(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 April 2014 (10.04.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/055790 A2

(51) International Patent Classification:
G06F 19/10 (2011.01)

(21) International Application Number:
PCT/US2013/063314

(22) International Filing Date:
3 October 2013 (03.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/709,909	4 October 2012 (04.10.2012)	US
13/797,930	12 March 2013 (12.03.2013)	US
13/829,164	14 March 2013 (14.03.2013)	US
13/829,373	14 March 2013 (14.03.2013)	US

(71) Applicant: **SEQUENOM, INC.** [US/US]; 3595 John Hopkins Court, San Diego, CA 92121 (US).

(72) Inventors: **DECIU, Cosmin**; 10545 Sea Mist Way, San Diego, CA 92121 (US). **DZAKULA, Zeljko**; 12830 Sundance Avenue, San Diego, CA 92129 (US).

(74) Agents: **FORCE, Walker, R.** et al.; Grant Anderson LLP, c/o Portfolioip, P.O. Box 52050, Minneapolis, MN 55402 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/055790 A2

(54) Title: METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS

(57) Abstract: Provided herein are methods, processes and apparatuses for non-invasive assessment of genetic variations.

METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS

Related Patent Applications

This patent application claims the benefit of U.S. Provisional Patent Application No. 61/709,909 filed on October 4, 2012, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu, Mathias Ehrich, Zeljko Dzakula and Amin Mazloom as inventors, and designated by Attorney Docket No. SEQ-6050-PV. This patent application also claims the benefit of (i) U.S. patent application no. 13/829,164 filed March 14, 2013, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu, Zeljko Dzakula, John Tynan and Grant Hogg as inventors, and designated by attorney docket no. SEQ-6034-CP2t; (ii) U.S. patent application no. 13/829,373 filed March 14, 2013, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu and Zeljko Dzakula as inventors, and designated by attorney docket no. SEQ-6034-CP3; and (iii) U.S. patent application no. 13/797,930 filed March 12, 2013, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu, Zeljko Dzakula, Mathias Ehrich and Taylor Jensen as inventors, and designated by attorney docket no. SEQ-6034-CP; each of which claims the benefit of U.S. patent application no. 13/669,136 filed November 5, 2012, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu, Zeljko Dzakula, Mathias Ehrich and Sung Kim as inventors, and designated by attorney docket no. SEQ-6034-CTt, which is a continuation of International PCT Application No. PCT/US2012/059123 filed October 5, 2012, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu, Zeljko Dzakula, Mathias Ehrich and Sung Kim as inventors, and designated by Attorney Docket No. SEQ-6034-PC, which claims the benefit of (i) U.S. Provisional Patent Application No. 61/709,899 filed on October 4, 2012, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Cosmin Deciu, Zeljko Dzakula, Mathias Ehrich and Sung Kim as inventors, and designated by Attorney Docket No. SEQ-6034-PV3; (ii) U.S. Provisional Patent Application No. 61/663,477 filed on June 22, 2012, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS, naming Zeljko Dzakula and Mathias Ehrich as inventors, and designated by Attorney Docket No. SEQ-6034-PV2; and (iii) U.S. Provisional Patent Application No. 61/544,251 filed on October 6, 2011, entitled METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC

VARIATIONS, naming Zeljko Dzakula and Mathias Ehrich as inventors, and designated by Attorney Docket No. SEQ-6034-PV. The entire content of the foregoing applications is incorporated herein by reference, including all text, tables and drawings.

Sequence Listing

The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on September 18, 2013, is named SEQ-6034-PC2_SL.txt and is 434,112 bytes in size.

Field

Technology provided herein relates in part to methods, processes and apparatuses for non-invasive assessment of genetic variations.

Background

Genetic information of living organisms (e.g., animals, plants and microorganisms) and other forms of replicating genetic information (e.g., viruses) is encoded in deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). Genetic information is a succession of nucleotides or modified nucleotides representing the primary structure of chemical or hypothetical nucleic acids. In humans, the complete genome contains about 30,000 genes located on twenty-four (24) chromosomes (see The Human Genome, T. Strachan, BIOS Scientific Publishers, 1992). Each gene encodes a specific protein, which after expression via transcription and translation fulfills a specific biochemical function within a living cell.

Many medical conditions are caused by one or more genetic variations. Certain genetic variations cause medical conditions that include, for example, hemophilia, thalassemia, Duchenne Muscular Dystrophy (DMD), Huntington's Disease (HD), Alzheimer's Disease and Cystic Fibrosis (CF) (Human Genome Mutations, D. N. Cooper and M. Krawczak, BIOS Publishers, 1993). Such genetic diseases can result from an addition, substitution, or deletion of a single nucleotide in DNA of a particular gene. Certain birth defects are caused by a chromosomal abnormality, also referred to as an aneuploidy, such as Trisomy 21 (Down's Syndrome), Trisomy 13 (Patau Syndrome), Trisomy 18 (Edward's Syndrome), Monosomy X (Turner's Syndrome) and certain sex chromosome

aneuploidies such as Klinefelter's Syndrome (XXY), for example. Another genetic variation is fetal gender, which can often be determined based on sex chromosomes X and Y. Some genetic variations may predispose an individual to, or cause, any of a number of diseases such as, for example, diabetes, arteriosclerosis, obesity, various autoimmune diseases and cancer (e.g., colorectal, breast, ovarian, lung).

Identifying one or more genetic variations or variances can lead to diagnosis of, or determining predisposition to, a particular medical condition. Identifying a genetic variance can result in facilitating a medical decision and/or employing a helpful medical procedure. In certain embodiments, identification of one or more genetic variations or variances involves the analysis of cell-free DNA. Cell-free DNA (CF-DNA) is composed of DNA fragments that originate from cell death and circulate in peripheral blood. High concentrations of CF-DNA can be indicative of certain clinical conditions such as cancer, trauma, burns, myocardial infarction, stroke, sepsis, infection, and other illnesses. Additionally, cell-free fetal DNA (CFF-DNA) can be detected in the maternal bloodstream and used for various noninvasive prenatal diagnostics.
Non-invasive prenatal testing is becoming a field of rapidly growing interest. Early detection of pregnancy-related conditions, including complications during pregnancy and genetic defects of the fetus is of crucial importance, as it allows early medical intervention necessary for the safety of both the mother and the fetus. Prenatal diagnosis has been conducted using cells isolated from the fetus through procedures such as chorionic villus sampling (CVS) or amniocentesis. However, these conventional methods are invasive and present an appreciable risk to both the mother and the fetus. The National Health Service currently cites a miscarriage rate of between 1 and 2 per cent following the invasive amniocentesis and chorionic villus sampling (CVS) tests.

An alternative to these invasive approaches has been developed for prenatal screening, e.g., to detecting fetal abnormalities, following the discovery that circulating cell-free fetal nucleic acid can be detected in maternal plasma and serum (Lo et al., Lancet 350:485-487, 1997; and U.S. Patent 6,258,540). Circulating cell free fetal nucleic acid (cffNA) has several advantages making it more applicable for non-invasive prenatal testing. For example, cell free nucleic acid is present at higher levels than fetal cells and at concentrations sufficient for genetic analysis. Also, cffNA is cleared from the maternal bloodstream within hours after delivery, preventing contamination from previous pregnancies. The presence of fetal nucleic acid in maternal plasma allows for non-invasive prenatal diagnosis through the analysis of a maternal blood sample. Hence, fetal nucleic

acid analysis in maternal plasma can be a useful mechanism for the monitoring of fetomaternal well-being.

Examples of prenatal tests performed by detecting fetal DNA in maternal plasma or serum include fetal rhesus D (RhD) genotyping (Lo et al., N. Engl. J. Med. 339:1734-1738, 1998), fetal sex determination (Costa et al., N. Engl. J. Med. 346:1502, 2002), and diagnosis of several fetal disorders (Amicucci et al., Clin. Chem. 46:301-302, 2000; Saito et al., Lancet 356:1170, 2000; and Chiu et al., Lancet 360:998-1000, 2002). In addition, quantitative abnormalities of fetal DNA in maternal plasma/serum have been reported in preeclampsia (Lo et al., Clin. Chem. 45:184-188, 1999 and Zhong et al., Am. J. Obstet. Gynecol. 184:414-419, 2001), fetal trisomy 21 (Lo et al., Clin. Chem. 45:1747-1751, 1999 and Zhong et al., Prenat. Diagn. 20:795-798, 2000) and hyperemesis gravidarum (Sekizawa et al., Clin. Chem. 47:2164-2165, 2001).

Summary

Provided herein, in some embodiments is a method for determining fetal ploidy according to nucleic acid sequence reads, comprising (a) determining a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus, (b) obtaining counts of sequence reads mapped to portions of a reference genome, which sequence reads are from the nucleic acid in the sample, (c) calculating a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels, and (d) determining fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a). In some embodiments the fetal fraction is determined from a first part of the test sample and the genomic section levels are determined from a second part of the test sample. In some embodiments calculating the genomic section level for each of the portions of the reference genome comprises normalizing counts of reads mapped to the reference genome according to guanine and cytosine (GC) content for each of the portions. In certain embodiments, the method comprises (1) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions, and (2) calculating the genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing

calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels. In some embodiments of the method, the subset of portions of the reference genome in (d)(i) is portions of a chromosome or a segment thereof. In some embodiments the chromosome is chosen from chromosome 13, chromosome 18 and chromosome 21. In some embodiments the method comprises determining a reference count and an uncertainty value according to the reference count. In some embodiments the reference count is determined according to calculated genomic section levels for a subset of portions of the reference genome for one or more pregnant females bearing a fetus. In some embodiments the reference count is determined where the subset of portions of the reference genome for one or more pregnant females are known to be euploid for the fetus and/or the mother, and where the reference count is not determined from the sample. In some embodiments the reference count is determined from the same subset of portions of the reference genome as in (d). In some embodiments the reference count is normalized by bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof. In some embodiments a maternal ploidy is determined. In certain embodiments the fetal ploidy is determined in (d) according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a), (iii) a maternal ploidy, (iv) the reference count and (v) an uncertainty value σ for the reference count. In some embodiments the fraction of fetal nucleic acid determined in (a) is fixed at its determined value and fetal ploidy *X* is determined according to Equation 8 below, or a derivation thereof:

$$y_i = (1 - F)M_i f_i + FXf_i \qquad (8)$$

where y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, *X* represents the fetal ploidy, and M_i represents the maternal ploidy of portion *i*. In some embodiments determining the fetal fraction comprises determining the sum of squared residuals according to equation (8) and for multiple bins *i* for a subset of portions of the reference genome. In some embodiments the fetal fraction is fixed at a value determined in (a) and the fetal ploidy is varied to optimize the sum of squared residuals according to equation (8) or a variation thereof. In some embodiments a linear regression is determined according to the sum of square residuals. In some embodiments the fetal ploidy is determined according to Equation 20 below:

$$X = \frac{\sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^{N} \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}} \quad . \qquad (20)$$

wherein y_i represents the calculated genomic section level for portion i of a reference genome, F represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for i, σ represents the uncertainty value for f_i, X represents the fetal ploidy, and M_i represents the maternal ploidy of portion i. In some embodiments the fetal ploidy is determined according to Equation 21 below:

$$X = \frac{\Xi_{fy} - (1-F)\Xi_{ff}}{F\Xi_{ff}} = \frac{\Xi_{fy}}{F\Xi_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Xi_{fy}}{\Xi_{ff}} - 1\right) \qquad (21)$$

wherein $\Xi_{ff} = \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}$, $\Xi_{fy} = \sum_{i=1}^{N} \frac{M_i f_i}{\sigma_i^2}$, y_i represents the calculated genomic section level for portion i of a reference genome, F represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for i, σ represents the uncertainty value for f_i, and X represents the fetal ploidy. In some embodiments the presence or absence of a fetal chromosome aneuploidy is determined according to the fetal ploidy determined in (d). In some embodiments the fetal ploidy determined in (d) is about 1.4 or greater and the presence of a fetal chromosome aneuploidy is determined. In some embodiments the fetal ploidy determined in (d) is about 1.2 or less and the absence of a fetal chromosome aneuploidy is determined. In some embodiments the fetal chromosome aneuploidy is a trisomy. In some embodiments the trisomy is selected from a trisomy of chromosome 13, 18 and 21. In some embodiments determining the fraction of fetal nucleic acid comprises analyzing the calculated genomic sections levels for a subset of portions of the reference genome, which subset is a first subset, the subset in (d) is a second subset, and the first subset of portions of the reference genome is portions of a Y chromosome or a segment thereof. In some embodiments determining the fraction of fetal nucleic acid comprises (1) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, (2) from the counts in (1), generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome,

and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, (3) determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (2) and a fitted relationship, wherein the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females, and the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus. In some embodiments determining the fraction of fetal nucleic acid comprises analyzing one or more loci in sample nucleic acid, wherein at least one of the one or more loci vary between fetal nucleic acid and maternal nucleic acid. In some embodiments the one or more loci comprise one or more polymorphic sites, comprising (1) enriching nucleic acid in a first part of the test sample for a plurality of polymorphic sites, (2) obtaining nucleotide sequences for some or all of the polymorphic sites by a sequencing process, (3) analyzing the nucleotide sequences of (2), and (4) determining the fraction of fetal nucleic acid based on the analysis of (3), wherein the polymorphic sites and number thereof result in at least five polymorphic sites being informative for determining the fetal fraction for at least 90% of samples. In some embodiments the one or more loci comprise one or more methylation regions, comprising, (1) contacting the test sample with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid, and (2) determining the fraction of fetal nucleic acid in the sample based on the differentially modified nucleic acid. In some embodiments the one or more agents are methylation sensitive restriction enzymes.

Also provided herein is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample obtained from a pregnant female bearing a fetus, and which instructions executable by the one or more processors are configured to (a) determine a fraction of fetal nucleic acid in the test sample, (b) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels, and (c) determine fetal ploidy according to (i) the calculated

genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

As used herein, the term “genomic sections” of a reference genome is the same as “portions of a reference genome”.

Certain aspects of the technology are described further in the following description, examples, claims and drawings.

Brief Description of the Drawings

The drawings illustrate embodiments of the technology and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be shown exaggerated or enlarged to facilitate an understanding of particular embodiments.

FIG. 1 graphically illustrates how increased uncertainty in bin counts within a genomic region sometimes reduces gaps between euploid and trisomy Z-values.

FIG. 2 graphically illustrates how decreased differences between triploid and euploid number of counts within a genomic region sometimes reduces predictive power of Z-scores. See Example 1 for experimental details and results.

FIG. 3 graphically illustrates the dependence of p-values on the position of genomic bins within chromosome 21.

FIG. 4 schematically represents a bin filtering procedure. A large number of euploid samples are lined up, bin count uncertainties (SD or MAD values) are evaluated, and bins with largest uncertainties sometimes are filtered out.

FIG. 5 graphically illustrates count profiles for chromosome 21 in two patients.

FIG. 6 graphically illustrates count profiles for patients used to filter out uninformative bins from chromosome 18. In FIG. 6, the two bottom traces show a patient with a large deletion in chromosome 18. See Example 1 for experimental details and results.

FIG. 7 graphically illustrates the dependence of p-values on the position of genomic bins within chromosome 18.

FIG. 8 schematically represents bin count normalization. The procedure first lines up known euploid count profiles, from a data set, and normalizes them with respect to total counts. For each bin, the median counts and deviations from the medians are evaluated. Bins with too much variability (exceeding 3 mean absolute deviations (e.g., MAD)), in certain embodiments, are eliminated. The remaining bins are normalized again with respect to residual total counts, and medians are re-evaluated following the renormalization, in some embodiments. Finally, the resulting reference profile (see bottom trace, left panel) is used to normalize bin counts in test samples (see top trace, left panel), smoothing the count contour (see trace on the right) and leaving gaps where uninformative bins have been excluded from consideration.

FIG. 9 graphically illustrates the expected behavior of normalized count profiles. The majority of normalized bin counts often will center on 1, with random noise superimposed. Maternal deletions and duplications sometimes shifts the elevation to an integer multiple of 0.5. Profile elevations corresponding to a triploid fetal chromosome often shifts upward in proportion to the fetal fraction. See Example 1 for experimental details and results.

FIG. 10 graphically illustrates a normalized T18 count profile with a heterozygous maternal deletion in chromosome 18. The light gray segment of the graph tracing shows a higher average elevation than the black segment of the graph tracing. See Example 1 for experimental details and results.

FIG. 11 graphically illustrates normalized binwise count profiles for two samples collected from the same patient with heterozygous maternal deletion in chromosome 18. The substantially identical tracings can be used to determine if two samples are from the same donor.

FIG. 12 graphically illustrates normalized binwise count profiles of a sample from one study, compared with two samples from a previous study. The duplication in chromosome 22 unambiguously points out the patient's identity.

FIG. 13 graphically illustrates normalized binwise count profiles of chromosome 4 in the same three patients presented in FIG. 12. The duplication in chromosome 4 confirms the patient's identity established in FIG. 12. See Example 1 for experimental details and results.

FIG. 14 graphically illustrates the distribution of normalized bin counts in chromosome 5 from a euploid sample.

FIG. 15 graphically illustrates two samples with different levels of noise in their normalized count profiles.

FIG. 16 schematically represents factors determining the confidence in peak elevation: noise standard deviation (e.g., σ) and average deviation from the reference baseline (e.g., Δ). See Example 1 for experimental details and results.

FIG. 17 graphically illustrates the results of applying a correlation function to normalized bin counts. The correlation function shown in FIG. 17 was used to normalize bin counts in chromosome 5 of an arbitrarily chosen euploid patient.

FIG. 18 graphically illustrates the standard deviation for the average stretch elevation in chromosome 5, evaluated as a sample estimate (square data points) and compared with the standard error of the mean (triangle data points) and with the estimate corrected for auto-correlation $\rho = 0.5$ (circular data points). The aberration depicted in FIG. 18 is about 18 bins long. See Example 1 for experimental details and results.

FIG. 19 graphically illustrates Z-values calculated for average peak elevation in chromosome 4. The patient has a heterozygous maternal duplication in chromosome 4 (see FIG. 13).

FIG. 20 graphically illustrates p-values for average peak elevation, based on a t-test and the Z-values from FIG. 19. The order of the t-distribution is determined by the length of the aberration. See Example 1 for experimental details and results.

FIG. 21 schematically represents edge comparisons between matching aberrations from different samples. Illustrated in FIG. 21 are overlaps, containment, and neighboring deviations.

FIG. 22 graphically illustrates matching heterozygous duplications in chromosome 4 (top trace and bottom trace), contrasted with a marginally touching aberration in an unrelated sample (middle trace). See Example 1 for experimental details and results.

FIG. 23 schematically represents edge detection by means of numerically evaluated first derivatives of count profiles.

FIG. 24 graphically illustrates that first derivative of count profiles, obtained from real data, are difficult to distinguish from noise.

FIG. 25 graphically illustrates the third power of the count profile, shifted by 1 to suppress noise and enhance signal (see top trace). Also illustrated in FIG. 25 (see bottom trace) is a first derivative of the top trace. Edges are unmistakably detectable. See Example 1 for experimental details and results.

FIG. 26 graphically illustrates histograms of median chromosome 21 elevations for various patients. The dotted histogram illustrates median chromosome 21 elevations for 86 euploid patients. The hatched histogram illustrates median chromosome 21 elevations for 35 trisomy 21 patients. The count profiles were normalized with respect to a euploid reference set prior to evaluating median elevations.

FIG. 27 graphically illustrates a distribution of normalized counts for chromosome 21 in a trisomy sample.

FIG. 28 graphically represents area ratios for various patients. The dotted histogram illustrates chromosome 21 area ratios for 86 euploid patients. The hatched histogram illustrates chromosome 21 area ratios for 35 trisomy 21 patients. The count profiles were normalized with respect to a euploid reference set prior to evaluating area ratios. See Example 1 for experimental details and results.

FIG. 29 graphically illustrates area ratio in chromosome 21 plotted against median normalized count elevations. The open circles represent about 86 euploid samples. The filled circles represent about 35 trisomy patients. See Example 1 for experimental details and results.

FIG. 30 graphically illustrates relationships among 9 different classification criteria, as evaluated for a set of trisomy patients. The criteria involve Z-scores, median normalized count elevations, area ratios, measured fetal fractions, fitted fetal fractions, the ratio between fitted and measured fetal fractions, sum of squared residuals for fitted fetal fractions, sum of squared residuals with fixed fetal fractions and fixed ploidy, and fitted ploidy values. See Example 1 for experimental details and results.

FIG. 31 graphically illustrates simulated functional Phi profiles for trisomy (dashed line) and euploid cases (solid line, bottom).

FIG. 32 graphically illustrates functional Phi values derived from measured trisomy (filled circles) and euploid data sets (open circles). See Example 2 for experimental details and results.

FIG. 33 graphically illustrates linearized sum of squared differences as a function of measured fetal fraction.

FIG. 34 graphically illustrates fetal fraction estimates based on Y-counts plotted against values obtained from a fetal quantifier assay (e.g., FQA) fetal fraction values.

FIG. 35 graphically illustrates Z-values for T21 patients plotted against FQA fetal fraction measurements. For FIG. 33-35 see Example 2 for experimental details and results.

FIG. 36 graphically illustrates fetal fraction estimates based on chromosome Y plotted against measured fetal fractions.

FIG. 37 graphically illustrates fetal fraction estimates based on chromosome 21 (Chr21) plotted against measured fetal fractions.

FIG. 38 graphically illustrates fetal fraction estimates derived from chromosome X counts plotted against measured fetal fractions.

FIG. 39 graphically illustrates medians of normalized bin counts for T21 cases plotted against measured fetal fractions. For FIG. 36-39 see Example 2 for experimental details and results.

FIG. 40 graphically illustrates simulated profiles of fitted triploid ploidy (e.g., X) as a function of F_0 with fixed errors ΔF = +/- 0.2%.

FIG. 41 graphically illustrates fitted triploid ploidy values as a function of measured fetal fractions. For FIG. 40 and 41 see Example 2 for experimental details and results.

FIG. 42 graphically illustrates probability distributions for fitted ploidy at different levels of errors in measured fetal fractions. The top panel in FIG. 42 sets measured fetal fraction error to 0.2%. The middle panel in FIG. 42 sets measured fetal fraction error to 0.4%. The bottom panel in FIG. 42 sets measured fetal fraction error to 0.6%. See Example 2 for experimental details and results.

FIG. 43 graphically illustrates euploid and trisomy distributions of fitted ploidy values for a data set derived from patient samples.

FIG. 44 graphically illustrates fitted fetal fractions plotted against measured fetal fractions. For FIG. 43 and 44 see Example 2 for experimental details and results.

FIG. 45 schematically illustrates the predicted difference between euploid and trisomy sums of squared residuals for fitted fetal fraction as a function of the measured fetal fraction.

FIG. 46 graphically illustrates the difference between euploid and trisomy sums of squared residuals as a function of the measured fetal fraction using a data set derived from patient samples. The data points are obtained by fitting fetal fraction values assuming fixed uncertainties in fetal fraction measurements.

FIG. 47 graphically illustrates the difference between euploid and trisomy sums of squared residuals as a function of the measured fetal fraction. The data points are obtained by fitting fetal fraction values assuming that uncertainties in fetal fraction measurements are proportional to fetal fractions: $\Delta F = 2/3 + F_0/6$. For FIG. 45-47 see Example 2 for experimental details and results.

FIG. 48 schematically illustrates the predicted dependence of the fitted fetal fraction plotted against measured fetal fraction profiles on systematic offsets in reference counts. The lower and upper branches represent euploid and triploids cases, respectively.

FIG. 49 graphically represents the effects of simulated systematic errors Δ artificially imposed on actual data. The main diagonal in the upper panel and the upper diagonal in the lower right panel represent ideal agreement. The dark gray line in all panels represents equations (51) and (53) for euploid and triploid cases, respectively. The data points represent actual measurements incorporating various levels of artificial systematic shifts. The systematic shifts are given as the offset above each panel. For FIG. 48 and 49 see Example 2 for experimental details and results.

FIG. 50 graphically illustrates fitted fetal fraction as a function of the systematic offset, obtained for a euploid and for a triploid data set.

FIG. 51 graphically illustrates simulations based on equation (61), along with fitted fetal fractions for actual data. Black lines represent two standard deviations (obtained as square root of equation (61)) above and below equation (40). ΔF is set to $2/3 + F_0/6$. For FIG. 50 and 51 see Example 2 for experimental details and results.

Example 3 addresses FIG. 52 to 61F.

FIG. 52 graphically illustrates an example of application of the cumulative sum algorithm to a heterozygous maternal microdeletion in chromosome 12, bin 1457. The difference between the intercepts associated with the left and the right linear models is 2.92, indicating that the heterozygous deletion is 6 bins wide.

FIG. 53 graphically illustrates a hypothetical heterozygous deletion, approximately 2 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is -1.

FIG. 54 graphically illustrates a hypothetical homozygous deletion, approximately 2 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is -2.

FIG. 55 graphically illustrates a hypothetical heterozygous deletion, approximately 6 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is -3.

FIG. 56 graphically illustrates a hypothetical homozygous deletion, approximately 6 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is -6.

FIG. 57 graphically illustrates a hypothetical heterozygous duplication, approximately 2 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is 1.

FIG. 58 graphically illustrates a hypothetical homozygous duplication, approximately 2 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is 2.

FIG. 59 graphically illustrates a hypothetical heterozygous duplication, approximately 6 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is 3.

FIG. 60 graphically illustrates a hypothetical homozygous duplication, approximately 6 genomic sections wide, and its associated cumulative sum profile. The difference between the left and the right intercepts is 6.

FIG. 61A-F graphically illustrate candidates for fetal heterozygous duplications in data obtained from women and infant clinical studies with high fetal fraction values (40-50%). To rule out the possibility that the aberrations originate from the mother and not the fetus, independent maternal profiles were used. The profile elevation in the affected regions is approximately 1.25, in accordance with the fetal fraction estimates.

FIG. 62 shows a profile of elevations for Chr20, Chr21 (~55750 to ~56750) and Chr22 obtained from a pregnant female bearing a euploid fetus.

FIG. 63 shows a profile of elevations for Chr20, Chr21 (~55750 to ~56750) and Chr22 obtained from a pregnant female bearing a trisomy 21 fetus.

FIG. 64 shows a profile of raw counts for Chr20, Chr21 (~55750 to ~56750) and Chr22 obtained from a pregnant female bearing a euploid fetus.

FIG. 65 shows a profile of raw counts for Chr20, Chr21 (~55750 to ~56750) and Chr22 obtained from a pregnant female bearing a trisomy 21 fetus.

FIG. 66 shows a profile of normalized counts for Chr20, Chr21 (~55750 to ~56750) and Chr22 obtained from pregnant females bearing a euploid fetus.

FIG. 67 shows a profile of normalized counts for Chr20, Chr21 (~55750 to ~56750) and Chr22 obtained from a pregnant female bearing a trisomy 21 fetus.

FIG. 68 shows a profile of normalized counts for Chr20, Chr21 (~47750 to ~48375) and Chr22 obtained from pregnant females bearing a euploid fetus.

FIG. 69 shows a profile of normalized counts for Chr20, Chr21 (~47750 to ~48375) and Chr22 obtained from a pregnant female bearing a trisomy 21 fetus.

FIG. 70 shows a graph of counts (y axis) versus GC content (X axis) before LOESS GC correction (upper panel) and after LOESS GC (lower panel).

FIG. 71 shows a graph of counts normalized by LOESS GC (Y axis) versus GC fraction for multiple samples of chromosome 1.

FIG. 72 shows a graph of counts normalized by LOESS GC and corrected for tilt (Y axis) versus GC fraction (X axis) for multiple samples of chromosome 1.

FIG. 73 shows a graph of variance (Y-axis) versus GC fraction (X axis) for chromosome 1 before tilting (filled symbols) and after tilting (open circles).

FIG. 74 shows a graph of frequency (Y-axis) versus GC fraction (X axis) for chromosome as well as a median (left vertical line) and mean (right vertical line).

FIG. 75A-F shows a graph of counts normalized by LOESS GC and corrected for tilt (Y axis) versus GC fraction (X axis) left panels and frequency (Y-axis) versus GC fraction (X axis)(right panels) for chromosomes 4, 13 and X (FIG. 75A, listed from top to bottom), chromosomes 5, 6 and

3 (FIG. 75B, listed from top to bottom), chromosomes 8, 2, 7 and 18 (FIG. 75C, listed from top to bottom), chromosomes 12, 14, 11 and 9 (FIG. 75D, listed from top to bottom), chromosomes 21, 1, 10, 15 and 20 (FIG. 75E, listed from top to bottom) and chromosomes 16, 17, 22 and 19 (FIG. 75F, listed from top to bottom). Median values (left vertical line) and mean values (right vertical line) are indicated in the right panels.

FIG. 76 shows a graph of counts normalized by LOESS GC and corrected for tilt (Y axis) versus GC fraction (X axis) for chromosome 19. The chromosome pivot is shown in the right boxed regions and the genome pivot is shown in the left boxed region.

FIG. 77 shows a graph of log-transformed p-values (Y axis) versus bins (X-axis) for chromosomes 13 (top left), 21 (top middle), and 18 (top right). The chromosomal position of certain bins is shown in the bottom panel.

FIG. 78 shows the Z-score for chromosome 21 where uninformative bins were excluded from the Z-score calculation (Y-axis) and Z-score for chromosome 21 for all bins (X-axis). Trisomy 21 cases are indicated by filled circles. Euploids are indicated by open circles.

FIG. 79 shows the Z-score for chromosome 18 where uninformative bins were excluded from the Z-score calculation (Y-axis) and Z-score for chromosome 18 for all bins (X-axis).

FIG. 80 shows a graph of selected bins (Y axis) verse all bins (X axis) for chromosome 18.

FIG. 81 shows a graph of selected bins (Y axis) verse all bins (X axis) for chromosome 21.

FIG. 82 shows a graph of counts (Y axis) verse GC content (X axis) for 7 samples.

FIG. 83 shows a graph of raw counts (Y axis) verse GC bias coefficients (X axis).

FIG. 84 shows a graph of frequency (Y axis) verse intercepts (X axis).

FIG. 85 shows a graph of frequency (Y axis) verse slopes (X axis).

FIG. 86 shows a graph of Log Median Count (Y axis) verse Log Intercept (X axis).

FIG. 87 shows a graph of frequency (Y axis) verse slope (X axis).

FIG. 88 shows a graph of frequency (Y axis) verse GC content (X axis).

FIG. 89 shows a graph of slope (Y axis) verse GC content (X axis).

FIG. 90 shows a graph of cross-validation errors (Y axis) verse R work (X axis) for bins chr2_2404.

FIG. 91 shows a graph of cross-validation errors (Y axis) verse R work (X axis) (Top Left), raw counts (Y axis) verse GC bias coefficients (X axis)(Top Right), frequency (Y axis) verse intercepts (X axis) (Bottom Left), and frequency (Y axis) verse slope (X axis)(Bottom Right) for bins chr2_2345.

FIG. 92 shows a graph of cross-validation errors (Y axis) verse R work (X axis) (Top Left), raw counts (Y axis) verse GC bias coefficients (X axis)(Top Right), frequency (Y axis) verse intercepts (X axis) (Bottom Left), and frequency (Y axis) verse slope (X axis)(Bottom Right) for bins chr1_31.

FIG. 93 shows a graph of cross-validation errors (Y axis) verse R work (X axis) (Top Left), raw counts (Y axis) verse GC bias coefficients (X axis)(Top Right), frequency (Y axis) verse intercepts (X axis) (Bottom Left), and frequency (Y axis) verse slope (X axis)(Bottom Right) for bins chr1_10.

FIG. 94 shows a graph of cross-validation errors (Y axis) verse R work (X axis) (Top Left), raw counts (Y axis) verse GC bias coefficients (X axis)(Top Right), frequency (Y axis) verse intercepts (X axis) (Bottom Left), and frequency (Y axis) verse slope (X axis)(Bottom Right) for bins chr1_9.

FIG. 95 shows a graph of cross-validation errors (Y axis) verse R work (X axis) (Top Left), raw counts (Y axis) verse GC bias coefficients (X axis)(Top Right), frequency (Y axis) verse intercepts (X axis) (Bottom Left), and frequency (Y axis) verse slope (X axis)(Bottom Right) for bins chr1_8.

FIG. 96 shows a graph of frequency (Y axis) verse max(R_{cv}, R_{work}) (X axis).

FIG. 97 shows a graph of technical replicates (X axis) verse Log10 cross-validation errors (X axis).

FIG. 98 shows a graph of Z score gap separation (Y axis) verse cross validation error threshold (X axis) for Chr21.

FIG. 99A (all bins) and FIG. 99B (cross-validated bins) demonstrates that the bin selection described in example 4 mostly removes bins with low mappability.

FIG. 100 shows a graph of normalized counts (Y axis) verse GC (X axis) bias for Chr18_6.

FIG. 101 show a graph of normalized counts (Y axis) verse GC bias (X axis) for Chr18_8.

FIG. 102 shows a histogram of frequency (Y axis) verse intercept error (X axis).

FIG. 103 shows a histogram of frequency (Y axis) verse slope error (X axis).

FIG. 104 shows a graph of slope error (Y axis) verse intercept (X axis).

FIG. 105 shows a normalized profile that includes Chr4 (about 12400 to about 15750) with elevation (Y axis) and bin number (X axis).

FIG. 106 shows a profile of raw counts (Top Panel) and normalized counts (Bottom Panel) for Chr20, Chr21 and Chr22. Also shown is a distribution of standard deviations (X axis) verse frequency (Y axis) for the profiles before (top) and after (bottom) PERUN normalization.

FIG. 107 shows a distribution of chromosome representations for euploids and trisomy cases for raw counts (top), repeat masking (middle) and normalized counts (bottom).

FIG. 108 shows a graph of results obtained with a linear additive model (Y axis) verse a GCRM for Chr13.

FIG. 109 shows a graph of results obtained with a linear additive model (Y axis) verse a GCRM for Chr18.

FIG. 110 and FIG. 111 show a graph of results obtained with a linear additive model (Y axis) verse a GCRM for Chr21.

Fig. 112A-C illustrates padding of a normalized autosomal profile for a euploid WI sample. FIG. 112A is an example of an unpadded profile. FIG. 112B is an example of a padded profile. FIG. 112C is an example of a padding correction (e.g., an adjusted profile, an adjusted elevation).

Fig. 113A-C illustrates padding of a normalized autosomal profile for a euploid WI sample. FIG. 113A is an example of an unpadded profile. FIG. 113B is an example of a padded profile. FIG. 113C is an example of a padding correction (e.g., an adjusted profile, an adjusted elevation).

Fig. 114A-C illustrates padding of a normalized autosomal profile for a trisomy 13 WI sample. FIG. 114A is an example of an unpadded profile. FIG. 114B is an example of a padded profile. FIG. 114C is an example of a padding correction (e.g., an adjusted profile, an adjusted elevation).

Fig. 115A-C illustrates padding of a normalized autosomal profile for a trisomy 18 WI sample. FIG. 115A is an example of an unpadded profile. FIG. 115B is an example of a padded profile. FIG. 115C is an example of a padding correction (e.g., an adjusted profile, an adjusted elevation).

FIG. 116-120, 122, 123, 126, 128, 129 and 131 show a maternal duplication within a profile.

FIG. 121, 124, 125, 127 and 130 show a maternal deletion within a profile.

FIG. 132 shows a linear relationship between (i) fetal fraction (%) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome according to equation AB (Y axis) and (ii) an experimental X chromosome representation (e.g., an MCR for ChrX) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome (X axis). The relationship is described according to equation AD.

FIG. 133 shows a linear relationship between (i) fetal fraction (%) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome according to equation AB (Y axis) and (ii) a fetal fraction determined from an MCR for ChrX as per equation AC (X axis).

FIG. 134 shows a linear relationship between (i) fetal fraction (%) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome

according to equation AB (Y axis) and (ii) an experimental Y chromosome representation (e.g., an MCR for ChrY) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome (X axis). The relationship is described according to equation AE.

FIG. 135 shows a linear relationship between (i) fetal fraction (%) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome according to equation AB (Y axis) and (ii) a fetal fraction determined from an MCR for ChrY (X axis).

FIG. 136 shows a linear relationship between (i) fetal fraction (%) determined from a pregnant female bearing a male fetus predicted from chromosome Y (Y axis) and (ii) fetal fraction (%) determined from a pregnant female bearing a male fetus predicted from chromosome X (X axis). The figure shows results obtained on euploid male pregnancies.

FIG. 137 shows a linear relationship between (i) fetal fraction (%) determined from a pregnant female bearing a male fetus comprising a trisomy 21, trisomy 18 or trisomy 13 chromosome according to equation AB (Y axis) and (ii) fetal fraction (%) determined for a pregnant female bearing a male fetus according to equation AF (x axis) using both ChrX and ChrY measured representations.

FIG. 138 shows the design of the recombinant MBD-Fc protein used to separate differentially methylated DNA.

FIG. 139 shows the methyl-CpG-binding, antibody-like protein has a high affinity and high avidity to its "antigen", which is preferably DNA that is methylated at CpG di-nucleotides.

FIG. 140 shows the methyl binding domain of MBD-FC binds all DNA molecules regardless of their methylation status. The strength of this protein/DNA interaction is defined by the level of DNA methylation. After binding genomic DNA, eluate solutions of increasing salt concentrations can be used to fractionate non-methylated and methylated DNA allowing for a controlled separation.

FIG. 141 shows the experiment used to identify differentially methylated DNA from a fetus and mother using the recombinant MBD-Fc protein and a microarray.

FIG. 142 shows typical results generated by Sequenom® EpiTYPER™ method, which was used to validate the results generated from the experiment illustrated in Figure 141.

FIG. 143 shows the correlation between the log ratios derived from microarray analysis (x axis) and methylation differences obtained by EpiTYPER™ analysis (y axis). Each data point represents the average for one region across all measured samples. The microarray analysis is comparative in nature because the highly methylated fraction of the maternal DNA is hybridized together with the highly methylated fraction of placenta DNA. Positive values indicate higher methylation of the placenta samples. In mass spectrometry each samples is measured individually. The difference in methylation was calculated by subtracting the maternal methylation values from the placenta methylation value. To compare the results with the microarray data the average of the differences for all maternal / placenta DNA pairs was calculated. FIG. 143 discloses SEQ ID NOS 374-375, respectively, in order of appearance.

FIG. 144 shows a correlation between microarray and EpiTYPER™ results.

FIG. 145 shows the correlation between the number of gDNA molecules that were expected and the number of molecules measured by competitive PCR in combination with mass spectrometry analysis. In this experiment, DNA derived from whole blood (black plus signs) and commercially available fully methylated DNA (grey crosses) were used in a 90 to 10 ratio. The MBD-FC fusion protein was used to separate the non-methylated and the methylated fraction of DNA. Each fraction was subject to competitive PCR analysis with mass spectrometry readout. The method has been described earlier for the analysis of copy number variations and is commercially available for gene expression analysis. The approach allows absolute quantification of DNA molecules with the help of a synthetic oligonucleotides of know concentration. In this experiment the MGMT locus was targeted, which was not methylated in the whole blood sample used here. Using an input of 300 total gDNA copies, 270 copies of non-methylated DNA and 30 copies of methylated DNA was expected. The measured copy numbers are largely in agreement with the expected values. The data point at 600 copies of input DNA indicates a bias in the reaction. This initial data indicates the feasibility of the approach for capturing and quantifying a few copies of methylated DNA in the presence of an excess of unmethylated DNA species.

FIG. 146A-146L show bar graph plots of the methylation differences obtained from the microarray analysis (dark bars) and the mass spectrometry analysis (light grey bars) with respect to their genomic location. For each of the 85 regions that were identified to be differentially methylated by microarray an individual plot is provided. The x axis for each plot shows the chromosomal position of the region. The y axis depicts the log ration (in case of the microarrays) and the methylation differences (in case of the mass spectrometry results). For the microarrays each hybridization probe in the area is shown as a single black (or dark grey) bar. For the mass spectrometry results each CpG site, is shown as a light grey bar. Bars showing values greater than zero indicate higher DNA methylation in the placenta samples compared to the maternal DNA. For some genes the differences are small (i.e. RB1 or DSCR6) but still statistically significant. Those regions would be less suitable for a fetal DNA enrichment strategy.

FIG. 147 shows one embodiment of the Fetal Quantifier Method. Maternal nucleic acid is selectively digested and the remaining fetal nucleic acid is quantified using a competitor of known concentration. In this schema, the analyte is separated and quantified by a mass spectrometer.

FIG. 148 shows one embodiment of the Methylation-Based Fetal Diagnostic Method. Maternal nucleic acid is selectively digested and the remaining fetal nucleic acid is quantified for three different chromosomes (13, 18 and 21). Parts 2 and 3 of the Figure illustrate the size distribution of the nucleic acid in the sample before and after digestion. The amplification reactions can be size-specific (e.g., greater than 100 base pair amplicons) such that they favor the longer, non-digested fetal nucleic acid over the digested maternal nucleic acid, thereby further enriching the fetal nucleic acid. The spectra at the bottom of the Figure show an increased amount of chromosome 21 fetal nucleic acid indicative of trisomy 21.

FIG. 149 shows the total number of amplifiable genomic copies from four different DNA samples isolated from the blood of non-pregnant women. Each sample was diluted to contain approximately 2500, 1250, 625 or 313 copies per reaction. Each measurement was obtained by taking the mean DNA/competitor ratio obtained from two total copy number assays (ALB and RNAseP in Table X). As shown, the total copy number is accurate and stable across the different samples, thus validating the usefulness of the competitor-based approach.

FIGS. 150A-150B show a model system that was created that contained a constant number of maternal non-methylated DNA with varying amounts of male placental methylated DNA spiked-in. The samples were spiked with male placental amounts ranging from approximately 0 to 25% relative to the maternal non-methylated DNA. The fraction of placental DNA was calculated using the ratios obtained from the methylation assays (Figure 150A) and the Y-chromosome marker (Figure 150B) as compared to the total copy number assay. The methylation and Y-chromosome markers are provided in Table X.

FIGS. 151A and 151B show the results of the total copy number assay from plasma samples. In Figure 151A, the copy number for each sample is shown. Two samples (no 25 and 26) have a significantly higher total copy number than all the other samples. A mean of approximately 1300 amplifiable copies/ml plasma was obtained (range 766-2055). Figure 151B shows a box-and-whisker plot of the given values, summarizing the results.

FIGS. 152A and 152B show the amount (or copy numbers) of fetal nucleic acid from 33 different plasma samples taken from pregnant women with male fetuses plotted. The copy numbers obtained were calculated using the methylation markers and the Y-chromosome-specific markers using the assays provided in Table X. As can be seen in Figure 152B, the box-and-whisker plot of the given values indicated minimal difference between the two different measurements, thus validating the accuracy and stability of the method.

FIG. 153 shows a paired correlation between the results obtained using the methylation markers versus the Y-chromosome marker from Figure 152A.

FIG. 154 shows the digestion efficiency of the restriction enzymes using the ratio of digestion for the control versus the competitor and comparing this value to the mean total copy number assays. Apart from sample 26 all reactions indicate the efficiency to be above about 99%.

FIG. 155 provides a specific method for calculating fetal DNA fraction (or concentration) in a sample using the Y-chromosome-specific markers for male pregnancies and the mean of the methylated fraction for all pregnancies (regardless of fetal sex).

FIG. 156 provides a specific method for calculating fetal DNA fraction (or concentration) in a sample without the Y-chromosome-specific markers. Instead, only the Assays for Methylation Quantification were used to determine the concentration of fetal DNA.

FIG. 157 shows a power calculation t-test for a simulated trisomy 21 diagnosis using the methods of the technology herein. The Figure shows the relationship between the coefficient of variation (CV) on the x-axis and the power to discriminate the assay populations using a simple t-test (y-axis). The data indicates that in 99% of all cases, one can discriminate the two population (euploid vs. aneuploid) on a significance level of 0.001 provided a CV of 5% or less.

FIG. 158 shows a scheme for ligating a PCR amplicon with Illumina sequencing adaptors.

FIG. 159 shows a modified ligation scheme.

FIG. 160 shows a comparison of copy numbers of individual markers determined by a fetal quantification assay using MPSS (FQA Sequencing; x-axis) with those obtained by a fetal quantification assay using MASSARRAY (FQA MA; y-axis). The results from both methods were highly correlated ($R^2 > 0.97$). In some cases, platform-specific allele bias resulted in slight copy number differences and slopes of the linear fit which deviated from 1.

FIG. 161 shows a comparison of mean copy numbers for each of the marker groups determined by a fetal quantification assay using MPSS (FQA Sequencing; x-axis) with those obtained by a fetal quantification assay using MASSARRAY (FQA MA; y-axis).

FIG. 162 shows a comparison of fetal fractions derived from either methylation (left) or Y-chromosome markers determined by a fetal quantification assay using MPSS (FQA Sequencing; x-axis) with those obtained by a fetal quantification assay using MASSARRAY (FQA MA; y-axis).

FIG. 163 shows an example of a likelihood chart for an informative fetal/maternal genotype combination.

FIG. 164 shows an embodiment of a possible distribution of maternal and paternal alleles. In such an embodiment a fetus with a homozygous mother (GG) and heterozygous father (GA) has a 50%

probability of having genotype GA. In some embodiments the paternal allele A is only observed in the fetal component of ccf DNA from plasma of a pregnant mother.

FIG. 165 shows one embodiment for calculating fetal fraction by MPSS. In this example ccf DNA from plasma of a pregnant mother with a 10% fetal fraction is targeted for 67 SNPs and sequenced at a coverage of 2000x. For the SNP shown, the mother has a genotype of GG and the fetus has a genotype of GA. In some embodiments the maternal allele G would be expected to be sequenced at a coverage of 1800x. In some embodiments the fetal alleles, G (derived from the mother) and A (derived from the father), would be expected to be sequenced at a coverage of 100x. In some embodiments the fetal fraction is calculated by calculating the fraction of paternal coverage to total coverage and multiplying by 2.

FIG. 166 illustrates a scheme for multiplexed amplicon library generation and sequencing. Loci-specific PCR amplifies targeted sequences each containing a single SNP. The Loci-specific PCR simultaneously incorporates a tag that can be used as a template for primers in a subsequent universal PCR. The universal PCR incorporates full length flowcell capture sequences and an index sequence to each amplicon.

FIG. 167 shows allele frequencies per SNP for multiple samples. Each data-point on the X-axis corresponds to a targeted SNP while the Y-axis shows the measured allele frequency for each SNP. The plot is grouped into allele frequencies measured for buffy coat (maternal genotypes only) and the paired pregnant plasma DNA (maternal and fetal genotypes). Dotted lines show allele frequencies of 0.01 and 0.99 which distinguish homozygote allele frequencies from putative Type 1 allele frequencies. Type 1 informative genotypes are expected to have allele frequencies between 0.01 - 0.25 or between 0.75 – 0.99. For example, SNP #12 is <0.01 in buffy coat sample but >0.01 in plasma DNA indicating that SNP #12 is an informative genotype.

FIG. 168 shows allele frequencies per SNP for multiple samples. Each data-point on the X-axis corresponds to a targeted SNP while the Y-axis shows the measured allele frequency for each SNP. The plot is grouped into allele frequencies measured for buffy coat (maternal genotypes only) and the paired pregnant plasma DNA (maternal and fetal genotypes). Dotted lines show allele frequencies of 0.01 and 0.99 which distinguish homozygote allele frequencies from putative Type 1 allele frequencies. Type 1 informative genotypes are expected to have allele frequencies

between 0.01 - 0.25 or between 0.75 – 0.99. For example, SNP #1 is >0.99 in buffy coat sample but <0.99 in plasma DNA suggesting that SNP #1 is an informative genotype.

FIG. 169 shows allele frequencies per sample for a collection of 46 samples. Each data-point on the X-axis corresponds to a pregnant plasma DNA sample while the Y-axis shows the measured allele frequencies for 67 targeted SNPs. Dotted lines show allele frequencies of 0.01, 0.25, 0.75 and 0.99. Type 1 informative genotypes (marker: ○) were defined to have allele frequencies between 0.01 - 0.25 or between 0.75 – 0.99. Non-informative homozygotes (marker: X) have allele frequencies <0.01 or >0.99. Non-informative heterozygotes (marker: +) have allele frequencies between 0.25 - 0.75.

FIG. 170 shows allele frequencies per sample (folded on 0.5, i.e., if the allele frequency is greater than 0.5, subtract the allele frequency value from 1) for a collection of 46 samples. Each data-point on the X-axis corresponds to a pregnant plasma DNA sample while the Y-axis shows the measured allele frequencies for 67 targeted SNPs folded on 0.5. Dotted line shows allele frequency of 0.01. Type 1 informative genotypes (marker: 0) have allele frequencies between 0.01 - 0.25. Non-informative homozygotes (marker: X) have allele frequencies <0.01. Non-informative heterozygotes (marker: +) have allele frequencies between 0.25 - 0.5.

FIG. 171 shows fetal fraction values calculated from informative genotypes for 46 samples. Each data-point on the X-axis corresponds to a pregnant plasma DNA sample while the Y-axis shows the measured fetal fraction. The error bars correspond to the standard deviation of the fetal fraction calculated by individual informative SNPs. The fetal fractions are ordered from low fetal fraction to high.

Fig. 172 shows a correlation plot for SNP-based fetal fraction estimates versus methylation-based fetal fraction estimates. Each data-point on the X-axis corresponds to the fetal fraction measured from 67 targeted SNPs. Each data-point on the Y-axis corresponds to the fetal fraction from a methylation-based estimate. There are 46 data-points. The line shows the linear fit of the correlation between the SNP-based estimate of the fetal fraction and the methylation-based estimate ($R^2 = 0.72$).

FIG. 173 shows a comparison of informative genotype measurements at varying sequencing coverage. Each group on the X-axis corresponds to a sequenced sample with mean amplicon

coverage of 71619x, 8557x and 1413x, respectively. Each data-point on the Y-axis corresponds to the allele frequency (folded on 0.5) of informative genotypes assigned from 67 targeted SNPs. Diamond plots show mean folded allele frequencies of informative genotypes ± 1 standard deviation. Each plot shows the result from 1 of 3 samples.

FIG. 174 shows probabilities of the number of informative SNPs for each of the selected thresholds (1-6 informative SNPs) at increasing numbers of total SNPs assayed. Each data-point on the X-axis corresponds to the total number of SNPs assayed. Each data-point on the y-axis corresponds to the probability of detecting N number of informative SNPs. The six curves from left to right show the probability of detecting 1 – 6 informative genotypes, respectively. Probabilities assume a minor allele frequency of 0.4 for each SNP in the population sampled.

FIG. 175 shows fitted ploidy (y axis) for chromosome 21 derived from PERUN profiles of chr21 in male LDTv2CE pregnancies. The input fetal fraction (x axis) was derived from PERUN chromosome Y profiles. For Figures 175 – 177, euploids are represented as circles (e.g., o); T21 as up triangles (e.g., ▲); T18 as down triangles (e.g., ▼); T13 as squares (e.g., ■); and complex genetic variations as an x or + symbol. QC failures are represented as hollow (open) symbols or an x symbol, and those passing QC as filled symbols or a + symbol.

FIG. 176 shows fitted ploidy (y axis) for chromosome 18 derived from PERUN profiles of chr18 in male LDTv2CE pregnancies. The input fetal fraction (x axis) was derived from PERUN chromosome Y profiles.

FIG. 177 show fitted ploidy (y axis) for chromosome 13 derived from PERUN profiles of chr21 in male LDTv2CE pregnancies. The input fetal fraction (x axis) was derived from PERUN chromosome Y profiles.

FIGS. 178A-F show six examples of a determination of the presence of a trisomy of chromosome 16 (i.e., T16, Trisomy 16, circled). FIG. 190B show an example of an absence of (e.g., a determination of the absence of) a chromosome aneuploidy (e.g., a chromosome duplication, deletion, microduplication, microdeletion or trisomy) for chromosomes 1 through 15.

FIGS. 179A-C show three examples of a determination of the presence of a trisomy of chromosome 20 (i.e., T20, Trisomy 20, circled).

FIG. 180 shows an example of a determination of the presence of a trisomy of chromosome 22 (i.e., T22, Trisomy 22, circled).

FIG. 181A-C shows an example of a determination of the presence of a double aneuploidy comprising a trisomy of chromosome 4 and chromosome 21 (circled regions). FIGS. 181A-C represent three independent determinations for the same patient.

FIG. 182 shows an example of a determination of the presence of a double aneuploidy comprising a trisomy of chromosome 9 and chromosome 21 (circled regions).

FIG. 183A-B shows an example of a determination of the presence of a trisomy of chromosome 16 (i.e., T16, Trisomy 16, circled). FIGS. 183A-B represent two independent determinations for the same patient.

FIG. 184A-C show examples of a determination of the presence of a microdeletion in chromosome X.

FIG. 185 shows an example of a determination of the presence of a microdeletion in chromosome 7 (shown within circled region). The deletion extends from chr7:92,350,001 to chr7:113,050,000 (length: 20.7 Mbp).

FIG. 186A-B shows an example of a determination of the presence of a microduplication in chromosome 18 (shown within circled region). FIG. 186A shows a whole genome plot. FIG. 186B shows an enlarged view of chromosome 18 as seen in FIG. 186A.

FIG. 187 shows an example of a determination of the presence of a microduplication in chromosome 3 (shown within circled region).

FIG. 188 shows an example of a determination of the presence of a microduplication in chromosome 4 (shown within circled region).

FIG. 189A-B shows an example of a determination of the presence of a complex genetic variation using methods described herein (e.g., PERUN). FIG. 189A and 189B show the results of two

separate independent measurements of the same sample. A duplication in chromosome 5 (the first 15 Mbp), a duplication in chromosome 8 (starting at chr8:119,750,000), a duplication in chromosome 11 and 14, a deletion in chromosome 14 and a trisomy of chromosome X are shown in the profile within the circled regions.

FIG. 190 shows an illustrative embodiment of a system in which certain embodiments of the technology may be implemented.

Definitions

The term "pregnancy-associated disorder," as used in this application, refers to any condition or disease that may affect a pregnant woman, the fetus, or both the woman and the fetus. Such a condition or disease may manifest its symptoms during a limited time period, e.g., during pregnancy or delivery, or may last the entire life span of the fetus following its birth. Some examples of a pregnancy-associated disorder include ectopic pregnancy, preeclampsia, preterm labor, RhD incompatibility, fetal chromosomal abnormalities such as trisomy 21, and genetically inherited fetal disorders such as cystic fibrosis, beta-thalassemia or other monogenic disorders. The compositions and processes described herein are particularly useful for diagnosis, prognosis and monitoring of pregnancy-associated disorders associated with quantitative abnormalities of fetal DNA in maternal plasma/serum, including but not limited to, preeclampsia (Lo et al., Clin. Chem. 45:184-188, 1999 and Zhong et al., Am. J. Obstet. Gynecol. 184:414-419, 2001), fetal trisomy (Lo et al., Clin. Chem. 45:1747-1751, 1999 and Zhong et al., Prenat. Diagn. 20:795-798, 2000) and hyperemesis gravidarum (Sekizawa et al., Clin. Chem. 47:2164-2165, 2001). For example, an elevated level of fetal nucleic acid in maternal blood (as compared to a normal pregnancy or pregnancies) may be indicative of a preeclamptic pregnancy. Further, the ability to enrich fetal nucleic from a maternal sample may prove particularly useful for the noninvasive prenatal diagnosis of autosomal recessive diseases such as the case when a mother and father share an identical disease causing mutation, an occurrence previously perceived as a challenge for maternal plasma-based non-trisomy prenatal diagnosis.

The terms “nucleic acid” and “nucleic acid molecule” may be used interchangeably throughout the disclosure. The terms refer to nucleic acids of any composition from, such as DNA (e.g., complementary DNA (cDNA), genomic DNA (gDNA) and the like), RNA (e.g., message RNA (mRNA), short inhibitory RNA (siRNA), ribosomal RNA (rRNA), tRNA, microRNA, RNA highly expressed by the fetus or placenta, and the like), and/or DNA or RNA analogs (e.g., containing

base analogs, sugar analogs and/or a non-native backbone and the like), RNA/DNA hybrids and polyamide nucleic acids (PNAs), all of which can be in single- or double-stranded form, and unless otherwise limited, can encompass known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. For example, the nucleic acids provided in SEQ ID NOs: 1-261 (see Tables 4A-4C) can be in any form useful for conducting processes herein (e.g., linear, circular, supercoiled, single-stranded, double-stranded and the like) or may include variations (e.g., insertions, deletions or substitutions) that do not alter their utility as part of the present technology. A nucleic acid may be, or may be from, a plasmid, phage, autonomously replicating sequence (ARS), centromere, artificial chromosome, chromosome, or other nucleic acid able to replicate or be replicated in vitro or in a host cell, a cell, a cell nucleus or cytoplasm of a cell in certain embodiments. A template nucleic acid in some embodiments can be from a single chromosome (e.g., a nucleic acid sample may be from one chromosome of a sample obtained from a diploid organism). Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions), alleles, orthologs, single nucleotide polymorphisms (SNPs), and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); and Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)). The term nucleic acid is used interchangeably with locus, gene, cDNA, and mRNA encoded by a gene. The term also may include, as equivalents, derivatives, variants and analogs of RNA or DNA synthesized from nucleotide analogs, single-stranded ("sense" or "antisense", "plus" strand or "minus" strand, "forward" reading frame or "reverse" reading frame) and double-stranded polynucleotides. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. For RNA, the base cytosine is replaced with uracil. A template nucleic acid may be prepared using a nucleic acid obtained from a subject as a template.

A “nucleic acid comprising one or more CpG sites” or a "CpG-containing genomic sequence" as used herein refers to a segment of DNA sequence at a defined location in the genome of an individual such as a human fetus or a pregnant woman. Typically, a "CpG-containing genomic sequence" is at least 15 nucleotides in length and contains at least one cytosine. Preferably, it can

be at least 30, 50, 80, 100, 150, 200, 250, or 300 nucleotides in length and contains at least 2, 5, 10, 15, 20, 25, or 30 cytosines. For anyone "CpG-containing genomic sequence" at a given location, e.g., within a region centering around a given genetic locus (see Tables 1A-1C), nucleotide sequence variations may exist from individual to individual and from allele to allele even for the same individual. Typically, such a region centering around a defined genetic locus (e.g., a CpG island) contains the locus as well as upstream and/or downstream sequences. Each of the upstream or downstream sequence (counting from the 5' or 3' boundary of the genetic locus, respectively) can be as long as 10 kb, in other cases may be as long as 5 kb, 2 kb, 1 kb, 500 bp, 200 bp, or 100 bp. Furthermore, a "CpG-containing genomic sequence" may encompass a nucleotide sequence transcribed or not transcribed for protein production, and the nucleotide sequence can be an inter-gene sequence, intra-gene sequence, protein-coding sequence, a non protein-coding sequence (such as a transcription promoter), or a combination thereof.

As used herein, a “methylated nucleotide” or a "methylated nucleotide base" refers to the presence of a methyl moiety on a nucleotide base, where the methyl moiety is not present in a recognized typical nucleotide base. For example, cytosine does not contain a methyl moiety on its pyrimidine ring, but 5-methylcytosine contains a methyl moiety at position 5 of its pyrimidine ring. Therefore, cytosine is not a methylated nucleotide and 5-methylcytosine is a methylated nucleotide. In another example, thymine contains a methyl moiety at position 5 of its pyrimidine ring, however, for purposes herein, thymine is not considered a methylated nucleotide when present in DNA since thymine is a typical nucleotide base of DNA. Typical nucleoside bases for DNA are thymine, adenine, cytosine and guanine. Typical bases for RNA are uracil, adenine, cytosine and guanine. Correspondingly a "methylation site" is the location in the target gene nucleic acid region where methylation has, or has the possibility of occurring. For example a location containing CpG is a methylation site where the cytosine may or may not be methylated.

As used herein, a “CpG site” or "methylation site" is a nucleotide within a nucleic acid that is susceptible to methylation either by natural occurring events in vivo or by an event instituted to chemically methylate the nucleotide in vitro.

As used herein, a “methylated nucleic acid molecule” refers to a nucleic acid molecule that contains one or more methylated nucleotides that is/are methylated.

A "CpG island" as used herein describes a segment of DNA sequence that comprises a functionally or structurally deviated CpG density. For example, Yamada et al. (Genome Research 14:247-266, 2004) have described a set of standards for determining a CpG island: it must be at least 400 nucleotides in length, has a greater than 50% GC content, and an OCF/ECF ratio greater than 0.6. Others (Takai et al., Proc. Natl. Acad. Sci. U.S.A. 99:3740-3745, 2002) have defined a CpG island less stringently as a sequence at least 200 nucleotides in length, having a greater than 50% GC content, and an OCF/ECF ratio greater than 0.6.

The term "epigenetic state" or "epigenetic status" as used herein refers to any structural feature at a molecular level of a nucleic acid (e.g., DNA or RNA) other than the primary nucleotide sequence. For instance, the epigenetic state of a genomic DNA may include its secondary or tertiary structure determined or influenced by, e.g., its methylation pattern or its association with cellular proteins.

The term "methylation profile" "methylation state" or "methylation status," as used herein to describe the state of methylation of a genomic sequence, refers to the characteristics of a DNA segment at a particular genomic locus relevant to methylation. Such characteristics include, but are not limited to, whether any of the cytosine (C) residues within this DNA sequence are methylated, location of methylated C residue(s), percentage of methylated C at any particular stretch of residues, and allelic differences in methylation due to, e.g., difference in the origin of the alleles. The term "methylation" profile" or "methylation status" also refers to the relative or absolute concentration of methylated C or unmethylated C at any particular stretch of residues in a biological sample. For example, if the cytosine (C) residue(s) within a DNA sequence are methylated it may be referred to as "hypermethylated"; whereas if the cytosine (C) residue(s) within a DNA sequence are not methylated it may be referred to as "hypomethylated". Likewise, if the cytosine (C) residue(s) within a DNA sequence (e.g., fetal nucleic acid) are methylated as compared to another sequence from a different region or from a different individual (e.g., relative to maternal nucleic acid), that sequence is considered hypermethylated compared to the other sequence. Alternatively, if the cytosine (C) residue(s) within a DNA sequence are not methylated as compared to another sequence from a different region or from a different individual (e.g., the mother), that sequence is considered hypomethylated compared to the other sequence. These sequences are said to be "differentially methylated", and more specifically, when the methylation status differs between mother and fetus, the sequences are considered "differentially methylated maternal and fetal nucleic acid".

The term “agent that binds to methylated nucleotides” as used herein refers to a substance that is capable of binding to methylated nucleic acid. The agent may be naturally-occurring or synthetic, and may be modified or unmodified. In one embodiment, the agent allows for the separation of different nucleic acid species according to their respective methylation states. An example of an agent that binds to methylated nucleotides is described in PCT Patent Application No. PCT/EP2005/012707, which published as WO06056480A2 and is hereby incorporated by reference. The described agent is a bifunctional polypeptide comprising the DNA-binding domain of a protein belonging to the family of Methyl-CpG binding proteins (MBDs) and an Fc portion of an antibody (see Fig. 138). The recombinant methyl-CpG-binding, antibody-like protein can preferably bind CpG methylated DNA in an antibody-like manner. That means, the methyl-CpG-binding, antibody-like protein has a high affinity and high avidity to its "antigen", which is preferably DNA that is methylated at CpG dinucleotides. The agent may also be a multivalent MBD (see Figure 139).

The term “polymorphism” or “polymorphic nucleic acid target” as used herein refers to a sequence variation within different alleles of the same genomic sequence. A sequence that contains a polymorphism is considered a “polymorphic sequence”. Detection of one or more polymorphisms allows differentiation of different alleles of a single genomic sequence or between two or more individuals. As used herein, the term "polymorphic marker" or “polymorphic sequence” refers to segments of genomic DNA that exhibit heritable variation in a DNA sequence between individuals. Such markers include, but are not limited to, single nucleotide polymorphisms (SNPs), restriction fragment length polymorphisms (RFLPs), short tandem repeats, such as di-, tri- or tetra-nucleotide repeats (STRs), deletions, duplications, and the like. Polymorphic markers according to the present technology can be used to specifically differentiate between a maternal and paternal allele in the enriched fetal nucleic acid sample.

The terms "single nucleotide polymorphism" or "SNP" as used herein refer to the polynucleotide sequence variation present at a single nucleotide residue within different alleles of the same genomic sequence. This variation may occur within the coding region or non-coding region (i.e., in the promoter or intronic region) of a genomic sequence, if the genomic sequence is transcribed during protein production. Detection of one or more SNP allows differentiation of different alleles of a single genomic sequence or between two or more individuals.

The term "allele" as used herein is one of several alternate forms of a gene or non-coding regions of DNA that occupy the same position on a chromosome. The term allele can be used to describe DNA from any organism including but not limited to bacteria, viruses, fungi, protozoa, molds, yeasts, plants, humans, non-humans, animals, and archeabacteria.

The terms "ratio of the alleles" or “allelic ratio” as used herein refer to the ratio of the population of one allele and the population of the other allele in a sample. In some trisomic cases, it is possible that a fetus may be tri-allelic for a particular locus. In such cases, the term "ratio of the alleles" refers to the ratio of the population of any one allele against one of the other alleles, or any one allele against the other two alleles.

The term “non-polymorphism-based quantitative method” as used herein refers to a method for determining the amount of an analyte (e.g., total nucleic acid, Y-chromosome nucleic acid, or fetal nucleic acid) that does not require the use of a polymorphic marker or sequence. Although a polymorphism may be present in the sequence, said polymorphism is not required to quantify the sequence. Examples of non-polymorphism-based quantitative methods include, but are not limited to, RT-PCR, digital PCR, array-based methods, sequencing methods, nanopore-based methods, nucleic acid-bound bead-based counting methods and competitor-based methods where one or more competitors are introduced at a known concentration(s) to determine the amount of one or more analytes. In some embodiments, some of the above exemplary methods (for example, sequencing) may need to be actively modified or designed such that one or more polymorphisms are not interrogated.

As used herein, a “competitor oligonucleotide” or “competitive oligonucleotide” or “competitor” is a nucleic acid polymer that competes with a target nucleotide sequence for hybridization of amplification primers. Often, a competitor has a similar nucleotide sequence as a corresponding target nucleotide sequence. In some cases, a competitor sequence and a corresponding target nucleotide sequence differ by one or more nucleotides. In some cases, a competitor sequence and a corresponding target nucleotide sequence are the same length. In some cases, the competitor optionally has an additional length of nucleotide sequence that is different from the target nucleotide sequence. In some embodiments, a known amount, or copy number, of competitor is used. In some embodiments, two or more competitors are used. In some cases, the two or more competitors possess similar characteristics (e.g., sequence, length, detectable label). In some cases, the two or more competitors possess different characteristics (e.g., sequence,

length, detectable label). In some embodiments, one or more competitors are used for a particular region. In some cases, the competitor possesses a characteristic that is unique for each set of competitors for a given region. Often, competitors for different regions possess different characteristics.

A competitor oligonucleotide may be composed of naturally occurring and/or non-naturally occurring nucleotides (e.g., labeled nucleotides), or a mixture thereof. Competitor oligonucleotides suitable for use with embodiments described herein, may be synthesized and labeled using known techniques. Competitor oligonucleotides may be chemically synthesized according to any suitable method known, for example, the solid phase phosphoramidite triester method first described by Beaucage and Caruthers, Tetrahedron Letts., 22:1859-1862, 1981, using an automated synthesizer, as described in Needham-VanDevanter et al., Nucleic Acids Res. 12:6159-6168, 1984. Purification of competitor oligonucleotides can be effected by any suitable method known, for example, native acrylamide gel electrophoresis or by anion-exchange high-performance liquid chromatography (HPLC), for example, as described in Pearson and Regnier, J. Chrom., 255:137-149, 1983.

The terms “absolute amount” or “copy number” as used herein refers to the amount or quantity of an analyte (e.g., total nucleic acid or fetal nucleic acid). The present technology provides compositions and processes for determining the absolute amount of fetal nucleic acid in a mixed maternal sample. Absolute amount or copy number represents the number of molecules available for detection, and may be expressed as the genomic equivalents per unit. The term “concentration” refers to the amount or proportion of a substance in a mixture or solution (e.g., the amount of fetal nucleic acid in a maternal sample that comprises a mixture of maternal and fetal nucleic acid). The concentration may be expressed as a percentage, which is used to express how large/small one quantity is, relative to another quantity as a fraction of 100. Platforms for determining the quantity or amount of an analyte (e.g., target nucleic acid) include, but are not limited to, mass spectrometry, digital PCR, sequencing by synthesis platforms (e.g., pyrosequencing), fluorescence spectroscopy and flow cytometry.

The term “sample” as used herein refers to a specimen containing nucleic acid. Examples of samples include, but are not limited to, tissue, bodily fluid (for example, blood, serum, plasma, saliva, urine, tears, peritoneal fluid, ascitic fluid, vaginal secretion, breast fluid, breast milk, lymph fluid, cerebrospinal fluid or mucosa secretion), umbilical cord blood, chorionic villi, amniotic fluid, an

embryo, a two-celled embryo, a four-celled embryo, an eight-celled embryo, a 16-celled embryo, a 32-celled embryo, a 64-celled embryo, a 128-celled embryo, a 256-celled embryo, a 512-celled embryo, a 1024-celled embryo, embryonic tissues, lymph fluid, cerebrospinal fluid, mucosa secretion, or other body exudate, fecal matter, an individual cell or extract of the such sources that contain the nucleic acid of the same, and subcellular structures such as mitochondria, using protocols well established within the art.

Fetal DNA can be obtained from sources including but not limited to maternal blood, maternal serum, maternal plasma, fetal cells, umbilical cord blood, chorionic villi, amniotic fluid, urine, saliva, lung lavage, cells or tissues.

The term "blood" as used herein refers to a blood sample or preparation from a pregnant woman or a woman being tested for possible pregnancy. The term encompasses whole blood or any fractions of blood, such as serum and plasma as conventionally defined.

The term "bisulfite" as used herein encompasses all types of bisulfites, such as sodium bisulfite, that are capable of chemically converting a cytosine (C) to a uracil (U) without chemically modifying a methylated cytosine and therefore can be used to differentially modify a DNA sequence based on the methylation status of the DNA.

As used herein, a reagent or agent that "differentially modifies" methylated or non-methylated DNA encompasses any reagent that modifies methylated and/or unmethylated DNA in a process through which distinguishable products result from methylated and non-methylated DNA, thereby allowing the identification of the DNA methylation status. Such processes may include, but are not limited to, chemical reactions (such as a C.fwdarw.U conversion by bisulfite) and enzymatic treatment (such as cleavage by a methylation-dependent endonuclease). Thus, an enzyme that preferentially cleaves or digests methylated DNA is one capable of cleaving or digesting a DNA molecule at a much higher efficiency when the DNA is methylated, whereas an enzyme that preferentially cleaves or digests unmethylated DNA exhibits a significantly higher efficiency when the DNA is not methylated.

The terms “non-bisulfite-based method” and “non-bisulfite-based quantitative method” as used herein refer to any method for quantifying methylated or non-methylated nucleic acid that does not require the use of bisulfite. The terms also refer to methods for preparing a nucleic acid to be

quantified that do not require bisulfite treatment. Examples of non-bisulfite-based methods include, but are not limited to, methods for digesting nucleic acid using one or more methylation sensitive enzymes and methods for separating nucleic acid using agents that bind nucleic acid based on methylation status.

The terms "methyl-sensitive enzymes" and "methylation sensitive restriction enzymes" are DNA restriction endonucleases that are dependent on the methylation state of their DNA recognition site for activity. For example, there are methyl-sensitive enzymes that cleave or digest at their DNA recognition sequence only if it is not methylated. Thus, an unmethylated DNA sample will be cut into smaller fragments than a methylated DNA sample. Similarly, a hypermethylated DNA sample will not be cleaved. In contrast, there are methyl-sensitive enzymes that cleave at their DNA recognition sequence only if it is methylated. As used herein, the terms "cleave", "cut" and "digest" are used interchangeably.

The term "target nucleic acid" as used herein refers to a nucleic acid examined using the methods disclosed herein to determine if the nucleic acid is part of a pregnancy-related disorder or chromosomal abnormality. For example, a target nucleic acid from chromosome 21 could be examined using the methods of the technology herein to detect Down's Syndrome.

The term "control nucleic acid" as used herein refers to a nucleic acid used as a reference nucleic acid according to the methods disclosed herein to determine if the nucleic acid is part of a chromosomal abnormality. For example, a control nucleic acid from a chromosome other than chromosome 21 (herein referred to as a "reference chromosome") could be as a reference sequence to detect Down's Syndrome. In some embodiments, the control sequence has a known or predetermined quantity.

The term "sequence-specific" or "locus-specific method" as used herein refers to a method that interrogates (for example, quantifies) nucleic acid at a specific location (or locus) in the genome based on the sequence composition. Sequence-specific or locus-specific methods allow for the quantification of specific regions or chromosomes.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) involved in the

transcription/translation of the gene product and the regulation of the transcription/translation, as well as intervening sequences (introns) between individual coding segments (exons).

In this application, the terms "polypeptide," "peptide," and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins (i.e., antigens), where the amino acid residues are linked by covalent peptide bonds.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, .gamma.-carboxyglutamate, and O-phosphoserine.

Amino acids may be referred to herein by either the commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Primers" as used herein refer to oligonucleotides that can be used in an amplification method, such as a polymerase chain reaction (PCR), to amplify a nucleotide sequence based on the polynucleotide sequence corresponding to a particular genomic sequence, e.g., one located within the CpG island CGI137, PDE9A, or CGI009 on chromosome 21, in various methylation status. At least one of the PCR primers for amplification of a polynucleotide sequence is sequence-specific for the sequence.

The term "template" refers to any nucleic acid molecule that can be used for amplification in the technology herein. RNA or DNA that is not naturally double stranded can be made into double stranded DNA so as to be used as template DNA. Any double stranded DNA or preparation containing multiple, different double stranded DNA molecules can be used as template DNA to amplify a locus or loci of interest contained in the template DNA.

The term “amplification reaction” as used herein refers to a process for copying nucleic acid one or more times. In embodiments, the method of amplification includes but is not limited to polymerase chain reaction, self-sustained sequence reaction, ligase chain reaction, rapid amplification of cDNA ends, polymerase chain reaction and ligase chain reaction, Q-beta phage amplification, strand displacement amplification, or splice overlap extension polymerase chain reaction. In some embodiments, a single molecule of nucleic acid is amplified, for example, by digital PCR.

The term “sensitivity” as used herein refers to the number of true positives divided by the number of true positives plus the number of false negatives, where sensitivity (sens) may be within the range of $0 \leq sens \leq 1$. Ideally, method embodiments herein have the number of false negatives equaling zero or close to equaling zero, so that no subject is wrongly identified as not having at least one chromosome abnormality or other genetic disorder when they indeed have at least one chromosome abnormality or other genetic disorder. Conversely, an assessment often is made of the ability of a prediction algorithm to classify negatives correctly, a complementary measurement to sensitivity. The term “specificity” as used herein refers to the number of true negatives divided by the number of true negatives plus the number of false positives, where sensitivity (spec) may be within the range of $0 \leq spec \leq 1$. Ideally, methods embodiments herein have the number of false positives equaling zero or close to equaling zero, so that no subject wrongly identified as having at least one chromosome abnormality other genetic disorder when they do not have the chromosome abnormality other genetic disorder being assessed. Hence, a method that has sensitivity and specificity equaling one, or 100%, sometimes is selected.

One or more prediction algorithms may be used to determine significance or give meaning to the detection data collected under variable conditions that may be weighed independently of or dependently on each other. The term “variable” as used herein refers to a factor, quantity, or function of an algorithm that has a value or set of values. For example, a variable may be the design of a set of amplified nucleic acid species, the number of sets of amplified nucleic acid species, percent fetal genetic contribution tested, percent maternal genetic contribution tested, type of chromosome abnormality assayed, type of genetic disorder assayed, type of sex-linked abnormalities assayed, the age of the mother and the like. The term “independent” as used herein refers to not being influenced or not being controlled by another. The term “dependent” as used herein refers to being influenced or controlled by another. For example, a particular chromosome and a trisomy event occurring for that particular chromosome that results in a viable being are variables that are dependent upon each other.

One of skill in the art may use any type of method or prediction algorithm to give significance to the data of the present technology within an acceptable sensitivity and/or specificity. For example, prediction algorithms such as Chi-squared test, z-test, t-test, ANOVA (analysis of variance), regression analysis, neural nets, fuzzy logic, Hidden Markov Models, multiple model state estimation, and the like may be used. One or more methods or prediction algorithms may be determined to give significance to the data having different independent and/or dependent variables of the present technology. And one or more methods or prediction algorithms may be determined not to give significance to the data having different independent and/or dependent variables of the present technology. One may design or change parameters of the different variables of methods described herein based on results of one or more prediction algorithms (e.g., number of sets analyzed, types of nucleotide species in each set). For example, applying the Chi-squared test to detection data may suggest that specific ranges of maternal age are correlated to a higher likelihood of having an offspring with a specific chromosome abnormality, hence the variable of maternal age may be weighed differently verses being weighed the same as other variables.

In certain embodiments, several algorithms may be chosen to be tested. These algorithms can be trained with raw data. For each new raw data sample, the trained algorithms will assign a classification to that sample (i.e. trisomy or normal). Based on the classifications of the new raw data samples, the trained algorithms' performance may be assessed based on sensitivity and specificity. Finally, an algorithm with the highest sensitivity and/or specificity or combination thereof may be identified.

Detailed Description

Provided are methods for determining the fraction of fetal nucleic acid in a test sample derived from a pregnant female with improved accuracy and/or precision. In some embodiments, provided herein are methods of using fetal fraction measurements to determine the presence or absence of a genetic variation in a fetus with improved accuracy and/or precision. In some embodiments the determination of fetal fraction and the determination of the presence or absence of a fetal genetic variation are obtained from the same sample, sequencing run, sequencing reads and/or the same data obtained from the same flow cell. Also provided herein are improved data manipulation methods as well as systems, apparatuses, modules and procedures that, in some embodiments, carry out the methods described herein. Provided herein are methods, processes and

apparatuses useful for identifying a genetic variation. In some embodiments identifying a genetic variation sometimes comprises detecting a copy number variation and/or sometimes comprises adjusting an elevation comprising a copy number variation. In some embodiments, an elevation is adjusted providing an identification of one or more genetic variations or variances with a reduced likelihood of a false positive or false negative diagnosis. In some embodiments, identifying a genetic variation by a method described herein can lead to a diagnosis of, or determining a predisposition to, a particular medical condition. Identifying a genetic variance can result in facilitating a medical decision and/or employing a helpful medical procedure. The term "a genetic variation", as referred to herein, means one or more genetic variations.

Samples

Provided herein are methods and compositions for analyzing nucleic acid. In some embodiments, nucleic acid fragments in a mixture of nucleic acid fragments are analyzed. A mixture of nucleic acids can comprise two or more nucleic acid fragment species having different nucleotide sequences, different fragment lengths, different origins (e.g., genomic origins, fetal vs. maternal origins, cell or tissue origins, sample origins, subject origins, and the like), or combinations thereof.

Nucleic acid or a nucleic acid mixture utilized in methods and apparatuses described herein often is isolated from a sample obtained from a subject. A subject can be any living or non-living organism, including but not limited to a human (e.g., a male human, female human, fetus, pregnant female, child, or the like), a non-human animal, a plant, a bacterium, a fungus or a protist. Any human or non-human animal can be selected, including but not limited to mammal, reptile, avian, amphibian, fish, ungulate, ruminant, bovine (e.g., cattle), equine (e.g., horse), caprine and ovine (e.g., sheep, goat), swine (e.g., pig), camelid (e.g., camel, llama, alpaca), monkey, ape (e.g., gorilla, chimpanzee), ursid (e.g., bear), poultry, dog, cat, mouse, rat, fish, dolphin, whale and shark. A subject may be a male or female (e.g., woman).

Nucleic acid may be isolated from any type of suitable biological specimen or sample (e.g., a test sample). A sample or test sample can be any specimen that is isolated or obtained from a subject (e.g., a human subject, a pregnant female). Non-limiting examples of specimens include fluid or tissue from a subject, including, without limitation, umbilical cord blood, chorionic villi, amniotic fluid, cerebrospinal fluid, spinal fluid, lavage fluid (e.g., bronchoalveolar, gastric, peritoneal, ductal, ear, arthroscopic), biopsy sample (e.g., from pre-implantation embryo), celocentesis sample, fetal

nucleated cells or fetal cellular remnants, washings of female reproductive tract, urine, feces, sputum, saliva, nasal mucous, prostate fluid, lavage, semen, lymphatic fluid, bile, tears, sweat, breast milk, breast fluid, embryonic cells and fetal cells (e.g., placental cells). In some embodiments, a biological sample is a cervical swab from a subject. In some embodiments, a biological sample may be blood and sometimes plasma or serum. As used herein, the term "blood" encompasses whole blood or any fractions of blood, such as serum and plasma as conventionally defined, for example. Blood or fractions thereof often comprise nucleosomes (e.g., maternal and/or fetal nucleosomes). Nucleosomes comprise nucleic acids and are sometimes cell-free or intracellular. Blood also comprises buffy coats. Buffy coats are sometimes isolated by utilizing a ficoll gradient. Buffy coats can comprise white blood cells (e.g., leukocytes, T-cells, B-cells, platelets, and the like). In certain embodiments buffy coats comprise maternal and/or fetal nucleic acid. Blood plasma refers to the fraction of whole blood resulting from centrifugation of blood treated with anticoagulants. Blood serum refers to the watery portion of fluid remaining after a blood sample has coagulated. Fluid or tissue samples often are collected in accordance with standard protocols hospitals or clinics generally follow. For blood, an appropriate amount of peripheral blood (e.g., between 3-40 milliliters) often is collected and can be stored according to standard procedures prior to or after preparation. A fluid or tissue sample from which nucleic acid is extracted may be acellular (e.g., cell-free). In some embodiments, a fluid or tissue sample may contain cellular elements or cellular remnants. In some embodiments fetal cells or cancer cells may be included in the sample.

A sample often is heterogeneous, by which is meant that more than one type of nucleic acid species is present in the sample. For example, heterogeneous nucleic acid can include, but is not limited to, (i) fetal derived and maternal derived nucleic acid, (ii) cancer and non-cancer nucleic acid, (iii) pathogen and host nucleic acid, and more generally, (iv) mutated and wild-type nucleic acid. A sample may be heterogeneous because more than one cell type is present, such as a fetal cell and a maternal cell, a cancer and non-cancer cell, or a pathogenic and host cell. In some embodiments, a minority nucleic acid species and a majority nucleic acid species is present.

For prenatal applications of technology described herein, fluid or tissue sample may be collected from a female at a gestational age suitable for testing, or from a female who is being tested for possible pregnancy. Suitable gestational age may vary depending on the prenatal test being performed. In certain embodiments, a pregnant female subject sometimes is in the first trimester of pregnancy, at times in the second trimester of pregnancy, or sometimes in the third trimester of

pregnancy. In certain embodiments, a fluid or tissue is collected from a pregnant female between about 1 to about 45 weeks of fetal gestation (e.g., at 1-4, 4-8, 8-12, 12-16, 16-20, 20-24, 24-28, 28-32, 32-36, 36-40 or 40-44 weeks of fetal gestation), and sometimes between about 5 to about 28 weeks of fetal gestation (e.g., at 6, 7, 8, 9,10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 or 27 weeks of fetal gestation). In certain embodiments a fluid or tissue sample is collected from a pregnant female during or just after (e.g., 0 to 72 hours after) giving birth (e.g., vaginal or non-vaginal birth (e.g., surgical delivery), or miscarriage).

Acquisition of Blood Samples and Extraction of DNA

The present technology relates to separating, enriching and analyzing fetal DNA found in maternal blood as a non-invasive means to detect the presence and/or to monitor the progress of a pregnancy-associated condition or disorder. Thus, the first steps of practicing the technology herein are to obtain a blood sample from a pregnant woman and extract DNA from the sample.

Acquisition of Blood Samples

A blood sample is obtained from a pregnant woman at a gestational age suitable for testing using a method of the present technology. The suitable gestational age may vary depending on the disorder tested, as discussed below. Collection of blood from a woman is performed in accordance with the standard protocol hospitals or clinics generally follow. An appropriate amount of peripheral blood, e.g., typically between 5-50 ml, is collected and may be stored according to standard procedure prior to further preparation. Blood samples may be collected, stored or transported in a manner known to the person of ordinary skill in the art to minimize degradation or the quality of nucleic acid present in the sample.

Preparation of Blood Samples

The analysis of fetal DNA found in maternal blood according to the present technology may be performed using, e.g., the whole blood, serum, or plasma. The methods for preparing serum or plasma from maternal blood are well known among those of skill in the art. For example, a pregnant woman's blood can be placed in a tube containing EDTA or a specialized commercial product such as Vacutainer SST (Becton Dickinson, Franklin Lakes, N.J.) to prevent blood clotting, and plasma can then be obtained from whole blood through centrifugation. On the other hand,

serum may be obtained with or without centrifugation-following blood clotting. If centrifugation is used then it is typically, though not exclusively, conducted at an appropriate speed, e.g., 1,500-3,000 times g. Plasma or serum may be subjected to additional centrifugation steps before being transferred to a fresh tube for DNA extraction.

In addition to the acellular portion of the whole blood, DNA may also be recovered from the cellular fraction, enriched in the buffy coat portion, which can be obtained following centrifugation of a whole blood sample from the woman and removal of the plasma.

Extraction of DNA

There are numerous known methods for extracting DNA from a biological sample including blood. The general methods of DNA preparation (e.g., described by Sambrook and Russell, Molecular Cloning: A Laboratory Manual 3d ed., 2001) can be followed; various commercially available reagents or kits, such as Qiagen's QIAamp Circulating Nucleic Acid Kit, QiaAmp DNA Mini Kit or QiaAmp DNA Blood Mini Kit (Qiagen, Hilden, Germany), GenomicPrep™ Blood DNA Isolation Kit (Promega, Madison, Wis.), and GFX™ Genomic Blood DNA Purification Kit (Amersham, Piscataway, N.J.), may also be used to obtain DNA from a blood sample from a pregnant woman. Combinations of more than one of these methods may also be used.

In some embodiments, the sample may first be enriched or relatively enriched for fetal nucleic acid by one or more methods. For example, the discrimination of fetal and maternal DNA can be performed using the compositions and processes of the present technology alone or in combination with other discriminating factors. Examples of these factors include, but are not limited to, single nucleotide differences between chromosome X and Y, chromosome Y-specific sequences, polymorphisms located elsewhere in the genome, size differences between fetal and maternal DNA and differences in methylation pattern between maternal and fetal tissues.

Other methods for enriching a sample for a particular species of nucleic acid are described in PCT Patent Application Number PCT/US07/69991, filed May 30, 2007, PCT Patent Application Number PCT/US2007/071232, filed June 15, 2007, US Provisional Application Numbers 60/968,876 and 60/968,878 (assigned to the Applicant), (PCT Patent Application Number PCT/EP05/012707, filed November 28, 2005) which are all hereby incorporated by reference. In certain embodiments,

maternal nucleic acid is selectively removed (either partially, substantially, almost completely or completely) from the sample.

Nucleic Acid Isolation and Processing

Nucleic acid may be derived from one or more sources (e.g., cells, serum, plasma, buffy coat, lymphatic fluid, skin, soil, and the like) by methods known in the art. Cell lysis procedures and reagents are known in the art and may generally be performed by chemical (e.g., detergent, hypotonic solutions, enzymatic procedures, and the like, or combination thereof), physical (e.g., French press, sonication, and the like), or electrolytic lysis methods. Any suitable lysis procedure can be utilized. For example, chemical methods generally employ lysing agents to disrupt cells and extract the nucleic acids from the cells, followed by treatment with chaotropic salts. Physical methods such as freeze/thaw followed by grinding, the use of cell presses and the like also are useful. High salt lysis procedures also are commonly used. For example, an alkaline lysis procedure may be utilized. The latter procedure traditionally incorporates the use of phenol-chloroform solutions, and an alternative phenol-chloroform-free procedure involving three solutions can be utilized. In the latter procedures, one solution can contain 15mM Tris, pH 8.0; 10mM EDTA and 100 ug/ml Rnase A; a second solution can contain 0.2N NaOH and 1% SDS; and a third solution can contain 3M KOAc, pH 5.5. These procedures can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989), incorporated herein in its entirety.

