cancer in the test subject, an origin of the cancer,

wherein the detection of a cancer is indicative of cancer recurrence and need of resuming treatment to the subject.

52. A method for assessing the efficacy of a cancer treatment in a subject suffering from cancer, the method comprising:

- a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
- b) using the bisulfite treated cfDNA to prepare (i) a first sequencing library for a plurality of specific target genomic regions and (ii) a second sequencing library for a genome of the species of the test subject from a flow through of the first sequencing library;
- c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
- d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test

- subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
- ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from a cohort of a healthy subject; and
- e) responsive to inputting into a model each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model:
- i. a categorical indication of a presence or absence of the cancer in the test subject, and
- in the case where the model determines presence of the cancer in the test subject, an origin of the cancer,

wherein the detection of a cancer is indicative of efficacy of treatment and need of continuing, modifying or discontinuing treatment of the subject.

53. A method for treating cancer in a subject in need thereof, the method comprising:

- a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
- b) using the bisulfite treated cfDNA to prepare a first sequencing library for (i) a plurality of specific target genomic regions and (ii) a second sequencing library for a genome of the species of the test subject from a flow through of the first sequencing library;

- c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
- d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from the cohort of a healthy subject; and
- e) responsive to inputting into a model each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model:
 - i. a categorical indication of a presence or absence of the cancer in the test subject, and
 - in the case where the model determines presence of the cancer in the test subject, an origin of the cancer,

wherein the detection of a cancer and the identification of the cancer origin are indicative of the need to treat the subject and the type of treatment that is the most efficacious given the tumor origin.