The terms “nucleic acid” and “nucleic acid molecule” are used interchangeably. The terms refer to nucleic acids of any composition form, such as deoxyribonucleic acid (DNA, e.g., complementary DNA (cDNA), genomic DNA (gDNA) and the like), ribonucleic acid (RNA, e.g., message RNA (mRNA), short inhibitory RNA (siRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA, RNA highly expressed by the fetus or placenta, and the like), and/or DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like), RNA/DNA hybrids and polyamide nucleic acids (PNAs), all of which can be in single- or double-stranded form. Unless otherwise limited, a nucleic acid can comprise known analogs of natural nucleotides, some of which can function in a similar manner as naturally occurring nucleotides. A nucleic acid can be in any form useful for conducting processes herein (e.g., linear, circular, supercoiled, single-stranded, double-stranded and the like). A nucleic acid may be, or may be from, a plasmid, phage, autonomously replicating sequence (ARS), centromere, artificial chromosome, chromosome, or other nucleic acid able to replicate or be replicated in vitro or in a host cell, a cell, a cell nucleus or

cytoplasm of a cell in certain embodiments. A nucleic acid in some embodiments can be from a single chromosome or fragment thereof (e.g., a nucleic acid sample may be from one chromosome of a sample obtained from a diploid organism). In certain embodiments nucleic acids comprise nucleosomes, fragments or parts of nucleosomes or nucleosome-like structures. Nucleic acids sometimes comprise protein (e.g., histones, DNA binding proteins, and the like). Nucleic acids analyzed by processes described herein sometimes are substantially isolated and are not substantially associated with protein or other molecules. Nucleic acids also include derivatives, variants and analogs of RNA or DNA synthesized, replicated or amplified from single-stranded ("sense" or "antisense", "plus" strand or "minus" strand, "forward" reading frame or "reverse" reading frame) and double-stranded polynucleotides. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. For RNA, the base cytosine is replaced with uracil and the sugar 2' position includes a hydroxyl moiety. A nucleic acid may be prepared using a nucleic acid obtained from a subject as a template.

Nucleic acid may be isolated at a different time point as compared to another nucleic acid, where each of the samples is from the same or a different source. A nucleic acid may be from a nucleic acid library, such as a cDNA or RNA library, for example. A nucleic acid may be a result of nucleic acid purification or isolation and/or amplification of nucleic acid molecules from the sample. Nucleic acid provided for processes described herein may contain nucleic acid from one sample or from two or more samples (e.g., from 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, or 20 or more samples).

Nucleic acids can include extracellular nucleic acid in certain embodiments. The term "extracellular nucleic acid" as used herein can refer to nucleic acid isolated from a source having substantially no cells and also is referred to as "cell-free" nucleic acid and/or "cell-free circulating" nucleic acid. Extracellular nucleic acid can be present in and obtained from blood (e.g., from the blood of a pregnant female). Extracellular nucleic acid often includes no detectable cells and may contain cellular elements or cellular remnants. Non-limiting examples of acellular sources for extracellular nucleic acid are blood, blood plasma, blood serum and urine. As used herein, the term "obtain cell-free circulating sample nucleic acid" includes obtaining a sample directly (e.g., collecting a sample, e.g., a test sample) or obtaining a sample from another who has collected a sample. Without being limited by theory, extracellular nucleic acid may be a product of cell

apoptosis and cell breakdown, which provides basis for extracellular nucleic acid often having a series of lengths across a spectrum (e.g., a "ladder").

Extracellular nucleic acid can include different nucleic acid species, and therefore is referred to herein as "heterogeneous" in certain embodiments. For example, blood serum or plasma from a person having cancer can include nucleic acid from cancer cells and nucleic acid from non-cancer cells. In another example, blood serum or plasma from a pregnant female can include maternal nucleic acid and fetal nucleic acid. In some instances, fetal nucleic acid sometimes is about 5% to about 50% of the overall nucleic acid (e.g., about 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, or 49% of the total nucleic acid is fetal nucleic acid). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 500 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 500 base pairs or less). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 250 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 250 base pairs or less). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 200 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 200 base pairs or less). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 150 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 150 base pairs or less). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 100 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 100 base pairs or less). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 50 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 50 base pairs or less). In some embodiments, the majority of fetal nucleic acid in nucleic acid is of a length of about 25 base pairs or less (e.g., about 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of fetal nucleic acid is of a length of about 25 base pairs or less).

Nucleic acid may be provided for conducting methods described herein without processing of the sample(s) containing the nucleic acid, in certain embodiments. In some embodiments, nucleic acid is provided for conducting methods described herein after processing of the sample(s) containing the nucleic acid. For example, a nucleic acid can be extracted, isolated, purified, partially purified

or amplified from the sample(s). The term “isolated” as used herein refers to nucleic acid removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered by human intervention (e.g., "by the hand of man") from its original environment. The term “isolated nucleic acid” as used herein can refer to a nucleic acid removed from a subject (e.g., a human subject). An isolated nucleic acid can be provided with fewer non-nucleic acid components (e.g., protein, lipid) than the amount of components present in a source sample. A composition comprising isolated nucleic acid can be about 50% to greater than 99% free of non-nucleic acid components. A composition comprising isolated nucleic acid can be about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of non-nucleic acid components. The term “purified” as used herein can refer to a nucleic acid provided that contains fewer non-nucleic acid components (e.g., protein, lipid, carbohydrate) than the amount of non-nucleic acid components present prior to subjecting the nucleic acid to a purification procedure. A composition comprising purified nucleic acid may be about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other non-nucleic acid components. The term “purified” as used herein can refer to a nucleic acid provided that contains fewer nucleic acid species than in the sample source from which the nucleic acid is derived. A composition comprising purified nucleic acid may be about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other nucleic acid species. For example, fetal nucleic acid can be purified from a mixture comprising maternal and fetal nucleic acid. In certain examples, nucleosomes comprising small fragments of fetal nucleic acid can be purified from a mixture of larger nucleosome complexes comprising larger fragments of maternal nucleic acid.

The term “amplified” as used herein refers to subjecting a target nucleic acid in a sample to a process that linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the target nucleic acid, or segment thereof. The term “amplified” as used herein can refer to subjecting a target nucleic acid (e.g., in a sample comprising other nucleic acids) to a process that selectively and linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the target nucleic acid, or segment thereof. The term “amplified” as used herein can refer to subjecting a population of nucleic acids to a process that non-selectively and linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as nucleic acids, or portions thereof, that were present in the sample prior to amplification. In

certain embodiments the term “amplified” refers to a method that comprises a polymerase chain reaction (PCR).

Nucleic acid also may be processed by subjecting nucleic acid to a method that generates nucleic acid fragments, in certain embodiments, before providing nucleic acid for a process described herein. In some embodiments, nucleic acid subjected to fragmentation or cleavage may have a nominal, average or mean length of about 5 to about 10,000 base pairs, about 100 to about 1,000 base pairs, about 100 to about 500 base pairs, or about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 or 9000 base pairs. Fragments can be generated by a suitable method known in the art, and the average, mean or nominal length of nucleic acid fragments can be controlled by selecting an appropriate fragment-generating procedure. In certain embodiments, nucleic acid of a relatively shorter length can be utilized to analyze sequences that contain little sequence variation and/or contain relatively large amounts of known nucleotide sequence information. In some embodiments, nucleic acid of a relatively longer length can be utilized to analyze sequences that contain greater sequence variation and/or contain relatively small amounts of nucleotide sequence information.

Nucleic acid fragments may contain overlapping nucleotide sequences, and such overlapping sequences can facilitate construction of a nucleotide sequence of the non-fragmented counterpart nucleic acid, or a segment thereof. For example, one fragment may have subsequences x and y and another fragment may have subsequences y and z, where x, y and z are nucleotide sequences that can be 5 nucleotides in length or greater. Overlap sequence y can be utilized to facilitate construction of the x-y-z nucleotide sequence in nucleic acid from a sample in certain embodiments. Nucleic acid may be partially fragmented (e.g., from an incomplete or terminated specific cleavage reaction) or fully fragmented in certain embodiments.

Nucleic acid can be fragmented by various methods known in the art, which include without limitation, physical, chemical and enzymatic processes. Non-limiting examples of such processes are described in U.S. Patent Application Publication No. 20050112590 (published on May 26, 2005, entitled “Fragmentation-based methods and systems for sequence variation detection and discovery,” naming Van Den Boom et al.). Certain processes can be selected to generate non-specifically cleaved fragments or specifically cleaved fragments. Non-limiting examples of processes that can generate non-specifically cleaved fragment nucleic acid include, without

limitation, contacting nucleic acid with apparatus that expose nucleic acid to shearing force (e.g., passing nucleic acid through a syringe needle; use of a French press); exposing nucleic acid to irradiation (e.g., gamma, x-ray, UV irradiation; fragment sizes can be controlled by irradiation intensity); boiling nucleic acid in water (e.g., yields about 500 base pair fragments) and exposing nucleic acid to an acid and base hydrolysis process.

As used herein, “fragmentation” or “cleavage” refers to a procedure or conditions in which a nucleic acid molecule, such as a nucleic acid template gene molecule or amplified product thereof, may be severed into two or more smaller nucleic acid molecules. Such fragmentation or cleavage can be sequence specific, base specific, or nonspecific, and can be accomplished by any of a variety of methods, reagents or conditions, including, for example, chemical, enzymatic, physical fragmentation.

As used herein, “fragments”, “cleavage products”, “cleaved products” or grammatical variants thereof, refers to nucleic acid molecules resultant from a fragmentation or cleavage of a nucleic acid template gene molecule or amplified product thereof. While such fragments or cleaved products can refer to all nucleic acid molecules resultant from a cleavage reaction, typically such fragments or cleaved products refer only to nucleic acid molecules resultant from a fragmentation or cleavage of a nucleic acid template gene molecule or the segment of an amplified product thereof containing the corresponding nucleotide sequence of a nucleic acid template gene molecule. For example, an amplified product can contain one or more nucleotides more than the amplified nucleotide region of a nucleic acid template sequence (e.g., a primer can contain "extra" nucleotides such as a transcriptional initiation sequence, in addition to nucleotides complementary to a nucleic acid template gene molecule, resulting in an amplified product containing "extra" nucleotides or nucleotides not corresponding to the amplified nucleotide region of the nucleic acid template gene molecule). Accordingly, fragments can include fragments arising from portions of amplified nucleic acid molecules containing, at least in part, nucleotide sequence information from or based on the representative nucleic acid template molecule.

As used herein, the term “complementary cleavage reactions” refers to cleavage reactions that are carried out on the same nucleic acid using different cleavage reagents or by altering the cleavage specificity of the same cleavage reagent such that alternate cleavage patterns of the same target or reference nucleic acid or protein are generated. In certain embodiments, nucleic acid may be treated with one or more specific cleavage agents (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more specific

cleavage agents) in one or more reaction vessels (e.g., nucleic acid is treated with each specific cleavage agent in a separate vessel).

Nucleic acid may be specifically cleaved or non-specifically cleaved by contacting the nucleic acid with one or more enzymatic cleavage agents (e.g., nucleases, restriction enzymes). The term "specific cleavage agent" as used herein refers to an agent, sometimes a chemical or an enzyme that can cleave a nucleic acid at one or more specific sites. Specific cleavage agents often cleave specifically according to a particular nucleotide sequence at a particular site. Non-specific cleavage agents often cleave nucleic acids at non-specific sites or degrade nucleic acids. Non-specific cleavage agents often degrade nucleic acids by removal of nucleotides from the end (either the 5' end, 3' end or both) of a nucleic acid strand.

Any suitable non-specific or specific enzymatic cleavage agent can be used to cleave or fragment nucleic acids. A suitable restriction enzyme can be used to cleave nucleic acids, in some embodiments. Examples of enzymatic cleavage agents include without limitation endonucleases (e.g., DNase (e.g., DNase I, II); RNase (e.g., RNase E, F, H, P); Cleavase™ enzyme; Taq DNA polymerase; E. coli DNA polymerase I and eukaryotic structure-specific endonucleases; murine FEN-1 endonucleases; type I, II or III restriction endonucleases such as Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bcl I, Bgl I. Bgl II, Bln I, Bsm I, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EclX I, EcoR I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I, Sca I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I; glycosylases (e.g., uracil-DNA glycosylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-etheno-adenine DNA glycosylase); exonucleases (e.g., exonuclease III); ribozymes, and DNAzymes. Nucleic acid may be treated with a chemical agent, and the modified nucleic acid may be cleaved. In non-limiting examples, nucleic acid may be treated with (i) alkylating agents such as methylnitrosourea that generate several alkylated bases, including N3-methyladenine and N3-methylguanine, which are recognized and cleaved by alkyl purine DNA-glycosylase; (ii) sodium bisulfite, which causes deamination of cytosine residues in DNA to form uracil residues that can be cleaved by uracil N-glycosylase; and (iii) a chemical agent that converts guanine to its oxidized form, 8-hydroxyguanine, which can be cleaved by formamidopyrimidine DNA N-glycosylase.

Examples of chemical cleavage processes include without limitation alkylation, (e.g., alkylation of phosphorothioate-modified nucleic acid); cleavage of acid lability of P3'-N5'-phosphoroamidate-containing nucleic acid; and osmium tetroxide and piperidine treatment of nucleic acid.

Nucleic acid also may be exposed to a process that modifies certain nucleotides in the nucleic acid before providing nucleic acid for a method described herein. A process that selectively modifies nucleic acid based upon the methylation state of nucleotides therein can be applied to nucleic acid, for example. In addition, conditions such as high temperature, ultraviolet radiation, x-radiation, can induce changes in the sequence of a nucleic acid molecule. Nucleic acid may be provided in any form useful for conducting a sequence analysis or manufacture process described herein, such as solid or liquid form, for example. In certain embodiments, nucleic acid may be provided in a liquid form optionally comprising one or more other components, including without limitation one or more buffers or salts.

Nucleic acid may be single or double stranded. Single stranded DNA, for example, can be generated by denaturing double stranded DNA by heating or by treatment with alkali, for example. In certain embodiments, nucleic acid is in a D-loop structure, formed by strand invasion of a duplex DNA molecule by an oligonucleotide or a DNA-like molecule such as peptide nucleic acid (PNA). D loop formation can be facilitated by addition of E. Coli RecA protein and/or by alteration of salt concentration, for example, using methods known in the art.

Genomic DNA Target Sequences

In some embodiments of the methods provided herein, one or more nucleic acid species, and sometimes one or more nucleotide sequence species, are targeted for amplification and quantification. In some embodiments, the targeted nucleic acids are genomic DNA sequences. Certain genomic DNA target sequences are used, for example, because they can allow for the determination of a particular feature for a given assay. Genomic DNA target sequences can be referred to herein as markers for a given assay. In some cases, genomic target sequences are polymorphic, as described herein. In some embodiments, more than one genomic DNA target sequence or marker can allow for the determination of a particular feature for a given assay. Such genomic DNA target sequences are considered to be of a particular “region”. As used herein, a “region” is not intended to be limited to a description of a genomic location, such as a particular chromosome, stretch of chromosomal DNA or genetic locus. Rather, the term “region” is used

herein to identify a collection of one or more genomic DNA target sequences or markers that can be indicative of a particular assay. Such assays can include, but are not limited to, assays for the detection and quantification of fetal nucleic acid, assays for the detection and quantification of maternal nucleic acid, assays for the detection and quantification of total DNA, assays for the detection and quantification of methylated DNA, assays for the detection and quantification of fetal specific nucleic acid (e.g., chromosome Y DNA), and assays for the detection and quantification of digested and/or undigested DNA, as an indicator of digestion efficiency. In some embodiments, the genomic DNA target sequence is described as being within a particular genomic locus. As used herein, a genomic locus can include any or a combination of open reading frame DNA, non-transcribed DNA, intronic sequences, extronic sequences, promoter sequences, enhancer sequences, flanking sequences, or any sequences considered by one of skill in the art to be associated with a given genomic locus.

Assays for the Determination of Methylated DNA

In some embodiments of the methods provided herein, one or more genomic DNA target sequences are used that can allow for the determination of methylated DNA. Generally, genomic DNA target sequences used for the determination of methylated DNA are differentially methylated in fetal and maternal nucleic acid, and thus, differentially digested according to the methods provided herein for methylation-sensitive restriction enzymes. In some cases, a genomic DNA target sequence is a single copy gene. In some cases, a genomic DNA target sequence is located on chromosome 13, chromosome 18, chromosome 21, chromosome X, or chromosome Y. In some cases, a genomic DNA target sequence is not located on chromosome 13. In some cases, a genomic DNA target sequence is not located on chromosome 18. In some cases, a genomic DNA target sequence is not located on chromosome 21. In some cases, a genomic DNA target sequence is not located on chromosome X. In some cases, a genomic DNA target sequence is not located on chromosome Y. In some cases, a genomic DNA target sequence is typically methylated in one DNA species such as, for example, placental DNA (i.e. at least about 50% or greater methylation). In some cases, the genomic DNA target sequence is minimally methylated in another DNA species such as, for example, maternal DNA (i.e. less than about 1% methylation). In some cases, the genomic DNA target sequence does not contain any known single nucleotide polymorphisms (SNPs) within the PCR primer hybridization sequences. In some cases, the genomic DNA target sequence does not contain any known mutations within the PCR primer hybridization sequences. In some cases, the genomic DNA target sequence does not contain any

known insertion or deletions within the PCR primer hybridization sequences. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not below 65 °C. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not above 75 °C. In some cases, the genomic DNA target sequence contains at least two restriction sites within the amplified region. In some embodiments, the genomic DNA target sequence length is about 50 base pairs to about 200 base pairs. In some cases, the genomic DNA target sequence length is 70 base pairs. In some cases, the genomic DNA target sequence does not possess any negative ΔG values for secondary structure of the complete amplicon prediction using mfold (M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406-15, (2003)). In some embodiments, the genomic DNA target sequence used for the determination of methylated DNA is within the TBX3 locus. In some embodiments, the genomic DNA target sequence used for the determination of methylated DNA is within the SOX14 locus. Additional genomic targets that can be used for the determination of methylated DNA in conjunction with the methods provided herein are presented in Example 3.

Assays for the Determination of Total DNA

In some embodiments of the methods provided herein, one or more genomic DNA target sequences are used that can allow for the determination of total DNA. Generally, genomic DNA target sequences used for the determination of total DNA are present in every genome copy (e.g., is present in fetal DNA and maternal DNA, cancer DNA and normal DNA, pathogen DNA and host DNA). In some cases, a genomic DNA target sequence is a single copy gene. In some cases, a genomic DNA target sequence is located on chromosome 13, chromosome 18, chromosome 21, chromosome X, or chromosome Y. In some cases, a genomic DNA target sequence is not located on chromosome 13. In some cases, a genomic DNA target sequence is not located on chromosome 18. In some cases, a genomic DNA target sequence is not located on chromosome 21. In some cases, a genomic DNA target sequence is not located on chromosome X. In some cases, a genomic DNA target sequence is not located on chromosome Y. In some cases, a genomic DNA target sequence does not contain any known single nucleotide polymorphisms (SNPs) within the PCR primer hybridization sequences. In some cases, a genomic DNA target sequence does not contain any known mutations within the PCR primer hybridization sequences. In some cases, a genomic DNA target sequence does not contain any known insertion or deletions within the PCR primer hybridization sequences. In some cases, the melting temperature of the

PCR primers that can hybridize to a genomic DNA target sequence is not below 65 °C. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not above 75 °C. In some embodiments, the genomic DNA target sequence length is about 50 base pairs to about 200 base pairs. In some cases, the genomic DNA target sequence length is 70 base pairs. In some cases, the genomic DNA target sequence does not possess any negative ΔG values for secondary structure of the complete amplicon prediction using mfold (M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406-15, (2003)). In some embodiments, the genomic DNA target sequence used for the determination of total DNA is within the ALB locus. In some embodiments, the genomic DNA target sequence used for the determination of total DNA is within the APOE or RNAseP locus.

Assays for the Determination of Fetal DNA

In some embodiments of the methods provided herein, one or more genomic DNA target sequences are used that can allow for the determination of fetal DNA. In some embodiments, genomic DNA target sequences used for the determination of fetal DNA are specific to the Y chromosome. In some cases, the genomic DNA target sequence is a single copy gene. In some cases, the genomic DNA target sequence does not contain any known single nucleotide polymorphisms (SNPs) within the PCR primer hybridization sequences. In some cases, the genomic DNA target sequence does not contain any known mutations within the PCR primer hybridization sequences. In some cases, the genomic DNA target sequence does not contain any known insertion or deletions within the PCR primer hybridization sequences. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not below 65 °C. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not above 75 °C. In some cases, the genomic DNA target sequence does not contain the restriction site GCGC within the amplified region. In some embodiments, the genomic DNA target sequence length is about 50 base pairs to about 200 base pairs. In some cases, the genomic DNA target sequence length is 70 base pairs. In some cases, the genomic DNA target sequence does not possess any negative ΔG values for secondary structure of the complete amplicon prediction using mfold (M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406-15, (2003)). In some embodiments, the genomic DNA target sequence used for the determination of fetal DNA is within the UTY locus. In some embodiments, the genomic DNA target sequence used for the determination of fetal DNA is within the SRY1 or SRY2 locus.

Determination of Fetal Gender

In some embodiments, the prediction of a fetal gender or gender related disorder (e.g., sex chromosome aneuploidy) can be determined by a method or apparatus described herein. Gender determination generally is based on a sex chromosome. In humans, there are two sex chromosomes, the X and Y chromosomes. The Y chromosome contains a gene, SRY, which triggers embryonic development as a male. The Y chromosomes of humans and other mammals also contain other genes needed for normal sperm production. Individuals with XX are female and XY are male and non-limiting variations, often referred to as sex chromosome aneuploidies, include X0, XYY, XXX and XXY. In certain embodiments, males have two X chromosomes and one Y chromosome (XXY; Klinefelter's Syndrome), or one X chromosome and two Y chromosomes (XYY syndrome; Jacobs Syndrome), and some females have three X chromosomes (XXX; Triple X Syndrome) or a single X chromosome instead of two (X0; Turner Syndrome). In certain embodiments, only a portion of cells in an individual are affected by a sex chromosome aneuploidy which may be referred to as a mosaicism (e.g., Turner mosaicism). Other cases include those where SRY is damaged (leading to an XY female), or copied to the X (leading to an XX male).

In certain cases, it can be beneficial to determine the gender of a fetus in utero. For example, a patient (e.g., pregnant female) with a family history of one or more sex-linked disorders may wish to determine the gender of the fetus she is carrying to help assess the risk of the fetus inheriting such a disorder. Sex-linked disorders include, without limitation, X-linked and Y-linked disorders. X-linked disorders include X-linked recessive and X-linked dominant disorders. Examples of X-linked recessive disorders include, without limitation, immune disorders (e.g., chronic granulomatous disease (CYBB), Wiskott–Aldrich syndrome, X-linked severe combined immunodeficiency, X-linked agammaglobulinemia, hyper-IgM syndrome type 1, IPEX, X-linked lymphoproliferative disease, Properdin deficiency), hematologic disorders (e.g., Hemophilia A, Hemophilia B, X-linked sideroblastic anemia), endocrine disorders (e.g., androgen insensitivity syndrome/Kennedy disease, KAL1 Kallmann syndrome, X-linked adrenal hypoplasia congenital), metabolic disorders (e.g., ornithine transcarbamylase deficiency, oculocerebrorenal syndrome, adrenoleukodystrophy, glucose-6-phosphate dehydrogenase deficiency, pyruvate dehydrogenase deficiency, Danon disease/glycogen storage disease Type IIb, Fabry's disease, Hunter syndrome, Lesch–Nyhan syndrome, Menkes disease/occipital horn syndrome), nervous system disorders (e.g., Coffin–Lowry syndrome, MASA syndrome, X-linked alpha thalassemia mental retardation

syndrome, Siderius X-linked mental retardation syndrome, color blindness, ocular albinism, Norrie disease, choroideremia, Charcot–Marie–Tooth disease (CMTX2-3), Pelizaeus–Merzbacher disease, SMAX2), skin and related tissue disorders (e.g., dyskeratosis congenital, hypohidrotic ectodermal dysplasia (EDA), X-linked ichthyosis, X-linked endothelial corneal dystrophy), neuromuscular disorders (e.g., Becker's muscular dystrophy/Duchenne, centronuclear myopathy (MTM1), Conradi–Hünermann syndrome, Emery–Dreifuss muscular dystrophy 1), urologic disorders (e.g., Alport syndrome, Dent's disease, X-linked nephrogenic diabetes insipidus), bone/tooth disorders (e.g., AMELX Amelogenesis imperfecta), and other disorders (e.g., Barth syndrome, McLeod syndrome, Smith-Fineman-Myers syndrome, Simpson–Golabi–Behmel syndrome, Mohr–Tranebjærg syndrome, Nasodigitoacoustic syndrome). Examples of X-linked dominant disorders include, without limitation, X-linked hypophosphatemia, Focal dermal hypoplasia, Fragile X syndrome, Aicardi syndrome, Incontinentia pigmenti, Rett syndrome, CHILD syndrome, Lujan–Fryns syndrome, and Orofaciodigital syndrome 1. Examples of Y-linked disorders include, without limitation, male infertility, retinits pigmentosa, and azoospermia.

Fetal gender can be determined by a suitable method, non-limiting examples of which include chorionic villus sampling, amniocentesis, obstetric ultrasonography, the like and methods described in International Patent Application No. PCT/US12/59592 and US Patent Application No. 13/656,328 both of which are incorporated herein by reference.

Assays for the Determination of Digested and/or Undigested DNA

In some embodiments of the methods provided herein, one or more genomic DNA target sequences are used that can allow for the determination of the amount of digested or undigested nucleic acid, as an indicator of digestion efficiency. Such genomic DNA target sequences are present in every genome in the sample (e.g., maternal and fetal species genomes). Generally, genomic DNA target sequences used for the determination of digested or undigested DNA contain at least one restriction site present in a genomic DNA target sequence used in another assay. Thus, the genomic DNA target sequences used for the determination of digested or undigested DNA serve as controls for assays that include differential digestion. Generally, the genomic DNA target sequence is unmethylated in all nucleic acid species tested (e.g., unmethylated in both maternal and fetal species genomes). In some cases, the genomic DNA target sequence is a single copy gene. In some cases, the genomic DNA target sequence is not located on

chromosome 13. In some cases, the genomic DNA target sequence is not located on chromosome 18. In some cases, the genomic DNA target sequence is not located on chromosome 21. In some cases, the genomic DNA target sequence is not located on chromosome X. In some cases, the genomic DNA target sequence is not located on chromosome Y. In some cases, the genomic DNA target sequence does not contain any known single nucleotide polymorphisms (SNPs) within the PCR primer hybridization sequences. In some cases, the genomic DNA target sequence does not contain any known mutations within the PCR primer hybridization sequences. In some cases, the genomic DNA target sequence does not contain any known insertion or deletions within the PCR primer hybridization sequences. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not below 65 °C. In some cases, the melting temperature of the PCR primers that can hybridize to a genomic DNA target sequence is not above 75 °C. In some embodiments, the genomic DNA target sequence length is about 50 base pairs to about 200 base pairs. In some cases, the genomic DNA target sequence length is 70 base pairs. In some cases, the genomic DNA target sequence does not possess any negative ΔG values for secondary structure of the complete amplicon prediction using mfold (M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 31 (13), 3406-15, (2003)). In some embodiments, the genomic DNA target sequence used for the determination of digested or undigested DNA is within the POP5 locus. In some embodiments, the genomic DNA target sequence used for the determination of digested or undigested DNA is within the LDHA locus.

Methylation Specific Separation of Nucleic Acid

The methods provided herein offer an alternative approach for the enrichment of fetal DNA based on the methylation-specific separation of differentially methylated DNA. It has recently been discovered that many genes involved in developmental regulation are controlled through epigenetics in embryonic stem cells. Consequently, multiple genes can be expected to show differential DNA methylation between nucleic acid of fetal origin and maternal origin. Once these regions are identified, a technique to capture methylated DNA can be used to specifically enrich fetal DNA. For identification of differentially methylated regions, a novel approach was used to capture methylated DNA. This approach uses a protein, in which the methyl binding domain of MBD2 is fused to the Fc fragment of an antibody (MBD-FC) (Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, Rehli M (2006) Genome wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res 66:6118-

6128). This fusion protein has several advantages over conventional methylation specific antibodies. The MBD-FC has a higher affinity to methylated DNA and it binds double stranded DNA. Most importantly the two proteins differ in the way they bind DNA. Methylation specific antibodies bind DNA stochastically, which means that only a binary answer can be obtained. The methyl binding domain of MBD-FC on the other hand binds all DNA molecules regardless of their methylation status. The strength of this protein - DNA interaction is defined by the level of DNA methylation. After binding genomic DNA, eluate solutions of increasing salt concentrations can be used to fractionate non-methylated and methylated DNA allowing for a more controlled separation (Gebhard C, Schwarzfischer L, Pham TH, Andreesen R, Mackensen A, Rehli M (2006) Rapid and sensitive detection of CpG-methylation using methyl-binding (MB)-PCR. Nucleic Acids Res 34:e82). Consequently this method, called Methyl-CpG immunoprecipitation (MCIP), cannot only enrich, but also fractionate genomic DNA according to methylation level, which is particularly helpful when the unmethylated DNA fraction should be investigated as well.

Methylation Sensitive Restriction Enzyme Digestion

The technology herein also provides compositions and processes for determining the amount of fetal nucleic acid from a maternal sample. The technology herein allows for the enrichment of fetal nucleic acid regions in a maternal sample by selectively digesting nucleic acid from said maternal sample with an enzyme that selectively and completely or substantially digests the maternal nucleic acid to enrich the sample for at least one fetal nucleic acid region. Preferably, the digestion efficiency is greater than about 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%. Following enrichment, the amount of fetal nucleic acid can be determined by quantitative methods that do not require polymorphic sequences or bisulfite treatment, thereby, offering a solution that works equally well for female fetuses and across different ethnicities and preserves the low copy number fetal nucleic acid present in the sample.

For example, there are methyl-sensitive enzymes that preferentially or substantially cleave or digest at their DNA recognition sequence if it is non-methylated. Thus, an unmethylated DNA sample will be cut into smaller fragments than a methylated DNA sample. Similarly, a hypermethylated DNA sample will not be cleaved. In contrast, there are methyl-sensitive enzymes that cleave at their DNA recognition sequence only if it is methylated.

Methyl-sensitive enzymes that digest unmethylated DNA suitable for use in methods of the technology herein include, but are not limited to, HpaII, HhaI, MaeII, BstUI and AciI. An enzyme that can be used is HpaII that cuts only the unmethylated sequence CCGG. Another enzyme that can be used is HhaI that cuts only the unmethylated sequence GCGC. Both enzymes are available from New England BioLabs®, Inc. Combinations of two or more methyl-sensitive enzymes that digest only unmethylated DNA can also be used. Suitable enzymes that digest only methylated DNA include, but are not limited to, DpnI, which cuts at a recognition sequence GATC, and McrBC, which belongs to the family of AAA.sup.+ proteins and cuts DNA containing modified cytosines and cuts at recognition site 5' . . . Pu.sup.mC(N.sub.40-3000) Pu.sup.mC . . . 3' (New England BioLabs, Inc., Beverly, Mass.).

Cleavage methods and procedures for selected restriction enzymes for cutting DNA at specific sites are well known to the skilled artisan. For example, many suppliers of restriction enzymes provide information on conditions and types of DNA sequences cut by specific restriction enzymes, including New England BioLabs, Pro-Mega Biochems, Boehringer-Mannheim, and the like. Sambrook et al. (See Sambrook et al., Molecular Biology: A laboratory Approach, Cold Spring Harbor, N.Y. 1989) provide a general description of methods for using restriction enzymes and other enzymes. Enzymes often are used under conditions that will enable cleavage of the maternal DNA with about 95%-100% efficiency, preferably with about 98%-100% efficiency.

Other Methods for Methylation Analysis

Various methylation analysis procedures are known in the art, and can be used in conjunction with the present technology. These assays allow for determination of the methylation state of one or a plurality of CpG islands within a DNA sequence. In addition, the methods maybe used to quantify methylated nucleic acid. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

Genomic sequencing is a technique that has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., Proc. Natl. Acad. Sci. USA 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA may be used, e.g., the method described by Sadri &

Hornsby (Nucl. Acids Res. 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird, Nucleic Acids Res. 25:2532-2534, 1997).

COBRA analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, Nucleic Acids Res. 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (Proc. Natl. Acad. Sci. USA 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples. Typical reagents (e.g., as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

The MethyLight™ assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan.RTM.) technology that requires no further manipulations after the PCR step (Eads et al., Cancer Res. 59:2302-2306, 1999). Briefly, the MethyLight™ process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

The MethyLight assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, where sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

The MethyLight process can by used with a "TaqMan" probe in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan.RTM. probes; e.g., with either biased primers and TaqMan.RTM. probe, or unbiased primers and TaqMan.RTM. probe. The TaqMan.RTM. probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10.degree. C. higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan.RTM. probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan.RTM. probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan.RTM. probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (e.g., as might be found in a typical MethyLight.TM.-based kit) for MethyLight.TM. analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan.RTM. probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

The Ms-SNuPE technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as

a template for methylation analysis at the CpG site(s) of interest. Small amounts of DNA can be analyzed (e.g., microdissected pathology sections), and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

Typical reagents (e.g., as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. Proc. Nat. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat. No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

The MCA technique is a method that can be used to screen for altered methylation patterns in genomic DNA, and to isolate specific sequences associated with these changes (Toyota et al., Cancer Res. 59:2307-12, 1999). Briefly, restriction enzymes with different sensitivities to cytosine methylation in their recognition sites are used to digest genomic DNAs from primary tumors, cell lines, and normal tissues prior to arbitrarily primed PCR amplification. Fragments that show differential methylation are cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments are then used as probes for Southern analysis to confirm differential methylation of these regions. Typical reagents (e.g., as might be found in a typical MCA-based kit) for MCA analysis may include, but are not limited to: PCR

primers for arbitrary priming Genomic DNA; PCR buffers and nucleotides, restriction enzymes and appropriate buffers; gene-hybridization oligos or probes; control hybridization oligos or probes.

Another method for analyzing methylation sites is a primer extension assay, including an optimized PCR amplification reaction that produces amplified targets for subsequent primer extension genotyping analysis using mass spectrometry. The assay can also be done in multiplex. This method (particularly as it relates to genotyping single nucleotide polymorphisms) is described in detail in PCT publication WO05012578A1 and US publication US20050079521A1. For methylation analysis, the assay can be adopted to detect bisulfite introduced methylation dependent C to T sequence changes. These methods are particularly useful for performing multiplexed amplification reactions and multiplexed primer extension reactions (e.g., multiplexed homogeneous primer mass extension (hME) assays) in a single well to further increase the throughput and reduce the cost per reaction for primer extension reactions.

Four additional methods for DNA methylation analysis include restriction landmark genomic scanning (RLGS, Costello et al., 2000), methylation-sensitive-representational difference analysis (MS-RDA), methylation-specific AP-PCR (MS-AP-PCR) and methyl-CpG binding domain column/segregation of partly melted molecules (MBD/SPM).

Additional methylation analysis methods that may be used in conjunction with the present technology are described in the following papers: Laird, P.W. Nature Reviews Cancer 3, 253-266 (2003); Biotechniques; Uhlmann, K. et al. Electrophoresis 23:4072-4079 (2002) - PyroMeth; Colella et al. Biotechniques. 2003 Jul;35(1):146-50; Dupont JM, Tost J, Jammes H, and Gut IG. Anal Biochem, Oct 2004; 333(1): 119-27; and Tooke N and Pettersson M. IVDT. Nov 2004; 41.

Nucleic Acid Quantification

In some embodiments, the amount of fetal nucleic acid in a sample is determined. In some cases, the amount of fetal nucleic acid is determined based on a quantification of sequence read counts described herein. Quantification may be achieved by direct counting of sequence reads covering particular methylation sites and/or target sites, or by competitive PCR (i.e., co-amplification of competitor oligonucleotides of known quantity, as described herein). The term "amount" as used herein with respect to nucleic acids refers to any suitable measurement, including, but not limited

to, absolute amount (e.g., copy number), relative amount (e.g., fraction or ratio), weight (e.g., grams), and concentration (e.g., grams per unit volume (e.g., milliliter); molar units).

Fraction Determination

In some embodiments, a fraction or ratio can be determined for the amount of one nucleic acid relative to the amount of another nucleic acid. In some embodiments, the fraction of fetal nucleic acid in a sample relative to the total amount of nucleic acid in the sample is determined. To calculate the fraction of fetal nucleic acid in a sample relative to the total amount of the nucleic acid in the sample, the following equation can be applied:

The fraction of fetal nucleic acid = (amount of fetal nucleic acid) / [(amount of total nucleic acid)].

Copy number Determination using Competitors

In some embodiments, the absolute amount (e.g., copy number) of fetal nucleic acid is determined. Often, the copy number of fetal nucleic acid is determined based on the amount of a competitor oligonucleotide used. In some embodiments, the copy number of maternal nucleic acid is determined. To calculate the copy number of fetal nucleic acid in a sample, the following equation can be applied:

Copy number (fetal nucleic acid) = [(amount of the fetal nucleic acid) /(amount of the fetal competitor)] x C

where C is the number of competitor oligonucleotides added into the reaction. In some cases, the amounts of the fetal nucleic acid and fetal competitor are obtained in a readout generated by a sequencing reaction (e.g., sequence read counts).

Additional Methods for Determining Fetal Nucleic Acid Content

The amount of fetal nucleic acid (e.g., concentration, relative amount, absolute amount, copy number, and the like) in nucleic acid is determined in some embodiments. In certain embodiments, the amount of fetal nucleic acid in a sample is referred to as “fetal fraction”. In certain embodiments “fetal fraction” refers to the fraction of fetal nucleic acid in circulating cell-free nucleic

acid in a sample (e.g., a blood sample, a serum sample, a plasma sample) obtained from a pregnant female. In some embodiments a fetal fraction is a percentage of fetal nucleic in a sample comprising fetal and maternal nucleic acid. In certain embodiments, the amount of fetal nucleic acid is determined according to markers specific to a male fetus (e.g., Y-chromosome STR markers (e.g., DYS 19, DYS 385, DYS 392 markers); RhD marker in RhD-negative females), allelic ratios of polymorphic sequences, or according to one or more markers specific to fetal nucleic acid and not maternal nucleic acid (e.g., differential epigenetic biomarkers (e.g., methylation; described in further detail below) between mother and fetus, or fetal RNA markers in maternal blood plasma (see e.g., Lo, 2005, Journal of Histochemistry and Cytochemistry 53 (3): 293-296)).

Determination of fetal nucleic acid content (e.g., fetal fraction) sometimes is performed using a fetal quantifier assay (FQA) as described, for example, in U.S. Patent Application Publication No. 2010/0105049, which is hereby incorporated by reference. This type of assay allows for the detection and quantification of fetal nucleic acid in a maternal sample based on the methylation status of the nucleic acid in the sample. In certain embodiments, the amount of fetal nucleic acid from a maternal sample can be determined relative to the total amount of nucleic acid present, thereby providing the percentage of fetal nucleic acid in the sample. In certain embodiments, the copy number of fetal nucleic acid can be determined in a maternal sample. In certain embodiments, the amount of fetal nucleic acid can be determined in a sequence-specific (or locus-specific) manner and sometimes with sufficient sensitivity to allow for accurate chromosomal dosage analysis (for example, to detect the presence or absence of a genetic variation, e.g., a fetal aneuploidy).

A fetal quantifier assay (FQA) can be performed in conjunction with any of the methods described herein. Such an assay can be performed by any method known in the art and/or described in U.S. Patent Application Publication No. 2010/0105049, such as, for example, by a method that can distinguish between maternal and fetal DNA based on differential methylation status, and quantify (i.e. determine the amount of) the fetal DNA. Methods for differentiating nucleic acid based on methylation status include, but are not limited to, methylation sensitive capture, for example, using a MBD2-Fc fragment in which the methyl binding domain of MBD2 is fused to the Fc fragment of an antibody (MBD-FC) (Gebhard et al. (2006) Cancer Res. 66(12):6118-28); methylation specific antibodies; bisulfite conversion methods, for example, MSP (methylation-sensitive PCR), COBRA, methylation-sensitive single nucleotide primer extension (Ms-SNuPE) or Sequenom MassCLEAVE™ technology; and the use of methylation sensitive restriction enzymes (e.g.,

digestion of maternal DNA in a maternal sample using one or more methylation sensitive restriction enzymes thereby enriching the fetal DNA). Methyl-sensitive enzymes also can be used to differentiate nucleic acid based on methylation status, which, for example, can preferentially or substantially cleave or digest at their DNA recognition sequence if the latter is non-methylated. Thus, an unmethylated DNA sample will be cut into smaller fragments than a methylated DNA sample and a hypermethylated DNA sample will not be cleaved. Except where explicitly stated, any method for differentiating nucleic acid based on methylation status can be used with the compositions and methods of the technology herein. The amount of fetal DNA can be determined, for example, by introducing one or more competitors at known concentrations during an amplification reaction. Determining the amount of fetal DNA also can be done, for example, by RT-PCR, primer extension, sequencing and/or counting. In certain instances, the amount of nucleic acid can be determined using BEAMing technology as described in U.S. Patent Application Publication No. 2007/0065823. In certain embodiments, the restriction efficiency can be determined and the efficiency rate is used to further determine the amount of fetal DNA.

In some embodiments, the determination of fetal fraction or determining the amount of fetal nucleic acid is not required or necessary for identifying the presence or absence of a chromosome aneuploidy. In some embodiments, identifying the presence or absence of a chromosome aneuploidy does not require the sequence differentiation of fetal versus maternal DNA. In some cases this is because the summed contribution of both maternal and fetal sequences in a particular chromosome, chromosome portion or segment thereof is analyzed. In some embodiments, identifying the presence or absence of a chromosome aneuploidy does not rely on a priori sequence information that would distinguish fetal DNA from maternal DNA.

Polymorphism-based fetal quantifier assay

Determination of fetal nucleic acid content (e.g., fetal fraction) sometimes is performed using a polymorphism-based fetal quantifier assay (FQA), as described herein. This type of assay allows for the detection and quantification of fetal nucleic acid in a maternal sample based on allelic ratios of polymorphic sequences (e.g., single nucleotide polymorphisms (SNPs)). In some cases, nucleotide sequence reads are obtained for a maternal sample and fetal fraction is determined by comparing the total number of nucleotide sequence reads that map to a first allele and the total number of nucleotide sequence reads that map to a second allele at an informative polymorphic site (e.g., SNP) in a reference genome. In some cases, fetal alleles are identified, for example, by

their relative minor contribution to the mixture of fetal and maternal nucleic acids in the sample when compared to the major contribution to the mixture by the maternal nucleic acids. In some cases, fetal alleles are identified by a deviation of allele frequency from an expected allele frequency, as described below. In some cases, the relative abundance of fetal nucleic acid in a maternal sample can be determined as a parameter of the total number of unique sequence reads mapped to a target nucleic acid sequence on a reference genome for each of the two alleles of a polymorphic site. In some cases, the relative abundance of fetal nucleic acid in a maternal sample can be determined as a parameter of the relative number of sequence reads for each allele from an enriched sample.

In some embodiments, determining fetal fraction comprises enriching a sample nucleic acid for one or more polymorphic nucleic acid targets. In some cases, a plurality of polymorphic targets is enriched. A plurality of polymorphic nucleic acid targets is sometimes referred to as a collection or a panel (e.g., target panel, SNP panel, SNP collection). A plurality of polymorphic targets can comprise two or more targets. For example, a plurality of polymorphic targets can comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more targets. In some cases, 10 or more polymorphic nucleic acid targets are enriched. In some cases, 50 or more polymorphic nucleic acid targets are enriched. In some cases, 100 or more polymorphic nucleic acid targets are enriched. In some cases, 500 or more polymorphic nucleic acid targets are enriched. In some cases, about 10 to about 500 polymorphic nucleic acid targets are enriched. In some cases, about 20 to about 400 polymorphic nucleic acid targets are enriched. In some cases, about 30 to about 200 polymorphic nucleic acid targets are enriched. In some cases, about 40 to about 100 polymorphic nucleic acid targets are enriched. In some cases, about 60 to about 90 polymorphic nucleic acid targets are enriched. For example, in certain embodiments, about 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 or 90 polymorphic nucleic acid targets are enriched.

In some embodiments, at least one polymorphic nucleic acid target of the plurality of polymorphic nucleic acid targets is informative for determining fetal fraction in a given sample. A polymorphic nucleic acid target that is informative for determining fetal fraction, sometimes referred to as an informative target, informative polymorphism, or informative SNP, typically differs in some aspect between the fetus and the mother. For example, an informative target may have one allele for the fetus and a different allele for the mother (e.g., the mother has allele A at the polymorphic target and the fetus has allele B at the polymorphic target site). Typically, a fetal allele that differs from

either of the maternal alleles is paternally inherited (i.e., is from the father). Thus, paternally inherited alleles that differ from maternal alleles can be useful for identifying and/or quantifying fetal nucleic acid (e.g., determining fetal fraction).

In some cases, polymorphic nucleic acid targets are informative in the context of certain maternal/fetal genotype combinations. For a biallelic polymorphic target (i.e., two possible alleles (e.g., A and B)), possible maternal/fetal genotype combinations include: 1) maternal AA, fetal AA; 2) maternal AA, fetal AB; 3) maternal AB, fetal AA; 4) maternal AB, fetal AB; 5) maternal AB; fetal BB; 6) maternal BB, fetal AB; and 7) maternal BB, fetal BB. Genotypes AA and BB are considered homozygous genotypes and genotype AB is considered a heterozygous genotype. In some cases, informative genotype combinations (i.e., genotype combinations for a polymorphic nucleic acid target that may be informative for determining fetal fraction) include combinations where the mother is homozygous and the fetus is heterozygous (e.g., maternal AA, fetal AB; or maternal BB, fetal AB). Such genotype combinations may be referred to as Type 1 informative genotypes or informative heterozygotes. In some cases, informative genotype combinations (i.e., genotype combinations for a polymorphic nucleic acid target that may be informative for determining fetal fraction) include combinations where the mother is heterozygous and the fetus is homozygous (e.g., maternal AB, fetal AA; or maternal AB, fetal BB). Such genotype combinations may be referred to as Type 2 informative genotypes or informative homozygotes. In some cases, non-informative genotype combinations (i.e., genotype combinations for a polymorphic nucleic acid target that may not be informative for determining fetal fraction) include combinations where the mother is heterozygous and the fetus is heterozygous (e.g., maternal AB, fetal AB). Such genotype combinations may be referred to as non-informative genotypes or non-informative heterozygotes. In some cases, non-informative genotype combinations (i.e., genotype combinations for a polymorphic nucleic acid target that may not be informative for determining fetal fraction) include combinations where the mother is homozygous and the fetus is homozygous (e.g., maternal AA, fetal AA; or maternal BB, fetal BB). Such genotype combinations may be referred to as non-informative genotypes or non-informative homozygotes.

In some embodiments, individual polymorphic nucleic acid targets and/or panels of polymorphic nucleic acid targets are selected based on certain criteria, such as, for example, minor allele population frequency, variance, coefficient of variance, MAD value, and the like. In some cases, polymorphic nucleic acid targets are selected so that at least one polymorphic nucleic acid target within a panel of polymorphic targets has a high probability of being informative for a majority of

samples tested. Additionally, in some cases, the number of polymorphic nucleic acid targets (i.e., number of targets in a panel) is selected so that least one polymorphic nucleic acid target has a high probability of being informative for a majority of samples tested. For example, selection of a larger number of polymorphic targets generally increases the probability that least one polymorphic nucleic acid target will be informative for a majority of samples tested (see, Figure 37, for example). In some cases, the polymorphic nucleic acid targets and number thereof (e.g., number of polymorphic targets selected for enrichment) result in at least about 2 to about 50 or more polymorphic nucleic acid targets being informative for determining the fetal fraction for at least about 80% to about 100% of samples. For example, the polymorphic nucleic acid targets and number thereof result in at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more polymorphic nucleic acid targets being informative for determining the fetal fraction for at least about 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of samples. In some cases, the polymorphic nucleic acid targets and number thereof result in at least five polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 90% of samples. In some cases, the polymorphic nucleic acid targets and number thereof result in at least five polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 95% of samples. In some cases, the polymorphic nucleic acid targets and number thereof result in at least five polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 99% of samples. In some cases, the polymorphic nucleic acid targets and number thereof result in at least ten polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 90% of samples. In some cases, the polymorphic nucleic acid targets and number thereof result in at least ten polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 95% of samples. In some cases, the polymorphic nucleic acid targets and number thereof result in at least ten polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 99% of samples.

In some embodiments, individual polymorphic nucleic acid targets are selected based, in part, on minor allele population frequency. In some cases, polymorphic nucleic acid targets having minor allele population frequencies of about 10% to about 50% are selected. For example, polymorphic nucleic acid targets having minor allele population frequencies of about 15%, 20%, 25%, 30%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, or 49% are selected. In some embodiments, polymorphic nucleic acid targets having a minor allele population frequency of about 40% or more are selected.

In some embodiments, individual polymorphic nucleic acid targets and/or panels of polymorphic nucleic acid targets are selected based, in part, on degree of variance for an individual polymorphic target or a panel of polymorphic targets. Variance, in come cases, can be specific for certain polymorphic targets or panels of polymorphic targets and can be from systematic, experimental, procedural, and or inherent errors or biases (e.g., sampling errors, sequencing errors, PCR bias, and the like). Variance of an individual polymorphic target or a panel of polymorphic targets can be determined by any method known in the art for assessing variance and may be expressed, for example, in terms of a calculated variance, an error, standard deviation, p-value, mean absolute deviation, median absolute deviation, median adjusted deviation (MAD score), coefficient of variance (CV), and the like. In some embodiments, measured allele frequency variance (i.e., background allele frequency) for certain SNPs (when homozygous, for example) can be from about 0.001 to about 0.01 (i.e., 0.1% to about 1.0%). For example, measured allele frequency variance can be about 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, or 0.009. In some cases, measured allele frequency variance is about 0.007.

In some cases, noisy polymorphic targets are excluded from a panel of polymorphic nucleic acid targets selected for determining fetal fraction. The term “noisy polymorphic targets” or “noisy SNPs” refers to (a) targets or SNPs that have significant variance between data points (e.g., measured fetal fraction, measured allele frequency) when analyzed or plotted, (b) targets or SNPs that have significant standard deviation (e.g., greater than 1, 2, or 3 standard deviations), (c) targets or SNPs that have a significant standard error of the mean, the like, and combinations of the foregoing. Noise for certain polymorphic targets or SNPs sometimes occurs due to the quantity and/or quality of starting material (e.g., nucleic acid sample), sometimes occurs as part of processes for preparing or replicating DNA used to generate sequence reads, and sometimes occurs as part of a sequencing process. In certain embodiments, noise for some polymorphic targets or SNPs results from certain sequences being over represented when prepared using PCR-based methods. In some cases, noise for some polymorphic targets or SNPs results from one or more inherent characteristics of the site such as, for example, certain nucleotide sequences and/or base compositions surrounding, or being adjacent to, a polymorphic target or SNP. A SNP having a measured allele frequency variance (when homozygous, for example) of about 0.005 or more may be considered noisy. For example, a SNP having a measured allele frequency variance of about 0.006, 0.007, 0.008, 0.009, 0.01 or more may be considered noisy.

In some embodiments, variance of an individual polymorphic target or a panel of polymorphic targets can be represented using coefficient of variance (CV). Coefficient of variance (i.e., standard deviation divided by the mean) can be determined, for example, by determining fetal fraction for several aliquots of a single maternal sample comprising maternal and fetal nucleic acid, and calculating the mean fetal fraction and standard deviation. In some cases, individual polymorphic nucleic acid targets and/or panels of polymorphic nucleic acid targets are selected so that fetal fraction is determined with a coefficient of variance (CV) of 0.30 or less. For example, fetal fraction may determined with a coefficient of variance (CV) of 0.25, 0.20, 0.19, 0.18, 0.17, 0.16, 0.15, 0.14, 0.13, 0.12, 0.11, 0.10, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01 or less, in some embodiments. In some cases, fetal fraction is determined with a coefficient of variance (CV) of 0.20 or less. In some cases, fetal fraction is determined with a coefficient of variance (CV) of 0.10 or less. In some cases, fetal fraction is determined with a coefficient of variance (CV) of 0.05 or less.

In some embodiments, an allele frequency is determined for each of the polymorphic nucleic acid targets in a sample. This sometimes is referred to as measured allele frequency. Allele frequency can be determined, for example, by counting the number of sequence reads for an allele (e.g., allele B) and dividing by the total number of sequence reads for that locus (e.g., allele B + allele A). In some cases, an allele frequency average, mean or median is determined. Fetal fraction can be determined based on the allele frequency mean (e.g., allele frequency mean multiplied by two), in some cases.

In some embodiments, determining whether a polymorphic nucleic acid target is informative comprises comparing its measured allele frequency to a fixed cutoff frequency. In some cases, determining which polymorphic nucleic acid targets are informative comprises identifying informative genotypes by comparing each allele frequency to one or more fixed cutoff frequencies. Fixed cutoff frequencies may be predetermined threshold values based on one or more qualifying data sets, for example. In some cases, the fixed cutoff for identifying informative genotypes from non-informative genotypes is expressed as a percent (%) shift in allele frequency from an expected allele frequency. Generally, expected allele frequencies for a given allele (e.g., allele A) are 0 (for a BB genotype), 0.5 (for an AB genotype) and 1.0 (for an AA genotype), or equivalent values on any numerical scale. A deviation from an expected allele frequency that is beyond one or more fixed cutoff frequencies may be considered informative. The degree of deviation generally is

proportional to fetal fraction (i.e., large deviations from expected allele frequency may be observed in samples having high fetal fraction).

In some cases, the fixed cutoff for identifying informative genotypes from non-informative homozygotes is about a 0.5% or greater shift in allele frequency. For example, a fixed cutoff may be about a 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 3%, 4%, 5%, 10% or greater shift in allele frequency. In some cases, the fixed cutoff for identifying informative genotypes from non-informative homozygotes is about a 1% or greater shift in allele frequency. In some cases, the fixed cutoff for identifying informative genotypes from non-informative homozygotes is about a 2% or greater shift in allele frequency. In some embodiments, the fixed cutoff for identifying informative genotypes from non-informative heterozygotes is about a 10% or greater shift in allele frequency. For example, a fixed cutoff may be about a 10%, 15%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 70%, 80% or greater shift in allele frequency. In some cases, the fixed cutoff for identifying informative genotypes from non-informative heterozygotes is about a 25% or greater shift in allele frequency. In some cases, the fixed cutoff for identifying informative genotypes from non-informative heterozygotes is about a 50% or greater shift in allele frequency.

In some embodiments, determining whether a polymorphic nucleic acid target is informative comprises comparing its measured allele frequency to a target-specific cutoff value. In some embodiments, target-specific cutoff frequencies are determined for each polymorphic nucleic acid target. Typically, target-specific cutoff frequency is determined based on the allele frequency variance for the corresponding polymorphic nucleic acid target. In some embodiments, variance of individual polymorphic targets can be represented by a median absolute deviation (MAD), for example. In some cases, determining a MAD value for each polymorphic nucleic acid target can generate unique (i.e., target-specific) cutoff values. To determine median absolute deviation, measured allele frequency can be determined, for example, for multiple replicates (e.g., 5, 6, 7, 8, 9, 10, 15, 20 or more replicates) of a maternal only nucleic acid sample (e.g., buffy coat sample). Each polymorphic target in each replicate will typically have a slightly different measured allele frequency due to PCR and/or sequencing errors, for example. A median allele frequency value can be identified for each polymorphic target. A deviation from the median for the remaining replicates can be calculated (i.e., the difference between the observed allele frequency and the median allele frequency). The absolute value of the deviations (i.e., negative values become positive) is taken and the median value of the absolute deviations is calculated to provide a median

absolute deviation (MAD) for each polymorphic nucleic acid target. A target-specific cutoff can be assigned, for example, as a multiple of the MAD (e.g., 1xMAD, 2xMAD, 3xMAD, 4xMAD or 5xMAD). Typically, polymorphic targets having less variance have a lower MAD and therefore a lower cutoff value than more variable targets.

In some embodiments, enriching comprises amplifying the plurality of polymorphic nucleic acid targets. In some cases, the enriching comprises generating amplification products in an amplification reaction. Amplification of polymorphic targets may be achieved by any method described herein or known in the art for amplifying nucleic acid (e.g., PCR). In some cases, the amplification reaction is performed in a single vessel (e.g., tube, container, well on a plate) which sometimes is referred to herein as multiplexed amplification.

In some embodiments, certain parental genotypes are known prior to the enriching of polymorphic nucleic acid targets. In some cases, the maternal genotype for one or more polymorphic targets is known prior to enriching. In some cases, the paternal genotype for one or more polymorphic targets is known prior to enriching. In some cases, the maternal genotype and the paternal genotype for one or more polymorphic targets are known prior to enriching. In some embodiments, certain parental genotypes are not known prior to the enriching of polymorphic nucleic acid targets. In some cases, the maternal genotype for one or more polymorphic targets is not known prior to enriching. In some cases, the paternal genotype for one or more polymorphic targets is not known prior to enriching. In some cases, the maternal genotype and the paternal genotype for one or more polymorphic targets are not known prior to enriching. In some embodiments, parental genotypes are not known for any of the polymorphic nucleic acid targets prior to enriching. In some cases, the maternal genotype for each of the polymorphic targets is not known prior to enriching. In some cases, the paternal genotype for each of the polymorphic targets is not known prior to enriching. In some cases, the maternal genotype and the paternal genotype for each of the polymorphic targets are not known prior to enriching.

In some embodiments, the polymorphic nucleic acid targets each comprise at least one single nucleotide polymorphism (SNP). In some embodiments, the SNPs are selected from: rs10413687, rs10949838, rs1115649, rs11207002, rs11632601, rs11971741, rs12660563, rs13155942, rs1444647, rs1572801, rs17773922, rs1797700, rs1921681, rs1958312, rs196008, rs2001778, rs2323659, rs2427099, rs243992, rs251344, rs254264, rs2827530, rs290387, rs321949, rs348971, rs390316, rs3944117, rs425002, rs432586, rs444016, rs4453265, rs447247,

rs4745577, rs484312, rs499946, rs500090, rs500399, rs505349, rs505662, rs516084, rs517316, rs517914, rs522810, rs531423, rs537330, rs539344, rs551372, rs567681,rs585487, rs600933, rs619208, rs622994, rs639298, rs642449, rs6700732, rs677866, rs683922, rs686851, rs6941942, rs7045684, rs7176924, rs7525374, rs870429, rs949312, rs9563831, rs970022, rs985462, rs1005241, rs1006101, rs10745725, rs10776856, rs10790342, rs11076499, rs11103233, rs11133637, rs11974817, rs12102203, rs12261, rs12460763, rs12543040, rs12695642, rs13137088, rs13139573, rs1327501, rs13438255, rs1360258, rs1421062, rs1432515, rs1452396, rs1518040, rs16853186, rs1712497, rs1792205, rs1863452, rs1991899, rs2022958, rs2099875, rs2108825, rs2132237, rs2195979, rs2248173, rs2250246, rs2268697, rs2270893, rs244887, rs2736966, rs2851428, rs2906237, rs2929724, rs3742257, rs3764584, rs3814332, rs4131376, rs4363444, rs4461567, rs4467511, rs4559013, rs4714802, rs4775899, rs4817609, rs488446, rs4950877, rs530913, rs6020434, rs6442703, rs6487229, rs6537064, rs654065, rs6576533, rs6661105, rs669161, rs6703320, rs675828, rs6814242, rs6989344, rs7120590, rs7131676, rs7214164, rs747583, rs768255, rs768708, rs7828904, rs7899772, rs7900911, rs7925270, rs7975781, rs8111589, rs849084, rs873870, rs9386151, rs9504197, rs9690525, and rs9909561.

In some embodiments, the SNPs are selected from: rs10413687, rs10949838, rs1115649, rs11207002, rs11632601, rs11971741, rs12660563, rs13155942, rs1444647, rs1572801, rs17773922, rs1797700, rs1921681, rs1958312, rs196008, rs2001778, rs2323659, rs2427099, rs243992, rs251344, rs254264, rs2827530, rs290387, rs321949, rs348971, rs390316, rs3944117, rs425002, rs432586, rs444016, rs4453265, rs447247, rs4745577, rs484312, rs499946, rs500090, rs500399, rs505349, rs505662, rs516084, rs517316, rs517914, rs522810, rs531423, rs537330, rs539344, rs551372, rs567681,rs585487, rs600933, rs619208, rs622994, rs639298, rs642449, rs6700732, rs677866, rs683922, rs686851, rs6941942, rs7045684, rs7176924, rs7525374, rs870429, rs949312, rs9563831, rs970022, and rs985462.

In some embodiments, SNPs are selected from: rs1005241, rs1006101, rs10745725, rs10776856, rs10790342, rs11076499, rs11103233, rs11133637, rs11974817, rs12102203, rs12261, rs12460763, rs12543040, rs12695642, rs13137088, rs13139573, rs1327501, rs13438255, rs1360258, rs1421062, rs1432515, rs1452396, rs1518040, rs16853186, rs1712497, rs1792205, rs1863452, rs1991899, rs2022958, rs2099875, rs2108825, rs2132237, rs2195979, rs2248173, rs2250246, rs2268697, rs2270893, rs244887, rs2736966, rs2851428, rs2906237, rs2929724, rs3742257, rs3764584, rs3814332, rs4131376, rs4363444, rs4461567, rs4467511, rs4559013, rs4714802, rs4775899, rs4817609, rs488446, rs4950877, rs530913, rs6020434, rs6442703,

rs6487229, rs6537064, rs654065, rs6576533, rs6661105, rs669161, rs6703320, rs675828, rs6814242, rs6989344, rs7120590, rs7131676, rs7214164, rs747583, rs768255, rs768708, rs7828904, rs7899772, rs7900911, rs7925270, rs7975781, rs8111589, rs849084, rs873870, rs9386151, rs9504197, rs9690525, and rs9909561.

SNPs may be selected from any SNP provided herein or known in the art that meets any one or all of the criteria described herein for SNP selection. In some cases, SNPs can be located on any chromosome (e.g., chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X and/or Y). In some cases, SNPs can be located on autosomes (e.g., chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22), and not on chromosome X or chromosome Y. In some cases, SNPs can be located on certain autosomes (e.g., chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 22 and not chromosome 13, 18 or 22). In some cases, SNPs can be located on certain chromosomes suspected of having a genetic variation (e.g., aneuploidy), such as, for example, chromosome 13, 18, 21, X and/or Y (i.e., test chromosome(s)). In some cases, SNPs are located on a reference chromosome. In some cases, fetal fraction and the presence or absence of a genetic variation (e.g., aneuploidy) are determined simultaneously using a method provided herein.

In some embodiments, enriched (e.g., amplified) polymorphic nucleic acid targets are sequenced by a sequencing process. In some cases, the sequencing process is a sequencing by synthesis method, as described herein. Typically, sequencing by synthesis methods comprise a plurality of synthesis cycles, whereby a complementary nucleotide is added to a single stranded template and identified during each cycle. The number of cycles generally corresponds to read length. In some cases, polymorphic targets are selected such that a minimal read length (i.e., minimal number of cycles) is required to include amplification primer sequence and the polymorphic target site (e.g., SNP) in the read. In some cases, amplification primer sequence includes about 10 to about 30 nucleotides. For example, amplification primer sequence may include about 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or 29 nucleotides, in some embodiments. In some cases, amplification primer sequence includes about 20 nucleotides. In some embodiments, a SNP site is located within 1 nucleotide base position (i.e., adjacent to) to about 30 base positions from the 3' terminus of an amplification primer. For example, a SNP site may be within 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, or 29 nucleotides of an amplification primer terminus. Read lengths can be any length that is inclusive of an amplification primer sequence and a polymorphic sequence or position. In some embodiments,

read lengths can be about 10 nucleotides in length to about 50 nucleotides in length. For example, read lengths can be about 15, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, or 45 nucleotides in length. In some cases, read length is about 36 nucleotides. In some cases, read length is about 27 nucleotides. Thus, in some cases, the sequencing by synthesis method comprises about 36 cycles and sometimes comprises about 27 cycles.

In some embodiments, a plurality of samples is sequenced in a single compartment (e.g., flow cell), which sometimes is referred to herein as sample multiplexing. Thus, in some embodiments, fetal fraction is determined for a plurality of samples in a multiplexed assay. For example, fetal fraction may be determined for about 10, 20, 30, 40 , 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000 or more samples. In some cases, fetal fraction is determined for about 10 or more samples. In some cases, fetal fraction is determined for about 100 or more samples. In some cases, fetal fraction is determined for about 1000 or more samples.

Determination of fetal nucleic acid content (e.g., fetal fraction) sometimes is performed using a methylation-based fetal quantifier assay (FQA) as described herein and, for example, in U.S. Patent Application Publication No. 2010/0105049, which is hereby incorporated by reference. In certain embodiments, a fetal quantifier assay (FQA) can be used to determine the concentration of fetal DNA in a maternal sample, for example, by the following method: a) determine the total amount of DNA present in a maternal sample; b) selectively digest the maternal DNA in a maternal sample using one or more methylation sensitive restriction enzymes thereby enriching the fetal DNA; c) determine the amount of fetal DNA from step b); and d) compare the amount of fetal DNA from step c) to the total amount of DNA from step a), thereby determining the concentration of fetal DNA in the maternal sample. In certain embodiments, the absolute copy number of fetal nucleic acid in a maternal sample can be determined, for example, using mass spectrometry and/or a system that uses a competitive PCR approach for absolute copy number measurements. See for example, Ding and Cantor (2003) PNAS USA 100:3059-3064, and U.S. Patent Application Publication No. 2004/0081993, both of which are hereby incorporated by reference.

In certain embodiments, fetal fraction can be determined based on allelic ratios of polymorphic sequences (e.g., single nucleotide polymorphisms (SNPs)), such as, for example, using a method described in U.S. Patent Application Publication No. 2011/0224087, which is hereby incorporated by reference. In such a method, nucleotide sequence reads are obtained for a maternal sample and fetal fraction is determined by comparing the total number of nucleotide sequence reads that

map to a first allele and the total number of nucleotide sequence reads that map to a second allele at an informative polymorphic site (e.g., SNP) in a reference genome. In certain embodiments, fetal alleles are identified, for example, by their relative minor contribution to the mixture of fetal and maternal nucleic acids in the sample when compared to the major contribution to the mixture by the maternal nucleic acids. Accordingly, the relative abundance of fetal nucleic acid in a maternal sample can be determined as a parameter of the total number of unique sequence reads mapped to a target nucleic acid sequence on a reference genome for each of the two alleles of a polymorphic site.

The amount of fetal nucleic acid in extracellular nucleic acid (e.g., fetal fraction) can be quantified and used in conjunction with other methods for assessing a genetic variation (e.g., fetal aneuploidy, fetal gender). Thus, in certain embodiments, methods for determining the presence or absence of a genetic variation, for example, comprise an additional step of determining the amount of fetal nucleic acid. The amount of fetal nucleic acid can be determined in a nucleic acid sample from a subject before or after processing to prepare sample nucleic acid. In certain embodiments, the amount of fetal nucleic acid is determined in a sample after sample nucleic acid is processed and prepared, which amount is utilized for further assessment. In some embodiments, an outcome comprises factoring the fraction of fetal nucleic acid in the sample nucleic acid (e.g., adjusting counts, removing samples, making a call or not making a call).

The determination of fetal nucleic acid content (e.g., fetal fraction) can be performed before, during, at any one point in a method described herein, for assessing a genetic variation (e.g., aneuploidy detection, fetal gender determination), or after such methods described herein. For example, to achieve a fetal gender or aneuploidy determination method with a given sensitivity or specificity, a fetal nucleic acid quantification method may be implemented prior to, during or after fetal gender or aneuploidy determination to identify those samples with greater than about 2%, 2.5%, 3%,3.5%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%,15%,16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25% or more fetal nucleic acid. In some embodiments, samples determined as having a certain threshold amount of fetal nucleic acid (e.g., about 15% or more fetal nucleic acid; about 4% or more fetal nucleic acid) are further analyzed for fetal gender or aneuploidy determination, or the presence or absence of an aneuploidy or genetic variation, for example. In certain embodiments, determinations of, for example, fetal gender or the presence or absence of aneuploidy are selected (e.g., selected and communicated to a patient) only for

samples having a certain threshold amount of fetal nucleic acid (e.g., about 15% or more fetal nucleic acid; about 4% or more fetal nucleic acid).

Methods for enriching for a subpopulation of nucleic acid

In some embodiments, nucleic acid (e.g., extracellular nucleic acid) is enriched or relatively enriched for a subpopulation or species of nucleic acid. Nucleic acid subpopulations can include, for example, fetal nucleic acid, maternal nucleic acid, nucleic acid comprising fragments of a particular length or range of lengths, or nucleic acid from a particular genome region (e.g., single chromosome, set of chromosomes, and/or certain chromosome regions). Such enriched samples can be used in conjunction with a method provided herein. Thus, in certain embodiments, methods of the technology comprise an additional step of enriching for a subpopulation of nucleic acid in a sample, such as, for example, fetal nucleic acid. In certain embodiments, a method for determining fetal fraction described above also can be used to enrich for fetal nucleic acid. In certain embodiments, maternal nucleic acid is selectively removed (partially, substantially, almost completely or completely) from the sample. In certain embodiments, enriching for a particular low copy number species nucleic acid (e.g., fetal nucleic acid) may improve quantitative sensitivity. Methods for enriching a sample for a particular species of nucleic acid are described, for example, in United States Patent No. 6,927,028, International Patent Application Publication No. WO2007/140417, International Patent Application Publication No. WO2007/147063, International Patent Application Publication No. WO2009/032779, International Patent Application Publication No. WO2009/032781, International Patent Application Publication No. WO2010/033639, International Patent Application Publication No. WO2011/034631, International Patent Application Publication No. WO2006/056480, and International Patent Application Publication No. WO2011/143659, all of which are incorporated by reference herein.

In some embodiments, nucleic acid is enriched for certain target fragment species and/or reference fragment species. In certain embodiments, nucleic acid is enriched for a specific nucleic acid fragment length or range of fragment lengths using one or more length-based separation methods described below. In certain embodiments, nucleic acid is enriched for fragments from a select genomic region (e.g., chromosome) using one or more sequence-based separation methods described herein and/or known in the art. Certain methods for enriching for a nucleic acid subpopulation (e.g., fetal nucleic acid) in a sample are described in detail below.

Some methods for enriching for a nucleic acid subpopulation (e.g., fetal nucleic acid) that can be used with a method described herein include methods that exploit epigenetic differences between maternal and fetal nucleic acid. For example, fetal nucleic acid can be differentiated and separated from maternal nucleic acid based on methylation differences. Methylation-based fetal nucleic acid enrichment methods are described in U.S. Patent Application Publication No. 2010/0105049, which is incorporated by reference herein. Such methods sometimes involve binding a sample nucleic acid to a methylation-specific binding agent (methyl-CpG binding protein (MBD), methylation specific antibodies, and the like) and separating bound nucleic acid from unbound nucleic acid based on differential methylation status. Such methods also can include the use of methylation-sensitive restriction enzymes (as described above; e.g., HhaI and HpaII), which allow for the enrichment of fetal nucleic acid regions in a maternal sample by selectively digesting nucleic acid from the maternal sample with an enzyme that selectively and completely or substantially digests the maternal nucleic acid to enrich the sample for at least one fetal nucleic acid region.

Another method for enriching for a nucleic acid subpopulation (e.g., fetal nucleic acid) that can be used with a method described herein is a restriction endonuclease enhanced polymorphic sequence approach, such as a method described in U.S. Patent Application Publication No. 2009/0317818, which is incorporated by reference herein. Such methods include cleavage of nucleic acid comprising a non-target allele with a restriction endonuclease that recognizes the nucleic acid comprising the non-target allele but not the target allele; and amplification of uncleaved nucleic acid but not cleaved nucleic acid, where the uncleaved, amplified nucleic acid represents enriched target nucleic acid (e.g., fetal nucleic acid) relative to non-target nucleic acid (e.g., maternal nucleic acid). In certain embodiments, nucleic acid may be selected such that it comprises an allele having a polymorphic site that is susceptible to selective digestion by a cleavage agent, for example.

Some methods for enriching for a nucleic acid subpopulation (e.g., fetal nucleic acid) that can be used with a method described herein include selective enzymatic degradation approaches. Such methods involve protecting target sequences from exonuclease digestion thereby facilitating the elimination in a sample of undesired sequences (e.g., maternal DNA). For example, in one approach, sample nucleic acid is denatured to generate single stranded nucleic acid, single stranded nucleic acid is contacted with at least one target-specific primer pair under suitable annealing conditions, annealed primers are extended by nucleotide polymerization generating

double stranded target sequences, and digesting single stranded nucleic acid using a nuclease that digests single stranded (i.e. non-target) nucleic acid. In certain embodiments, the method can be repeated for at least one additional cycle. In certain embodiments, the same target-specific primer pair is used to prime each of the first and second cycles of extension, and in some cases, different target-specific primer pairs are used for the first and second cycles.

Some methods for enriching for a nucleic acid subpopulation (e.g., fetal nucleic acid) that can be used with a method described herein include massively parallel signature sequencing (MPSS) approaches. MPSS typically is a solid phase method that uses adapter (i.e. tag) ligation, followed by adapter decoding, and reading of the nucleic acid sequence in small increments. Tagged PCR products are typically amplified such that each nucleic acid generates a PCR product with a unique tag. Tags are often used to attach the PCR products to microbeads. After several rounds of ligation-based sequence determination, for example, a sequence signature can be identified from each bead. Each signature sequence (MPSS tag) in a MPSS dataset is analyzed, compared with all other signatures, and all identical signatures are counted.

In certain embodiments, certain MPSS-based enrichment methods can include amplification (e.g., PCR)-based approaches. In certain embodiments, loci-specific amplification methods can be used (e.g., using loci-specific amplification primers). In certain embodiments, a multiplex SNP allele PCR approach can be used. In certain embodiments, a multiplex SNP allele PCR approach can be used in combination with uniplex sequencing. For example, such an approach can involve the use of multiplex PCR (e.g., MASSARRAY system) and incorporation of capture probe sequences into the amplicons followed by sequencing using, for example, the Illumina MPSS system. In certain embodiments, a multiplex SNP allele PCR approach can be used in combination with a three-primer system and indexed sequencing. For example, such an approach can involve the use of multiplex PCR (e.g., MASSARRAY system) with primers having a first capture probe incorporated into certain loci-specific forward PCR primers and adapter sequences incorporated into loci-specific reverse PCR primers, to thereby generate amplicons, followed by a secondary PCR to incorporate reverse capture sequences and molecular index barcodes for sequencing using, for example, the Illumina MPSS system. In certain embodiments, a multiplex SNP allele PCR approach can be used in combination with a four-primer system and indexed sequencing. For example, such an approach can involve the use of multiplex PCR (e.g., MASSARRAY system) with primers having adaptor sequences incorporated into both loci-specific forward and loci-specific reverse PCR primers, followed by a secondary PCR to incorporate both forward and reverse

capture sequences and molecular index barcodes for sequencing using, for example, the Illumina MPSS system. In certain embodiments, a microfluidics approach can be used. In certain embodiments, an array-based microfluidics approach can be used. For example, such an approach can involve the use of a microfluidics array (e.g., Fluidigm) for amplification at low plex and incorporation of index and capture probes, followed by sequencing. In certain embodiments, an emulsion microfluidics approach can be used, such as, for example, digital droplet PCR.

In certain embodiments, universal amplification methods can be used (e.g., using universal or non-loci-specific amplification primers). In certain embodiments, universal amplification methods can be used in combination with pull-down approaches. In certain embodiments, a method can include biotinylated ultramer pull-down (e.g., biotinylated pull-down assays from Agilent or IDT) from a universally amplified sequencing library. For example, such an approach can involve preparation of a standard library, enrichment for selected regions by a pull-down assay, and a secondary universal amplification step. In certain embodiments, pull-down approaches can be used in combination with ligation-based methods. In certain embodiments, a method can include biotinylated ultramer pull down with sequence specific adapter ligation (e.g., HALOPLEX PCR, Halo Genomics). For example, such an approach can involve the use of selector probes to capture restriction enzyme-digested fragments, followed by ligation of captured products to an adaptor, and universal amplification followed by sequencing. In certain embodiments, pull-down approaches can be used in combination with extension and ligation-based methods. In certain embodiments, a method can include molecular inversion probe (MIP) extension and ligation. For example, such an approach can involve the use of molecular inversion probes in combination with sequence adapters followed by universal amplification and sequencing. In certain embodiments, complementary DNA can be synthesized and sequenced without amplification.

In certain embodiments, extension and ligation approaches can be performed without a pull-down component. In certain embodiments, a method can include loci-specific forward and reverse primer hybridization, extension and ligation. Such methods can further include universal amplification or complementary DNA synthesis without amplification, followed by sequencing. Such methods can reduce or exclude background sequences during analysis, in some cases.

In certain embodiments, pull-down approaches can be used with an optional amplification component or with no amplification component. In certain embodiments, a method can include a modified pull-down assay and ligation with full incorporation of capture probes without universal

amplification. For example, such an approach can involve the use of modified selector probes to capture restriction enzyme-digested fragments, followed by ligation of captured products to an adaptor, optional amplification, and sequencing. In certain embodiments, a method can include a biotinylated pull-down assay with extension and ligation of adaptor sequence in combination with circular single stranded ligation. For example, such an approach can involve the use of selector probes to capture regions of interest (i.e. target sequences), extension of the probes, adaptor ligation, single stranded circular ligation, optional amplification, and sequencing. In certain embodiments, the analysis of the sequencing result can separate target sequences form background.

In some embodiments, nucleic acid is enriched for fragments from a select genomic region (e.g., chromosome) using one or more sequence-based separation methods described herein. Sequence-based separation generally is based on nucleotide sequences present in the fragments of interest (e.g., target and/or reference fragments) and substantially not present in other fragments of the sample or present in an insubstantial amount of the other fragments (e.g., 5% or less). In some embodiments, sequence-based separation can generate separated target fragments and/or separated reference fragments. Separated target fragments and/or separated reference fragments typically are isolated away from the remaining fragments in the nucleic acid sample. In certain embodiments, the separated target fragments and the separated reference fragments also are isolated away from each other (e.g., isolated in separate assay compartments). In certain embodiments, the separated target fragments and the separated reference fragments are isolated together (e.g., isolated in the same assay compartment). In some embodiments, unbound fragments can be differentially removed or degraded or digested.

In some embodiments, a selective nucleic acid capture process is used to separate target and/or reference fragments away from the nucleic acid sample. Commercially available nucleic acid capture systems include, for example, Nimblegen sequence capture system (Roche NimbleGen, Madison, WI); Illumina BEADARRAY platform (Illumina, San Diego, CA); Affymetrix GENECHIP platform (Affymetrix, Santa Clara, CA); Agilent SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA); and related platforms. Such methods typically involve hybridization of a capture oligonucleotide to a segment or all of the nucleotide sequence of a target or reference fragment and can include use of a solid phase (e.g., solid phase array) and/or a solution based platform. Capture oligonucleotides (sometimes referred to as “bait”) can be selected or designed such that they preferentially hybridize to nucleic acid fragments from selected

genomic regions or loci (e.g., a chromosome, e.g., one of chromosomes 21, 18, 13, X or Y, or a reference chromosome).

In some embodiments, nucleic acid is enriched for a particular nucleic acid fragment length, range of lengths, or lengths under or over a particular threshold or cutoff using one or more length-based separation methods. Nucleic acid fragment length typically refers to the number of nucleotides in the fragment. Nucleic acid fragment length also is sometimes referred to as nucleic acid fragment size. In some embodiments, a length-based separation method is performed without measuring lengths of individual fragments. In some embodiments, a length based separation method is performed in conjunction with a method for determining length of individual fragments. In some embodiments, length-based separation refers to a size fractionation procedure where all or part of the fractionated pool can be isolated (e.g., retained) and/or analyzed. Size fractionation procedures are known in the art (e.g., separation on an array, separation by a molecular sieve, separation by gel electrophoresis, separation by column chromatography (e.g., size-exclusion columns), and microfluidics-based approaches). In certain embodiments, length-based separation approaches can include fragment circularization, chemical treatment (e.g., formaldehyde, polyethylene glycol (PEG)), mass spectrometry and/or size-specific nucleic acid amplification, for example.

Certain length-based separation methods that can be used with methods described herein employ a selective sequence tagging approach, for example. The term “sequence tagging” refers to incorporating a recognizable and distinct sequence into a nucleic acid or population of nucleic acids. The term “sequence tagging” as used herein has a different meaning than the term “sequence tag” described later herein. In such sequence tagging methods, a fragment size species (e.g., short fragments) nucleic acids are subjected to selective sequence tagging in a sample that includes long and short nucleic acids. Such methods typically involve performing a nucleic acid amplification reaction using a set of nested primers which include inner primers and outer primers. In certain embodiments, one or both of the inner can be tagged to thereby introduce a tag onto the target amplification product. The outer primers generally do not anneal to the short fragments that carry the (inner) target sequence. The inner primers can anneal to the short fragments and generate an amplification product that carries a tag and the target sequence. Typically, tagging of the long fragments is inhibited through a combination of mechanisms which include, for example, blocked extension of the inner primers by the prior annealing and extension of the outer primers. Enrichment for tagged fragments can be accomplished by any of a variety of

methods, including for example, exonuclease digestion of single stranded nucleic acid and amplification of the tagged fragments using amplification primers specific for at least one tag.

Another length-based separation method that can be used with methods described herein involves subjecting a nucleic acid sample to polyethylene glycol (PEG) precipitation. Examples of methods include those described in International Patent Application Publication Nos. WO2007/140417 and WO2010/115016. This method in general entails contacting a nucleic acid sample with PEG in the presence of one or more monovalent salts under conditions sufficient to substantially precipitate large nucleic acids without substantially precipitating small (e.g., less than 300 nucleotides) nucleic acids.

Another size-based enrichment method that can be used with methods described herein involves circularization by ligation, for example, using circligase. Short nucleic acid fragments typically can be circularized with higher efficiency than long fragments. Non-circularized sequences can be separated from circularized sequences, and the enriched short fragments can be used for further analysis.

Nucleic Acid Amplification and Detection

Following separation of nucleic acid in a methylation-differential manner, nucleic acid may be amplified and/or subjected to a detection process (e.g., sequence-based analysis, mass spectrometry). Furthermore, once it is determined that one particular genomic sequence of fetal origin is hypermethylated or hypomethylated compared to the maternal counterpart, the amount of this fetal genomic sequence can be determined. Subsequently, this amount can be compared to a standard control value and serve as an indication for the potential of certain pregnancy-associated disorder.

Nucleotide sequences, or amplified nucleic acid sequences, or detectable products prepared from the foregoing, can be detected by a suitable detection process. Non-limiting examples of methods of detection, quantification, sequencing and the like include mass detection of mass modified amplicons (e.g., matrix-assisted laser desorption ionization (MALDI) mass spectrometry and electrospray (ES) mass spectrometry), a primer extension method (e.g., iPLEX™; Sequenom, Inc.), direct DNA sequencing, Molecular Inversion Probe (MIP) technology from

Affymetrix, restriction fragment length polymorphism (RFLP analysis), allele specific oligonucleotide (ASO) analysis, methylation-specific PCR (MSPCR), pyrosequencing analysis, acycloprime analysis, Reverse dot blot, GeneChip microarrays, Dynamic allele-specific hybridization (DASH), Peptide nucleic acid (PNA) and locked nucleic acids (LNA) probes, TaqMan, Molecular Beacons, Intercalating dye, FRET primers, AlphaScreen, SNPstream, genetic bit analysis (GBA), Multiplex minisequencing, SNaPshot, GOOD assay, Microarray miniseq, arrayed primer extension (APEX), Microarray primer extension, Tag arrays, Coded microspheres, Template-directed incorporation (TDI), fluorescence polarization, Colorimetric oligonucleotide ligation assay (OLA), Sequence-coded OLA, Microarray ligation, Ligase chain reaction, Padlock probes, Invader assay, hybridization using at least one probe, hybridization using at least one fluorescently labeled probe, cloning and sequencing, electrophoresis, the use of hybridization probes and quantitative real time polymerase chain reaction (QRT-PCR), digital PCR, nanopore sequencing, chips and combinations thereof. In some embodiments the amount of each amplified nucleic acid species is determined by mass spectrometry, primer extension, sequencing (e.g., any suitable method, for example nanopore or pyrosequencing), Quantitative PCR (Q-PCR or QRT-PCR), digital PCR, combinations thereof, and the like.

Nucleic acid detection and/or quantification also may include, for example, solid support array based detection of fluorescently labeled nucleic acid with fluorescent labels incorporated during or after PCR, single molecule detection of fluorescently labeled molecules in solution or captured on a solid phase, or other sequencing technologies such as, for example, sequencing using ION TORRENT or MISEQ platforms or single molecule sequencing technologies using instrumentation such as, for example, PACBIO sequencers, HELICOS sequencer, or nanopore sequencing technologies.

In some cases, nucleotide sequences, or amplified nucleic acid sequences, or detectable products prepared from the foregoing, are detected using a sequencing process (e.g., such as a sequencing process described herein). Nucleic acid quantifications generated by a method comprising a sequencing detection process may be compared to nucleic acid quantifications generated by a method comprising a different detection process (e.g., mass spectrometry). Such comparisons may be expressed using an R^2 value, which is a measure of correlation between two outcomes (e.g., nucleic acid quantifications). In some cases, nucleic acid quantifications (e.g., fetal copy number quantifications) are highly correlated (i.e., have high R^2 values) for quantifications generated using different detection processes (e.g., sequencing and mass spectrometry). In some

cases, R^2 values for nucleic acid quantifications generated using different detection processes may be between about 0.90 and about 1.0. For example, R^2 values may be about 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, or 0.99.

Amplification of Nucleotide Sequences

In many instances, it is desirable to amplify a nucleic acid sequence of the technology herein using any of several nucleic acid amplification procedures which are well known in the art (listed above and described in greater detail below). Specifically, nucleic acid amplification is the enzymatic synthesis of nucleic acid amplicons (copies) which contain a sequence that is complementary to a nucleic acid sequence being amplified. Nucleic acid amplification is especially beneficial when the amount of target sequence present in a sample is very low. By amplifying the target sequences and detecting the amplicon synthesized, the sensitivity of an assay can be vastly improved, since fewer target sequences are needed at the beginning of the assay to better ensure detection of nucleic acid in the sample belonging to the organism or virus of interest.

A variety of polynucleotide amplification methods are well established and frequently used in research. For instance, the general methods of polymerase chain reaction (PCR) for polynucleotide sequence amplification are well known in the art and are thus not described in detail herein. For a review of PCR methods, protocols, and principles in designing primers, see, e.g., Innis, et al., PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc. N.Y., 1990. PCR reagents and protocols are also available from commercial vendors, such as Roche Molecular Systems.

PCR is most usually carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer annealing region, and an extension reaction region automatically. Machines specifically adapted for this purpose are commercially available.

Although PCR amplification of a polynucleotide sequence is typically used in practicing the present technology, one of skill in the art will recognize that the amplification of a genomic sequence found in a maternal blood sample may be accomplished by any known method, such as ligase chain reaction (LCR), transcription-mediated amplification, and self-sustained sequence replication or nucleic acid sequence-based amplification (NASBA), each of which provides sufficient

amplification. More recently developed branched-DNA technology may also be used to qualitatively demonstrate the presence of a particular genomic sequence of the technology herein, which represents a particular methylation pattern, or to quantitatively determine the amount of this particular genomic sequence in the maternal blood. For a review of branched-DNA signal amplification for direct quantitation of nucleic acid sequences in clinical samples, see Nolte, Adv. Clin. Chem. 33:201-235, 1998.

The compositions and processes of the technology herein are also particularly useful when practiced with digital PCR. Digital PCR was first developed by Kalinina and colleagues (Kalinina et al., "Nanoliter scale PCR with TaqMan detection." Nucleic Acids Research. 25; 1999-2004, (1997)) and further developed by Vogelstein and Kinzler (Digital PCR. PNAS USA. 96; 9236-41, (1999)). The application of digital PCR for use with fetal diagnostics was first described by Cantor et al. (PCT Patent Publication No. WO05023091A2) and subsequently described by Quake et al. (US Patent Publication No. US 20070202525), which are both hereby incorporated by reference. Digital PCR takes advantage of nucleic acid (DNA, cDNA or RNA) amplification on a single molecule level, and offers a highly sensitive method for quantifying low copy number nucleic acid. Fluidigm® Corporation offers systems for the digital analysis of nucleic acids.

The terms "amplify", "amplification", "amplification reaction", or "amplifying" refer to any in vitro process for multiplying the copies of a nucleic acid. Amplification sometimes refers to an "exponential" increase in nucleic acid. However, "amplifying" as used herein can also refer to linear increases in the numbers of a select nucleic acid, but is different than a one-time, single primer extension step. In some embodiments a limited amplification reaction, also known as pre-amplification, can be performed. Pre-amplification is a method in which a limited amount of amplification occurs due to a small number of cycles, for example 10 cycles, being performed. Pre-amplification can allow some amplification, but stops amplification prior to the exponential phase, and typically produces about 500 copies of the desired nucleotide sequence(s). Use of pre-amplification may also limit inaccuracies associated with depleted reactants in standard PCR reactions, for example, and also may reduce amplification biases due to nucleotide sequence or abundance of the nucleic acid. In some embodiments a one-time primer extension may be performed as a prelude to linear or exponential amplification.

Any suitable amplification technique can be utilized. Amplification of polynucleotides include, but are not limited to, polymerase chain reaction (PCR); ligation amplification (or ligase chain reaction

(LCR)); amplification methods based on the use of Q-beta replicase or template-dependent polymerase (see US Patent Publication Number US20050287592); helicase-dependent isothermal amplification (Vincent et al., "Helicase-dependent isothermal DNA amplification". EMBO reports 5 (8): 795–800 (2004)); strand displacement amplification (SDA); thermophilic SDA nucleic acid sequence based amplification (3SR or NASBA) and transcription-associated amplification (TAA). Non-limiting examples of PCR amplification methods include standard PCR, AFLP-PCR, Allele-specific PCR, Alu-PCR, Asymmetric PCR, Colony PCR, Hot start PCR, Inverse PCR (IPCR), In situ PCR (ISH), Intersequence-specific PCR (ISSR-PCR), Long PCR, Multiplex PCR, Nested PCR, Quantitative PCR, Reverse Transcriptase PCR (RT-PCR), Real Time PCR, Single cell PCR, Solid phase PCR, digital PCR, combinations thereof, and the like. For example, amplification can be accomplished using digital PCR, in certain embodiments (see e.g., Kalinina et al., “Nanoliter scale PCR with TaqMan detection.” Nucleic Acids Research. 25; 1999-2004, (1997); Vogelstein and Kinzler (Digital PCR. PNAS USA. 96; 9236-41, (1999); PCT Patent Publication No. WO05023091A2; US Patent Publication No. US 20070202525). Digital PCR takes advantage of nucleic acid (DNA, cDNA or RNA) amplification on a single molecule level, and offers a highly sensitive method for quantifying low copy number nucleic acid. Systems for digital amplification and analysis of nucleic acids are available (e.g., Fluidigm® Corporation). Reagents and hardware for conducting PCR are commercially available.

A generalized description of an amplification process is presented herein. Primers and nucleic acid are contacted, and complementary sequences anneal to one another, for example. Primers can anneal to a nucleic acid, at or near (e.g., adjacent to, abutting, and the like) a sequence of interest. In some embodiments, the primers in a set hybridize within about 10 to 30 nucleotides from a nucleic acid sequence of interest and produce amplified products. In some embodiments, the primers hybridize within the nucleic acid sequence of interest.

A reaction mixture, containing components necessary for enzymatic functionality, is added to the primer-nucleic acid hybrid, and amplification can occur under suitable conditions. Components of an amplification reaction may include, but are not limited to, e.g., primers (e.g., individual primers, primer pairs, primer sets and the like) a polynucleotide template, polymerase, nucleotides, dNTPs and the like. In some embodiments, non-naturally occurring nucleotides or nucleotide analogs, such as analogs containing a detectable label (e.g., fluorescent or colorimetric label), may be used for example. Polymerases can be selected by a person of ordinary skill and include polymerases for thermocycle amplification (e.g., Taq DNA Polymerase; Q-Bio™ Taq DNA Polymerase

(recombinant truncated form of Taq DNA Polymerase lacking 5'-3'exo activity); SurePrime™ Polymerase (chemically modified Taq DNA polymerase for "hot start" PCR); Arrow™ Taq DNA Polymerase (high sensitivity and long template amplification)) and polymerases for thermostable amplification (e.g., RNA polymerase for transcription-mediated amplification (TMA) described at World Wide Web URL "gen-probe.com/pdfs/tma_whiteppr.pdf"). Other enzyme components can be added, such as reverse transcriptase for transcription mediated amplification (TMA) reactions, for example.

PCR conditions can be dependent upon primer sequences, abundance of nucleic acid, and the desired amount of amplification, and therefore, one of skill in the art may choose from a number of PCR protocols available (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds, 1990. Digital PCR is also known in the art; see, e.g., United States Patent Application Publication no. 20070202525, filed February 2, 2007, which is hereby incorporated by reference). PCR is typically carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing step, a primer-annealing step, and an extension reaction step automatically. Some PCR protocols also include an activation step and a final extension step. Machines specifically adapted for this purpose are commercially available. A non-limiting example of a PCR protocol that may be suitable for embodiments described herein is, treating the sample at 95 °C for 5 minutes; repeating thirty-five cycles of 95 °C for 45 seconds and 68 °C for 30 seconds; and then treating the sample at 72 °C for 3 minutes. A completed PCR reaction can optionally be kept at 4 °C until further action is desired. Multiple cycles frequently are performed using a commercially available thermal cycler. Suitable isothermal amplification processes known and selected by the person of ordinary skill in the art also may be applied, in certain embodiments.

In some embodiments, an amplification product may include naturally occurring nucleotides, non-naturally occurring nucleotides, nucleotide analogs and the like and combinations of the foregoing. An amplification product often has a nucleotide sequence that is identical to or substantially identical to a nucleic acid sequence herein, or complement thereof. A "substantially identical" nucleotide sequence in an amplification product will generally have a high degree of sequence identity to the nucleotide sequence species being amplified or complement thereof (e.g., about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% sequence identity), and

variations sometimes are a result of infidelity of the polymerase used for extension and/or amplification, or additional nucleotide sequence(s) added to the primers used for amplification.

Primers

Primers useful for detection, amplification, quantification, sequencing and analysis of nucleic acid are provided. The term "primer" as used herein refers to a nucleic acid that includes a nucleotide sequence capable of hybridizing or annealing to a target nucleic acid, at or near (e.g., adjacent to) a specific region of interest. Primers can allow for specific determination of a target nucleic acid nucleotide sequence or detection of the target nucleic acid (e.g., presence or absence of a sequence or copy number of a sequence), or feature thereof, for example. A primer may be naturally occurring or synthetic. The term "specific" or "specificity", as used herein, refers to the binding or hybridization of one molecule to another molecule, such as a primer for a target polynucleotide. That is, "specific" or "specificity" refers to the recognition, contact, and formation of a stable complex between two molecules, as compared to substantially less recognition, contact, or complex formation of either of those two molecules with other molecules. As used herein, the term "anneal" refers to the formation of a stable complex between two molecules. The terms "primer", "oligo", or "oligonucleotide" may be used interchangeably throughout the document, when referring to primers.

A primer nucleic acid can be designed and synthesized using suitable processes, and may be of any length suitable for hybridizing to a nucleotide sequence of interest (e.g., where the nucleic acid is in liquid phase or bound to a solid support) and performing analysis processes described herein. Primers may be designed based upon a target nucleotide sequence. A primer in some embodiments may be about 10 to about 100 nucleotides, about 10 to about 70 nucleotides, about 10 to about 50 nucleotides, about 15 to about 30 nucleotides, or about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 nucleotides in length. A primer may be composed of naturally occurring and/or non-naturally occurring nucleotides (e.g., labeled nucleotides), or a mixture thereof. Primers suitable for use with embodiments described herein, may be synthesized and labeled using known techniques. Primers may be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Caruthers, Tetrahedron Letts., 22:1859-1862, 1981, using an automated synthesizer, as described in Needham-VanDevanter et al., Nucleic Acids Res. 12:6159-6168, 1984. Purification of primers can be effected by native acrylamide gel electrophoresis or by

anion-exchange high-performance liquid chromatography (HPLC), for example, as described in Pearson and Regnier, J. Chrom., 255:137-149, 1983.

All or a portion of a primer nucleic acid sequence (naturally occurring or synthetic) may be substantially complementary to a target nucleic acid, in some embodiments. As referred to herein, "substantially complementary" with respect to sequences refers to nucleotide sequences that will hybridize with each other. The stringency of the hybridization conditions can be altered to tolerate varying amounts of sequence mismatch. Included are target and primer sequences that are 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more, 79% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more or 99% or more complementary to each other.

Primers that are substantially complimentary to a target nucleic acid sequence are also substantially identical to the compliment of the target nucleic acid sequence. That is, primers are substantially identical to the anti-sense strand of the nucleic acid. As referred to herein, "substantially identical" with respect to sequences refers to nucleotide sequences that are 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more, 79% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more or 99% or more identical to each other. One test for determining whether two nucleotide sequences are substantially identical is to determine the percent of identical nucleotide sequences shared.

Primer sequences and length may affect hybridization to target nucleic acid sequences. Depending on the degree of mismatch between the primer and target nucleic acid, low, medium or high stringency conditions may be used to effect primer/target annealing. As used herein, the term "stringent conditions" refers to conditions for hybridization and washing. Methods for hybridization

reaction temperature condition optimization are known to those of skill in the art, and may be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. , 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. Non-limiting examples of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Stringent hybridization temperatures can also be altered (i.e. lowered) with the addition of certain organic solvents, formamide for example. Organic solvents, like formamide, reduce the thermal stability of double-stranded polynucleotides, so that hybridization can be performed at lower temperatures, while still maintaining stringent conditions and extending the useful life of nucleic acids that may be heat labile. Features of primers can be applied to probes and oligonucleotides, such as, for example, the competitive and inhibitory oligonucleotides provided herein.

As used herein, the phrase “hybridizing” or grammatical variations thereof, refers to binding of a first nucleic acid molecule to a second nucleic acid molecule under low, medium or high stringency conditions, or under nucleic acid synthesis conditions. Hybridizing can include instances where a first nucleic acid molecule binds to a second nucleic acid molecule, where the first and second nucleic acid molecules are complementary. As used herein, “specifically hybridizes” refers to preferential hybridization under nucleic acid synthesis conditions of a primer, to a nucleic acid molecule having a sequence complementary to the primer compared to hybridization to a nucleic acid molecule not having a complementary sequence. For example, specific hybridization includes the hybridization of a primer to a target nucleic acid sequence that is complementary to the primer.

In some embodiments primers can include a nucleotide subsequence that may be complementary to a solid phase nucleic acid primer hybridization sequence or substantially complementary to a solid phase nucleic acid primer hybridization sequence (e.g., about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99% or greater than 99% identical to the primer hybridization sequence complement when aligned). A primer may contain a nucleotide subsequence not complementary to or not substantially complementary to a solid phase nucleic acid primer hybridization sequence (e.g., at the 3' or 5' end of the nucleotide subsequence in the primer complementary to or substantially complementary to the solid phase primer hybridization sequence).

A primer, in certain embodiments, may contain a modification such as one or more inosines, abasic sites, locked nucleic acids, minor groove binders, duplex stabilizers (e.g., acridine, spermidine), Tm modifiers or any modifier that changes the binding properties of the primers or probes. A primer, in certain embodiments, may contain a detectable molecule or entity (e.g., a fluorophore, radioisotope, colorimetric agent, particle, enzyme and the like, as described above for labeled competitor oligonucleotides).

A primer also may refer to a polynucleotide sequence that hybridizes to a subsequence of a target nucleic acid or another primer and facilitates the detection of a primer, a target nucleic acid or both, as with molecular beacons, for example. The term "molecular beacon" as used herein refers to detectable molecule, where the detectable property of the molecule is detectable only under certain specific conditions, thereby enabling it to function as a specific and informative signal. Non-limiting examples of detectable properties are, optical properties, electrical properties, magnetic properties, chemical properties and time or speed through an opening of known size.

In some embodiments, the primers are complementary to genomic DNA target sequences. In some cases, the forward and reverse primers hybridize to the 5' and 3' ends of the genomic DNA target sequences. In some embodiments, primers that hybridize to the genomic DNA target sequences also hybridize to competitor oligonucleotides that were designed to compete with corresponding genomic DNA target sequences for binding of the primers. In some cases, the primers hybridize or anneal to the genomic DNA target sequences and the corresponding competitor oligonucleotides with the same or similar hybridization efficiencies. In some cases the hybridization efficiencies are different. The ratio between genomic DNA target amplicons and competitor amplicons can be measured during the reaction. For example if the ratio is 1:1 at 28 cycles but 2:1 at 35, this could indicate that during the end of the amplification reaction the primers for one target (i.e. genomic DNA target or competitor) are either reannealing faster than the other, or the denaturation is less effective than the other.

In some embodiments primers are used in sets. As used herein, an amplification primer set is one or more pairs of forward and reverse primers for a given region. Thus, for example, primers that amplify genomic targets for region 1 (i.e. targets 1a and 1b) are considered a primer set. Primers that amplify genomic targets for region 2 (i.e. targets 2a and 2b) are considered a different primer set. In some embodiments, the primer sets that amplify targets within a particular region also amplify the corresponding competitor oligonucleotide(s). A plurality of primer pairs may constitute a primer set in certain embodiments (e.g., about 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 pairs). In some embodiments a plurality of primer sets, each set comprising pair(s) of primers, may be used.

Determination of Polynucleotide Sequences

Techniques for polynucleotide sequence determination are also well established and widely practiced in the relevant research field. For instance, the basic principles and general techniques for polynucleotide sequencing are described in various research reports and treatises on molecular biology and recombinant genetics, such as Wallace et al., supra; Sambrook and Russell, supra, and Ausubel et al., supra. DNA sequencing methods routinely practiced in research laboratories, either manual or automated, can be used for practicing the present technology. Additional means suitable for detecting changes in a polynucleotide sequence for practicing the methods of the present technology include but are not limited to mass spectrometry, primer extension, polynucleotide hybridization, real-time PCR, and electrophoresis.

Use of a primer extension reaction also can be applied in methods of the technology herein. A primer extension reaction operates, for example, by discriminating the SNP alleles by the incorporation of deoxynucleotides and/or dideoxynucleotides to a primer extension primer which hybridizes to a region adjacent to the SNP site. The primer is extended with a polymerase. The primer extended SNP can be detected physically by mass spectrometry or by a tagging moiety such as biotin. As the SNP site is only extended by a complementary deoxynucleotide or dideoxynucleotide that is either tagged by a specific label or generates a primer extension product with a specific mass, the SNP alleles can be discriminated and quantified.

Reverse transcribed and amplified nucleic acids may be modified nucleic acids. Modified nucleic acids can include nucleotide analogs, and in certain embodiments include a detectable label and/or a capture agent. Examples of detectable labels include without limitation fluorophores,

radioisotopes, colormetric agents, light emitting agents, chemiluminescent agents, light scattering agents, enzymes and the like. Examples of capture agents include without limitation an agent from a binding pair selected from antibody/antigen, antibody/antibody, antibody/antibody fragment, antibody/antibody receptor, antibody/protein A or protein G, hapten/anti-hapten, biotin/avidin, biotin/streptavidin, folic acid/folate binding protein, vitamin B12/intrinsic factor, chemical reactive group/complementary chemical reactive group (e.g., sulfhydryl/maleimide, sulfhydryl/haloacetyl derivative, amine/isotriocyanate, amine/succinimidyl ester, and amine/sulfonyl halides) pairs, and the like. Modified nucleic acids having a capture agent can be immobilized to a solid support in certain embodiments

Mass spectrometry is a particularly effective method for the detection of a polynucleotide of the technology herein, for example a PCR amplicon, a primer extension product or a detector probe that is cleaved from a target nucleic acid. The presence of the polynucleotide sequence is verified by comparing the mass of the detected signal with the expected mass of the polynucleotide of interest. The relative signal strength, e.g., mass peak on a spectra, for a particular polynucleotide sequence indicates the relative population of a specific allele, thus enabling calculation of the allele ratio directly from the data. For a review of genotyping methods using Sequenom® standard iPLEX™ assay and MassARRAY® technology, see Jurinke, C., Oeth, P., van den Boom, D., "MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis." Mol. Biotechnol. 26, 147-164 (2004); and Oeth, P. et al., "iPLEX™ Assay: Increased Plexing Efficiency and Flexibility for MassARRAY® System through single base primer extension with mass-modified Terminators." SEQUENOM Application Note (2005), both of which are hereby incorporated by reference. For a review of detecting and quantifying target nucleic using cleavable detector probes that are cleaved during the amplification process and detected by mass spectrometry, see US Patent Application Number 11/950,395, which was filed December 4, 2007, and is hereby incorporated by reference.

Sequencing technologies are improving in terms of throughput and cost. Sequencing technologies, such as that achievable on the 454 platform (Roche) (Margulies, M. et al. 2005 Nature 437, 376-380), Illumina Genome Analyzer (or Solexa platform) or SOLiD System (Applied Biosystems) or the Helicos True Single Molecule DNA sequencing technology (Harris T D et al. 2008 Science, 320, 106-109), the single molecule, real-time (SMRT.TM.) technology of Pacific Biosciences, and nanopore sequencing (Soni GV and Meller A. 2007 Clin Chem 53: 1996-2001),

allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion (Dear Brief Funct Genomic Proteomic 2003; 1: 397-416).

Each of these platforms allow sequencing of clonally expanded or non-amplified single molecules of nucleic acid fragments. Certain platforms involve, for example, (i) sequencing by ligation of dye-modified probes (including cyclic ligation and cleavage), (ii) pyrosequencing, and (iii) single-molecule sequencing. Nucleotide sequence species, amplification nucleic acid species and detectable products generated there from can be considered a "study nucleic acid" for purposes of analyzing a nucleotide sequence by such sequence analysis platforms.

Sequencing by ligation is a nucleic acid sequencing method that relies on the sensitivity of DNA ligase to base-pairing mismatch. DNA ligase joins together ends of DNA that are correctly base paired. Combining the ability of DNA ligase to join together only correctly base paired DNA ends, with mixed pools of fluorescently labeled oligonucleotides or primers, enables sequence determination by fluorescence detection. Longer sequence reads may be obtained by including primers containing cleavable linkages that can be cleaved after label identification. Cleavage at the linker removes the label and regenerates the 5' phosphate on the end of the ligated primer, preparing the primer for another round of ligation. In some embodiments primers may be labeled with more than one fluorescent label (e.g., 1 fluorescent label, 2, 3, or 4 fluorescent labels). An example of a system that can be used by a person of ordinary skill based on sequencing by ligation generally involves the following steps. Clonal bead populations can be prepared in emulsion microreactors containing study nucleic acid ("template"), amplification reaction components, beads and primers. After amplification, templates are denatured and bead enrichment is performed to separate beads with extended templates from undesired beads (e.g., beads with no extended templates). The template on the selected beads undergoes a 3' modification to allow covalent bonding to the slide, and modified beads can be deposited onto a glass slide. Deposition chambers offer the ability to segment a slide into one, four or eight chambers during the bead loading process. For sequence analysis, primers hybridize to the adapter sequence. A set of four color dye-labeled probes competes for ligation to the sequencing primer. Specificity of probe ligation is achieved by interrogating every 4th and 5th base during the ligation series. Five to seven rounds of ligation, detection and cleavage record the color at every 5th position with the number of rounds determined by the type of library used. Following each round of ligation, a new complimentary primer offset by one base in the 5' direction is laid down for another series of ligations. Primer reset and ligation rounds (5–7 ligation cycles per round) are

repeated sequentially five times to generate 25–35 base pairs of sequence for a single tag. With mate-paired sequencing, this process is repeated for a second tag. Such a system can be used to exponentially amplify amplification products generated by a process described herein, e.g., by ligating a heterologous nucleic acid to the first amplification product generated by a process described herein and performing emulsion amplification using the same or a different solid support originally used to generate the first amplification product. Such a system also may be used to analyze amplification products directly generated by a process described herein by bypassing an exponential amplification process and directly sorting the solid supports described herein on the glass slide.

Pyrosequencing is a nucleic acid sequencing method based on sequencing by synthesis, which relies on detection of a pyrophosphate released on nucleotide incorporation. Generally, sequencing by synthesis involves synthesizing, one nucleotide at a time, a DNA strand complimentary to the strand whose sequence is being sought. Study nucleic acids may be immobilized to a solid support, hybridized with a sequencing primer, incubated with DNA polymerase, ATP sulfurylase, luciferase, apyrase, adenosine 5' phosphosulfate and luciferin. Nucleotide solutions are sequentially added and removed. Correct incorporation of a nucleotide releases a pyrophosphate, which interacts with ATP sulfurylase and produces ATP in the presence of adenosine 5' phosphosulfate, fueling the luciferin reaction, which produces a chemiluminescent signal allowing sequence determination.

An example of a system that can be used by a person of ordinary skill based on pyrosequencing generally involves the following steps: ligating an adaptor nucleic acid to a study nucleic acid and hybridizing the study nucleic acid to a bead; amplifying a nucleotide sequence in the study nucleic acid in an emulsion; sorting beads using a picoliter multiwell solid support; and sequencing amplified nucleotide sequences by pyrosequencing methodology (e.g., Nakano et al., "Single-molecule PCR using water-in-oil emulsion;" Journal of Biotechnology 102: 117-124 (2003)). Such a system can be used to exponentially amplify amplification products generated by a process described herein, e.g., by ligating a heterologous nucleic acid to the first amplification product generated by a process described herein.

Certain single-molecule sequencing embodiments are based on the principal of sequencing by synthesis, and utilize single-pair Fluorescence Resonance Energy Transfer (single pair FRET) as a mechanism by which photons are emitted as a result of successful nucleotide incorporation. The

emitted photons often are detected using intensified or high sensitivity cooled charge-couple-devices in conjunction with total internal reflection microscopy (TIRM). Photons are only emitted when the introduced reaction solution contains the correct nucleotide for incorporation into the growing nucleic acid chain that is synthesized as a result of the sequencing process. In FRET based single-molecule sequencing, energy is transferred between two fluorescent dyes, sometimes polymethine cyanine dyes Cy3 and Cy5, through long-range dipole interactions. The donor is excited at its specific excitation wavelength and the excited state energy is transferred, non-radiatively to the acceptor dye, which in turn becomes excited. The acceptor dye eventually returns to the ground state by radiative emission of a photon. The two dyes used in the energy transfer process represent the "single pair", in single pair FRET. Cy3 often is used as the donor fluorophore and often is incorporated as the first labeled nucleotide. Cy5 often is used as the acceptor fluorophore and is used as the nucleotide label for successive nucleotide additions after incorporation of a first Cy3 labeled nucleotide. The fluorophores generally are within 10 nanometers of each for energy transfer to occur successfully.

An example of a system that can be used based on single-molecule sequencing generally involves hybridizing a primer to a study nucleic acid to generate a complex; associating the complex with a solid phase; iteratively extending the primer by a nucleotide tagged with a fluorescent molecule; and capturing an image of fluorescence resonance energy transfer signals after each iteration (e.g., U.S. Patent No. 7,169,314; Braslavsky et al., PNAS 100(7): 3960-3964 (2003)). Such a system can be used to directly sequence amplification products generated by processes described herein. In some embodiments the released linear amplification product can be hybridized to a primer that contains sequences complementary to immobilized capture sequences present on a solid support, a bead or glass slide for example. Hybridization of the primer--released linear amplification product complexes with the immobilized capture sequences, immobilizes released linear amplification products to solid supports for single pair FRET based sequencing by synthesis. The primer often is fluorescent, so that an initial reference image of the surface of the slide with immobilized nucleic acids can be generated. The initial reference image is useful for determining locations at which true nucleotide incorporation is occurring. Fluorescence signals detected in array locations not initially identified in the "primer only" reference image are discarded as non-specific fluorescence. Following immobilization of the primer--released linear amplification product complexes, the bound nucleic acids often are sequenced in parallel by the iterative steps of, a) polymerase extension in the presence of one fluorescently labeled nucleotide, b) detection of

fluorescence using appropriate microscopy, TIRM for example, c) removal of fluorescent nucleotide, and d) return to step a with a different fluorescently labeled nucleotide.

In some embodiments, nucleotide sequencing may be by solid phase single nucleotide sequencing methods and processes. Solid phase single nucleotide sequencing methods involve contacting sample nucleic acid and solid support under conditions in which a single molecule of sample nucleic acid hybridizes to a single molecule of a solid support. Such conditions can include providing the solid support molecules and a single molecule of sample nucleic acid in a "microreactor." Such conditions also can include providing a mixture in which the sample nucleic acid molecule can hybridize to solid phase nucleic acid on the solid support. Single nucleotide sequencing methods useful in the embodiments described herein are described in United States Provisional Patent Application Serial Number 61/021,871 filed January 17, 2008.

In certain embodiments, nanopore sequencing detection methods include (a) contacting a nucleic acid for sequencing ("base nucleic acid," e.g., linked probe molecule) with sequence-specific detectors, under conditions in which the detectors specifically hybridize to substantially complementary subsequences of the base nucleic acid; (b) detecting signals from the detectors and (c) determining the sequence of the base nucleic acid according to the signals detected. In certain embodiments, the detectors hybridized to the base nucleic acid are disassociated from the base nucleic acid (e.g., sequentially dissociated) when the detectors interfere with a nanopore structure as the base nucleic acid passes through a pore, and the detectors disassociated from the base sequence are detected. In some embodiments, a detector disassociated from a base nucleic acid emits a detectable signal, and the detector hybridized to the base nucleic acid emits a different detectable signal or no detectable signal. In certain embodiments, nucleotides in a nucleic acid (e.g., linked probe molecule) are substituted with specific nucleotide sequences corresponding to specific nucleotides ("nucleotide representatives"), thereby giving rise to an expanded nucleic acid (e.g., U.S. Patent No. 6,723,513), and the detectors hybridize to the nucleotide representatives in the expanded nucleic acid, which serves as a base nucleic acid. In such embodiments, nucleotide representatives may be arranged in a binary or higher order arrangement (e.g., Soni and Meller, Clinical Chemistry 53(11): 1996-2001 (2007)). In some embodiments, a nucleic acid is not expanded, does not give rise to an expanded nucleic acid, and directly serves a base nucleic acid (e.g., a linked probe molecule serves as a non-expanded base nucleic acid), and detectors are directly contacted with the base nucleic acid. For example, a first detector may hybridize to a first subsequence and a second detector may hybridize to a second

subsequence, where the first detector and second detector each have detectable labels that can be distinguished from one another, and where the signals from the first detector and second detector can be distinguished from one another when the detectors are disassociated from the base nucleic acid. In certain embodiments, detectors include a region that hybridizes to the base nucleic acid (e.g., two regions), which can be about 3 to about 100 nucleotides in length (e.g., about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 55, 60, 65, 70, 75, 80, 85, 90, or 95 nucleotides in length). A detector also may include one or more regions of nucleotides that do not hybridize to the base nucleic acid. In some embodiments, a detector is a molecular beacon. A detector often comprises one or more detectable labels independently selected from those described herein. Each detectable label can be detected by any convenient detection process capable of detecting a signal generated by each label (e.g., magnetic, electric, chemical, optical and the like). For example, a CD camera can be used to detect signals from one or more distinguishable quantum dots linked to a detector.

In certain sequence analysis embodiments, reads may be used to construct a larger nucleotide sequence, which can be facilitated by identifying overlapping sequences in different reads and by using identification sequences in the reads. Such sequence analysis methods and software for constructing larger sequences from reads are known to the person of ordinary skill (e.g., Venter et al., Science 291: 1304-1351 (2001)). Specific reads, partial nucleotide sequence constructs, and full nucleotide sequence constructs may be compared between nucleotide sequences within a sample nucleic acid (i.e., internal comparison) or may be compared with a reference sequence (i.e., reference comparison) in certain sequence analysis embodiments. Internal comparisons sometimes are performed in situations where a sample nucleic acid is prepared from multiple samples or from a single sample source that contains sequence variations. Reference comparisons sometimes are performed when a reference nucleotide sequence is known and an objective is to determine whether a sample nucleic acid contains a nucleotide sequence that is substantially similar or the same, or different, than a reference nucleotide sequence. Sequence analysis is facilitated by sequence analysis apparatus and components known to the person of ordinary skill in the art.

Methods provided herein allow for high-throughput detection of nucleic acid species in a plurality of nucleic acids (e.g., nucleotide sequence species, amplified nucleic acid species and detectable products generated from the foregoing). Multiplexing refers to the simultaneous detection of more than one nucleic acid species. General methods for performing multiplexed reactions in

conjunction with mass spectrometry, are known (see, e.g., U.S. Pat. Nos. 6,043,031, 5,547,835 and International PCT application No. WO 97/37041). Multiplexing provides an advantage that a plurality of nucleic acid species (e.g., some having different sequence variations) can be identified in as few as a single mass spectrum, as compared to having to perform a separate mass spectrometry analysis for each individual target nucleic acid species. Methods provided herein lend themselves to high-throughput, highly-automated processes for analyzing sequence variations with high speed and accuracy, in some embodiments. In some embodiments, methods herein may be multiplexed at high levels in a single reaction.

In certain embodiments, the number of nucleic acid species multiplexed include, without limitation, about 1 to about 500 (e.g., about 1-3, 3-5, 5-7, 7-9, 9-11, 11-13, 13-15, 15-17, 17-19, 19-21, 21-23, 23-25, 25-27, 27-29, 29-31, 31-33, 33-35, 35-37, 37-39, 39-41, 41-43, 43-45, 45-47, 47-49, 49-51, 51-53, 53-55, 55-57, 57-59, 59-61, 61-63, 63-65, 65-67, 67-69, 69-71, 71-73, 73-75, 75-77, 77-79, 79-81, 81-83, 83-85, 85-87, 87-89, 89-91, 91-93, 93-95, 95-97, 97-101, 101-103, 103-105, 105-107, 107-109, 109-111, 111-113, 113-115, 115-117, 117-119, 121-123, 123-125, 125-127, 127-129, 129-131, 131-133, 133-135, 135-137, 137-139, 139-141, 141-143, 143-145, 145-147, 147-149, 149-151, 151-153, 153-155, 155-157, 157-159, 159-161, 161-163, 163-165, 165-167, 167-169, 169-171, 171-173, 173-175, 175-177, 177-179, 179-181, 181-183, 183-185, 185-187, 187-189, 189-191, 191-193, 193-195, 195-197, 197-199, 199-201, 201-203, 203-205, 205-207, 207-209, 209-211, 211-213, 213-215, 215-217, 217-219, 219-221, 221-223, 223-225, 225-227, 227-229, 229-231, 231-233, 233-235, 235-237, 237-239, 239-241, 241-243, 243-245, 245-247, 247-249, 249-251, 251-253, 253-255, 255-257, 257-259, 259-261, 261-263, 263-265, 265-267, 267-269, 269-271, 271-273, 273-275, 275-277, 277-279, 279-281, 281-283, 283-285, 285-287, 287-289, 289-291, 291-293, 293-295, 295-297, 297-299, 299-301, 301- 303, 303- 305, 305- 307, 307-309, 309- 311, 311- 313, 313- 315, 315- 317, 317- 319, 319-321, 321-323, 323-325, 325-327, 327-329, 329-331, 331-333, 333- 335, 335-337, 337-339, 339-341, 341-343, 343-345, 345-347, 347-349, 349-351, 351-353, 353-355, 355-357, 357-359, 359-361, 361-363, 363-365, 365-367, 367-369, 369-371, 371-373, 373-375, 375-377, 377-379, 379-381, 381-383, 383-385, 385-387, 387-389, 389-391, 391-393, 393-395, 395-397, 397-401, 401- 403, 403- 405, 405- 407, 407- 409, 409-411, 411- 413, 413- 415, 415- 417, 417- 419, 419-421, 421-423, 423-425, 425-427, 427-429, 429-431, 431-433, 433- 435, 435-437, 437-439, 439-441, 441-443, 443-445, 445-447, 447-449, 449-451, 451-453, 453-455, 455-457, 457-459, 459-461, 461-463, 463-465, 465-467, 467-469, 469-471, 471-473, 473-475, 475-477, 477-479, 479-481, 481-483, 483-485, 485-487, 487-489, 489-491, 491-493, 493-495, 495-497, 497-501).

Design methods for achieving resolved mass spectra with multiplexed assays can include primer and oligonucleotide design methods and reaction design methods. See, for example, the multiplex schemes provided in Tables X and Y. For primer and oligonucleotide design in multiplexed assays, the same general guidelines for primer design applies for uniplexed reactions, such as avoiding false priming and primer dimers, only more primers are involved for multiplex reactions. For mass spectrometry applications, analyte peaks in the mass spectra for one assay are sufficiently resolved from a product of any assay with which that assay is multiplexed, including pausing peaks and any other by-product peaks. Also, analyte peaks optimally fall within a user-specified mass window, for example, within a range of 5,000-8,500 Da. In some embodiments multiplex analysis may be adapted to mass spectrometric detection of chromosome abnormalities, for example. In certain embodiments multiplex analysis may be adapted to various single nucleotide or nanopore based sequencing methods described herein. Commercially produced micro-reaction chambers or devices or arrays or chips may be used to facilitate multiplex analysis, and are commercially available.

Additional methods for obtaining sequence reads

In some embodiments, nucleic acids (e.g., nucleic acid fragments, sample nucleic acid, cell-free nucleic acid) may be sequenced. In certain embodiments, a full or substantially full sequence is obtained and sometimes a partial sequence is obtained. Sequencing, mapping and related analytical methods are known in the art (e.g., United States Patent Application Publication US2009/0029377, incorporated by reference). Certain aspects of such processes are described hereafter.

As used herein, “reads” (i.e., “a read”, “a sequence read”) are short nucleotide sequences produced by any sequencing process described herein or known in the art. Reads can be generated from one end of nucleic acid fragments ("single-end reads"), and sometimes are generated from both ends of nucleic acids (e.g., paired-end reads, double-end reads).

In some embodiments the nominal, average, mean or absolute length of single-end reads sometimes is about 20 contiguous nucleotides to about 50 contiguous nucleotides, sometimes about 30 contiguous nucleotides to about 40 contiguous nucleotides, and sometimes about 35 contiguous nucleotides or about 36 contiguous nucleotides. In certain embodiments the nominal,

average, mean or absolute length of single-end reads is about 20 to about 30 bases in length. In certain embodiments the nominal, average, mean or absolute length of single-end reads is about 24 to about 28 bases in length. In certain embodiments the nominal, average, mean or absolute length of single-end reads is about 21, 22, 23, 24, 25, 26, 27, 28 or about 29 bases in length.

In certain embodiments, the nominal, average, mean or absolute length of the paired-end reads sometimes is about 10 contiguous nucleotides to about 25 contiguous nucleotides (e.g., about 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 nucleotides in length), sometimes is about 15 contiguous nucleotides to about 20 contiguous nucleotides, and sometimes is about 17 contiguous nucleotides or about 18 contiguous nucleotides.

Reads generally are representations of nucleotide sequences in a physical nucleic acid. For example, in a read containing an ATGC depiction of a sequence, "A" represents an adenine nucleotide, "T" represents a thymine nucleotide, "G" represents a guanine nucleotide and "C" represents a cytosine nucleotide, in a physical nucleic acid. Sequence reads obtained from the blood of a pregnant female can be reads from a mixture of fetal and maternal nucleic acid. A mixture of relatively short reads can be transformed by processes described herein into a representation of a genomic nucleic acid present in the pregnant female and/or in the fetus. A mixture of relatively short reads can be transformed into a representation of a copy number variation (e.g., a maternal and/or fetal copy number variation), genetic variation or an aneuploidy, for example. Reads of a mixture of maternal and fetal nucleic acid can be transformed into a representation of a composite chromosome or a segment thereof comprising features of one or both maternal and fetal chromosomes. In certain embodiments, “obtaining” nucleic acid sequence reads of a sample from a subject and/or “obtaining” nucleic acid sequence reads of a biological specimen from one or more reference persons can involve directly sequencing nucleic acid to obtain the sequence information. In some embodiments, “obtaining” can involve receiving sequence information obtained directly from a nucleic acid by another.

Sequence reads can be mapped and the number of reads or sequence tags mapping to a specified nucleic acid region (e.g., a chromosome, a bin, a genomic section) are referred to as counts. In some embodiments, counts can be manipulated or transformed (e.g., normalized, combined, added, filtered, selected, averaged, derived as a mean, the like, or a combination thereof). In some embodiments, counts can be transformed to produce normalized counts. Normalized counts for multiple genomic sections can be provided in a profile (e.g., a genomic

profile, a chromosome profile, a profile of a segment or portion of a chromosome). One or more different elevations in a profile also can be manipulated or transformed (e.g., counts associated with elevations can be normalized) and elevations can be adjusted.

In some embodiments, one nucleic acid sample from one individual is sequenced. In certain embodiments, nucleic acid samples from two or more biological samples, where each biological sample is from one individual or two or more individuals, are pooled and the pool is sequenced. In the latter embodiments, a nucleic acid sample from each biological sample often is identified by one or more unique identification tags.

In some embodiments, a fraction of the genome is sequenced, which sometimes is expressed in the amount of the genome covered by the determined nucleotide sequences (e.g., "fold" coverage less than 1). When a genome is sequenced with about 1-fold coverage, roughly 100% of the nucleotide sequence of the genome is represented by reads. A genome also can be sequenced with redundancy, where a given region of the genome can be covered by two or more reads or overlapping reads (e.g., "fold" coverage greater than 1). In some embodiments, a genome is sequenced with about 0.1-fold to about 100-fold coverage, about 0.2-fold to 20-fold coverage, or about 0.2-fold to about 1-fold coverage (e.g., about 0.2-, 0.3-, 0.4-, 0.5-, 0.6-, 0.7-, 0.8-, 0.9-, 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 15-, 20-, 30-, 40-, 50-, 60-, 70-, 80-, 90-fold coverage).

In certain embodiments, a fraction of a nucleic acid pool that is sequenced in a run is further sub-selected prior to sequencing. In certain embodiments, hybridization-based techniques (e.g., using oligonucleotide arrays) can be used to first sub-select for nucleic acid sequences from certain chromosomes (e.g., a potentially aneuploid chromosome and other chromosome(s) not involved in the aneuploidy tested). In some embodiments, nucleic acid can be fractionated by size (e.g., by gel electrophoresis, size exclusion chromatography or by microfluidics-based approach) and in certain instances, fetal nucleic acid can be enriched by selecting for nucleic acid having a lower molecular weight (e.g., less than 300 base pairs, less than 200 base pairs, less than 150 base pairs, less than 100 base pairs). In some embodiments, fetal nucleic acid can be enriched by suppressing maternal background nucleic acid, such as by the addition of formaldehyde. In some embodiments, a portion or subset of a pre-selected pool of nucleic acids is sequenced randomly. In some embodiments, the nucleic acid is amplified prior to sequencing. In some embodiments, a portion or subset of the nucleic acid is amplified prior to sequencing.

In certain embodiments, a sequencing library is prepared prior to or during a sequencing process. Methods for preparing a sequencing library are known in the art and commercially available platforms may be used for certain applications. Certain commercially available library platforms may be compatible with certain nucleotide sequencing processes described herein. For example, one or more commercially available library platforms may be compatible with a sequencing by synthesis process. In certain embodiments, a ligation-based library preparation method is used (e.g., ILLUMINA TRUSEQ, Illumina, San Diego CA). Ligation-based library preparation methods typically use a methylated adaptor design which can incorporate an index sequence at the initial ligation step and often can be used to prepare samples for single-read sequencing, paired-end sequencing and multiplexed sequencing. In certain embodiments, a transposon-based library preparation method is used (e.g., EPICENTRE NEXTERA, Epicentre, Madison WI). Transposon-based methods typically use in vitro transposition to simultaneously fragment and tag DNA in a single-tube reaction (often allowing incorporation of platform-specific tags and optional barcodes), and prepare sequencer-ready libraries.

Any sequencing method suitable for conducting methods described herein can be utilized. In some embodiments, a high-throughput sequencing method is used. High-throughput sequencing methods generally involve clonally amplified DNA templates or single DNA molecules that are sequenced in a massively parallel fashion within a flow cell (e.g., as described in Metzker M Nature Rev 11:31-46 (2010); Volkerding et al. Clin Chem 55:641-658 (2009)). Such sequencing methods also can provide digital quantitative information, where each sequence read is a countable "sequence tag" or "count" representing an individual clonal DNA template, a single DNA molecule, bin or chromosome. Next generation sequencing techniques capable of sequencing DNA in a massively parallel fashion are collectively referred to herein as "massively parallel sequencing" (MPS). High-throughput sequencing technologies include, for example, sequencing-by-synthesis with reversible dye terminators, sequencing by oligonucleotide probe ligation, pyrosequencing and real time sequencing. Non-limiting examples of MPS include Massively Parallel Signature Sequencing (MPSS), Polony sequencing, Pyrosequencing, Illumina (Solexa) sequencing, SOLiD sequencing, Ion semiconductor sequencing, DNA nanoball sequencing, Helioscope single molecule sequencing, single molecule real time (SMRT) sequencing, nanopore sequencing, ION Torrent and RNA polymerase (RNAP) sequencing.

Systems utilized for high-throughput sequencing methods are commercially available and include, for example, the Roche 454 platform, the Applied Biosystems SOLID platform, the Helicos True

Single Molecule DNA sequencing technology, the sequencing-by-hybridization platform from Affymetrix Inc., the single molecule, real-time (SMRT) technology of Pacific Biosciences, the sequencing-by-synthesis platforms from 454 Life Sciences, Illumina/Solexa and Helicos Biosciences, and the sequencing-by-ligation platform from Applied Biosystems. The ION TORRENT technology from Life technologies and nanopore sequencing also can be used in high-throughput sequencing approaches.

In some embodiments, first generation technology, such as, for example, Sanger sequencing including the automated Sanger sequencing, can be used in a method provided herein. Additional sequencing technologies that include the use of developing nucleic acid imaging technologies (e.g., transmission electron microscopy (TEM) and atomic force microscopy (AFM)), also are contemplated herein. Examples of various sequencing technologies are described below.

A nucleic acid sequencing technology that may be used in a method described herein is sequencing-by-synthesis and reversible terminator-based sequencing (e.g., Illumina's Genome Analyzer; Genome Analyzer II; HISEQ 2000; HISEQ 2500 (Illumina, San Diego CA)). With this technology, millions of nucleic acid (e.g., DNA) fragments can be sequenced in parallel. In one example of this type of sequencing technology, a flow cell is used which contains an optically transparent slide with 8 individual lanes on the surfaces of which are bound oligonucleotide anchors (e.g., adaptor primers). A flow cell often is a solid support that can be configured to retain and/or allow the orderly passage of reagent solutions over bound analytes. Flow cells frequently are planar in shape, optically transparent, generally in the millimeter or sub-millimeter scale, and often have channels or lanes in which the analyte/reagent interaction occurs.

In certain sequencing by synthesis procedures, for example, template DNA (e.g., circulating cell-free DNA (ccfDNA)) sometimes can be fragmented into lengths of several hundred base pairs in preparation for library generation. In some embodiments, library preparation can be performed without further fragmentation or size selection of the template DNA (e.g., ccfDNA). Sample isolation and library generation may be performed using automated methods and apparatus, in certain embodiments. Briefly, template DNA is end repaired by a fill-in reaction, exonuclease reaction or a combination of a fill-in reaction and exonuclease reaction. The resulting blunt-end repaired template DNA is extended by a single nucleotide, which is complementary to a single nucleotide overhang on the 3' end of an adapter primer, and often increases ligation efficiency. Any complementary nucleotides can be used for the extension/overhang nucleotides (e.g., A/T,

C/G), however adenine frequently is used to extend the end-repaired DNA, and thymine often is used as the 3' end overhang nucleotide.

In certain sequencing by synthesis procedures, for example, adapter oligonucleotides are complementary to the flow-cell anchors, and sometimes are utilized to associate the modified template DNA (e.g., end-repaired and single nucleotide extended) with a solid support, such as the inside surface of a flow cell, for example. In some embodiments, the adapter also includes identifiers (i.e., indexing nucleotides, or "barcode" nucleotides (e.g., a unique sequence of nucleotides usable as an identifier to allow unambiguous identification of a sample and/or chromosome)), one or more sequencing primer hybridization sites (e.g., sequences complementary to universal sequencing primers, single end sequencing primers, paired end sequencing primers, multiplexed sequencing primers, and the like), or combinations thereof (e.g., adapter/sequencing, adapter/identifier, adapter/identifier/sequencing). Identifiers or nucleotides contained in an adapter often are six or more nucleotides in length, and frequently are positioned in the adaptor such that the identifier nucleotides are the first nucleotides sequenced during the sequencing reaction. In certain embodiments, identifier nucleotides are associated with a sample but are sequenced in a separate sequencing reaction to avoid compromising the quality of sequence reads. Subsequently, the reads from the identifier sequencing and the DNA template sequencing are linked together and the reads de-multiplexed. After linking and de-multiplexing the sequence reads and/or identifiers can be further adjusted or processed as described herein.

In certain sequencing by synthesis procedures, utilization of identifiers allows multiplexing of sequence reactions in a flow cell lane, thereby allowing analysis of multiple samples per flow cell lane. The number of samples that can be analyzed in a given flow cell lane often is dependent on the number of unique identifiers utilized during library preparation and/or probe design. Non limiting examples of commercially available multiplex sequencing kits include Illumina's multiplexing sample preparation oligonucleotide kit and multiplexing sequencing primers and PhiX control kit (e.g., Illumina's catalog numbers PE-400-1001 and PE-400-1002, respectively). A method described herein can be performed using any number of unique identifiers (e.g., 4, 8, 12, 24, 48, 96, or more). The greater the number of unique identifiers, the greater the number of samples and/or chromosomes, for example, that can be multiplexed in a single flow cell lane. Multiplexing using 12 identifiers, for example, allows simultaneous analysis of 96 samples (e.g., equal to the number of wells in a 96 well microwell plate) in an 8 lane flow cell. Similarly,

multiplexing using 48 identifiers, for example, allows simultaneous analysis of 384 samples (e.g., equal to the number of wells in a 384 well microwell plate) in an 8 lane flow cell.

In certain sequencing by synthesis procedures, adapter-modified, single-stranded template DNA is added to the flow cell and immobilized by hybridization to the anchors under limiting-dilution conditions. In contrast to emulsion PCR, DNA templates are amplified in the flow cell by "bridge" amplification, which relies on captured DNA strands "arching" over and hybridizing to an adjacent anchor oligonucleotide. Multiple amplification cycles convert the single-molecule DNA template to a clonally amplified arching "cluster," with each cluster containing approximately 1000 clonal molecules. Approximately 50 x 10^6 separate clusters can be generated per flow cell. For sequencing, the clusters are denatured, and a subsequent chemical cleavage reaction and wash leave only forward strands for single-end sequencing. Sequencing of the forward strands is initiated by hybridizing a primer complementary to the adapter sequences, which is followed by addition of polymerase and a mixture of four differently colored fluorescent reversible dye terminators. The terminators are incorporated according to sequence complementarity in each strand in a clonal cluster. After incorporation, excess reagents are washed away, the clusters are optically interrogated, and the fluorescence is recorded. With successive chemical steps, the reversible dye terminators are unblocked, the fluorescent labels are cleaved and washed away, and the next sequencing cycle is performed. This iterative, sequencing-by-synthesis process sometimes requires approximately 2.5 days to generate read lengths of 36 bases. With 50 x 10^6 clusters per flow cell, the overall sequence output can be greater than 1 billion base pairs (Gb) per analytical run.

Another nucleic acid sequencing technology that may be used with a method described herein is 454 sequencing (Roche). 454 sequencing uses a large-scale parallel pyrosequencing system capable of sequencing about 400-600 megabases of DNA per run. The process typically involves two steps. In the first step, sample nucleic acid (e.g., DNA) is sometimes fractionated into smaller fragments (300-800 base pairs) and polished (made blunt at each end). Short adaptors are then ligated onto the ends of the fragments. These adaptors provide priming sequences for both amplification and sequencing of the sample-library fragments. One adaptor (Adaptor B) contains a 5'-biotin tag for immobilization of the DNA library onto streptavidin-coated beads. After nick repair, the non-biotinylated strand is released and used as a single-stranded template DNA (sstDNA) library. The sstDNA library is assessed for its quality and the optimal amount (DNA copies per bead) needed for emPCR is determined by titration. The sstDNA library is immobilized onto beads.

The beads containing a library fragment carry a single sstDNA molecule. The bead-bound library is emulsified with the amplification reagents in a water-in-oil mixture. Each bead is captured within its own microreactor where PCR amplification occurs. This results in bead-immobilized, clonally amplified DNA fragments.

In the second step of 454 sequencing, single-stranded template DNA library beads are added to an incubation mix containing DNA polymerase and are layered with beads containing sulfurylase and luciferase onto a device containing pico-liter sized wells. Pyrosequencing is performed on each DNA fragment in parallel. Addition of one or more nucleotides generates a light signal that is recorded by a CCD camera in a sequencing instrument. The signal strength is proportional to the number of nucleotides incorporated. Pyrosequencing exploits the release of pyrophosphate (PPi) upon nucleotide addition. PPi is converted to ATP by ATP sulfurylase in the presence of adenosine 5′ phosphosulfate. Luciferase uses ATP to convert luciferin to oxyluciferin, and this reaction generates light that is discerned and analyzed (see, for example, Margulies, M. et al. Nature 437:376-380 (2005)).

Another nucleic acid sequencing technology that may be used in a method provided herein is Applied Biosystems' SOLiD™ technology. In SOLiD™ sequencing-by-ligation, a library of nucleic acid fragments is prepared from the sample and is used to prepare clonal bead populations. With this method, one species of nucleic acid fragment will be present on the surface of each bead (e.g., magnetic bead). Sample nucleic acid (e.g., genomic DNA) is sheared into fragments, and adaptors are subsequently attached to the 5′ and 3′ ends of the fragments to generate a fragment library. The adapters are typically universal adapter sequences so that the starting sequence of every fragment is both known and identical. Emulsion PCR takes place in microreactors containing all the necessary reagents for PCR. The resulting PCR products attached to the beads are then covalently bound to a glass slide. Primers then hybridize to the adapter sequence within the library template. A set of four fluorescently labeled di-base probes compete for ligation to the sequencing primer. Specificity of the di-base probe is achieved by interrogating every 1st and 2nd base in each ligation reaction. Multiple cycles of ligation, detection and cleavage are performed with the number of cycles determining the eventual read length. Following a series of ligation cycles, the extension product is removed and the template is reset with a primer complementary to the n-1 position for a second round of ligation cycles. Often, five rounds of primer reset are completed for each sequence tag. Through the primer reset process, each base is interrogated in two

independent ligation reactions by two different primers. For example, the base at read position 5 is assayed by primer number 2 in ligation cycle 2 and by primer number 3 in ligation cycle 1.

Another nucleic acid sequencing technology that may be used in a method described herein is the Helicos True Single Molecule Sequencing (tSMS). In the tSMS technique, a polyA sequence is added to the 3′ end of each nucleic acid (e.g., DNA) strand from the sample. Each strand is labeled by the addition of a fluorescently labeled adenosine nucleotide. The DNA strands are then hybridized to a flow cell, which contains millions of oligo-T capture sites that are immobilized to the flow cell surface. The templates can be at a density of about 100 million templates/cm^2. The flow cell is then loaded into a sequencing apparatus and a laser illuminates the surface of the flow cell, revealing the position of each template. A CCD camera can map the position of the templates on the flow cell surface. The template fluorescent label is then cleaved and washed away. The sequencing reaction begins by introducing a DNA polymerase and a fluorescently labeled nucleotide. The oligo-T nucleic acid serves as a primer. The polymerase incorporates the labeled nucleotides to the primer in a template directed manner. The polymerase and unincorporated nucleotides are removed. The templates that have directed incorporation of the fluorescently labeled nucleotide are detected by imaging the flow cell surface. After imaging, a cleavage step removes the fluorescent label, and the process is repeated with other fluorescently labeled nucleotides until the desired read length is achieved. Sequence information is collected with each nucleotide addition step (see, for example, Harris T. D. et al., Science 320:106-109 (2008)).

Another nucleic acid sequencing technology that may be used in a method provided herein is the single molecule, real-time (SMRT™) sequencing technology of Pacific Biosciences. With this method, each of the four DNA bases is attached to one of four different fluorescent dyes. These dyes are phospholinked. A single DNA polymerase is immobilized with a single molecule of template single stranded DNA at the bottom of a zero-mode waveguide (ZMW). A ZMW is a confinement structure which enables observation of incorporation of a single nucleotide by DNA polymerase against the background of fluorescent nucleotides that rapidly diffuse in an out of the ZMW (in microseconds). It takes several milliseconds to incorporate a nucleotide into a growing strand. During this time, the fluorescent label is excited and produces a fluorescent signal, and the fluorescent tag is cleaved off. Detection of the corresponding fluorescence of the dye indicates which base was incorporated. The process is then repeated.

Another nucleic acid sequencing technology that may be used in a method described herein is ION TORRENT (Life Technologies) single molecule sequencing which pairs semiconductor technology with a simple sequencing chemistry to directly translate chemically encoded information (A, C, G, T) into digital information (0, 1) on a semiconductor chip. ION TORRENT uses a high-density array of micro-machined wells to perform nucleic acid sequencing in a massively parallel way. Each well holds a different DNA molecule. Beneath the wells is an ion-sensitive layer and beneath that an ion sensor. Typically, when a nucleotide is incorporated into a strand of DNA by a polymerase, a hydrogen ion is released as a byproduct. If a nucleotide, for example a C, is added to a DNA template and is then incorporated into a strand of DNA, a hydrogen ion will be released. The charge from that ion will change the pH of the solution, which can be detected by an ion sensor. A sequencer can call the base, going directly from chemical information to digital information. The sequencer then sequentially floods the chip with one nucleotide after another. If the next nucleotide that floods the chip is not a match, no voltage change will be recorded and no base will be called. If there are two identical bases on the DNA strand, the voltage will be double, and the chip will record two identical bases called. Because this is direct detection (i.e. detection without scanning, cameras or light), each nucleotide incorporation is recorded in seconds.

Another nucleic acid sequencing technology that may be used in a method described herein is the chemical-sensitive field effect transistor (CHEMFET) array. In one example of this sequencing technique, DNA molecules are placed into reaction chambers, and the template molecules can be hybridized to a sequencing primer bound to a polymerase. Incorporation of one or more triphosphates into a new nucleic acid strand at the 3′ end of the sequencing primer can be detected by a change in current by a CHEMFET sensor. An array can have multiple CHEMFET sensors. In another example, single nucleic acids are attached to beads, and the nucleic acids can be amplified on the bead, and the individual beads can be transferred to individual reaction chambers on a CHEMFET array, with each chamber having a CHEMFET sensor, and the nucleic acids can be sequenced (see, for example, U.S. Patent Application Publication No. 2009/0026082).

Another nucleic acid sequencing technology that may be used in a method described herein is electron microscopy. In one example of this sequencing technique, individual nucleic acid (e.g., DNA) molecules are labeled using metallic labels that are distinguishable using an electron microscope. These molecules are then stretched on a flat surface and imaged using an electron microscope to measure sequences (see, for example, Moudrianakis E. N. and Beer M. PNAS USA. 1965 March; 53:564-71). In certain embodiments, transmission electron microscopy (TEM) is used

(e.g., Halcyon Molecular's TEM method). This method, termed Individual Molecule Placement Rapid Nano Transfer (IMPRNT), includes utilizing single atom resolution transmission electron microscope imaging of high-molecular weight (e.g., about 150 kb or greater) DNA selectively labeled with heavy atom markers and arranging these molecules on ultra-thin films in ultra-dense (3nm strand-to-strand) parallel arrays with consistent base-to-base spacing. The electron microscope is used to image the molecules on the films to determine the position of the heavy atom markers and to extract base sequence information from the DNA (see, for example, International Patent Application No. WO 2009/046445).

Other sequencing methods that may be used to conduct methods herein include digital PCR and sequencing by hybridization. Digital polymerase chain reaction (digital PCR or dPCR) can be used to directly identify and quantify nucleic acids in a sample. Digital PCR can be performed in an emulsion, in some embodiments. For example, individual nucleic acids are separated, e.g., in a microfluidic chamber device, and each nucleic acid is individually amplified by PCR. Nucleic acids can be separated such that there is no more than one nucleic acid per well. In some embodiments, different probes can be used to distinguish various alleles (e.g., fetal alleles and maternal alleles). Alleles can be enumerated to determine copy number. In sequencing by hybridization, the method involves contacting a plurality of polynucleotide sequences with a plurality of polynucleotide probes, where each of the plurality of polynucleotide probes can be optionally tethered to a substrate. The substrate can be a flat surface with an array of known nucleotide sequences, in some embodiments. The pattern of hybridization to the array can be used to determine the polynucleotide sequences present in the sample. In some embodiments, each probe is tethered to a bead, e.g., a magnetic bead or the like. Hybridization to the beads can be identified and used to identify the plurality of polynucleotide sequences within the sample.

In some embodiments, nanopore sequencing can be used in a method described herein. Nanopore sequencing is a single-molecule sequencing technology whereby a single nucleic acid molecule (e.g., DNA) is sequenced directly as it passes through a nanopore. A nanopore is a small hole or channel, of the order of 1 nanometer in diameter. Certain transmembrane cellular proteins can act as nanopores (e.g., alpha-hemolysin). In certain embodiments, nanopores can be synthesized (e.g., using a silicon platform). Immersion of a nanopore in a conducting fluid and application of a potential across it results in a slight electrical current due to conduction of ions through the nanopore. The amount of current which flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs

the nanopore to a different degree and generates characteristic changes to the current. The amount of current which can pass through the nanopore at any given moment therefore varies depending on whether the nanopore is blocked by an A, a C, a G, a T, or in some cases, methyl-C. The change in the current through the nanopore as the DNA molecule passes through the nanopore represents a direct reading of the DNA sequence. In certain embodiments a nanopore can be used to identify individual DNA bases as they pass through the nanopore in the correct order (see, for example, Soni GV and Meller A. Clin Chem 53: 1996-2001 (2007); International Patent Application No. WO2010/004265).

There are a number of ways that nanopores can be used to sequence nucleic acid molecules. In some embodiments, an exonuclease enzyme, such as a deoxyribonuclease, is used. In this case, the exonuclease enzyme is used to sequentially detach nucleotides from a nucleic acid (e.g., DNA) molecule. The nucleotides are then detected and discriminated by the nanopore in order of their release, thus reading the sequence of the original strand. For such an embodiment, the exonuclease enzyme can be attached to the nanopore such that a proportion of the nucleotides released from the DNA molecule is capable of entering and interacting with the channel of the nanopore. The exonuclease can be attached to the nanopore structure at a site in close proximity to the part of the nanopore that forms the opening of the channel. In certain embodiments, the exonuclease enzyme can be attached to the nanopore structure such that its nucleotide exit trajectory site is orientated towards the part of the nanopore that forms part of the opening.

In some embodiments, nanopore sequencing of nucleic acids involves the use of an enzyme that pushes or pulls the nucleic acid (e.g., DNA) molecule through the pore. In this case, the ionic current fluctuates as a nucleotide in the DNA molecule passes through the pore. The fluctuations in the current are indicative of the DNA sequence. For such an embodiment, the enzyme can be attached to the nanopore structure such that it is capable of pushing or pulling the target nucleic acid through the channel of a nanopore without interfering with the flow of ionic current through the pore. The enzyme can be attached to the nanopore structure at a site in close proximity to the part of the structure that forms part of the opening. The enzyme can be attached to the subunit, for example, such that its active site is orientated towards the part of the structure that forms part of the opening.

In some embodiments, nanopore sequencing of nucleic acids involves detection of polymerase bi-products in close proximity to a nanopore detector. In this case, nucleoside phosphates

(nucleotides) are labeled so that a phosphate labeled species is released upon the addition of a polymerase to the nucleotide strand and the phosphate labeled species is detected by the pore. Typically, the phosphate species contains a specific label for each nucleotide. As nucleotides are sequentially added to the nucleic acid strand, the bi-products of the base addition are detected. The order that the phosphate labeled species are detected can be used to determine the sequence of the nucleic acid strand.

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). Nanopore sequencing, for example, can provide sequence reads that can vary in size from tens to hundreds to thousands of base pairs. 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 or more.

In some embodiments, chromosome-specific sequencing is performed. In some embodiments, chromosome-specific sequencing is performed utilizing DANSR (digital analysis of selected regions). Digital analysis of selected regions enables simultaneous quantification of hundreds of loci by cfDNA-dependent catenation of two locus-specific oligonucleotides via an intervening 'bridge' oligo to form a PCR template. In some embodiments, chromosome-specific sequencing is performed by generating a library enriched in chromosome-specific sequences. In some embodiments, sequence reads are obtained only for a selected set of chromosomes. In some embodiments, sequence reads are obtained only for chromosomes 21, 18 and 13.

In some embodiments, nucleic acids may include a fluorescent signal or sequence tag information. Quantification of the signal or tag may be used in a variety of techniques such as, for example, flow cytometry, quantitative polymerase chain reaction (qPCR), gel electrophoresis, gene-chip analysis, microarray, mass spectrometry, cytofluorimetric analysis, fluorescence microscopy, confocal laser scanning microscopy, laser scanning cytometry, affinity chromatography, manual batch mode separation, electric field suspension, sequencing, and combination thereof.

Mapping reads

Mapping nucleotide sequence reads (i.e., sequence information from a fragment whose physical genomic position is unknown) can be performed in a number of ways, and often comprises alignment of the obtained sequence reads with a matching sequence in a reference genome (e.g., Li et al., "Mapping short DNA sequencing reads and calling variants using mapping quality score," Genome Res., 2008 Aug 19.) In such alignments, sequence reads generally are aligned to a reference sequence and those that align are designated as being "mapped" or a "sequence tag." In certain embodiments, a mapped sequence read is referred to as a “hit” or a “count”. In some embodiments, mapped sequence reads are grouped together according to various parameters and assigned to particular genomic sections, which are discussed in further detail below.

As used herein, the terms “aligned”, “alignment”, or “aligning” refer to two or more nucleic acid sequences that can be identified as a match (e.g., 100% identity) or partial match. Alignments can be done manually or by a computer algorithm, examples including the Efficient Local Alignment of Nucleotide Data (ELAND) computer program distributed as part of the Illumina Genomics Analysis pipeline. The alignment of a sequence read can be a 100% sequence match. In some cases, an alignment is less than a 100% sequence match (i.e., non-perfect match, partial match, partial alignment). In some embodiments an alignment is about a 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76% or 75% match. In some embodiments, an alignment comprises a mismatch. In some embodiments, an alignment comprises 1, 2, 3, 4 or 5 mismatches. Two or more sequences can be aligned using either strand. In certain embodiments a nucleic acid sequence is aligned with the reverse complement of another nucleic acid sequence.

Various computational methods can be used to map each sequence read to a genomic section. Non-limiting examples of computer algorithms that can be used to align sequences include, without limitation, BLAST, BLITZ, FASTA, BOWTIE 1, BOWTIE 2, ELAND, MAQ, PROBEMATCH, SOAP or SEQMAP, or variations thereof or combinations thereof. In some embodiments, sequence reads can be aligned with sequences in a reference genome. In some embodiments, the sequence reads can be found and/or aligned with sequences in nucleic acid databases known in the art including, for example, GenBank, dbEST, dbSTS, EMBL (European Molecular Biology Laboratory) and DDBJ (DNA Databank of Japan). BLAST or similar tools can be used to search

the identified sequences against a sequence database. Search hits can then be used to sort the identified sequences into appropriate genomic sections (described hereafter), for example.

The term “sequence tag” is herein used interchangeably with the term “mapped sequence tag” to refer to a sequence read that has been specifically assigned i.e. mapped, to a larger sequence e.g., a reference genome, by alignment. Mapped sequence tags are uniquely mapped to a reference genome i.e. they are assigned to a single location to the reference genome. Tags that can be mapped to more than one location on a reference genome i.e. tags that do not map uniquely, are not included in the analysis. A "sequence tag" can be a nucleic acid (e.g., DNA) sequence (i.e. read) assigned specifically to a particular genomic section and/or chromosome (i.e. one of chromosomes 1-22, X or Y for a human subject). A sequence tag may be repetitive or non-repetitive within a single segment of the reference genome (e.g., a chromosome). In some embodiments, repetitive sequence tags are eliminated from further analysis (e.g., quantification). In some embodiments, a read may uniquely or non-uniquely map to portions in the reference genome. A read is considered to be “uniquely mapped” if it aligns with a single sequence in the reference genome. A read is considered to be “non-uniquely mapped” if it aligns with two or more sequences in the reference genome. In some embodiments, non-uniquely mapped reads are eliminated from further analysis (e.g., quantification). A certain, small degree of mismatch (0-1) may be allowed to account for single nucleotide polymorphisms that may exist between the reference genome and the reads from individual samples being mapped, in certain embodiments. In some embodiments, no degree of mismatch is allowed for a read to be mapped to a reference sequence.

As used herein, the term “reference genome” can refer to any particular known, sequenced or characterized genome, whether partial or complete, of any organism or virus which may be used to reference identified sequences from a subject. For example, a reference genome used for human subjects as well as many other organisms can be found at the National Center for Biotechnology Information at www.ncbi.nlm.nih.gov. A “genome” refers to the complete genetic information of an organism or virus, 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. In some embodiments, a reference genome comprises sequences assigned to chromosomes.

In certain embodiments, where a sample nucleic acid is from a pregnant female, a reference sequence sometimes is not from the fetus, the mother of the fetus or the father of the fetus, and is referred to herein as an "external reference." A maternal reference may be prepared and used in some embodiments. When a reference from the pregnant female is prepared ("maternal reference sequence") based on an external reference, reads from DNA of the pregnant female that contains substantially no fetal DNA often are mapped to the external reference sequence and assembled. In certain embodiments the external reference is from DNA of an individual having substantially the same ethnicity as the pregnant female. A maternal reference sequence may not completely cover the maternal genomic DNA (e.g., it may cover about 50%, 60%, 70%, 80%, 90% or more of the maternal genomic DNA), and the maternal reference may not perfectly match the maternal genomic DNA sequence (e.g., the maternal reference sequence may include multiple mismatches).

In certain embodiments, mappability is assessed for a genomic region (e.g., genomic section, genomic portion, bin). Mappability is the ability to unambiguously align a nucleotide sequence read to a portion of a reference genome, typically up to a specified number of mismatches, including, for example, 0, 1, 2 or more mismatches. For a given genomic region, the expected mappability can be estimated using a sliding-window approach of a preset read length and averaging the resulting read-level mappability values. Genomic regions comprising stretches of unique nucleotide sequence sometimes have a high mappability value.

Genomic sections

In some embodiments, mapped sequence reads (i.e. sequence tags) are grouped together according to various parameters and assigned to particular genomic sections. Often, the individual mapped sequence reads can be used to identify an amount of a genomic section present in a sample. In some embodiments, the amount of a genomic section can be indicative of the amount of a larger sequence (e.g., a chromosome) in the sample. The term “genomic section” can also be referred to herein as a “sequence window”, “section”, “bin”, “locus”, “region”, “partition”, “portion” (e.g., portion of a reference genome, portion of a chromosome) or "genomic portion." In some embodiments, a genomic section is an entire chromosome, portion of a chromosome, portion of a reference genome, multiple chromosome portions, multiple chromosomes, portions from multiple chromosomes, and/or combinations thereof. In some embodiments, a genomic section is predefined based on specific parameters. In some embodiments, a genomic section is arbitrarily

defined based on partitioning of a genome (e.g., partitioned by size, portions, contiguous regions, contiguous regions of an arbitrarily defined size, and the like).

In some embodiments, a genomic section is delineated based on one or more parameters which include, for example, length or a particular feature or features of the sequence. Genomic sections can be selected, filtered and/or removed from consideration using any suitable criteria know in the art or described herein. 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 sequence reads to a plurality of genomic sections. Genomic sections can be approximately the same length or the genomic sections can be different lengths. In some embodiments, genomic sections are of about equal length. In some embodiments genomic sections of different lengths are adjusted or weighted. In some embodiments, a genomic section is about 10 kilobases (kb) to about 100 kb, about 20 kb to about 80 kb, about 30 kb to about 70 kb, about 40 kb to about 60 kb, and sometimes about 50 kb. In some embodiments, a genomic section is about 10 kb to about 20 kb. A genomic section is not limited to contiguous runs of sequence. Thus, genomic sections can be made up of contiguous and/or non-contiguous sequences. A genomic section is not limited to a single chromosome. In some embodiments, a genomic section includes all or part of one chromosome or all or part of two or more chromosomes. In some embodiments, genomic sections may span one, two, or more entire chromosomes. In addition, the genomic sections may span joint or disjointed portions of multiple chromosomes.

In some embodiments, genomic sections can be particular chromosome portion in a chromosome of interest, such as, for example, chromosomes where a genetic variation is assessed (e.g., an aneuploidy of chromosomes 13, 18 and/or 21 or a sex chromosome). A genomic section can also be a pathogenic genome (e.g., bacterial, fungal or viral) or fragment thereof. Genomic sections can be genes, gene fragments, regulatory sequences, introns, exons, and the like.

In some embodiments, a genome (e.g., human genome) is partitioned into genomic sections based on the information content of the regions. The resulting genomic regions may contain sequences for multiple chromosomes and/or may contain sequences for portions of multiple chromosomes. In some embodiments, the partitioning may eliminate similar locations across the genome and only keep unique regions. The eliminated regions may be within a single chromosome or may span multiple chromosomes. The resulting genome is thus trimmed down and optimized for faster alignment, often allowing for focus on uniquely identifiable sequences.

In some embodiments, the partitioning may down weight similar regions. The process for down weighting a genomic section is discussed in further detail below. In some embodiments, the partitioning of the genome into regions transcending chromosomes may be based on information gain produced in the context of classification. For example, the information content may be quantified using the p-value profile measuring the significance of particular genomic locations for distinguishing between groups of confirmed normal and abnormal subjects (e.g., euploid and trisomy subjects, respectively). In some embodiments, the partitioning of the genome into regions transcending chromosomes may be based on any other criterion, such as, for example, speed/convenience while aligning tags, high or low GC content, uniformity of GC content, other measures of sequence content (e.g., fraction of individual nucleotides, fraction of pyrimidines or purines, fraction of natural vs. non-natural nucleic acids, fraction of methylated nucleotides, and CpG content), methylation state, duplex melting temperature, amenability to sequencing or PCR, uncertainty value assigned to individual bins, and/or a targeted search for particular features.

A "segment" of a chromosome generally is part of a chromosome, and typically is a different part of a chromosome than a genomic section (e.g., bin). A segment of a chromosome sometimes is in a different region of a chromosome than a genomic section, sometimes does not share a polynucleotide with a genomic section, and sometimes includes a polynucleotide that is in a genomic section. A segment of a chromosome often contains a larger number of nucleotides than a genomic section (e.g., a segment sometimes includes a genomic section), and sometimes a segment of a chromosome contains a smaller number of nucleotides than a genomic section (e.g., a segment sometimes is within a genomic section).

Sequence tag density

"Sequence tag density" refers to the normalized value of sequence tags or reads for a defined genomic section where the sequence tag density is used for comparing different samples and for subsequent analysis. The value of the sequence tag density often is normalized within a sample. In some embodiments, normalization can be performed by counting the number of tags falling within each genomic section; obtaining a median value of the total sequence tag count for each chromosome; obtaining a median value of all of the autosomal values; and using this value as a normalization constant to account for the differences in total number of sequence tags obtained for different samples. A sequence tag density sometimes is about 1 for a disomic chromosome.

Sequence tag densities can vary according to sequencing artifacts, most notably G/C bias, which can be corrected by use of an external standard or internal reference (e.g., derived from substantially all of the sequence tags (genomic sequences), which may be, for example, a single chromosome or a calculated value from all autosomes, in some embodiments). Thus, dosage imbalance of a chromosome or chromosomal regions can be inferred from the percentage representation of the locus among other mappable sequenced tags of the specimen. Dosage imbalance of a particular chromosome or chromosomal regions therefore can be quantitatively determined and be normalized. Methods for sequence tag density normalization and quantification are discussed in further detail below.

In some embodiments a proportion of all of the sequence reads are from a chromosome involved in an aneuploidy (e.g., chromosome 13, 18, 21, 4, 9, 16, 20, 22, the like, or a sex chromosome), and other sequence reads are from other chromosomes. By taking into account the relative size of the chromosome involved in the aneuploidy (e.g., "target chromosome": e.g., chromosome 21 or 9) compared to other chromosomes, one could obtain a normalized frequency, within a reference range, of target chromosome-specific sequences, in some embodiments. If the fetus has an aneuploidy in a target chromosome, then the normalized frequency of the target chromosome-derived sequences is statistically greater than the normalized frequency of non-target chromosome-derived sequences, thus allowing the detection of the aneuploidy. The degree of change in the normalized frequency will be dependent on the fractional concentration of fetal nucleic acids in the analyzed sample, in some embodiments.

Counts

Sequence reads that are mapped or partitioned based on a selected feature or variable can be quantified to determine the number of reads that are mapped to a genomic section (e.g., bin, partition, genomic portion, portion of a reference genome, portion of a chromosome and the like), in some embodiments. In certain embodiments the quantity of sequence reads that are mapped to a genomic section are termed counts (e.g., a count). Often a count is associated with a genomic section. In certain embodiments counts for two or more genomic sections (e.g., a set of genomic sections) are mathematically manipulated (e.g., averaged, added, normalized, the like or a combination thereof). In some embodiments a count is determined from some or all of the sequence reads mapped to (i.e., associated with) a genomic section. In certain embodiments, a count is determined from a pre-defined subset of mapped sequence reads. Pre-defined subsets of

mapped sequence reads can be defined or selected utilizing any suitable feature or variable. In some embodiments, pre-defined subsets of mapped sequence reads can include from 1 to n sequence reads, where n represents a number equal to the sum of all sequence reads generated from a test subject or reference subject sample.

In certain embodiments a count is derived from sequence reads that are processed or manipulated by a suitable method, operation or mathematical process known in the art. In certain embodiments a count is derived from sequence reads associated with a genomic section where some or all of the sequence reads are weighted, removed, filtered, normalized, adjusted, averaged, derived as a mean, added, or subtracted or processed by a combination thereof. In some embodiments, a count is derived from raw sequence reads and or filtered sequence reads. A count (e.g., counts) can be determined by a suitable method, operation or mathematical process. In certain embodiments a count value is determined by a mathematical process. In certain embodiments a count value is an average, mean or sum of sequence reads mapped to a genomic section. Often a count is a mean number of counts. In some embodiments, a count is associated with an uncertainty value. Counts can be processed (e.g., normalized) by a method known in the art and/or as described herein (e.g., bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, GC LOESS, LOWESS, PERUN, RM, GCRM, cQn and/or combinations thereof).

Counts (e.g., raw, filtered and/or normalized counts) can be processed and normalized to one or more elevations. Elevations and profiles are described in greater detail hereafter. In certain embodiments counts can be processed and/or normalized to a reference elevation. Reference elevations are addressed later herein. Counts processed according to an elevation (e.g., processed counts) can be associated with an uncertainty value (e.g., a calculated variance, an error, standard deviation, p-value, mean absolute deviation, etc.). An uncertainty value typically defines a range above and below an elevation. A value for deviation can be used in place of an uncertainty value, and non-limiting examples of measures of deviation include standard deviation, average absolute deviation, median absolute deviation, standard score (e.g., Z-score, Z-value, normal score, standardized variable) and the like.

Counts are often obtained from a nucleic acid sample from a pregnant female bearing a fetus. Counts of nucleic acid sequence reads mapped to a genomic section often are counts representative of both the fetus and the mother of the fetus (e.g., a pregnant female subject). In

certain embodiments some of the counts mapped to a genomic section are from a fetal genome and some of the counts mapped to the same genomic section are from the maternal genome.

Data processing, Normalization & PERUN

Mapped sequence reads that have been counted are referred to herein as raw data, since the data represents unmanipulated counts (e.g., raw counts). In some embodiments, sequence read data in a data set can be processed further (e.g., mathematically and/or statistically manipulated) and/or displayed to facilitate providing an outcome. In certain embodiments, data sets, including larger data sets, may benefit from pre-processing to facilitate further analysis. Pre-processing of data sets sometimes involves removal of redundant and/or uninformative genomic sections or bins (e.g., bins with uninformative data, redundant mapped reads, genomic sections or bins with zero median counts, over represented or under represented sequences). Without being limited by theory, data processing and/or preprocessing may (i) remove noisy data, (ii) remove uninformative data, (iii) remove redundant data, (iv) reduce the complexity of larger data sets, and/or (v) facilitate transformation of the data from one form into one or more other forms. The terms “pre-processing” and “processing” when utilized with respect to data or data sets are collectively referred to herein as “processing”. Processing can render data more amenable to further analysis, and can generate an outcome in some embodiments.

The term “noisy data” as used herein refers to (a) data that has a significant variance between data points when analyzed or plotted, (b) data that has a significant standard deviation (e.g., greater than 3 standard deviations), (c) data that has a significant standard error of the mean, the like, and combinations of the foregoing. Noisy data sometimes occurs due to the quantity and/or quality of starting material (e.g., nucleic acid sample), and sometimes occurs as part of processes for preparing or replicating DNA used to generate sequence reads. In certain embodiments, noise results from certain sequences being over represented when prepared using PCR-based methods. Methods described herein can reduce or eliminate the contribution of noisy data, and therefore reduce the effect of noisy data on the provided outcome.

The terms “uninformative data”, “uninformative bins”, and “uninformative genomic sections” as used herein refer to genomic sections, or data derived therefrom, having a numerical value that is significantly different from a predetermined threshold value or falls outside a predetermined cutoff range of values. The terms “threshold” and “threshold value” herein refer to any number that is

calculated using a qualifying data set and serves as a limit of diagnosis of a genetic variation (e.g., a copy number variation, an aneuploidy, a chromosomal aberration, and the like). In certain embodiments a threshold is exceeded by results obtained by methods described herein and a

[redacted]

subject is diagnosed with a genetic variation (e.g., trisomy 21). A threshold value or range of values often is calculated by mathematically and/or statistically manipulating sequence read data (e.g., from a reference and/or subject), in some embodiments, and in certain embodiments, sequence read data manipulated to generate a threshold value or range of values is sequence read data (e.g., from a reference and/or subject). In some embodiments, an uncertainty value is determined. An uncertainty value generally is a measure of variance or error and can be any suitable measure of variance or error. An uncertainty value can be a standard deviation, standard error, calculated variance, p-value, or mean absolute deviation (MAD), in some embodiments. In some embodiments an uncertainty value can be calculated according to a formula in Example 6.

Any suitable procedure can be utilized for processing data sets described herein. Non-limiting examples of procedures suitable for use for processing data sets include filtering, normalizing, weighting, monitoring peak heights, monitoring peak areas, monitoring peak edges, determining area ratios, mathematical processing of data, statistical processing of data, application of statistical algorithms, analysis with fixed variables, analysis with optimized variables, plotting data to identify patterns or trends for additional processing, the like and combinations of the foregoing. In some embodiments, data sets are processed based on various features (e.g., GC content, redundant mapped reads, centromere regions, telomere regions, the like and combinations thereof) and/or variables (e.g., fetal gender, maternal age, maternal ploidy, percent contribution of fetal nucleic acid, the like or combinations thereof). In certain embodiments, processing data sets as described herein can reduce the complexity and/or dimensionality of large and/or complex data sets. A non-limiting example of a complex data set includes sequence read data generated from one or more test subjects and a plurality of reference subjects of different ages and ethnic backgrounds. In some embodiments, data sets can include from thousands to millions of sequence reads for each test and/or reference subject.

Data processing can be performed in any number of steps, in certain embodiments. For example, data may be processed using only a single processing procedure in some embodiments, and in certain embodiments data may be processed using 1 or more, 5 or more, 10 or more or 20 or more

steps, 7 or more processing steps, 8 or more processing steps, 9 or more processing steps, 10 or more processing steps, 11 or more processing steps, 12 or more processing steps, 13 or more processing steps, 14 or more processing steps, 15 or more processing steps, 16 or more processing steps, 17 or more processing steps, 18 or more processing steps, 19 or more processing steps, or 20 or more processing steps). In some embodiments, processing steps may be the same step repeated two or more times (e.g., filtering two or more times, normalizing two or more times), and in certain embodiments, processing steps may be two or more different processing steps (e.g., filtering, normalizing; normalizing, monitoring peak heights and edges; filtering, normalizing, normalizing to a reference, statistical manipulation to determine p-values, and the like), carried out simultaneously or sequentially. In some embodiments, any suitable number and/or combination of the same or different processing steps can be utilized to process sequence read data to facilitate providing an outcome. In certain embodiments, processing data sets by the criteria described herein may reduce the complexity and/or dimensionality of a data set.

In some embodiments, one or more processing steps can comprise one or more filtering steps. The term "filtering" as used herein refers to removing genomic sections or bins from consideration. Bins can be selected for removal based on any suitable criteria, including but not limited to redundant data (e.g., redundant or overlapping mapped reads), non-informative data (e.g., bins with zero median counts), bins with over represented or under represented sequences, noisy data, the like, or combinations of the foregoing. A filtering process often involves removing one or more bins from consideration and subtracting the counts in the one or more bins selected for removal from the counted or summed counts for the bins, chromosome or chromosomes, or genome under consideration. In some embodiments, bins can be removed successively (e.g., one at a time to allow evaluation of the effect of removal of each individual bin), and in certain embodiments all bins marked for removal can be removed at the same time. In some embodiments, genomic sections characterized by a variance above or below a certain level are removed, which sometimes is referred to herein as filtering "noisy" genomic sections. In certain embodiments, a filtering process comprises obtaining data points from a data set that deviate from the mean profile elevation of a genomic section, a chromosome, or segment of a chromosome by a predetermined multiple of the profile variance, and in certain embodiments, a filtering process comprises removing data points from a data set that do not deviate from the mean profile elevation of a genomic section, a chromosome or segment of a chromosome by a predetermined multiple of the profile variance. In some embodiments, a filtering process is utilized to reduce the number of candidate genomic sections analyzed for the presence or absence of a genetic variation. Reducing the number of

candidate genomic sections analyzed for the presence or absence of a genetic variation (e.g., micro-deletion, micro-duplication) often reduces the complexity and/or dimensionality of a data set, and sometimes increases the speed of searching for and/or identifying genetic variations and/or genetic aberrations by two or more orders of magnitude.

In some embodiments one or more processing steps can comprise one or more normalization steps. Normalization can be performed by a suitable method known in the art. In certain embodiments normalization comprises adjusting values measured on different scales to a notionally common scale. In certain embodiments normalization comprises a sophisticated mathematical adjustment to bring probability distributions of adjusted values into alignment. In some embodiments normalization comprises aligning distributions to a normal distribution. In certain embodiments normalization comprises mathematical adjustments that allow comparison of corresponding normalized values for different datasets in a way that eliminates the effects of certain gross influences (e.g., error and anomalies). In certain embodiments normalization comprises scaling. Normalization sometimes comprises division of one or more data sets by a predetermined variable or formula. Non-limiting examples of normalization methods include bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS (locally weighted scatterplot smoothing), PERUN, repeat masking (RM), GC-normalization and repeat masking (GCRM), cQn and/or combinations thereof. In some embodiments, the determination of a presence or absence of a genetic variation (e.g., an aneuploidy) utilizes a normalization method (e.g., bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS (locally weighted scatterplot smoothing), PERUN, repeat masking (RM), GC-normalization and repeat masking (GCRM), cQn, a normalization method known in the art and/or a combination thereof).

For example, LOESS is a regression modeling method known in the art that combines multiple regression models in a k-nearest-neighbor-based meta-model. LOESS is sometimes referred to as a locally weighted polynomial regression. GC LOESS, in some embodiments, applies an LOESS model to the relation between fragment count (e.g., sequence reads, counts) and GC composition for genomic sections. Plotting a smooth curve through a set of data points using LOESS is sometimes called an LOESS curve, particularly when each smoothed value is given by a weighted quadratic least squares regression over the span of values of the y-axis scattergram criterion variable. For each point in a data set, the LOESS method fits a low-degree polynomial to a subset of the data, with explanatory variable values near the point whose response is being estimated.

The polynomial is fitted using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. The value of the regression function for a point is then obtained by evaluating the local polynomial using the explanatory variable values for that data point. The LOESS fit is sometimes considered complete after regression function values have been computed for each of the data points. Many of the details of this method, such as the degree of the polynomial model and the weights, are flexible.

Any suitable number of normalizations can be used. In some embodiments, data sets can be normalized 1 or more, 5 or more, 10 or more or even 20 or more times. Data sets can be normalized to values (e.g., normalizing value) representative of any suitable feature or variable (e.g., sample data, reference data, or both). Non-limiting examples of types of data normalizations that can be used include normalizing raw count data for one or more selected test or reference genomic sections to the total number of counts mapped to the chromosome or the entire genome on which the selected genomic section or sections are mapped; normalizing raw count data for one or more selected genomic sections to a median reference count for one or more genomic sections or the chromosome on which a selected genomic section or segments is mapped; normalizing raw count data to previously normalized data or derivatives thereof; and normalizing previously normalized data to one or more other predetermined normalization variables. Normalizing a data set sometimes has the effect of isolating statistical error, depending on the feature or property selected as the predetermined normalization variable. Normalizing a data set sometimes also allows comparison of data characteristics of data having different scales, by bringing the data to a common scale (e.g., predetermined normalization variable). In some embodiments, one or more normalizations to a statistically derived value can be utilized to minimize data differences and diminish the importance of outlying data. Normalizing genomic sections, or bins, with respect to a normalizing value sometimes is referred to as “bin-wise normalization”.

In certain embodiments, a processing step comprising normalization includes normalizing to a static window, and in some embodiments, a processing step comprising normalization includes normalizing to a moving or sliding window. The term “window” as used herein refers to one or more genomic sections chosen for analysis, and sometimes used as a reference for comparison (e.g., used for normalization and/or other mathematical or statistical manipulation). The term “normalizing to a static window” as used herein refers to a normalization process using one or more genomic sections selected for comparison between a test subject and reference subject data set. In some embodiments the selected genomic sections are utilized to generate a profile. A

static window generally includes a predetermined set of genomic sections that do not change during manipulations and/or analysis. The terms "normalizing to a moving window" and "normalizing to a sliding window" as used herein refer to normalizations performed to genomic sections localized to the genomic region (e.g., immediate genetic surrounding, adjacent genomic section or sections, and the like) of a selected test genomic section, where one or more selected test genomic sections are normalized to genomic sections immediately surrounding the selected test genomic section. In certain embodiments, the selected genomic sections are utilized to generate a profile. A sliding or moving window normalization often includes repeatedly moving or sliding to an adjacent test genomic section, and normalizing the newly selected test genomic section to genomic sections immediately surrounding or adjacent to the newly selected test genomic section, where adjacent windows have one or more genomic sections in common. In certain embodiments, a plurality of selected test genomic sections and/or chromosomes can be analyzed by a sliding window process.

In some embodiments, normalizing to a sliding or moving window can generate one or more values, where each value represents normalization to a different set of reference genomic sections selected from different regions of a genome (e.g., chromosome). In certain embodiments, the one or more values generated are cumulative sums (e.g., a numerical estimate of the integral of the normalized count profile over the selected genomic section, domain (e.g., part of chromosome), or chromosome). The values generated by the sliding or moving window process can be used to generate a profile and facilitate arriving at an outcome. In some embodiments, cumulative sums of one or more genomic sections can be displayed as a function of genomic position. Moving or sliding window analysis sometimes is used to analyze a genome for the presence or absence of micro-deletions and/or micro-insertions. In certain embodiments, displaying cumulative sums of one or more genomic sections is used to identify the presence or absence of regions of genetic variation (e.g., micro-deletions, micro-duplications). In some embodiments, moving or sliding window analysis is used to identify genomic regions containing micro-deletions and in certain embodiments, moving or sliding window analysis is used to identify genomic regions containing micro-duplications.

A particularly useful normalization methodology for reducing error associated with nucleic acid indicators is referred to herein as Parameterized Error Removal and Unbiased Normalization (PERUN). PERUN methodology can be applied to a variety of nucleic acid indicators (e.g., nucleic

acid sequence reads) for the purpose of reducing effects of error that confound predictions based on such indicators.

For example, PERUN methodology can be applied to nucleic acid sequence reads from a sample and reduce the effects of error that can impair nucleic acid elevation determinations (e.g., genomic section elevation determinations). Such an application is useful for using nucleic acid sequence reads to assess the presence or absence of a genetic variation in a subject manifested as a varying elevation of a nucleotide sequence (e.g., genomic section). Non-limiting examples of variations in genomic sections are chromosome aneuploidies (e.g., trisomy 21, trisomy 18, trisomy 13, trisomy 4, trisomy 9, trisomy 16, and the like) and presence or absence of a sex chromosome (e.g., XX in females versus XY in males). A trisomy of an autosome (e.g., a chromosome other than a sex chromosome) can be referred to as an affected autosome. Other non-limiting examples of variations in genomic section elevations include microdeletions, microinsertions, duplications and mosaicism.

In certain applications, PERUN methodology can reduce experimental bias by normalizing nucleic acid indicators for particular genomic groups, the latter of which are referred to as bins. Bins include a suitable collection of nucleic acid indicators, a non-limiting example of which includes a length of contiguous nucleotides, which is referred to herein as a genomic section or portion of a reference genome. Bins can include other nucleic acid indicators as described herein. In such applications, PERUN methodology generally normalizes nucleic acid indicators at particular bins across a number of samples in three dimensions. A detailed description of particular PERUN applications is described in Example 4 and Example 5 herein.

In certain embodiments, PERUN methodology includes calculating a genomic section elevation for each bin from a fitted relation between (i) experimental bias for a bin of a reference genome to which sequence reads are mapped and (ii) counts of sequence reads mapped to the bin. Experimental bias for each of the bins can be determined across multiple samples according to a fitted relation for each sample between (i) the counts of sequence reads mapped to each of the bins, and (ii) a mapping feature fore each of the bins. This fitted relation for each sample can be assembled for multiple samples in three dimensions. The assembly can be ordered according to the experimental bias in certain embodiments (e.g., FIG. 82, Example 4), although PERUN methodology may be practiced without ordering the assembly according to the experimental bias.

A relation can be generated by a method known in the art. A relation in two dimensions can be generated for each sample in certain embodiments, and a variable probative of error, or possibly probative of error, can be selected for one or more of the dimensions. A relation can be generated, for example, using graphing software known in the art that plots a graph using values of two or more variables provided by a user. A relation can be fitted using a method known in the art (e.g., graphing software). Certain relations can be fitted by linear regression, and the linear regression can generate a slope value and intercept value. Certain relations sometimes are not linear and can be fitted by a non-linear function, such as a parabolic, hyperbolic or exponential function, for example.

In PERUN methodology, one or more of the fitted relations may be linear. For an analysis of cell-free circulating nucleic acid from pregnant females, where the experimental bias is GC bias and the mapping feature is GC content, the fitted relation for a sample between the (i) the counts of sequence reads mapped to each bin, and (ii) GC content for each of the bins, can be linear. For the latter fitted relation, the slope pertains to GC bias, and a GC bias coefficient can be determined for each bin when the fitted relations are assembled across multiple samples. In such embodiments, the fitted relation for multiple samples and a bin between (i) GC bias coefficient for the bin, and (ii) counts of sequence reads mapped to bin, also can be linear. An intercept and slope can be obtained from the latter fitted relation. In such applications, the slope addresses sample-specific bias based on GC-content and the intercept addresses a bin-specific attenuation pattern common to all samples. PERUN methodology can significantly reduce such sample-specific bias and bin-specific attenuation when calculating genomic section elevations for providing an outcome (e.g., presence or absence of genetic variation; determination of fetal sex).

Thus, application of PERUN methodology to sequence reads across multiple samples in parallel can significantly reduce error caused by (i) sample-specific experimental bias (e.g., GC bias) and (ii) bin-specific attenuation common to samples. Other methods in which each of these two sources of error are addressed separately or serially often are not able to reduce these as effectively as PERUN methodology. Without being limited by theory, it is expected that PERUN methodology reduces error more effectively in part because its generally additive processes do not magnify spread as much as generally multiplicative processes utilized in other normalization approaches (e.g., GC-LOESS).

Additional normalization and statistical techniques may be utilized in combination with PERUN methodology. An additional process can be applied before, after and/or during employment of PERUN methodology. Non-limiting examples of processes that can be used in combination with PERUN methodology are described hereafter.

In some embodiments, a secondary normalization or adjustment of a genomic section elevation for GC content can be utilized in conjunction with PERUN methodology. A suitable GC content adjustment or normalization procedure can be utilized (e.g., GC-LOESS, GCRM). In certain embodiments, a particular sample can be identified for application of an additional GC normalization process. For example, application of PERUN methodology can determine GC bias for each sample, and a sample associated with a GC bias above a certain threshold can be selected for an additional GC normalization process. In such embodiments, a predetermined threshold elevation can be used to select such samples for additional GC normalization.

In certain embodiments, a bin filtering or weighting process can be utilized in conjunction with PERUN methodology. A suitable bin filtering or weighting process can be utilized and non-limiting examples are described herein. Examples 4 and 5 describe utilization of R-factor measures of error for bin filtering.

In some embodiments, a normalization technique that reduces error associated with maternal insertions, duplications and/or deletions (e.g., maternal and/or fetal copy number variations), is utilized in conjunction with PERUN methodology.

Genomic section elevations calculated by PERUN methodology can be utilized directly for providing an outcome. In some embodiments, genomic section elevations can be utilized directly to provide an outcome for samples in which fetal fraction is about 2% to about 6% or greater (e.g., fetal fraction of about 4% or greater). Genomic section elevations calculated by PERUN methodology sometimes are further processed for the provision of an outcome. In some embodiments, calculated genomic section elevations are standardized. In certain embodiments, the sum, mean or median of calculated genomic section elevations for a test genomic section (e.g., chromosome 21) can be divided by the sum, mean or median of calculated genomic section elevations for genomic sections other than the test genomic section (e.g., autosomes other than chromosome 21), to generate an experimental genomic section elevation. An experimental genomic section elevation or a raw genomic section elevation can be used as part of a

standardization analysis, such as calculation of a Z-score or Z-value. A Z-score can be generated for a sample by subtracting an expected genomic section elevation from an experimental genomic section elevation or raw genomic section elevation and the resulting value may be divided by a standard deviation for the samples. Resulting Z-scores can be distributed for different samples and analyzed, or can be related to other variables, such as fetal fraction and others, and analyzed, to provide an outcome, in certain embodiments.

As noted herein, PERUN methodology is not limited to normalization according to GC bias and GC content per se, and can be used to reduce error associated with other sources of error. A non-limiting example of a source of non-GC content bias is mappability. When normalization parameters other than GC bias and content are addressed, one or more of the fitted relations may be non-linear (e.g., hyperbolic, exponential). Where experimental bias is determined from a non-linear relation, for example, an experimental bias curvature estimation may be analyzed in some embodiments.

PERUN methodology can be applied to a variety of nucleic acid indicators. Non-limiting examples of nucleic acid indicators are nucleic acid sequence reads and nucleic acid elevations at a particular location on a microarray. Non-limiting examples of sequence reads include those obtained from cell-free circulating DNA, cell-free circulating RNA, cellular DNA and cellular RNA. PERUN methodology can be applied to sequence reads mapped to suitable reference sequences, such as genomic reference DNA, cellular reference RNA (e.g., transcriptome), and portions thereof (e.g., part(s) of a genomic complement of DNA or RNA transcriptome, part(s) of a chromosome).

Thus, in certain embodiments, cellular nucleic acid (e.g., DNA or RNA) can serve as a nucleic acid indicator. Cellular nucleic acid reads mapped to reference genome portions can be normalized using PERUN methodology.

Cellular nucleic acid, in some embodiments, is an association with one or more proteins, and an agent that captures protein-associated nucleic acid can be utilized to enrich for the latter, in some embodiments. An agent in certain cases is an antibody or antibody fragment that specifically binds to a protein in association with cellular nucleic acid (e.g., an antibody that specifically binds to a chromatin protein (e.g., histone protein)). Processes in which an antibody or antibody fragment is used to enrich for cellular nucleic acid bound to a particular protein sometimes are referred to chromatin immunoprecipitation (ChIP) processes. ChIP-enriched nucleic acid is a nucleic acid in

association with cellular protein, such as DNA or RNA for example. Reads of ChIP-enriched nucleic acid can be obtained using technology known in the art. Reads of ChIP-enriched nucleic acid can be mapped to one or more portions of a reference genome, and results can be normalized using PERUN methodology for providing an outcome.

Thus, provided in certain embodiments are methods for calculating with reduced bias genomic section elevations for a test sample, comprising: (a) obtaining counts of sequence reads mapped to bins of a reference genome, which sequence reads are reads of cellular nucleic acid from a test sample obtained by isolation of a protein to which the nucleic acid was associated; (b) determining experimental bias for each of the bins across multiple samples from a fitted relation between (i) the counts of the sequence reads mapped to each of the bins, and (ii) a mapping feature for each of the bins; and (c) calculating a genomic section elevation for each of the bins from a fitted relation between the experimental bias and the counts of the sequence reads mapped to each of the bins, thereby providing calculated genomic section elevations, whereby bias in the counts of the sequence reads mapped to each of the bins is reduced in the calculated genomic section elevations.

In certain embodiments, cellular RNA can serve as nucleic acid indicators. Cellular RNA reads can be mapped to reference RNA portions and normalized using PERUN methodology for providing an outcome. Known sequences for cellular RNA, referred to as a transcriptome, or a segment thereof, can be used as a reference to which RNA reads from a sample can be mapped. Reads of sample RNA can be obtained using technology known in the art. Results of RNA reads mapped to a reference can be normalized using PERUN methodology for providing an outcome.

Thus, provided in some embodiments are methods for calculating with reduced bias genomic section elevations for a test sample, comprising: (a) obtaining counts of sequence reads mapped to bins of reference RNA (e.g., reference transcriptome or segment(s) thereof), which sequence reads are reads of cellular RNA from a test sample; (b) determining experimental bias for each of the bins across multiple samples from a fitted relation between (i) the counts of the sequence reads mapped to each of the bins, and (ii) a mapping feature for each of the bins; and (c) calculating a genomic section elevation for each of the bins from a fitted relation between the experimental bias and the counts of the sequence reads mapped to each of the bins, thereby providing calculated genomic section elevations, whereby bias in the counts of the sequence reads mapped to each of the bins is reduced in the calculated genomic section elevations.

In some embodiments, microarray nucleic acid levels can serve as nucleic acid indicators. Nucleic acid levels across samples for a particular address, or hybridizing nucleic acid, on an array can be analyzed using PERUN methodology, thereby normalizing nucleic acid indicators provided by microarray analysis. In this manner, a particular address or hybridizing nucleic acid on a microarray is analogous to a bin for mapped nucleic acid sequence reads, and PERUN methodology can be used to normalize microarray data to provide an improved outcome.

Thus, provided in certain embodiments are methods for reducing microarray nucleic acid level error for a test sample, comprising: (a) obtaining nucleic acid levels in a microarray to which test sample nucleic acid has been associated, which microarray includes an array of capture nucleic acids; (b) determining experimental bias for each of the capture nucleic acids across multiple samples from a fitted relation between (i) the test sample nucleic acid levels associated with each of the capture nucleic acids, and (ii) an association feature for each of the capture nucleic acids; and (c) calculating a test sample nucleic acid level for each of the capture nucleic acids from a fitted relation between the experimental bias and the levels of the test sample nucleic acid associated with each of the capture nucleic acids, thereby providing calculated levels, whereby bias in the levels of test sample nucleic acid associated with each of the capture nucleic acids is reduced in the calculated levels. The association feature mentioned above can be any feature correlated with hybridization of a test sample nucleic acid to a capture nucleic acid that gives rise to, or may give rise to, error in determining the level of test sample nucleic acid associated with a capture nucleic acid.

In some embodiments, a processing step comprises a weighting. The terms “weighted”, “weighting” or “weight function” or grammatical derivatives or equivalents thereof, as used herein, refer to a mathematical manipulation of a portion or all of a data set sometimes utilized to alter the influence of certain data set features or variables with respect to other data set features or variables (e.g., increase or decrease the significance and/or contribution of data contained in one or more genomic sections or bins, based on the quality or usefulness of the data in the selected bin or bins). A weighting function can be used to increase the influence of data with a relatively small measurement variance, and/or to decrease the influence of data with a relatively large measurement variance, in some embodiments. For example, bins with under represented or low quality sequence data can be “down weighted” to minimize the influence on a data set, whereas selected bins can be “up weighted” to increase the influence on a data set. A non-limiting example

of a weighting function is [1 / (standard deviation)^2]. A weighting step sometimes is performed in a manner substantially similar to a normalizing step. In some embodiments, a data set is divided by a predetermined variable (e.g., weighting variable). A predetermined variable (e.g., minimized target function, Phi) often is selected to weigh different parts of a data set differently (e.g., increase the influence of certain data types while decreasing the influence of other data types).

In certain embodiments, a processing step can comprise one or more mathematical and/or statistical manipulations. Any suitable mathematical and/or statistical manipulation, alone or in combination, may be used to analyze and/or manipulate a data set described herein. Any suitable number of mathematical and/or statistical manipulations can be used. In some embodiments, a data set can be mathematically and/or statistically manipulated 1 or more, 5 or more, 10 or more or 20 or more times. Non-limiting examples of mathematical and statistical manipulations that can be used include addition, subtraction, multiplication, division, algebraic functions, least squares estimators, curve fitting, differential equations, rational polynomials, double polynomials, orthogonal polynomials, z-scores, p-values, chi values, phi values, analysis of peak elevations, determination of peak edge locations, calculation of peak area ratios, analysis of median chromosomal elevation, calculation of mean absolute deviation, sum of squared residuals, mean, standard deviation, standard error, the like or combinations thereof. A mathematical and/or statistical manipulation can be performed on all or a portion of sequence read data, or processed products thereof. Non-limiting examples of data set variables or features that can be statistically manipulated include raw counts, filtered counts, normalized counts, peak heights, peak widths, peak areas, peak edges, lateral tolerances, P-values, median elevations, mean elevations, count distribution within a genomic region, relative representation of nucleic acid species, the like or combinations thereof.

In some embodiments, a processing step can include the use of one or more statistical algorithms. Any suitable statistical algorithm, alone or in combination, may be used to analyze and/or manipulate a data set described herein. Any suitable number of statistical algorithms can be used. In some embodiments, a data set can be analyzed using 1 or more, 5 or more, 10 or more or 20 or more statistical algorithms. Non-limiting examples of statistical algorithms suitable for use with methods described herein include decision trees, counternulls, multiple comparisons, omnibus test, Behrens-Fisher problem, bootstrapping, Fisher's method for combining independent tests of significance, null hypothesis, type I error, type II error, exact test, one-sample Z test, two-sample Z test, one-sample t-test, paired t-test, two-sample pooled t-test having equal variances, two-sample

unpooled t-test having unequal variances, one-proportion z-test, two-proportion z-test pooled, two-proportion z-test unpooled, one-sample chi-square test, two-sample F test for equality of variances, confidence interval, credible interval, significance, meta analysis, simple linear regression, robust linear regression, the like or combinations of the foregoing. Non-limiting examples of data set variables or features that can be analyzed using statistical algorithms include raw counts, filtered counts, normalized counts, peak heights, peak widths, peak edges, lateral tolerances, P-values, median elevations, mean elevations, count distribution within a genomic region, relative representation of nucleic acid species, the like or combinations thereof.

In certain embodiments, a data set can be analyzed by utilizing multiple (e.g., 2 or more) statistical algorithms (e.g., least squares regression, principle component analysis, linear discriminant analysis, quadratic discriminant analysis, bagging, neural networks, support vector machine models, random forests, classification tree models, K-nearest neighbors, logistic regression and/or loss smoothing) and/or mathematical and/or statistical manipulations (e.g., referred to herein as manipulations). The use of multiple manipulations can generate an N-dimensional space that can be used to provide an outcome, in some embodiments. In certain embodiments, analysis of a data set by utilizing multiple manipulations can reduce the complexity and/or dimensionality of the data set. For example, the use of multiple manipulations on a reference data set can generate an N-dimensional space (e.g., probability plot) that can be used to represent the presence or absence of a genetic variation, depending on the genetic status of the reference samples (e.g., positive or negative for a selected genetic variation). Analysis of test samples using a substantially similar set of manipulations can be used to generate an N-dimensional point for each of the test samples. The complexity and/or dimensionality of a test subject data set sometimes is reduced to a single value or N-dimensional point that can be readily compared to the N-dimensional space generated from the reference data. Test sample data that fall within the N-dimensional space populated by the reference subject data are indicative of a genetic status substantially similar to that of the reference subjects. Test sample data that fall outside of the N-dimensional space populated by the reference subject data are indicative of a genetic status substantially dissimilar to that of the reference subjects. In some embodiments, references are euploid or do not otherwise have a genetic variation or medical condition.

After data sets have been counted, optionally filtered and normalized, the processed data sets can be further manipulated by one or more filtering and/or normalizing procedures, in some embodiments. A data set that has been further manipulated by one or more filtering and/or

normalizing procedures can be used to generate a profile, in certain embodiments. The one or more filtering and/or normalizing procedures sometimes can reduce data set complexity and/or dimensionality, in some embodiments. An outcome can be provided based on a data set of reduced complexity and/or dimensionality.

Non-limiting examples of genomic section filtering is provided herein in Example 4 with respect to PERUN methods. Genomic sections may be filtered based on, or based on part on, a measure of error. A measure of error comprising absolute values of deviation, such as an R-factor, can be used for genomic section removal or weighting in certain embodiments. An R-factor, in some embodiments, is defined as the sum of the absolute deviations of the predicted count values from the actual measurements divided by the predicted count values from the actual measurements (e.g., Equation B herein). While a measure of error comprising absolute values of deviation may be used, a suitable measure of error may be alternatively employed. In certain embodiments, a measure of error not comprising absolute values of deviation, such as a dispersion based on squares, may be utilized. In some embodiments, genomic sections are filtered or weighted according to a measure of mappability (e.g., a mappability score; Example 5). A genomic section sometimes is filtered or weighted according to a relatively low number of sequence reads mapped to the genomic section (e.g., 0, 1, 2, 3, 4, 5 reads mapped to the genomic section). Genomic sections can be filtered or weighted according to the type of analysis being performed. For example, for autosome aneuploidy analysis (e.g., chromosome 13, 18, 9, 16, 20, 22, 21 and the like), sex chromosomes may be filtered, and only autosomes, or a subset of autosomes, may be analyzed.

In particular embodiments, the following filtering process may be employed. The same set of genomic sections (e.g., bins) within a given chromosome (e.g., chromosome 21) are selected and the number of reads in affected and unaffected samples are compared. The gap relates trisomy 21 and euploid samples and it involves a set of genomic sections covering most of chromosome 21. The set of genomic sections is the same between euploid and T21 samples. The distinction between a set of genomic sections and a single section is not crucial, as a genomic section can be defined. The same genomic region is compared in different patients. This process can be utilized for a trisomy analysis, such as for T4, T9, T13, T16, T20, T22 or T18 in addition to, or instead of, T21.

After data sets have been counted, optionally filtered and normalized, the processed data sets can be manipulated by weighting, in some embodiments. One or more genomic sections can be selected for weighting to reduce the influence of data (e.g., noisy data, uninformative data) contained in the selected genomic sections, in certain embodiments, and in some embodiments, one or more genomic sections can be selected for weighting to enhance or augment the influence of data (e.g., data with small measured variance) contained in the selected genomic sections. In some embodiments, a data set is weighted utilizing a single weighting function that decreases the influence of data with large variances and increases the influence of data with small variances. A weighting function sometimes is used to reduce the influence of data with large variances and augment the influence of data with small variances (e.g., $[1/(\text{standard deviation})^2]$). In some embodiments, a profile plot of processed data further manipulated by weighting is generated to facilitate classification and/or providing an outcome. An outcome can be provided based on a profile plot of weighted data

Filtering or weighting of genomic sections can be performed at one or more suitable points in an analysis. For example, genomic sections may be filtered or weighted before or after sequence reads are mapped to portions of a reference genome. Genomic sections may be filtered or weighted before or after an experimental bias for individual genome portions is determined in some embodiments. In certain embodiments, genomic sections may be filtered or weighted before or after genomic section elevations are calculated.

After data sets have been counted, optionally filtered, normalized, and optionally weighted, the processed data sets can be manipulated by one or more mathematical and/or statistical (e.g., statistical functions or statistical algorithm) manipulations, in some embodiments. In certain embodiments, processed data sets can be further manipulated by calculating Z-scores for one or more selected genomic sections, chromosomes, or portions of chromosomes. In some embodiments, processed data sets can be further manipulated by calculating P-values. Formulas for calculating Z-scores and P-values are presented in Example 1. In certain embodiments, mathematical and/or statistical manipulations include one or more assumptions pertaining to ploidy and/or fetal fraction. In some embodiments, a profile plot of processed data further manipulated by one or more statistical and/or mathematical manipulations is generated to facilitate classification and/or providing an outcome. An outcome can be provided based on a profile plot of statistically and/or mathematically manipulated data. An outcome provided based on a profile plot of

statistically and/or mathematically manipulated data often includes one or more assumptions pertaining to ploidy and/or fetal fraction.

In certain embodiments, multiple manipulations are performed on processed data sets to generate an N-dimensional space and/or N-dimensional point, after data sets have been counted, optionally filtered and normalized. An outcome can be provided based on a profile plot of data sets analyzed in N-dimensions.

In some embodiments, data sets are processed utilizing one or more peak elevation analysis, peak width analysis, peak edge location analysis, peak lateral tolerances, the like, derivations thereof, or combinations of the foregoing, as part of or after data sets have processed and/or manipulated. In some embodiments, a profile plot of data processed utilizing one or more peak elevation analysis, peak width analysis, peak edge location analysis, peak lateral tolerances, the like, derivations thereof, or combinations of the foregoing is generated to facilitate classification and/or providing an outcome. An outcome can be provided based on a profile plot of data that has been processed utilizing one or more peak elevation analysis, peak width analysis, peak edge location analysis, peak lateral tolerances, the like, derivations thereof, or combinations of the foregoing.

In some embodiments, the use of one or more reference samples known to be free of a genetic variation in question can be used to generate a reference median count profile, which may result in a predetermined value representative of the absence of the genetic variation, and often deviates from a predetermined value in areas corresponding to the genomic location in which the genetic variation is located in the test subject, if the test subject possessed the genetic variation. In test subjects at risk for, or suffering from a medical condition associated with a genetic variation, the numerical value for the selected genomic section or sections is expected to vary significantly from the predetermined value for non-affected genomic locations. In certain embodiments, the use of one or more reference samples known to carry the genetic variation in question can be used to generate a reference median count profile, which may result in a predetermined value representative of the presence of the genetic variation, and often deviates from a predetermined value in areas corresponding to the genomic location in which a test subject does not carry the genetic variation. In test subjects not at risk for, or suffering from a medical condition associated with a genetic variation, the numerical value for the selected genomic section or sections is expected to vary significantly from the predetermined value for affected genomic locations.

In some embodiments, analysis and processing of data can include the use of one or more assumptions. A suitable number or type of assumptions can be utilized to analyze or process a data set. Non-limiting examples of assumptions that can be used for data processing and/or analysis include maternal ploidy, fetal contribution, prevalence of certain sequences in a reference population, ethnic background, prevalence of a selected medical condition in related family members, parallelism between raw count profiles from different patients and/or runs after GC-normalization and repeat masking (e.g., GCRM), identical matches represent PCR artifacts (e.g., identical base position), assumptions inherent in a fetal quantifier assay (e.g., FQA), assumptions regarding twins (e.g., if 2 twins and only 1 is affected the effective fetal fraction is only 50% of the total measured fetal fraction (similarly for triplets, quadruplets and the like)), fetal cell free DNA (e.g., cfDNA) uniformly covers the entire genome, the like and combinations thereof.

In those instances where the quality and/or depth of mapped sequence reads does not permit an outcome prediction of the presence or absence of a genetic variation at a desired confidence level (e.g., 95% or higher confidence level), based on the normalized count profiles, one or more additional mathematical manipulation algorithms and/or statistical prediction algorithms, can be utilized to generate additional numerical values useful for data analysis and/or providing an outcome. The term "normalized count profile" as used herein refers to a profile generated using normalized counts. Examples of methods that can be used to generate normalized counts and normalized count profiles are described herein. As noted, mapped sequence reads that have been counted can be normalized with respect to test sample counts or reference sample counts. In some embodiments, a normalized count profile can be presented as a plot.

Profiles

In some embodiments, a processing step can comprise generating one or more profiles (e.g., profile plot) from various aspects of a data set or derivation thereof (e.g., product of one or more mathematical and/or statistical data processing steps known in the art and/or described herein). The term "profile" as used herein refers to a product of a mathematical and/or statistical manipulation of data that can facilitate identification of patterns and/or correlations in large quantities of data. A "profile" often includes values resulting from one or more manipulations of data or data sets, based on one or more criteria. A profile often includes multiple data points. Any suitable number of data points may be included in a profile depending on the nature and/or complexity of a data set. In certain embodiments, profiles may include 2 or more data points, 3 or

more data points, 5 or more data points, 10 or more data points, 24 or more data points, 25 or more data points, 50 or more data points, 100 or more data points, 500 or more data points, 1000 or more data points, 5000 or more data points, 10,000 or more data points, or 100,000 or more data points.

In some embodiments, a profile is representative of the entirety of a data set, and in certain embodiments, a profile is representative of a portion or subset of a data set. That is, a profile sometimes includes or is generated from data points representative of data that has not been filtered to remove any data, and sometimes a profile includes or is generated from data points representative of data that has been filtered to remove unwanted data. In some embodiments, a data point in a profile represents the results of data manipulation for a genomic section. In certain embodiments, a data point in a profile includes results of data manipulation for groups of genomic sections. In some embodiments, groups of genomic sections may be adjacent to one another, and in certain embodiments, groups of genomic sections may be from different parts of a chromosome or genome.

Data points in a profile derived from a data set can be representative of any suitable data categorization. Non-limiting examples of categories into which data can be grouped to generate profile data points include: genomic sections based on size, genomic sections based on sequence features (e.g., GC content, AT content, position on a chromosome (e.g., short arm, long arm, centromere, telomere), and the like), levels of expression, chromosome, the like or combinations thereof. In some embodiments, a profile may be generated from data points obtained from another profile (e.g., normalized data profile renormalized to a different normalizing value to generate a renormalized data profile). In certain embodiments, a profile generated from data points obtained from another profile reduces the number of data points and/or complexity of the data set. Reducing the number of data points and/or complexity of a data set often facilitates interpretation of data and/or facilitates providing an outcome.

A profile often is a collection of normalized or non-normalized counts for two or more genomic sections. A profile often includes at least one elevation, and often comprises two or more elevations (e.g., a profile often has multiple elevations). An elevation generally is for a set of genomic sections having about the same counts or normalized counts. Elevations are described in greater detail herein. In certain embodiments, a profile comprises one or more genomic sections, which genomic sections can be weighted, removed, filtered, normalized, adjusted, averaged,

derived as a mean, added, subtracted, processed or transformed by any combination thereof. A profile often comprises normalized counts mapped to genomic sections defining two or more elevations, where the counts are further normalized according to one of the elevations by a suitable method. Often counts of a profile (e.g., a profile elevation) are associated with an uncertainty value.

A profile comprising one or more elevations can include a first elevation and a second elevation. In certain embodiments a first elevation is different (e.g., significantly different) than a second elevation. In some embodiments a first elevation comprises a first set of genomic sections, a second elevation comprises a second set of genomic sections and the first set of genomic sections is not a subset of the second set of genomic sections. In certain embodiments, a first set of genomic sections is different than a second set of genomic sections from which a first and second elevation are determined. In certain embodiments a profile can have multiple first elevations that are different (e.g., significantly different, e.g., have a significantly different value) than a second elevation within the profile. In certain embodiments a profile comprises one or more first elevations that are significantly different than a second elevation within the profile and one or more of the first elevations are adjusted. In certain embodiments a profile comprises one or more first elevations that are significantly different than a second elevation within the profile, each of the one or more first elevations comprise a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation and one or more of the first elevations are adjusted. In certain embodiments a first elevation within a profile is removed from the profile or adjusted (e.g., padded). A profile can comprise multiple elevations that include one or more first elevations significantly different than one or more second elevations and often the majority of elevations in a profile are second elevations, which second elevations are about equal to one another. In certain embodiments greater than 50%, greater than 60%, greater than 70%, greater than 80%, greater than 90% or greater than 95% of the elevations in a profile are second elevations.

A profile sometimes is displayed as a plot. For example, one or more elevations representing counts (e.g., normalized counts) of genomic sections can be plotted and visualized. Non-limiting examples of profile plots that can be generated include raw count (e.g., raw count profile or raw profile), normalized count, bin-weighted, z-score, p-value, area ratio versus fitted ploidy, median elevation versus ratio between fitted and measured fetal fraction, principle components, the like, or combinations thereof. Profile plots allow visualization of the manipulated data, in some

embodiments. In certain embodiments, a profile plot can be utilized to provide an outcome (e.g., area ratio versus fitted ploidy, median elevation versus ratio between fitted and measured fetal fraction, principle components). The terms "raw count profile plot" or "raw profile plot" as used herein refer to a plot of counts in each genomic section in a region normalized to total counts in a region (e.g., genome, genomic section, chromosome, chromosome bins or a segment of a chromosome). In some embodiments, a profile can be generated using a static window process, and in certain embodiments, a profile can be generated using a sliding window process.

A profile generated for a test subject sometimes is compared to a profile generated for one or more reference subjects, to facilitate interpretation of mathematical and/or statistical manipulations of a data set and/or to provide an outcome. In some embodiments, a profile is generated based on one or more starting assumptions (e.g., maternal contribution of nucleic acid (e.g., maternal fraction), fetal contribution of nucleic acid (e.g., fetal fraction), ploidy of reference sample, the like or combinations thereof). In certain embodiments, a test profile often centers around a predetermined value representative of the absence of a genetic variation, and often deviates from a predetermined value in areas corresponding to the genomic location in which the genetic variation is located in the test subject, if the test subject possessed the genetic variation. In test subjects at risk for, or suffering from a medical condition associated with a genetic variation, the numerical value for a selected genomic section is expected to vary significantly from the predetermined value for non-affected genomic locations. Depending on starting assumptions (e.g., fixed ploidy or optimized ploidy, fixed fetal fraction or optimized fetal fraction or combinations thereof) the predetermined threshold or cutoff value or threshold range of values indicative of the presence or absence of a genetic variation can vary while still providing an outcome useful for determining the presence or absence of a genetic variation. In some embodiments, a profile is indicative of and/or representative of a phenotype.

By way of a non-limiting example, normalized sample and/or reference count profiles can be obtained from raw sequence read data by (a) calculating reference median counts for selected chromosomes, genomic sections or segments thereof from a set of references known not to carry a genetic variation, (b) removal of uninformative genomic sections from the reference sample raw counts (e.g., filtering); (c) normalizing the reference counts for all remaining bins to the total residual number of counts (e.g., sum of remaining counts after removal of uninformative bins) for the reference sample selected chromosome or selected genomic location, thereby generating a normalized reference subject profile; (d) removing the corresponding genomic sections from the

test subject sample; and (e) normalizing the remaining test subject counts for one or more selected genomic locations to the sum of the residual reference median counts for the chromosome or chromosomes containing the selected genomic locations, thereby generating a normalized test subject profile. In certain embodiments, an additional normalizing step with respect to the entire genome, reduced by the filtered genomic sections in (b), can be included between (c) and (d). A data set profile can be generated by one or more manipulations of counted mapped sequence read data. Some embodiments include the following. Sequence reads are mapped and the number of sequence tags mapping to each genomic bin are determined (e.g., counted). A raw count profile is generated from the mapped sequence reads that are counted. An outcome is provided by comparing a raw count profile from a test subject to a reference median count profile for chromosomes, genomic sections or segments thereof from a set of reference subjects known not to possess a genetic variation, in certain embodiments.

In some embodiments, sequence read data is optionally filtered to remove noisy data or uninformative genomic sections. After filtering, the remaining counts typically are summed to generate a filtered data set. A filtered count profile is generated from a filtered data set, in certain embodiments.

After sequence read data have been counted and optionally filtered, data sets can be normalized to generate elevations or profiles. A data set can be normalized by normalizing one or more selected genomic sections to a suitable normalizing reference value. In some embodiments, a normalizing reference value is representative of the total counts for the chromosome or chromosomes from which genomic sections are selected. In certain embodiments, a normalizing reference value is representative of one or more corresponding genomic sections, portions of chromosomes or chromosomes from a reference data set prepared from a set of reference subjects known not to possess a genetic variation. In some embodiments, a normalizing reference value is representative of one or more corresponding genomic sections, portions of chromosomes or chromosomes from a test subject data set prepared from a test subject being analyzed for the presence or absence of a genetic variation. In certain embodiments, the normalizing process is performed utilizing a static window approach, and in some embodiments the normalizing process is performed utilizing a moving or sliding window approach. In certain embodiments, a profile comprising normalized counts is generated to facilitate classification and/or providing an outcome. An outcome can be provided based on a plot of a profile comprising normalized counts (e.g., using a plot of such a profile).

Elevations

In some embodiments, a value is ascribed to an elevation (e.g., a number). An elevation can be determined by a suitable method, operation or mathematical process (e.g., a processed elevation). An elevation often is, or is derived from, counts (e.g., normalized counts) for a set of genomic sections. In certain embodiments an elevation of a genomic section is substantially equal to the total number of counts mapped to a genomic section (e.g., normalized counts). Often an elevation is determined from counts that are processed, transformed or manipulated by a suitable method, operation or mathematical process known in the art. In certain embodiments an elevation is derived from counts that are processed and non-limiting examples of processed counts include weighted, removed, filtered, normalized, adjusted, averaged, derived as a mean (e.g., mean elevation), added, subtracted, transformed counts or combination thereof. In certain embodiments an elevation comprises counts that are normalized (e.g., normalized counts of genomic sections). An elevation can be for counts normalized by a suitable process, non-limiting examples of which include bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, GC LOESS, LOWESS, PERUN, RM, GCRM, cQn, the like and/or combinations thereof. An elevation can comprise normalized counts or relative amounts of counts. In certain embodiments an elevation is for counts or normalized counts of two or more genomic sections that are averaged and the elevation is referred to as an average elevation. In certain embodiments an elevation is for a set of genomic sections having a mean count or mean of normalized counts which is referred to as a mean elevation. In certain embodiments an elevation is derived for genomic sections that comprise raw and/or filtered counts. In some embodiments, an elevation is based on counts that are raw. In certain embodiments an elevation is associated with an uncertainty value. An elevation for a genomic section, or a "genomic section elevation," is synonymous with a "genomic section level" herein.

Normalized or non-normalized counts for two or more elevations (e.g., two or more elevations in a profile) can sometimes be mathematically manipulated (e.g., added, multiplied, averaged, normalized, the like or combination thereof) according to elevations. For example, normalized or non-normalized counts for two or more elevations can be normalized according to one, some or all of the elevations in a profile. In certain embodiments normalized or non-normalized counts of all elevations in a profile are normalized according to one elevation in the profile. In certain

embodiments normalized or non-normalized counts of a fist elevation in a profile are normalized according to normalized or non-normalized counts of a second elevation in the profile.

Non-limiting examples of an elevation (e.g., a first elevation, a second elevation) are an elevation for a set of genomic sections comprising processed counts, an elevation for a set of genomic sections comprising a mean, median or average of counts, an elevation for a set of genomic sections comprising normalized counts, the like or any combination thereof. In some embodiments, a first elevation and a second elevation in a profile are derived from counts of genomic sections mapped to the same chromosome. In some embodiments, a first elevation and a second elevation in a profile are derived from counts of genomic sections mapped to different chromosomes.

In some embodiments an elevation is determined from normalized or non-normalized counts mapped to one or more genomic sections. In some embodiments, an elevation is determined from normalized or non-normalized counts mapped to two or more genomic sections, where the normalized counts for each genomic section often are about the same. There can be variation in counts (e.g., normalized counts) in a set of genomic sections for an elevation. In a set of genomic sections for an elevation there can be one or more genomic sections having counts that are significantly different than in other genomic sections of the set (e.g., peaks and/or dips). Any suitable number of normalized or non-normalized counts associated with any suitable number of genomic sections can define an elevation.

In certain embodiments one or more elevations can be determined from normalized or non-normalized counts of all or some of the genomic sections of a genome. Often an elevation can be determined from all or some of the normalized or non-normalized counts of a chromosome, or segment thereof. In certain embodiments, two or more counts derived from two or more genomic sections (e.g., a set of genomic sections) determine an elevation. In certain embodiments two or more counts (e.g., counts from two or more genomic sections) determine an elevation. In some embodiments, counts from 2 to about 100,000 genomic sections determine an elevation. In some embodiments, counts from 2 to about 50,000, 2 to about 40,000, 2 to about 30,000, 2 to about 20,000, 2 to about 10,000, 2 to about 5000, 2 to about 2500, 2 to about 1250, 2 to about 1000, 2 to about 500, 2 to about 250, 2 to about 100 or 2 to about 60 genomic sections determine an elevation. In some embodiments counts from about 10 to about 50 genomic sections determine an elevation. In some embodiments counts from about 20 to about 40 or more genomic sections

determine an elevation. In some embodiments, an elevation comprises counts from about 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, 45, 50, 55, 60 or more genomic sections. In some embodiments, an elevation corresponds to a set of genomic sections (e.g., a set of genomic sections of a reference genome, a set of genomic sections of a chromosome or a set of genomic sections of a segment of a chromosome).

In some embodiments, an elevation is determined for normalized or non-normalized counts of genomic sections that are contiguous. In certain embodiments genomic sections (e.g., a set of genomic sections) that are contiguous represent neighboring segments of a genome or neighboring segments of a chromosome or gene. For example, two or more contiguous genomic sections, when aligned by merging the genomic sections end to end, can represent a sequence assembly of a DNA sequence longer than each genomic section. For example two or more contiguous genomic sections can represent of an intact genome, chromosome, gene, intron, exon or segment thereof. In certain embodiments an elevation is determined from a collection (e.g., a set) of contiguous genomic sections and/or non-contiguous genomic sections.

Significantly Different Elevations

In some embodiments, a profile of normalized counts comprises an elevation (e.g., a first elevation) significantly different than another elevation (e.g., a second elevation) within the profile. A first elevation may be higher or lower than a second elevation. In some embodiments, a first elevation is for a set of genomic sections comprising one or more reads comprising a copy number variation (e.g., a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation) and the second elevation is for a set of genomic sections comprising reads having substantially no copy number variation. In some embodiments, significantly different refers to an observable difference. In certain embodiments significantly different refers to statistically different or a statistically significant difference. A statistically significant difference is sometimes a statistical assessment of an observed difference. A statistically significant difference can be assessed by a suitable method in the art. Any suitable threshold or range can be used to determine that two elevations are significantly different. In certain embodiments two elevations (e.g., mean elevations) that differ by about 0.01 percent or more (e.g., 0.01 percent of one or either of the elevation values) are significantly different. In certain embodiments two elevations (e.g., mean elevations) that differ by about 0.1 percent or

more are significantly different. In certain embodiments, two elevations (e.g., mean elevations) that differ by about 0.5 percent or more are significantly different. In certain embodiments two elevations (e.g., mean elevations) that differ by about 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or more than about 10% are significantly different. In certain embodiments two elevations (e.g., mean elevations) are significantly different and there is no overlap in either elevation and/or no overlap in a range defined by an uncertainty value calculated for one or both elevations. In certain embodiments the uncertainty value is a standard deviation expressed as sigma. In certain embodiments two elevations (e.g., mean elevations) are significantly different and they differ by about 1 or more times the uncertainty value (e.g., 1 sigma). In certain embodiments two elevations (e.g., mean elevations) are significantly different and they differ by about 2 or more times the uncertainty value (e.g., 2 sigma), about 3 or more, about 4 or more, about 5 or more, about 6 or more, about 7 or more, about 8 or more, about 9 or more, or about 10 or more times the uncertainty value. In certain embodiments two elevations (e.g., mean elevations) are significantly different when they differ by about 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, or 4.0 times the uncertainty value or more. In some embodiments, the confidence level increases as the difference between two elevations increases. In certain embodiments, the confidence level decreases as the difference between two elevations decreases and/or as the uncertainty value increases. For example, sometimes the confidence level increases with the ratio of the difference between elevations and the standard deviation (e.g., MADs).

In some embodiments, a first set of genomic sections often includes genomic sections that are different than (e.g., non-overlapping with) a second set of genomic sections. For example, sometimes a first elevation of normalized counts is significantly different than a second elevation of normalized counts in a profile, and the first elevation is for a first set of genomic sections, the second elevation is for a second set of genomic sections and the genomic sections do not overlap in the first set and second set of genomic sections. In certain embodiments, a first set of genomic sections is not a subset of a second set of genomic sections from which a first elevation and second elevation are determined, respectively. In certain embodiments a first set of genomic sections is different and/or distinct from a second set of genomic sections from which a first elevation and second elevation are determined, respectively.

In certain embodiments a first set of genomic sections is a subset of a second set of genomic sections in a profile. For example, sometimes a second elevation of normalized counts for a

second set of genomic sections in a profile comprises normalized counts of a first set of genomic sections for a first elevation in the profile and the first set of genomic sections is a subset of the second set of genomic sections in the profile. In certain embodiments an average, mean or median elevation is derived from a second elevation where the second elevation comprises a first elevation. In certain embodiments, a second elevation comprises a second set of genomic sections representing an entire chromosome and a first elevation comprises a first set of genomic sections where the first set is a subset of the second set of genomic sections and the first elevation represents a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation that is present in the chromosome.

In some embodiments, a value of a second elevation is closer to the mean, average or median value of a count profile for a chromosome, or segment thereof, than the first elevation. In some embodiments, a second elevation is a mean elevation of a chromosome, a portion of a chromosome or a segment thereof. In some embodiments, a first elevation is significantly different from a predominant elevation (e.g., a second elevation) representing a chromosome, or segment thereof. A profile may include multiple first elevations that significantly differ from a second elevation, and each first elevation independently can be higher or lower than the second elevation. In some embodiments, a first elevation and a second elevation are derived from the same chromosome and the first elevation is higher or lower than the second elevation, and the second elevation is the predominant elevation of the chromosome. In certain embodiments, a first elevation and a second elevation are derived from the same chromosome, a first elevation is indicative of a copy number variation (e.g., a maternal and/or fetal copy number variation, deletion, insertion, duplication) and a second elevation is a mean elevation or predominant elevation of genomic sections for a chromosome, or segment thereof.

In certain embodiments, a read in a second set of genomic sections for a second elevation substantially does not include a genetic variation (e.g., a copy number variation, a maternal and/or fetal copy number variation). Often, a second set of genomic sections for a second elevation includes some variability (e.g., variability in elevation, variability in counts for genomic sections). In certain embodiments, one or more genomic sections in a set of genomic sections for an elevation associated with substantially no copy number variation include one or more reads having a copy number variation present in a maternal and/or fetal genome. For example, sometimes a set of genomic sections include a copy number variation that is present in a small segment of a chromosome (e.g., less than 10 genomic sections) and the set of genomic sections is for an

elevation associated with substantially no copy number variation. Thus a set of genomic sections that include substantially no copy number variation still can include a copy number variation that is present in less than about 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 genomic sections of an elevation.

In certain embodiments a first elevation is for a first set of genomic sections and a second elevation is for a second set of genomic sections and the first set of genomic sections and second set of genomic sections are contiguous (e.g., adjacent with respect to the nucleic acid sequence of a chromosome or segment thereof). In certain embodiments the first set of genomic sections and second set of genomic sections are not contiguous.

Relatively short sequence reads from a mixture of fetal and maternal nucleic acid can be utilized to provide counts which can be transformed into an elevation and/or a profile. Counts, elevations and profiles can be depicted in electronic or tangible form and can be visualized. Counts mapped to genomic sections (e.g., represented as elevations and/or profiles) can provide a visual representation of a fetal and/or a maternal genome, chromosome, or a portion or a segment of a chromosome that is present in a fetus and/or pregnant female.

Reference Elevation and Normalized Reference Value

In certain embodiments a profile comprises a reference elevation (e.g., an elevation used as a reference). Often a profile of normalized counts provides a reference elevation from which expected elevations and expected ranges are determined (see discussion below on expected elevations and ranges). A reference elevation often is for normalized counts of genomic sections comprising mapped reads from both a mother and a fetus. A reference elevation is often the sum of normalized counts of mapped reads from a fetus and a mother (e.g., a pregnant female). In certain embodiments a reference elevation is for genomic sections comprising mapped reads from a euploid mother and/or a euploid fetus. In certain embodiments a reference elevation is for genomic sections comprising mapped reads having a fetal genetic variation (e.g., an aneuploidy (e.g., a trisomy)), and/or reads having a maternal genetic variation (e.g., a copy number variation, insertion, deletion). In certain embodiments a reference elevation is for genomic sections that include substantially no maternal and/or fetal copy number variations. In certain embodiments a second elevation is used as a reference elevation. In certain embodiments a profile comprises a first elevation of normalized counts and a second elevation of normalized counts, the first elevation is significantly different from the second elevation and the second elevation is the reference

elevation. In certain embodiments a profile comprises a first elevation of normalized counts for a first set of genomic sections, a second elevation of normalized counts for a second set of genomic sections, the first set of genomic sections includes mapped reads having a maternal and/or fetal copy number variation, the second set of genomic sections comprises mapped reads having substantially no maternal copy number variation and/or fetal copy number variation, and the second elevation is a reference elevation.

In some embodiments counts mapped to genomic sections for one or more elevations of a profile are normalized according to counts of a reference elevation. In some embodiments, normalizing counts of an elevation according to counts of a reference elevation comprise dividing counts of an elevation by counts of a reference elevation or a multiple or fraction thereof. Counts normalized according to counts of a reference elevation often have been normalized according to another process (e.g., PERUN) and counts of a reference elevation also often have been normalized (e.g., by PERUN). In certain embodiments the counts of an elevation are normalized according to counts of a reference elevation and the counts of the reference elevation are scalable to a suitable value either prior to or after normalizing. The process of scaling the counts of a reference elevation can comprise any suitable constant (i.e., number) and any suitable mathematical manipulation may be applied to the counts of a reference elevation.

A normalized reference value (NRV) is often determined according to the normalized counts of a reference elevation. Determining an NRV can comprise any suitable normalization process (e.g., mathematical manipulation) applied to the counts of a reference elevation where the same normalization process is used to normalize the counts of other elevations within the same profile. Determining an NRV often comprises dividing a reference elevation by itself. Determining an NRV often comprises dividing a reference elevation by a multiple of itself. Determining an NRV often comprises dividing a reference elevation by the sum or difference of the reference elevation and a constant (e.g., any number).

An NRV is sometimes referred to as a null value. An NRV can be any suitable value. In some embodiments, an NRV is any value other than zero. In certain embodiments an NRV is a whole number. In certain embodiments an NRV is a positive integer. In some embodiments, an NRV is 1, 10, 100 or 1000. Often, an NRV is equal to 1. In certain embodiments an NRV is equal to zero. The counts of a reference elevation can be normalized to any suitable NRV. In some

embodiments, the counts of a reference elevation are normalized to an NRV of zero. Often the counts of a reference elevation are normalized to an NRV of 1.

Expected Elevations

An expected elevation is sometimes a pre-defined elevation (e.g., a theoretical elevation, predicted elevation). An "expected elevation" is sometimes referred to herein as a "predetermined elevation value". In some embodiments, an expected elevation is a predicted value for an elevation of normalized counts for a set of genomic sections that include a copy number variation. In certain embodiments, an expected elevation is determined for a set of genomic sections that include substantially no copy number variation. An expected elevation can be determined for a chromosome ploidy (e.g., 0, 1, 2 (i.e., diploid), 3 or 4 chromosomes) or a microploidy (homozygous or heterozygous deletion, duplication, insertion or absence thereof). Often an expected elevation is determined for a maternal microploidy (e.g., a maternal and/or fetal copy number variation).

An expected elevation for a genetic variation or a copy number variation can be determined by any suitable manner. Often an expected elevation is determined by a suitable mathematical manipulation of an elevation (e.g., counts mapped to a set of genomic sections for an elevation). In certain embodiments an expected elevation is determined by utilizing a constant sometimes referred to as an expected elevation constant. An expected elevation for a copy number variation is sometimes calculated by multiplying a reference elevation, normalized counts of a reference elevation or an NRV by an expected elevation constant, adding an expected elevation constant, subtracting an expected elevation constant, dividing by an expected elevation constant, or by a combination thereof. Often an expected elevation (e.g., an expected elevation of a maternal and/or fetal copy number variation) determined for the same subject, sample or test group is determined according to the same reference elevation or NRV.

Often an expected elevation is determined by multiplying a reference elevation, normalized counts of a reference elevation or an NRV by an expected elevation constant where the reference elevation, normalized counts of a reference elevation or NRV is not equal to zero. In certain embodiments an expected elevation is determined by adding an expected elevation constant to reference elevation, normalized counts of a reference elevation or an NRV that is equal to zero. In some embodiments, an expected elevation, normalized counts of a reference elevation, NRV and expected elevation constant are scalable. The process of scaling can comprise any suitable

constant (i.e., number) and any suitable mathematical manipulation where the same scaling process is applied to all values under consideration.

Expected Elevation Constant

An expected elevation constant can be determined by a suitable method. In certain embodiments an expected elevation constant is arbitrarily determined. Often an expected elevation constant is determined empirically. In certain embodiments an expected elevation constant is determined according to a mathematical manipulation. In certain embodiments an expected elevation constant is determined according to a reference (e.g., a reference genome, a reference sample, reference test data). In some embodiments, an expected elevation constant is predetermined for an elevation representative of the presence or absence of a genetic variation or copy number variation (e.g., a duplication, insertion or deletion). In some embodiments, an expected elevation constant is predetermined for an elevation representative of the presence or absence of a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation. An expected elevation constant for a copy number variation can be any suitable constant or set of constants.

In some embodiments, the expected elevation constant for a homozygous duplication (e.g., a homozygous duplication) can be from about 1.6 to about 2.4, from about 1.7 to about 2.3, from about 1.8 to about 2.2, or from about 1.9 to about 2.1. In certain embodiments the expected elevation constant for a homozygous duplication is about 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3 or about 2.4. Often the expected elevation constant for a homozygous duplication is about 1.90, 1.92, 1.94, 1.96, 1.98, 2.0, 2.02, 2.04, 2.06, 2.08 or about 2.10. Often the expected elevation constant for a homozygous duplication is about 2.

In some embodiments, the expected elevation constant for a heterozygous duplication (e.g., a homozygous duplication) is from about 1.2 to about 1.8, from about 1.3 to about 1.7, or from about 1.4 to about 1.6. In certain embodiments the expected elevation constant for a heterozygous duplication is about 1.2, 1.3, 1.4, 1.5, 1.6, 1.7 or about 1.8. Often the expected elevation constant for a heterozygous duplication is about 1.40, 1.42, 1.44, 1.46, 1.48, 1.5, 1.52, 1.54, 1.56, 1.58 or about 1.60. In some embodiments, the expected elevation constant for a heterozygous duplication is about 1.5.

In some embodiments, the expected elevation constant for the absence of a copy number variation (e.g., the absence of a maternal copy number variation and/or fetal copy number variation) is from about 1.3 to about 0.7, from about 1.2 to about 0.8, or from about 1.1 to about 0.9. In certain embodiments the expected elevation constant for the absence of a copy number variation is about 1.3, 1.2, 1.1, 1.0, 0.9, 0.8 or about 0.7. Often the expected elevation constant for the absence of a copy number variation is about 1.09, 1.08, 1.06, 1.04, 1.02, 1.0, 0.98, 0.96, 0.94, or about 0.92. In some embodiments, the expected elevation constant for the absence of a copy number variation is about 1.

In some embodiments, the expected elevation constant for a heterozygous deletion (e.g., a maternal, fetal, or a maternal and a fetal heterozygous deletion) is from about 0.2 to about 0.8, from about 0.3 to about 0.7, or from about 0.4 to about 0.6. In certain embodiments the expected elevation constant for a heterozygous deletion is about 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 or about 0.8. Often the expected elevation constant for a heterozygous deletion is about 0.40, 0.42, 0.44, 0.46, 0.48, 0.5, 0.52, 0.54, 0.56, 0.58 or about 0.60. In some embodiments, the expected elevation constant for a heterozygous deletion is about 0.5.

In some embodiments, the expected elevation constant for a homozygous deletion (e.g., a homozygous deletion) can be from about -0.4 to about 0.4, from about -0.3 to about 0.3, from about -0.2 to about 0.2, or from about -0.1 to about 0.1. In certain embodiments the expected elevation constant for a homozygous deletion is about -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3 or about 0.4. Often the expected elevation constant for a homozygous deletion is about -0.1, -0.08, -0.06, -0.04, -0.02, 0.0, 0.02, 0.04, 0.06, 0.08 or about 0.10. Often the expected elevation constant for a homozygous deletion is about 0.

Expected Elevation Range

In certain embodiments the presence or absence of a genetic variation or copy number variation (e.g., a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation) is determined by an elevation that falls within or outside of an expected elevation range. An expected elevation range is often determined according to an expected elevation. In certain embodiments an expected elevation range is determined for an elevation comprising substantially no genetic variation or substantially no copy number variation. A suitable method can be used to determine an expected elevation range.

In certain embodiments, an expected elevation range is defined according to a suitable uncertainty value calculated for an elevation. Non-limiting examples of an uncertainty value are a standard deviation, standard error, calculated variance, p-value, and mean absolute deviation (MAD). In certain embodiments, an expected elevation range for a genetic variation or a copy number variation is determined, in part, by calculating the uncertainty value for an elevation (e.g., a first elevation, a second elevation, a first elevation and a second elevation). In certain embodiments an expected elevation range is defined according to an uncertainty value calculated for a profile (e.g., a profile of normalized counts for a chromosome or segment thereof). In some embodiments, an uncertainty value is calculated for an elevation comprising substantially no genetic variation or substantially no copy number variation. In some embodiments, an uncertainty value is calculated for a first elevation, a second elevation or a first elevation and a second elevation. In some embodiments an uncertainty value is determined for a first elevation, a second elevation or a second elevation comprising a first elevation.

An expected elevation range is sometimes calculated, in part, by multiplying, adding, subtracting, or dividing an uncertainty value by a constant (e.g., a predetermined constant) *n*. A suitable mathematical procedure or combination of procedures can be used. The constant *n* (e.g., predetermined constant *n*) is sometimes referred to as a confidence interval. A selected confidence interval is determined according to the constant *n* that is selected. The constant *n* (e.g., the predetermined constant *n*, the confidence interval) can be determined by a suitable manner. The constant *n* can be a number or fraction of a number greater than zero. The constant *n* can be a whole number. Often the constant *n* is a number less than 10. In certain embodiments the constant *n* is a number less than about 10, less than about 9, less than about 8, less than about 7, less than about 6, less than about 5, less than about 4, less than about 3, or less than about 2. In certain embodiments the constant *n* is about 10, 9.5, 9, 8.5, 8, 7.5, 7, 6.5, 6, 5.5, 5, 4.5, 4, 3.5, 3, 2.5, 2 or 1. The constant *n* can be determined empirically from data derived from subjects (a pregnant female and/or a fetus) with a known genetic disposition.

Often an uncertainty value and constant *n* defines a range (e.g.,, an uncertainty cutoff). For example, sometimes an uncertainty value is a standard deviation (e.g., +/- 5) and is multiplied by a constant *n* (e.g., a confidence interval) thereby defining a range or uncertainty cutoff (e.g., 5*n* to -5*n*).

In some embodiments, an expected elevation range for a genetic variation (e.g., a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and fetal copy number variation) is the sum of an expected elevation plus a constant *n* times the uncertainty (e.g., *n* x sigma (e.g., 6 sigma)). In certain embodiments the expected elevation range for a genetic variation or copy number variation designated by *k* can be defined by the formula:

Formula R: $(\text{Expected Elevation Range})_k = (\text{Expected Elevation})_k + n\sigma$

where σ is an uncertainty value, *n* is a constant (e.g., a predetermined constant) and the expected elevation range and expected elevation are for the genetic variation *k* (e.g., *k* = a heterozygous deletion, e.g., *k* = the absence of a genetic variation). For example, for an expected elevation equal to 1 (e.g., the absence of a copy number variation), an uncertainty value (i.e. σ) equal to +/- 0.05, and *n*=3, the expected elevation range is defined as 1.15 to 0.85. In some embodiments, the expected elevation range for a heterozygous duplication is determined as 1.65 to 1.35 when the expected elevation for a heterozygous duplication is 1.5, *n* = 3, and the uncertainty value σ is +/- 0.05. In some embodiments the expected elevation range for a heterozygous deletion is determined as 0.65 to 0.35 when the expected elevation for a heterozygous duplication is 0.5, *n* = 3, and the uncertainty value σ is +/- 0.05. In some embodiments the expected elevation range for a homozygous duplication is determined as 2.15 to 1.85 when the expected elevation for a heterozygous duplication is 2.0, *n* = 3 and the uncertainty value σ is +/- 0.05. In some embodiments the expected elevation range for a homozygous deletion is determined as 0.15 to -0.15 when the expected elevation for a heterozygous duplication is 0.0, *n* = 3 and the uncertainty value σ is +/- 0.05.

In certain embodiments an expected elevation range for a homozygous copy number variation (e.g., a maternal, fetal or maternal and fetal homozygous copy number variation) is determined, in part, according to an expected elevation range for a corresponding heterozygous copy number variation. For example, sometimes an expected elevation range for a homozygous duplication comprises all values greater than an upper limit of an expected elevation range for a heterozygous duplication. In certain embodiments an expected elevation range for a homozygous duplication comprises all values greater than or equal to an upper limit of an expected elevation range for a heterozygous duplication. In certain embodiments an expected elevation range for a homozygous duplication comprises all values greater than an upper limit of an expected elevation range for a heterozygous duplication and less than the upper limit defined by the formula R where σ is an

uncertainty value and is a positive value, *n* is a constant and *k* is a homozygous duplication. In certain embodiments an expected elevation range for a homozygous duplication comprises all values greater than or equal to an upper limit of an expected elevation range for a heterozygous duplication and less than or equal to the upper limit defined by the formula R where σ is an uncertainty value, σ is a positive value, *n* is a constant and *k* is a homozygous duplication.

In some embodiments, an expected elevation range for a homozygous deletion comprises all values less than a lower limit of an expected elevation range for a heterozygous deletion. In certain embodiments an expected elevation range for a homozygous deletion comprises all values less than or equal to a lower limit of an expected elevation range for a heterozygous deletion. In certain embodiments an expected elevation range for a homozygous deletion comprises all values less than a lower limit of an expected elevation range for a heterozygous deletion and greater than the lower limit defined by the formula R where σ is an uncertainty value, σ is a negative value, *n* is a constant and *k* is a homozygous deletion. In certain embodiments an expected elevation range for a homozygous deletion comprises all values less than or equal to a lower limit of an expected elevation range for a heterozygous deletion and greater than or equal to the lower limit defined by the formula R where σ is an uncertainty value, σ is a negative value, *n* is a constant and *k* is a homozygous deletion.

An uncertainty value can be utilized to determine a threshold value. In some embodiments, a range (e.g., a threshold range) is obtained by calculating the uncertainty value determined from a raw, filtered and/or normalized counts. A range can be determined by multiplying the uncertainty value for an elevation (e.g., normalized counts of an elevation) by a predetermined constant (e.g., 1, 2, 3, 4, 5, 6, etc.) representing the multiple of uncertainty (e.g., number of standard deviations) chosen as a cutoff threshold (e.g., multiply by 3 for 3 standard deviations), whereby a range is generated, in some embodiments. A range can be determined by adding and/or subtracting a value (e.g., a predetermined value, an uncertainty value, an uncertainty value multiplied by a predetermined constant) to and/or from an elevation whereby a range is generated, in some embodiments. For example, for an elevation equal to 1, a standard deviation of +/-0.2, where a predetermined constant is 3, the range can be calculated as (1 + 3(0.2)) to (1 + 3(-0.2)), or 1.6 to 0.4. A range sometimes can define an expected range or expected elevation range for a copy number variation. In certain embodiments, some or all of the genomic sections exceeding a threshold value, falling outside a range or falling inside a range of values, are removed as part of, prior to, or after a normalization process. In some embodiments, some or all of the genomic

sections exceeding a calculated threshold value, falling outside a range or falling inside a range are weighted or adjusted as part of, or prior to the normalization or classification process. Examples of weighting are described herein. The terms “redundant data”, and “redundant mapped reads” as used herein refer to sample derived sequence reads that are identified as having already been assigned to a genomic location (e.g., base position) and/or counted for a genomic section.

In some embodiments an uncertainty value is determined according to the formula below:

$$Z = \frac{L_A - L_o}{\sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_o^2}{N_o}}}$$

Where Z represents the standardized deviation between two elevations, L is the mean (or median) elevation and sigma is the standard deviation (or MAD). The subscript O denotes a segment of a profile (e.g., a second elevation, a chromosome, an NRV, a “euploid elevation”, an elevation absent a copy number variation), and A denotes another segment of a profile (e.g., a first elevation, an elevation representing a copy number variation, an elevation representing an aneuploidy (e.g., a trisomy). The variable N_o represents the total number of genomic sections in the segment of the profile denoted by the subscript O. N_A represents the total number of genomic sections in the segment of the profile denoted by subscript A.

Categorizing a Copy Number Variation

An elevation (e.g., a first elevation) that significantly differs from another elevation (e.g., a second elevation) can often be categorized as a copy number variation (e.g., a maternal and/or fetal copy number variation, a fetal copy number variation, a deletion, duplication, insertion) according to an expected elevation range. In some embodiments, the presence of a copy number variation is categorized when a first elevation is significantly different from a second elevation and the first elevation falls within the expected elevation range for a copy number variation. For example, a copy number variation (e.g., a maternal and/or fetal copy number variation, a fetal copy number variation) can be categorized when a first elevation is significantly different from a second elevation and the first elevation falls within the expected elevation range for a copy number variation. In certain embodiments a heterozygous duplication (e.g., a maternal or fetal, or maternal and fetal, heterozygous duplication) or heterozygous deletion (e.g., a maternal or fetal, or maternal and fetal,

heterozygous deletion) is categorized when a first elevation is significantly different from a second elevation and the first elevation falls within the expected elevation range for a heterozygous duplication or heterozygous deletion, respectively. In certain embodiments a homozygous duplication or homozygous deletion is categorized when a first elevation is significantly different from a second elevation and the first elevation falls within the expected elevation range for a homozygous duplication or homozygous deletion, respectively.

Fetal Fraction Determination Based on Elevation

In some embodiments, a fetal fraction is determined according to an elevation categorized as representative of a maternal and/or fetal copy number variation. For example determining fetal fraction often comprises assessing an expected elevation for a maternal and/or fetal copy number variation utilized for the determination of fetal fraction. In certain embodiments a fetal fraction is determined for an elevation (e.g., a first elevation) categorized as representative of a copy number variation according to an expected elevation range determined for the same type of copy number variation. Often a fetal fraction is determined according to an observed elevation that falls within an expected elevation range and is thereby categorized as a maternal and/or fetal copy number variation. In certain embodiments a fetal fraction is determined when an observed elevation (e.g., a first elevation) categorized as a maternal and/or fetal copy number variation is different than the expected elevation determined for the same maternal and/or fetal copy number variation.

In some embodiments an elevation (e.g., a first elevation, an observed elevation), is significantly different than a second elevation, the first elevation is categorized as a maternal and/or fetal copy number variation, and a fetal fraction is determined according to the first elevation. In certain embodiments a first elevation is an observed and/or experimentally obtained elevation that is significantly different than a second elevation in a profile and a fetal fraction is determined according to the first elevation. In certain embodiments the first elevation is an average, mean or summed elevation and a fetal fraction is determined according to the first elevation. In certain embodiments a first elevation and a second elevation are observed and/or experimentally obtained elevations and a fetal fraction is determined according to the first elevation. In some instances a first elevation comprises normalized counts for a first set of genomic sections and a second elevation comprises normalized counts for a second set of genomic sections and a fetal fraction is determined according to the first elevation. In certain embodiments a first set of genomic sections of a first elevation includes a copy number variation (e.g., the first elevation is representative of a

copy number variation) and a fetal fraction is determined according to the first elevation. In certain embodiments the first set of genomic sections of a first elevation includes a homozygous or heterozygous maternal copy number variation and a fetal fraction is determined according to the first elevation. In certain embodiments a profile comprises a first elevation for a first set of genomic sections and a second elevation for a second set of genomic sections, the second set of genomic sections includes substantially no copy number variation (e.g., a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation) and a fetal fraction is determined according to the first elevation.

In some embodiments an elevation (e.g., a first elevation, an observed elevation), is significantly different than a second elevation, the first elevation is categorized as for a maternal and/or fetal copy number variation, and a fetal fraction is determined according to the first elevation and/or an expected elevation of the copy number variation. In certain embodiments a first elevation is categorized as for a copy number variation according to an expected elevation for a copy number variation and a fetal fraction is determined according to a difference between the first elevation and the expected elevation. In certain embodiments an elevation (e.g., a first elevation, an observed elevation) is categorized as a maternal and/or fetal copy number variation, and a fetal fraction is determined as twice the difference between the first elevation and expected elevation of the copy number variation. In certain embodiments an elevation (e.g., a first elevation, an observed elevation) is categorized as a maternal and/or fetal copy number variation, the first elevation is subtracted from the expected elevation thereby providing a difference, and a fetal fraction is determined as twice the difference. In certain embodiments an elevation (e.g., a first elevation, an observed elevation) is categorized as a maternal and/or fetal copy number variation, an expected elevation is subtracted from a first elevation thereby providing a difference, and the fetal fraction is determined as twice the difference.

Often a fetal fraction is provided as a percent. For example, a fetal fraction can be multiplied or divided by 100 thereby providing a percent value. For example, for a first elevation representative of a maternal homozygous duplication and having an elevation of 155 and an expected elevation for a maternal homozygous duplication having an elevation of 150, a fetal fraction can be determined as 10% (e.g., (fetal fraction = 2 x (155 – 150)).

In some embodiments a fetal fraction is determined from two or more elevations within a profile that are categorized as copy number variations. For example, sometimes two or more elevations

(e.g., two or more first elevations) in a profile are identified as significantly different than a reference elevation (e.g., a second elevation, an elevation that includes substantially no copy number variation), the two or more elevations are categorized as representative of a maternal and/or fetal copy number variation and a fetal fraction is determined from each of the two or more elevations. In certain embodiments a fetal fraction is determined from about 3 or more, about 4 or more, about 5 or more, about 6 or more, about 7 or more, about 8 or more, or about 9 or more fetal fraction determinations within a profile. In certain embodiments a fetal fraction is determined from about 10 or more, about 20 or more, about 30 or more, about 40 or more, about 50 or more, about 60 or more, about 70 or more, about 80 or more, or about 90 or more fetal fraction determinations within a profile. In certain embodiments a fetal fraction is determined from about 100 or more, about 200 or more, about 300 or more, about 400 or more, about 500 or more, about 600 or more, about 700 or more, about 800 or more, about 900 or more, or about 1000 or more fetal fraction determinations within a profile. In certain embodiments a fetal fraction is determined from about 10 to about 1000, about 20 to about 900, about 30 to about 700, about 40 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, or about 50 to about 100 fetal fraction determinations within a profile.

In some embodiments a fetal fraction is determined as the average or mean of multiple fetal fraction determinations within a profile. In certain embodiments, a fetal fraction determined from multiple fetal fraction determinations is a mean (e.g., an average, a mean, a standard average, a median, or the like) of multiple fetal fraction determinations. Often a fetal fraction determined from multiple fetal fraction determinations is a mean value determined by a suitable method known in the art or described herein. In certain embodiments a mean value of a fetal fraction determination is a weighted mean. In certain embodiments a mean value of a fetal fraction determination is an unweighted mean. In some embodiments, a fetal fraction is determined from multiple first elevations significantly different than a second elevation where the elevation and/or independent fetal fraction determinations are weighted. In certain embodiments a fetal fraction determination is obtained from counts for multiple genomic sections where the counts for each genomic section or the counts for multiple sets of genomic sections are weighted. A mean, median or average fetal fraction determination (i.e., a mean, median or average fetal fraction determination value) generated from multiple fetal fraction determinations is sometimes associated with an uncertainty value (e.g., a variance, standard deviation, MAD, or the like). Before determining a mean, median or average fetal fraction value from multiple determinations, one or more deviant determinations are removed in some embodiments (described in greater detail herein).

Some fetal fraction determinations within a profile sometimes are not included in the overall determination of a fetal fraction (e.g., mean or average fetal fraction determination). In certain embodiments a fetal fraction determination is derived from a first elevation (e.g., a first elevation that is significantly different than a second elevation) in a profile and the first elevation is not indicative of a genetic variation. For example, some first elevations (e.g., spikes or dips) in a profile are generated from anomalies or unknown causes. Such values often generate fetal fraction determinations that differ significantly from other fetal fraction determinations obtained from true copy number variations. In certain embodiments fetal fraction determinations that differ significantly from other fetal fraction determinations in a profile are identified and removed from a fetal fraction determination. For example, some fetal fraction determinations obtained from anomalous spikes and dips are identified by comparing them to other fetal fraction determinations within a profile and are excluded from the overall determination of fetal fraction.

In certain embodiments, an independent fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination is an identified, recognized and/or observable difference. In certain embodiments, the term “differs significantly” can mean statistically different and/or a statistically significant difference. An “independent” fetal fraction determination can be a fetal fraction determined (e.g., in some cases a single determination) from a specific elevation categorized as a copy number variation. Any suitable threshold or range can be used to determine that a fetal fraction determination differs significantly from a mean, median or average fetal fraction determination. In certain embodiments a fetal fraction determination differs significantly from a mean, median or average fetal fraction determination and the determination can be expressed as a percent deviation from the average or mean value. In certain embodiments a fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination differs by about 10 percent or more. In certain embodiments a fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination differs by about 15 percent or more. In certain embodiments a fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination differs by about 15% to about 100% or more.

In certain embodiments a fetal fraction determination differs significantly from a mean, median or average fetal fraction determination according to a multiple of an uncertainty value associated with the mean or average fetal fraction determination. Often an uncertainty value and constant *n* (e.g., a confidence interval) defines a range (e.g., an uncertainty cutoff). For example, sometimes an

uncertainty value is a standard deviation for fetal fraction determinations (e.g., +/- 5) and is multiplied by a constant *n* (e.g., a confidence interval) thereby defining a range or uncertainty cutoff (e.g., 5*n* to -5*n,* sometimes referred to as 5 sigma). In certain embodiments an independent fetal fraction determination falls outside a range defined by the uncertainty cutoff and is considered significantly different from a mean, median or average fetal fraction determination. For example, for a mean value of 10 and an uncertainty cutoff of 3, an independent fetal fraction greater than 13 or less than 7 is significantly different. In certain embodiments a fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination differs by more than *n* times the uncertainty value (e.g., *n* x sigma) where *n* is about equal to or greater than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10. In certain embodiments a fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination differs by more than *n* times the uncertainty value (e.g., *n* x sigma) where *n* is about equal to or greater than 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, or 4.0.

In some embodiments, an elevation is representative of a fetal and/or maternal microploidy. In certain embodiments an elevation (e.g., a first elevation, an observed elevation), is significantly different than a second elevation, the first elevation is categorized as a maternal and/or fetal copy number variation, and the first elevation and/or second elevation is representative of a fetal microploidy and/or a maternal microploidy. In certain embodiments a first elevation is representative of a fetal microploidy, In certain embodiments a first elevation is representative of a maternal microploidy. Often a first elevation is representative of a fetal microploidy and a maternal microploidy. In certain embodiments an elevation (e.g., a first elevation, an observed elevation), is significantly different than a second elevation, the first elevation is categorized as a maternal and/or fetal copy number variation, the first elevation is representative of a fetal and/or maternal microploidy and a fetal fraction is determined according to the fetal and/or maternal microploidy. In some instances a first elevation is categorized as a maternal and/or fetal copy number variation, the first elevation is representative of a fetal microploidy and a fetal fraction is determined according to the fetal microploidy. In certain embodiments a first elevation is categorized as a maternal and/or fetal copy number variation, the first elevation is representative of a maternal microploidy and a fetal fraction is determined according to the maternal microploidy. In certain embodiments a first elevation is categorized as a maternal and/or fetal copy number variation, the first elevation is representative of a maternal and a fetal microploidy and a fetal fraction is determined according to the maternal and fetal microploidy.

In some embodiments, a determination of a fetal fraction comprises determining a fetal and/or maternal microploidy. In certain embodiments an elevation (e.g., a first elevation, an observed elevation), is significantly different than a second elevation, the first elevation is categorized as a maternal and/or fetal copy number variation, a fetal and/or maternal microploidy is determined according to the first elevation and/or second elevation and a fetal fraction is determined. In certain embodiments a first elevation is categorized as a maternal and/or fetal copy number variation, a fetal microploidy is determined according to the first elevation and/or second elevation and a fetal fraction is determined according to the fetal microploidy. In certain embodiments a first elevation is categorized as a maternal and/or fetal copy number variation, a maternal microploidy is determined according to the first elevation and/or second elevation and a fetal fraction is determined according to the maternal microploidy. In certain embodiments a first elevation is categorized as a maternal and/or fetal copy number variation, a maternal and fetal microploidy is determined according to the first elevation and/or second elevation and a fetal fraction is determined according to the maternal and fetal microploidy.

A fetal fraction often is determined when the microploidy of the mother is different from (e.g., not the same as) the microploidy of the fetus for a given elevation or for an elevation categorized as a copy number variation. In certain embodiments a fetal fraction is determined when the mother is homozygous for a duplication (e.g., a microploidy of 2) and the fetus is heterozygous for the same duplication (e.g., a microploidy of 1.5). In certain embodiments a fetal fraction is determined when the mother is heterozygous for a duplication (e.g., a microploidy of 1.5) and the fetus is homozygous for the same duplication (e.g., a microploidy of 2) or the duplication is absent in the fetus (e.g., a microploidy of 1). In certain embodiments a fetal fraction is determined when the mother is homozygous for a deletion (e.g., a microploidy of 0) and the fetus is heterozygous for the same deletion (e.g., a microploidy of 0.5). In certain embodiments a fetal fraction is determined when the mother is heterozygous for a deletion (e.g., a microploidy of 0.5) and the fetus is homozygous for the same deletion (e.g., a microploidy of 0) or the deletion is absent in the fetus (e.g., a microploidy of 1).

In certain embodiments, a fetal fraction cannot be determined when the microploidy of the mother is the same (e.g., identified as the same) as the microploidy of the fetus for a given elevation identified as a copy number variation. For example, for a given elevation where both the mother and fetus carry the same number of copies of a copy number variation, a fetal fraction is not determined, in some embodiments. For example, a fetal fraction cannot be determined for an

elevation categorized as a copy number variation when both the mother and fetus are homozygous for the same deletion or homozygous for the same duplication. In certain embodiments, a fetal fraction cannot be determined for an elevation categorized as a copy number variation when both the mother and fetus are heterozygous for the same deletion or heterozygous for the same duplication. In embodiments where multiple fetal fraction determinations are made for a sample, determinations that significantly deviate from a mean, median or average value can result from a copy number variation for which maternal ploidy is equal to fetal ploidy, and such determinations can be removed from consideration.

In some embodiments the microploidy of a maternal copy number variation and fetal copy number variation is unknown. In certain embodiments, in cases when there is no determination of fetal and/or maternal microploidy for a copy number variation, a fetal fraction is generated and compared to a mean, median or average fetal fraction determination. A fetal fraction determination for a copy number variation that differs significantly from a mean, median or average fetal fraction determination is sometimes because the microploidy of the mother and fetus are the same for the copy number variation. A fetal fraction determination that differs significantly from a mean, median or average fetal fraction determination is often excluded from an overall fetal fraction determination regardless of the source or cause of the difference. In some embodiments, the microploidy of the mother and/or fetus is determined and/or verified by a method known in the art (e.g., by targeted sequencing methods).

Elevation Adjustments

In some embodiments, one or more elevations are adjusted. A process for adjusting an elevation often is referred to as padding. In some embodiments, multiple elevations in a profile (e.g., a profile of a genome, a chromosome profile, a profile of a portion or segment of a chromosome) are adjusted. In certain embodiments, about 1 to about 10,000 or more elevations in a profile are adjusted. In certain embodiments about 1 to about a 1000, 1 to about 900, 1 to about 800, 1 to about 700, 1 to about 600, 1 to about 500, 1 to about 400, 1 to about 300, 1 to about 200, 1 to about 100, 1 to about 50, 1 to about 25, 1 to about 20, 1 to about 15, 1 to about 10, or 1 to about 5 elevations in a profile are adjusted. In certain embodiments one elevation is adjusted. In some embodiments, an elevation (e.g., a first elevation of a normalized count profile) that significantly differs from a second elevation is adjusted. In certain embodiments an elevation categorized as a copy number variation is adjusted. In certain embodiments an elevation (e.g., a first elevation of a

normalized count profile) that significantly differs from a second elevation is categorized as a copy number variation (e.g., a copy number variation, e.g., a maternal copy number variation) and is adjusted. In some embodiments, an elevation (e.g., a first elevation) is within an expected elevation range for a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation and the elevation is adjusted. In certain embodiments, one or more elevations (e.g., elevations in a profile) are not adjusted. In some embodiments, an elevation (e.g., a first elevation) is outside an expected elevation range for a copy number variation and the elevation is not adjusted. Often, an elevation within an expected elevation range for the absence of a copy number variation is not adjusted. Any suitable number of adjustments can be made to one or more elevations in a profile. In some embodiments, one or more elevations are adjusted. In certain embodiments 2 or more, 3 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more and sometimes 10 or more elevations are adjusted.

In some embodiments, a value of a first elevation is adjusted according to a value of a second elevation. In certain embodiments a first elevation, identified as representative of a copy number variation, is adjusted to the value of a second elevation, where the second elevation is often associated with no copy number variation. In certain embodiments, a value of a first elevation, identified as representative of a copy number variation, is adjusted so the value of the first elevation is about equal to a value of a second elevation.

An adjustment can comprise a suitable mathematical operation. In certain embodiments an adjustment comprises one or more mathematical operations. In certain embodiments an elevation is adjusted by normalizing, filtering, averaging, multiplying, dividing, adding or subtracting or combination thereof. In certain embodiments an elevation is adjusted by a predetermined value or a constant. In certain embodiments an elevation is adjusted by modifying the value of the elevation to the value of another elevation. For example, a first elevation may be adjusted by modifying its value to the value of a second elevation. A value in such cases may be a processed value (e.g., mean, normalized value and the like).

In certain embodiments an elevation is categorized as a copy number variation (e.g., a maternal copy number variation) and is adjusted according to a predetermined value referred to herein as a predetermined adjustment value (PAV). Often a PAV is determined for a specific copy number variation. Often a PAV determined for a specific copy number variation (e.g., homozygous duplication, homozygous deletion, heterozygous duplication, heterozygous deletion) is used to

adjust an elevation categorized as a specific copy number variation (e.g., homozygous duplication, homozygous deletion, heterozygous duplication, heterozygous deletion). In certain embodiments, an elevation is categorized as a copy number variation and is then adjusted according to a PAV specific to the type of copy number variation categorized. In certain embodiments an elevation (e.g., a first elevation) is categorized as a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation and is adjusted by adding or subtracting a PAV from the elevation. Often an elevation (e.g., a first elevation) is categorized as a maternal copy number variation and is adjusted by adding a PAV to the elevation. For example, an elevation categorized as a duplication (e.g., a maternal, fetal or maternal and fetal homozygous duplication) can be adjusted by adding a PAV determined for a specific duplication (e.g., a homozygous duplication) thereby providing an adjusted elevation. Often a PAV determined for a copy number duplication is a negative value. In some embodiments providing an adjustment to an elevation representative of a duplication by utilizing a PAV determined for a duplication results in a reduction in the value of the elevation. In some embodiments, an elevation (e.g., a first elevation) that significantly differs from a second elevation is categorized as a copy number deletion (e.g., a homozygous deletion, heterozygous deletion, homozygous duplication, homozygous duplication) and the first elevation is adjusted by adding a PAV determined for a copy number deletion. Often a PAV determined for a copy number deletion is a positive value. In some embodiments providing an adjustment to an elevation representative of a deletion by utilizing a PAV determined for a deletion results in an increase in the value of the elevation.

A PAV can be any suitable value. Often a PAV is determined according to and is specific for a copy number variation (e.g., a categorized copy number variation). In certain embodiments a PAV is determined according to an expected elevation for a copy number variation (e.g., a categorized copy number variation) and/or a PAV factor. A PAV sometimes is determined by multiplying an expected elevation by a PAV factor. For example, a PAV for a copy number variation can be determined by multiplying an expected elevation determined for a copy number variation (e.g., a heterozygous deletion) by a PAV factor determined for the same copy number variation (e.g., a heterozygous deletion). For example, PAV can be determined by the formula below:

PAV_k = (Expected Elevation)$_k$ **x** (PAV factor)$_k$

for the copy number variation k (e.g., k = a heterozygous deletion)

A PAV factor can be any suitable value. In certain embodiments a PAV factor for a homozygous duplication is between about -0.6 and about -0.4. In certain embodiments a PAV factor for a homozygous duplication is about -0.60, -0.59, -0.58, -0.57, -0.56, -0.55, -0.54, -0.53, -0.52, -0.51, -0.50, -0.49, -0.48, -0.47, -0.46, -0.45, -0.44, -0.43, -0.42, -0.41 and -0.40. Often a PAV factor for a homozygous duplication is about -0.5.

For example, for an NRV of about 1 and an expected elevation of a homozygous duplication equal to about 2, the PAV for the homozygous duplication is determined as about -1 according to the formula above. In this case, a first elevation categorized as a homozygous duplication is adjusted by adding about -1 to the value of the first elevation, for example.

In certain embodiments a PAV factor for a heterozygous duplication is between about -0.4 and about -0.2. In certain embodiments a PAV factor for a heterozygous duplication is about -0.40, -0.39, -0.38, -0.37, -0.36, -0.35, -0.34, -0.33, -0.32, -0.31, -0.30, -0.29, -0.28, -0.27, -0.26, -0.25, -0.24, -0.23, -0.22, -0.21 and -0.20. Often a PAV factor for a heterozygous duplication is about -0.33.

For example, for an NRV of about 1 and an expected elevation of a heterozygous duplication equal to about 1.5, the PAV for the homozygous duplication is determined as about -0.495 according to the formula above. In this case, a first elevation categorized as a heterozygous duplication is adjusted by adding about -0.495 to the value of the first elevation, for example.

In certain embodiments a PAV factor for a heterozygous deletion is between about 0.4 and about 0.2. In certain embodiments a PAV factor for a heterozygous deletion is about 0.40, 0.39, 0.38, 0.37, 0.36, 0.35, 0.34, 0.33, 0.32, 0.31, 0.30, 0.29, 0.28, 0.27, 0.26, 0.25, 0.24, 0.23, 0.22, 0.21 and 0.20. Often a PAV factor for a heterozygous deletion is about 0.33.

For example, for an NRV of about 1 and an expected elevation of a heterozygous deletion equal to about 0.5, the PAV for the heterozygous deletion is determined as about 0.495 according to the formula above. In this case, a first elevation categorized as a heterozygous deletion is adjusted by adding about 0.495 to the value of the first elevation, for example.

In certain embodiments a PAV factor for a homozygous deletion is between about 0.6 and about 0.4. In certain embodiments a PAV factor for a homozygous deletion is about 0.60, 0.59, 0.58,

0.57, 0.56, 0.55, 0.54, 0.53, 0.52, 0.51, 0.50, 0.49, 0.48, 0.47, 0.46, 0.45, 0.44, 0.43, 0.42, 0.41 and 0.40. Often a PAV factor for a homozygous deletion is about 0.5.

For example, for an NRV of about 1 and an expected elevation of a homozygous deletion equal to about 0, the PAV for the homozygous deletion is determined as about 1 according to the formula above. In this case, a first elevation categorized as a homozygous deletion is adjusted by adding about 1 to the value of the first elevation, for example.

In certain embodiments, a PAV is about equal to or equal to an expected elevation for a copy number variation (e.g., the expected elevation of a copy number variation).

In some embodiments, counts of an elevation are normalized prior to making an adjustment. In certain embodiments, counts of some or all elevations in a profile are normalized prior to making an adjustment. For example, counts of an elevation can be normalized according to counts of a reference elevation or an NRV. In certain embodiments, counts of an elevation (e.g., a second elevation) are normalized according to counts of a reference elevation or an NRV and the counts of all other elevations (e.g., a first elevation) in a profile are normalized relative to the counts of the same reference elevation or NRV prior to making an adjustment.

In some embodiments, an elevation of a profile results from one or more adjustments. In certain embodiments, an elevation of a profile is determined after one or more elevations in the profile are adjusted. In some embodiments, an elevation of a profile is re-calculated after one or more adjustments are made.

In some embodiments, a copy number variation (e.g., a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation) is determined (e.g., determined directly or indirectly) from an adjustment. For example, an elevation in a profile that was adjusted (e.g., an adjusted first elevation) can be identified as a maternal copy number variation. In some embodiments, the magnitude of the adjustment indicates the type of copy number variation (e.g., heterozygous deletion, homozygous duplication, and the like). In certain embodiments, an adjusted elevation in a profile can be identified as representative of a copy number variation according to the value of a PAV for the copy number variation. For example, for a given profile, PAV is about -1 for a homozygous duplication, about -0.5 for a heterozygous duplication, about 0.5 for a heterozygous deletion and about 1 for a homozygous

deletion. In the preceding example, an elevation adjusted by about -1 can be identified as a homozygous duplication, for example. In some embodiments, one or more copy number variations can be determined from a profile or an elevation comprising one or more adjustments.

In certain embodiments, adjusted elevations within a profile are compared. In certain embodiments anomalies and errors are identified by comparing adjusted elevations. For example, often one or more adjusted elevations in a profile are compared and a particular elevation may be identified as an anomaly or error. In certain embodiments an anomaly or error is identified within one or more genomic sections making up an elevation. An anomaly or error may be identified within the same elevation (e.g., in a profile) or in one or more elevations that represent genomic sections that are adjacent, contiguous, adjoining or abutting. In certain embodiments one or more adjusted elevations are elevations of genomic sections that are adjacent, contiguous, adjoining or abutting where the one or more adjusted elevations are compared and an anomaly or error is identified. An anomaly or error can be a peak or dip in a profile or elevation where a cause of the peak or dip is known or unknown. In certain embodiments adjusted elevations are compared and an anomaly or error is identified where the anomaly or error is due to a stochastic, systematic, random or user error. In certain embodiments adjusted elevations are compared and an anomaly or error is removed from a profile. In certain embodiments, adjusted elevations are compared and an anomaly or error is adjusted.

In certain embodiments an outcome is determined according to one or more elevations. In some embodiments, a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy) is determined according to one or more adjusted elevations. In certain embodiments, a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy) is determined according to a profile comprising 1 to about 10,000 adjusted elevations. Often a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy) is determined according to a profile comprising about 1 to about a 1000, 1 to about 900, 1 to about 800, 1 to about 700, 1 to about 600, 1 to about 500, 1 to about 400, 1 to about 300, 1 to about 200, 1 to about 100, 1 to about 50, 1 to about 25, 1 to about 20, 1 to about 15, 1 to about 10, or 1 to about 5 adjustments. In certain embodiments a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy) is determined according to a profile comprising about 1 adjustment (e.g., one adjusted elevation). In certain embodiments an outcome is determined according to one or more profiles (e.g., a profile of a chromosome or segment thereof) comprising one or more, 2 or more, 3 or more, 5 or more, 6 or

more, 7 or more, 8 or more, 9 or more or sometimes 10 or more adjustments. In certain embodiments, a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy) is determined according to a profile where some elevations in a profile are not adjusted. In certain embodiments, a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy) is determined according to a profile where adjustments are not made.

In some embodiments, an adjustment of an elevation (e.g., a first elevation) in a profile reduces a false determination or false outcome. In some embodiments, an adjustment of an elevation (e.g., a first elevation) in a profile reduces the frequency and/or probability (e.g., statistical probability, likelihood) of a false determination or false outcome. A false determination or outcome can be a determination or outcome that is not accurate. A false determination or outcome can be a determination or outcome that is not reflective of the actual or true genetic make-up or the actual or true genetic disposition (e.g., the presence or absence of a genetic variation) of a subject (e.g., a pregnant female, a fetus and/or a combination thereof). In certain embodiments a false determination or outcome is a false negative determination. In some embodiments a negative determination or negative outcome is the absence of a genetic variation (e.g., aneuploidy, copy number variation). In certain embodiments a false determination or false outcome is a false positive determination or false positive outcome. In some embodiments a positive determination or positive outcome is the presence of a genetic variation (e.g., aneuploidy, copy number variation). In some embodiments, a determination or outcome is utilized in a diagnosis. In some embodiments, a determination or outcome is for a fetus.

Determining a Chromosome Representation

A fetal fraction determination can be based in part on an expected chromosome representation and/or a measured chromosome representation.

Expected Chromosome Representation (ECR)

In some embodiments, fetal fraction is determined in part by generating an expected chromosome representation (ECR, e.g., an expected euploid chromosome representation) for a chromosome or segment thereof. An ECR is often for a euploid representation of a chromosome, or segment thereof. An ECR can be determined for an autosome, a sex chromosome or segment thereof. In

certain embodiments an ECR is determined for an affected autosome (e.g., in the case of a trisomy, e.g., chromosome 13 is the affected autosome in the case of a trisomy 13, chromosome 18 is the affected autosome in the case of trisomy 18, or chromosome 21 is the affected autosome in the case of a trisomy 21). An ECR for chromosome *n*, or segment thereof, can be referred to as an "expected *n* chromosome representation". For example, an ECR for chromosome X can be referred to as an "expected X chromosome representation". In certain embodiments an ECR is determined according to the number of genomic sections in a normalized count profile. In certain embodiments the ECR for chromosome *n* is the ratio between the total number of genomic sections for chromosome *n*, or a segment thereof, and the total number of genomic sections in a profile (e.g., a profile of all autosomal chromosomes, a profile of most all autosomal chromosomes, a profile of a genome or segment of a genome). A profile can comprise one or more of any chromosome or segments thereof, a subset of chromosomes (e.g., including any chromosome or segment thereof), all autosomes, a subset of autosomes. A profile can comprise some or all chromosomes sequenced in a sample, from a single sequencing run and/or from a single flow cell. In certain embodiments an ECR is the ratio between the total area under an expected elevation representative of the genomic sections for chromosome *n*, or a segment thereof, and the total area under the expected elevation for all genomic sections of an entire profile (e.g., a profile of all autosomal chromosomes, a profile of most all autosomal chromosomes, a profile of a genome or segment of a genome). In certain embodiments, an ECR is determined according to an expected median or mean value of an expected elevation and/or profile. An ECR can be determined, in some cases, by equation Z described in Example 9. Often, an ECR is determined for chromosome *n*, or a segment thereof, where chromosome *n* is an aneuploid chromosome (e.g., a trisomy). In certain embodiments, an ECR is determined for chromosome X and/or chromosome Y for a pregnant female bearing a male fetus. In certain embodiments, an ECR is determined for chromosome X and/or chromosome Y for a pregnant female bearing a male fetus comprising a sex aneuploidy (e.g., Turner's Syndrome, Klinefelter syndrome, Double Y syndrome, Trisomy X syndrome, Four X syndrome). In some embodiments, an expected euploid chromosome representation for ChrX is the median or mean ChrX representation obtained from a female pregnancy or from a set of female pregnancies. In some embodiments, an expected chromosome representation for ChrX in a male pregnancy is the median or mean ChrX representation obtained from a female pregnancy or from a set of female pregnancies. In some embodiments, the MCR of ChrX in a male pregnancy differs from the ECR of ChrX in a female pregnancy and the deviation is proportional to fetal fraction. In certain embodiments fetal fraction is estimated from the MCR of ChrX in a male pregnancy.

Measured Chromosome Representation (MCR)

In some embodiments, a fetal fraction is determined, in part, by generating a measured (i.e., experimentally measured) chromosome representation (MCR). Often an MCR is an experimentally derived value. An MCR can be referred to as an experimental chromosome representation. An MCR for chromosome *n* can be referred to as an "experimental *n* chromosome representation". For example, an MCR for chromosome X can be referred to as an "experimental X chromosome representation". In certain embodiments an MCR is determined according to counts mapped to genomic sections of a chromosome or a segment thereof. In certain embodiments an MCR is determined from normalized counts. In certain embodiments an MCR is determined from raw counts. Often an MCR is determined from counts normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, the like or a combination thereof. In certain embodiments an MCR is determined according to counts mapped to genomic sections of a sex chromosome (e.g., an X or Y chromosome) or a chromosome representing an aneuploidy (e.g., an affected autosome, a trisomy). In certain embodiments an MCR is determined according to a measured elevation of a chromosome or segment thereof. In certain embodiments, an MCR for a chromosome can be determined according to a median, average or mean value of one or more elevations in a profile. In certain embodiments, a fetal fraction is determined, in part, by generating an MCR for a chromosome, or segment thereof, representing an aneuploidy. In certain embodiments, a fetal fraction is determined, in part, by generating an MCR for an X and/or a Y chromosome, or segment thereof, for a pregnant female bearing a male fetus. In certain embodiments an MCR is determined according to counts mapped to genomic sections of a sex chromosome (e.g., an X or Y chromosome) for a pregnant female bearing a male fetus comprising a sex aneuploidy (e.g., Turner's Syndrome, Klinefelter syndrome, Double Y syndrome, Trisomy X syndrome, Four X syndrome). In some embodiments, an MCR for a sex chromosome is not determined for a pregnant female bearing a euploid female fetus. In certain embodiments an MCR for chromosome *n* is the ratio between the total number of counts (e.g., total normalized counts, mean, average or median of total counts) mapped to genomic sections of chromosome *n,* or a segment thereof, and the total number of counts (e.g., total normalized counts, mean, average or median of total counts) mapped to genomic sections of some or all chromosomes represented in a profile (e.g., a profile of a genome or segment thereof) where chromosome *n* can be any chromosome. In certain embodiments an MCR for chromosome *n* is the ratio between the total number of counts (e.g., total normalized counts, mean, average or median of total counts) mapped to genomic sections of chromosome *n,* or a segment thereof, and the total number of counts (e.g.,

total normalized counts, mean, average or median of total counts) mapped to genomic sections of some or all autosomes represented in a profile (e.g., a profile of a genome or segment thereof) where chromosome *n* can be any chromosome. In certain embodiments an MCR for chromosome *n* is the ratio between the total area under an elevation representative of chromosome *n*, or a segment thereof, and the total area under an elevation of an entire profile (e.g., a profile of all autosomal chromosomes, a profile of most all autosomal chromosomes, a profile of a genome or segment of a genome).

In certain embodiments a median chromosome representation is a median MCR for chromosome *n*, or a segment thereof, derived from two or more samples (e.g., two or more subjects). For example, a median chromosome representation can be a median of multiple MCRs for a chromosome or segment thereof.

Additional methods for Determining an MCR for a Sex Chromosome

In some embodiments the measured experimental Y chromosomal representation is determined for a test sample obtained from a pregnant female bearing a male fetus. In some embodiments the measured experimental Y chromosomal representation is determined for a test sample obtained from a pregnant female bearing a female fetus. The measured experimental Y chromosomal representation for a pregnant female bearing a female fetus is often zero or close to zero. In some embodiments the measured experimental Y chromosomal representation for a pregnant female bearing a female fetus is not zero due to noise. In certain embodiments a measured Y chromosome representation from a pregnant female bearing a female fetus represents noise and is used to normalize and/or correct (e.g., subtract out background) when determining fetal fraction and/or measured chromosome representations of other chromosomes. In some embodiments a measured Y chromosome representation from a pregnant female bearing a fetus is an average, mean or median measured Y chromosome representation. In some embodiments a measured Y chromosome representation from a pregnant female bearing a fetus is an average, mean or median measured Y chromosome representation where the average, mean or median is determined from multiple bins of a Y chromosome and/or from multiple samples (e.g., samples from multiple subjects). In certain embodiments an average, mean or median measured Y chromosome representation is used to determine a measure of noise and/or background.

In some embodiments the presence or absence of a Y chromosome, ploidy, counts or an MCR of a Y chromosome in a fetus is determined from sequence reads that map to substantially unique portions or segments of a Y chromosome. In certain embodiments the presence or absence of an X chromosome, ploidy, counts or an MCR of an X chromosome is determined according to sequence reads that map to substantially unique portions or segments of an X chromosome. The X and Y chromosomes comprise nucleic acid segments that are very similar and/or homologous (e.g., comprise high homology). The origin (e.g., from ChrX or ChrY) of sequence reads mapping to these high homology regions is sometimes difficult, if not impossible, to determine. In certain embodiments reads that map to both chromosome X and chromosome Y are not considered substantially unique to either the X or Y chromosome. In some embodiments bins comprising greater than about 40%, 30%, 25%, 20%, 15% or greater than about 10% of reads that map to both the X and Y chromosome are not substantially unique to either chromosome X or Y. In some embodiments bins that are not substantially unique to either chromosome X or Y are filtered and/or removed from any determination of fetal aneuploidy, fetal gender and/or fetal fraction. In some embodiments only sequence reads that map to portions or segments that are substantially unique to chromosome X and/or chromosome Y are utilized to determine fetal gender, fetal fraction and/or the presence of a fetal sex chromosome aneuploidy.

In some embodiments reads that map to the X chromosome and not to the Y chromosome are substantially unique to chromosome X. In some embodiments reads that map to the Y chromosome and not to the X chromosome are substantially unique to chromosome Y. In some embodiments sequence reads that map within the first 28 Mb from the 5' end of the Y chromosome are substantially unique to chromosome Y. In some embodiments a selected set of bins in the Y chromosome are used when determining the presence or absence of a Y chromosome, ploidy, counts, an ECR or an MCR of a Y chromosome. In certain embodiments a selected set of 20-30 bins of the Y chromosome are used. In some embodiments 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 selected bins are used when determining the presence or absence of a Y chromosome, ploidy, counts, an ECR or an MCR of a Y chromosome. In some embodiments bins comprising greater than about 60%, 70%, 75%, 80%, 85%, 90%, 95%, or about 100% of mapped reads that are unique to the X or Y chromosome are bins that are substantially unique to the X or Y chromosome, respectively.

In some embodiments a measured chromosome representation for chromosome X and/or chromosome Y is determined according to counts that map to substantially unique segments or

portions of chromosome X and/or chromosome Y. In some embodiments fetal gender, fetal fraction and/or a fetal aneuploidy is determined according, in part, to an MCR for a sex chromosome where the measured counts are determined from counts that map to substantially unique segments or portions of chromosome X and/or chromosome Y.

Determining Fetal Fraction from a Chromosomal Representation

In some embodiments, a difference in an elevation representing a chromosomes in a fetus comprising an aneuploidy (e.g., trisomy, a sex chromosome aneuploidy) and/or chromosome X and Y in the case of male fetuses can be used to determine fetal fraction. In certain embodiments raw counts are used to determine fetal fraction. In some embodiments normalized counts are used to determine fetal fraction. In some embodiments prior to determining fetal fraction by a method described herein, counts are normalized by a suitable method, non-limiting examples of which include bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and/or combinations thereof. In some embodiments, a fetal fraction is determined from a chromosomal representation (e.g., a chromosome representation of an aneuploid chromosome, X chromosome or Y chromosome). In certain embodiments, a fetal fraction is determined when fetal ploidy differs from maternal ploidy. In certain embodiments, a fetal fraction is not determined when fetal ploidy is zero (e.g., the absence of a Y chromosome in a female fetus). In some embodiments, a fetal fraction is determined from a chromosome representation of an aneuploid chromosome when a pregnant female bears a fetus comprising a chromosome aneuploidy (e.g., a trisomy). In some embodiments, a fetal fraction is determined from a chromosome representation of an aneuploid chromosome when a pregnant female bears a male fetus comprising a chromosome aneuploidy (e.g., a trisomy). In certain embodiments, an MCR for an aneuploidy (e.g., a fetal trisomy) of chromosome *n* differs from an ECR for chromosome *n* according to the fraction of fetal nucleic acids present in a sample of cell-free nucleic acid obtained from a pregnant female. Thus, in some embodiments, a fetal fraction can be determined according to a ratio of an MCR for an aneuploidy and an ECR for the same affected chromosome. In certain embodiments, for a pregnant female bearing a fetus comprising an aneuploidy of an autosomal chromosome (e.g., a trisomy), a fetal fraction can be determined according to equation AA and/or according to equation AB in Example 9. In some embodiments, for a pregnant female bearing a fetus comprising a sex aneuploidy (e.g., TripleX syndrome), a fetal fraction can be determined according to equation AA and/or according to equation AB in Example 9 using an X chromosome representation. In equations AA and AB, the

symbol C_{in} represents an MCR and the symbol C^0_n represents an ECR. For example, in the case of a trisomy, a fetal fraction can be determined by subtracting 1 from the ratio of an MCR and ECR and multiplying the result by 2. In certain embodiments the result of equation AB is multiplied by 100 to determine a percent fetal fraction (e.g., the percentage of fetal DNA in a mixture of maternal and fetal nucleic acid).

In some embodiments, a fetal fraction can be determined according to an ECR for chromosome X and an MCR for chromosome X for a pregnant female bearing a male fetus (XY), a fetus with Jacobs syndrome (i.e., XYY) or a fetus with Turner syndrome (i.e., X). In some embodiments, a fetal fraction can be determined according to the ratio of an ECR for chromosome X and an MCR for chromosome X. In some embodiments, for a pregnant female bearing a male fetus (XY), a fetus with Jacobs (i.e., XYY) or a fetus with Turner syndrome (i.e., X), a fetal fraction can be determined according to equation AC in Example 9. In equations AC the symbol C_{in} represents an MCR and the symbol C^0_n represents an ECR. In certain embodiments, C^0_n represents the median chromosomal representation of ChrX in a female pregnancy. For example, in the case of a male fetus, a fetal fraction can be determined by subtracting the ratio of an MCR and ECR from 1 and multiplying the result by 2. In certain embodiments the result of equation AC is multiplied by 100 to determine a percent fetal fraction (e.g., the percentage of fetal DNA in a mixture of maternal and fetal nucleic acid).

In some embodiments, a fetal fraction can be determined according to an ECR for chromosome Y and an MCR for chromosome Y for a pregnant female bearing a male fetus (XY) or a fetus with Klinefelter's syndrome (i.e., XXY). In some embodiments, a fetal fraction can be determined according to the ratio of an ECR for chromosome Y and an MCR for chromosome Y. In some embodiments, for a pregnant female bearing a male fetus (XY) or a fetus with Klinefelter's syndrome (i.e., XXY), a fetal fraction can be determined according to equation AG in Example 9. The term (C_y) in equation AG represents a median chromosomal representation of chromosome Y (e.g., an ECR for ChrY) and C_y represents an MCR for chromosome Y. In certain embodiments, C_y represents the median chromosomal representation of ChrX in a female pregnancy. For example, in the case of a male fetus, a fetal fraction can be determined by subtracting the ratio of an MCR and ECR from 1 and multiplying the result by 2. In certain embodiments the result of equation AC is multiplied by 100 to determine a percent fetal fraction (e.g., the percentage of fetal DNA in a mixture of maternal and fetal nucleic acid). In certain embodiments a modified version of equation AG is used to determine fetal fraction from ChrY in Jacobs syndrome (XYY). In some

embodiments the fetal fraction resulting from equation AG is divided by 2 to arrive at a fetal fraction for Jacobs syndrome. In some embodiments the C_y value in equation AG is divided by 2 to arrive at a fetal fraction for Jacobs syndrome.

Determining relationships

In some embodiments, a fetal fraction is, in part, generated according to a relationship. A relationship can be a mathematical relationship. In some embodiments, a relationship is a geometric and/or graphical relationship. In some embodiments, a relationship is plotted. In some embodiments a relationship is a linear relationship. In certain embodiments a linear relationship is an inverse relationship and sometimes a linear relationship is a direct relationship. In some embodiments, a relationship is a bivariate relationship. A relationship can be expressed by a mathematical equation. Often a relationship defines one or more constants.

In some embodiments a relationship is generated for a fetal fraction determination and an MCR (e.g., an MCR of a chromosome, an MCR of an X or a Y chromosome, an MCR of an affected autosome). In certain embodiments a fetal fraction determination from which a relationship is generated is a fetal fraction obtained from a chromosome representation of an aneuploid chromosome. For example, a fetal fraction obtained from a chromosome representation of an aneuploid chromosome can be determined from equation AA or AB or by a method described herein. In certain embodiments a fetal fraction determination from which a relationship is generated is a fetal fraction obtained from a chromosome representation of a trisomy 18, a trisomy 21 or a trisomy 13. Often a fetal fraction determination from which a relationship is generated is a fetal fraction obtained from a chromosome representation of a trisomy 18, a trisomy 21 or a trisomy 13 obtained from a pregnant female bearing a male fetus. Fetal fraction can be determined for such aneuploidy pregnancies by a convenient method known in the art or described herein. For example, a fetal fraction can be determined by an FQA. In certain embodiments, a fetal fraction can be determined according to an elevation and/or according to a copy number variation as described herein.

In some embodiments an MCR from which a relationship is generated is an MCR for an X or Y chromosome or segment thereof. In some embodiments an MCR from which a relationship is generated is an MCR for an X or Y chromosome or segment thereof obtained from equation AC or by a method described herein. In some embodiments an MCR from which a relationship is

generated is an MCR for an X or Y chromosome obtained from a pregnant female bearing a male fetus. Often an MCR from which a relationship is generated is an MCR for an X or Y chromosome obtained from a pregnant female bearing a male fetus comprising a chromosome aneuploidy (e.g., a trisomy 13, 18, or 21). In certain embodiments an MCR from which a relationship is generated is an MCR for an X or Y chromosome obtained from a pregnant female bearing a male fetus comprising a sex chromosome aneuploidy (e.g., a sex chromosome aneuploidy as in Table 1A-1G). In some embodiments a relationship is generated for (i) a fetal fraction determination obtained from an aneuploid chromosome and (ii) an MCR of an X or a Y chromosome where both the fetal fraction determination and MCR are obtained from a pregnant female bearing a male fetus comprising an aneuploid chromosome.

In some embodiments an MCR from which a relationship is generated is an MCR for an autosome or segment thereof (e.g., an affected autosome). In certain embodiments an MCR from which a relationship is generated is an MCR for an aneuploid chromosome (e.g., where the fetal representation of the chromosome is an aneuploid).

Often a relationship is generated from a fetal fraction determination and an MCR determination obtained from multiple subjects. In certain embodiments, a relationship is generated from greater than about 10, greater than about 100, greater than about 500 or greater than about 1000 subjects. In certain embodiments, a relationship is generated from about 500 to about 50,000, about 500 to about 25,000, about 500 to about 10,000, about 500 to about 5000, or about 500 to about 2500 subjects.

In some embodiments a relationship generated for a fetal fraction determination from a pregnant female bearing a male fetus comprising an aneuploidy (i.e., a male aneuploid pregnancy) and an MCR of an X chromosome from a pregnant female bearing a male fetus (i.e. a male pregnancy) is expressed mathematically. Often a relationship between a fetal fraction determination from a male aneuploid pregnancy and an MCR determination obtained from a male pregnancy is a linear relationship. Sometime an MCR determination obtained from a male pregnancy is obtained from a male aneuploid pregnancy. In certain embodiments a linear relationship generated for a fetal fraction determination from a male aneuploid pregnancy and an MCR of an X chromosome from a male pregnancy is represented by Equation AD2 below:

$$F_i = k - r(MCR_{ix}) \qquad \text{(AD2)}$$

where *k* (e.g., y intercept) and *r* (e.g., slope) are constants defining the relationship for F_i (the fetal fraction of sample *i*) and MCR_{ix} (an MCR determined for chromosome X, or segment thereof, in sample *i*). In some embodiments, the relationship represented by Equation AD2 is generated for a fetal fraction determination obtained from an aneuploid chromosome and an MCR of an X chromosome where both the fetal fraction determination and MCR are obtained from a pregnant female bearing a male fetus comprising an aneuploid chromosome. In some embodiments, constants *k* and/or *r* are determined empirically from a fetal fraction determination and an MCR obtained from multiple subjects. Constants *k* or *r* can be a suitable number that defines the relationship between a fetal fraction determination and an MCR obtained from multiple subjects. Constants *k* or *r* can vary according to differences in experimental parameters and difference in methods of obtaining sequence reads (e.g., sequencing platform, sequencing recipe (e.g., how many bases are called), library/clustering, chemistry, normalization methods and/or normalization parameters, bin filtering, bin selection, the like or combinations thereof). In some embodiments *k* in equation AD2 equals about 150 to about 210, about 155 to about 205, about 160 to about 200, about 165 to about 195, about 170 to about 190, about 175 to about 185. In certain embodiments *k* is equal to about 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, or about 185. In some embodiments *k* is equal to about 178.1, 178.2, 178.3, 178.4, 178.5, 178.6, 178.7, 178.8, 178.9, 179, 179.1, 179.2, 179.3, 179.4, 179.5, 179.6, 179.7, 179.8, or about 179.9. In certain embodiments *k* is equal to about 179.1. In some embodiments *r* in equation AD2 equals about 2500 to about 3500, about 2600 to about 3400, about 2700 to about 3300, about 2800 to about 3200, about 2900 to about 3100, or about 3000 to about 3100. In certain embodiments *r* in equation AD2 equals about 3005 to about 3085, about 3010 to about 3080, about 3015 to about 3075, about 3020 to about 3070, about 3025 to about 3065, about 3030 to about 3060, about 3035 to about 3055, or about 3040 to about 3050. In certain embodiments *r* in equation AD2 equals about 3040, 3041, 3042, 3043, 3044, 3045, 3046, 3047, 3048, 3049 or about 3050. In certain embodiments *r* in equation AD2 equals about 3045.1, 3045.2, 3045.3, 3045.4, 3045.5, 3045.6, 3045.7, 3045.8, 3045.9, 3046.0, 3046.1, 3046.2, 3046.3, 3046.4, or about 3046.5. In certain embodiments *r* in equation AD2 equals about 3045.74, 3045.75, 3045.76, 3045.77, 3045.78, 3045.79, 3045.80, 3045.81, 3045.82, 3045.83, 3045.84, 3045.85, 3045.86, 3045.87, 3045.88, 3045.89, or about 3045.90. In certain embodiments *r* in equation AD2 equals about 3045.82. In certain embodiments a relationship generated for a fetal fraction determination and an MCR of an X chromosome is represented by Equation AD.

In certain embodiments a fetal fraction is determined from a pregnant female bearing a male fetus from the relationship represented by equation AD2 or AD. In some embodiments, a fetal fraction is determined from a pregnant female bearing a male euploid fetus from the relationship represented by equation AD2 or AD. In certain embodiments an MCR for an X chromosome is provided for a pregnant female subject bearing a male fetus comprising a genetic variation (e.g., an aneuploidy) and a fetal fraction is determined according to Equation AD2 or AD. In certain embodiments an MCR for an X chromosome is provided for a pregnant female subject bearing a euploid male fetus and a fetal fraction is determined according to Equation AD2 or AD.

In some embodiments a relationship generated for a fetal fraction determination from a male aneuploid pregnancy and an MCR of a Y chromosome, or segment thereof, from a male pregnancy is expressed mathematically. In certain embodiments a linear relationship generated for a fetal fraction determination and an MCR of a Y chromosome is represented by Equation AE2:

$$F_i = k + r(MCR_{iy}) \qquad \text{(AE2)}$$

where k (e.g., y intercept) and r (e.g., slope) are constants defining the relationship for F_i (the fetal fraction of sample i) and MCR_{iy} (an MCR determined for chromosome Y, or segment thereof, in sample i). In some embodiments, the relationship represented by Equation AE2 is generated for (i) a fetal fraction determination obtained from an aneuploid chromosome and (ii) an MCR of a Y chromosome where both the fetal fraction determination and MCR are obtained from a pregnant female bearing a male fetus comprising an aneuploid chromosome. In some embodiments, constants k and/or r are determined empirically from a fetal fraction determination and an MCR obtained from multiple subjects. Constants k or r can be a suitable number that defines the relationship between a fetal fraction determination and an MCR obtained from multiple subjects. Constants k or r can vary according to differences in experimental parameters and difference in methods of obtaining sequence reads (e.g., sequencing platform, sequencing recipe (e.g., how many bases are called), library/clustering, chemistry, normalization methods and/or normalization parameters, bin filtering, bin selection, the like or combinations thereof). In some embodiments k in equation AE2 equals about .43 to about .63, about .44 to about .62, about .45 to about .61, about .46 to about .60, about .47 to about .59, about .48 to about .58, about .49 to about .57, about .50 to about .56, about .51 to about .55 or about .52 to about .54. In certain embodiments k is equal to about .526 to about .546, about .527 to about .545, about .528 to about .544, about .529 to about .543, about .530 to about .542, about .531 to about .541, about .532 to about .540, about .533 to

about .539, about .534 to about .538, about .535 to about .537. In some embodiments *k* is equal to about .5360, .5361, .5362, .5363, .5364, .5365, .5366, .5367, .5368, .5369, .5370, .5371, .5372, .5373, .5374, .5375, or about .5376,. In certain embodiments *k* is equal to about .5368. In some embodiments *r* in equation AE2 equals about 1348 to about 1350, about 1347 to about 1349, about 1346 to about 1348, about 1345 to about 1347, about 1344 to about 1346 or about 1343 to about 1345,. In certain embodiments *r* in equation AE2 equals about 1343.1 to about 1344.9, about 1343.2 to about 1344.8, about 1343.3 to about 1344.7, about 1343.4 to about 1344.6, about 1343.5 to about 1344.5, about 1343.6 to about 1344.4, about 1343.7 to about 1344.3, about 1343.8 to about 1344.2 or about 1343.9 to about 1344.1. In certain embodiments *r* in equation AE2 equals about 1343.96 to about 1344.06, about 1343.97 to about 1344.05, about 1343.98 to about 1344.04, about 1343.99 to about 1344.03, or about 1344.00 to about 1344.02. In certain embodiments *r* in equation AE2 equals about 1344.010, 1344.011, 1344.012, 1344.013, 1344.014, 1344.015, 1344.016, 1344.017, 1344.018, or about 1344.019. In certain embodiments *r* in equation AE2 equals about 1344.0158, 1344.0159, 1344.0160, 1344.0161, 1344.0162, 1344.0163, 1344.0164, 1344.0165, 1344.0166, or about 1344.0167. In certain embodiments *r* in equation AE2 equals about 1344.0162. In certain embodiments a relationship generated for a fetal fraction determination and an MCR of a Y chromosome is represented by Equation AE.

In certain embodiments a fetal fraction is determined from a pregnant female bearing a male fetus from the relationship represented by equation AE2 or AE. In some embodiments, a fetal fraction is determined from a pregnant female bearing a male euploid fetus from the relationship represented by equation AE2 or AE. In certain embodiments an MCR for a Y chromosome is provided for a pregnant female subject bearing a male fetus comprising a genetic variation (e.g., an aneuploidy) and a fetal fraction is determined according to Equation AE2 or AE. In certain embodiments an MCR for a Y chromosome is provided for a pregnant female subject bearing a euploid male fetus and a fetal fraction is determined according to Equation AE2 or AE.

In certain embodiments a relationship is generated for a fetal fraction determined by a first method and a fetal fraction determined by a second method. In certain embodiments a relationship is generated for a fetal fraction determined by a first method and a fetal fraction determined by a second method where the relationship is generated for multiple fetal fraction determinations. In certain embodiments a relationship is generated for greater than about 10, greater than about 100, greater than about 500 or greater than about 1000 fetal fraction determinations. In certain embodiments, a relationship is generated from about 500 to about 50,000, about 500 to about

25,000, about 500 to about 10,000, about 500 to about 5000, or about 500 to about 2500 fetal fraction determinations.

In some embodiments, a fetal fraction determined by a first method and/or a fetal fraction determined by a second method are determined by an FQA. In some embodiments, a fetal fraction determined by a first method and/or a fetal fraction determined by a second method are determined by a process that does not utilize sequence reads mapped to genomic sections of a reference genome. In some embodiments, a fetal fraction determined by a first method and/or a fetal fraction determined by a second method are determined by a process comprising mass spectrometry. In some embodiments, a fetal fraction determined by a first method and/or a fetal fraction determined by a second method are determined by a process utilizing MPS. In some embodiments, a fetal fraction of nucleic acids in the blood of a pregnant female is determined by a first method and a fetal fraction of nucleic acids in the blood of a different pregnant female is determined by a second method. In certain embodiments the first method and the second method are the same method. In certain embodiments the first method and the second method are different methods. In some embodiments, a fetal fraction determined by a first method is determined at a different time than a fetal fraction determined by a second method. For example, sometimes a fetal fraction determined by a first method is determined before or after a fetal fraction determined by a second method.

In certain embodiments a fetal fraction determined by a first method is determined from a first subject (e.g., from a sample obtained from a pregnant female subject) and a fetal fraction determined by a second method is determined from a second subject (e.g., from a sample obtained from a pregnant female subject) where the first subject and second subject are different subjects. In certain embodiments a fetal fraction determined by a first method is determined from a first set of multiple subjects (e.g., a pregnant female subjects) and a fetal fraction determined by a second method is determined from a second set of subjects (e.g., a pregnant female subjects) where the first set of subjects and second set of subjects are different subjects. In certain embodiments, a first set of multiple subjects (e.g., different than a second set of subjects) is greater than about 10, greater than about 100, greater than about 500 or greater than about 1000 subjects. In certain embodiments, a first set of multiple subjects (e.g., different than a second set of subjects) is from about 500 to about 50,000, about 500 to about 25,000, about 500 to about 10,000, about 500 to about 5000, or about 500 to about 2500 subjects. In some embodiments, a second set of subjects (e.g., different that a first set of multiple subjects) is 1 subject or greater than about 1,

greater than about 10, greater than about 100, greater than about 500 or greater than about 1000 subjects. In certain embodiments, a second set of subjects (e.g., different that a first set of multiple subjects) is from about 1 to about 50,000, about 1 to about 25,000, about 1 to about 10,000, about 1 to about 5000, about 1 to about 2500, 1 to about 1000, or about 1 to about 500 subjects. In certain embodiments a second set of subjects (e.g., different that a first set of multiple subjects) is 1 subject.

In some embodiments, a fetal fraction determined by a first method is determined by a relationship for (i) a fetal fraction determination obtained from an aneuploid chromosome and (ii) an MCR of an X chromosome where both the fetal fraction determination and MCR are obtained from a pregnant female bearing a male fetus comprising an aneuploid chromosome. In certain embodiments a fetal fraction determined by a first method is a fetal fraction determined by Equation AD2 or AD. In some embodiments, a fetal fraction determined by a second method is determined by a relationship for (i) a fetal fraction determination obtained from an aneuploid chromosome and an (i) MCR of a Y chromosome where both the fetal fraction determination and MCR are obtained from a pregnant female bearing a male fetus comprising an aneuploid chromosome. In certain embodiments a fetal fraction determined by a second method is determined by Equation AE2 or AE.

In certain embodiments a relationship is generated for (i) a fetal fraction determined by a first method and (ii) a fetal fraction determined by a second method where the fetal fraction in (i) and (ii) is determined from an MCR for an X and a Y chromosome, respectively, obtained from a pregnant female bearing a male euploid fetus. In certain embodiments the fetal fraction in (i) and (ii) is determined from an MCR for an X and a Y chromosome where the MCR was determined for the same sample (e.g., same subject). In some embodiments a relationship

In some embodiments a relationship generated for (i) a fetal fraction determined by a first method and (ii) a fetal fraction determined by a second method is expressed mathematically. In certain embodiments a relationship generated for (i) a fetal fraction determined by a first method and (ii) a fetal fraction determined by a second method is represented by Equation AF2:

$$F_i = k - r(MCR_{ix}) + t(MCR_{iy}) \qquad \text{(AF2)}$$

where *k*, *r* and *t* are constants defining the relationship for F_i (the fetal fraction of sample *i*), MCR_{ix} (an MCR determined for chromosome X in sample *i*) and MCR_{iy} (an MCR determined for chromosome Y in sample *i*). In some embodiments, the relationship in Equation AF2 is for sample *i* obtained from a pregnant female bearing a male fetus. In some embodiments, constants *k, r* and/or *t* are determined empirically. Each constant *k, r* and/or *t* can be a suitable number that defines the relationship between (i) a fetal fraction determined by a first method and (ii) a fetal fraction determined by a second method. Constants *k, r* and/or *t* can vary according to differences in experimental parameters and difference in methods of obtaining sequence reads (e.g., sequencing platform, sequencing recipe (e.g., how many bases are called), library/clustering, chemistry, normalization methods and/or normalization parameters, bin filtering, bin selection, the like or combinations thereof). Constants *k, r* and *t* often are derived for a linear relationship. In some embodiments *k* in equation AF2 equals about 20 to about 29, about 21 to about 28, about 22 to about 27, about 23 to about 26 or about 24 to about 25. In some embodiments *k* in equation AF2 equals about 24.4 to about 25.3, about 24.5 to about 25.2, about 24.6 to about 25.1, about 24.7 to about 25.0, or about 24.8 to about 24.9. In some embodiments *k* in equation AF2 equals about 24.78, 24.79, 24.80, 24.81, 24.82, 24.83, 24.84, 24.85, 24.86, 24.87, 24.88, 24.89, 24.90, 24.91, 24.92, 24.93, 24.94, 24.95, 24.96 or about 24.97. In some embodiments *k* in equation AF2 equals about 24.88. In some embodiments *r* in equation AF2 equals about 411 to about 421, about 412 to about 420, about 413 to about 419, about 414 to about 418 or about 415 to about 417. In some embodiments *r* in equation AF2 equals about 416.0 to about 416.9, about 416.1 to about 416.8, about 416.2 to about 416.7, about 416.3 to about 416.6, about 416.4 to about 416.5. In some embodiments *r* in equation AF2 equals about 416.32, 416.33, 416.34, 416.35, 416.36, 416.37, 416.38, 416.39, 416.40, 416.41, 416.42, 416.43, 416.44, 416.45, 416.46, 416.47, 416.48, 416.49, 416.50, 416.51 or 416.52. In some embodiments *r* in equation AF2 equals about 416.42. In some embodiments *t* in equation AF2 equals about 1164 to about 1174, about 1165 to about 1173, about 1166 to about 1172, about 1167 to about 1171, about 1168 to about 1170. In some embodiments *t* in equation AF2 equals about 1169.0 to about 1169.9, about 1169.1 to about 1169.8, about 1169.2 to about 1169.7, about 1163.0 to about 1169.6 or about 1169.4 to about 1169.5. In some embodiments *t* in equation AF2 equals about 1169.36, about 1169.37, about 1169.38, about 1169.39, about 1169.40, about 1169.41, about 1169.42, about 1169.43, about 1169.44, about 1169.45, about 1169.46, about 1169.47, about 1169.48, about 1169.49, about 1169.50, about 1169.51, about 1169.52, about 1169.53, about 1169.54, about 1169.54 or about 1169.56. In some embodiments *t* in equation AF2 equals about 1169.46. In certain embodiments

a relationship generated for (i) a fetal fraction determined by a first method and (ii) a fetal fraction determined by a second method is represented by Equation AF.

In certain embodiments the fraction of fetal nucleic acid in circulating cell-free nucleic acid obtained from the blood of a pregnant female is determined from the relationship represented by equation AF2 or AF. In certain embodiments a fetal fraction is determined for a pregnant female bearing a male fetus from the relationship represented by equation AF2 or AF. In certain embodiments the male fetus is euploid. In certain embodiments the male fetus comprises and aneuploidy (e.g., a trisomy 21, trisomy 13, trisomy 18). In certain embodiments an MCR for a Y chromosome and an MCR for an X chromosome is provided for a pregnant female subject bearing a male fetus and a fetal fraction is determined according to Equation AF2 or AF. In certain embodiments an MCR for a Y chromosome and an MCR for an X chromosome is provided for a pregnant female subject bearing a euploid male fetus and a fetal fraction is determined according to Equation AF2 or AF.

In certain embodiments a first relationship is generated for a fetal fraction determination according to (i) an MCR for chromosome Y for a pregnant female bearing a male fetus, (ii) an MCR for chromosome Y for a pregnant female bearing a female fetus and (iii) *K* a constant according to Equation AG. The constant *K* is determined by a second relationship relating (a) a fetal fraction determined from an aneuploid chromosome (e.g., a trisomy 13, 18 or 21) and (b) a fetal fraction determined from an MCR for chromosome Y determined from a pregnant female bearing a male fetus comprising an aneuploidy (e.g., a trisomy). In some embodiments, the constant K in Equation AG is the empirical slope of the second relationship. In certain embodiments K equals from about 0.001 to about 0.003. In certain embodiments K equals about 0.00017, 0.00018, 0.00019, 0.00020, 0.00021, 0.00022, 0.00023, 0.00024, 0.00025, or about 0.00026. In certain embodiments K equals about 0.0002179630. In certain embodiments the second relationship is a linear relationship. In some embodiments, a fetal fraction F for a male fetus is determined from an MCR of chromosome Y for a pregnant female bearing a male fetus using formula AG below:

$$F = \frac{C_y - (C_y)}{K} \qquad \text{(AG)}$$

where the term (C_y) represents a median chromosomal representation of chromosome Y for a female fetus, C_y represents an MCR for chromosome Y for a pregnant female bearing a male fetus and *K* is the constant described above for Equation AG. In certain embodiments the term (C_y) is

obtained from multiple measurements. In certain embodiments the term (C_y) represents noise. Often (C_y) is different from zero due to noise.

In some embodiments a fetal fraction is determined from the blood of a pregnant female according to a relationship described herein with an accuracy of equal to 90% or greater than 90% and/or a precision equal to 90% or greater than 90%. In some embodiments a fetal fraction is determined from the blood of a pregnant female according to a relationship described by equation AD, AD2, AE, AE2, AF, AF2 or AG with an accuracy of equal to 90% or greater than 90% and/or a precision equal to 90% or greater than 90%. In some embodiments a fetal fraction is provided by a fetal fraction module with an accuracy of equal to 90% or greater than 90% and/or a precision equal to 90% or greater than 90%. In some embodiments a fetal fraction is determined from the blood of a pregnant female according to a relationship described herein with an accuracy of about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or about 100% and/or a precision of about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or about 100%.

Determining fetal fraction from a fixed relationship

In some embodiments the fraction of the fetal nucleic acid in the blood of the pregnant female bearing a male fetus can be determined according to an experimental Y chromosome representation and a relationship (e.g., a linear relationship) that is fitted. An embodiment of this method is illustrated in Example 10. In certain embodiments the relationship is linear. In some embodiments the relationship is not linear. In some embodiments the relationship is fitted to a representation of chromosome X derived from one or more pregnant females bearing a female fetus. In some embodiments the fetal fraction *f* is determined according to equation (62) below or a variation thereof.

$$f = 2\frac{I+S\langle x\rangle - y}{S\langle x\rangle} \qquad (62)$$

where *f* is the fraction of fetal nucleic acid in a test sample, *y* is the measured experimental Y chromosomal representation from a test sample obtained from a pregnant female bearing a male fetus, *I* is a slope, *S* is a slope and $\langle x\rangle$ is a median MCR of chromosome X obtained from multiple pregnant females bearing a female fetus. The parameters *I* (the intercept) and S (the slope) are derived from a fitted linear relationship (see Example 10). Parameters *I* and S are sometimes derived from a linear relationship for chromosome X (MCR of chromosome X) and chromosome Y

(MCR of chromosome X) representations obtained from pregnant female bearing a male fetus after the linear relationship is fitted to a point (ie., forced to go through a point) representing a median chromosome X (median MCR of chromosome X, $\langle x \rangle$) and a median chromosome Y (median MCR of chromosome Y, $\langle y \rangle$) representation obtained from pregnant female bearing a female fetus. The resulting relationship (62) can determine the fraction of fetal nucleic acid in a test sample from a measured experimental Y chromosomal representation from a test sample where the test sample was obtained from a pregnant female bearing a male fetus. In some embodiments the fraction of the fetal nucleic acid is determined according to the slope and intercept of the fitted relationship, the measured experimental Y chromosomal representation from a test sample where the test sample was obtained from a pregnant female bearing a male fetus and a median X chromosome representation for a set of pregnant females bearing a female fetus.

In some embodiments $\langle x \rangle$, $\langle y \rangle$, *S* and *I* are derived from multiple MCRs for chromosome X and Y obtained from pregnant females bearing a male fetus and/or multiple pregnant females bearing a female fetus. In some embodiments a median of multiple MCRs for chromosome X and Y is determined from multiple MCRs obtained from pregnant females bearing a female fetus. In some embodiments, MCRs for chromosome X and Y are determined for multiple subjects. In some embodiments, *S*, *I* and $\langle x \rangle$ in equation (62) are constants or coefficients and are determined by methods described herein. In some embodiments, *S*, *I* and $\langle x \rangle$ in equation (62) are constants or coefficients and are determined from data obtained from multiple subjects. Multiple subjects is sometimes greater than about 10, greater than about 100, greater than about 500, greater than about 1000 or greater than about 10,000 subjects. In certain embodiments, multiple subjects is about 500 to about 50,000, about 500 to about 25,000, about 500 to about 10,000, about 500 to about 5000, or about 500 to about 2500 subjects.

In some embodiments, the mean chromosome X representation $\langle x \rangle$ is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof. In some embodiments, the sequence reads in (ii) are mapped to all chromosomes in a profile, all autosomes in a profile, a selected chromosome or a selected autosome in a profile. In some embodiments the sequence reads in (ii) are mapped to any subset of chromosomes, any chromosome or a segment thereof.

In some embodiments $\langle x \rangle$ is a constant with a value between about 0.027 and about 0.067, about 0.037 and about 0.057, or about 0.042 and about 0.053. In certain embodiments $\langle x \rangle$ is a constant with a value of about 0.042, 0.043, 0.044, 0.045, 0.046, 0.047, 0.048, 0.049, 0.050, 0.051, 0.052 or about 0.053. In some embodiments $\langle x \rangle$ is a constant with a value of about 0.04765159.

In some embodiments $\langle y \rangle$ is a constant with a value between about 0.00001 and about 0.0002, 0.00005 and about 0.00015, or between about 0.00005 and about 0.000125. In certain embodiments $\langle y \rangle$ is a constant with a value of about 0.00004, 0.00005, 0.00006, 0.00007, 0.00008, 0.00009, 0.00010, 0.00011, 0.00012, 0.00013, 0.00014 or about 0.00015. In some embodiments $\langle y \rangle$ is a constant with a value of about 0.0001054401.

In some embodiments *S* is a constant with a value between about 0.005 and about 0.03, about 0.0075 and about 0.025, or between about 0.010 and about 0.020. In certain embodiments *S* is a constant with a value of about 0.010, 0.011, 0.012, 0.013, 0.014, 0.015, 0.016, 0.017, 0.018, 0.019 or about 0.020. In some embodiments *S* is a constant with a value of about 0.01560178.

In some embodiments *I* is a constant with a value between about -0.1 and about -0.5, about -0.2 and about -0.4, or between about -0.275 and about -0.375. In certain embodiments *I* is a constant with a value of about -0.321, -0.322, -0.323, -0.324, -0.325, -0.326, -0.327, -0.328, -0.329, -0.330 or about -0.331. In some embodiments *I* is a constant with a value of about -0.3252008.

In some embodiments $\langle x \rangle$, $\langle y \rangle$, *S* and *I* are constants with values of about 0.04765159, 0.0001054401, 0.01560178 and -0.3252008, respectively.

Determining fetal ploidy from fetal fraction

A fetal ploidy can be determined, in part, from a measure of fetal fraction and the fetal ploidy determination is used to make a determination of the presence or absence of a genetic variation (e.g., a chromosome aneuploidy, a trisomy). For example, FIGS. 175-177 illustrate fetal ploidy determinations derived, in part, from fetal fractions measurements. A fetal ploidy can be determined, in part, from a measure of fetal fraction determined by any suitable method of fetal fraction determination including methods described herein. In some embodiments fetal ploidy is determined according to a fetal fraction determination and equation (8), (20), (21) or a variation or derivation thereof (see Example 2). In some embodiments, fetal ploidy is determined by a method

described below. In some embodiments each method described below requires a calculated reference count F_i (sometimes represented as f_i) determined for a portion (i.e. a bin, i) of a genome for multiple samples where the ploidy of the fetus for portion i of the genome is known to be euploid. In some embodiments an uncertainty value (e.g., a standard deviation, σ) is determined for the reference count f_i. In some embodiments a reference count f_i, an uncertainty value, a test sample count and/or a measured fetal fraction (F) are used to determine fetal ploidy according to a method described below. In some embodiments a reference count (e.g., an average, mean or median reference count) is normalized by a method described herein (e.g., bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and/or combinations thereof). In some embodiments a reference count of a segment of a genome known to be euploid is equal to 1 when the reference count is normalized by PERUN. In some embodiments both the reference count (e.g., for a fetus known to be euploid) and the counts of a test sample for a portion or segment of a genome are normalized by PERUN and the reference count is equal to 1. Likewise, in some embodiments, a reference count of a portion or segment of a genome known to be euploid is equal to 1 when the counts are normalized by (i.e., divided by) a median of the reference count. For example, in some embodiments both the reference count (e.g., for a fetus known to be euploid) and the counts of a test sample for a portion or segment of a genome are normalized by a median reference count, the normalized reference count is equal to 1 and the test sample count is normalized (e.g., divided by) the median reference count. In some embodiments both the reference count (e.g., for a fetus known to be euploid) and the counts of a test sample for a portion or segment of a genome are normalized by GCRM, GC, RM or a suitable method. In some embodiments a reference count is an average, mean or median reference count. A reference count is often a normalized count for a bin (e.g., a normalized genomic section elevation). In some embodiments a reference count and the counts for a test sample are raw counts. A reference count, in some embodiments, is determined from an average, mean or median count profile. In some embodiments, a reference count is a calculated genomic section level. In some embodiments a reference count of a reference sample and a count of a test sample (e.g., a patient sample, e.g., y_i) are normalized by the same method or process.

A Non-Limiting Example of a Fetal Ploidy Determination

In some embodiments a measurement of fetal fraction (F) is determined. This fetal fraction value is then used to determine fetal ploidy according to equation (8), a derivation or a variation thereof. In

some embodiments, a negative value is returned if the fetus is euploid and a positive value is returned if the fetus is not euploid. In some embodiments a negative value indicates the fetus is euploid for the segment of the genome considered. In certain embodiments, a value that is not negative indicates the fetus comprises an aneuploidy (e.g., a duplication). In certain embodiments, a value that is not negative indicates the fetus comprises a trisomy. In certain embodiments, any positive value indicates the fetus comprises an aneuploidy (e.g., a trisomy, a duplication).

In some embodiments a sum of square residuals is determined. For example, an equation representing the sum of square residuals derived from equation (8) is illustrated in equation (18). In some embodiments a sum of square residuals is determined from equation (8) for a ploidy value X set to a value of 1 (see equation (9)) and for a ploidy value set to a value of 3/2 (see equation (13)). In some embodiments the sum of square residuals (equations (9) and (13)) are determined for a segment of a genome or chromosome (e.g., for all bins *i* in a segment of the genome). For example, the sum of square residuals (e.g., equations (9) and (13)) can be determined for chromosome 21, 13, 18 or a portion thereof. In some embodiments, to determine a ploidy status of a fetus, the result of equation (13) is subtracted from equation (9) to arrive at a value, phi (e.g., see equation (14)). In certain embodiments, the sign (i.e. positive or negative) of the value phi determines the presence or absence of a fetal aneuploidy. In certain embodiments, a phi value (e.g., from equation (14)) that is negative indicates the absence of an aneuploidy (e.g., the fetus is euploid for bins *i*) and a phi value that is not negative indicates the presence of an aneuploidy (e.g., a trisomy).

In some embodiments the reference count f_i, the uncertainty value for the reference count σ and/or the measured fetal fraction (F) are used in equations (9) and (13) to determine the sum of square residuals for the sum of all bins *i*. In some embodiments the reference count f_i, the uncertainty value for the reference count σ and/or the measured fetal fraction (F) are used in equations (9) and (13) to determine fetal ploidy. In some embodiments the counts (e.g., normalized counts, e.g., calculated genomic section level), represented by y_i for bin *i*, for a test sample are used to determine the ploidy status of a fetus for bin *i*. For example, in certain embodiments, the ploidy status for a segment of a genome is determined according to a reference count f_i, an uncertainty value (e.g., from the reference count), a feta fraction (F) determined for a test sample and the counts y_i determined for the test sample where the ploidy status is determined according to equation (14) or a derivation or variation thereof. In some embodiments the counts y_i and/or

reference counts are normalized by a method described herein (e.g., bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof). In some embodiments a fetal ploidy status (e.g., euploid, aneuploid, trisomy) for a portion or segment of a genome or chromosome is determined by the non-limiting example described above and in Example 2.

Another Non-Limiting Example of a Fetal Ploidy Determination

In some embodiments a fetal fraction is determined from a test sample, counts *y* are determined for a test sample and both are used to determine a ploidy for a fetus from a test sample. In certain embodiments of the method described here, the value of fetal ploidy represented by X is not fixed or assumed. In certain embodiments of the method described here, fetal fraction F is fixed. In some embodiments, a ploidy (e.g., a ploidy value) is determined for a portion or segment of a genome according to equation (20) or (21)(see Example 2). In some embodiments of this method, a ploidy value is determined, where the value is close to 1, 3/2, or 5/4. In some embodiments a ploidy value of about 1 indicates a euploid fetus, a value of about 3/2 indicates a fetal trisomy and, in the case of twins, a value of about 5/4 indicates that one fetus comprises a trisomy and the other is euploid for the portion or segment of the genome considered. Additional information regarding determining the presence or absence of a fetal aneuploidy from a fetal ploidy determination is discussed in another section below.

In some embodiments, fetal fraction is determined, fixed at its determined value and fetal ploidy is determined from a regression. Any suitable regression can be utilized, non-limiting examples of which include a linear regression, a non-linear regression (e.g., a polynomial regression), and the like. In some embodiments, a linear regression is used according to equation (8), (20), (21) and/or a derivation or variation thereof. In some embodiments, the linear regression used is according to a sum of square residuals derived from equation (8), (20), (21) and/or a derivation or variation thereof. In some embodiments, fetal ploidy is determined according to equation (8), (20), (21) and/or a derivation or variation thereof and a regression is not used. In some embodiments, fetal ploidy is determined according to a sum of square residuals derived from equation (8), (20), (21) and/or a derivation or variation thereof for multiple bins *i* and a regression is not used. A derivation of an equation is any variation of the equation obtained from a mathematical proof of an equation.

In some embodiments a reference count f_i (described previously herein), an uncertainty value σ and/or a measured fetal fraction (F) are used in equations (20) and (21) to determine a fetal ploidy. In some embodiments a reference count f_i, an uncertainty value σ and/or a measured fetal fraction (F) are used in equations (20) or (21) to determine a fetal ploidy X for bin *i* or for a sum of multiple bins *i* (e.g., for the sum of all bins *i* for a chromosome or segment thereof). In some embodiments the counts (e.g., normalized counts, calculated genomic section level), represented by y_i for bin *i*, for a test sample are used to determine the ploidy of a fetus for a segment of a genome represented by multiple bins *i*. For example, in certain embodiments, the ploidy X for a segment of a genome is determined according to a reference count f_i, an uncertainty value, a feta fraction (F) determined for a test sample and the counts y_i determined for the test sample where the ploidy is determined according to equation (20), (21) or a derivation or variation thereof. In some embodiments the counts y_i and/or reference counts are normalized by a method described herein (e.g., bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof). In some embodiments the counts y_i and/or reference counts are normalized and/or processed by the same method (e.g., bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM, a method described herein or combinations thereof). In some embodiments counts y_i and f_i are counts mapped to the same portion or segment of a genome or chromosome.

The uncertainty value σ can be a suitable measure of error, non-limiting examples of which include standard deviation, standard error, calculated variance, p-value, and/or mean absolute deviation (MAD). The uncertainty value σ can be determined for any suitable measurement, non-limiting examples of which include Z-scores, Z-values, t-values, p-values, cross-validation error, genomic section level, calculated genomic section levels, elevations, counts, the like, or combinations thereof. In some embodiments σ is set to a value of 1. In some embodiments σ is not set to a value of 1. In some embodiments the value of σ is estimated and sometimes it is measured and/or calculated.

In some embodiments M_i is the ploidy of the mother (i.e., maternal ploidy) for a portion of the genome *i*. In some embodiments M_i is determined for the same patient (e.g., same test sample) from which y_i is determined. In some embodiments the maternal ploidy M_i is known or determined according to a method described herein. In some embodiments maternal ploidy is determined before or after padding (e.g., after making elevation adjustments). In certain embodiments M_i is

estimated or determined from visualizing a profile. In some embodiments the maternal ploidy M_i is not known. In some embodiments the maternal ploidy M_i is assumed. For example, in some embodiments it is assumed or known that the mother has no deletions and/or duplications in the segment of the genome being evaluated. In some embodiments it is assumed or known that maternal ploidy is 1. In some embodiments maternal ploidy is set to a value of 1 after padding (e.g., after making elevations adjustments). In some embodiments maternal ploidy is ignored and is set to a value of 1. In some embodiments equation (21) is derived from equation (20) with the assumption that the mother has no deletions and/or duplications in the segment of the genome being evaluated.

In some embodiments a method for determining fetal ploidy is according to nucleic acid sequence reads for a test sample obtained from a pregnant female. In some embodiments the sequence reads are reads of circulating cell-free nucleic acid from a sample (e.g., a test sample). In some embodiments, a method for determining fetal ploidy comprises obtaining counts of sequence reads mapped to portions of a reference genome. In some embodiments the sequence reads are mapped to a subset of portions of the reference genome. In some embodiments determining fetal ploidy comprises determining a fetal fraction. In some embodiments determining fetal ploidy comprises calculating or determining genomic section levels. In certain embodiments determining fetal ploidy comprises determining a fetal fraction and calculating or determining genomic section levels. In some embodiments the fetal fraction and the calculated genomic section levels are determined from the same test sample (e.g., same part of the test sample). In some embodiments the fetal fraction and the calculated genomic section levels are determined from the same reads obtained from the same test sample (e.g., same part of the test sample). In some embodiments the fetal fraction and the calculated genomic section levels are determined from the same reads obtained from the same sequencing run and/or from the same flow cell. In some embodiments the fetal fraction and the calculated genomic section levels are determined from the same equipment and/or machine (e.g., sequencing apparatus, flow cell, or the like).

In some embodiments a method for determining fetal ploidy is determined according to a fetal fraction determination and normalized counts (e.g., calculated genomic section levels) wherein the fetal fraction determination and the normalized counts (e.g., calculated genomic section levels) are determined from different parts of a test sample (e.g., different aliquots, or e.g., different test samples taken at about the same time from the same subject or patient). For example, sometimes a fetal fraction is determined from a first part of a test sample and normalized counts and/or

genomic section levels are determined from a second part of the test sample. In some embodiments the fetal fraction and the calculated genomic section levels are determined from different test samples (e.g., different parts of a test sample) taken from the same subject (e.g., patient). In some embodiments the fetal fraction and the calculated genomic section levels are determined from reads obtained at different times. In some embodiments the fetal fraction determination and the normalized counts (e.g., calculated genomic section levels) are determined from different equipment and/or from different machines (e.g., sequencing apparatus, flow cell, or the like).

Some Embodiments of Determining a Chromosome Aneuploidy

In some embodiments the presence or absence of a fetal chromosomal aneuploidy (e.g., a trisomy) is determined from a fetal ploidy determination. In some embodiments a fetal ploidy is determined by a suitable method described herein. In some certain embodiments a fetal ploidy determination of about 1.20 or greater, 1.25 or greater, 1.30 or greater, about 1.35 or greater, about 1.4 or greater, or about 1.45 or greater indicates the presence of a fetal chromosome aneuploidy (e.g., the presence of a fetal trisomy). In some embodiments a fetal ploidy determination of about 1.20 to about 2.0, about 1.20 to about 1.9, about 1.20 to about 1.85, about 1.20 to about 1.8, about 1.25 to about 2.0, about 1.25 to about 1.9, about 1.25 to about 1.85, about 1.25 to about 1.8, about 1.3 to about 2.0, about 1.3 to about 1.9, about 1.3 to about 1.85, about 1.3 to about 1.8, about 1.35 to about 2.0, about 1.35 to about 1.9, about 1.35 to about 1.8, about 1.4 to about 2.0, about 1.4 to about 1.85 or about 1.4 to about 1.8 indicates the presence of a fetal chromosome aneuploidy (e.g., the presence of a fetal trisomy). In some embodiments the fetal aneuploidy is trisomy. In some embodiments the fetal aneuploidy is trisomy of chromosome 13, 18 and/or 21.

In some embodiments a fetal ploidy of less than about 1.35, less than about 1.30, less than about 1.25, less than about 1.20 or less than about 1.15 indicates the absence of a fetal aneuploidy (e.g., the absence of a fetal trisomy, e.g., euploid). In some embodiments a fetal ploidy determination of about 0.7 to about 1.35, about 0.7 to about 1.30, about 0.7 to about 1.25, about 0.7 to about 1.20, about 0.7 to about 1.15, about 0.75 to about 1.35, about 0.75 to about 1.30, about 0.75 to about 1.25, about 0.75 to about 1.20, about 0.75 to about 1.15, about 0.8 to about 1.35, about 0.8 to about 1.30, about 0.8 to about 1.25, about 0.8 to about 1.20, or about 0.8 to about 1.15 indicates the absence of a fetal chromosome aneuploidy (e.g., the absence of a fetal trisomy, e.g., euploid).

In some embodiments a fetal ploidy of less than about 0.8, less than about 0.75, less than about 0.70 or less than about 0.6 indicates the presence of a fetal aneuploidy (e.g., the presence of a chromosome deletion). In some embodiments a fetal ploidy determination of about 0 to about 0.8, about 0 to about 0.75, about 0 to about 0.70, about 0 to about 0.65, about 0 to about 0.60, about 0.1 to about 0.8, about 0.1 to about 0.75, about 0.1 to about 0.70, about 0.1 to about 0.65, about 0.1 to about 0.60, about 0.2 to about 0.8, about 0.2 to about 0.75, about 0.2 to about 0.70, about 0.2 to about 0.65, about 0.2 to about 0.60, about 0.25 to about 0.8, about 0.25 to about 0.75, about 0.25 to about 0.70, about 0.25 to about 0.65, about 0.25 to about 0.60, about 0.3 to about 0.8, about 0.3 to about 0.75, about 0.3 to about 0.70, about 0.3 to about 0.65, about 0.3 to about 0.60 indicates the presence of a fetal chromosome aneuploidy (e.g., the presence of a chromosome deletion). In some embodiments the fetal aneuploidy determined is a whole chromosome deletion.

In some embodiments a determination of the presence or absence of a fetal aneuploidy (e.g., according to one or more of the ranges of a ploidy determination above) is determined according to a call zone. In certain embodiments a call is made (e.g., a call determining the presence or absence of a genetic variation, e.g., an outcome) when a value (e.g., a ploidy value, a fetal fraction value, a level of uncertainty) or collection of values falls within a pre-defined range (e.g., a zone, a call zone). In some embodiments a call zone is defined according to a collection of values that are obtained from the same patient sample. In certain embodiments a call zone is defined according to a collection of values that are derived from the same chromosome or segment thereof. In some embodiments a call zone based on a ploidy determination is defined according a level of confidence (e.g., high level of confidence, e.g., low level of uncertainty) and/or a fetal fraction. In some embodiments a call zone is defined according to a ploidy determination and a fetal fraction of about 2.0% or greater, about 2.5% or greater, about 3% or greater, about 3.25% or greater, about 3.5% or greater, about 3.75% or greater, or about 4.0 % or greater. For example, in some embodiments a call is made that a fetus comprises a trisomy 21 based on a ploidy determination of greater than 1.25 with a fetal fraction determination of 2% or greater or 4% or greater for a sample obtained from a pregnant female bearing a fetus. In certain embodiments, for example, a call is made that a fetus is euploid based on a ploidy determination of less than 1.25 with a fetal fraction determination of 2% or greater or 4% or greater for a sample obtained from a pregnant female bearing a fetus. In some embodiments a call zone is defined by a confidence level of about 99% or greater, about 99.1% or greater, about 99.2% or greater, about 99.3% or greater, about 99.4% or greater, about 99.5% or greater, about 99.6% or greater, about 99.7% or greater, about 99.8%

or greater or about 99.9% or greater. In some embodiments a call is made without using a call zone. In some embodiments a call is made using a call zone and additional data or information. In some embodiments a call is made based on a ploidy value without the use of a call zone. In some embodiments a call is made without calculating a ploidy value. In some embodiments a call is made based on visual inspection of a profile (e.g., visual inspection of genomic section levels). A call can be made by any suitable method based in full, or in part, upon determinations, values and/or data obtained by methods described herein, non-limiting examples of which include a fetal ploidy determination, a fetal fraction determination, maternal ploidy, uncertainty and/or confidence determinations, genomic sections levels, profiles, z-scores, expected chromosome representations, measured chromosome representations, counts (e.g., normalized counts, raw counts), fetal or maternal copy number variations (e.g., categorized copy number variations), significantly different elevations, adjusted elevations (e.g., padding), the like or combinations thereof.

In some embodiments a no-call zone is where a call is not made. In some embodiments a no-call zone is defined by a value or collection of values that indicate low accuracy, high risk, high error, low level of confidence, high level of uncertainty, the like or a combination thereof. In some embodiments a no-call zone is defined, in part, by a fetal fraction of about 5% or less, about 4% or less, about 3% or less, about 2.5% or less, about 2.0% or less, about 1.5% or less or about 1.0% or less.

Outcome

Methods described herein can provide a determination of the presence or absence of one or more genetic variations (e.g., a fetal aneuploidy, microduplication, microdeletion, trisomy, the like or combinations thereof) for a sample, thereby providing an outcome (e.g., thereby providing an outcome determinative of the presence or absence of one or more genetic variations (e.g., one or more fetal aneuploidies)). A genetic variation often includes a gain, a loss and/or alteration (e.g., duplication, deletion, fusion, insertion, mutation, reorganization, substitution or aberrant methylation) of genetic information (e.g., chromosomes, segments of chromosomes, polymorphic regions, translocated regions, altered nucleotide sequence, the like or combinations of the foregoing) that results in a detectable change in the genome or genetic information of a test subject with respect to a reference. Presence or absence of one or more genetic variations can be

determined by transforming, analyzing and/or manipulating sequence reads that have been mapped to genomic sections (e.g., genomic bins).

Methods described herein sometimes determine presence or absence of one or more of a fetal aneuploidy (e.g., full chromosome aneuploidy, partial chromosome aneuploidy or segmental chromosomal aberration (e.g., mosaicism, deletion and/or insertion)) for a test sample from a pregnant female bearing a fetus. Methods described herein sometimes determine presence or absence of one or more of a fetal aneuploidy (e.g., full chromosome aneuploidy, partial chromosome aneuploidy or segmental chromosomal aberration (e.g., mosaicism, deletion and/or insertion)) in a mother bearing a fetus, a fetus and/or in fetal tissue (e.g., placenta). In certain embodiments methods described herein detect euploidy or lack of euploidy (non-euploidy) for a sample from a pregnant female bearing a fetus. Methods described herein, in some embodiments, detect a trisomy for one or more chromosomes in a sample, non-limiting examples of which include chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, Y or a combinations thereof.

Methods described herein, in some embodiments, determine the presence or absence of one or more deletions, duplications, microdeletions, microduplications or a combinations thereof from a test sample from a pregnant female bearing a fetus. Methods described herein, in some embodiments, determine the presence or absence of one or more deletions, duplications, microdeletions, microduplications or a combinations thereof in a mother bearing a fetus, a fetus and/or in fetal tissue (e.g., placenta). Methods described herein, in some embodiments, detect one or more deletions, duplications, microdeletions, microduplications or a combination thereof in one or more chromosomes, non-limiting examples of which include chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, and Y.

In some embodiments, presence or absence of one or more genetic variations (e.g., a fetal aneuploidy) is determined by a method described herein, by a method known in the art or by a combination thereof. Presence or absence of one or more genetic variations generally is determined from counts of sequence reads mapped to genomic sections of a reference genome. Counts of sequence reads utilized to determine presence or absence of one or more genetic variations sometimes are raw counts and/or filtered counts, and often are normalized counts. A suitable normalization process or processes can be used to generate normalized counts, non-limiting examples of which include bin-wise normalization, normalization by GC content, linear and

nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof. Normalized counts sometimes are expressed as one or more elevations in a profile for a particular set or sets of genomic sections. Normalized counts sometimes are adjusted or padded prior to determining presence or absence of one or more genetic variations.

The presence or absence of one or more genetic variations (e.g., fetal aneuploidy) sometimes is determined without comparing counts for a set of genomic sections to a reference. Counts measured for a test sample and are in a test region (e.g., a set of genomic sections of interest) are referred to as "test counts" herein. Test counts sometimes are processed counts, averaged or summed counts, a representation, normalized counts, or one or more elevations, as described herein. In certain embodiments test counts are averaged or summed (e.g., an average, mean, median, mode or sum is calculated) for a set of genomic sections, and the averaged or summed counts are compared to a threshold or range. Test counts sometimes are expressed as a representation, which can be expressed as a ratio or percentage of counts for a first set of genomic sections to counts for a second set of genomic sections. In certain embodiments the first set of genomic sections is for one or more test chromosomes (e.g., chromosome 4, 9, 13, 16, 18, 20, 21 and 22, the like or a combination thereof) and sometimes the second set of genomic sections is for the genome or a part of the genome (e.g., autosomes or autosomes and sex chromosomes). In certain embodiments a representation is compared to a threshold or range. In certain embodiments test counts are expressed as one or more elevations for normalized counts over a set of genomic sections, and the one or more elevations are compared to a threshold or range. Test counts (e.g., averaged or summed counts, representation, normalized counts, one or more elevations) above or below a particular threshold, in a particular range or outside a particular range sometimes are determinative of the presence of a genetic variation or lack of euploidy (e.g., not euploidy). Test counts (e.g., averaged or summed counts, representation, normalized counts, one or more elevations) below or above a particular threshold, in a particular range or outside a particular range sometimes are determinative of the absence of a genetic variation or euploidy.

The presence or absence of one or more genetic variations (e.g., fetal aneuploidy) sometimes is determined by comparing test counts (e.g., raw counts, filtered counts, averaged or summed counts, representation, normalized counts, one or more elevations, for a set of genomic sections) to a reference. A reference can be a suitable determination of counts. Counts for a reference sometimes are raw counts, filtered counts, averaged or summed counts, representation,

normalized counts, one or more elevations, for a set of genomic sections. Reference counts often are counts for a euploid test region.

In certain embodiments, test counts sometimes are for a first set of genomic sections and a reference includes counts for a second set of genomic sections different than the first set of genomic sections. Reference counts sometimes are for a nucleic acid sample from the same pregnant female from which the test sample is obtained. In certain embodiments reference counts are for a nucleic acid sample from one or more pregnant females different than the female from which the test sample was obtained. In some embodiments, a first set of genomic sections is in a chromosome (e.g., a target chromosome, e.g., 4, 9, 13, 16, 18, 20, 21 and/or 22, a segment thereof or combination of the foregoing), and the second set of genomic sections is in another chromosome or chromosomes or segment thereof. In a non-limiting example, where a first set of genomic sections is in a target chromosome (e.g., chromosome 21) or segment thereof, a second set of genomic sections often is in another chromosome (e.g., chromosome 1, chromosome 13, chromosome 14, chromosome 18, chromosome 19, a segment thereof or combination of the foregoing). A reference often is located in a chromosome or segment thereof that is typically euploid. For example, chromosome 1 and chromosome 19 often are euploid in fetuses owing to a high rate of early fetal mortality associated with chromosome 1 and chromosome 19 aneuploidies. A measure of deviation between the test counts and the reference counts can be generated.

In certain embodiments a reference comprises counts for the same set of genomic sections as for the test counts, where the counts for the reference are from one or more reference samples (e.g., often multiple reference samples from multiple reference subjects). A reference sample often is from one or more pregnant females different than the female from which a test sample is obtained. A measure of deviation between the test counts and the reference counts can be generated.

A suitable measure of deviation between test counts and reference counts can be selected, non-limiting examples of which include standard deviation, average absolute deviation, median absolute deviation, maximum absolute deviation, standard score (e.g., z-value, z-score, normal score, standardized variable) and the like. In some embodiments, reference samples are euploid for a test region and deviation between the test counts and the reference counts is assessed. A deviation of less than three between test counts and reference counts (e.g., 3-sigma for standard deviation) often is indicative of a euploid test region (e.g., absence of a genetic variation). A deviation of greater than three between test counts and reference counts often is indicative of a

non-euploid test region (e.g., presence of a genetic variation). Test counts significantly below reference counts, which reference counts are indicative of euploidy, sometimes are determinative of a monosomy. Test counts significantly above reference counts, which reference counts are indicative of euploidy, sometimes are determinative of a trisomy. A measure of deviation between test counts for a test sample and reference counts for multiple reference subjects can be plotted and visualized (e.g., z-score plot).

Any other suitable reference can be factored with test counts for determining presence or absence of a genetic variation (or determination of euploid or non-euploid) for a test region of a test sample. For example, a fetal fraction determination can be factored with test counts to determine the presence or absence of a genetic variation. A suitable process for quantifying fetal fraction can be utilized, non-limiting examples of which include a mass spectrometric process, sequencing process or combination thereof.

Laboratory personnel (e.g., a laboratory manager) can analyze values (e.g., test counts, reference counts, level of deviation) underlying a determination of the presence or absence of one or more genetic variations (or determination of euploid or non-euploid for a test region). For calls pertaining to presence or absence of one or more genetic variations that are close or questionable, laboratory personnel can re-order the same test, and/or order a different test (e.g., karyotyping and/or amniocentesis in the case of fetal aneuploidy determinations), that makes use of the same or different sample nucleic acid from a test subject.

One or more genetic variations sometimes are associated with a medical condition. An outcome determinative of a genetic variation is sometimes an outcome determinative of the presence or absence of a condition (e.g., a medical condition), disease, syndrome or abnormality, or includes, detection of a condition, disease, syndrome or abnormality (e.g., non-limiting examples listed in Tables 1A-1G). In some embodiments a diagnosis comprises assessment of an outcome. An outcome determinative of the presence or absence of a condition (e.g., a medical condition), disease, syndrome or abnormality by methods described herein can sometimes be independently verified by further testing (e.g., by karyotyping and/or amniocentesis).

Analysis and processing of data can provide one or more outcomes. The term “outcome” as used herein can refer to a result of data processing that facilitates determining the presence or absence of one or more genetic variations (e.g., an aneuploidy, a copy number variation). In certain

embodiments the term “outcome” as used herein refers to a conclusion that predicts and/or determines the presence or absence of one or more genetic variations (e.g., an aneuploidy, a copy number variation). In certain embodiments the term “outcome” as used herein refers to a conclusion that predicts and/or determines a risk or probability of the presence or absence of one or more genetic variations (e.g., an aneuploidy, a copy number variation) in a subject (e.g., a fetus). A diagnosis sometimes comprises use of an outcome. For example, a health practitioner may analyze an outcome and provide a diagnosis bases on, or based in part on, the outcome. In some embodiments, determination, detection or diagnosis of a condition, syndrome or abnormality (e.g., listed in Tables 1A-1G) comprises use of an outcome determinative of the presence or absence of one or more genetic variations. In some embodiments, an outcome based on counted mapped sequence reads or transformations thereof is determinative of the presence or absence of one or more genetic variations. In certain embodiments, an outcome generated utilizing one or more methods (e.g., data processing methods) described herein is determinative of the presence or absence of one or more conditions, syndromes or abnormalities listed in Tables 1A-1G. In certain embodiments, an outcome generated utilizing one or more methods described herein is determinative of the presence or absence of one or more conditions, syndromes or abnormalities, non-limiting examples of which include cancers (e.g., certain lymphoproliferative disorders, leukemia), autoimmune disorders, immune deficiencies, metabolic disorders or diseases, neurological disorders, bone and structural deformities, blood disorders, growth related disorders, the like or combinations thereof. In certain embodiments, an outcome generated utilizing one or more methods described herein is determinative of the presence or absence of any condition, syndromes, disorder, disease or abnormality known to be associated or caused by a genetic variation detectable by a method described herein. In some embodiments, an outcome generated utilizing one or more methods described herein is predictive of the death of a fetus or impending abortion of a pregnancy. In some embodiments, an outcome generated utilizing one or more methods described herein is determinative of the presence or absence of a health risk to a mother of a fetus. For example, in some embodiments a genetic variation is detected in a fetus utilizing one or more methods described herein and the genetic variation is predictive of a health risk (e.g., a risk a impending health complications, a risk of death) to the mother of the fetus. In certain embodiments a diagnosis comprises a determination of a presence or absence of a condition, syndrome or abnormality. Often a diagnosis comprises a determination of one or more genetic variations as the nature and/or cause of a condition, syndrome or abnormality. In certain embodiments an outcome is not a diagnosis. An outcome often comprises one or more numerical values generated using a processing method described herein in the context of one or more

considerations of probability. A consideration of risk or probability can include, but is not limited to: an uncertainty value, a measure of variability, confidence level, sensitivity, specificity, standard deviation, coefficient of variation (CV) and/or confidence level, Z-scores, Chi values, Phi values, ploidy values, fitted fetal fraction, area ratios, median elevation, the like or combinations thereof. A consideration of probability can facilitate determining whether a subject is at risk of having, or has, a genetic variation, and an outcome determinative of a presence or absence of a genetic disorder often includes such a consideration.

An outcome sometimes is a phenotype. An outcome sometimes is a phenotype with an associated level of confidence (e.g., an uncertainty value, e.g., a fetus is positive for trisomy 21 with a confidence level of 99%, a test subject is negative for a cancer associated with a genetic variation at a confidence level of 95%). Different methods of generating outcome values sometimes can produce different types of results. Generally, there are four types of possible scores or calls that can be made based on outcome values generated using methods described herein: true positive, false positive, true negative and false negative. The terms “score”, “scores”, “call” and “calls” as used herein refer to calculating the probability that a particular genetic variation is present or absent in a subject/sample. The value of a score may be used to determine, for example, a variation, difference, or ratio of mapped sequence reads that may correspond to a genetic variation. For example, calculating a positive score for a selected genetic variation or genomic section from a data set, with respect to a reference genome can lead to an identification of the presence or absence of a genetic variation, which genetic variation sometimes is associated with a medical condition (e.g., cancer, preeclampsia, trisomy, monosomy, and the like). In some embodiments, an outcome comprises an elevation, a profile and/or a plot (e.g., a profile plot). In those embodiments in which an outcome comprises a profile, a suitable profile or combination of profiles can be used for an outcome. Non-limiting examples of profiles that can be used for an outcome include z-score profiles, p-value profiles, chi value profiles, phi value profiles, the like, and combinations thereof

An outcome generated for determining the presence or absence of a genetic variation sometimes includes a null result (e.g., a data point between two clusters, a numerical value with a standard deviation that encompasses values for both the presence and absence of a genetic variation, a data set with a profile plot that is not similar to profile plots for subjects having or free from the genetic variation being investigated). In some embodiments, an outcome indicative of a null result still is a determinative result, and the determination can include the need for additional information

and/or a repeat of the data generation and/or analysis for determining the presence or absence of a genetic variation.

An outcome can be generated after performing one or more processing steps described herein, in some embodiments. In certain embodiments, an outcome is generated as a result of one of the processing steps described herein, and in some embodiments, an outcome can be generated after each statistical and/or mathematical manipulation of a data set is performed. An outcome pertaining to the determination of the presence or absence of a genetic variation can be expressed in a suitable form, which form comprises without limitation, a probability (e.g., odds ratio, p-value), likelihood, value in or out of a cluster, value over or under a threshold value, value within a range (e.g., a threshold range), value with a measure of variance or confidence, or risk factor, associated with the presence or absence of a genetic variation for a subject or sample. In certain embodiments, comparison between samples allows confirmation of sample identity (e.g., allows identification of repeated samples and/or samples that have been mixed up (e.g., mislabeled, combined, and the like)).

In some embodiments, an outcome comprises a value above or below a predetermined threshold or cutoff value (e.g., greater than 1, less than 1), and an uncertainty or confidence level associated with the value. In certain embodiments a predetermined threshold or cutoff value is an expected elevation or an expected elevation range. An outcome also can describe an assumption used in data processing. In certain embodiments, an outcome comprises a value that falls within or outside a predetermined range of values (e.g., a threshold range) and the associated uncertainty or confidence level for that value being inside or outside the range. In some embodiments, an outcome comprises a value that is equal to a predetermined value (e.g., equal to 1, equal to zero), or is equal to a value within a predetermined value range, and its associated uncertainty or confidence level for that value being equal or within or outside a range. An outcome sometimes is graphically represented as a plot (e.g., profile plot).

As noted above, an outcome can be characterized as a true positive, true negative, false positive or false negative. The term “true positive” as used herein refers to a subject correctly diagnosed as having a genetic variation. The term “false positive” as used herein refers to a subject wrongly identified as having a genetic variation. The term “true negative” as used herein refers to a subject correctly identified as not having a genetic variation. The term “false negative” as used herein refers to a subject wrongly identified as not having a genetic variation. Two measures of

performance for any given method can be calculated based on the ratios of these occurrences: (i) a sensitivity value, which generally is the fraction of predicted positives that are correctly identified as being positives; and (ii) a specificity value, which generally is the fraction of predicted negatives correctly identified as being negative. The term "sensitivity" as used herein refers to the number of true positives divided by the number of true positives plus the number of false negatives, where sensitivity (sens) may be within the range of $0 \leq sens \leq 1$. Ideally, the number of false negatives equal zero or close to zero, so that no subject is wrongly identified as not having at least one genetic variation when they indeed have at least one genetic variation. Conversely, an assessment often is made of the ability of a prediction algorithm to classify negatives correctly, a complementary measurement to sensitivity. The term "specificity" as used herein refers to the number of true negatives divided by the number of true negatives plus the number of false positives, where sensitivity (spec) may be within the range of $0 \leq spec \leq 1$. Ideally, the number of false positives equal zero or close to zero, so that no subject is wrongly identified as having at least one genetic variation when they do not have the genetic variation being assessed.

In certain embodiments, one or more of sensitivity, specificity and/or confidence level are expressed as a percentage. In some embodiments, the percentage, independently for each variable, is greater than about 90% (e.g., about 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99%, or greater than 99% (e.g., about 99.5%, or greater, about 99.9% or greater, about 99.95% or greater, about 99.99% or greater)). Coefficient of variation (CV) in some embodiments is expressed as a percentage, and sometimes the percentage is about 10% or less (e.g., about 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1%, or less than 1% (e.g., about 0.5% or less, about 0.1% or less, about 0.05% or less, about 0.01% or less)). A probability (e.g., that a particular outcome is not due to chance) in certain embodiments is expressed as a Z-score, a p-value, or the results of a t-test. In some embodiments, a measured variance, confidence interval, sensitivity, specificity and the like (e.g., referred to collectively as confidence parameters) for an outcome can be generated using one or more data processing manipulations described herein. Specific examples of generating outcomes and associated confidence levels are described in the Example section.

A method that has sensitivity and specificity equaling one, or 100%, or near one (e.g., between about 90% to about 99%) sometimes is selected. In some embodiments, a method having a sensitivity equaling 1, or 100% is selected, and in certain embodiments, a method having a sensitivity near 1 is selected (e.g., a sensitivity of about 90%, a sensitivity of about 91%, a sensitivity of about 92%, a sensitivity of about 93%, a sensitivity of about 94%, a sensitivity of

about 95%, a sensitivity of about 96%, a sensitivity of about 97%, a sensitivity of about 98%, or a sensitivity of about 99%). In some embodiments, a method having a specificity equaling 1, or 100% is selected, and in certain embodiments, a method having a specificity near 1 is selected (e.g., a specificity of about 90%, a specificity of about 91%, a specificity of about 92%, a specificity of about 93%, a specificity of about 94%, a specificity of about 95%, a specificity of about 96%, a specificity of about 97%, a specificity of about 98%, or a specificity of about 99%).

In some embodiments, presence or absence of a genetic variation (e.g., chromosome aneuploidy) is determined for a fetus. In such embodiments, presence or absence of a fetal genetic variation (e.g., fetal chromosome aneuploidy) is determined.

In certain embodiments, presence or absence of a genetic variation (e.g., chromosome aneuploidy) is determined for a sample. In such embodiments, presence or absence of a genetic variation in sample nucleic acid (e.g., chromosome aneuploidy) is determined. In some embodiments, a variation detected or not detected resides in sample nucleic acid from one source but not in sample nucleic acid from another source. Non-limiting examples of sources include placental nucleic acid, fetal nucleic acid, maternal nucleic acid, cancer cell nucleic acid, non-cancer cell nucleic acid, the like and combinations thereof. In non-limiting examples, a particular genetic variation detected or not detected (i) resides in placental nucleic acid but not in fetal nucleic acid and not in maternal nucleic acid; (ii) resides in fetal nucleic acid but not maternal nucleic acid; or (iii) resides in maternal nucleic acid but not fetal nucleic acid.

After one or more outcomes have been generated, an outcome often is used to provide a determination of the presence or absence of a genetic variation and/or associated medical condition. An outcome typically is provided to a health care professional (e.g., laboratory technician or manager; physician or assistant). Often an outcome is provided by an outcome module. In certain embodiments an outcome is provided by a plotting module. In certain embodiments an outcome is provided on a peripheral or component of an apparatus. For example, sometimes an outcome is provided by a printer or display. In some embodiments, an outcome determinative of the presence or absence of a genetic variation is provided to a healthcare professional in the form of a report, and in certain embodiments the report comprises a display of an outcome value and an associated confidence parameter. Generally, an outcome can be displayed in a suitable format that facilitates determination of the presence or absence of a genetic variation and/or medical condition. Non-limiting examples of formats suitable for use for

reporting and/or displaying data sets or reporting an outcome include digital data, a graph, a 2D graph, a 3D graph, and 4D graph, a picture, a pictograph, a chart, a bar graph, a pie graph, a diagram, a flow chart, a scatter plot, a map, a histogram, a density chart, a function graph, a circuit diagram, a block diagram, a bubble map, a constellation diagram, a contour diagram, a cartogram, spider chart, Venn diagram, nomogram, and the like, and combination of the foregoing. Various examples of outcome representations are shown in the drawings and are described in the Examples.

Generating an outcome can be viewed as a transformation of nucleic acid sequence read data, or the like, into a representation of a subject's cellular nucleic acid, in certain embodiments. For example, analyzing sequence reads of nucleic acid from a subject and generating a chromosome profile and/or outcome can be viewed as a transformation of relatively small sequence read fragments to a representation of relatively large chromosome structure. In some embodiments, an outcome results from a transformation of sequence reads from a subject (e.g., a pregnant female), into a representation of an existing structure (e.g., a genome, a chromosome or segment thereof) present in the subject (e.g., a maternal and/or fetal nucleic acid). In some embodiments, an outcome comprises a transformation of sequence reads from a first subject (e.g., a pregnant female), into a composite representation of structures (e.g., a genome, a chromosome or segment thereof), and a second transformation of the composite representation that yields a representation of a structure present in a first subject (e.g., a pregnant female) and/or a second subject (e.g., a fetus).

Use of Outcomes

A health care professional, or other qualified individual, receiving a report comprising one or more outcomes determinative of the presence or absence of a genetic variation can use the displayed data in the report to make a call regarding the status of the test subject or patient. The healthcare professional can make a recommendation based on the provided outcome, in some embodiments. A health care professional or qualified individual can provide a test subject or patient with a call or score with regards to the presence or absence of the genetic variation based on the outcome value or values and associated confidence parameters provided in a report, in some embodiments. In certain embodiments, a score or call is made manually by a healthcare professional or qualified individual, using visual observation of the provided report. In certain embodiments, a score or call is made by an automated routine, sometimes embedded in software, and reviewed by a healthcare

professional or qualified individual for accuracy prior to providing information to a test subject or patient. The term "receiving a report" as used herein refers to obtaining, by a communication means, a written and/or graphical representation comprising an outcome, which upon review allows a healthcare professional or other qualified individual to make a determination as to the presence or absence of a genetic variation in a test subject or patient. The report may be generated by a computer or by human data entry, and can be communicated using electronic means (e.g., over the internet, via computer, via fax, from one network location to another location at the same or different physical sites), or by a other method of sending or receiving data (e.g., mail service, courier service and the like). In some embodiments the outcome is transmitted to a health care professional in a suitable medium, including, without limitation, in verbal, document, or file form. The file may be, for example, but not limited to, an auditory file, a computer readable file, a paper file, a laboratory file or a medical record file.

The term "providing an outcome" and grammatical equivalents thereof, as used herein also can refer to a method for obtaining such information, including, without limitation, obtaining the information from a laboratory (e.g., a laboratory file). A laboratory file can be generated by a laboratory that carried out one or more assays or one or more data processing steps to determine the presence or absence of the medical condition. The laboratory may be in the same location or different location (e.g., in another country) as the personnel identifying the presence or absence of the medical condition from the laboratory file. For example, the laboratory file can be generated in one location and transmitted to another location in which the information therein will be transmitted to the pregnant female subject. The laboratory file may be in tangible form or electronic form (e.g., computer readable form), in certain embodiments.

In some embodiments, an outcome can be provided to a health care professional, physician or qualified individual from a laboratory and the health care professional, physician or qualified individual can make a diagnosis based on the outcome. In some embodiments, an outcome can be provided to a health care professional, physician or qualified individual from a laboratory and the health care professional, physician or qualified individual can make a diagnosis based, in part, on the outcome along with additional data and/or information and other outcomes.

A healthcare professional or qualified individual, can provide a suitable recommendation based on the outcome or outcomes provided in the report. Non-limiting examples of recommendations that can be provided based on the provided outcome report includes, surgery, radiation therapy,

chemotherapy, genetic counseling, after birth treatment solutions (e.g., life planning, long term assisted care, medicaments, symptomatic treatments), pregnancy termination, organ transplant, blood transfusion, the like or combinations of the foregoing. In some embodiments the recommendation is dependent on the outcome based classification provided (e.g., Down's syndrome, Turner syndrome, medical conditions associated with genetic variations in T13, medical conditions associated with genetic variations in T18).

Genetic Variations and Medical Conditions

The presence or absence of a genetic variance can be determined using a method or apparatus described herein. In certain embodiments, the presence or absence of one or more genetic variations is determined according to an outcome provided by methods and apparatuses described herein. In some embodiments the presence or absence of a complex genetic variation is determined by methods described herein. A complex genetic variation is the presence of two or more genetic variations. A genetic variation generally is a particular genetic phenotype present in certain individuals, and often a genetic variation is present in a statistically significant sub-population of individuals. In some embodiments, a genetic variation is a chromosome abnormality (e.g., aneuploidy), partial chromosome abnormality or mosaicism, each of which is described in greater detail herein. Non-limiting examples of genetic variations include one or more deletions (e.g., micro-deletions), duplications (e.g., micro-duplications), insertions, mutations, polymorphisms (e.g., single-nucleotide polymorphisms), fusions, repeats (e.g., short tandem repeats), distinct methylation sites, distinct methylation patterns, the like and combinations thereof. An insertion, repeat, deletion, duplication, mutation or polymorphism can be of any length, and in some embodiments, is about 1 base or base pair (bp) to about 250 megabases (Mb) in length. In some embodiments, an insertion, repeat, deletion, duplication, mutation or polymorphism is about 1 base or base pair (bp) to about 1,000 kilobases (kb) in length (e.g., about 10 bp, 50 bp, 100 bp, 500 bp, 1kb, 5 kb, 10kb, 50 kb, 100 kb, 500 kb, or 1000 kb in length).

A genetic variation is sometime a deletion. In certain embodiments a deletion is a mutation (e.g., a genetic aberration) in which a part of a chromosome or a sequence of DNA is missing. A deletion is often the loss of genetic material. Any number of nucleotides can be deleted. A deletion can comprise the deletion of one or more entire chromosomes, a segment of a chromosome, an allele, a gene, an intron, an exon, any non-coding region, any coding region, a segment thereof or

combination thereof. A deletion can comprise a microdeletion. A deletion can comprise the deletion of a single base.

A genetic variation is sometimes a genetic duplication. In certain embodiments a duplication is a mutation (e.g., a genetic aberration) in which a part of a chromosome or a sequence of DNA is copied and inserted back into the genome. In certain embodiments a genetic duplication (i.e. duplication) is any duplication of a region of DNA. In some embodiments a duplication is a nucleic acid sequence that is repeated, often in tandem, within a genome or chromosome. In some embodiments a duplication can comprise a copy of one or more entire chromosomes, a segment of a chromosome, an allele, a gene, an intron, an exon, any non-coding region, any coding region, segment thereof or combination thereof. A duplication can comprise a microduplication. A duplication sometimes comprises one or more copies of a duplicated nucleic acid. A duplication sometimes is characterized as a genetic region repeated one or more times (e.g., repeated 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 times). Duplications can range from small regions (thousands of base pairs) to whole chromosomes in some instances. Duplications frequently occur as the result of an error in homologous recombination or due to a retrotransposon event. Duplications have been associated with certain types of proliferative diseases. Duplications can be characterized using genomic microarrays or comparative genetic hybridization (CGH).

A genetic variation is sometimes an insertion. An insertion is sometimes the addition of one or more nucleotide base pairs into a nucleic acid sequence. An insertion is sometimes a microinsertion. In certain embodiments an insertion comprises the addition of a segment of a chromosome into a genome, chromosome, or segment thereof. In certain embodiments an insertion comprises the addition of an allele, a gene, an intron, an exon, any non-coding region, any coding region, segment thereof or combination thereof into a genome or segment thereof. In certain embodiments an insertion comprises the addition (i.e., insertion) of nucleic acid of unknown origin into a genome, chromosome, or segment thereof. In certain embodiments an insertion comprises the addition (i.e. insertion) of a single base.

As used herein a “copy number variation” generally is a class or type of genetic variation or chromosomal aberration. As used herein a “copy number variation” is one or more copy number variations. A copy number variation can be a deletion (e.g., micro-deletion), duplication (e.g., a micro-duplication) or insertion (e.g., a micro-insertion). Often, the prefix “micro” as used herein sometimes is a segment of nucleic acid less than 5 Mb in length. A copy number variation can

include one or more deletions (e.g., micro-deletion), duplications and/or insertions (e.g., a micro-duplication, micro-insertion) of a segment of a chromosome. In certain embodiments a duplication comprises an insertion. In certain embodiments an insertion is a duplication. In certain embodiments an insertion is not a duplication. For example, often a duplication of a sequence in a genomic section increases the counts for a genomic section in which the duplication is found. Often a duplication of a sequence in a genomic section increases the elevation. In certain embodiments, a duplication present in genomic sections making up a first elevation increases the elevation relative to a second elevation where a duplication is absent. In certain embodiments an insertion increases the counts of a genomic section and a sequence representing the insertion is present (i.e., duplicated) at another location within the same genomic section. In certain embodiments an insertion does not significantly increase the counts of a genomic section or elevation and the sequence that is inserted is not a duplication of a sequence within the same genomic section. In certain embodiments an insertion is not detected or represented as a duplication and a duplicate sequence representing the insertion is not present in the same genomic section.

In some embodiments a copy number variation is a fetal copy number variation. Often, a fetal copy number variation is a copy number variation in the genome of a fetus. In some embodiments a copy number variation is a maternal and/or fetal copy number variation. In certain embodiments a maternal and/or fetal copy number variation is a copy number variation within the genome of a pregnant female (e.g., a female subject bearing a fetus), a female subject that gave birth or a female capable of bearing a fetus. A copy number variation can be a heterozygous copy number variation where the variation (e.g., a duplication or deletion) is present on one allele of a genome. A copy number variation can be a homozygous copy number variation where the variation is present on both alleles of a genome. In some embodiments a copy number variation is a heterozygous or homozygous fetal copy number variation. In some embodiments a copy number variation is a heterozygous or homozygous maternal and/or fetal copy number variation. A copy number variation sometimes is present in a maternal genome and a fetal genome, a maternal genome and not a fetal genome, or a fetal genome and not a maternal genome.

"Ploidy" refers to the number of chromosomes present in a fetus or mother. In certain embodiments "Ploidy" is the same as "chromosome ploidy". In humans, for example, autosomal chromosomes are often present in pairs. For example, in the absence of a genetic variation, most humans have two of each autosomal chromosome (e.g., chromosomes 1-22). The presence of the

normal complement of 2 autosomal chromosomes in a human is often referred to as euploid. "Microploidy" is similar in meaning to ploidy. "Microploidy" often refers to the ploidy of a segment of a chromosome. The term "microploidy" sometimes refers to the presence or absence of a copy number variation (e.g., a deletion, duplication and/or an insertion) within a chromosome (e.g., a homozygous or heterozygous deletion, duplication, or insertion, the like or absence thereof). "Ploidy" and "microploidy" sometimes are determined after normalization of counts of an elevation in a profile (e.g., after normalizing counts of an elevation to an NRV of 1). Thus, an elevation representing an autosomal chromosome pair (e.g., a euploid) is often normalized to an NRV of 1 and is referred to as a ploidy of 1. Similarly, an elevation within a segment of a chromosome representing the absence of a duplication, deletion or insertion is often normalized to an NRV of 1 and is referred to as a microploidy of 1. Ploidy and microploidy are often bin-specific (e.g., genomic section specific) and sample-specific. Ploidy is often defined as integral multiples of ½, with the values of 1, ½, 0, 3/2, and 2 representing euploid (e.g., euploidy, e.g., 2 chromosomes), 1 chromosome present (e.g., a chromosome deletion), no chromosome present, 3 chromosomes (e.g., a trisomy) and 4 chromosomes, respectively. Likewise, microploidy is often defined as integral multiples of ½, with the values of 1, ½, 0, 3/2, and 2 representing euploidy (e.g., no copy number variation), a heterozygous deletion, homozygous deletion, heterozygous duplication and homozygous duplication, respectively. Some examples of ploidy values for a fetus are provided in Table 2 for an NRV of 1.

In certain embodiments the microploidy of a fetus matches the microploidy of the mother of the fetus (i.e., the pregnant female subject). In certain embodiments the microploidy of a fetus matches the microploidy of the mother of the fetus and both the mother and fetus carry the same heterozygous copy number variation, homozygous copy number variation or both are euploid. In certain embodiments the microploidy of a fetus is different than the microploidy of the mother of the fetus. For example, sometimes the microploidy of a fetus is heterozygous for a copy number variation, the mother is homozygous for a copy number variation and the microploidy of the fetus does not match (e.g., does not equal) the microploidy of the mother for the specified copy number variation.

A microploidy is often associated with an expected elevation. For example, sometimes an elevation (e.g., an elevation in a profile, sometimes an elevation that includes substantially no copy number variation) is normalized to an NRV of 1 and the microploidy of a homozygous duplication is

2, a heterozygous duplication is 1.5, a heterozygous deletion is 0.5 and a homozygous deletion is zero.

A genetic variation for which the presence or absence is identified for a subject is associated with a medical condition in certain embodiments. Thus, technology described herein can be used to identify the presence or absence of one or more genetic variations that are associated with a medical condition or medical state. Non-limiting examples of medical conditions include those associated with intellectual disability (e.g., Down Syndrome), aberrant cell-proliferation (e.g., cancer), presence of a micro-organism nucleic acid (e.g., virus, bacterium, fungus, yeast), and preeclampsia.

Non-limiting examples of genetic variations, medical conditions and states are described hereafter.

Chromosome Abnormalities

In some embodiments, the presence or absence of a fetal chromosome abnormality can be determined by using a method or apparatus described herein. In some embodiments, the presence or absence of a genetic variation is determined by methods described herein. A genetic variation refers to a genetic abnormality or difference, non-limiting examples of which include a copy number variation, a mutation (e.g., a mutation of one or more, and/or a series of nucleic acid bases), a deletion (e.g., a microdeletion, a deletion of a chromosome or part thereof), an insertion (e.g., a microinsertion, insertion of heterologous DNA, e.g., viral DNA), a duplication (e.g., microduplication, a duplication of a chromosome or part thereof), a chromosome abnormality, the like or combinations thereof. In some embodiments a genetic variation is a fetal genetic variation, a maternal genetic variation, a tissue genetic variation (e.g., a genetic variation of placental tissue, a genetic variation in placenta), the like or a combination thereof. In some embodiments a fetal chromosome abnormality is determined by using a method or apparatus described herein. A chromosome abnormality is sometimes a chromosome aneuploidy. Chromosome abnormalities include, without limitation, a gain or loss of an entire chromosome or a region of a chromosome comprising one or more genes (e.g., a chromosome aneuploidy). A double chromosome aneuploidy means a chromosome aneuploidy of two different chromosomes (e.g., a trisomy 21 and a trisomy 9). In some embodiments a double chromosome aneuploidy is a chromosome aneuploidy of two different chromosome in the same genome (e.g., a genome of a fetus). In some embodiments a double chromosome aneuploidy is a chromosome aneuploidy of two different

chromosome in different genomes (e.g., a genome of a fetus and a mother). In some embodiments a double chromosome aneuploidy is determined for the same sample (e.g., a sample from a pregnant female). For example, a double chromosome aneuploidy is sometimes determined in the same sample where one aneuploidy is present in the genome of a fetus and the other aneuploidy is present in the mother, placenta and/or the fetus. Chromosome abnormalities include monosomies, trisomies, polysomies, loss of heterozygosity, deletions and/or duplications of one or more nucleotide sequences (e.g., one or more genes), including deletions and duplications caused by unbalanced translocations. The term "chromosomal abnormality", “aneuploid” and “aneuploidy” as used herein refers to a deviation between the structure of the subject chromosome and a normal homologous chromosome. The terms "aneuploidy" and "aneuploid" as used herein often refers to an abnormal number of chromosomes in cells of an organism. The term "normal" refers to the predominate karyotype or banding pattern found in healthy individuals of a particular species, for example, a euploid genome (in humans, 46XX or 46XY). As different organisms have widely varying chromosome complements, the term "aneuploidy" does not refer to a particular number of chromosomes, but rather to the situation in which the chromosome content within a given cell or cells of an organism is abnormal. In some embodiments, the term “aneuploidy” herein refers to an imbalance of genetic material caused by a loss or gain of a whole chromosome, or part of a chromosome. An “aneuploidy” can refer to one or more deletions and/or insertions (e.g., microdeletions, microduplications, microinsertions) of a segment of a chromosome. The term “euploid”, in some embodiments, refers a normal complement of chromosomes. A chromosome aneuploidy is sometimes referred to as a chromosome abnormality.

The term "monosomy" as used herein refers to lack of one chromosome of the normal complement. Partial monosomy can occur in unbalanced translocations or deletions, in which only a segment of the chromosome is present in a single copy. Monosomy of sex chromosomes (45, X) causes Turner syndrome, for example. The term "disomy" refers to the presence of two copies of a chromosome. For organisms such as humans that have two copies of each chromosome (those that are diploid or "euploid"), disomy is the normal condition. For organisms that normally have three or more copies of each chromosome (those that are triploid or above), disomy is an aneuploid chromosome state. In uniparental disomy, both copies of a chromosome come from the same parent (with no contribution from the other parent).

The term "trisomy" as used herein refers to the presence of three copies, instead of two copies, of a particular chromosome. The presence of an extra chromosome 21, which is found in human

Down syndrome, is referred to as "Trisomy 21." Trisomy 18 and Trisomy 13 are two other human autosomal trisomies. Trisomy of sex chromosomes can be seen in females (e.g., 47, XXX in Triple X Syndrome) or males (e.g., 47, XXY in Klinefelter's Syndrome; or 47, XYY in Jacobs Syndrome). In some embodiments, a trisomy is a duplication of most or all of an autosome. In certain embodiments a trisomy is a whole chromosome aneuploidy resulting in three chromosomes.

The terms "tetrasomy" and "pentasomy" as used herein refer to the presence of four or five copies of a chromosome, respectively. Although rarely seen with autosomes, sex chromosome tetrasomy and pentasomy have been reported in humans, including XXXX, XXXY, XXYY, XYYY, XXXXX, XXXXY, XXXYY, XXYYY and XYYYY.

Chromosome abnormalities can be caused by a variety of mechanisms. Mechanisms include, but are not limited to (i) nondisjunction occurring as the result of a weakened mitotic checkpoint, (ii) inactive mitotic checkpoints causing non-disjunction at multiple chromosomes, (iii) merotelic attachment occurring when one kinetochore is attached to both mitotic spindle poles, (iv) a multipolar spindle forming when more than two spindle poles form, (v) a monopolar spindle forming when only a single spindle pole forms, and (vi) a tetraploid intermediate occurring as an end result of the monopolar spindle mechanism.

The terms "partial monosomy" and "partial trisomy" as used herein refer to an imbalance of genetic material caused by loss or gain of part of a chromosome. A partial monosomy or partial trisomy can result from an unbalanced translocation, where an individual carries a derivative chromosome formed through the breakage and fusion of two different chromosomes. In this situation, the individual would have three copies of part of one chromosome (two normal copies and the segment that exists on the derivative chromosome) and only one copy of part of the other chromosome involved in the derivative chromosome.

The term "mosaicism" as used herein refers to aneuploidy in some cells, but not all cells, of an organism. Certain chromosome abnormalities can exist as mosaic and non-mosaic chromosome abnormalities. For example, certain trisomy 21 individuals have mosaic Down syndrome and some have non-mosaic Down syndrome. Different mechanisms can lead to mosaicism. For example, (i) an initial zygote may have three 21st chromosomes, which normally would result in simple trisomy 21, but during the course of cell division one or more cell lines lost one of the 21st chromosomes; and (ii) an initial zygote may have two 21st chromosomes, but during the course of cell division one

of the 21st chromosomes were duplicated. Somatic mosaicism likely occurs through mechanisms distinct from those typically associated with genetic syndromes involving complete or mosaic aneuploidy. Somatic mosaicism has been identified in certain types of cancers and in neurons, for example. In certain instances, trisomy 12 has been identified in chronic lymphocytic leukemia (CLL) and trisomy 8 has been identified in acute myeloid leukemia (AML). Also, genetic syndromes in which an individual is predisposed to breakage of chromosomes (chromosome instability syndromes) are frequently associated with increased risk for various types of cancer, thus highlighting the role of somatic aneuploidy in carcinogenesis. Methods and protocols described herein can identify presence or absence of non-mosaic and mosaic chromosome abnormalities.

Tables 1A-1G present a non-limiting list of chromosome conditions, syndromes and/or abnormalities (e.g., chromosome aneuploidies) that can be potentially identified by methods and apparatus described herein. Table 1B is from the DECIPHER database as of October 6, 2011 (e.g., version 5.1, based on positions mapped to GRCh37; available at uniform resource locator (URL) dechipher.sanger.ac.uk).

TABLE 1A

Chromosome	Abnormality	Disease Association
X	XO	Turner's Syndrome
Y	XXY	Klinefelter syndrome
Y	XYY	Double Y syndrome
Y	XXX	Trisomy X syndrome
Y	XXXX	Four X syndrome
Y	Xp21 deletion	Duchenne's/Becker syndrome, congenital adrenal hypoplasia, chronic granulomatus disease
Y	Xp22 deletion	steroid sulfatase deficiency
Y	Xq26 deletion	X-linked lymphoproliferative disease
1	1p (somatic) monosomy trisomy	neuroblastoma
2	monosomy trisomy 2q	growth retardation, developmental and mental delay, and minor physical abnormalities
3	monosomy trisomy (somatic)	Non-Hodgkin's lymphoma

Chromosome	Abnormality	Disease Association
4	monosomy trisomy (somatic)	Acute non lymphocytic leukemia (ANLL)
5	5p	Cri du chat; Lejeune syndrome
5	5q (somatic) monosomy trisomy	myelodysplastic syndrome
6	monosomy trisomy (somatic)	clear-cell sarcoma
7	7q11.23 deletion	William's syndrome
7	monosomy trisomy	monosomy 7 syndrome of childhood; somatic: renal cortical adenomas; myelodysplastic syndrome
8	8q24.1 deletion	Langer-Giedon syndrome
8	monosomy trisomy	myelodysplastic syndrome; Warkany syndrome; somatic: chronic myelogenous leukemia
9	monosomy 9p	Alfi's syndrome
9	monosomy 9p partial trisomy	Rethore syndrome
9	trisomy	complete trisomy 9 syndrome; mosaic trisomy 9 syndrome
10	Monosomy trisomy (somatic)	ALL or ANLL
11	11p-	Aniridia; Wilms tumor
11	11q-	Jacobsen Syndrome
11	monosomy (somatic) trisomy	myeloid lineages affected (ANLL, MDS)
12	monosomy trisomy (somatic)	CLL, Juvenile granulosa cell tumor (JGCT)
13	13q-	13q-syndrome; Orbeli syndrome
13	13q14 deletion	retinoblastoma
13	monosomy trisomy	Patau's syndrome
14	monosomy trisomy (somatic)	myeloid disorders (MDS, ANLL, atypical CML)
15	15q11-q13 deletion monosomy	Prader-Willi, Angelman's syndrome

Chromosome	Abnormality	Disease Association
15	trisomy (somatic)	myeloid and lymphoid lineages affected, e.g., MDS, ANLL, ALL, CLL)
16	16q13.3 deletion	Rubenstein-Taybi
3	monosomy trisomy (somatic)	papillary renal cell carcinomas (malignant)
17	17p-(somatic)	17p syndrome in myeloid malignancies
17	17q11.2 deletion	Smith-Magenis
17	17q13.3	Miller-Dieker
17	monosomy trisomy (somatic)	renal cortical adenomas
17	17p11.2-12 trisomy	Charcot-Marie Tooth Syndrome type 1; HNPP
18	18p-	18p partial monosomy syndrome or Grouchy Lamy Thieffry syndrome
18	18q-	Grouchy Lamy Salmon Landry Syndrome
18	monosomy trisomy	Edwards Syndrome
19	monosomy trisomy	
20	20p-	trisomy 20p syndrome
20	20p11.2-12 deletion	Alagille
20	20q-	somatic: MDS, ANLL, polycythemia vera, chronic neutrophilic leukemia
20	monosomy trisomy (somatic)	papillary renal cell carcinomas (malignant)
21	monosomy trisomy	Down's syndrome
22	22q11.2 deletion	DiGeorge's syndrome, velocardiofacial syndrome, conotruncal anomaly face syndrome, autosomal dominant Opitz G/BBB syndrome, Caylor cardiofacial syndrome
22	monosomy trisomy	complete trisomy 22 syndrome

TABLE 1B

Syndrome	Chromosome	Start	End	Interval (Mb)	Grade
12q14 microdeletion syndrome	12	65,071,919	68,645,525	3.57	
15q13.3 microdeletion syndrome	15	30,769,995	32,701,482	1.93	
15q24 recurrent microdeletion syndrome	15	74,377,174	76,162,277	1.79	
15q26 overgrowth syndrome	15	99,357,970	102,521,392	3.16	
16p11.2 microduplication syndrome	16	29,501,198	30,202,572	0.70	
16p11.2-p12.2 microdeletion syndrome	16	21,613,956	29,042,192	7.43	
16p13.11 recurrent microdeletion (neurocognitive disorder susceptibility locus)	16	15,504,454	16,284,248	0.78	
16p13.11 recurrent microduplication (neurocognitive disorder susceptibility locus)	16	15,504,454	16,284,248	0.78	
17q21.3 recurrent microdeletion syndrome	17	43,632,466	44,210,205	0.58	1
1p36 microdeletion syndrome	1	10,001	5,408,761	5.40	1
1q21.1 recurrent microdeletion (susceptibility locus for neurodevelopmental disorders)	1	146,512,930	147,737,500	1.22	3
1q21.1 recurrent microduplication (possible susceptibility locus for neurodevelopmental disorders)	1	146,512,930	147,737,500	1.22	3
1q21.1 susceptibility locus for	1	145,401,253	145,928,123	0.53	3

Syndrome	Chromosome	Start	End	Interval (Mb)	Grade
Thrombocytopenia-Absent Radius (TAR) syndrome					
22q11 deletion syndrome (Velocardiofacial / DiGeorge syndrome)	22	18,546,349	22,336,469	3.79	1
22q11 duplication syndrome	22	18,546,349	22,336,469	3.79	3
22q11.2 distal deletion syndrome	22	22,115,848	23,696,229	1.58	
22q13 deletion syndrome (Phelan-Mcdermid syndrome)	22	51,045,516	51,187,844	0.14	1
2p15-16.1 microdeletion syndrome	2	57,741,796	61,738,334	4.00	
2q33.1 deletion syndrome	2	196,925,089	205,206,940	8.28	1
2q37 monosomy	2	239,954,693	243,102,476	3.15	1
3q29 microdeletion syndrome	3	195,672,229	197,497,869	1.83	
3q29 microduplication syndrome	3	195,672,229	197,497,869	1.83	
7q11.23 duplication syndrome	7	72,332,743	74,616,901	2.28	
8p23.1 deletion syndrome	8	8,119,295	11,765,719	3.65	
9q subtelomeric deletion syndrome	9	140,403,363	141,153,431	0.75	1
Adult-onset autosomal dominant leukodystrophy (ADLD)	5	126,063,045	126,204,952	0.14	
Angelman syndrome (Type 1)	15	22,876,632	28,557,186	5.68	1
Angelman syndrome (Type 2)	15	23,758,390	28,557,186	4.80	1
ATR-16 syndrome	16	60,001	834,372	0.77	1
AZFa	Y	14,352,761	15,154,862	0.80	
AZFb	Y	20,118,045	26,065,197	5.95	
AZFb+AZFc	Y	19,964,826	27,793,830	7.83	
AZFc	Y	24,977,425	28,033,929	3.06	
Cat-Eye Syndrome	22	1	16,971,860	16.97	

Syndrome	Chromosome	Start	End	Interval (Mb)	Grade
(Type I)					
Charcot-Marie-Tooth syndrome type 1A (CMT1A)	17	13,968,607	15,434,038	1.47	1
Cri du Chat Syndrome (5p deletion)	5	10,001	11,723,854	11.71	1
Early-onset Alzheimer disease with cerebral amyloid angiopathy	21	27,037,956	27,548,479	0.51	
Familial Adenomatous Polyposis	5	112,101,596	112,221,377	0.12	
Hereditary Liability to Pressure Palsies (HNPP)	17	13,968,607	15,434,038	1.47	1
Leri-Weill dyschondrostosis (LWD) - SHOX deletion	X	751,878	867,875	0.12	
Leri-Weill dyschondrostosis (LWD) - SHOX deletion	X	460,558	753,877	0.29	
Miller-Dieker syndrome (MDS)	17	1	2,545,429	2.55	1
NF1-microdeletion syndrome	17	29,162,822	30,218,667	1.06	1
Pelizaeus-Merzbacher disease	X	102,642,051	103,131,767	0.49	
Potocki-Lupski syndrome (17p11.2 duplication syndrome)	17	16,706,021	20,482,061	3.78	
Potocki-Shaffer syndrome	11	43,985,277	46,064,560	2.08	1
Prader-Willi syndrome (Type 1)	15	22,876,632	28,557,186	5.68	1
Prader-Willi Syndrome (Type 2)	15	23,758,390	28,557,186	4.80	1
RCAD (renal cysts and diabetes)	17	34,907,366	36,076,803	1.17	
Rubinstein-Taybi Syndrome	16	3,781,464	3,861,246	0.08	1
Smith-Magenis Syndrome	17	16,706,021	20,482,061	3.78	1

Syndrome	Chromosome	Start	End	Interval (Mb)	Grade
Sotos syndrome	5	175,130,402	177,456,545	2.33	1
Split hand/foot malformation 1 (SHFM1)	7	95,533,860	96,779,486	1.25	
Steroid sulphatase deficiency (STS)	X	6,441,957	8,167,697	1.73	
WAGR 11p13 deletion syndrome	11	31,803,509	32,510,988	0.71	
Williams-Beuren Syndrome (WBS)	7	72,332,743	74,616,901	2.28	1
Wolf-Hirschhorn Syndrome	4	10,001	2,073,670	2.06	1
Xq28 (MECP2) duplication	X	152,749,900	153,390,999	0.64	

Grade 1 conditions often have one or more of the following characteristics; pathogenic anomaly; strong agreement amongst geneticists; highly penetrant; may still have variable phenotype but some common features; all cases in the literature have a clinical phenotype; no cases of healthy individuals with the anomaly; not reported on DVG databases or found in healthy population; functional data confirming single gene or multi-gene dosage effect; confirmed or strong candidate genes; clinical management implications defined; known cancer risk with implication for surveillance; multiple sources of information (OMIM, Genereviews, Orphanet, Unique, Wikipedia); and/or available for diagnostic use (reproductive counseling).

Grade 2 conditions often have one or more of the following characteristics; likely pathogenic anomaly; highly penetrant; variable phenotype with no consistent features other than DD; small number of cases/ reports in the literature; all reported cases have a clinical phenotype; no functional data or confirmed pathogenic genes; multiple sources of information (OMIM, Genereviews, Orphanet, Unique, Wikipedia); and/or may be used for diagnostic purposes and reproductive counseling.

Grade 3 conditions often have one or more of the following characteristics; susceptibility locus; healthy individuals or unaffected parents of a proband described; present in control populations; non penetrant; phenotype mild and not specific; features less consistent; no functional data or confirmed pathogenic genes; more limited sources of data; possibility of second diagnosis remains a possibility for cases deviating from the majority or if novel clinical finding present; and/or caution when using for diagnostic purposes and guarded advice for reproductive counseling.

Chromosome abnormalities (e.g., a chromosome aneuploidy) include microduplications, mircoinsertions and/or microdeletions of all or a portion of a chromosome. Tables 1E-1G present a non-limiting list of genes and disorder related to microdeletions and microduplications.

Microdeletions in chromosome X (e.g., as shown in FIG. 184) include clinically relevant conditions such as, Ornithine Transcarbamylase Polymorphism, Syndromic X-Linked Mental Retardation type 7 and 12, X-Linked Mental Retardation type 9, 14, 20, 45, 78, 81, 84, and 89, X-Linked Nonspecific Mental Retardation Type 50, X-Linked Exudative Vitreoretinopathy 2, X-Linked Spinal Muscular Atrophy 2, Properdin Deficiency Type I, X-Linked Epilepsy with Variable Learning Disabilities and Behavior Disorders, Leber Optic Atrophy, X-Linked Optic Atrophy, Susceptibility to X-Linked Diabetes Mellitus, Susceptibility to Graves Disease, Congenital X-Linked Nystagmus 5, X-Linked Angioma Serptiginosum, and the entire Xp11.3 Deletion Syndrome. Table 1C below list additional examples of genes and disorders related to microdeletions in Chromosome X.

TABLE 1C

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
314850	Kell blood group precursor	XK, MCLDS	•McLeod syndrome with or without chronic granulomatous disease (phenotype 300842 (3))
300481	Cytochrome b-245, beta polypeptide	CYBB, CGD, AMCBX2	•Atypical mycobacteriosis, familial, X-linked 2 (phenotype 300645 (3)) •Chronic granulomatous disease, X-linked (phenotype 306400 (3))
312610	Retinitis pigmentosa GTPase regulator	RPGR, RP3, CRD, RP15, COD1, CORDX1	•Cone-rod dystrophy-1 (phenotype 304020 (3)) •Macular degeneration, X-linked atrophic (phenotype 300834 (3)) •Retinitis pigmentosa 3 (phenotype 300029 (3)) •Retinitis pigmentosa, X-linked, and sinorespiratory infections, with or without deafness (phenotype 300455 (3))
300461	Ornithine transcarbamylase	OTC	•Ornithine transcarbamylase deficiency (phenotype 311250 (3))
300096	Tetraspanin 7	TSPAN7,	•Mental retardation, X-linked

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
		TM4SF2, MXS1, A15, MRX58	58 (phenotype 300210 (3))
300485	BCL6 corepressor	BCOR, KIAA1575, MCOPS2, MAA2, ANOP2	•Microphthalmia, syndromic 2 (phenotype 300166 (3))
300556	ATPase, H+ transporting, lysosomal, accessory protein 2	ATP6AP2, ATP6M8-9, XMRE, MRXE	•Mental retardation, X-linked, with epilepsy (phenotype 300423 (3))
300278	Nyctalopin	NYX, CSNB1A, NBM1	•Night blindness, congenital stationary, X-linked, type 1A (phenotype 310500 (3))
300172	Calcium/calmodulin-dependent serine protein kinase	CASK, MICPCH, FGS4, CMG, MRXSNA	•FG syndrome 4 (phenotype 300422 (3)) •Mental retardation and microcephaly with pontine and cerebellar hypoplasia (phenotype 300749 (3)) •Mental retardation, with or without nystagmus (phenotype 300422 (3))
309850	Monoamine oxidase A	MAOA	•Brunner syndrome (phenotype 300615 (3))
300658	Norrin	NDP, ND	•Exudative vitreoretinopathy, X-linked (phenotype 305390 (3)) •Norrie disease (phenotype 310600 (3))
300128	Lysine (K)-specific demethylase 6A	KDM6A, UTX, KABUK2	•Kabuki syndrome 2 (phenotype 300867 (3))
300573	Zinc finger protein 674	ZNF674, MRX92	•Mental retardation, X-linked 92 (phenotype 300851 (3))
300757	Retinitis pigmentosa-2, X-linked recessive	RP2	•Retinitis pigmentosa 2 (phenotype 312600 (3))
300080	RNA-binding motif protein 10	RBM10, DXS8237E, KIAA0122, TARPS	•TARP syndrome (phenotype 311900 (3))
314370	Ubiquitin-like modifier-activating enzyme 1	UBA1, UBE1, GXP1, A1ST, SMAX2, AMCX1	•Spinal muscular atrophy, X-linked 2, infantile (phenotype 301830 (3))

Table 1D below list examples of genes and disorders related to microdeletions in Chromosome 7.

5

TABLE 1D

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
610456	Sterile alpha motif domain-containing protein 9	SAMD9, NFTC	•Tumoral calcinosis, familial, normophosphatemic (phenotype 610455 (3))
114131	Calcitonin receptor	CALCR, CRT	•{Osteoporosis, postmenopausal, susceptibility} (phenotype 166710 (3))
120160	Collagen I, alpha-2 polypeptide	COL1A2	•Ehlers-Danlos syndrome, cardiac valvular form (phenotype 225320 (3)) •Ehlers-Danlos syndrome, type VIIB (phenotype 130060 (3)) •Osteogenesis imperfecta, type II (phenotype 166210 (3)) •Osteogenesis imperfecta, type III (phenotype 259420 (3)) •Osteogenesis imperfecta, type IV (phenotype 166220 (3)) •{Osteoporosis, postmenopausal} (phenotype 166710 (3))
604149	Sarcoglycan, epsilon	SGCE, DYT11	•Dystonia-11, myoclonic (phenotype 159900 (3))
168820	Paraoxonase-1	PON1, PON, ESA, MVCD5	•{Coronary artery disease, susceptibility to} (3) •{Coronary artery spasm 2, susceptibility to (3) •{Microvascular complications of diabetes 5} (phenotype 612633 (3)) •{Organophosphate poisoning, sensitivity to} (3)
602447	Paraoxonase-2	PON2	•{Coronary artery disease, susceptibility to} (3)
603859	Solute carrier family 25 (mitochondrial carrier, citrin), member 13	SLC25A13, CTLN2	•Citrullinemia, adult-onset type II (phenotype 603471 (3)) •Citrullinemia, type II, neonatal-onset (phenotype 605814 (3))
600028	Distal-less homeo box-5	DLX5, SHFM1D	•Split-hand/foot malformation 1 with sensorineural hearing loss (phenotype 220600 (3))
605325	Cytochrome P450, subfamily IIIA, polypeptide 5	CYP3A5, P450PCN3	•{Hypertension, salt-sensitive essential, susceptibility to} (phenotype 145500 (3))
602296	Adaptor-related protein complex 4,	AP4M1, SPG50, CPSQ3	•Spastic paraplegia 50, autosomal recessive

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
	mu-1 subunit		(phenotype 612936 (3))
604720	Transferrin receptor 2	TFR2, HFE3	•Hemochromatosis, type 3 (phenotype 604250 (3))
133170	Erythropoietin	EPO, MVCD2	•{Microvascular complications of diabetes 2} (phenotype 612623 (3))
100740	Acetylcholinesterase (YT blood group)	ACHE, YT	•[Blood group, Yt system] (phenotype 112100 (3))
173360	Plasminogen activator inhibitor, type I	PAI1, PLANH1, SERPINE1	•Plasminogen activator inhibitor-1 deficiency (phenotype 613329 (3)) •{Transcription of plasminogen activator inhibitor, modulator of} (3)
603066	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3 (lysyl hydroxylase 3)	PLOD3, LH3	•Lysyl hydroxylase 3 deficiency (phenotype 612394 (3))
604943	Solute carrier family 26, member 5	SLC26A5, PRES	•Deafness, autosomal recessive 61 (phenotype 613865 (3))
600514	Reelin	RELN, RL, LIS2	•Lissencephaly 2 (Norman-Roberts type) (phenotype 257320 (3))
606821	Component of oligomeric golgi complex 5	COG5, GOLTC1, GTC90, CDG2I	•Congenital disorder of glycosylation, type IIi (phenotype 613612 (3))
605646	Solute carrier family 26 (sulfate transporter), member 4	SLC26A4, PDS, DFNB4, EVA, TDH2B	•Deafness, autosomal recessive 4, with enlarged vestibular aqueduct (phenotype 600791 (3)) •Pendred syndrome (phenotype 274600 (3))
126650	Solute carrier family 26 (sulfate transporter), member 3	SLC26A3, DRA, CLD	•Colon cancer (1) •Chloride diarrhea, congenital, Finnish type (phenotype 214700 (3))
238331	Dihydrolipoamide dehydrogenase (E3 component of pyruvate dehydrogenase complex, 2-oxo-glutarate complex)	DLD, LAD, PHE3	•Leigh syndrome (phenotype 256000 (3)) •Maple syrup urine disease, type III (phenotype 248600 (3))

Table 1E below list examples of genes and disorders related to microduplications in Chromosome 18.

TABLE 1E

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
605519	Lipin 2	LPIN2	•Majeed syndrome (phenotype 609628 (3))
602630	TG-interacting factor	TGIF, HPE4	•Holoprosencephaly-4 (phenotype 142946 (3))
600532	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kD	NDUFV2	•Mitochondrial complex I deficiency (phenotype 252010 (3))
607479	Adenomatosis polyposis coli down-regulated 1	APCDD1, HHS, HTS	•Hypotrichosis simplex (phenotype 605389 (3))
604581	ATPase family gene 3-like 2	AFG3L2, SCA28, SPAX5	•Ataxia, spastic, 5, autosomal recessive (phenotype 614487 (3)) •Spinocerebellar ataxia 28 (phenotype 610246 (3))
607397	Melanocortin-2 receptor (ACTH receptor)	MC2R	•Glucocorticoid deficiency, due to ACTH unresponsiveness (phenotype 202200 (3))

Table 1F below list examples of genes and disorders related to microduplications in Chromosome 3.

TABLE 1F

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
609262	Cereblon	CRBN, MRT2	•Mental retardation, autosomal recessive 2 (phenotype 607417 (3))
607939	Sulfatase-modifying factor-1	SUMF1, FGE	•Multiple sulfatase deficiency (phenotype 272200 (3))
147265	Inositol 1,4,5-triphosphate receptor, type 1	ITPR1, SCA15, SCA16	•Spinocerebellar ataxia 15 (phenotype 606658 (3))
601253	Caveolin-3	CAV3, LGMD1C, LQT9	•Cardiomyopathy, familial hypertrophic (phenotype 192600 (3)) •Creatine phosphokinase, elevated serum (phenotype 123320 (3))

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
			•Long QT syndrome-9 (phenotype 611818 (3)) •Muscular dystrophy, limb-girdle, type IC (phenotype 607801 (3)) •Myopathy, distal, Tateyama type (phenotype 614321 (3)) •Rippling muscle disease (phenotype 606072 (3))
611089	Myotubularin-related protein 14	MTMR14, C3orf29, HJUMPY	•{Centronuclear myopathy, autosomal, modifier of} (phenotype 160150 (3))
601982	8-hydroxyguanine DNA glycosylase	OGG1	•Renal cell carcinoma, clear cell, somatic (phenotype 144700 (3))
607170	Cysteine-rich protein with EGF-like domains 1	CRELD1, AVSD2	•Atrioventricular septal defect, partial, with heterotaxy syndrome (phenotype 606217 (3)) •{Atrioventricular septal defect, susceptibility to, 2} (phenotype 606217 (3))
613984	Fanconi anemia, complementation group D2	FANCD2, FANCD, FACD, FAD	•Fanconi anemia, complementation group D2 (phenotype 227646 (3))
608537	VHL gene	VHL	•Erythrocytosis, familial, 2 (phenotype 263400 (3)) •Hemangioblastoma, cerebellar, somatic (3) •Pheochromocytoma (phenotype 171300 (3)) •Renal cell carcinoma, somatic (phenotype 144700 (3)) •von Hippel-Lindau syndrome (phenotype 193300 (3))
605353	Ghrelin	GHRL	•{Obesity, susceptibility to} (phenotype 601665 (3))
108733	ATPase, Ca++ transporting, plasma membrane, 2	ATP2B2, PMCA2	•{Deafness, autosomal recessive 12, modifier of} (phenotype 601386 (3))

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
600755	Synapsin II	SYN2	•{Schizophrenia, susceptibility to} (phenotype 181500 (3))
601487	Peroxisome proliferator activated receptor, gamma	PPARG, PPARG1, PPARG2, CIMT1, GLM1	•Carotid intimal medial thickness 1 (phenotype 609338 (3)) •Insulin resistance, severe, digenic (phenotype 604367 (3)) •Lipodystrophy, familial partial, type 3 (phenotype 604367 (3)) •Obesity, severe (phenotype 601665 (3)) •[Obesity, resistance to] (3) •{Diabetes, type 2} (phenotype 125853 (3)) •{Glioblastoma, susceptibility to} (phenotype 137800 (3))
608753	tRNA splicing endonuclease 2, S. cerevisiae, homolog of	TSEN2, SEN2, PCH2B	•Pontocerebellar hypoplasia type 2B (phenotype 612389 (3))
164760	Oncogene RAF1	RAF1, CRAF, NS5	•LEOPARD syndrome 2 (phenotype 611554 (3)) •Noonan syndrome 5 (phenotype 611553 (3))
601570	Wingless-type MMTV integration site family, member 7A	WNT7A	•Fuhrmann syndrome (phenotype 228930 (3)) •Ulna and fibula, absence of, with sever limb deficiency (phenotype 276820 (3))
612048	Transmembrane protein 43	TMEM43, ARVD5, ARVC5	•Arrhythmogenic right ventricular dysplasia 5 (phenotype 604400 (3))
613208	XPC gene	XPC, XPCC	•Xeroderma pigmentosum, group C (phenotype 278720 (3))
603033	Collagenic tail of endplate acetylcholinesterase	COLQ, EAD	•Endplate acetylcholinesterase deficiency (phenotype 603034 (3))
609019	Biotinidase	BTD	•Biotinidase deficiency (phenotype 253260 (3))
601486	Deleted in azoospermia-like	DAZL, DAZH, SPGYLA	•{Spermatogenic failure, susceptibility to} (3)
190160	Thyroid hormone	THRB, ERBA2, THR1,	•Thyroid hormone

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
	receptor, beta (avian erythroblastic leukemia viral (v-erb-a) oncogene homolog-2)	PRTH	resistance (phenotype 188570 (3)) •Thyroid hormone resistance, autosomal recessive (phenotype 274300 (3)) •Thyroid hormone resistance, selective pituitary (phenotype 145650 (3))
190182	Transforming growth factor, beta receptor II, 70-80kD	TGFBR2, HNPCC6, AAT3, MFS2, LDS1B, LDS2B	•Colorectal cancer, hereditary nonpolyposis, type 6 (phenotype 614331 (3)) •Esophageal cancer, somatic (phenotype 133239 (3)) •Loeys-Dietz syndrome, type 1B (phenotype 610168 (3)) •Loeys-Dietz syndrome, type 2B (phenotype 610380 (3))
611778	Glycerol-3-phosphate dehydrogenase 1-like	GPD1L, KIAA0089	•Brugada syndrome 2 (phenotype 611777 (3))
611458	Galactosidase, beta-1	GLB1, MPS4B	•GM1-gangliosidosis, type I (phenotype 230500 (3)) •GM1-gangliosidosis, type II (phenotype 230600 (3)) •GM1-gangliosidosis, type III (phenotype 230650 (3)) •Mucopolysaccharidosis type IVB (Morquio) (phenotype 253010 (3))
605497	Cartilage-associated protein	CRTAP, CASP, OI7	•Osteogenesis imperfecta, type VII (phenotype 610682 (3))
120436	mutL, E. coli, homolog of, 1	MLH1, COCA2, HNPCC2	•Colorectal cancer, hereditary nonpolyposis, type 2 (phenotype 609310 (3)) •Mismatch repair cancer syndrome (phenotype 276300 (3)) •Muir-Torre syndrome

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
			(phenotype 158320 (3))
602142	Phospholipase C, delta-1	PLCD1, NDNC3	•Nail disorder, nonsyndromic congenital, 3, (leukonychia) (phenotype 151600 (3))
602170	Myeloid differentiation primary response gene 88	MYD88, MYD88D	•Macroglobulinemia, Waldenstrom, somatic (phenotype 153600 (3)) •Pyogenic bacterial infections, recurrent, due to MYD88 deficiency (phenotype 612260 (3))
602730	Activin A receptor, type IIB	ACVR2B, ACTRIIB, HTX4	•Heterotaxy, visceral, 4, autosomal (phenotype 613751 (3))
600163	Sodium channel, voltage-gated, type V, alpha polypeptide	SCN5A, LQT3, VF1, HB1, SSS1, CMD1E, CDCD2	•Atrial fibrillation, familial, 10 (phenotype 614022 (3)) •Brugada syndrome 1 (phenotype 601144 (3)) •Cardiomyopathy, dilated, 1E (phenotype 601154 (3)) •Heart block, nonprogressive (phenotype 113900 (3)) •Heart block, progressive, type IA (phenotype 113900 (3)) •Long QT syndrome-3 (phenotype 603830 (3)) •Sick sinus syndrome 1 (phenotype 608567 (3)) •Ventricular fibrillation, familial, 1 (phenotype 603829 (3)) •{Sudden infant death syndrome, susceptibility to} (phenotype 272120 (3))
601470	Chemokine (C-X3-C) receptor 1 (G protein-coupled receptor-13)	CX3CR1, GPR13, V28	•{Coronary artery disease, resistance to} (phenotype 607339 (3)) •{Macular degeneration, age-related, 12} (phenotype 613784 (3)) •{Rapid progression to

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
			AIDS from HIV1 infection} (phenotype 609423 (3))
610819	Solute carrier family 25, member 38	SLC25A38	•Anemia, sideroblastic, pyridoxine-refractory, autosomal recessive (phenotype 205950 (3))

Table 1G below list examples of genes and disorders related to microduplications in Chromosome 4.

TABLE 1G

OMIM Identifier	Short Description	Gene symbol(s):	Disorder(s)
613228	Aspartylglucosaminidase	AGA	•Aspartylglucosaminuria (phenotype 208400 (3))
103220	Solute carrier family 25 (mitochondrial carrier) member 4 (adenine nucleotide translocator-1, skeletal muscle)	SLC25A4, ANT1, T1, PEO3, PEO2	•Cardiomyopathy, familial hypertrophic (phenotype 192600 (3)) •Progressive external ophthalmoplegia with mitochondrial DNA deletions 3 (phenotype 609283 (3))
603029	Toll-like receptor-3	TLR3, IIAE2	•Herpes simplex encephalitis, susceptibility to, 2 (phenotype 613002 (3))
608614	Cytochrome P450, family 4, subfamily V, poypeptide 2	CYP4V2, BCD	•Bietti crystalline corneoretinal dystrophy (phenotype 210370 (3))
229000	Kallikrein B plasma 1 (Fletcher factor)	KLKB1, KLK3	•Fletcher factor deficiency (phenotype 612423 (3))
264900	Coagulation factor XI (plasma thromboplastin antecedent)	F11	•Factor XI deficiency, autosomal dominant (phenotype 612416 (3)) •Factor XI deficiency, autosomal recessive (phenotype 612416 (3))

Preeclampsia

In some embodiments, the presence or absence of preeclampsia is determined by using a method or apparatus described herein. Preeclampsia is a condition in which hypertension arises in pregnancy (i.e. pregnancy-induced hypertension) and is associated with significant amounts of protein in the urine. In certain embodiments, preeclampsia also is associated with elevated levels of extracellular nucleic acid and/or alterations in methylation patterns. For example, a positive correlation between extracellular fetal-derived hypermethylated RASSF1A levels and the severity of pre-eclampsia has been observed. In certain examples, increased DNA methylation is observed for the H19 gene in preeclamptic placentas compared to normal controls.

Preeclampsia is one of the leading causes of maternal and fetal/neonatal mortality and morbidity worldwide. Circulating cell-free nucleic acids in plasma and serum are novel biomarkers with promising clinical applications in different medical fields, including prenatal diagnosis. Quantitative changes of cell-free fetal (cff)DNA in maternal plasma as an indicator for impending preeclampsia have been reported in different studies, for example, using real-time quantitative PCR for the male-specific SRY or DYS 14 loci. In cases of early onset preeclampsia, elevated levels may be seen in the first trimester. The increased levels of cffDNA before the onset of symptoms may be due to hypoxia/reoxygenation within the intervillous space leading to tissue oxidative stress and increased placental apoptosis and necrosis. In addition to the evidence for increased shedding of cffDNA into the maternal circulation, there is also evidence for reduced renal clearance of cffDNA in preeclampsia. As the amount of fetal DNA is currently determined by quantifying Y-chromosome specific sequences, alternative approaches such as measurement of total cell-free DNA or the use of gender-independent fetal epigenetic markers, such as DNA methylation, offer an alternative. Cell-free RNA of placental origin is another alternative biomarker that may be used for screening and diagnosing preeclampsia in clinical practice. Fetal RNA is associated with subcellular placental particles that protect it from degradation. Fetal RNA levels sometimes are ten-fold higher in pregnant females with preeclampsia compared to controls, and therefore is an alternative biomarker that may be used for screening and diagnosing preeclampsia in clinical practice.

Pathogens

In some embodiments, the presence or absence of a pathogenic condition is determined by a method or apparatus described herein. A pathogenic condition can be caused by infection of a

host by a pathogen including, but not limited to, a bacterium, virus or fungus. Since pathogens typically possess nucleic acid (e.g., genomic DNA, genomic RNA, mRNA) that can be distinguishable from host nucleic acid, methods and apparatus provided herein can be used to determine the presence or absence of a pathogen. Often, pathogens possess nucleic acid with characteristics unique to a particular pathogen such as, for example, epigenetic state and/or one or more sequence variations, duplications and/or deletions. Thus, methods provided herein may be used to identify a particular pathogen or pathogen variant (e.g., strain).

Cancers

In some embodiments, the presence or absence of a cell proliferation disorder (e.g., a cancer) is determined by using a method or apparatus described herein. For example, levels of cell-free nucleic acid in serum can be elevated in patients with various types of cancer compared with healthy patients. Patients with metastatic diseases, for example, can sometimes have serum DNA levels approximately twice as high as non-metastatic patients. Patients with metastatic diseases may also be identified by cancer-specific markers and/or certain single nucleotide polymorphisms or short tandem repeats, for example. Non-limiting examples of cancer types that may be positively correlated with elevated levels of circulating DNA include breast cancer, colorectal cancer, gastrointestinal cancer, hepatocellular cancer, lung cancer, melanoma, non-Hodgkin lymphoma, leukemia, multiple myeloma, bladder cancer, hepatoma, cervical cancer, esophageal cancer, pancreatic cancer, and prostate cancer. Various cancers can possess, and can sometimes release into the bloodstream, nucleic acids with characteristics that are distinguishable from nucleic acids from non-cancerous healthy cells, such as, for example, epigenetic state and/or sequence variations, duplications and/or deletions. Such characteristics can, for example, be specific to a particular type of cancer. Thus, it is further contemplated that a method provided herein can be used to identify a particular type of cancer.

Software can be used to perform one or more steps in the processes described herein, including but not limited to; counting, data processing, generating an outcome, and/or providing one or more recommendations based on generated outcomes, as described in greater detail hereafter.

Machines, Software and Interfaces

Certain processes and methods described herein (e.g., quantifying, mapping, normalizing, range setting, adjusting, categorizing, counting and/or determining sequence reads, counts, elevations (e.g., elevations) and/or profiles) often cannot be performed without a computer, processor, software, module or other apparatus. Methods described herein typically are computer-implemented methods, and one or more portions of a method sometimes are performed by one or more processors. Embodiments pertaining to methods described in this document generally are applicable to the same or related processes implemented by instructions in systems, apparatus and computer program products described herein. In some embodiments, processes and methods described herein (e.g., quantifying, counting and/or determining sequence reads, counts, elevations and/or profiles) are performed by automated methods. In some embodiments, an automated method is embodied in software, modules, processors, peripherals and/or an apparatus comprising the like, that determine sequence reads, counts, mapping, mapped sequence tags, elevations, profiles, normalizations, comparisons, range setting, categorization, adjustments, plotting, outcomes, transformations and identifications. As used herein, software refers to computer readable program instructions that, when executed by a processor, perform computer operations, as described herein.

Sequence reads, counts, elevations, and profiles derived from a test subject (e.g., a patient, a pregnant female) and/or from a reference subject can be further analyzed and processed to determine the presence or absence of a genetic variation. Sequence reads, counts, elevations and/or profiles sometimes are referred to as “data” or “data sets”. In some embodiments, data or data sets can be characterized by one or more features or variables (e.g., sequence based [e.g., GC content, specific nucleotide sequence, the like], function specific [e.g., expressed genes, cancer genes, the like], location based [genome specific, chromosome specific, genomic section or bin specific], the like and combinations thereof). In certain embodiments, data or data sets can be organized into a matrix having two or more dimensions based on one or more features or variables. Data organized into matrices can be organized using any suitable features or variables. A non-limiting example of data in a matrix includes data that is organized by maternal age, maternal ploidy, and fetal contribution. In certain embodiments, data sets characterized by one or more features or variables sometimes are processed after counting.

Apparatuses, software and interfaces may be used to conduct methods described herein. Using apparatuses, software and interfaces, a user may enter, request, query or determine options for using particular information, programs or processes (e.g., mapping sequence reads, processing mapped data and/or providing an outcome), which can involve implementing statistical analysis algorithms, statistical significance algorithms, statistical algorithms, iterative steps, validation algorithms, and graphical representations, for example. In some embodiments, a data set may be entered by a user as input information, a user may download one or more data sets by a suitable hardware media (e.g., flash drive), and/or a user may send a data set from one system to another for subsequent processing and/or providing an outcome (e.g., send sequence read data from a sequencer to a computer system for sequence read mapping; send mapped sequence data to a computer system for processing and yielding an outcome and/or report).

A system typically comprises one or more apparatus. Each apparatus comprises one or more of memory, one or more processors, and instructions. Where a system includes two or more apparatus, some or all of the apparatus may be located at the same location, some or all of the apparatus may be located at different locations, all of the apparatus may be located at one location and/or all of the apparatus may be located at different locations. Where a system includes two or more apparatus, some or all of the apparatus may be located at the same location as a user, some or all of the apparatus may be located at a location different than a user, all of the apparatus may be located at the same location as the user, and/or all of the apparatus may be located at one or more locations different than the user.

Apparatuses (multiple apparatuses, also referred to herein in plural as apparatus), software and interfaces may be used to conduct methods described herein. Using apparatuses, software and interfaces, a user may enter, request, query or determine options for using particular information, programs or processes (e.g., mapping sequence reads, processing mapped data and/or providing an outcome), which can involve implementing statistical analysis algorithms, statistical significance algorithms, statistical algorithms, iterative steps, validation algorithms, and graphical representations, for example. In some embodiments, a data set may be entered by a user as input information, a user may download one or more data sets by a suitable hardware media (e.g., flash drive), and/or a user may send a data set from one system to another for subsequent processing and/or providing an outcome (e.g., send sequence read data from a sequencer to a computer system for sequence read mapping; send mapped sequence data to a computer system for processing and yielding an outcome and/or report).

A system sometimes comprises a computing apparatus or a sequencing apparatus, or a computing apparatus and a sequencing apparatus (i.e., sequencing machine and/or computing machine). A sequencing apparatus generally is configured to receive physical nucleic acid and generate signals corresponding to nucleotide bases of the nucleic acid. A sequencing apparatus is often "loaded" with a sample comprising nucleic acid and the nucleic acid of the sample loaded in the sequencing apparatus generally is subjected to a nucleic acid sequencing process. The term "loading a sequence apparatus" as used herein refers to contacting a portion of a sequencing apparatus (e.g., a flow cell) with a nucleic acid sample, which portion of the sequencing apparatus is configured to receive a sample for conducting a nucleic acid sequencing process. In some embodiments a sequencing apparatus is loaded with a variant of a sample nucleic acid. A variant sometimes is produced by a process that modifies the sample nucleic acid to a form suitable for sequencing the nucleic acid (e.g., by ligation (e.g., adding adaptors to ends of sample nucleic acid by ligation), amplification, restriction digest, the like or combinations thereof). A sequencing apparatus is often configured, in part, to perform a suitable DNA sequencing method that generates signals (e.g., electronic signals, detector signals, images, the like, or combinations thereof) corresponding to nucleotide bases of the loaded nucleic acid.

One or more signals corresponding to each base of a DNA sequence are often processed and/or transformed into base calls (e.g., a specific nucleotide base, e.g., guanine, cytosine, thymine, uracil, adenine, and the like) by a suitable process. A collection of base calls derived from a loaded nucleic acid often are processed and/or assembled into one or more sequence reads. In embodiments in which multiple sample nucleic acids are sequenced at one time (i.e., multiplexing), a suitable de-multiplexing process can be utilized to associated particular reads with the sample nucleic acid from which they originated. Sequence reads can be aligned by a suitable process to a reference genome and reads aligned to portions of the reference genome can be counted, as described herein.

A sequencing apparatus sometimes is associated with and/or comprises one or more computing apparatus in a system. The one or more computing apparatus sometimes are configured to perform one or more of the following processes: generating base calls from sequencing apparatus signals, assembling reads (e.g., generating reads), de-multiplexing reads, aligning reads to a reference genome, counting reads aligned to genomic portions in the reference genome, and the like. The one or more computing apparatus sometimes are configured to perform one or more of

the following additional processes: normalize read counts (e.g., reduce or remove bias), generate one or more determinations (e.g., determine fetal fraction, fetal ploidy, fetal sex, fetal chromosome count, outcome, presence or absence of a genetic variation (e.g., presence or absence of a fetal chromosome aneuploidy (e.g., chromosome 13, 18 and/or 21 trisomy)), and the like.

In some embodiments, one computing apparatus is associated with a sequencing apparatus, and in certain embodiments, the one computing apparatus performs the majority or all of the following processes: generate base calls from sequencing apparatus signals, assemble reads, de-multiplex reads, align reads and count the reads aligned to genomic portions of a reference genome, normalize read counts and generate one or more outcomes (e.g., fetal fraction, presence or absence of a particular genetic variation). In the latter embodiments, in which one computing apparatus is associated with a sequencing apparatus, the computing apparatus often includes one or more processors (e.g., microprocessors) and memory having instructions that are carried out by the one or more processors to perform the processes. In some embodiments, the one computing apparatus can be a single or multi-core computing device local to the sequencing apparatus (e.g., located in the same location (e.g., the same address, the same building, same floor, same room or the like)). In some embodiments the one computing apparatus is integrated with the sequencing apparatus.

In some embodiments, multiple computing apparatus in a system are associated with a sequencing apparatus, and a subset of the total processes performed by the system may be allocated to or divided among particular computing apparatus in the system. Subsets of the total number of processes can be divided among two or more computing apparatus, or groups thereof, in any suitable combination. In certain embodiments, generating base calls from sequencing apparatus signals, assembling reads and de-multiplexing reads are performed by a first computing apparatus or group thereof, aligning and counting reads mapped to portions of a reference genome are performed by a second computing apparatus or group thereof, and normalizing read counts and providing one or more outcomes are performed by a third computing apparatus or group thereof. In systems comprising two or more computing apparatus or groups thereof, each particular computing apparatus may include memory, one or more processors or a combination thereof. A multi-computing apparatus system sometimes includes one or more suitable servers local to a sequencing apparatus, and sometimes includes one or more suitable servers not local to the sequencing apparatus (e.g., web servers, on-line servers, application servers, remote file servers, cloud servers (e.g., cloud environment, cloud computing)).

Apparatus in different system configurations can generate different types of output data. For example, a sequencing apparatus can output base signals and the base signal output data can be transferred to a computing apparatus that converts the base signal data to base calls. In some embodiments, the base calls are output data from one computing apparatus and are transferred to another computing apparatus for generating sequence reads. In certain embodiments, base calls are not output data from a particular apparatus, and instead, are utilized in the same apparatus that received sequencing apparatus base signals to generate sequence reads. In some embodiments, one apparatus receives sequencing apparatus base signals, generates base calls, sequence reads and de-multiplexes sequence reads, and outputs de-multiplexed sequence reads for a sample that can be transferred to another apparatus or group thereof that aligns the sequence reads to a reference genome. In some embodiments, one apparatus or group thereof can output aligned sequence reads mapped to portions of a reference genome (e.g., SAM or BAM files), and such output data can be transferred to a second computing apparatus or group thereof that normalizes the sequence reads (e.g., normalizes the counts of the sequence reads) and generates an outcome (e.g., fetal fraction and/or presence or absence of a fetal trisomy). Output data from one apparatus can be transferred to a second apparatus in any suitable manner. For example, output data from one apparatus sometimes is placed on a physical storage device and the storage device is transported and connected to a second apparatus to which the output data is transferred. Output data sometimes is stored by one apparatus in a database, and a second apparatus accesses the output data from the same database.

In some embodiments a user interacts with an apparatus (e.g., a computing apparatus, a sequencing apparatus). A user may, for example, place a query to software which then may acquire a data set via internet access, and in certain embodiments, a programmable processor may be prompted to acquire a suitable data set based on given parameters. A programmable processor also may prompt a user to select one or more data set options selected by the processor based on given parameters. A programmable processor may prompt a user to select one or more data set options selected by the processor based on information found via the internet, other internal or external information, or the like. Options may be chosen for selecting one or more data feature selections, one or more statistical algorithms, one or more statistical analysis algorithms, one or more statistical significance algorithms, iterative steps, one or more validation algorithms, and one or more graphical representations of methods, apparatuses, or computer programs.

Systems addressed herein may comprise general components of computer systems, such as, for example, network servers, laptop systems, desktop systems, handheld systems, personal digital assistants, computing kiosks, and the like. A computer system may comprise one or more input means such as a keyboard, touch screen, mouse, voice recognition or other means to allow the user to enter data into the system. A system may further comprise one or more outputs, including, but not limited to, a display screen (e.g., CRT or LCD), speaker, FAX machine, printer (e.g., laser, ink jet, impact, black and white or color printer), or other output useful for providing visual, auditory and/or hardcopy output of information (e.g., outcome and/or report).

In a system, input and output means may be connected to a central processing unit which may comprise among other components, a microprocessor for executing program instructions and memory for storing program code and data. In some embodiments, processes may be implemented as a single user system located in a single geographical site. In certain embodiments, processes may be implemented as a multi-user system. In the case of a multi-user implementation, multiple central processing units may be connected by means of a network. The network may be local, encompassing a single department in one portion of a building, an entire building, span multiple buildings, span a region, span an entire country or be worldwide. The network may be private, being owned and controlled by a provider, or it may be implemented as an internet based service where the user accesses a web page to enter and retrieve information. Accordingly, in certain embodiments, a system includes one or more machines, which may be local or remote with respect to a user. More than one machine in one location or multiple locations may be accessed by a user, and data may be mapped and/or processed in series and/or in parallel. Thus, a suitable configuration and control may be utilized for mapping and/or processing data using multiple machines, such as in local network, remote network and/or "cloud" computing platforms.

A system can include a communications interface in some embodiments. A communications interface allows for transfer of software and data between a computer system and one or more external devices. Non-limiting examples of communications interfaces include a modem, a network interface (such as an Ethernet card), a communications port, a PCMCIA slot and card, and the like. Software and data transferred via a communications interface generally are in the form of signals, which can be electronic, electromagnetic, optical and/or other signals capable of being received by a communications interface. Signals often are provided to a communications interface via a channel. A channel often carries signals and can be implemented using wire or cable, fiber

optics, a phone line, a cellular phone link, an RF link and/or other communications channels. Thus, in an example, a communications interface may be used to receive signal information that can be detected by a signal detection module.

Data may be input by a suitable device and/or method, including, but not limited to, manual input devices or direct data entry devices (DDEs). Non-limiting examples of manual devices include keyboards, concept keyboards, touch sensitive screens, light pens, mouse, tracker balls, joysticks, graphic tablets, scanners, digital cameras, video digitizers and voice recognition devices. Non-limiting examples of DDEs include bar code readers, magnetic strip codes, smart cards, magnetic ink character recognition, optical character recognition, optical mark recognition, and turnaround documents.

In some embodiments, output from a sequencing apparatus may serve as data that can be input via an input device. In certain embodiments, mapped sequence reads may serve as data that can be input via an input device. In certain embodiments, simulated data is generated by an in silico process and the simulated data serves as data that can be input via an input device. The term "in silico" refers to research and experiments performed using a computer. In silico processes include, but are not limited to, mapping sequence reads and processing mapped sequence reads according to processes described herein.

A system may include software useful for performing a process described herein, and software can include one or more modules for performing such processes (e.g., sequencing module, logic processing module, data display organization module). The term "software" refers to computer readable program instructions that, when executed by a computer, perform computer operations. Instructions executable by the one or more processors sometimes are provided as executable code, that when executed, can cause one or more processors to implement a method described herein. A module described herein can exist as software, and instructions (e.g., processes, routines, subroutines) embodied in the software can be implemented or performed by a processor. For example, a module (e.g., a software module) can be a part of a program that performs a particular process or task. The term “module” refers to a self-contained functional unit that can be used in a larger apparatus or software system. A module can comprise a set of instructions for carrying out a function of the module. A module can transform data and/or information. Data and/or information can be in a suitable form. For example, data and/or information can be digital or analogue. In certain embodiments, data and/or information can be packets, bytes, characters, or

bits. In some embodiments, data and/or information can be any gathered, assembled or usable data or information. Non-limiting examples of data and/or information include a suitable media, pictures, video, sound (e.g., frequencies, audible or non-audible), numbers, constants, a value, objects, time, functions, instructions, maps, references, sequences, reads, mapped reads, elevations, ranges, thresholds, signals, displays, representations, or transformations thereof. A module can accept or receive data and/or information, transform the data and/or information into a second form, and provide or transfer the second form to an apparatus, peripheral, component or another module. A module can perform one or more of the following non-limiting functions: mapping sequence reads, providing counts, assembling genomic sections, providing or determining an elevation, providing a count profile, normalizing (e.g., normalizing reads, normalizing counts, and the like), providing a normalized count profile or elevations of normalized counts, comparing two or more elevations, providing uncertainty values, providing or determining expected elevations and expected ranges(e.g., expected elevation ranges, threshold ranges and threshold elevations), providing adjustments to elevations (e.g., adjusting a first elevation, adjusting a second elevation, adjusting a profile of a chromosome or a segment thereof, and/or padding), providing identification (e.g., identifying a copy number variation, genetic variation or aneuploidy), categorizing, plotting, and/or determining an outcome, for example. A processor can, in some cases, carry out the instructions in a module. In some embodiments, one or more processors are required to carry out instructions in a module or group of modules. A module can provide data and/or information to another module, apparatus or source and can receive data and/or information from another module, apparatus or source.

A computer program product sometimes is embodied on a tangible computer-readable medium, and sometimes is tangibly embodied on a non-transitory computer-readable medium. A module sometimes is stored on a computer readable medium (e.g., disk, drive) or in memory (e.g., random access memory). A module and processor capable of implementing instructions from a module can be located in an apparatus or in different apparatus. A module and/or processor capable of implementing an instruction for a module can be located in the same location as a user (e.g., local network) or in a different location from a user (e.g., remote network, cloud system). In embodiments in which a method is carried out in conjunction with two or more modules, the modules can be located in the same apparatus, one or more modules can be located in different apparatus in the same physical location, and one or more modules may be located in different apparatus in different physical locations.

An apparatus, in some embodiments, comprises at least one processor for carrying out the instructions in a module. Counts of sequence reads mapped to genomic sections of a reference genome sometimes are accessed by a processor that executes instructions configured to carry out a method described herein. Counts that are accessed by a processor can be within memory of a system, and the counts can be accessed and placed into the memory of the system after they are obtained. In some embodiments, an apparatus includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from a module. In some embodiments, an apparatus includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, an apparatus operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, an apparatus comprises a module. In certain embodiments an apparatus comprises one or more modules. An apparatus comprising a module often can receive and transfer one or more of data and/or information to and from other modules. In certain embodiments, an apparatus comprises peripherals and/or components. In certain embodiments an apparatus can comprise one or more peripherals or components that can transfer data and/or information to and from other modules, peripherals and/or components. In certain embodiments an apparatus interacts with a peripheral and/or component that provides data and/or information. In certain embodiments peripherals and components assist an apparatus in carrying out a function or interact directly with a module. Non-limiting examples of peripherals and/or components include a suitable computer peripheral, I/O or storage method or device including but not limited to scanners, printers, displays (e.g., monitors, LED, LCT or CRTs), cameras, microphones, pads (e.g., ipads, tablets), touch screens, smart phones, mobile phones, USB I/O devices, USB mass storage devices, keyboards, a computer mouse, digital pens, modems, hard drives, jump drives, flash drives, a processor, a server, CDs, DVDs, graphic cards, specialized I/O devices (e.g., sequencers, photo cells, photo multiplier tubes, optical readers, sensors, etc.), one or more flow cells, fluid handling components, network interface controllers, ROM, RAM, wireless transfer methods and devices (Bluetooth, WiFi, and the like,), the world wide web (www), the internet, a computer and/or another module.

One or more of a sequencing module, logic processing module and data display organization module can be utilized in a method described herein. In certain embodiments a logic processing module, sequencing module or data display organization module, or an apparatus comprising one or more such modules, gather, assemble, receive, provide and/or transfer data and/or information to or from another module, apparatus, component, peripheral or operator of an apparatus. For

example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a logic processing module, sequencing module or data display organization module. A logic processing module, sequencing module or data display organization module can receive data and/or information from another module, non-limiting examples of which include a logic processing module, sequencing module, data display organization module, sequencing module, sequencing module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module and/or logic processing module, the like or combination thereof. Data and/or information derived from or transformed by a logic processing module, sequencing module or data display organization module can be transferred from a logic processing module, sequencing module or data display organization module to a sequencing module, sequencing module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module or other suitable apparatus and/or module. A sequencing module can receive data and/or information form a logic processing module and/or sequencing module and transfer data and/or information to a logic processing module and/or a mapping module, for example. In certain embodiments a logic processing module orchestrates, controls, limits, organizes, orders, distributes, partitions, transforms and/or regulates data and/or information or the transfer of data and/or information to and from one or more other modules, peripherals or devices. A data display organization module can receive data and/or information form a logic processing module and/or plotting module and transfer data and/or information to a logic processing module, plotting module, display, peripheral or device. An apparatus comprising a logic processing module, sequencing module or data display organization module can comprise at least one processor. In some embodiments, data and/or information are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the logic processing module, sequencing module and/or data display organization module. In some embodiments, a logic processing module, sequencing module or data display organization module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)).

Software often is provided on a program product containing program instructions recorded on a computer readable medium, including, but not limited to, magnetic media including floppy disks,

hard disks, and magnetic tape; and optical media including CD-ROM discs, DVD discs, magneto-optical discs, flash drives, RAM, floppy discs, the like, and other such media on which the program instructions can be recorded. In online implementation, a server and web site maintained by an organization can be configured to provide software downloads to remote users, or remote users may access a remote system maintained by an organization to remotely access software. Software may obtain or receive input information. Software may include a module that specifically obtains or receives data (e.g., a data receiving module that receives sequence read data and/or mapped read data) and may include a module that specifically processes the data (e.g., a processing module that processes received data (e.g., filters, normalizes, provides an outcome and/or report). The terms "obtaining" and "receiving" input information refers to receiving data (e.g., sequence reads, mapped reads) by computer communication means from a local, or remote site, human data entry, or any other method of receiving data. The input information may be generated in the same location at which it is received, or it may be generated in a different location and transmitted to the receiving location. In some embodiments, input information is modified before it is processed (e.g., placed into a format amenable to processing (e.g., tabulated)).

In some embodiments, provided are computer program products, such as, for example, a computer program product comprising a computer usable medium having a computer readable program code embodied therein, the computer readable program code adapted to be executed to implement a method comprising: (a) obtaining sequence reads of sample nucleic acid from a test subject; (b) mapping the sequence reads obtained in (a) to a known genome, which known genome has been divided into genomic sections; (c) counting the mapped sequence reads within the genomic sections; (d) generating a sample normalized count profile by normalizing the counts for the genomic sections obtained in (c); and (e) determining the presence or absence of a genetic variation from the sample normalized count profile in (d).

Software can include one or more algorithms in certain embodiments. An algorithm may be used for processing data and/or providing an outcome or report according to a finite sequence of instructions. An algorithm often is a list of defined instructions for completing a task. Starting from an initial state, the instructions may describe a computation that proceeds through a defined series of successive states, eventually terminating in a final ending state. The transition from one state to the next is not necessarily deterministic (e.g., some algorithms incorporate randomness). By way of example, and without limitation, an algorithm can be a search algorithm, sorting algorithm, merge algorithm, numerical algorithm, graph algorithm, string algorithm, modeling algorithm, computational genometric algorithm, combinatorial algorithm, machine learning algorithm,

cryptography algorithm, data compression algorithm, parsing algorithm and the like. An algorithm can include one algorithm or two or more algorithms working in combination. An algorithm can be of any suitable complexity class and/or parameterized complexity. An algorithm can be used for calculation and/or data processing, and in some embodiments, can be used in a deterministic or probabilistic/predictive approach. An algorithm can be implemented in a computing environment by use of a suitable programming language, non-limiting examples of which are C, C++, Java, Perl, Python, Fortran, and the like. In some embodiments, an algorithm can be configured or modified to include margin of errors, statistical analysis, statistical significance, and/or comparison to other information or data sets (e.g., applicable when using a neural net or clustering algorithm).

In certain embodiments, several algorithms may be implemented for use in software. These algorithms can be trained with raw data in some embodiments. For each new raw data sample, the trained algorithms may produce a representative processed data set or outcome. A processed data set sometimes is of reduced complexity compared to the parent data set that was processed. Based on a processed set, the performance of a trained algorithm may be assessed based on sensitivity and specificity, in some embodiments. An algorithm with the highest sensitivity and/or specificity may be identified and utilized, in certain embodiments.

In certain embodiments, simulated (or simulation) data can aid data processing, for example, by training an algorithm or testing an algorithm. In some embodiments, simulated data includes hypothetical various samplings of different groupings of sequence reads. Simulated data may be based on what might be expected from a real population or may be skewed to test an algorithm and/or to assign a correct classification. Simulated data also is referred to herein as “virtual” data. Simulations can be performed by a computer program in certain embodiments. One possible step in using a simulated data set is to evaluate the confidence of an identified results, e.g., how well a random sampling matches or best represents the original data. One approach is to calculate a probability value (p-value), which estimates the probability of a random sample having better score than the selected samples. In some embodiments, an empirical model may be assessed, in which it is assumed that at least one sample matches a reference sample (with or without resolved variations). In some embodiments, another distribution, such as a Poisson distribution for example, can be used to define the probability distribution.

A system may include one or more processors in certain embodiments. A processor can be connected to a communication bus. A computer system may include a main memory, often

random access memory (RAM), and can also include a secondary memory. Memory in some embodiments comprises a non-transitory computer-readable storage medium. Secondary memory can include, for example, a hard disk drive and/or a removable storage drive, representing a floppy disk drive, a magnetic tape drive, an optical disk drive, memory card and the like. A removable storage drive often reads from and/or writes to a removable storage unit. Non-limiting examples of removable storage units include a floppy disk, magnetic tape, optical disk, and the like, which can be read by and written to by, for example, a removable storage drive. A removable storage unit can include a computer-usable storage medium having stored therein computer software and/or data.

A processor may implement software in a system. In some embodiments, a processor may be programmed to automatically perform a task described herein that a user could perform. Accordingly, a processor, or algorithm conducted by such a processor, can require little to no supervision or input from a user (e.g., software may be programmed to implement a function automatically). In some embodiments, the complexity of a process is so large that a single person or group of persons could not perform the process in a timeframe short enough for determining the presence or absence of a genetic variation.

In some embodiments, secondary memory may include other similar means for allowing computer programs or other instructions to be loaded into a computer system. For example, a system can include a removable storage unit and an interface device. Non-limiting examples of such systems include a program cartridge and cartridge interface (such as that found in video game devices), a removable memory chip (such as an EPROM, or PROM) and associated socket, and other removable storage units and interfaces that allow software and data to be transferred from the removable storage unit to a computer system.

One entity can generate counts of sequence reads, map the sequence reads to genomic sections, count the mapped reads, and utilize the counted mapped reads in a method, system, apparatus or computer program product described herein, in some embodiments. Counts of sequence reads mapped to genomic sections sometimes are transferred by one entity to a second entity for use by the second entity in a method, system, apparatus or computer program product described herein, in certain embodiments.

In some embodiments, one entity generates sequence reads and a second entity maps those sequence reads to genomic sections in a reference genome in some embodiments. The second entity sometimes counts the mapped reads and utilizes the counted mapped reads in a method, system, apparatus or computer program product described herein. In certain embodiments the second entity transfers the mapped reads to a third entity, and the third entity counts the mapped reads and utilizes the mapped reads in a method, system, apparatus or computer program product described herein. In certain embodiments the second entity counts the mapped reads and transfers the counted mapped reads to a third entity, and the third entity utilizes the counted mapped reads in a method, system, apparatus or computer program product described herein. In embodiments involving a third entity, the third entity sometimes is the same as the first entity. That is, the first entity sometimes transfers sequence reads to a second entity, which second entity can map sequence reads to genomic sections in a reference genome and/or count the mapped reads, and the second entity can transfer the mapped and/or counted reads to a third entity. A third entity sometimes can utilize the mapped and/or counted reads in a method, system, apparatus or computer program product described herein, wherein the third entity sometimes is the same as the first entity, and sometimes the third entity is different from the first or second entity.

In some embodiments, one entity obtains blood from a pregnant female, optionally isolates nucleic acid from the blood (e.g., from the plasma or serum), and transfers the blood or nucleic acid to a second entity that generates sequence reads from the nucleic acid.

FIG. 190 illustrates a non-limiting example of a computing environment 510 in which various systems, methods, algorithms, and data structures described herein may be implemented. The computing environment 510 is only one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of the systems, methods, and data structures described herein. Neither should computing environment 510 be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in computing environment 510. A subset of systems, methods, and data structures shown in FIG. 190 can be utilized in certain embodiments. Systems, methods, and data structures described herein are operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of known computing systems, environments, and/or configurations that may be suitable include, but are not limited to, personal computers, server computers, thin clients, thick clients, hand-held or laptop devices, multiprocessor systems, microprocessor-based systems, set top boxes, programmable consumer electronics,

network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

The operating environment 510 of FIG. 190 includes a general purpose computing device in the form of a computer 520, including a processing unit 521, a system memory 522, and a system bus 523 that operatively couples various system components including the system memory 522 to the processing unit 521. There may be only one or there may be more than one processing unit 521, such that the processor of computer 520 includes a single central-processing unit (CPU), or a plurality of processing units, commonly referred to as a parallel processing environment. The computer 520 may be a conventional computer, a distributed computer, or any other type of computer.

The system bus 523 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. The system memory may also be referred to as simply the memory, and includes read only memory (ROM) 524 and random access memory (RAM). A basic input/output system (BIOS) 526, containing the basic routines that help to transfer information between elements within the computer 520, such as during start-up, is stored in ROM 524. The computer 520 may further include a hard disk drive interface 527 for reading from and writing to a hard disk, not shown, a magnetic disk drive 528 for reading from or writing to a removable magnetic disk 529, and an optical disk drive 530 for reading from or writing to a removable optical disk 531 such as a CD ROM or other optical media.

The hard disk drive 527, magnetic disk drive 528, and optical disk drive 530 are connected to the system bus 523 by a hard disk drive interface 532, a magnetic disk drive interface 533, and an optical disk drive interface 534, respectively. The drives and their associated computer-readable media provide nonvolatile storage of computer-readable instructions, data structures, program modules and other data for the computer 520. Any type of computer-readable media that can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, Bernoulli cartridges, random access memories (RAMs), read only memories (ROMs), and the like, may be used in the operating environment.

A number of program modules may be stored on the hard disk, magnetic disk 529, optical disk 531, ROM 524, or RAM, including an operating system 535, one or more application programs

536, other program modules 537, and program data 538. A user may enter commands and information into the personal computer 520 through input devices such as a keyboard 540 and pointing device 542. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, scanner, or the like. These and other input devices are often connected to the processing unit 521 through a serial port interface 546 that is coupled to the system bus, but may be connected by other interfaces, such as a parallel port, game port, or a universal serial bus (USB). A monitor 547 or other type of display device is also connected to the system bus 523 via an interface, such as a video adapter 548. In addition to the monitor, computers typically include other peripheral output devices (not shown), such as speakers and printers.

The computer 520 may operate in a networked environment using logical connections to one or more remote computers, such as remote computer 549. These logical connections may be achieved by a communication device coupled to or a part of the computer 520, or in other manners. The remote computer 549 may be another computer, a server, a router, a network PC, a client, a peer device or other common network node, and typically includes many or all of the elements described above relative to the computer 520, although only a memory storage device 550 has been illustrated in FIG. 190. The logical connections depicted in FIG. 190 include a local-area network (LAN) 551 and a wide-area network (WAN) 552. Such networking environments are commonplace in office networks, enterprise-wide computer networks, intranets and the Internet, which all are types of networks.

When used in a LAN-networking environment, the computer 520 is connected to the local network 551 through a network interface or adapter 553, which is one type of communications device. When used in a WAN-networking environment, the computer 520 often includes a modem 554, a type of communications device, or any other type of communications device for establishing communications over the wide area network 552. The modem 554, which may be internal or external, is connected to the system bus 523 via the serial port interface 546. In a networked environment, program modules depicted relative to the personal computer 520, or portions thereof, may be stored in the remote memory storage device. It is appreciated that the network connections shown are non-limiting examples and other communications devices for establishing a communications link between computers may be used.

Modules

Sequencing Module

Sequencing and obtaining sequencing reads can be provided by a sequencing module or by an apparatus comprising a sequencing module. A "sequence receiving module" as used herein is the same as a "sequencing module". An apparatus comprising a sequencing module can be any apparatus that determines the sequence of a nucleic acid from a sequencing technology known in the art. In certain embodiments, an apparatus comprising a sequencing module performs a sequencing reaction known in the art. A sequencing module generally provides a nucleic acid sequence read according to data from a sequencing reaction (e.g., signals generated from a sequencing apparatus). In some embodiments, a sequencing module or an apparatus comprising a sequencing module is required to provide sequencing reads. In some embodiments a sequencing module can receive, obtain, access or recover sequence reads from another sequencing module, computer peripheral, operator, server, hard drive, apparatus or from a suitable source. In certain embodiments a sequencing module can manipulate sequence reads. For example, a sequencing module can align, assemble, fragment, complement, reverse complement, error check, or error correct sequence reads. An apparatus comprising a sequencing module can comprise at least one processor. In some embodiments, sequencing reads are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the sequencing module. In some embodiments, sequencing reads are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a sequencing module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In certain embodiments a sequencing module gathers, assembles and/or receives data and/or information from another module, apparatus, peripheral, component or specialized component (e.g., a sequencer). In some embodiments, sequencing reads are provided by an apparatus comprising one or more of the following: one or more flow cells, a camera, a photo detector, a photo cell, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. Often a sequencing module receives, gathers and/or assembles sequence reads. In certain embodiments a sequencing module accepts and gathers input data and/or information from an operator of an apparatus. For example, sometimes an operator of an apparatus provides instructions, a constant, a threshold value, a formula or a predetermined value to a module. In certain embodiments a sequencing module can transform data and/or information that it receives

into a contiguous nucleic acid sequence. In some embodiments, a nucleic acid sequence provided by a sequencing module is printed or displayed. In some embodiments, sequence reads are provided by a sequencing module and transferred from a sequencing module to an apparatus or an apparatus comprising any suitable peripheral, component or specialized component. In some embodiments, data and/or information are provided from a sequencing module to an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In certain embodiments, data and/or information related to sequence reads can be transferred from a sequencing module to any other suitable module. A sequencing module can transfer sequence reads to a mapping module or counting module, in some embodiments.

Mapping Module

Sequence reads can be mapped by a mapping module or by an apparatus comprising a mapping module, which mapping module generally maps reads to a reference genome or segment thereof. A mapping module can map sequencing reads by a suitable method known in the art. In some embodiments, a mapping module or an apparatus comprising a mapping module is required to provide mapped sequence reads. An apparatus comprising a mapping module can comprise at least one processor. In some embodiments, mapped sequencing reads are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the mapping module. In some embodiments, sequencing reads are mapped by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a mapping module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). An apparatus may comprise a mapping module and a sequencing module. In some embodiments, sequence reads are mapped by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. A mapping module can receive sequence reads from a sequencing module, in some embodiments. Mapped sequencing reads can be transferred from a mapping module to a counting module or a normalization module, in some embodiments.

Counting Module

Counts can be provided by a counting module or by an apparatus comprising a counting module. A counting module can determine, assemble, and/or display counts according to a counting method known in the art. A counting module generally determines or assembles counts according to counting methodology known in the art. In some embodiments, a counting module or an apparatus comprising a counting module is required to provide counts. An apparatus comprising a counting module can comprise at least one processor. In some embodiments, counts are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the counting module. In some embodiments, reads are counted by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a counting module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, reads are counted by an apparatus comprising one or more of the following: a sequencing module, a mapping module, one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. A counting module can receive data and/or information from a sequencing module and/or a mapping module, transform the data and/or information and provide counts (e.g., counts mapped to genomic sections). A counting module can receive mapped sequence reads from a mapping module. A counting module can receive normalized mapped sequence reads from a mapping module or from a normalization module. A counting module can transfer data and/or information related to counts (e.g., counts, assembled counts and/or displays of counts) to any other suitable apparatus, peripheral, or module. In certain embodiments data and/or information related to counts are transferred from a counting module to a normalization module, a plotting module, a categorization module and/or an outcome module.

Filtering Module

Filtering genomic sections can be provided by a filtering module (e.g., by an apparatus comprising a filtering module). In some embodiments, a filtering module is required to provide filtered genomic section data (e.g., filtered genomic sections) and/or to remove genomic sections from consideration. In certain embodiments a filtering module removes counts mapped to a genomic section from consideration. In certain embodiments a filtering module removes counts mapped to

a genomic section from a determination of an elevation or a profile. A filtering module can filter data (e.g., counts, counts mapped to genomic sections, genomic sections, genomic sections elevations, normalized counts, raw counts, and the like) by one or more filtering procedures known in the art or described herein. An apparatus comprising a filtering module can comprise at least one processor. In some embodiments, filtered data is provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the filtering module. In some embodiments, filtered data is provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a filtering module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, filtered data is provided by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. A filtering module can receive data and/or information from a suitable apparatus or module. In certain embodiments a filtering module can receive data and/or information from a sequencing module, a normalization module, a weighting module, a mapping module or counting module. A filtering module can receive sequencing reads from a sequencing module, mapped sequencing reads from a mapping module and/or counts from a counting module, in some embodiments. Often a filtering module receives data and/or information from another apparatus or module, transforms the data and/or information and provides filtered data and/or information (e.g., filtered counts, filtered values, filtered genomic sections, and the like). Filtered data and/or information can be transferred from a filtering module to a comparison module, a normalization module, a weighting module, a range setting module, an adjustment module, a categorization module, and/or an outcome module, in certain embodiments.

Weighting Module

Weighting genomic sections can be provided by a weighting module (e.g., by an apparatus comprising a weighting module). In some embodiments, a weighting module is required to weight genomics sections and/or provide weighted genomic section values. A weighting module can weight genomic sections by one or more weighting procedures known in the art or described herein. An apparatus comprising a weighting module can comprise at least one processor. In some embodiments, weighted genomic sections are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or

more instructions (e.g., processes, routines and/or subroutines) from the weighting module. In some embodiments, weighted genomic sections are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a weighting module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, weighted genomic sections are provided by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. A weighting module can receive data and/or information from a suitable apparatus or module. In certain embodiments a weighting module can receive data and/or information from a sequencing module, a normalization module, a filtering module, a mapping module and/or a counting module. A weighting module can receive sequencing reads from a sequencing module, mapped sequencing reads from a mapping module and/or counts from a counting module, in some embodiments. In some embodiments a weighting module receives data and/or information from another apparatus or module, transforms the data and/or information and provides data and/or information (e.g., weighted genomic sections, weighted values, and the like). Weighted genomic section data and/or information can be transferred from a weighting module to a comparison module, a normalization module, a filtering module, a range setting module, an adjustment module, a categorization module, and/or an outcome module, in certain embodiments.

Normalization Module

Normalized data (e.g., normalized counts) can be provided by a normalization module (e.g., by an apparatus comprising a normalization module). In some embodiments, a normalization module is required to provide normalized data (e.g., normalized counts) obtained from sequencing reads. A normalization module can normalize data (e.g., counts, filtered counts, raw counts) by one or more normalization procedures known in the art. An apparatus comprising a normalization module can comprise at least one processor. In some embodiments, normalized data is provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the normalization module. In some embodiments, normalized data is provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a normalization module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In

some embodiments, normalized data is provided by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. A normalization module can receive data and/or information from a suitable apparatus or module. In certain embodiments a normalization module can receive data and/or information from a sequencing module, a normalization module, a mapping module or counting module. A normalization module can receive sequencing reads from a sequencing module, mapped sequencing reads from a mapping module and/or counts from a counting module, in some embodiments. Often a normalization module receives data and/or information from another apparatus or module, transforms the data and/or information and provides normalized data and/or information (e.g., normalized counts, normalized values, normalized reference values (NRVs), and the like). Normalized data and/or information can be transferred from a normalization module to a comparison module, a normalization module, a range setting module, an adjustment module, a categorization module, and/or an outcome module, in certain embodiments. In certain embodiments normalized counts (e.g., normalized mapped counts) are transferred to an expected representation module and/or to a representation module from a normalization module.

GC Bias Module

Determining GC bias (e.g., determining GC bias for each of the portions of a reference genome (e.g., genomic sections)) can be provided by a GC bias module (e.g., by an apparatus comprising a GC bias module). In some embodiments, a GC bias module is required to provide a determination of GC bias. In certain embodiments a GC bias module provides a determination of GC bias from a fitted relationship (e.g., a fitted linear relationship) between counts of sequence reads mapped to each of the portions of a reference genome and GC content of each portion. An apparatus comprising a GC bias module can comprise at least one processor. In some embodiments, GC bias determinations (i.e., GC bias data) are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the GC bias module. In some embodiments, GC bias data is provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a GC bias module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, GC bias data is provided by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like.

A GC bias module can receive data and/or information from a suitable apparatus or module. In certain embodiments a GC bias module can receive data and/or information from a sequencing module, a normalization module, a weighting module, a mapping module or counting module. A GC bias module sometimes is part of a normalization module (e.g., PERUN normalization module). A GC bias module can receive sequencing reads from a sequencing module, mapped sequencing reads from a mapping module and/or counts from a counting module, in some embodiments. Often a GC bias module receives data and/or information from an apparatus or another module (e.g., a counting module), transforms the data and/or information and provides GC bias data and/or information (e.g., a determination of GC bias, a linear fitted relationship, and the like). GC bias data and/or information can be transferred from a GC bias module to an elevation module, filtering module, comparison module, a normalization module, a weighting module, a range setting module, an adjustment module, a categorization module, and/or an outcome module, in certain embodiments.

Elevation Module

Determining elevations (e.g., genomic section elevations) and/or calculating genomic section levels (e.g., genomic section elevations) for portions of a reference genome can be provided by an elevation module (e.g., by an apparatus comprising an elevation module). In some embodiments, an elevation module is required to provide an elevation or a calculated genomic section level. In certain embodiments an elevation module provides an elevation from a fitted relationship (e.g., a fitted linear relationship) between a GC bias and counts of sequence reads mapped to each of the portions of a reference genome. In certain embodiments an elevation module calculates a genomic section level as part of PERUN. In some embodiments, an elevation module provides a genomic section level (i.e., L_i) according to equation $L_i = (m_i - G_iS)\ I^{-1}$ wherein G_i is the GC bias, m_i is measured counts mapped to each portion of a reference genome, i is a sample, and I is the intercept and S is the slope of the a fitted relationship (e.g., a fitted linear relationship) between a GC bias and counts of sequence reads mapped to each of the portions of a reference genome. An apparatus comprising an elevation module can comprise at least one processor. In some embodiments, an elevation determination (i.e., elevation data) is provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the elevation module. In some embodiments, elevation data is provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some

embodiments, an elevation module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, elevation data is provided by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the like. An elevation module can receive data and/or information from a suitable apparatus or module. In certain embodiments an elevation module can receive data and/or information from a GC bias module, a sequencing module, a normalization module, a weighting module, a mapping module or counting module. An elevation module can receive sequencing reads from a sequencing module, mapped sequencing reads from a mapping module and/or counts from a counting module, in some embodiments. An elevation module sometimes is part of a normalization module (e.g., PERUN normalization module). Often an elevation module receives data and/or information from an apparatus or another module (e.g., a GC bias module), transforms the data and/or information and provides elevation data and/or information (e.g., a determination of an elevation, a linear fitted relationship, and the like). Elevation data and/or information can be transferred from an elevation module to a comparison module, a normalization module, a weighting module, a range setting module, an adjustment module, a categorization module, a module in a normalization module and/or an outcome module, in certain embodiments.

Comparison Module

A first elevation can be identified as significantly different from a second elevation by a comparison module or by an apparatus comprising a comparison module. In some embodiments, a comparison module or an apparatus comprising a comparison module is required to provide a comparison between two elevations. An apparatus comprising a comparison module can comprise at least one processor. In some embodiments, elevations are determined to be significantly different by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the comparison module. In some embodiments, elevations are determined to be significantly different by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a comparison module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, elevations are determined to be significantly different by an apparatus comprising one or more of the following: one or more flow cells, a camera, fluid handling components, a printer, a display (e.g., an LED, LCT or CRT) and the

like. A comparison module can receive data and/or information from a suitable module. A comparison module can receive data and/or information from a sequencing module, a mapping module, a counting module, or a normalization module. A comparison module can receive normalized data and/or information from a normalization module. Data and/or information derived from, or transformed by, a comparison module can be transferred from a comparison module to a range setting module, a plotting module, an adjustment module, a categorization module or an outcome module. A comparison between two or more elevations and/or an identification of an elevation as significantly different from another elevation can be transferred from (e.g., provided to) a comparison module to a categorization module, range setting module or adjustment module.

Range Setting Module

Expected ranges (e.g., expected elevation ranges) for various copy number variations (e.g., duplications, insertions and/or deletions) or ranges for the absence of a copy number variation can be provided by a range setting module or by an apparatus comprising a range setting module. In certain embodiments, expected elevations are provided by a range setting module or by an apparatus comprising a range setting module. In some embodiments, a range setting module or an apparatus comprising a range setting module is required to provide expected elevations and/or ranges. In certain embodiments a range setting module gathers, assembles and/or receives data and/or information from another module or apparatus. In certain embodiments a range setting module or an apparatus comprising a range setting module provides and/or transfers data and/or information to another module or apparatus. In certain embodiments a range setting module accepts and gathers data and/or information from a component or peripheral. Often a range setting module gathers and assembles elevations, reference elevations, uncertainty values, and/or constants. In certain embodiments a range setting module accepts and gathers input data and/or information from an operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a module. An apparatus comprising a range setting module can comprise at least one processor. In some embodiments, expected elevations and expected ranges are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the range setting module. In some embodiments, expected ranges and elevations are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, a range setting module operates with one or more external

processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In some embodiments, expected ranges are provided by an apparatus comprising a suitable peripheral or component. A range setting module can receive normalized data from a normalization module or comparison data from a comparison module. Data and/or information derived from or transformed by a range setting module (e.g., set ranges, range limits, expected elevation ranges, thresholds, and/or threshold ranges) can be transferred from a range setting module to an adjustment module, an outcome module, a categorization module, plotting module or other suitable apparatus and/or module.

Categorization Module

A copy number variation (e.g., a maternal and/or fetal copy number variation, a fetal copy number variation, a duplication, insertion, deletion) can be categorized by a categorization module or by an apparatus comprising a categorization module. In certain embodiments a copy number variation (e.g., a maternal and/or fetal copy number variation) is categorized by a categorization module. In certain embodiments an elevation (e.g., a first elevation) determined to be significantly different from another elevation (e.g., a second elevation) is identified as representative of a copy number variation by a categorization module. In certain embodiments the absence of a copy number variation is determined by a categorization module. In some embodiments, a determination of a copy number variation can be determined by an apparatus comprising a categorization module. A categorization module can be specialized for categorizing a maternal and/or fetal copy number variation, a fetal copy number variation, a duplication, deletion or insertion or lack thereof or combination of the foregoing. For example, a categorization module that identifies a maternal deletion can be different than and/or distinct from a categorization module that identifies a fetal duplication. In some embodiments, a categorization module or an apparatus comprising a categorization module is required to identify a copy number variation or an outcome determinative of a copy number variation. An apparatus comprising a categorization module can comprise at least one processor. In some embodiments, a copy number variation or an outcome determinative of a copy number variation is categorized by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the categorization module. In some embodiments, a copy number variation or an outcome determinative of a copy number variation is categorized by an apparatus that may include multiple processors, such as processors coordinated and working in parallel. In some embodiments, a categorization module operates with one or more external

processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In certain embodiments a categorization module transfers or receives and/or gathers data and/or information to or from a component or peripheral. Often a categorization module receives, gathers and/or assembles counts, elevations, profiles, normalized data and/or information, reference elevations, expected elevations, expected ranges, uncertainty values, adjustments, adjusted elevations, plots, comparisons and/or constants. In certain embodiments a categorization module accepts and gathers input data and/or information from an operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a module. In some embodiments, data and/or information are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, identification or categorization of a copy number variation or an outcome determinative of a copy number variation is provided by an apparatus comprising a suitable peripheral or component. In certain embodiments a categorization module gathers, assembles and/or receives data and/or information from another module or apparatus. A categorization module can receive normalized data from a normalization module, expected elevations and/or ranges from a range setting module, comparison data from a comparison module, plots from a plotting module, and/or adjustment data from an adjustment module. A categorization module can transform data and/or information that it receives into a determination of the presence or absence of a copy number variation. A categorization module can transform data and/or information that it receives into a determination that an elevation represents a genomic section comprising a copy number variation or a specific type of copy number variation (e.g., a maternal homozygous deletion). Data and/or information related to a copy number variation or an outcome determinative of a copy number variation can be transferred from a categorization module to a suitable apparatus and/or module. A copy number variation or an outcome determinative of a copy number variation categorized by methods described herein can be independently verified by further testing (e.g., by targeted sequencing of maternal and/or fetal nucleic acid).

Adjustment Module

In some embodiments, adjustments (e.g., adjustments to elevations or profiles) are made by an adjustment module or by an apparatus comprising an adjustment module. In some embodiments, an adjustment module or an apparatus comprising an adjustment module is required to adjust an elevation. An apparatus comprising an adjustment module can comprise at least one processor.

In some embodiments, an adjusted elevation is provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the adjustment module. In some embodiments, an elevation is adjusted by an apparatus that may include multiple processors, such as processors coordinated and working in parallel. In some embodiments, an adjustment module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In certain embodiments an apparatus comprising an adjustment module gathers, assembles and/or receives data and/or information from another module or apparatus. In certain embodiments an apparatus comprising an adjustment module provides and/or transfers data and/or information to another module or apparatus.

In certain embodiments an adjustment module receives and gathers data and/or information from a component or peripheral. Often an adjustment module receives, gathers and/or assembles counts, elevations, profiles, reference elevations, expected elevations, expected elevation ranges, uncertainty values, adjustments and/or constants. Often an adjustment module receives gathers and/or assembles elevations (e.g., first elevations) that are categorized or determined to be copy number variations (e.g., a maternal copy number variation, fetal copy number variation, or a maternal copy number variation and a fetal copy number variation). In certain embodiments an adjustment module accepts and gathers input data and/or information from an operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a module. In some embodiments, data and/or information are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, an elevation is adjusted by an apparatus comprising a suitable peripheral or component. An apparatus comprising an adjustment module can receive normalized data from a normalization module, ranges from a range setting module, comparison data from a comparison module, elevations identified (e.g., identified as a copy number variation) from a categorization module, and/or adjustment data from another adjustment module. An adjustment module can receive data and/or information, transform the received data and/or information and provide adjustments. Data and/or information derived from, or transformed by, an adjustment module can be transferred from an adjustment module to a categorization module or to a suitable apparatus and/or module. An elevation adjusted by methods described herein can be independently verified and/or adjusted by further testing (e.g., by targeted sequencing of maternal and or fetal nucleic acid).

Plotting Module

In some embodiments a count, an elevation, and/or a profile is plotted (e.g., graphed). In certain embodiments a plot (e.g., a graph) comprises an adjustment. In certain embodiments a plot comprises an adjustment of a count, an elevation, and/or a profile. In certain embodiments a count, an elevation, and/or a profile is plotted and a count, elevation, and/or a profile comprises an adjustment. Often a count, an elevation, and/or a profile is plotted and a count, elevation, and/or a profile are compared. In certain embodiments one or more genetic variations (e.g., an aneuploidy, a copy number variation) are identified and/or categorized from a plot of a count, an elevation, and/or a profile. In certain embodiments an outcome is determined from a plot of a count, an elevation, and/or a profile. In some embodiments, a plot (e.g., a graph) is made (e.g., generated) by a plotting module or an apparatus comprising a plotting module. In some embodiments, a plotting module or an apparatus comprising a plotting module is required to plot a count, an elevation or a profile. A plotting module may display a plot or send a plot to a display (e.g., a display module). An apparatus comprising a plotting module can comprise at least one processor. In some embodiments, a plot is provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the plotting module. In some embodiments, a plot is made by an apparatus that may include multiple processors, such as processors coordinated and working in parallel. In some embodiments, a plotting module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In certain embodiments an apparatus comprising a plotting module gathers, assembles and/or receives data and/or information from another module or apparatus. In certain embodiments a plotting module receives and gathers data and/or information from a component or peripheral. Often a plotting module receives, gathers, assembles and/or plots sequence reads, genomic sections, mapped reads, counts, elevations, profiles, reference elevations, expected elevations, expected elevation ranges, uncertainty values, comparisons, categorized elevations (e.g., elevations identified as copy number variations) and/or outcomes, adjustments and/or constants. In certain embodiments a plotting module accepts and gathers input data and/or information from an operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a plotting module. In some embodiments, data and/or information are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some

embodiments, a count, an elevation and/or a profile is plotted by an apparatus comprising a suitable peripheral or component. An apparatus comprising a plotting module can receive normalized data from a normalization module, ranges from a range setting module, comparison data from a comparison module, categorization data from a categorization module, and/or adjustment data from an adjustment module. A plotting module can receive data and/or information, transform the data and/or information and provided plotted data. In certain embodiments an apparatus comprising a plotting module provides and/or transfers data and/or information to another module or apparatus. An apparatus comprising a plotting module can plot a count, an elevation and/or a profile and provide or transfer data and/or information related to the plotting to a suitable apparatus and/or module. Often a plotting module receives, gathers, assembles and/or plots elevations (e.g., profiles, first elevations) and transfers plotted data and/or information to and from an adjustment module and/or comparison module. Plotted data and/or information is sometimes transferred from a plotting module to a categorization module and/or a peripheral (e.g., a display or printer). In some embodiments, plots are categorized and/or determined to comprise a genetic variation (e.g., an aneuploidy) or a copy number variation (e.g., a maternal and/or fetal copy number variation). A count, an elevation and/or a profile plotted by methods described herein can be independently verified and/or adjusted by further testing (e.g., by targeted sequencing of maternal and or fetal nucleic acid).

Representation Module

In certain embodiments, a chromosome representation is determined by a representation module. In certain embodiments, an ECR is determined by an expected representation module. In certain embodiments, an MCR is determined by a representation module. A representation module can be a representation module or an expected representation module. In some embodiments, a representation module determines one or more ratios. As used herein the term “ratio” refers to a numerical value (e.g., a number arrived at) by dividing a first numerical value by a second numerical value. For example, a ratio between A and B can be expressed mathematically as A/B or B/A and a numerical value for the ratio can be obtained by dividing A by B or by dividing B by A. In certain embodiments, a representation module (e.g., a representation module) determines an MCR by generating a ratio of counts. In certain embodiments a representation module determines an MCR for an affected autosome (e.g., chromosome 13 in the case of a trisomy 13, chromosome 18 in the case of a trisomy 18 or chromosome 21 in the case of a trisomy 21). For example, sometimes a representation module (e.g., a representation module) determines an MCR by

generating a ratio of counts mapped to genomic sections of chromosome *n* to the total number of counts mapped to genomic sections of all autosomal chromosomes represented in a profile. In certain embodiments a representation module (e.g., a representation module) determines an MCR by generating a ratio of counts mapped to genomic sections of a sex chromosome (e.g., chromosome X or Y) to the total number of counts mapped to genomic sections of all autosomal chromosomes represented in a profile. In certain embodiments, a representation module (e.g., an expected representation module) determines an ECR by generating a ratio of genomic sections. In certain embodiments an expected representation module determines an ECR for an affected autosome (e.g., chromosome 13 the case of a trisomy 13, chromosome 18 in the case of a trisomy 18 or chromosome 21 in the case of a trisomy 21). For example, sometimes a representation module (e.g., an expected representation module) determines an ECR by generating a ratio of genomic sections for chromosome *n* to all autosomal genomic sections in a profile. In some embodiments, a representation module can provide a ratio of an MCR to an ECR. In certain embodiments a representation module or an apparatus comprising a representation module gathers, assembles, receives, provides and/or transfers data and/or information to or from another module, apparatus, component, peripheral or operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a representation module. A representation module can receive data and/or information from a sequencing module, sequencing module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module and/or logic processing module. In certain embodiments normalized mapped counts are transferred to a representation module from a normalization module. In certain embodiments normalized mapped counts are transferred to an expected representation module from a normalization module. Data and/or information derived from or transformed by a representation module can be transferred from a representation module to a normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module, fetal fraction module or other suitable apparatus and/or module. In certain embodiments an MCR for chromosome 21, 18, 15, an X and/or a Y chromosome is transferred to a fetal fraction module from a representation module (e.g., a representation module). In certain embodiments an ECR for chromosome 21, 18, 15, an X and/or a Y chromosome is transferred to a fetal fraction module from a representation module (e.g., an expected representation module). An apparatus comprising a representation module can comprise at least one processor. In some embodiments, a representation is provided by an apparatus that

includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the representation module. In some embodiments, a representation module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)).

Relationship Module

In certain embodiments, a relationship is determined by a relationship module. In some embodiments a relationship is generated for a fetal fraction determination and an MCR of an X or a Y chromosome by a relationship module. In some embodiments a relationship is generated for (i) a fetal fraction determined by a first method and (ii) a fetal fraction determined by a second method by a relationship module. In certain embodiments a relationship module or an apparatus comprising a relationship module gathers, assembles, receives, provides and/or transfers data and/or information to or from another module, apparatus, component, peripheral or operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a relationship module. A relationship module can receive data and/or information from a sequencing module, sequencing module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module and/or a representation module. Data and/or information derived from or transformed by a relationship module can be transferred from a relationship module to a normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module, representation module, fetal fraction module or other suitable apparatus and/or module. An apparatus comprising a relationship module can comprise at least one processor. In some embodiments, data and/or information are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the relationship module. In some embodiments, a relationship module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)).

Fetal Fraction Module

In certain embodiments, a fetal fraction is determined by a fetal fraction module. In certain embodiments a fetal fraction module or an apparatus comprising a fetal fraction module gathers, assembles, receives, provides and/or transfers data and/or information to or from another module, apparatus, component, peripheral or operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to a fetal fraction module. A fetal fraction module can receive data and/or information from a sequencing module, sequencing module, mapping module, weighting module, filtering module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module, a representation module and/or a relationship module. Data and/or information derived from or transformed by a fetal fraction module can be transferred from a fetal fraction module to a normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module, representation module, relationship module, fetal fraction module or other suitable apparatus and/or module. An apparatus comprising a fetal fraction module can comprise at least one processor. In some embodiments, data and/or information are provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the fetal fraction module. In some embodiments, a fetal fraction module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)).

In some embodiments an apparatus (e.g., a first apparatus) comprises a normalization module, a representation module, an expected representation module, a fetal fraction module and a relationship module. In some embodiments an apparatus (e.g., a second apparatus) comprises a mapping module and a counting module. In certain embodiments an apparatus (e.g., a third apparatus) comprises a sequencing module.

Outcome Module

The presence or absence of a genetic variation (an aneuploidy, a fetal aneuploidy, a copy number variation) can be identified by an outcome module or by an apparatus comprising an outcome

module. In certain embodiments a genetic variation is identified by an outcome module. Often a determination of the presence or absence of an aneuploidy is identified by an outcome module. In some embodiments, an outcome determinative of a genetic variation (an aneuploidy, a copy number variation) can be identified by an outcome module or by an apparatus comprising an outcome module. An outcome module can be specialized for determining a specific genetic variation (e.g., a trisomy, a trisomy 21, a trisomy 18). For example, an outcome module that identifies a trisomy 21 can be different than and/or distinct from an outcome module that identifies a trisomy 18. In some embodiments, an outcome module or an apparatus comprising an outcome module is required to identify a genetic variation or an outcome determinative of a genetic variation (e.g., an aneuploidy, a copy number variation). An apparatus comprising an outcome module can comprise at least one processor. In some embodiments, a genetic variation or an outcome determinative of a genetic variation is provided by an apparatus that includes a processor (e.g., one or more processors) which processor can perform and/or implement one or more instructions (e.g., processes, routines and/or subroutines) from the outcome module. In some embodiments, a genetic variation or an outcome determinative of a genetic variation is identified by an apparatus that may include multiple processors, such as processors coordinated and working in parallel. In some embodiments, an outcome module operates with one or more external processors (e.g., an internal or external network, server, storage device and/or storage network (e.g., a cloud)). In certain embodiments an apparatus comprising an outcome module gathers, assembles and/or receives data and/or information from another module or apparatus. In certain embodiments an apparatus comprising an outcome module provides and/or transfers data and/or information to another module or apparatus. In certain embodiments an outcome module transfers, receives or gathers data and/or information to or from a component or peripheral. Often an outcome module receives, gathers and/or assembles counts, elevations, profiles, normalized data and/or information, reference elevations, expected elevations, expected ranges, uncertainty values, adjustments, adjusted elevations, plots, categorized elevations, comparisons and/or constants. In certain embodiments an outcome module accepts and gathers input data and/or information from an operator of an apparatus. For example, sometimes an operator of an apparatus provides a constant, a threshold value, a formula or a predetermined value to an outcome module. In some embodiments, data and/or information are provided by an apparatus that includes multiple processors, such as processors coordinated and working in parallel. In some embodiments, identification of a genetic variation or an outcome determinative of a genetic variation is provided by an apparatus comprising a suitable peripheral or component. An apparatus comprising an outcome module can receive normalized data from a normalization module, expected elevations

and/or ranges from a range setting module, comparison data from a comparison module, categorized elevations from a categorization module, plots from a plotting module, and/or adjustment data from an adjustment module. An outcome module can receive data and/or information, transform the data and/or information and provide an outcome. An outcome module can provide or transfer data and/or information related to a genetic variation or an outcome determinative of a genetic variation to a suitable apparatus and/or module. A genetic variation or an outcome determinative of a genetic variation identified by methods described herein can be independently verified by further testing (e.g., by targeted sequencing of maternal and/or fetal nucleic acid).

Transformations

As noted above, data sometimes is transformed from one form into another form. The terms "transformed", "transformation", and grammatical derivations or equivalents thereof, as used herein refer to an alteration of data from a physical starting material (e.g., test subject and/or reference subject sample nucleic acid) into a digital representation of the physical starting material (e.g., sequence read data), and in some embodiments includes a further transformation into one or more numerical values or graphical representations of the digital representation that can be utilized to provide an outcome. In certain embodiments, the one or more numerical values and/or graphical representations of digitally represented data can be utilized to represent the appearance of a test subject's physical genome (e.g., virtually represent or visually represent the presence or absence of a genomic insertion, duplication or deletion; represent the presence or absence of a variation in the physical amount of a sequence associated with medical conditions). A virtual representation sometimes is further transformed into one or more numerical values or graphical representations of the digital representation of the starting material. These procedures can transform physical starting material into a numerical value or graphical representation, or a representation of the physical appearance of a test subject's genome.

In some embodiments, transformation of a data set facilitates providing an outcome by reducing data complexity and/or data dimensionality. Data set complexity sometimes is reduced during the process of transforming a physical starting material into a virtual representation of the starting material (e.g., sequence reads representative of physical starting material). A suitable feature or variable can be utilized to reduce data set complexity and/or dimensionality. Non-limiting examples of features that can be chosen for use as a target feature for data processing include GC

content, fetal gender prediction, identification of chromosomal aneuploidy, identification of particular genes or proteins, identification of cancer, diseases, inherited genes/traits, chromosomal abnormalities, a biological category, a chemical category, a biochemical category, a category of genes or proteins, a gene ontology, a protein ontology, co-regulated genes, cell signaling genes, cell cycle genes, proteins pertaining to the foregoing genes, gene variants, protein variants, co-regulated genes, co-regulated proteins, amino acid sequence, nucleotide sequence, protein structure data and the like, and combinations of the foregoing. Non-limiting examples of data set complexity and/or dimensionality reduction include; reduction of a plurality of sequence reads to profile plots, reduction of a plurality of sequence reads to numerical values (e.g., normalized values, Z-scores, p-values); reduction of multiple analysis methods to probability plots or single points; principle component analysis of derived quantities; and the like or combinations thereof.

Genomic Section Normalization Systems, Apparatus and Computer Program Products

In certain aspects provided is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads of circulating, cell-free sample nucleic acid from a test subject mapped to genomic sections of a reference genome; and which instructions executable by the one or more processors are configured to: (a) generate a sample normalized count profile by normalizing counts of the sequence reads for each of the genomic sections; and (b) determine the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both from the sample normalized count profile in (a).

Provided also in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads of circulating, cell-free sample nucleic acid from a test subject mapped to genomic sections of a reference genome; and which instructions executable by the one or more processors are configured to: (a) generate a sample normalized count profile by normalizing counts of the sequence reads for each of the genomic sections; and (b) determine the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both from the sample normalized count profile in (a).

Also provided in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are

configured to: (a) access counts of sequence reads of circulating, cell-free sample nucleic acid from a test subject mapped to genomic sections of a reference genome; (b) generate a sample normalized count profile by normalizing counts of the sequence reads for each of the genomic sections; and (c) determine the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both from the sample normalized count profile in (b).

In some embodiments, the counts of the sequence reads for each of the genomic sections in a segment of the reference genome (e.g., the segment is a chromosome) individually are normalized according to the total counts of sequence reads in the genomic sections in the segment. Certain genomic sections in the segment sometimes are removed (e.g., filtered) and the remaining genomic sections in the segment are normalized.

In certain embodiments, the system, apparatus and/or computer program product comprises a: (i) a sequencing module configured to obtain nucleic acid sequence reads; (ii) a mapping module configured to map nucleic acid sequence reads to portions of a reference genome; (iii) a weighting module configured to weight genomic sections, (iv) a filtering module configured to filter genomic sections or counts mapped to a genomic section, (v) a counting module configured to provide counts of nucleic acid sequence reads mapped to portions of a reference genome; (vi) a normalization module configured to provide normalized counts; (vii) a comparison module configured to provide an identification of a first elevation that is significantly different than a second elevation; (viii) a range setting module configured to provide one or more expected elevation ranges; (ix) a categorization module configured to identify an elevation representative of a copy number variation; (x) an adjustment module configured to adjust an elevation identified as a copy number variation; (xi) a plotting module configured to graph and display an elevation and/or a profile; (xii) an outcome module configured to determine an outcome (e.g., outcome determinative of the presence or absence of a fetal aneuploidy); (xiii) a data display organization module configured to indicate the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both; (xiv) a logic processing module configured to perform one or more of map sequence reads, count mapped sequence reads, normalize counts and generate an outcome; or (xv) combination of two or more of the foregoing.

In some embodiments the sequencing module and mapping module are configured to transfer sequence reads from the sequencing module to the mapping module. The mapping module and counting module sometimes are configured to transfer mapped sequence reads from the mapping

module to the counting module. The counting module and filtering module sometimes are configured to transfer counts from the counting module to the filtering module. The counting module and weighting module sometimes are configured to transfer counts from the counting module to the weighting module. The mapping module and filtering module sometimes are configured to transfer mapped sequence reads from the mapping module to the filtering module. The mapping module and weighting module sometimes are configured to transfer mapped sequence reads from the mapping module to the weighting module. In certain embodiments the weighting module, filtering module and counting module are configured to transfer filtered and/or weighted genomic sections from the weighting module and filtering module to the counting module. The weighting module and normalization module sometimes are configured to transfer weighted genomic sections from the weighting module to the normalization module. The filtering module and normalization module sometimes are configured to transfer filtered genomic sections from the filtering module to the normalization module. In some embodiments, the normalization module and/or comparison module are configured to transfer normalized counts to the comparison module and/or range setting module. The comparison module, range setting module and/or categorization module independently are configured to transfer (i) an identification of a first elevation that is significantly different than a second elevation and/or (ii) an expected elevation range from the comparison module and/or range setting module to the categorization module, in some embodiments. In certain embodiments, the categorization module and the adjustment module are configured to transfer an elevation categorized as a copy number variation from the categorization module to the adjustment module. In some embodiments, the adjustment module, plotting module and the outcome module are configured to transfer one or more adjusted elevations from the adjustment module to the plotting module or outcome module. The normalization module sometimes is configured to transfer mapped normalized sequence read counts to one or more of the comparison module, range setting module, categorization module, adjustment module, outcome module or plotting module.

Parameterized Error Removal and Unbiased Normalization Systems, Apparatus and Computer Program Products

Provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample; and which instructions

executable by the one or more processors are configured to: (a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and (b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

Also provided in some aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample; and which instructions executable by the one or more processors are configured to: (a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and (b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

Also provided in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample; (b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and (c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels,

whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

Provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female bearing a fetus; and which instructions executable by the one or more processors are configured to: (a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; (b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between the GC bias and the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels; and (c) identify the presence or absence of an aneuploidy for the fetus according to the calculated genomic section levels with a sensitivity of 95% or greater and a specificity of 95% or greater.

Also provided in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female bearing a fetus; and which instructions executable by the one or more processors are configured to: (a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; (b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between the GC bias and the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels; and (c) identify the presence or absence of an aneuploidy for the fetus according to the calculated genomic section levels with a sensitivity of 95% or greater and a specificity of 95% or greater.

Provided also in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of sequence reads mapped to portions of a reference genome,

which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female bearing a fetus; (b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; (c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between the GC bias and the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels; and (d) identify the presence or absence of an aneuploidy for the fetus according to the calculated genomic section levels with a sensitivity of 95% or greater and a specificity of 95% or greater.

Also provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female bearing a fetus; and which instructions executable by the one or more processors are configured to: (a) determine experimental bias for each of the portions of the reference genome for multiple samples from a fitted relation between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) a mapping feature for each of the portions; and (b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between the experimental bias and the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

Provided also in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female bearing a fetus; and which instructions executable by the one or more processors are configured to: (a) determine experimental bias for each of the portions of the reference genome for multiple samples from a fitted relation between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) a mapping feature for each of the portions; and (b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation

between the experimental bias and the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

Also provided in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample; (b) determine experimental bias for each of the portions of the reference genome for multiple samples from a fitted relation between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) a mapping feature for each of the portions; and (c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between the experimental bias and the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

In certain embodiments, the system, apparatus and/or computer program product comprises a: (i) a sequencing module configured to obtain nucleic acid sequence reads; (ii) a mapping module configured to map nucleic acid sequence reads to portions of a reference genome; (iii) a weighting module configured to weight genomic sections; (iv) a filtering module configured to filter genomic sections or counts mapped to a genomic section; (v) a counting module configured to provide counts of nucleic acid sequence reads mapped to portions of a reference genome; (vi) a normalization module configured to provide normalized counts; (vii) a comparison module configured to provide an identification of a first elevation that is significantly different than a second elevation; (viii) a range setting module configured to provide one or more expected elevation ranges; (ix) a categorization module configured to identify an elevation representative of a copy number variation; (x) an adjustment module configured to adjust an elevation identified as a copy number variation; (xi) a plotting module configured to graph and display an elevation and/or a profile; (xii) an outcome module configured to determine an outcome (e.g., outcome determinative of the presence or absence of a fetal aneuploidy); (xiii) a data display organization module configured to indicate the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both; (xiv) a logic processing module configured to perform one or more of map

sequence reads, count mapped sequence reads, normalize counts and generate an outcome; or (xv) combination of two or more of the foregoing.

In some embodiments the sequencing module and mapping module are configured to transfer sequence reads from the sequencing module to the mapping module. The mapping module and counting module sometimes are configured to transfer mapped sequence reads from the mapping module to the counting module. The counting module and filtering module sometimes are configured to transfer counts from the counting module to the filtering module. The counting module and weighting module sometimes are configured to transfer counts from the counting module to the weighting module. The mapping module and filtering module sometimes are configured to transfer mapped sequence reads from the mapping module to the filtering module. The mapping module and weighting module sometimes are configured to transfer mapped sequence reads from the mapping module to the weighting module. In certain embodiments the weighting module, filtering module and counting module are configured to transfer filtered and/or weighted genomic sections from the weighting module and filtering module to the counting module. The weighting module and normalization module sometimes are configured to transfer weighted genomic sections from the weighting module to the normalization module. The filtering module and normalization module sometimes are configured to transfer filtered genomic sections from the filtering module to the normalization module. In some embodiments, the normalization module and/or comparison module are configured to transfer normalized counts to the comparison module and/or range setting module. The comparison module, range setting module and/or categorization module independently are configured to transfer (i) an identification of a first elevation that is significantly different than a second elevation and/or (ii) an expected elevation range from the comparison module and/or range setting module to the categorization module, in some embodiments. In certain embodiments, the categorization module and the adjustment module are configured to transfer an elevation categorized as a copy number variation from the categorization module to the adjustment module. In some embodiments, the adjustment module, plotting module and the outcome module are configured to transfer one or more adjusted elevations from the adjustment module to the plotting module or outcome module. The normalization module sometimes is configured to transfer mapped normalized sequence read counts to one or more of the comparison module, range setting module, categorization module, adjustment module, outcome module or plotting module.

Adjustment Systems, Apparatus and Computer Program Products

Provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; and which instructions executable by the one or more processors are configured to: (a) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (b) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (c) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; (d) adjust the first elevation by a predetermined value when the first elevation is within one of the expected elevation ranges, thereby providing an adjustment of the first elevation; and (e) determine the presence or absence of a chromosome aneuploidy in the fetus according to the elevations of genomic sections comprising the adjustment of (d), whereby the outcome determinative of the presence or absence of the chromosome aneuploidy is generated from the nucleic acid sequence reads.

Also provided in some aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; and which instructions executable by the one or more processors are configured to: (a) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (b) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (c) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; (d) adjust the first elevation by a predetermined value when the first elevation is within one of the expected elevation ranges, thereby providing an adjustment of the first elevation; and (e) determine the presence or absence of a chromosome aneuploidy in the fetus according to the

elevations of genomic sections comprising the adjustment of (d), whereby the outcome determinative of the presence or absence of the chromosome aneuploidy is generated from the nucleic acid sequence reads.

Provided also in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; (b) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (c) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (d) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; (e) adjust the first elevation by a predetermined value when the first elevation is within one of the expected elevation ranges, thereby providing an adjustment of the first elevation; and (f) determine the presence or absence of a chromosome aneuploidy in the fetus according to the elevations of genomic sections comprising the adjustment of (e), whereby the outcome determinative of the presence or absence of the chromosome aneuploidy is generated from the nucleic acid sequence reads.

Also provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; and which instructions executable by the one or more processors are configured to: (a) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (b) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (c) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; and (d) identify a maternal and/or fetal copy number variation within the genomic section based on one of the expected elevation ranges, whereby the maternal and/or fetal copy number variation is identified from the nucleic acid sequence reads.

Provided also in some aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; and which instructions executable by the one or more processors are configured to: (a) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (b) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (c) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; and (d) identify a maternal and/or fetal copy number variation within the genomic section based on one of the expected elevation ranges, whereby the maternal and/or fetal copy number variation is identified from the nucleic acid sequence reads.

Also provided in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; (b) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (c) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (d) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; and (e) identify a maternal and/or fetal copy number variation within the genomic section based on one of the expected elevation ranges, whereby the maternal and/or fetal copy number variation is identified from the nucleic acid sequence reads.

Provided also in some aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference

genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; and which instructions executable by the one or more processors are configured to: (a) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (b) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (c) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; (d) adjust the first elevation according to the second elevation, thereby providing an adjustment of the first elevation; and (e) determine the presence or absence of a chromosome aneuploidy in the fetus according to the elevations of genomic sections comprising the adjustment of (d), whereby the outcome determinative of the presence or absence of the chromosome aneuploidy is generated from the nucleic acid sequence reads.

In certain aspects provided is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; and which instructions executable by the one or more processors are configured to: (a) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (b) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (c) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; (d) adjust the first elevation according to the second elevation, thereby providing an adjustment of the first elevation; and (e) determine the presence or absence of a chromosome aneuploidy in the fetus according to the elevations of genomic sections comprising the adjustment of (d), whereby the outcome determinative of the presence or absence of the chromosome aneuploidy is generated from the nucleic acid sequence reads.

Provided in some aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a

reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a pregnant female; (b) normalize the counts mapped to the genomic sections of the reference genome, thereby providing a profile of normalized counts for the genomic sections; (c) identify a first elevation of the normalized counts significantly different than a second elevation of the normalized counts in the profile, which first elevation is for a first set of genomic sections, and which second elevation is for a second set of genomic sections; (d) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome; (e) adjust the first elevation according to the second elevation, thereby providing an adjustment of the first elevation; and (f) determine the presence or absence of a chromosome aneuploidy in the fetus according to the elevations of genomic sections comprising the adjustment of (e), whereby the outcome determinative of the presence or absence of the chromosome aneuploidy is generated from the nucleic acid sequence reads.

In certain embodiments, the system, apparatus and/or computer program product comprises a: (i) a sequencing module configured to obtain nucleic acid sequence reads; (ii) a mapping module configured to map nucleic acid sequence reads to portions of a reference genome; (iii) a weighting module configured to weight genomic sections; (iv) a filtering module configured to filter genomic sections or counts mapped to a genomic section; (v) a counting module configured to provide counts of nucleic acid sequence reads mapped to portions of a reference genome; (vi) a normalization module configured to provide normalized counts; (vii) a comparison module configured to provide an identification of a first elevation that is significantly different than a second elevation; (viii) a range setting module configured to provide one or more expected elevation ranges; (ix) a categorization module configured to identify an elevation representative of a copy number variation; (x) an adjustment module configured to adjust an elevation identified as a copy number variation; (xi) a plotting module configured to graph and display an elevation and/or a profile; (xii) an outcome module configured to determine an outcome (e.g., outcome determinative of the presence or absence of a fetal aneuploidy); (xiii) a data display organization module configured to indicate the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both; (xiv) a logic processing module configured to perform one or more of map sequence reads, count mapped sequence reads, normalize counts and generate an outcome; or (xv) combination of two or more of the foregoing.

Fetal Fraction Determination Systems, Apparatus and Computer Program Products

Provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and which instructions executable by the one or more processors are configured to: (a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

Also provided in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and which instructions executable by the one or more processors are configured to: (a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

Provided also in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the

blood of a pregnant female bearing a male fetus; (b) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

Also provided in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and which instructions executable by the one or more processors are configured to: (a) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

Provided also in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and which instructions executable by the one or more processors are configured to: (a) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for

nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

Also provided in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; (b) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

Provided also in certain aspects is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and which instructions executable by the one or more processors are configured to: (a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

Also provided in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and which instructions executable by the one or more processors are configured to: (a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

Also provided in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; (b) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

Provided also is a system comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome; and which instructions executable by the one or more processors are configured to: (a) generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

Also provided in certain aspects is an apparatus comprising one or more processors and memory, which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome; and which instructions executable by the one or more processors are configured to: (a) generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

Provided also in certain aspects is a computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to: (a) access counts of nucleic acid sequence reads mapped to genomic sections of a

reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome; (b) generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and (c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

In certain embodiments, the system, apparatus and/or computer program product comprises a: (i) a sequencing module configured to obtain nucleic acid sequence reads; (ii) a mapping module configured to map nucleic acid sequence reads to portions of a reference genome; (iii) a weighting module configured to weight genomic sections, (iv) a filtering module configured to filter genomic sections or counts mapped to a genomic section; (v) a counting module configured to provide counts of nucleic acid sequence reads mapped to portions of a reference genome; (vi) a normalization module configured to provide normalized counts; (vii) a comparison module configured to provide an identification of a first elevation that is significantly different than a second elevation; (viii) a range setting module configured to provide one or more expected elevation ranges; (ix) a categorization module configured to identify an elevation representative of a copy number variation; (x) an adjustment module configured to adjust an elevation identified as a copy number variation; (xi) a plotting module configured to graph and display an elevation and/or a profile; (xii) an outcome module configured to determine an outcome (e.g., outcome determinative of the presence or absence of a fetal aneuploidy); (xiii) a data display organization module configured to indicate the presence or absence of a segmental chromosomal aberration or a fetal aneuploidy or both; (xiv) a logic processing module configured to perform one or more of map sequence reads, count mapped sequence reads, normalize counts and generate an outcome; (xv) a representation module configured to determine an experimental chromosome representation (e.g., X chromosome representation, Y chromosome representation, autosome representation); (xvi) a relationship module configured to determine a relationship between (a) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (b) an experimental affected autosome representation for a sample; (xvii) a fetal fraction module configured to determine fetal fraction from an experimental

chromosome representation; or (xviii) combination of two or more of the foregoing. In certain embodiments, the copy number variation categorized from the first elevation is a maternal copy number variation. In some embodiments, the copy number variation categorized from the first elevation is a fetal copy number variation.

In some embodiments the sequencing module and mapping module are configured to transfer sequence reads from the sequencing module to the mapping module. The mapping module and counting module sometimes are configured to transfer mapped sequence reads from the mapping module to the counting module. The counting module and filtering module sometimes are configured to transfer counts from the counting module to the filtering module. The counting module and weighting module sometimes are configured to transfer counts from the counting module to the weighting module. The mapping module and filtering module sometimes are configured to transfer mapped sequence reads from the mapping module to the filtering module. The mapping module and weighting module sometimes are configured to transfer mapped sequence reads from the mapping module to the weighting module. In certain embodiments the weighting module, filtering module and counting module are configured to transfer filtered and/or weighted genomic sections from the weighting module and filtering module to the counting module. The weighting module and normalization module sometimes are configured to transfer weighted genomic sections from the weighting module to the normalization module. The filtering module and normalization module sometimes are configured to transfer filtered genomic sections from the filtering module to the normalization module. In some embodiments, the normalization module and/or comparison module are configured to transfer normalized counts to the comparison module and/or range setting module. The comparison module, range setting module and/or categorization module independently are configured to transfer (i) an identification of a first elevation that is significantly different than a second elevation and/or (ii) an expected elevation range from the comparison module and/or range setting module to the categorization module, in some embodiments. In certain embodiments, the categorization module and the adjustment module are configured to transfer an elevation categorized as a copy number variation from the categorization module to the adjustment module and/or fetal fraction module. In some embodiments, the adjustment module, plotting module and the outcome module are configured to transfer one or more adjusted elevations from the adjustment module to the plotting module, outcome module or fetal fraction module. The normalization module sometimes is configured to transfer mapped normalized sequence read counts to one or more of the comparison module, range setting module, categorization module, adjustment module, outcome module, plotting module, fetal fraction module

or representation module. In some embodiments, a relationship module is configured to receive information from the representation module, and is configured to transfer information to the fetal fraction module.

Examples

The following examples are provided by way of illustration only and not by way of limitation. Thus, the examples set forth below illustrate certain embodiments and do not limit the technology. Those of skill in the art will readily recognize a variety of non-critical parameters that could be changed or modified to yield essentially the same or similar results.

Example 1: General methods for detecting conditions associated with genetic variations.

The methods and underlying theory described herein can be utilized to detect various conditions associated with genetic variation and provide an outcome determinative of, or determine the presence or absence of a genetic variation. Non-limiting examples of genetic variations that can be detected with a method described herein include, segmental chromosomal aberrations (e.g., deletions, duplications), aneuploidy, gender, sample identification, disease conditions associated with genetic variation, the like or combinations of the foregoing.

Bin Filtering

The information content of a genomic region in a target chromosome can be visualized by plotting the result of the average separation between euploid and trisomy counts normalized by combined uncertainties, as a function of chromosome position. Increased uncertainty (see FIG. 1) or reduced gap between triploids and euploids (e.g., triploid pregnancies and euploid pregnancies)(see FIG. 2) both result in decreased Z-values for affected cases, sometimes reducing the predictive power of Z-scores.

FIG. 3 graphically illustrates a p-value profile, based on t-distribution, plotted as a function of chromosome position along chromosome 21. Analysis of the data presented in FIG. 3 identifies 36 uninformative chromosome 21 bins, each about 50 kilo-base pairs (kbp) in length. The uninformative region is located in the p-arm, close to centromere (21p11.2-21p11.1). Removing all 36 bins from the calculation of Z-scores, as schematically outlined in FIG. 4, sometimes can

significantly increase the Z-values for all trisomy cases, while introducing only random variations into euploid Z-values.

The improvement in predictive power afforded by removal of the 36 uninformative bins can be explained by examining the count profile for chromosome 21 (see FIG. 5). In FIG. 5, two arbitrarily chosen samples demonstrate the general tendency of count versus (vs) bin profiles to follow substantially similar trends, apart from short-range noise. The profiles shown in FIG. 5 are substantially parallel. The highlighted region of the profile plot presented in FIG. 5 (e.g., the region in the ellipse), while still exhibiting parallelism, also exhibit large fluctuations relative to the rest of chromosome. Removal of the fluctuating bins (e.g., the 36 uninformative bins) can improve precision and consistency of Z statistics, in some embodiments.

Bin Normalization

Filtering out uninformative bins, as described in Example 1, sometimes does not provide the desired improvement to the predictive power of Z-values. When chromosome 18 data is filtered to remove uninformative bins, as described in Example 1, the z-values did not substantially improve (see FIG. 6). As seen with the chromosome 21 count profiles presented in Example 1, the chromosome 18 count profiles also are substantially parallel, disregarding short range noise. However, two chromosome 18 samples used to evaluate binwise count uncertainties (see the bottom of FIG. 6) significantly deviate from the general parallelism of count profiles. The dips in the middle of the two traces, highlighted by the ellipse, represent large deletions. Other samples examined during the course of the experiment did not exhibit this deletion. The deletion coincides with the location of a dip in p-value profiles for chromosome 18, illustrated in by the ellipse shown in FIG. 7. That is, the dip observed in the p-value profiles for chromosome 18 are explained by the presence of the deletion in the chromosome 18 samples, which cause an increase in the variance of counts in the affected region. The variance in counts is not random, but represents a rare event (e.g., the deletion of a segment of chromosome 18), which, if included with other, random fluctuations from other samples, decreases the predictive power bin filtering procedure.

Two questions arise from this example; (1) how are p-value signals determined to be meaningful and/or useful, and (2) can the p-value approach described herein be generalized for use with any bin data (e.g., from within any chromosome, not only bins from within chromosomes 13, 18 or 21). A generalized procedure could be used to remove variability in the total counts for the entire

genome, which can often be used as the normalization factor when evaluating Z-scores. The data presented in FIG. 8 can be used to investigate the answers to the questions above by reconstructing the general contour of the data by assigning the median reference count to each bin, and normalizing each bin count in the test sample with respect to the assigned median reference count.

The medians are extracted from a set of known euploid references. Prior to computing the reference median counts, uninformative bins throughout the genome are filtered out. The remaining bin counts are normalized with respect to the total residual number of counts. The test sample is also normalized with respect to the sum of counts observed for bins that are not filtered out. The resulting test profile often centers around a value of 1, except in areas of maternal deletions or duplication, and areas in which the fetus is triploid (see FIG. 9). The bin-wise normalized profile illustrated in FIG. 10 confirms the validity of the normalization procedure, and clearly reveals the heterozygous maternal deletion (e.g., central dip in the gray segment of the profile tracing) in chromosome 18 and the elevated chromosomal representation of chromosome 18 of the tested sample (see the gray area of profile tracing in FIG. 10). As can be seen from FIG. 10, the median value for the gray segment of the tracing centers around about 1.1, where the median value for the black segment of the tracing centers around 1.0.

Peak Elevation

FIG. 11 graphically illustrates the results of analyzing multiple samples using bin-wise normalization, from a patient with a discernible feature or trait (e.g., maternal duplication, maternal deletion, the like or combinations thereof). The identities of the samples often can be determined by comparing their respective normalized count profiles. In the example illustrated in FIG. 11, the location of the dip in the normalized profile and its elevation, as well as its rarity, indicate that both samples originate from the same patient. Forensic panel data often can be used to substantiate these findings.

FIG. 12 and 13 graphically illustrate the results of the use of normalized bin profiles for identifying patient identity, or sample identity. The samples analyzed in FIG. 12 and 13 carry wide maternal aberrations in chromosomes 4 and 22, which are absent in the other samples in the profile tracings, confirming the shared origin of the top and bottom traces. Results such as this can lead

to the determination that a particular sample belongs to a specific patient, and also can be used to determine if a particular sample has already been analyzed.

Bin-wise normalization facilitates the detection of aberrations, however, comparison of peaks from different samples often is further facilitated by analyzing quantitative measures of peak elevations and locations (e.g., peak edges). The most prominent descriptor of a peak often is its elevation, followed by the locations of its edges. Features from different count profiles often can be compared using the following non-limiting analysis.

(a) Determine the confidence in a features detected peaks in a single test sample. If the feature is distinguishable from background noise or processing artifacts, the feature can be further analyzed against the general population.
(b) Determine the prevalence of the detected feature in the general population. If the feature is rare, it can be used as a marker for rare aberrations. Features that are found frequently in the general population are less useful for analysis. Ethnic origins can play a role in determining the relevance of a detected features peak elevation. Thus, some features provide useful information for samples from certain ethnic origins.
(c) Derive the confidence in the comparison between features observed in different samples.

Illustrated in FIG. 14 are the normalized bin counts in chromosome 5, from a euploid subject. The average elevation generally is the reference baseline from which the elevations of aberrations are measured, in some embodiments. Small and/or narrow deviations are less reliable predictors than wide, pronounced aberrations. Thus, the background noise or variance from low fetal contribution and/or processing artifacts is an important consideration when aberrations are not large or do not have a significant peak elevation above the background. An example of this is presented in FIG. 15, where a peak that would be significant in the upper trace, can be masked in the background noise observed in the bottom profile trace. The confidence in the peak elevation (see FIG. 16) can be determined by the average deviation from the reference (shown as the delta symbol), relative to the width of the euploid distribution (e.g., combined with the variance (shown as the sigma symbol) in the average deviation). The error in the average stretch elevation can be derived from the known formula for the error of the mean. If a stretch longer than one bin is treated as a random (non-contiguous) sample of all bins within a chromosome, the error in the average elevation decreases with the square root of the number of bins within the aberration. This reasoning neglects

the correlation between neighboring bins, an assumption confirmed by the correlation function shown in FIG. 17 (e.g., the equation for G(n)). Non-normalized profiles sometimes exhibit strong medium-range correlations (e.g., the wavelike variation of the baseline), however, the normalized profiles smooth out the correlation, leaving only random noise. The close match between the standard error of the mean, the correction for autocorrelation, and the actual sample estimates of the standard deviation of the mean elevation in chromosome 5 (see FIG. 18) confirms the validity of the assumed lack of correlation. Z-scores (see FIG. 19) and p-values calculated from Z-scores associated with deviations from the expected elevation of 1 (see FIG. 20) can then be evaluated in light of the estimate for uncertainty in the average elevation. The p-values are based on a t-distribution whose order is determined by the number of bins in a peak. Depending on the desired level of confidence, a cutoff can suppress noise and allow unequivocal detection of the actual signal.

$$Z = \frac{\Delta_1 - \Delta_2}{\sqrt{\sigma_1^2\left(\frac{1}{N_1}+\frac{1}{n_1}\right)+\sigma_2^2\left(\frac{1}{N_2}+\frac{1}{n_2}\right)}} \quad (1)$$

Equation 1 can be used to directly compare peak elevation from two different samples, where N and n refer to the numbers of bins in the entire chromosome and within the aberration, respectively. The order of the t-test that will yield a p-value measuring the similarity between two samples is determined by the number of bins in the shorter of the two deviant stretches.

Peak Edge

In addition to comparing average elevations of aberrations in a sample, the beginning and end of the compared stretches also can provide useful information for statistical analysis. The upper limit of resolution for comparisons of peak edges often is determined by the bin size (e.g., 50 kbps in the examples described herein). FIG. 21 illustrates 3 possible peak edge scenarios; (a) a peak from one sample can be completely contained within the matching peak from another sample, (b) the edges from one sample can partially overlap the edges of another sample, or (c) the leading edge from one sample can just marginally touch or overlap the trailing edge of another sample. FIG. 22 illustrates and example of the scenario described in (c) (e.g., see the middle, light gray trace, where the trailing edge of the middle trace marginally touches the leading edge of the upper trace).

The lateral tolerance associated with an edge often can be used to distinguish random variations from true, aberration edges. The position and the width of an edge can be quantified by numerically evaluating the first derivative of the aberrant count profile, as shown in FIG. 23. If the aberration is represented as a composite of two Heaviside functions, its derivative will be the sum of two Dirac's delta functions. The starting edge corresponds to an upward absorption-shaped peak, while the ending edge is a downward, 180 degree-shifted absorption peak. If the aberration is narrow, the two spikes are close to one another, forming a dispersion-like contour. The locations of the edges can be approximated by the extrema of the first derivative spikes, while the edge tolerance is determined by their widths.

Comparison between different samples often can be reduced to determining the difference between two matching edge locations, divided by the combined edge uncertainties. However, the derivatives sometimes are lost in background noise, as illustrated in FIG. 24. While the aberration itself benefits from the collective information contributed from all its bins, the first derivative only can afford information from the few points at the edge of the aberration, which can be insufficient to overcome the noise. Sliding window averaging, used to create FIG. 24, is of limited value in this situation. Noise can be suppressed by combining the first derivative (e.g., akin to a point estimate) with the peak elevation (e.g., comparable to an integral estimate). In some embodiments the first derivative and the peak elevation can be combined by multiplying them together, which is equivalent to taking the first derivative of a power of the peak elevation, as shown in FIG. 25. The results presented in FIG. 25 successfully suppress noise outside of the aberration, however, noise within the aberration is enhanced by the manipulation. The first derivative peaks are still clearly discernible, allowing them to be used to extract edge locations and lateral tolerances, thereby allowing the aberration to be clearly identified in the lower profile tracing.

Median chromosomal elevation

The median normalized elevation within the target chromosome in a euploid patient is expected to remain close to 1 regardless of the fetal fraction. However, as shown in FIG. 9 and 10, median elevations in trisomy patients increase with the fetal fraction. The increase generally is substantially linear with a slope of 0.5. Experimental measurements confirm these expectations. FIG. 26 illustrates a histogram of median elevations for 86 euploid samples (shown in dotted bars in FIG. 26). The median values are tightly clustered around 1 (median = 1.0000, median absolute deviation (MAD) = 0.0042, mean = 0.9996, standard deviation (SD) = 0.0046). None of the euploid

median elevations exceeds 1.012, as shown in the histogram presented in FIG. 26. In contrast, out of 35 trisomy samples shown (hatched bars) in Fig 26, all but one have median elevations exceeding 1.02, significantly above the euploid range. The gap between the two groups of patients in this example is large enough to allow classification as euploid or aneuploid.

Fetal fraction as the limiting factor in classification accuracy

The ratio between the fetal fraction and the width of the distribution of median normalized counts in euploids (e.g., euploid pregnancies) can be used to determine the reliability of classification using median normalized elevations, in some embodiments. Since median normalized counts, as well as other descriptors such as Z-values, linearly increase with the fetal fraction with the proportionality constant of 0.5, the fetal fraction must exceed four standard deviations of the distribution of median normalized counts to achieve 95% confidence in classification, or six standard deviations to achieve 99% confidence in classification. Increasing the number of aligned sequences tags can serve to decrease the error in measured profiles and sharpen the distribution of median normalized elevations, in certain embodiments. Thus, the effect of increasingly precise measurements is to improve the ratio between fetal fraction and the width of the distribution of euploid median normalized elevations.

Area Ratio

The median of the distribution of normalized counts generally is a point estimate and, as such, often is a less reliable estimate than integral estimates, such as areas under the distribution (e.g., area under the curve. Samples containing high fetal level fractions are not as affected by using a point estimate, however at low fetal fraction values, it becomes difficult to distinguish a truly elevated normalized profile from a euploid sample that has a slightly increased median count due to random errors. A histogram illustrating the median distribution of normalized counts from a trisomy case with a relatively low fetal fraction (e.g., F = about 7%; F(7%)) is shown in FIG. 27. The median of the distribution is 1.021, not far from 1 + F/2 = 1.035. However, the width of the distribution (MAD = 0.054, SD = 0.082) far exceeds the deviation of the median from the euploid value of 1, precluding any claims that the sample is abnormal. Visual inspection of the distribution suggests an alternative analysis: although the shift of the peak to the right is relatively small, it significantly perturbs the balance between the areas to the left (backward slashed) and to the right (forward slashed) from the euploid expectation of 1. Thus the ratio between the two areas, being

an integral estimate, can be advantageous in cases where classification is difficult due to low fetal fraction values. Calculation of the integral estimate for the forward slashed and backward slashed areas under the curve is explained in more detail below.

If a Gaussian distribution of normalized counts is assumed, then

$$P(q) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-{(q-q_0)}/{(2\sigma^2)}\right] \quad (2).$$

In euploid cases, the expectation for the normalized counts is 1. For trisomy patients, the expectation is

$$q_0 = 1 + {F}/{2} \quad (3).$$

Since the reference point for calculating the area ratio is 1, the argument to the exponential function is z^2, where

$$z = -{F}/{(2\sigma\sqrt{2})} \quad (4).$$

The area to the left of the reference point is

$$B = \int_{-\infty}^{1} P(q)dq = \frac{1}{2}[1 + erf(z)] \quad (5).$$

The error function erf(z) can be evaluated using its Taylor expansion:

$$erf(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!\,(2n+1)} \quad (6).$$

The area to the right from the reference point is 1 – B. The ratio between two areas is therefore

$$R = \frac{1-B}{B} = \frac{1-erf(z)}{1+erf(z)} = \frac{1-erf\left[-{F}/{(2\sigma\sqrt{2})}\right]}{1+erf\left[-{F}/{(2\sigma\sqrt{2})}\right]} \quad (7).$$

Error propagation from measured fetal fractions into area ratios R can be estimated by simply replacing F in equation 7 with F – ΔF and F + ΔF. FIG. 28 shows the frequencies of euploid and trisomy area ratios in a set of 480 samples. The overlap between two groups involves trisomy samples with low fetal fractions.

Combined classification criteria

FIG. 29 illustrates the interrelation and interdependence of median elevations and area ratios, both of which described substantially similar phenomena. Similar relationships connect median elevations and area ratios with other classification criteria, such as Z-scores, fitted fetal fractions, various sums of squared residuals, and Bayesian p-values (see FIG. 30). Individual classification criteria can suffer from ambiguity stemming from partial overlap between euploid and trisomy distributions in gap regions, however, a combination of multiple criteria can reduce or eliminate any ambiguities. Spreading the signal along multiple dimensions can have the same effect as measuring NMR frequencies of different nuclei, in some embodiments, resolving overlapping peaks into well-defined, readily identifiable entities. Since no attempt is made to quantitatively predict any theoretical parameter using mutually correlated descriptors, the cross-correlations observed between different classification criteria do not interfere. Defining a region in multidimensional space that is exclusively populated by euploids, allows classification of any sample that is located outside of the limiting surface of that region. Thus the classification scheme is reduced to a consensus vote for euploid.

In some embodiments utilizing a combined classification criteria approach, classification criteria described herein can be combined with additional classification criteria known in the art. Certain embodiments can use a subset of the classification criteria listed here. Certain embodiments can mathematically combine (e.g., add, subtract, divide, multiply, and the like) one or more classification criteria among themselves and/or with fetal fraction to derive new classification criteria. Some embodiments can apply principal components analysis to reduce the dimensionality of the multidimensional classification space. Some embodiments can use one or more classification criteria to define the gap between affected and unaffected patients and to classify new data sets. Any combination of classification criteria can be used to define the gap between affected and unaffected patients and to classify new data sets. Non-limiting examples of classification criteria that can be used in combination with other classification criteria to define the gap between affected and unaffected patients and to classify new data sets include: linear discriminant analysis, quadratic discriminant analysis, flexible discriminant analysis, mixture discriminant analysis, k Nearest Neighbors, classification tree, bagging, boosting, neural networks, support vector machines, and/or random forest.

Example 2: Methods for detection of genetic variations associated with fetal aneuploidy using measured fetal fractions and bin-weighted sums of squared residuals

Z-value statistics and other statistical analysis of sequence read data frequently are suitable for determining, or providing an outcome determinative of, the presence or absence of a genetic variation with respect to fetal aneuploidy, however, in some instances it can be useful to include additional analysis based on fetal fraction contribution and ploidy assumptions. When including fetal fraction contribution in a classification scheme, a reference median count profile from a set of known euploids (e.g., euploid pregnancies) generally is utilized for comparison. A reference median count profile can be generated by dividing the entire genome into *N* bins, where *N* is the number of bins. Each bin *i* is assigned two numbers: (i) a reference count F_i and (ii) the uncertainty (e.g., standard deviation or σ) for the bin reference counts.

The following relationship can be utilized to incorporate fetal fraction, maternal ploidy, and median reference counts into a classification scheme for determining the presence or absence of a genetic variation with respect to fetal aneuploidy,

$$Y_i = (1 - F)M_i f_i + FXf_i \qquad (8)$$

where Y_i represents the measured counts for a bin in the test sample corresponding to the bin in the median count profile, *F* represents the fetal fraction, *X* represents the fetal ploidy, and M_i represents maternal ploidy assigned to each bin. Possible values used for *X* in equation (8) are: 1 if the fetus is euploid; 3/2, if the fetus is triploid; and, 5/4, if there are twin fetuses and one is affected and one is not. 5/4 is used in the case of twins where one fetus is affected and the other not, because the term *F* in equation (8) represents total fetal DNA, therefore all fetal DNA must be taken into account. In some embodiments, large deletions and/or duplications in the maternal genome can be accounted for by assigning maternal ploidy, M_i, to each bin or genomic section. Maternal ploidy often is assigned as a multiple of 1/2, and can be estimated using bin-wise normalization, in some embodiments. Because maternal ploidy often is a multiple of 1/2, maternal ploidy can be readily accounted for, and therefore will not be included in further equations to simplify derivations.

Fetal ploidy can be assessed using any suitable approach. In some embodiments, fetal ploidy can be assessed using equation (8), or derivations thereof. In certain embodiments, fetal ploidy can be classified using one of the following, equation (8) based, non-limiting approaches:

1) Measure fetal fraction F and use the value to form two sums of squared residuals. To calculate the sum of squared residuals, subtract the right hand side (RHS) of equation (8) from its left hand side (LHS), square the difference, and sum over selected genomic bins, or in those embodiments using all bins, sum over all bins. This process is performed to calculate each of the two sums of squared residuals. One sum of square residuals is evaluated with fetal ploidy set to 1 (e.g., $X = 1$) and the other sum of squared residuals is evaluated with fetal ploidy set to 3/2 (e.g., $X = 3/2$). If the fetal test subject is euploid, the difference between the two sums of squared residuals is negative, otherwise the difference is positive.

2) Fix fetal fraction at its measured value and optimize ploidy value. Fetal ploidy generally can take on only 1 of two discrete values, 1 or 3/2, however, the ploidy sometimes can be treated as a continuous function. Linear regression can be used to generate an estimate for ploidy. If the estimate resulting from linear regression analysis is close to 1, the fetal test sample can be classified as euploid. If the estimate is close to 3/2, the fetus can be classified as triploid.

3) Fix fetal ploidy and optimize fetal fraction using linear regression analysis. The fetal fraction can be measured and a restraint term can be included to keep the fitted fetal fraction close to the measured fetal fraction value, with a weighting function that is reciprocally proportional to the estimated error in the measure fetal fraction. Equation (8) is solved twice, once with ploidy set at 3/2, and once for fetal ploidy set to 1. When solving equation (8) with ploidy set to 1, the fetal fraction need not be fitted. A sum of square residuals is formed for each result and the sum of squared residuals subtracted. If the difference is negative, the fetal test subject is euploid. If the difference is positive, the fetal test subject is triploid.

The generalized approaches described in 1), 2) and 3) are described in further detail herein.

Fixed ploidy, fixed fetal fraction: sums of squared residuals

In some embodiments, fetal aneuploidy can be determined using a model which analyzes two variables, fetal ploidy (e.g., *X*) and fetal nucleic acid fraction (e.g., fetal fraction; *F*). In certain embodiments, fetal ploidy can take on discrete values, and in some embodiments, fetal fraction can be a continuum of values. Fetal fraction can be measured, and the measured valued used to generate a result for equation (8), for each possible value for fetal ploidy. Fetal ploidy values that can be used to generate a result for equation (8) include 1and 3/2 for a single fetus pregnancy, and in the case of a twin fetus pregnancy where one fetus is affected and the other fetus unaffected, 5/4 can be used. The sum of squared residuals obtained for each fetal ploidy value measures the success with which the method reproduces the measurements, in some embodiments. When evaluating equation (8) at $X = 1$, (e.g., euploid assumption), the fetal fraction is canceled out and the following equation results for the sum of squared residuals:

$$\varphi_E = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - f_i)^2 = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2} - 2\sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2} + \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2} = \Xi_{yy} - 2\Xi_{fy} + \Xi_{ff} \quad (9)$$

To simplify equation (9) and subsequent calculations, the following notion is utilized:

$$\Xi_{yy} = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2} \quad (10)$$

$$\Xi_{ff} = \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2} \quad (11)$$

$$\Xi_{fy} = \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2} \quad (12)$$

When evaluating equation (8) at $X = 3/2$ (e.g., triploid assumption), the following equation results for the sum of the squared residuals:

$$\varphi_T = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left(y_i - f_i - \frac{1}{2}F f_i\right)^2 = \Xi_{yy} - 2\Xi_{fy} + \Xi_{ff} + F(\Xi_{ff} - \Xi_{fy}) + \frac{1}{4}F^2\Xi_{ff} \quad (13)$$

The difference between equations (9) and (13) forms the functional result (e.g., phi) that can be used to test the null hypothesis (e.g., euploid, $X = 1$) against the alternative hypothesis (e.g., trisomy singleton, $X = 3/2$):

$$\varphi = \varphi_g - \varphi_r = F(\Xi_{fy} - \Xi_{ff}) - \frac{1}{4}F^2\Xi_{ff} \quad (14)$$

The profile of phi with respect to F is a parabola defined to the right of the ordinate (since F is greater than or equal to 0). Phi converges to the origin as F approaches zero, regardless of experimental errors and uncertainties in the model parameters.

In some embodiments, the functional Phi is dependent on the measured fetal fraction F with a negative second-order quadratic coefficient (see equation (14)). Phi's dependence on the measured fetal fraction would seem to imply a convex shape for both euploid and triploid cases. If this analysis were correct, trisomy cases would reverse the sign at high F values, however equation (12) depends on F. Combining equations (8) and (14), disregarding maternal ploidy, setting X = 3/2 and neglecting experimental errors, the equation for trisomy cases becomes:

$$\Xi_{fy} = \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2} = \sum_{i=1}^{N} \frac{f_i}{\sigma_i^2}[(1-F)f_i + FXf_i] = \left(1 + \frac{1}{2}F\right)\sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2} = \left(1 + \frac{1}{2}F\right)\Xi_{ff} \quad (15)$$

The relationship between equations (11) and (12) for triploids holds under ideal circumstances, in the absence of any measurement errors. Combining equations (14) and (15) results in the following expression, which often yields a concave parabola in triploid cases:

$$\varphi = F(\Xi_{fy} - \Xi_{ff}) - \frac{1}{4}F^2\Xi_{ff} = F\left[\left(1 + \frac{1}{2}F\right)\Xi_{ff} - \Xi_{ff}\right] - \frac{1}{4}F^2\Xi_{ff} = \frac{1}{4}F^2\Xi_{ff} \text{ (Trisomy)} \quad (16)$$

For euploids, equations (11) and (12) should have the same value, with the exception of measurement errors, which sometimes yields a convex parabola:

$$\varphi = F(\Xi_{fy} - \Xi_{ff}) - \frac{1}{4}F^2\Xi_{ff} = -\frac{1}{4}F^2\Xi_{ff} \quad \text{(Euploids)} \quad (17)$$

Simulated functional phi profiles for typical model parameter values are shown in FIG. 31, for trisomy (dashed line) and euploid (solid line, bottom) cases. FIG. 32 shows an example using actual data. In FIG. 31 and 32, data points below the abscissa generally represent cases classified as euploids. Data points above the abscissa generally represent cases classified as trisomy 21 (T21) cases. In FIG. 32, the solitary data point in the fourth quadrant (e.g., middle lower quadrant) is a twin pregnancy with one affected fetus. The data set utilized to generate FIG. 32 includes other affected twin samples as well, explaining the spread of T21 data points toward the abscissa.

Equations (9) and (10) often can be interpreted as follows: For triploids, the euploid model sometimes generates larger errors, implying that phi_E (see equation (9)) is greater than phi_T (see equation (13)). As a result, functional phi (see equation (7)) occupies the first quadrant (e.g., upper left quadrant). For euploids, the trisomy model sometimes generates larger errors, the rank of equations (2) and (6) reverses and functional phi (equation (7)) occupies in the fourth quadrant. Thus, in principle, classification of a sample as euploid or triploid sometimes reduces to evaluating the sign of phi.

In some embodiments, the curvature of the data points shown in FIG. 31 and 32 can be reduced or eliminated by replacing functional phi (equation (7)) with the square root of functional phi's absolute value, multiplied by its sign. The linear relationship generated with respect to *F* sometimes can improve separation between triploids and euploids at low fetal fraction values, as shown in FIG. 33. Linearizing the relationship with respect to *F* sometimes results in increase uncertainty intervals at low fetal fraction (e.g., *F*) values, therefore, the gains realized from this process are related to making visual inspection of the differences substantially easier; the gray area remains unchanged. Extension of the process to analysis of twin pregnancies is relatively straightforward. The reason used to generate equation (9) implies that in a twin pregnancy with one affected and one normal fetus, functional phi should reduce to zero, plus or minus experimental error, regardless of *F*. Twin pregnancies generally produce more fetal DNA than single pregnancies.

Optimized ploidy, fixed fetal fraction: linear regression

In certain embodiments, fetal aneuploidy can be determined using a model in which the fetal fraction is fixed at its measured value and ploidy is varied to optimize the sum of squared residuals. In some embodiments, the resulting fitted fetal fraction value can be used to classify a case as trisomy or euploid, depending on whether the value is close to 1, 3/2, or 5/4 in the case of twins.

Starting from equation (8), the sum of squared residuals can be formed as follows:

$$\varphi = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} [y_i - (1-F)M_i f_i - FXf_i]^2$$

$$= \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left[y_i^2 - 2(1-F)M_i f_i y_i - 2FXf_i y_i + (1-F)^2 M_i^2 f_i^2 + 2F(1-F)XM_i f_i^2 + F^2 X^2 f_i^2 \right]$$

(18)

To minimize phi as a function of X, the first derivative of phi with respect to X is generated, set equal to zero, and the resulting equation solved for X. The resulting expression is presented in equation (19).

$$\frac{1}{2}\left(\frac{d\varphi}{dX}\right) = 0 = XF^2 \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2} - F \sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} + F(1-F) \sum_{i=1}^{N} \frac{M_i f_i^2}{\sigma_i^2} \quad (19)$$

The optimal ploidy value sometimes is given by the following expression:

$$X = \frac{\sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^{N} \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}} \quad (20)$$

As noted previously, the term for maternal ploidy, M_i, can be omitted from further mathematical derivations. The resulting expression for X corresponds to the relatively simple, and often most frequently occurring, special case of when the mother has no deletions or duplications in the chromosome or chromosomes being evaluated. The resulting expression is presented in FIG. 21.

$$X = \frac{\Xi_{fy} - (1-F)\Xi_{ff}}{F\Xi_{ff}} = \frac{\Xi_{fy}}{F\Xi_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Xi_{fy}}{\Xi_{ff}} - 1\right) \quad (21)$$

Xi_{ff} and Xi_{fy} are given by equations (11) and (12), respectively. In embodiments where all experimental errors are negligible, solving equation (21) results in a value of 1 for euploids where $Xi_{ff} = Xi_{fy}$. In certain embodiments where all experimental errors are negligible, solving equation

(21) results in a value of 3/2 for triploids (see equation (15) for triploid relationship between Xi_{ff} and Xi_{fy}.

Optimized ploidy, fixed fetal fraction: error propagation

Optimized ploidy often is inexact due to various sources of error. Three, non-limiting examples of error sources include: reference bin counts f_i, measured bin counts y_i, and fetal fraction F. The contribution of the non-limiting examples of error will be examined separately.

Errors in measured fetal fractions: quality of fitted fetal fraction

Fetal fraction estimates based on the number of sequence tags mapped to the Y chromosome (e.g., Y-counts) sometimes show relatively large deviations with respect to FQA fetal fraction values (see FIG. 34). Z-values for triploid often also exhibit a relatively wide spread around the diagonal shown in FIG. 35. The diagonal line in FIG. 35 represents a theoretically expected increase of the chromosomal representation for chromosome 21 with increasing fetal fraction in trisomy 21 cases. Fetal fraction can be estimated using a suitable method. A non-limiting example of a method that can be utilized to estimate fetal fraction is the fetal quantifier assay (e.g., FQA). Other methods for estimating fetal fraction are known in the art. Various methods utilized to estimate fetal fraction sometimes also show a substantially similar spread around the central diagonal, as shown in FIG. 36-39. In FIG. 36, the deviations are substantially similar (e.g., negative at high F_0) to those observed in fitted fetal fraction (see equation (33)). In some embodiments, the slope of the linear approximation to the average chromosome Y (e.g., chromosome Y) fetal fraction (see the middle histogram in FIG. 36) in the range between 0% and 20% is about 3/4. In certain embodiments, the linear approximation for standard deviation (see FIG. 36, upper and lower histogram lines) is about $2/3 + F_0/6$. In some embodiments, fetal fraction estimates based on chromosome 21 (e.g., chromosome 21) are substantially similar to those obtained by fitting fetal fractions (see FIG. 37). Another qualitatively similar set of gender-based fetal fraction estimates is shown in FIG. 38. FIG. 39 illustrates the medians of normalized bin counts for T21 cases, which are expected to have a slope whose linear approximation is substantially similar to $1 + F_0/2$ (see gray line from origin to the midpoint of the top of the graph in FIG. 39).

FIG. 36-39 share the following common features:

a) slope not equal to 1(either greater or less than 1, depending on the method, with the exception of Z-values),
b) large spread fetal fraction estimation, and
c) the extent of spread increases with fetal fraction.

To account for these observations, errors in measured fetal fraction will be modeled using the formula $\Delta F = 2/3 + F_0/6$, in some embodiments.

Errors in measured fetal fractions: error propagation from measured fetal fractions to fitted ploidy

If the assumption is made that f_i and y_i are errorless, to simplify analysis, the measured fetal fraction F is composed of F_V (e.g., the true fetal fraction) and ΔF (e.g., the error in measured fetal fraction):

$$F = F_V + \Delta F \quad (22).$$

In some instances, uncertainties in fitted X values originate from errors in measured fetal fraction, F. Optimized values for X are given by equation (21), however the true ploidy value is given by X_V, where X_V =1 or 3/2. X_V varies discretely, whereas X varies continuously and only accumulates around X_V under favorable conditions (e.g., relatively low error).

Assuming again that f_i and y_i are errorless, equation (8) becomes:

$$y_i = (1 - F_V)M_i f_i + F_V X f_i \quad (23).$$

Combining equations (21) to (23) generates the following relationship between true ploidy X_V and the ploidy estimate X that includes the error ΔF. The relationship also includes the assumption that maternal ploidy equals 1 (e.g., euploid), and the term for maternal ploidy, M_i, is replaced by 1.

$$X = 1 + \frac{1}{F_V + \Delta F}\left\{\frac{\sum_{i=1}^{N}\frac{f_i}{\sigma_i^2}((1-F_V)f_i + F_V X_V f_i)}{\sum_{i=1}^{N}\frac{f_i^2}{\sigma_i^2}} - 1\right\} = 1 + \frac{F_V(X_V - 1)}{F_V + \Delta F} \quad (24)$$

In some instances, the term $X_V - 1$ is substantially identical to zero in euploids, and ΔF does not contribute to errors in X. In triploid cases, the error term does not reduce to zero (e.g., is not substantially identical to zero). Thus, in some embodiments, ploidy estimates can be viewed as a function of the error ΔF:

$$X = g(\Delta F) \quad (25)$$

Simulated profiles of fitted triploid X as a function of F_0 with fixed errors ΔF = plus or minus 0.2% are shown in Fig 40. Results obtained using actual data are shown in FIG. 41. The data points generally conform to the asymmetric trumpet-shaped contour predicted by equation (24).

Smaller fetal fractions often are qualitatively associated with larger ploidy errors. Underestimated fetal fraction sometimes is compensated by ploidy overestimates; overestimated fetal fraction often is linked to underestimates in ploidy. The effect frequently is stronger when fetal fraction is underestimated. This is consistent with the asymmetry seen in the graphs presented in FIG. 40 and 41, (e.g., as F decreases, the growth of the upper branch is substantially faster than the decay of the lower branch). Simulations with different levels of error in F follow the same pattern, with the extent of the deviations from X_V increasing with ΔF.

A probability distribution for X can be used to quantify these observations. In some embodiments, the distribution of ΔF can be used to derive the density function for X using the following expression:

$$f_Y(y) = \left| \frac{1}{g'(g^{-1}(y))} \right| f_X(g^{-1}(y)) \quad (26)$$

where,

$f_Y(y)$ is the unknown density function for $y = g(x)$

$f_X(x)$ is the given density function for x

$g'(x)$ is the first derivative of the given function $y = g(x)$

$g^{-1}(y)$ is the inverse of the given function g : $x = g^{-1}(y)$

$g'(g^{-1}(y))$ is the value of the derivative at the point $g^{-1}(y)$

In equation 26 x is ΔF, y is X (e.g., ploidy estimate), and $g(x)$ is given by equation (24). The derivative is evaluated according to the following expression:

$$\frac{dg}{d\Delta F} = -\frac{F_V(X_V-1)}{(F_V+\Delta F)^2} \quad (27)$$

The inverse $g^{-1}(y)$ can be obtained from equation (24), in some embodiments:

$$\Delta F = \frac{F_V(X_V-x)}{x-1} \quad (28)$$

If the error in *F* conforms to a Gaussian distribution, $f_x(x)$ in equation (26) can be replaced with the following expression:

$$P(\Delta F) = \frac{\exp\left[-(\Delta F)^2/(2\sigma^2)\right]}{\sigma\sqrt{2\pi}} \quad (29)$$

In certain embodiments, combining equations (26) to (29) results in a probability distribution for X at different levels of ΔF, as shown in FIG. 42.

In some instances, a bias towards higher ploidy values, which sometimes are prominent at high levels of errors in F, often is reflected in the asymmetric shape of the density function: a relatively long, slowly decaying tail to the right of the right vertical line, vertically in line with *X*, along the X axis, as shown in FIG. 42, panels A-C. In some embodiments, for any value of ΔF, the area under the probability density function to the left of the right vertical line (X_V = 3/2) equals the area to the right of the right vertical line. That is, one half of all fitted ploidy values often are overestimates, while the other half of all fitted ploidy values sometimes are underestimates. In some instances, the bias generally only concerns the extent of errors in *X*, not the prevalence of one or the other direction. The median of the distribution remains equal to X_V, in some embodiments. FIG. 43 illustrates euploid and trisomy distributions obtained for actual data. Uncertainties in measured fetal fraction values sometimes explain part of the variance seen in the fitted ploidy values for triploids, however errors in estimated X values for euploids often require examining error propagation from bin counts.

Fixed ploidy, optimized fetal fraction: linear regression

A continuously varying fetal fraction often can be optimized while keeping ploidy fixed at one of its possible values (e.g., 1 for euploids, 3/2 for singleton triploids, 5/4 for twin triploids), as opposed to

fitting ploidy that often can take on a limited number of known discrete values. In embodiments in which the measured fetal fraction (F_0) is known, optimization of the fetal fraction can be restrained such that the fitted F remains close to F_0, within experimental error (e.g., ΔF). In some instances, the observed (e.g., measured) fetal fraction F_0, sometimes differs from fetal fraction, F_V, described in equations (22) to (28). A robust error propagation analysis should be able to distinguish between F_0 and F_V. To simplify the following derivations, difference between the observed fetal fraction and the true fetal fraction will be ignored.

Equation (8) is presented below in a rearranged format that also omits the maternal ploidy term (e.g., M_i).

$$y_i = F(X-1)f_i + f_i \quad (30)$$

A functional term that needs to be minimized is defined as follows, in some embodiments:

$$\varphi(F) = \frac{(F-F_0)^2}{(\Delta F)^2} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}[y_i - F(X-1)f_i - f_i]^2$$

$$= \frac{(F-F_0)^2}{(\Delta F)^2} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[y_i^2 + F^2(X-1)^2 f_i^2 + f_i^2 - 2F(X-1)f_i y_i - 2f_i y_i + 2F(X-1)f_i^2\right]$$

$$= \frac{(F-F_0)^2}{(\Delta F)^2} + F^2(X-1)^2 \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2} + 2F(X-1)\sum_{i=1}^{N} \frac{f_i^2 - f_i y_i}{\sigma_i^2} + \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{\sigma_i^2} \quad (31)$$

When equation (31) is evaluated for euploids (e.g., $X = 1$), the term $\frac{(F-F_0)^2}{(\Delta F)^2}$ often depends on F, thus fitted F frequently equals F_0. In some instances, when equation (24) is evaluated for euploids, the equation sometimes reduces to $\sum_{i=1}^{N} \frac{(y_i - f_i)^2}{\sigma_i^2}$.

When equation (24) is evaluated for singleton trisomy cases (e.g., $X = 3/2$), the coefficients that multiply F contain both fetal fraction measurements and bin counts, therefore the optimized value for F often depends on both parameters. The first derivative of equation (24) with respect to F reduces to zero in some instances:

$$\frac{1}{2}\left(\frac{d\varphi}{dF}\right) = 0 = \frac{(F-F_0)}{(\Delta F)^2} + F(X-1)^2\sum_{i=1}^{N}\frac{f_i^2}{\sigma_i^2} + (X-1)\sum_{i=1}^{N}\frac{f_i^2 - f_i y_i}{\sigma_i^2} \quad (32)$$

In some embodiments, replacing $X = 3/2$ and solving equation (32) for F yields an optimized value for F:

$$F = \frac{F_0 + \frac{(\Delta F)^2}{2}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}(f_i y_i - f_i^2)}{1 + \frac{(\Delta F)^2}{4}\sum_{i=1}^{N} f_i^2/\sigma_i^2} \quad (33).$$

To simplify further calculations and/or derivations, the following auxiliary variables will be utilized:

$$S_0 = \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{1}{\sigma_i^2} \quad (34)$$

$$S_f = \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{f_i}{\sigma_i^2} \quad (35)$$

$$S_y = \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{y_i}{\sigma_i^2} \quad (36)$$

$$S_{yy} = \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{y_i^2}{\sigma_i^2} \quad (37)$$

$$S_{ff} = \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{f_i^2}{\sigma_i^2} \quad (38)$$

$$S_{fy} = \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{y_i f_i}{\sigma_i^2} \quad (39)$$

Utilizing the auxiliary variables, the optimized fetal fraction for $X = 3/2$ for equation (33) then reduces to:

$$F = \frac{F_0 + 2S_{fy} - 2S_{ff}}{1 + S_{ff}} \quad (40)$$

Fitted *F* often is linearly proportional to the measured value F_0, but sometimes is not necessarily equal to F_0. The ratio between errors in fetal fraction measurements and uncertainties in bin counts determines the relative weight given to the measured F_0 versus individual bins, in some embodiments. In some instances, the larger the error ΔF, the stronger the influence that bin counts will exert on the fitted *F*. Alternatively, small ΔF generally implies that the fitted value F will be dominated by F_0. In some embodiments, if a data set comes from a trisomy sample, and all errors are negligible, equation (40) reduces to identity between *F* and F_0. By way of mathematic proof, using fetal ploidy set to X = 3/2, and assuming that F_0 (observed) and F_V (true) have the same value, equation (30) becomes:

$$y_i = \frac{1}{2} F_0 f_i + f_i \quad (41)$$

The assumption that F_0 and F_V generally is an acceptable assumption for the sake of the qualitative analysis presented herein. Combing equations (39) and (41) yields

$$S_{fy} = \frac{(\Delta F)^2}{4} \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2} = \frac{(\Delta F)^2}{4} \sum_{i=1}^{N} \frac{\left(\frac{1}{2} F_0 f_i + f_i\right) f_i}{\sigma_i^2} = \left(\frac{1}{2} F_0 + 1\right) S_{ff} \quad (42)$$

Combining equations (40) and (42) results in identity between F_0 and F_V:

$$F = \frac{F_0 + 2S_{fy} - 2S_{ff}}{1 + S_{ff}} = \frac{F_0 + 2\left(\frac{1}{2} F_0 + 1\right) S_{ff} - 2S_{ff}}{1 + S_{ff}} = \frac{F_0 (1 + S_{ff})}{1 + S_{ff}} \equiv F_0 \quad \text{QED} \quad (43)$$

To further illustrate the theoretical model, if the true ploidy is 1 (e.g., euploid) but the ploidy value use in equation (40) is set to X = 3/2 (e.g., triploid singleton), the resulting fitted *F* does not equal F_0, nor does it reduce to zero, and the following expression generally is true:

$$y_i = f_i \Rightarrow S_{fy} = \frac{(\Delta F)^2}{4} \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2} = \frac{(\Delta F)^2}{4} \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2} = S_{ff} \Rightarrow F = \frac{F_0 + 2S_{fy} - 2S_{ff}}{1 + S_{ff}} = \frac{F_0}{1 + S_{ff}} \quad (44).$$

Thus, application of triploid equations when testing a euploid case generally results in a non-zero fitted F that is proportional to F_0 with a coefficient of proportionality between 0 and 1 (exclusive), depending on reference bin counts and associated uncertainties (cf. equation (38)), in certain embodiments. A similar analysis is shown in FIG. 44, using actual data from 86 know euploids as reference. The slope of the straight line from equation (44) is close to 20 degrees, as shown in FIG. 44.

The solitary data point between euploid and T21 cases (e.g., measured fetal fraction approximately 40%, fitted fraction approximately 20%) represents a T21 twin. When a constant ΔF is assumed the euploid branch of the graph shown in FIG. 44 generally is sloped, however when $\Delta F = 2/3 + F_0/6$ is used the euploid branch of the graph often becomes substantially horizontal, as described herein in the section entitled "Fixed ploidy, optimized fetal fraction, error propagation: fitted fetal fractions".

Fixed ploidy, optimized fetal fraction: sums of squared residuals

In some instances for euploid cases, were fitted F for equation (32) equals F_0 and X = 1, the sum of square residuals for a euploid model follows from equation (31):

$$\varphi_0 = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (y_i - f_i)^2 = \Xi_{yy} - 2\Xi_{fy} + \Xi_{ff} \quad (45)$$

which is substantially the same result as equation (9). In certain instances for euploid cases, equation (40) can be combined into equation (31). The resulting mathematical expression quadratically depends on F_0, in some embodiments. In certain embodiments, classification of a genetic variation is performed by subtracting the triploid sum of squared residuals from the euploid sum of squared residuals. The result of the classification obtained by subtracting the triploid sum of squared residuals from the euploid sum of squared residuals also frequently depends on F_0:

$$\varphi_E - \varphi_T$$

$$= \frac{-1}{(\Delta F)^2}\left[\left(\frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}} - F_0\right)^2 + \left(\frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}}\right)^2 \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}\frac{f_i^2}{\sigma_i^2} + \left(\frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}}\right)(\Delta F)^2\sum_{i=1}^{N}\frac{f_i^2-f_i y_i}{\sigma_i^2}\right]$$

$$= \frac{-1}{(\Delta F)^2}\left[\left(\frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}} - F_0\right)^2 + \left(\frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}}\right)^2 S_{ff} + 4\left(\frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}}\right)(S_{ff}-S_{fy})\right]$$

$$= \frac{-\left[(2S_{fy}-2S_{ff}-F_0S_{ff})^2+(F_0+2S_{fy}-2S_{ff})^2S_{ff}+4(F_0+2S_{fy}-2S_{ff})(1+S_{ff})(S_{ff}-S_{fy})\right]}{(\Delta F)^2(1+S_{ff})^2}$$

$$= \frac{-1}{(\Delta F)^2(1+S_{ff})^2}\left[\left(4S_{fy}^2+4S_{ff}^2+F_0^2S_{ff}^2-8S_{fy}S_{ff}-4F_0S_{fy}S_{ff}+4F_0S_{ff}^2\right)\right.$$

$$+\left(F_0^2S_{ff}+4S_{fy}^2S_{ff}+4S_{ff}^3+4F_0S_{fy}S_{ff}-4F_0S_{ff}^2-8S_{fy}S_{ff}^2\right)$$

$$+\left(4F_0S_{ff}+8S_{fy}S_{ff}-8S_{ff}^2-4F_0S_{fy}-8F_0S_{fy}+8S_{fy}S_{ff}\right.$$

$$\left.\left.+4F_0S_{ff}^2+8S_{fy}S_{ff}^2-8S_{ff}^3-4F_0S_{fy}S_{ff}-8S_{fy}^2S_{ff}+8S_{fy}S_{ff}^2\right)\right]$$

$$= \frac{-1}{(\Delta F)^2(1+S_{ff})}\left[F_0^2S_{ff}+4F_0(S_{ff}-S_{fy})-4(S_{ff}-S_{fy})^2\right]$$

(46)

The term S_{fy} generally depends on fetal fraction, as also seen for equation (14). The dependence of $\varphi_E - \varphi_T$ on the measured fetal fraction can be analyzed by accounting for the fetal fraction, in some embodiments. The fetal fraction often can be accounted for by assuming that measured fetal fraction F_0 equals true fetal fraction F_V. In some embodiments, if the sample's karyotype is euploid, S_{fy} and S_{ff} have the same values (e.g., with the exception of experimental errors). As a result, the difference between the two sums of squared residuals often reduces to:

$$\varphi_E - \varphi_T = \frac{-F_0^2S_{ff}}{(\Delta F)^2(1+S_{ff})} \qquad \text{(Euploids)} \qquad (47)$$

In certain embodiments, if the sample's karyotype is triploid, equations (41) and (42) can be combined with equation (46), yielding:

$$\varphi_E - \varphi_T = \frac{F_0^2S_{ff}}{(\Delta F)^2} \qquad \text{(Triploids)} \qquad (48)$$

Thus, if the difference of $\Phi_S - \Phi_T$ is positive, the fetus is triploid, in some embodiments, and in certain embodiments, if the difference is negative, the fetus is unaffected. The graphical representation for the positive or negative result frequently is a parabola; concave for triploids and convex for euploids. Both branches tend towards zero as F_0 decreases, with experimental error having little effect on the shape of the graph. Neither branch has a substantially linear or free term, but the second order coefficients differ in size in addition to having different signs, in many instances. With ΔF approximately 2%, the value of the term S_{ff} is close to 3.7, using the reference counts and uncertainties extracted from the 86 euploid set (see FIG. 45).

In the example shown in FIG. 45, the two branches often are asymmetric due to the different coefficients multiplying the square of the measured fetal fraction in equations (47) and (48). The triploid (e.g., positive) branch increases relatively quickly, becoming distinguishable from zero substantially earlier than the euploid branch. FIG. 46, obtained using a real data set, confirms the qualitative results shown in FIG. 45. In FIG. 46 the solitary dark gray point in the fourth quadrant (e.g., lower middle quadrant) is an affected twin. In the data set used to generate FIG. 46, both the euploid and T21 branches of the graph show curvature because both show quadratic dependence on F_0 from the trisomy version of equation (31)

In some embodiments, both branches of the graph can be linearized to facilitate visual inspection. The value of the linearization often is conditioned on the error propagation analysis. The results presented in FIG. 45 and 46 were based on the assumption that the error in measured fetal fractions is uniform the entire range of fetal fractions. However, the assumption is not always the case. In some instances, the more realistic assumption, based on a linear relationship between error ΔF and measured fetal fraction $F_0(\Delta F = 2/3 + F_0/6)$, produces the results presented in FIG. 47. In FIG. 47, the euploid branch is substantially flat, almost constant (e.g., the parabolic character is substantially lost), however, the trisomy branch remains parabolic. The three light gray points interspersed in the dark gray points of the trisomy branch represent data from twins. Twin data sometimes are elevated relative to the fixed error model.

Classification of whether or not a sample is affected by a genetic variation often is carried out using one of three processes: (1) classification based on parabolic differences of summed squares of residuals, (see FIG. 45 and 46), (2) classification based on linear differences of summed squares of residuals, (see FIG. 47 and 48), and (3) classification based on fitted fetal fraction (see equation (33)). In some embodiments, the chosen approach takes error propagation into account.

Fixed ploidy, optimized fetal fraction: systematic error – reference offset

Ideally, reference and measured bin counts should contain zero systematic error (e.g., offset), however, in practice, reference and measured bin counts sometimes are shifted with respect to one another. In some instances, the effect of the shift with respect to one another can be analyzed using equation (33), assuming the shift Δ is constant across the chromosome of interest. For euploid cases, if random errors are neglected, the following relationships hold, in some embodiments:

$$f_i = f_i^0 + \Delta \quad (49)$$

$$y_i = f_i^0 = f_i - \Delta \quad (50)$$

f_i^0 represents the true reference bin count i, and f_i represents the reference bin counts used, including any systematic error Δ. In certain embodiments, replacing equations (49) and (50) into equation (33) generates the following expression for the euploid branch of the fitted fetal fraction graph:

$$F_E = \frac{F_0 + \frac{(\Delta F)^2}{2}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}(f_i y_i - f_i^2)}{1 + \frac{(\Delta F)^2}{4}\sum_{i=1}^{N} f_i^2/\sigma_i^2} = \frac{F_0 + \frac{(\Delta F)^2}{2}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left[(f_i^0+\Delta)f_i^0 - (f_i^0+\Delta)^2\right]}{1 + \frac{(\Delta F)^2}{4}\sum_{i=1}^{N}(f_i^0+\Delta)^2/\sigma_i^2}$$

$$= \frac{F_0 - \frac{(\Delta F)^2}{2}\left(\Delta\sum_{i=1}^{N}\frac{f_i^0}{\sigma_i^2} + \Delta^2\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)}{1 + \frac{(\Delta F)^2}{4}\left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}(f_i^0)^2 + 2\Delta\sum_{i=1}^{N}\frac{f_i^0}{\sigma_i^2} + \Delta^2\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)} = \frac{F_0 - 2S_f^0\Delta - 2S_0^0\Delta^2}{1 + S_{ff}^0 + 2S_f^0\Delta + S_0^0\Delta^2} \quad (51)$$

The coefficients S_0^0, S_f^0 and S_{ff}^0, are generated from equations (33) to (39) by replacing f_i with f_i^0, in some embodiments. In certain embodiments, the reciprocal slope of the linear functional relationship between fitted euploid value F_E and measured F_0 equals $1 + S_{ff}^0 + 2\,S_f^0\,\Delta + S_0^0\,\Delta^2$, which often allows estimation of the systematic error Δ by solving a relatively simple quadratic equation. For triploids, assuming F_0 equals F_V, measured bin counts sometimes become:

$$y_i = f_i^0 + \frac{1}{2} F_0 f_i^0 \qquad (52)$$

Combining equations (52), (49) and (33) generates the following expression for the triploid branch of the fitted fetal fraction graph:

$$F_T = \frac{F_0 + \frac{(\Delta F)^2}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (r_i x_i - f_i^0)}{1 + \frac{(\Delta F)^2}{4} \sum_{i=1}^{N} f_i^0 / \sigma_i^2} = \frac{F_0 + \frac{(\Delta F)^2}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left[(f_i^0 + \Delta)\left(f_i^0 + \frac{1}{2} F_0 f_i^0\right) - (f_i^0 + \Delta)^2 \right]}{1 + \frac{(\Delta F)^2}{4} \sum_{i=1}^{N} (f_i^0 + \Delta)^2 / \sigma_i^2}$$

$$= \frac{F_0 + \frac{(\Delta F)^2}{2} \left(\frac{1}{2} F_0 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (f_i^0)^2 + \frac{1}{2} F_0 \Delta \sum_{i=1}^{N} \frac{f_i^0}{\sigma_i^2} - \Delta \sum_{i=1}^{N} \frac{f_i^0}{\sigma_i^2} - \Delta^2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)}{1 + \frac{(\Delta F)^2}{4} \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} (f_i^0)^2 + 2\Delta \sum_{i=1}^{N} \frac{f_i^0}{\sigma_i^2} + \Delta^2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)} = \frac{F_0 \left(1 + S_f^0 f_i^0 + S_f^0 \Delta\right) - S_f^0 \Delta - S_0^0 \Delta^2}{1 + S_f^0 f_i^0 + 2 S_f^0 \Delta + S_0^0 \Delta^2} \qquad (53)$$

In some embodiments, equations (51) and (53) predict that fitted triploid and euploid fetal fractions will behave as shown in FIG. 48. In FIG. 48 black lines (e.g., upper lines in each set of 3 lines) correspond to negative offset Δ, dark gray lines (e.g., bottom lines in each set of 3 lines) correspond to positive offset Δ, and light gray lines (e.g., middle lines in each set of 3 lines), correspond to the absence of offset. FIG. 49 illustrates the effects of simulated systematic errors Δ artificially imposed on actual data.

FIG. 50 illustrates the dependence of fitted fetal fraction on systematic error offset for euploid and triploid data sets. For both euploid and triploid cases, the theoretical expressions of equations (51) and (53) often capture the qualitative dependence of fitted fetal fraction on measured fetal fraction and on systematic error offset. Coefficients used for the graphs in FIG. 49 and 50 were obtained from raw reference bin counts, without removing any potential systematic bias.

Fixed ploidy, optimized fetal fraction, error propagation: fitted fetal fraction

Contributions to errors in fitted fetal fractions often fall into one of two types of errors: 1) from measured fetal fractions, and 2) from measured and reference bin counts. The two types of errors will be analyzed separately, using different approaches, and later combined to generate final error ranges. Errors propagated from measure fetal fractions can be evaluated by replacing F_0 in equation (40) first with $F_0 - 2\Delta F$ (e.g., for the lower error boundary) and then with $F_0 + 2\Delta F$ (e.g., for

the upper error boundary). This relatively simple approach produces correct qualitative behavior at 95% confidence intervals, in certain embodiments. For a different desired level of confidence, a more general pair of bounds, $F_0 - n\Delta F$ and $F_0 + n\Delta F$, can be utilized. The terms used to generate upper and lower error boundaries sometimes underestimates the total error because the contributions from errors in measure and reference bin counts often are neglected.

To better assess the contribution from measured and reference bin counts on error in fitted fetal fraction, equations (38) to (40) can be utilized, in some embodiments. In certain embodiments, equation (33) can be expanded for fitted fetal fraction into a Taylor series with respect to f_i and y_i, truncated to the first order, square and average. In some instances, it can be assumed that uncertainties in y_i often are the same as uncertainties in f_i. To simply analysis, cross-terms and higher-order terms are assumed to reduce to zero upon averaging. Taylor expansion coefficients often are obtained utilizing the chain rule. The mean squared variation in the fitted fetal fraction is then given by equation (54) shown below. The model represented by equation ignores contributions from estimates for ΔF, in some embodiments. Partial derivatives can be evaluated using the expressions presented below equation (54).

$$(\partial F)^2 = \sum_{i=1}^{N} \left(\frac{\partial F}{\partial f_i}\right)^2 \sigma_i^2 + \sum_{i=1}^{N} \left(\frac{\partial F}{\partial y_i}\right)^2 \sigma_i^2$$

$$= \sum_{i=1}^{N} \left[\left(\frac{\partial F}{\partial s_{ff}}\right)\left(\frac{\partial s_{ff}}{\partial f_i}\right) + \left(\frac{\partial F}{\partial s_{fy}}\right)\left(\frac{\partial s_{fy}}{\partial f_i}\right)\right]^2 \sigma_i^2 + \sum_{i=1}^{N} \left[\left(\frac{\partial F}{\partial s_{fy}}\right)\left(\frac{\partial s_{fy}}{\partial y_i}\right)\right]^2 \sigma_i^2 \quad (54)$$

$$\left(\frac{\partial F}{\partial s_{ff}}\right) = -\frac{F_0 + 2s_{fy} + 2}{(1 + s_{ff})^2} \quad (55)$$

$$\left(\frac{\partial F}{\partial s_{fy}}\right) = \frac{2}{1 + s_{ff}} \quad (56)$$

$$\left(\frac{\partial s_{ff}}{\partial f_i}\right) = \frac{(\Delta F)^2}{2}\left(\frac{f_i}{\sigma_i^2}\right) \quad (57)$$

$$\left(\frac{\partial s_{fy}}{\partial f_i}\right) = \frac{(\Delta F)^2}{4}\left(\frac{y_i}{\sigma_i^2}\right) \quad (58)$$

$$\left(\frac{\partial S_{fy}}{\partial y_i}\right) = \frac{(\Delta F)^2}{4}\left(\frac{f_i}{\sigma_i^2}\right) \quad (59)$$

Combining equations (54) to (59) generates the following expression:

$$(\delta F)^2 = \left[\frac{(\Delta F)^2}{4}\right]^2 \left\{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[\frac{2y_i}{1+S_{ff}} - 2f_i\frac{F_0+2S_{fy}+2}{(1+S_{ff})^2}\right]^2 + \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left(\frac{2f_i}{1+S_{ff}}\right)^2\right\}$$

$$= \left[\frac{(\Delta F)^2}{4}\right]^2 \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left[\left(\frac{2y_i}{1+S_{ff}}\right)^2 - 8f_iy_i\frac{F_0+2S_{fy}+2}{(1+S_{ff})^3} + 4f_i^2\frac{(F_0+2S_{fy}+2)^2}{(1+S_{ff})^4} + \left(\frac{2f_i}{1+S_{ff}}\right)^2\right]$$

$$= \left[\frac{(\Delta F)^2}{4}\right]^2 \left\{\frac{4}{(1+S_{ff})^2}\sum_{i=1}^{N}\frac{y_i^2}{\sigma_i^2} - 8\frac{F_0+2S_{fy}+2}{(1+S_{ff})^3}\sum_{i=1}^{N}\frac{f_iy_i}{\sigma_i^2} + 4\left[\frac{(F_0+2S_{fy}+2)^2}{(1+S_{ff})^4} + \frac{1}{(1+S_{ff})^2}\right]\sum_{i=1}^{N}\frac{f_i^2}{\sigma_i^2}\right\}$$

$$= (\Delta F)^2\left\{\frac{S_{yy}}{(1+S_{ff})^2} - 2S_{fy}\frac{F_0+2S_{fy}+2}{(1+S_{ff})^3} + S_{ff}\left[\frac{(F_0+2S_{fy}+2)^2}{(1+S_{ff})^4} + \frac{1}{(1+S_{ff})^2}\right]\right\}$$

(60)

To evaluate equation (60) at a 95% confidence interval, the following upper and lower bounds can be used, in some embodiments:

$$\begin{bmatrix}F_{Lower}\\F_{Upper}\end{bmatrix} = \frac{F_0+2S_{fy}-2S_{ff}}{1+S_{ff}} + \begin{bmatrix}-2\\2\end{bmatrix}\Delta F\left\{\frac{1}{1+S_{ff}} + \sqrt{\frac{S_{yy}}{(1+S_{ff})^2} - 2S_{fy}\frac{F_0+2S_{fy}+2}{(1+S_{ff})^3} + S_{ff}\left[\frac{(F_0+2S_{fy}+2)^2}{(1+S_{ff})^4} + \frac{1}{(1+S_{ff})^2}\right]}\right\}$$

(61)

In embodiments in which substantially all possible sources of error (e.g., F_0, f_i, y_i) are included in the Taylor expansion series, the same equation often is obtained. In some instances, dependence of F on F_o, can be accounted for through S_{fy}. In some embodiments, power series terms corresponding to F_0 often take the form;

$\left[\left(\frac{\partial F}{\partial F_0}\right)+\left(\frac{\partial F}{\partial S_{fy}}\right)\left(\frac{\partial S_{fy}}{\partial F_0}\right)\right]^2 (\Delta F)^2$, but $\left[\left(\frac{\partial F}{\partial F_0}\right)+\left(\frac{\partial F}{\partial S_{fy}}\right)\left(\frac{\partial S_{fy}}{\partial F_0}\right)\right]^2$ equals 1 for triploids. Thus, relatively simple subtraction and addition of ΔF to F_0 often is justified, even though ΔF often increases with F_0 and becomes large at high F_0. The outcome is due to both F and S_{fy} depending linearly on F_0, in some embodiments. Simulations based on equation (61) are shown in FIG. 51, along with fitted fetal fractions obtained from test subject derived data. In the simulations presented in Fig, 51, $\Delta F = 2/3 + F_0/6$, as described herein.

Example 3: Sliding window analysis and cumulative sums as a function of genomic position

Identification of recognizable features (e.g., regions of genetic variation, regions of copy number variation) in a normalized count profile sometimes is a relatively time consuming and/or relatively expensive process. The process of identifying recognizable features often is complicated by data sets containing noisy data and/or low fetal nucleic acid contribution. Identification of recognizable features that represent true genetic variations or copy number variations can help avoid searching large, featureless regions of a genome. Identification of recognizable features can be achieved by removing highly variable genomic sections from a data set being searched and obtaining, from the remaining genomic sections, data points that deviate from the mean profile elevation by a predetermined multiple of the profile variance.

In some embodiments, obtaining data points that deviate from the mean profile elevation by a predetermined multiple of the profile variance can be used to reduce the number of candidate genomic sections from greater than 50,000 or 100, 000 genomic sections to in the range of about 100 to about 1000 candidate genomic sections that represent true signals or solitary noise spikes (e.g., about 100 genomic sections, about 200 genomic sections, about 300 genomic sections, about 400 genomic sections, about 500 genomic sections, about 600 genomic sections, about 700 genomic sections, about 800 genomic sections, about 900 genomic sections, or about 1000 genomic sections). The reduction in the number of candidate genomic sections can be achieved relatively quickly and easily and often speeds up the search for and/or identification of genetic aberrations by two or more orders of magnitude. Reduction in the number of genomic sections searched for the presence or absence of candidate regions of genomic variation often reduces the complexity and/or dimensionality of a data set.

After a reduced data set containing data points that deviate from the mean profile elevation by a predetermined multiple of the profile variance is generated, the reduced data set is filtered to

eliminate solitary noise spikes, in some embodiments. Filtering a reduced data set to remove solitary noise spikes often generates a filtered, reduced data set. In some embodiments, a filtered, reduced data set retains contiguous clusters of data points, and in certain embodiments, a filtered, reduced data set retains clusters of data points that are largely contiguous with allowance for a predetermined number and/or size of gaps. Data points from the filtered, reduced data set that deviate from the average profile elevation in substantially the same direction are grouped together, in some embodiments.

Due to the background noise often present in nucleic acid samples (e.g., ratio of regions of interest compared to the total nucleic acid in a sample), distinguishing regions of genetic variation or genetic aberration from background noise often is challenging. Methods that improve the signal-to-noise ratio often are useful for facilitating the identification of candidate regions representative of regions of true genetic variation and/or genetic aberration. Any method that improves the signal-to-noise ratio of regions of true genetic variation with respect to the genomic background noise can be used. A non-limiting example of a method suitable for use in improving the signal-to-noise ratio of regions of true genetic variation with respect to the genomic background noise is the use of integrals over the suspected aberration and its immediate surroundings. In some embodiments, the use of integrals over the suspected aberration and its immediate surroundings is beneficial, because summation cancel out random noise. After noise has been reduced or eliminated, even relatively minor signals can become readily detectable using a cumulative sum of the candidate peak and its surroundings, in some embodiments. A cumulative sum sometimes is defined with respect to an arbitrarily chosen origin outside (e.g., on one side or the other) of the peak. A cumulative sum often is a numerical estimate of the integral of the normalized count profile over the selected genetic section or sections.

In the absence of aberrations, the cumulative sum as a function of the genomic position often behaves as a straight line with unit slope (e.g., slope equal to 1). If deletions or duplications are present, the cumulative sum profile often consists of two or more line segments. In some embodiments, areas outside of aberrations map to line segments with unit slopes. For areas within aberrations, the line segments are connected by other line segments whose slopes equal the count profile elevation or depression within the aberration, in certain embodiments.

In those samples having maternal aberrations, the slopes (e.g., equivalent to the count profile elevation) are relatively easily determined: 0 for homozygous maternal deletions, 0.5 for

heterozygous maternal deletions, 1.5 for heterozygous maternal duplications, 2.0 for homozygous maternal duplications. In those samples having fetal aberrations, the actual slopes depend both on the type of the aberration (e.g., homozygous deletion, heterozygous deletion, homozygous duplication or heterozygous duplication) and on the fetal fraction. In some embodiments, inheritance of a maternal aberration by the fetus also is taken into account when evaluating fetal samples for genetic variations.

In some embodiments, line segments with unit slopes, corresponding to normal genomic areas to the left and to the right of an aberration, are vertically shifted with respect to one another. The difference (e.g., subtractive result) between their intercepts equals the product between the width of the aberration (number of affected genomic sections) and the aberration level (e.g., -1 for homozygous maternal deletion, -0.5 for heterozygous maternal deletion, +0.5 for heterozygous maternal duplication, +1 for homozygous maternal duplication, and the like). Refer to FIGS. 52-61F for examples of data sets processed using cumulative sums as a function of genomic position (e.g., sliding window analysis).

Example 4: Parameterized Error Removal and Unbiased Normalization (PERUN)

Variability of Measured Counts

Ideally, the measured chromosomal elevation is a straight horizontal line with the elevation of 1 for euploids, as in FIG. 62. For trisomy pregnancies, the desired behavior of the measured chromosomal elevation is a step-function, with the deviation from 1 proportional to the fetal fraction, as simulated in FIG. 63 for fetal fraction equal to 15%. Exceptions arise out of maternal deletions/duplications, which are readily recognized and distinguished from fetal abnormalities based on their magnitudes, which are multiples of one-half.

What was actually measured was not ideal. FIG. 64 shows overlaid raw counts for chromosomes 20, 21, and 22 collected from 1093 euploid pregnancies and FIG. 65 shows overlaid raw counts for chromosomes 20, 21, and 22 collected from 134 trisomy 21 pregnancies. Visual inspection of the two sets of profiles failed to confirm that chromosome 21 traces in trisomy cases were elevated. Stochastic noise and systematic bias both made the elevation of chromosome 21 difficult to visualize. Furthermore, the far right segment of chromosome 21 incorrectly suggested that euploid chromosome 21 traces were elevated, rather than the trisomy profiles. A large portion of the systematic bias originated from the GC content associated with a particular genomic region.

Attempts to remove the systematic bias due to GC content included multiplicative LOESS GC smoothing, Repeat Masking (RM), combination of LOESS and RM (GCRM), and others, such as cQN. FIG. 66 shows the results of a GCRM procedure as applied to 1093 euploid traces and FIG. 67 shows the GCRM profiles for 134 trisomy cases. GCRM successfully flattened the elevated, GC-rich, rightmost segment of chromosome 21 in euploids. However, the procedure evidently increased the overall stochastic noise. Moreover, it created a new systematic bias, absent from the raw measurements (leftmost region of chromosome 20 (Chr20)). The improvements that were due to GCRM were offset by increased noise and bias, rendering the usefulness of the procedure questionable. The tiny elevation from chromosome 21 as observed in FIG. 63 was lost in the high noise as shown in FIG. 66 and FIG. 67.

PERUN (Parameterized Error Removal and Unbiased Normalization) was developed as a viable alternative to previously described GC normalization methods. FIG. 68 and FIG. 69 contrast the PERUN method results against those presented in FIG. 64 through 67. PERUN results were obtained on the same two subpopulations of data that was analyzed in FIG. 64 through 67. Most of the systematic bias was absent from PERUN traces, only leaving stochastic noise and biological variation, such as the prominent deletion in chromosome 20 of one of the euploid samples (FIG. 68). The chromosome 20 deletion was also observable in raw count profiles (FIG. 64), but completely masked in the GCRM traces. The inability of GCRM to reveal this huge deviation clearly disqualifies it for the purposes of measuring the miniscule fetal T21 elevations. PERUN traces contain fewer bins than raw or GCRM profiles. As shown in FIG. 62-63, the PERUN results look at least as good as the measurement errors permit.

Normalization with Respect to Reference Median Count Profile

Conventional GC normalization procedures can perform suboptimally. A part of the reason has been that GC bias is not the only source of variation. A stack plot of many individual raw count profiles revealed parallelism between different samples. While some genomic regions were consistently over-represented, others were consistently under-represented, as illustrated by the traces from a 480v2 study (FIG. 6). While GC bias varied from one sample to another, the systematic, bin-specific bias observed in these profiles followed the same pattern for all samples. All the profiles in FIG. 6 zigzagged in a coordinated fashion. The only exceptions were the middle portions of the bottom two samples, which turned out to originate from maternal deletions. To

correct for this bin-specific bias, a median reference profile was used. The median reference profile was constructed from a set of known euploids (e.g., euploid pregnancies) or from all the samples in a flow cell. The procedure generated the reference profile by evaluating median counts per bin for a set of reference samples. The MAD associated with a bin measured the reliability of a bin. Highly variable bins and bins that consistently have vanishing representations were removed from further analysis (FIG. 4). The measured counts in a test data set were then normalized with respect to the median reference profile, as illustrated in FIG. 8. The highly variable bins are removed from the normalized profile, leaving a trace that is approximately 1 in the diploid sections, 1.5 in the regions of maternal heterozygous duplication, 0.5 in the areas of maternal heterozygous deletion, and so on (FIG. 9). The resulting normalized profiles reasonably reduced the variability, enabling detection of maternal deletions and duplications and tracing of sample identities (FIG. 12, 22, 13, 11). Normalization based on median count profile can clarify outcomes, but GC bias still has a negative effect on such methods. PERUN methods described here can be used to address GC bias and provide outcomes with higher sensitivity and specificity.

Detrimental Effects of Multiplicative LOESS Correction

FIG. 11. illustrated why binwise counts fluctuate more after application of GC-LOESS or GCRM (FIG. 66-67) than before (FIG. 64-65). LOESS GC correction removed the trend from the raw counts (FIG. 70, upper panel) by dividing the raw counts with the regression line (straight line, FIG. 70, upper panel). The point defined by the median counts and the median genome GC content was kept immobile. On average, counts below the median count were divided by small numbers, while counts exceeding the median count were divided by large numbers. In either case, on average, counts were scaled up or down to match 1 (FIG. 70, lower panel). The scaling of small counts, in addition to inflating the counts, also inflated their variability. The end result (FIG. 70, lower panel) to the left from the median GC genome content displayed a larger spread than the corresponding raw counts (FIG. 70, upper panel), forming the typical triangular shape (FIG. 70, lower panel, triangle). To detrend the counts, GC LOESS/GCRM sacrificed precision as such corrective processes generally are multiplicative and not additive. Normalization provided by PERUN generally is additive in nature and enhances precision over multiplicative techniques.

Inadequacy of a Genome-Wide Pivot for GC-Bias Scaling

An alternative approach applied the LOESS correction separately to individual chromosomes instead of subjecting the entire genome to a collective GC-Bias scaling. The scaling of individual

chromosomes was impractical for purposes of classifying samples as euploid or trisomy because it canceled out the signal from over-represented chromosomes. However, the conclusions from this study were eventually useful as catalyzers for developing the PERUN algorithm. FIG. 71 illustrates the fact that LOESS curves obtained for the same chromosome from multiple samples share a common intersection (pivot).

FIG. 72 demonstrated that tilting chromosome-specific LOESS curves around the pivot by an angle proportional to the GC bias coefficients measured in those samples caused all the curves to coalesce. The tilting of the chromosome-specific LOESS curves by the sample-specific GC bias coefficients significantly reduced the spread of the family of LOESS curves obtained for multiple samples, as shown in FIG. 73 (filled circles (before tilting) and open circles (after tilting)). The point where the filled circles and open circle curves touch coincided with the pivot. In addition, it became evident that the location on the GC content axis of the chromosome-specific pivot coincided with the median GC content of the given chromosome (FIG. 74, left vertical line: median, right vertical line: mean). Similar results were obtained for all chromosomes, as shown in FIG. 75A through FIG. 75F (left vertical line: median, right vertical line: mean). All autosomes and chromosome X were ordered according to their median GC content.

The genome-wide GC LOESS scaling pivoted the transformation on the median GC content of the entire genome, as shown in FIG. 76. That pivot was acceptable for chromosomes that have median GC content similar to the GC content of the entire genome, but became suboptimal for chromosomes with extreme GC contents, such as chromosomes 19, 20, 17, and 16 (extremely high GC content). The pivoting of those chromosomes centered on the median GC content of the entire genome maintained the spread observed within the left box in FIG. 76, missing the low-variability region enclosed by the right box in FIG. 76 (the chromosome-specific pivot).

Pivoting on the chromosome-specific median GC content, however, significantly reduced the variability (FIG. 75). The following observations were made:

1) GC correction should be done on small genomic sections or segments, rather than on the entire genome, to reduce the variability. The smaller the section or segment, the more focused GC correction becomes, minimizing the residual error.

2) In this particular instance, those small genomic sections or segments are identical to chromosomes. In principle, the concept is more general: the sections or segments could be any genomic regions, including 50 kbp bins.

3) The GC bias within individual genomic regions can be rectified using the sample-specific, genome-wide GC coefficient evaluated for the entire genome. This concept is important: while some descriptors of the genomic sections (such as the location of the pivot point, GC content distribution, median GC content, shape of the LOESS curve, and so on) are specific to each section and independent of the sample, the GC coefficient value used to rectify the bias is the same for all the sections and different for each sample.

These general conclusions guided the development of PERUN, as will become apparent from the detailed description of its processes.

Separability of Sources of Systematic Bias

Careful inspection of a multitude of raw count profiles measured using different library preparation chemistries, clustering environments, sequencing technologies, and sample cohorts consistently confirmed the existence of at least two independent sources of systematic variability:

1) sample-specific bias based on GC-content, affecting all bins within a given sample in the same manner, varying from sample to sample, and

2) bin-specific attenuation pattern common to all samples.

The two sources of variability are intermingled in the data. Thorough removal of both required their deconvolution. The deficiencies of the error-removal procedures predating PERUN stem from the fact that they only correct for one of the two sources of systematic bias, while neglecting the other.

For example, the GCRM (or GC LOESS) method treated identically all the bins with GC content values falling within a narrow GC content range. The bins belonging to that subset may be characterized by a wide range of different *intrinsic* elevations, as reflected by the reference median count profile. However, GCRM was blind to their inherent properties other than their GC content. GCRM therefore maintains (or even enlarges) the spread already present in the bin subset. On the other hand, the binwise reference median count disregarded the modulation of the bin-specific attenuation pattern by the GC bias, maintaining the spread caused by the varying GC content.

The sequential application of methods dealing with the opposite extremes of the error spectrum unsuccessfully attempts to resolve the two biases globally (genome-wide), ignoring the need to dissociate the two biases on the bin elevation. Without being limited by theory, PERUN apparently owes its success to the fact that it separates the two sources of bias locally, on the bin elevation.

Removal of Uninformative Bins

Multiple attempts to remove uninformative bins have indicated that bin selection has the potential to improve classification. The first such approach evaluated the mean chromosome 21, chromosome 18, and chromosome 13 counts per bin for all 480v2 trisomy cases and compared it with the mean counts per bin for all 480v2 euploids. The gap between affected and unaffected cases was scaled with the combined binwise uncertainty derived from bin counts measured in both groups. The resulting t-statistic was used to evaluate binwise p-value profile, shown in FIG. 77. In the case of chromosome 21, the procedure identified 36 uninformative bins (center panel, labeled with ellipse on FIG. 77). Elimination of those bins from calculation of Z scores noticeably increased the Z-values for affected cases, while randomly perturbing the unaffected Z-scores (FIG. 78), thereby increasing the gap between euploids and trisomy 21 cases.

In chromosome 18, the procedure only improved Z scores for two affected cases (FIG. 79).

A post-hoc analysis showed that the improvement of the Z-scores in those two samples resulted from removal of the large maternal deletion in chromosome 18 (FIG. 11) and that the two samples actually come from the same patient. These improvements were sample-specific, with no generalizing power. In chromosome 13, the procedure did not lead to any improvements of Z-scores.

An alternative bin filtering scheme removes bins with extremely low or extremely high GC content. This approach yielded mixed results, with noticeably reduced variance in chromosomes 9, 15, 16, 19, and 22 (depending on the cutoffs), but adverse effects on chromosomes 13 and 18.

Yet another simple bin selection scheme eliminates bins with consistently low counts. The procedure corrected two LDTv2CE chromosome 18 false negatives (FIG. 80) and two chromosome 21 false negatives (FIG. 81). It also corrected at least three chromosome 18 false positives, but created at least one new chromosome 18 false positive (FIG. 80):

In conclusion, the different criteria used to filter out uninformative bins made it clear that data processing will benefit from bin selection based on how much useful information the bins contribute to the classification.

Separation of GC Bias from Systematic Binwise Bias

To resolve and eliminate the different systematic biases found in the measured counts, the data processing workflow needed to optimally combine the partial procedures described from the previous section entitled "Normalization with Respect to Reference Median Count Profile" to the section entitled "Removal of Uninformative Bias". The first step is to order different samples according to their GC bias coefficient values and then stack their plots of counts-vs.-GC content. The result is a three-dimensional surface that twists like a propeller, schematically shown on FIG. 82.

Thus arranged, the measurements suggest that a set of sample-specific GC bias coefficient can be applied to rectify errors within an individual genomic section or segment. In FIG. 82, the sections or segments are defined by their GC content. An alternative partition of the genome gives contiguous, non-overlapping bins. The successive starting locations of the bins uniformly cover the genome. For one such 50 kbp long bin, FIG. 83 explores the behavior of the count values measured within that bin for a set of samples. The counts are plotted against the GC bias coefficients observed in those samples. The counts within the bin evidently increase linearly with the sample-specific GC bias. The same pattern in observed in an overwhelming majority of bins. The observations can be modeled using the simple linear relationship:

$$M = LJ + GS \qquad \text{(A)}$$

The various terms in Eq. A have the following meanings:

- ***M***: measured counts, representing the primary information polluted by unwanted variation.
- ***L***: chromosomal elevation – this is the desired output from the data processing procedure. *L* indicates fetal and/or maternal aberrations from euploid. This is the quantity that is masked both by stochastic errors and by the systematic biases. The chromosomal elevation *L* is both sample specific and bin-specific.
- ***G***: GC bias coefficient measured using linear model, LOESS, or any equivalent approach. *G* represents secondary information, extracted from M and from a set of bin-specific GC

content values, usually derived from the reference genome (but may be derived from actually observed GC contents as well). *G* is sample specific and does not vary along the genomic position. It encapsulates a portion of the unwanted variation.

- ***I***: Intercept of the linear model (diagonal line in FIG. 83). This model parameter is fixed for a given experimental setup, independent on the sample, and bin-specific.

- ***S***: Slope of the linear model (diagonal line in FIG. 83). This model parameter is fixed for a given experimental setup, independent on the sample, and bin specific.

The quantities *M* and *G* are measured. Initially, the bin-specific values *I* and *S* are unknown. To evaluate unknown *I* and *S*, we must assume that $L = 1$ for all bins in euploid samples. The assumption is not always true, but one can reasonably expect that any samples with deletions/duplications will be overwhelmed by samples with normal chromosomal elevations. A linear model applied to the euploid samples extracts the *I* and *S* parameter values specific for the selected bin (assuming $L = 1$). The same procedure is applied to all the bins in the human genome, yielding a set of intercepts *I* and slopes *S* for every genomic location. Cross-validation randomly selects a work set containing 90% of all LDTv2CE euploids and uses that subset to train the model. The random selection is repeated 100 times, yielding a set of 100 slopes and 100 intercepts for every bin. The previous section entitled "Cross-Validation of PERUN Parameters" describes the cross-validation procedure in more detail.

FIG. 84-85 show 100 intercept values and 100 slope values, respectively, evaluated for bin # 2404 in chromosome 2. The two distributions correspond to 100 different 90% subsets of 1093 LDTv2CE euploids shown in FIG. 83. Both distributions are relatively narrow and irregularly shaped. Their spreads are similar to the errors in the coefficient as reported by the linear model. As a rule, the slope is less reliable than the intercept because fewer samples populate the extreme sections of the GC-bias range.

Interpretation of PERUN Parameters I and S

The meaning of the intercept *I* is illustrated by FIG. 86. The graph correlates the estimated bin intercepts with the data extracted from a set of technical replicates, obtained when one LDTv2CE flow cell was subjected to three separate sequencing runs. The y-axis contains median values of binwise counts from those three measurements. These median values are related conceptually to

the median reference profile, previously used to normalize profiles as described in the section entitled “Normalization with Respect to Reference Median Count Profile”. The binwise intercepts are plotted along the x-axis. The striking correlation between the two quantities reveals the true meaning of the intercepts as the expected counts per bin in the absence of GC bias. The problem with the median reference count profile is that it fails to account for the GC bias (see section entitled “Normalization with Respect to Reference Median Count Profile”). In PERUN, without being limited by theory, the task of an intercept *I* is to deal with the bin-specific attenuation, while the GC bias is relegated to the other model parameter, the slope *S*.

FIG. 86 excludes chromosome Y from the correlation because the set of technical replicates does not reflect the general population of male pregnancies.

The distribution of the slope *S* (FIG. 87) illustrates the meaning of that model parameter.

The marked semblance between the distribution from FIG. 87 and the distribution of the genome-wide GC content (FIG. 88) indicates that the slope *S* approximates the GC content of a bin, shifted by the median GC content of the containing chromosome. The thin vertical line in FIG. 88 marks the median GC content of the entire genome.

FIG. 89 reaffirms the close relationship between the slope *S* and the GC content per bin. While slightly bent, the observed trend is extremely tight and consistent, with only a handful of notable outlier bins.

Extraction of Chromosomal Elevation from Measured Counts

Assuming that the model parameter values *I* and *S* are available for every bin, measurements *M* collected on a new test sample are used to evaluate the chromosomal elevation according to the following expression:

$$L = (M - GS)/I \qquad \text{(B)}$$

As in Eq. A, the GC bias coefficient *G* is evaluated as the slope of the regression between the binwise measured raw counts *M* and the GC content of the reference genome. The chromosomal elevation *L* is then used for further analyses (Z-values, maternal deletions/duplications, fetal microdeletions/ microduplications, fetal gender, sex aneuploidies, and so on). The procedure

encapsulated by Eq. B is named Parameterized Error Removal and Unbiased Normalization (PERUN).

Cross-Validation of PERUN Parameters

As inferred in the section entitled "Separation of GC Bias from Systematic Binwise Bias", the evaluation of *I* and *S* randomly selects 10% of known euploids (a set of 1093 LDTv2 in FIG. 83) and sets them aside for cross-validation. Linear model applied to the remaining 90% of euploids extracts the *I* and *S* parameter values specific for the selected bin (assuming *L* = 1). Cross validation then uses the *I* and *S* estimates for a given bin to reproduce measured *M* values from measured *G* values both in the work set and in the remaining 10% euploids (again assuming *L* = 1). The random selection of the cross-validation subset is repeated many times (100 times in FIG. 83, although 10 repetitions would suffice). 100 diagonal straight lines in FIG. 83 represent the linear models for 100 different 90% work subset selections. The same procedure is applied to all the bins in the human genome, yielding a set of intercepts *I* and slopes *S* for every genomic location.

To quantify the success of the model and avoid biasing the results, we use the *R*-factor, defined as follows:

$$R = \frac{\sum_{i=1}^{N} |M_i - P_i|}{\sum_{i=1}^{N} |M_i|} \qquad \text{(C)}$$

The numerator in Eq. B sums up the absolute deviations of the predicted count values (*P*, Eq. B) from the actual measurements (*M*). The numerator simply sums up the measurements. The *R* factor may be interpreted as the residual error in the model, or the unexplained variation. The *R* factor is directly borrowed from the crystallographic model refinement practice, which is vulnerable to bias. In crystallography, the bias is detected and measured by the R-factor evaluated within the cross-validation subset of observables. The same concepts are applied in the context of genome-wide count bias removal.

FIG. 90 shows the *R*-factors evaluated for the cross-validation subset (y-axis) plotted against *R*-factors evaluated for the work (training) set for bin # 2404 from chromosome 2. There are 100 data points since the random selection of the cross-validation subset was repeated 100 times.

Typical linear relationship is observed, with the increasing R_{cv} values (measuring bias) accompanying the decreasing R_{work}.

FIG. 90 may be interpreted in terms of the percentage error (or relative error) of the model for this particular bin. R_{cv} always exceeds R_{work}, usually by ~1%. Here, both R_{cv} and R_{work} remain below 6%, meaning that one can expect ~6% error in the predicted *M* values using the measured GC bias coefficient *G* and the model parameters *I* and *S* from the procedure described above.

Cross-Validation Error Values

FIG. 90-91 show cross-validation errors for bins chr2_2404 and chr2_2345, respectively. For those and many other bins, the errors never exceed 6%. Some bins, such as chr1_31 (FIG. 92) have cross-validation errors approaching 8%. Still others (FIG. 93-95) have much larger cross-validation errors, at times exceeding 100% (40% for chr1_10 in FIG. 93, 350% for chr1_9 in FIG. 94, and 800% for chr1_8 in FIG. 95).

FIG. 96 shows the distribution of max(R_{cv}, R_{work}) for all bins. Only a handful of bins have errors below 5%. Most bins have errors below 7% (48956 autosomes out of 61927 total including X and Y). A few bins have errors between 7% and 10%. The tail consists of bins with errors exceeding 10%.

FIG. 97 correlates the cross-validation errors with the relative errors per bin estimated from the set of technical replicates. Data points in the center region (i.e., data points located between the two vertical lines) correspond to cross-validation errors between 7% and 10%. Data points in the region to the right of the two vertical lines denote bins with cross-validation error exceeding 10%. Data points in the region to the left of the two vertical lines (error < 7%) represent the bulk of bins.

In FIG. 91-95, the number in parentheses following the bin name above the top right inset indicates the ratio between the intercept found for that particular bin and the genome-wise median count per bin. The cross-validation errors evidently increase with the decreasing value of that ratio. For example, the bin chr1_8 never gets more than 3 counts and its relative error approaches 800%. The smaller the expected number of counts for a given bin, the less reliable that bin becomes.

Bin Selection Based on Cross-Validation

Based on the observations described in the previous section entitled "Removal of Uninformative Bins" (FIG. 78 and FIG. 80-81), cross-validation errors were used as a criterion for bin filtering. The selection procedure throws away all bins with cross-validation errors exceeding 7%. The filtering also eliminates all bins that consistently contain zero counts. The remaining subset contains 48956 autosomal bins. Those are the bins used to evaluate chromosomal representations and to classify samples as affected or euploid. The cutoff of 7% is justified by the fact that the gap separating euploid Z-scores from trisomy Z-scores plateaus at the 7% cross-validation error (FIG. 98).

FIG. 99A (all bins) and 99B (cross-validated bins) demonstrate that the bin selection described above mostly removes bins with low mappability.

As expected, most removed bins have intercepts far smaller than the genome-wide median bin count. Not surprisingly, the bin selection largely overlaps with the selection described in the previous section entitled "Removal of Uninformative Bins" (FIG. 25 and 27-28).

Errors in Model Parameters

FIG. 100-101 show the 95% confidence intervals (curved lines) of the fitted linear model (thin straight line) for two bins (chr18_6 and chr18_8). The thick grey straight lines are obtained by replacing the S parameter with the difference between the GC contents of these two bins and the median GC content of chromosome 18. The error range is evaluated based on errors in the model parameters I and S for those two bins, as reported by the linear model. In addition, larger GC bias coefficients also contain larger errors. The large uncertainty corresponding to extremely large GC bias coefficients suggests that the range of applicability of the unmodified PERUN is limited to modest GC bias coefficients. Beyond that range, additional measures need to be taken to remove the residual GC bias. Fortunately, only very few samples are affected (roughly 10% of the LDTv2CE population).

FIG. 102-104 show the errors in the model parameters I and S and the correlation between the error in S and the value of the intercept.

Secondary Normalization

High values of GC bias coefficients exceed the linear range assumed by the PERUN model and are remedied by an additional LOESS GC normalization step after PERUN normalization. The multiplicative nature of the LOESS procedure does not significantly inflate the variability since the normalized counts are already very close to 1. Alternatively, LOESS can be replaced with an additive procedure that subtracts residuals. The optional secondary normalization often is utilized only required for a minority of samples (roughly 10%).

Hole Padding (Padding)

FIG. 68-69 confirm the presence of a large number of maternal deletions and duplications that have the potential to create false positives or false negatives, depending on their sizes and locations. An optional procedure called hole-padding has been devised to eliminate the interferences from these maternal aberrations. The procedure simply pads the normalized profile to remain close to 1 when it deviates above 1.3 or below 0.7. In LDTv2CE, hole padding (i.e., padding) did not significantly affect the classification. However, FIG. 105 shows a WI profile that contains a large deletion in chromosome 4. Hole padding converts that profile from chromosome 13 false positive to chromosome 13 true negative.

Results

This section discusses PERUN results for trisomy 13, trisomy 18 and trisomy 21 (T13, T18 and T21, respectively), gender determination, and sex aneuploidy.

Reduced Variability

FIG. 106 compares the distribution of standard deviations of the binwise count profiles before and after PERUN normalization. The resulting distributions of chromosome representations for euploids and trisomy cases are shown in FIG. 107.

Improved T13, T18, and T21 Classification

FIG. 108-111 compare LDTv2CE PERUN classification results with those obtained using GCRM counts. In addition to removing two chromosome 18 false positives, two chromosome 18 false

negatives, and two chromosome 21 false negatives, PERUN almost doubles the gap between the euploids and the affected cases, in spite of the fact that the higher plexing elevation decreased the number of counts per sample (ELAND data). Similar results are obtained when PERUN parameters trained on LDTv2CE Eland data are applied to WI measurements. Bowtie alignments require a different set of parameters and additional bin filtering, accounting for low mappability in some bins, but its results approach those seen with ELAND alignments.

Example 5: Additional Description of PERUN

Examples of parameterized Error Removal and Unbiased Normalization (PERUN) methods are described in Example 4, and an additional description of such methods is provided in this Example 5.

Massive parallel sequencing of cell-free circulating DNA (e.g., from maternal plasma) can, under ideal conditions, quantify chromosomal elevations by counting sequenced reads if unambiguously aligned to a reference human genome. Such methods that incorporate massive amounts of replicate data can, in some cases, show statistically significant deviations between the measured and expected chromosomal elevations that can imply aneuploidy [Chiu et al., Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *PNAS USA*. 2008;105:20458–20463; Fan et al., Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *PNAS USA*. 2008;105:16266–16271; Ehrich et al., Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting, *American Journal of Obstetrics and Gynecology - AMER J OBSTET GYNECOL*, vol. 204, no. 3, pp. 205.e1-205.e11, 2011 DOI: 10.1016/j.ajog.2010.12.060]. Ideally, the distribution of aligned reads should cover euploid sections of the genome at a constant level (FIG. 62 and FIG. 63). In practice, uniformity can be difficult to attain because multiplexed Next Generation Sequencing (NGS) measurements typically yield low coverage (about 0.1) with sparsely scattered read start positions. In some embodiments, this problem is partially overcome by partitioning the genome into non-overlapping sections (bins) of equal lengths and assigning to each bin the number of the reads that align within it. In some embodiments, residual unevenness stemming from GC bias [Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008 Sep;36(16):e105. Epub 2008 Jul 26.] is largely suppressed using multiplicative detrending with respect to the binwise GC content (Fan HC, Quake SR (2010) Sensitivity of Noninvasive Prenatal

Detection of Fetal Aneuploidy from Maternal Plasma Using Shotgun Sequencing Is Limited Only by Counting Statistics. PLoS ONE 5(5): e10439. doi:10.1371/journal.pone.0010439). In some embodiments, the resulting flattening of the count profile allows for successful classification of fetal trisomies in a clinical setting using quadruplex barcoding [Palomaki et al., DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. *Genet Med.,* 2011 Nov;13(11):913-20.].

The transition from a quadruplex (i.e. 4 simultaneous sample reads) to higher sample plexing levels (e.g., dodecaplex (i.e. 12 simultaneous sample reads)) pushes the limits of NGS-based detection of genetic variations (e.g., aneuploidy, trisomy, and the like) in a test subject (e.g., a pregnant female), reducing both the number of reads per sample and the gap separating genetic variations (e.g., euploid from trisomy samples). The downsampling driven by increased multiplexing can impose new, more stringent requirements on data processing algorithms (FIG. 64, FIG. 65 and Example 4). In some embodiments, GC detrending, even when coupled with repeat masking, requires some improvement (FIG. 66, FIG. 67 and Example 4). In some embodiments, to maintain the sensitivity achieved with quadruplex barcoding (e.g., quadruplex indexing), methods and algorithms are presented that are capable of extracting a minute signal of interest from an overwhelming background noise as illustrated and described below and in FIG. 7, FIG. 8 and Example 4. In some embodiments, a novel method termed “PERUN” (Parameterized Error Removal and Unbiased Normalization) is described.

Conventional GC detrending can be multiplicative in nature (FIG. 17 and Example 4) and may not address additional sources of systematic bias, illustrated in FIG. 6. In certain embodiments, a reference median count profile constructed from a set of known euploid samples can eliminate additional bias and lead to qualitative improvements. In certain embodiments, a reference median count profile constructed from a set of known euploid samples can inherit a mixture of residual GC biases from the reference samples. In some embodiments, a normalization removes one or more orthogonal types of bias by separating them from one another at the bin elevation, rather than tackling them in bulk. In some embodiments GC bias is removed and binwise separation of the GC bias from the position-dependent attenuation is achieved (FIG. 68. FIG. 69 and Example 4). In some embodiments, substantially increased gaps between euploid and trisomy Z-scores are obtained relative to both quadruplex and dodecaplex GCRM results. In some embodiments, maternal and fetal microdeletions and duplications are detected. In some embodiments fetal

fractions are accurately measured. In some embodiments gender is determined reliably. In some embodiments sex aneuploidy (e.g., fetal sex aneuploidy) is identified.

PERUN Method and Definitions

In some embodiments the entire reference genome is partitioned into an ordered set *B* of *J* bins:

$$B = \{b_j | j = 1, \ldots, J\} \quad \text{(D)}$$

Bin lengths can be constrained to accommodate genomic stretches of relatively uniform GC content. In some embodiments adjacent bins can overlap. In some embodiments adjacent bins do not overlap. In some embodiments the bin edges can be equidistant or can vary to offset systematic biases, such as nucleotide composition or signal attenuation. In some embodiments a bin comprises genomic positions within a single chromosome. Each bin b_j is characterized by the GC content g_j^0 of the corresponding portion of the reference genome. In some embodiments, the entire genome is assigned a reference GC content profile:

$$g^0 = \lfloor g_1^0 \quad g_2^0 \quad \cdots \quad g_J^0 \rceil \quad \text{(E)}$$

The same g^0 profile can apply to all samples aligned to the chosen reference genome.

A proper or trivial subset of bins b,

$$b \subseteq B \quad \text{(F)}$$

can be selected to satisfy certain criteria, such as to exclude bins with $g_j^0 = 0$, bins with extreme g_j^0 values, bins characterized by low complexity or low mappability (Derrien T, Estelle´ J, Marco Sola S, Knowles DG, Raineri E, et al. (2012) Fast Computation and Applications of Genome Mappability. PLoS ONE 7(1): e30377, doi:10.1371/journal.pone.0030377), highly variable or otherwise uninformative bins, regions with consistently attenuated signal, observed maternal aberrations, or entire chromosomes (X, Y, triploid chromosomes, and/or chromosomes with extreme GC content). The symbol $\|b\|$ denotes the size of b.

All sequenced reads from sample i unambiguously aligned within a bin b_j form a set a_{ij} whose cardinality M_{ij} represents raw measured counts assigned to that bin. In some embodiments, the vector of measured bin counts for sample i constitutes the raw count profile for that sample. In some embodiments this is the primary observation for the purposes of PERUN:

$$M_i = [M_{i1} \quad M_{i2} \quad \cdots \quad M_{iJ}] \tag{G}$$

To enable comparisons among different samples, the scaling constant N_i is evaluated as the sum of raw bin counts over a subset of the bins:

$$N_i = \sum_{b \subseteq B} M_{ij} \tag{H}$$

In some embodiments b in Eq. H is restricted to autosomal bins. In some embodiments b in Eq. H is not restricted to autosomal bins. Division of M_i by the total counts N_i yields the scaled raw bin counts m_{ij}:

$$m_i = [m_{i1} \quad m_{i2} \quad \cdots \quad m_{iJ}] = M_i / N_i \tag{I}$$

The nucleotide composition of the set a_{ij} is described by the bin's observed GC content g_{ij}. The sample-specific observed GC content profile g_i gathers individual bin-specific GC contents into a vector:

$$g_i = [g_{i1} \quad g_{i2} \quad \cdots \quad g_{iJ}] \tag{J}$$

In some embodiments, $g_i \neq g^0$ and $g_{i_1} \neq g_{i_2 \neq i_1}$. The symbol g denotes the GC content profile regardless of its origin, i.e. whether it is derived from the reference genome or from the sample-specific read alignments. In some embodiments model equations use g. In some embodiments, actual implementations can substitute g with either g^0 or g_i.

For a single sample i, a linear relationship between m_i and g is assumed, with G_i and r_i denoting the sample-specific slope of the regression line and the array of residuals, respectively:

$$m_i = G_i g + r_i \quad \text{(K)}$$

The regression can extend over the entire set B (Eq. D) or its proper subset b (Eq. F). The observed slope G_i is also referred to as the scaled GC bias coefficient. G_i expresses the bulk of the vulnerability of the sample i to the systematic GC bias. In some embodiments, to minimize the number of model parameters, higher-order terms, linked with curvature of the relationship $m_i(g)$ and encapsulated in the residuals r_i are not explicitly addressed. In some embodiments, since sample-specific total counts N_i confound the interactions among observables recorded on different samples, the unscaled equivalent of G_i, relating M_i to g, is less useful and will not be considered.

The vector of true chromosomal elevations l_{ij} corresponding to bins $b_j \in b$ in sample i form the sample-specific chromosomal elevation profile:

$$l_i = \begin{bmatrix} l_{i1} & l_{i2} & \cdots & l_{iJ} \end{bmatrix} \quad \text{(L)}$$

In some embodiments, the goal is to derive estimates for l_i from m_i by removing systematic biases present in m_i.

The values l_{ij} are bin-specific and also sample-specific. They comprise both maternal and fetal contributions, proportional to their respective ploidy P_{ij}^M and P_{ij}^F. The bin-specific and sample-specific ploidy P_{ij} can be defined as an integral multiple of one-half, with the values of 1, 1/2, 0, 3/2, and 2 representing euploid, heterozygous deletion, homozygous deletion, heterozygous duplication, and homozygous duplication, respectively. In some instances, trisomy of a given chromosome implies ploidy values of 3/2 along the entire chromosome or its substantial portion.

When both the mother and the fetus are diploid ($P_{ij}^{M} = P_{ij}^{F} = 1$), l_{ij} equals some arbitrarily chosen euploid elevation E. In some embodiments, a convenient choice sets E to $1/\|b\|$, thus ensuring that the profile l_i is normalized. In the absence of bin selection, $\|b\| = \|B\| = J \Rightarrow E = 1/J$. In some embodiments, E can be set to 1 for visualization. In some embodiments, the following relationship is satisfied:

$$l_{ij} = E\left[(1-f_i)P_{ij}^{M} + f_i P_{ij}^{F}\right] \quad \text{(M)}$$

The symbol f_i stands for the fraction of the fetal DNA present in the cell-free circulating DNA from maternal plasma in sample i. Any deviations from euploid, either fetal ($P_{ij}^{F} \neq 1$) or maternal ($P_{ij}^{M} \neq 1$), cause differences between l_{ij} and E that can be exploited to estimate f_i and detect microdeletions/microduplications or trisomy.

To achieve the goal of extracting l_i from m_i , a linear relationship is postulated between the bin-specific scaled raw counts m_{ij} measured on a given sample and the sample-specific scaled GC bias coefficients:

$$m_i = l_i I + G_i S \quad \text{(N)}$$

The diagonal matrix I and the vector S gather bin-specific intercepts and slopes of the set of linear equations summarized by Eq. N:

$$I = \begin{bmatrix} I_1 & 0 & \cdots & 0 \\ 0 & I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_J \end{bmatrix} \quad \text{(O)}$$

$$S = \begin{bmatrix} S_1 & S_2 & \cdots & S_J \end{bmatrix} \quad \text{(P)}$$

Both I and S are sample-independent. The intercepts I_j can be viewed as expected euploid values for scaled row counts in the absence of GC bias (i.e. when $G_i = 0$). Their actual values

reflect the convention adopted for E (*vide supra*). The intercepts S_j are non-linearly related to the differences $g_j^0 - \langle g_k^0 \rangle$, where $\langle g_k^0 \rangle$ represents the median GC content of the chromosome containing the bin j.

Once the values for the parameters I and S are known, the true chromosomal elevation profile l_i is estimated from the scaled raw count profile m_i and the scaled GC bias coefficient G_i by rearranging Eq. N:

$$l_i = (m_i - G_i S) I^{-1} \qquad \text{(Q)}$$

The diagonal character of the intercept matrix *I* provides for the matrix inversion in Eq. Q.

Parameter estimation

Model parameters I and S are evaluated from a set of N scaled raw count profiles collected on samples karyotyped as euploid pregnancies. N is of the order of 10^3. Scaled GC bias coefficients G_i are determined for each sample $(i = 1,..., N)$. All samples are segregated into a small number of classes according to the sizes and signs of their G_i values. The stratification balances the opposing needs to include sufficiently large numbers of representatives and a sufficiently small range of G_i values within each shell. The compromise of four strata accommodates negative, near-zero, moderately positive, and extreme positive GC biases, with the near-zero shell being most densely populated. A fraction of samples (typically 10%) from each stratum can be randomly selected and set aside for cross-validation. The remaining samples make up the work set, used to train the model. Both the training and the subsequent cross-validation assume that all samples are free of maternal and fetal deletions or duplications along the entire genome:

$$P_{ij}^M = P_{ij}^F = 1, \forall i = 1,\ldots N, \forall j = 1,\ldots, J \qquad \text{(R)}$$

The large number of samples compensates for the occasional maternal deviations from the assumption R. For each bin j, l_{ij} is set to E, allowing evaluation of the intercept I_j and the slope

S_j as the coefficients of the linear regression applied to the training set according to Eq. N. The uncertainty estimates for I_j and S_j are recorded as well.

The random partitioning into the working and the cross-validation subsets is repeated multiple times (e.g., 10^2), yielding distributions of values for the I_j and S_j parameters. In some embodiments the random partitioning is repeated between about 10 and about 10^5 times. In some embodiments the random partitioning is repeated about 10, about 10^2, about 10^3, about 10^4 or about 10^5 times.

Cross-validation

Once derived from the work set, the model parameters I_j and S_j are employed to back-calculate scaled raw counts from the scaled GC bias coefficients using Eq. N and assumption R. The symbol p_{ij} denotes the predicted scaled raw counts for the bin b_j in the sample i. The indices W and CV in further text designate the work and the cross-validation subsets, respectively. The back-calculation is applied to all samples, both from W and CV. R-factors, borrowed from the crystallographic structure refinement practice (Brünger, Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature* 355, 472 - 475 (30 January 1992); doi:10.1038/355472a0), are separately defined for the two subsets of samples:

$$R_j^W = \frac{\sum_{i \in W} \left| m_{ij} - p_{ij} \right|}{\sum_{i \in W} \left| m_{ij} \right|} \qquad \text{(S)}$$

$$R_j^{CV} = \frac{\sum_{i \in CV} \left| m_{ij} - p_{ij} \right|}{\sum_{i \in CV} \left| m_{ij} \right|} \qquad \text{(T)}$$

Both R-factors are bin-specific. As in crystallography, R-factors 16-17 can be interpreted as residual relative errors in the model. Having been excluded from the parameter estimation, the cross-validation R-factor R_j^{CV} provides a true measure of the error for the given W/CV division,

while the difference between R_j^{CV} and R_j^W reflects the model bias for the bin j. A separate pair of R-values is evaluated for each bin and for each random partitioning of the set of samples into W and CV. The maximum of all R_j^{CV} and R_j^W values obtained for the different random partitionings into W and CV is assigned to the bin j as its overall model error ε_j.

Bin selection

All the bins with zero GC content g_j^0 are eliminated from further consideration, as is the set $\{b_j : M_{ij} \equiv 0, \forall i = 1,\ldots,N\}$ of bins that consistently receive zero counts across a large number of samples. In addition, a maximum tolerable cross-validation error value ε can be imposed on all bins. In some embodiments the bins with model errors ε_j exceeding the upper limit ε are rejected. In some embodiments, filtering uses bin mappability scores $\mu_j \in [0,1]$ and imposes a minimum acceptable mappability μ, rejecting bins with $\mu_j < \mu$ (Derrien T, Estelle´ J, Marco Sola S, Knowles DG, Raineri E, et al. (2012) Fast Computation and Applications of Genome Mappability. PLoS ONE 7(1): e30377, doi:10.1371/journal.pone.0030377). For the purposes of determining fetal trisomy of chromosomes 21, 18, and 13, the sex chromosomes can be excluded as well. The subset β of bins that survive all the phases of the bin selection can undergo further computations. In some embodiments, the same subset β is used for all samples.

Normalization and standardization

In some embodiments, for a given sample i, the chromosomal elevations l_{ij} corresponding to the bin selection β are estimated according to Eq. Q. In some embodiments, a secondary normalization is applied to remove any curvature from the l_{ij}-vs.-GC content correlation. In some embodiments l_{ij} is already nearly unbiased, the secondary detrending is robust and is immune to error boosting. In some embodiments, standard textbook procedures suffice.

In some embodiments, the results of the normalization are summed up within each chromosome:

$$L_{in} = \sum_{b_j \in \beta \cap Chr_n} l_{ij} \ , \quad n = 1,\ldots,22 \qquad \text{(U)}$$

The total autosomal material in sample *i* can be evaluated as the sum of all individual L_{in} terms:

$$L_i = \sum_{n=1}^{22} L_{in} \qquad \text{(V)}$$

The chromosomal representation of each chromosome of interest can be obtained by dividing L_{in} with L_i:

$$\chi_{in} = L_{in} / L_i \qquad \text{(W)}$$

The variability σ_n of the representation of the chromosome n can be estimated as an uncensored MAD of χ_{in} values across a selection of samples spanning multiple flow cells. In some embodiments, the expectation $\langle \chi_n \rangle$ is evaluated as the median of χ_{in} values corresponding to a selection of samples from the same flow cell as the tested sample. Both sample selections can exclude high positive controls, low positive controls, high negative controls, blanks, samples that fail QC criteria, and samples with SD(l_i) exceeding a predefined cutoff (typically 0.10). Together, the values σ_n and $\langle \chi_n \rangle$ can provide the context for standardization and comparison of chromosomal representations among different samples using Z-scores:

$$Z_{in} = (\chi_{in} - \langle \chi_n \rangle) / \sigma_n \qquad \text{(X)}$$

In some embodiments, aberrations such as trisomies 13, 18, and 21 are indicated by Z-values exceeding a predefined value, dictated by the desired confidence level.

Example 6: Examples of formulas

Provided below are non-limiting examples of mathematical and/or statistical formulas that can be used in methods described herein.

$$Z = \frac{\Delta_1 - \Delta_2}{\sqrt{\sigma_1^2\left(\frac{1}{N_1}+\frac{1}{n_1}\right)+\sigma_2^2\left(\frac{1}{N_2}+\frac{1}{n_2}\right)}}$$

$$P(q) = \frac{1}{\sigma\sqrt{2\pi}} exp[-(q-q_0)/(2\sigma^2)]$$

$$q_0 = 1 + F/2$$

$$z = -F/(2\sigma\sqrt{2})$$

$$B = \int_{-\infty}^{1} P(q)dq = \frac{1}{2}[1 + erf(z)]$$

$$erf(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)}$$

$$R = \frac{1-B}{B} = \frac{1-erf(z)}{1+erf(z)} = \frac{1-erf[-F/(2\sigma\sqrt{2})]}{1+erf[-F/(2\sigma\sqrt{2})]}$$

Example 7: Identifying and Adjusting (Padding) Elevations

Maternal deletions and duplications, often represented as first elevations in a profile, can be removed from count profiles normalized with PERUN to reduce variability when detecting T21, T18, or T13. The removal of deletions and duplication from a profile can reduce the variability (e.g., biological variability) found in measured chromosomal representations that originates from maternal aberrations.

All bins that significantly deviate from the expected chromosomal elevation of 1 are first identified. In this example some isolated bins are removed from the selection. This is optional. In this example only large enough groups of contiguous outlier bins are kept. This is also optional. Depending on the elevation assigned to an outlier bin or a group of contiguous outlier bins, a correction factor is added to the measured elevation to adjust it closer to the expected elevation of 1. The PAV values used in this example are +1 (for homozygous maternal deletions), +0.5 (for heterozygous maternal deletions), -0.5 (for heterozygous maternal duplications), -1 (for homozygous maternal duplications), or more (for large spikes). Large spikes are often not identified as maternal deletions and duplications.

This padding procedure corrected the classification (e.g., the classification as an aneuploidy, e.g., a trisomy) for samples that contains large maternal aberrations. Padding converted the WI sample from false positive T13 to true negative due to removal of a large maternal deletion in Chr4 (FIG. 112 - 115).

Past simulations with experimental data have shown that depending on the chromosome, fetal fraction, and the type of aberration (homozygous or heterozygous, duplication or deletion), maternal aberrations in 20-40 bins long may push the Z-value over the classification edge (e.g., threshold) and result in a false positive or a false negative. Padding (e.g., adjusting) can circumvent this risk.

This padding procedure can remove uninteresting maternal aberrations (a confounding factor), reduce euploid variability, create tighter sigma-values used to standardize Z-scores and therefore enlarge the gap between euploids and trisomy cases.

Example 8: Determining Fetal Fractions from Maternal and/or Fetal Copy Number Variations

A distinguishing feature of a method described herein is the use of maternal aberrations (e.g., maternal and/or fetal copy number variations) as a probe providing insight into the fetal fraction in the case of a pregnant female bearing a fetus (e.g., a euploid fetus). The detection and quantitation of maternal aberrations typically is aided by normalization of raw counts. In this example raw counts are normalized using PERUN. Alternatively, normalization with respect to a reference median count profile can be used in a similar manner and for the same purpose.

PERUN normalization of raw counts yields sample-specific binwise chromosomal levels l_{ij} (*i* counts samples, *j* counts bins). They comprise both maternal and fetal contributions, proportional to their respective ploidies P_{ij}^{M} and P_{ij}^{F}. The bin-specific and sample-specific ploidy P_{ij} is defined as an integral multiple of 1/2, with the values of 1, 1/2, 0, 3/2, and 2 representing euploid, heterozygous deletion, homozygous deletion, heterozygous duplication, and homozygous duplication, respectively. In particular, trisomy of a given chromosome implies ploidy values of 3/2 along the entire chromosome or its substantial portion.

When both the mother and the fetus are diploid ($P_{ij}^{M} = P_{ij}^{F} = 1$), l_{ij} equals some arbitrarily chosen euploid level *E*. A convenient choice sets *E* to $1/\|b\|$, where *b* denotes a proper or trivial subset of the set of all bins (*B*). thus ensuring that the profile l_i is normalized. In the absence of bin selection, $\|b\| = \|B\| = J \Rightarrow E = 1/J$. Alternatively and preferentially, *E* may be set to 1 for visualization. In general, the following relationship is satisfied:

$$l_{ij} = E\left[(1-f_i)P_{ij}^{M} + f_i P_{ij}^{F}\right] \qquad \text{(Y)}$$

The symbol f_i stands for the fraction of the fetal DNA present in the cell-free circulating DNA from maternal plasma in sample *i*. Any deviations from euploid, either fetal ($P_{ij}^{F} \neq 1$) or maternal ($P_{ij}^{M} \neq 1$), cause differences between l_{ij} and *E* that can be exploited to estimate f_i and detect microdeletions/microduplications or trisomy.

Four different types of maternal aberrations are considered separately. All four account for possible fetal genotypes, as the fetus may (or in homozygous cases must) inherit the maternal aberration. In addition, the fetus may inherit a matching aberration from the father as well. In general, fetal fraction can only be measured when $P_{ij}^{M} \neq P_{ij}^{F}$.

A) Homozygous maternal deletion ($P_{ij}^{M} = 0$). Two possible accompanying fetal ploidies include:
   a. $P_{ij}^{F} = 0$, in which case $l_{ij} = 0$ and the fetal fraction cannot be evaluated from the deletion.
   b. $P_{ij}^{F} = 1/2$, in which case $l_{ij} = f_i/2$ and the fetal fraction is evaluated as twice the average elevation within the deletion.

B) Heterozygous maternal deletion ($P_{ij}^{M} = 1/2$). Three possible accompanying fetal ploidies include:
   a. $P_{ij}^{F} = 0$, in which case $l_{ij} = (1-f_i)/2$ and the fetal fraction is evaluated as twice the difference between ½ and the average elevation within the deletion.

b. $P_{ij}^{F} = 1/2$, in which case $l_{ij} = 1/2$ and the fetal fraction cannot be evaluated from the deletion.

c. $P_{ij}^{F} = 1$, in which case $l_{ij} = (1 + f_i)/2$ and the fetal fraction is evaluated as twice the difference between ½ and the average elevation within the deletion.

C) Heterozygous maternal duplication ($P_{ij}^{M} = 3/2$). Three possible accompanying fetal ploidies include:

a. $P_{ij}^{F} = 1$, in which case $l_{ij} = (3 - f_i)/2$ and the fetal fraction is evaluated as twice the difference between 3/2 and the average elevation within the duplication.

b. $P_{ij}^{F} = 3/2$, in which case $l_{ij} = 3/2$ and the fetal fraction cannot be evaluated from the duplication.

c. $P_{ij}^{F} = 2$, in which case $l_{ij} = (3 + f_i)/2$ and the fetal fraction is evaluated as twice the difference between 3/2 and the average elevation within the duplication.

D) Homozygous maternal duplication ($P_{ij}^{M} = 2$). Two possible accompanying fetal ploidies include:

a. $P_{ij}^{F} = 2$, in which case $l_{ij} = 2$ and the fetal fraction cannot be evaluated from the duplication.

b. $P_{ij}^{F} = 3/2$, in which case $l_{ij} = 2 - f_i/2$ and the fetal fraction is evaluated as twice the difference between 2 and the average elevation within the duplication.

The following LDTv2CE samples (FIG. 116 – 131) illustrate the application of determining fetal fraction from maternal and/or fetal copy number variations. The patients were not selected randomly and any agreement with FQA fetal fraction values should not be construed as the measure of merit of either technique.

Example 9: *Determining Fetal Fractions from chromosomal representations of Chromosome X and Chromosome Y*

Measurements of fetal fraction (e.g., the measurement of the fraction of fetal DNA in the bodily fluids of a pregnant female) can be performed by an FQA, which is based on mass spectroscopy which does not make use of next generation sequencing (NGS) data obtained for detection of

trisomies T21, T18, and T13 (e.g., PCT/US2010/027879 filed March 18, 2010 entitled PROCESSES AND COMPOSITIONS FOR METHYLATION-BASED ENRICHMENT OF FETAL NUCLEIC ACID FROM A MATERNAL SAMPLE USEFUL FOR NON-INVASIVE PRENATAL DIAGNOSES, which is hereby incorporated by reference). Fetal fraction determinations can be made from untargeted NGS data collected on male euploids and/or aneuploidy pregnancies, including sex aneuploidies and trisomies T21, T18 and/or T13 as described here. Counts obtained from chromosome X (Chr X) and/or chromosome Y (ChrY) measurements (in the case of male pregnancies or sex aneuploidies) can be combined with chromosome 21, 18, and/or chromosome 13 measurements for trisomy pregnancies to determine fetal fraction.

In this example, the fraction of the fetal DNA in the circulating cell-free DNA from maternal plasma is determined using massively parallel sequencing data. Count profiles normalized with PERUN are used to measure fetal fractions from chromosomal representations of ChrX and ChrY in male pregnancies, pregnancies with sex aneuploidies, and/or chromosomal representations of chromosome 21 (Chr21), chromosome 18 (Chr18) and/or chromosome (Chr13) in trisomy pregnancies.

Detection and quantification of maternal aberrations is aided by normalization of raw counts. In this example, raw counts are normalized using PERUN. Alternatively, GCRM counts or normalization with respect to a reference median count profile can be used in a similar manner and for the same purpose. PERUN normalization of raw counts yields sample-specific binwise chromosomal levels l_{ij} (*i* counts samples, *j* counts bins). Such levels are attributed to maternal and fetal contributions, proportional to their respective ploidies P_{ij}^{M} and P_{ij}^{F}. The bin-specific and sample-specific ploidy P_{ij} is defined as an integral multiple of 1/2, with the values of 1, 1/2, 0, 3/2, and 2 representing euploid, heterozygous deletion, homozygous deletion, heterozygous duplication, and homozygous duplication, respectively. In particular, trisomy of a given chromosome implies ploidy values of 3/2 along the entire chromosome or its substantial portion.

When both the mother and the fetus are diploid ($P_{ij}^{M} = P_{ij}^{F} = 1$), l_{ij} equals arbitrarily chosen euploid level *E*. A convenient choice sets *E* to $1/\|b\|$, where *b* denotes a proper or trivial subset of the set of all bins (*B*), thus ensuring that the profile l_i is normalized. In the absence of bin selection,

$\|b\| = \|B\| = J \Rightarrow E = 1/J$. Alternatively and preferentially, *E* may be set to 1 for visualization. In general, the following relationship is satisfied:

$$l_{ij} = E\left[(1-f_i)P_{ij}^M + f_i P_{ij}^F\right] \qquad \text{(Y)}$$

The symbol f_i stands for the fraction of the fetal DNA present in the cell-free circulating DNA from maternal plasma in sample *i*. Any deviations from euploid, either fetal ($P_{ij}^F \neq 1$) or maternal ($P_{ij}^M \neq 1$), cause differences between l_{ij} and *E* that can be exploited to estimate f_i and detect microdeletions/microduplications or trisomy.

Table 2 summarizes fetal ploidy values for various types of pregnancies. Maternal contribution is fixed at $P_{ij}^M = 1$ for every bin *j* belonging to chromosomes 21, 18, 13, and X. Maternal ploidy for ChrY is zero.

TABLE 2

Pregnancy Status	Fetal Chr21	Fetal Chr18	Fetal Chr13	Fetal ChrX	Fetal ChrY
Female T21	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
Female T18	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
Female T13	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
Male T21	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Male T18	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Male T13	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Male Euploid	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Turner	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 0$
Jacobs	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1$
Klinefelter	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$
TripleX	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 0$

When fetal ploidy differs from maternal ploidy and from zero, the imbalance can be used to evaluate the fetal fraction for male fetuses. In this example, the imbalance is not used to evaluate

fetal fraction in pregnant females bearing a euploid female fetus. The following illustrates the principles described above.

Let N_n and N denote the number of bins in a given chromosome *n* and the total number of autosomal bins in the normalized profile, respectively. The expected euploid chromosomal representation is the ratio between the total area under the segment of the chromosome in question and the area under all autosomal bins:

$$c_n^0 = N_n / N \qquad \text{(Z)}$$

In practice, Equation Z may be substituted by the median (or mean) value of a large number of euploid chromosomal representations for a given chromosome. In a trisomy pregnancy, such as T21, T18, or T13, $P_{ij}^F = 3/2$ and $P_{ij}^M = 1$. The expected chromosomal representation is then modified by the fetal fraction as in Equation AA.

$$c_{in} = \frac{N_n}{N}\left(1+\frac{f_i}{2}\right) = c_n^0\left(1+\frac{f_i}{2}\right) \qquad \text{(AA)}$$

The fetal fraction is therefore obtained from the measured chromosomal representation and the expected euploid value as per Equation AB.

$$f_i = 2\left(\frac{c_{in}}{c_n^0}-1\right) \qquad \text{(AB)}$$

ChrX in male pregnancies follows a different relationship as shown in Equation AC.

$$f_i = 2\left(1-\frac{c_{in}}{c_n^0}\right) \qquad \text{(AC)}$$

ChrY shows a strong linear correlation with fetal fraction. ChrY in male pregnancies can follow a relationship as shown in Equation AG. Due to large variability of ChrY normalized profiles, prediction of fetal fractions from measured ChrY values can be conveniently performed by calibrating the relationship ChrY-vs-fetal fraction in trisomy male pregnancies and then applying the calibration line to all male pregnancies. A similar procedure is also convenient for ChrX. Alternatively, both ChrX and ChrY in male trisomy pregnancies can be used together to generate a

bivariate linear model, with the fetal fraction extracted from trisomy chromosomes as the response variable. Derivation of the appropriate relationships corresponding to Turner (X), Jacobs (XYY), Klinefelter (XXY), and TripleX (XXX) syndromes follows from the same line of reasoning. For example, for TripleX syndrome the relationship described in equation AB can be utilized and the X chromosome is treated as aneuploid. For Turner syndrome, equation AC is utilized to determine the fetal fraction using an X chromosome representation. For Klinefelter syndrome, equation AG is used to determine the fetal fraction from ChrY representation and equation AC is not used. For Jacobs syndrome, equation AC is utilized to determine the fetal fraction from ChrX representation and a variation of equation AG is used to determine the fetal fraction from ChrY representation. In some embodiments the fetal fraction resulting from equation AG is divided by 2 to arrive at a fetal fraction for Jacobs syndrome. In some embodiments the C_y value in equation AG is divided by 2 to arrive at a fetal fraction for Jacobs syndrome.

WI samples normalized with PERUN are used to illustrate the application of a method described in this Example. Fetal fractions were obtained from Chr21 representations in 107 male T21 pregnancies, from Chr18 representations in 33 male T18 pregnancies, and from 5 male T13 pregnancies using Equation AB. The graph in FIG. 132 correlates ChrX chromosomal representations in those 145 trisomy male pregnancies. The linear model based on FIG. 132 yields the following linear relationship between the fetal fraction (multiplied by 100) and the chromosome X representation in male pregnancies as shown in Equation AD.

$$f_i = 179.1 - 3045.82 c_{iX} \qquad \text{(AD)}$$

The agreement between fetal fractions obtained from ChrX using Equation AD and the fetal fractions obtained from Chr21, 18, and 13 in the male trisomy pregnancies is shown in FIG. 133. FIG. 134 correlates chromosome Y representations in those same 145 trisomy male pregnancies. The resulting regression allows estimation of the fetal fraction from ChrY representation from Equation AE.

$$f_i = 0.5368 + 1344.0162 c_{iY} \qquad \text{(AE)}$$

The agreement between fetal fractions obtained from ChrY using Equation AE and the fetal fractions obtained from Chr21, 18, and 13 in the male trisomy pregnancies is shown in FIG. 135. Equations AD and AE were used to evaluate fetal fractions in 858 euploid male pregnancies. FIG.

136 illustrates the agreement between fetal fractions derived from ChrX and from ChrY. Finally, when both ChrX and ChrY are trained on fetal fraction values derived from Chr21, 18, and 13 in trisomy male pregnancies, the following regression coefficients are obtained as per Equation AF below.

$$f_i = 24.88 - 416.42c_{iX} + 1169.46c_{iY} \quad \text{(AF)}$$

The model that combines ChrX and ChrY predicts fetal fractions shown in FIG. 137, along with the corresponding values obtained from Chr 21, 18, and 13 (male trisomy pregnancies). Extension to sex aneuploidies is straightforward as discussed previously.

Example 10: Determining fetal fraction from a measured ChrY representation

The fetal fraction *f* for a male pregnancy was determined from the measured chromosomal representation y of ChrY using the following formula:

$$f = 2\frac{I + S\langle x\rangle - y}{S\langle x\rangle} \quad (71)$$

The term $\langle x\rangle$ represents the median chromosomal representation of ChrX for female pregnancies. This was different from zero due to noise. I and S quantified the linear relationship between ChrX and ChrY representations in male pregnancies. The parameters I and S were evaluated using constrained linear regression between ChrX and ChrY representations in male pregnancies. The regression line was forced through the point representing median ChrX and ChrY representations for female fetuses ($\langle x\rangle$ and $\langle y\rangle$, respectively), leaving only one adjustable parameter – the slope S. The value of the model parameter S was derived from the following assumption:

$$y - \langle y\rangle = S(x - \langle x\rangle) \quad (72)$$

To optimize the model parameter S, functional *F* was defined as the sum of squared residuals between the model (Eq. 72) and the actually measured ChrY representations:

$$F = \sum_i [y_i - S(x_i - \langle x\rangle) - \langle y\rangle]^2$$
$$= \sum_i \left(y_i^2 + S^2x_i^2 + S^2\langle x\rangle^2 + \langle y\rangle^2 - 2Sx_iy_i + 2S\langle x\rangle y_i - 2\langle y\rangle y_i - 2S^2\langle x\rangle x_i + 2S\langle y\rangle x_i - 2S\langle x\rangle\langle y\rangle\right) \quad (73)$$

Deviation between the observed and predicted ChrY representations was minimized as the first derivative of *F* with respect to *S* vanishes:

$$\frac{dF}{dS} = 0 = \frac{d}{dS}\sum_i \left(y_i^2 + S^2x_i^2 + S^2\langle x\rangle^2 + \langle y\rangle^2 - 2Sx_iy_i + 2S\langle x\rangle y_i\right.$$
$$\left. -2\langle y\rangle y_i - 2S^2\langle x\rangle x_i + 2S\langle y\rangle x_i - 2S\langle x\rangle\langle y\rangle\right)$$
$$= 2\sum_i \left[S\left(x_i^2 - 2\langle x\rangle x_i + \langle x\rangle^2\right) + \langle x\rangle y_i + \langle y\rangle x_i - \langle x\rangle\langle y\rangle - x_iy_i\right]$$
$$= 2S\sum_i (x_i - \langle x\rangle)^2 + 2\sum_i (x_i - \langle x\rangle)(y_i - \langle y\rangle) \quad (74)$$

Solving Eq. 74 for *S* yielded the optimal slope:

$$S = \frac{\sum_i (x_i - \langle x\rangle)(y_i - \langle y\rangle)}{\sum_i (x_i - \langle x\rangle)^2} \quad (75)$$

The intercept I was introduced to simplify Eq. 72 and was defined as follows:

$$I = \langle y\rangle - S\langle x\rangle \quad (76)$$

Combining Eqs. 72 and 76 yielded the following expression:

$$y = I + Sx \quad (77)$$

The expression for fetal fraction based on ChrX has already been described in an earlier technology disclosure. It reads as follows:

$$f = -2\left(\frac{x}{\langle x\rangle} - 1\right) \quad (78)$$

The fetal fraction was evaluated from measured ChrY representation by introducing Eq. 77 into Eq. 78 and rearranging the resulting expression:

$$f = -2\left(\frac{y-I}{S\langle x\rangle} - 1\right) = 2\frac{I + S\langle x\rangle - y}{S\langle x\rangle} \qquad \text{QED} \quad (79)$$

This proved Eq. 71.

The R code used to evaluate model parameters I and S is shown below:

```
slope <- sum( ( chrRepresentations[ as.character( selectorBoys ), "Y" ] - girlsMedianY ) *
        ( chrRepresentations[ as.character( selectorBoys ), "X" ] - girlsMedianX ), na.rm=T ) /
    sum( ( chrRepresentations[ as.character( selectorBoys ), "X" ] - girlsMedianX ) *
        ( chrRepresentations[ as.character( selectorBoys ), "X" ] - girlsMedianX ), na.rm=T );
```

```
intercept <- girlsMedianY - slope * girlsMedianX;
slope
#[1] -0.3252008
intercept
#[1] 0.01560178
```

The model parameters $\langle x \rangle$, $\langle y \rangle$, I, and S have the following values:

$$\langle x \rangle = 0.04765159$$

$$\langle y \rangle = 0.0001054401$$

$$I = 0.01560178$$

$$S = -0.3252008$$

The above values were derived from LDTv2CE PERUN data. The parameter values applied to v2 chemistry and the current PERUN parameterization (extended to cover chromosomes X and Y). The R code that evaluated fetal fractions from the measured chromosome representations of chromosomes 13, 18, 21, X, and Y is shown below:

```
evaluateFetalFractions <- function( chrRepresentations,
    median21=0.01265891,
    median18=0.02887715,
    median13=0.03599132,
    girlsMedianX=0.04765159,
    girlsMedianY=0.0001054401,
    #
    boysInterceptY=0.01560178,
    boysSlopeY= -0.3252008
    ) {
    fetalFractions <- c(
        200 * ( chrRepresentations[ "21" ] / median21 - 1 ),
        200 * ( chrRepresentations[ "18" ] / median18 - 1 ),
        200 * ( chrRepresentations[ "13" ] / median13 - 1 ),
        -200 * ( chrRepresentations[ "X" ] / girlsMedianX - 1 ),
        -200 * ( chrRepresentations[ "Y" ] - boysInterceptY - boysSlopeY * girlsMedianX ) /
            ( boysSlopeY * girlsMedianX )
        );
    names( fetalFractions ) <- c( "21", "18", "13", "X", "Y" );
    return( fetalFractions );
} # evaluateFetalFractions
```

Example 11:

In this Example, a new fusion protein was used that captures methylated DNA in combination with CpG Island array to identify genomic regions that are differentially methylated between fetal placenta tissue and maternal blood. A stringent statistical approach was used to only select regions which show little variation between the samples, and hence suggest an underlying biological mechanism. Eighty-five differentially methylated genomic regions predominantly located on chromosomes 13, 18 and 21 were validated. For this validation, a quantitative mass spectrometry based approach was used that interrogated 261 PCR amplicons covering these 85 regions. The results are in very good concordance (95% confirmation), proving the feasibility of the approach.

Ten paired maternal and placental DNA samples were used to identify differentially methylated regions. These results were validated using a mass spectrometry-based quantitative methylation assay. First, genomic DNA from maternal buffy coat and corresponding placental tissue was first extracted. Next the MBD-FC was used to capture the methylated fraction of each DNA sample. See FIGS. 138-142. The two tissue fractions were labeled with different fluorescent dyes and hybridized to an Agilent® CpG Island microarray. See FIG. 141. This was done to identify differentially methylated regions that could be utilized for prenatal diagnoses. Therefore, two criteria were employed to select genomic regions as potential enrichment markers: the observed methylation difference had to be present in all tested sample pairs, and the region had to be more than 200 bp in length.

DNA preparation and fragmentation

Genomic DNA (gDNA) from maternal buffy coat and placental tissue was prepared using the QIAamp DNA Mini Kit™ and QIAamp DNA Blood Mini Kit™, respectively, from Qiagen® (Hilden, Germany). For MCIp, gDNA was quantified using the NanoDrop ND 1000™ spectrophotometer (Thermo Fisher®, Waltham, MA,USA). Ultrasonication of 2.5 µg DNA in 500 µl TE buffer to a mean fragment size of 300-500 bp was carried out with the Branson Digital Sonifier 450™ (Danbury, CT, USA) using the following settings: amplitude 20%, sonication time 110 seconds, pulse on/pulse off time 1.4/0.6 seconds. Fragment range was monitored using gel electrophoresis.

Methyl-CpG Immunoprecipitation

Per sample, 56 µg purified MBD-Fc protein and 150 µl of Protein A Sepharose 4 Fast Flow beads (Amersham Biosciences®, Piscataway, NJ, USA) were rotated in 15 ml TBS overnight at 4°C. Then, theMBD-Fc beads (150 µl/assay) were transferred and dispersed in to 2 ml Ultrafree-CL centrifugal filter devices (Millipore®, Billerica, MA, USA) and spin-washed three times with Buffer A (20 mM Tris-HCl, pH8.0, 2 mM MgCl2, 0.5 mM EDTA 300 mM NaCl, 0.1% NP-40). Sonicated DNA (2 µg) was added to the washed MBD-Fc beads in 2 ml Buffer A and rotated for 3 hours at 4°C. Beads were centrifuged to recover unbound DNA fragments (300 mM fraction) and subsequently washed twice with 600 µl of buffers containing increasing NaCl concentrations (400, 500, 550, 600, and 1000 mM). The flow through of each wash step was collected in separate tubes and desalted using a MinElute PCR Purification Kit™ (Qiagen®). In parallel, 200 ng sonicated input DNA was processed as a control using the MinElute PCR Purification Kit™ (Qiagen®).

Microarray handling and analysis

To generate fluorescently labeled DNA for microarray hybridization, the 600 mM and 1M NaCl fractions (enriched methylated DNA) for each sample were combined and labeled with either Alexa Fluor 555–aha–dCTP (maternal) or Alexa Fluor 647–aha–dCTP (placental) using the BioPrime Total Genomic Labeling System™ (Invitrogen®, Carlsbad, CA, USA). The labeling reaction was carried out according to the manufacturer's manual. The differently labeled genomic DNA fragments of matched maternal/placental pairs were combined to a final volume of 80 µl, supplemented with 50 µg Cot-1 DNA (Invitrogen®), 52 µl of Agilent 10X blocking reagent (Agilent Technologies®, Santa Clara, CA, USA), 78 µl of deionized formamide, and 260 µl Agilent 2X hybridization buffer. The samples were heated to 95°C for 3 min, mixed, and subsequently incubated at 37°C for 30 min. Hybridization on Agilent CpG Island Microarray Kit™ was then carried out at 67°C for 40 hours using an Agilent SureHyb™ chamber and an Agilent hybridization oven. Slides were washed in Wash I (6X SSPE, 0.005% N-lauroylsarcosine) at room temperature for 5 min and in Wash II (0.06X SSPE) at 37°C for an additional 5 min. Next, the slides were submerged in acetonitrile and Agilent Ozone Protection Solution™, respectively, for 30 seconds. Images were scanned immediately and analyzed using an Agilent DNA Microarray Scanner™. Microarray images were processed using Feature Extraction Software v9.5 and the standard CGH protocol.

Bisulfite Treatment

Genomic DNA sodium bisulfite conversion was performed using EZ-96 DNA Methylation Kit™ (ZymoResearch, Orange County, CA). The manufacturer's protocol was followed using 1ug of genomic DNA and the alternative conversion protocol (a two temperature DNA denaturation).

Quantitative Methylation Analysis

Sequenom's MassARRAY® System was used to perform quantitative methylation analysis. This system utilizes matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry in combination with RNA base specific cleavage (Sequenom® MassCLEAVE™). A detectable pattern is then analyzed for methylation status. PCR primers were designed using Sequenom® EpiDESIGNER™ (www.epidesigner.com). A total of 261 amplicons, covering 85 target regions, were used for validation (median amplification length = 367 bp, min = 108, max = 500; median number of CpG's per amplicon =23, min = 4, max = 65). For each reverse primer, an additional T7 promoter tag for in-vivo transcription was added, as well as a 10mer tag on the forward primer to adjust for melting temperature differences. The MassCLEAVE(tm) biochemistry was performed as previously described (Ehrich M, et al. (2005) Quantitative high-throughput analysis of DNA methylation patterns by base specific cleavage and mass spectrometry. PNAS USA 102:15785-15790). Mass spectra were acquired using a MassARRAY™ Compact MALDI-TOF (Sequenom®, San Diego) and methylation ratios were generated by the EpiTYPER™ software v1.0 (Sequenom®, San Diego).

Statistical analysis

All statistical calculations were performed using the R statistical software package (www.r-project.org). First, the array probes were grouped based on their genomic location. Subsequent probes that were less than 1000 bp apart were grouped together. To identify differentially methylated regions, a control sample was used as reference. In the control sample, the methylated fraction of a blood derived control DNA was hybridized against itself. Ideally this sample should show log ratios of the two color channels around 0. However because of the variability in hybridization behavior, the probes show a mean log ratio of 0.02 and a standard deviation of 0.18. Next the log ratios observed in the samples were compared to the control sample. A two way, paired t-test was used to test the NULL hypothesis that the groups are identical. Groups that contained less than 4 probes were excluded from the analysis. For groups

including four or five probes, all probes were used in a paired t-test. For Groups with six or more probes, a sliding window test consisting of five probes at a time was used, whereby the window was moved by one probe increments. Each test sample was compared to the control sample and the p-values were recorded. Genomic regions were selected as being differentially methylated if eight out of ten samples showed a p value < 0.01, or if six out of ten samples showed a p value < 0.001. The genomic regions were classified as being not differentially methylated when the group showed less than eight samples with a p value < 0.01 and less than six samples with a p value < 0.001. Samples that didn't fall in either category were excluded from the analysis. For a subset of genomic regions that have been identified as differentially methylated, the results were confirmed using quantitative methylation analysis.

The Go analysis was performed using the online GOstat tool (http://gostat.wehi.edu.au/cgibin/-goStat.pl). P values were calculated using Fisher's exact test.

Microarray-based marker discovery results

To identify differentially methylated regions a standard sample was used, in which the methylated DNA fraction of monocytes was hybridized against itself. This standard provided a reference for the variability of fluorescent measurements in a genomic region. Differentially methylated regions were then identified by comparing the log ratios of each of the ten placental/maternal samples against this standard. Because the goal of this study was to identify markers that allow the reliable separation of maternal and fetal DNA, the target selection was limited to genes that showed a stable, consistent methylation difference over a contiguous stretch of genomic DNA. This focused the analysis on genomic regions where multiple probes indicated differential methylation. The selection was also limited to target regions where all samples showed differential methylation, excluding those with strong inter-individual differences. Two of the samples showed generally lower log ratios in the microarray analysis. Because a paired test was used for target selection, this did not negatively impact the results.

Based on these selection criteria, 3043 genomic regions were identified that were differentially methylated between maternal and fetal DNA. 21778 regions did not show a methylation difference. No inter-chromosomal bias in the distribution of differentially methylated regions was observed. The differentially methylated regions were located next to or within 2159 known genes. The majority of differentially methylated regions are located in the promoter area (18%) and inside

the coding region (68%), while only few regions are located downstream of the gene (7%) or at the transition from promoter to coding region (7%). Regions that showed no differential methylation showed a similar distribution for promoter (13%) and downstream (5%) locations, but the fraction of regions located in the transition of promoter to coding region was higher (39%) and the fraction inside the coding region was lower (43%).

It has been shown in embryonic stem cells (ES) that genes targeted by the polycomb repressive complex2 (PRC2) are enriched for genes regulating development (Lee TI, et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125:301-313). It has also been shown that differentially methylated genes are enriched for genes targeted by PRC2 in many cancer types (Ehrich M, et al. (2008) Cytosine methylation profiling of cancer cell lines. PNAS USA 105:4844-48). The set of genes identified as differentially methylated in this study is also enriched for genes targeted by PRC2 (p-value < 0.001, odds ratio = 3.6, 95% CI for odds ratio= 3.1 – 4.2). A GO analysis of the set of differentially methylated genes reveals that this set is significantly enriched for functions important during development. Six out of the ten most enriched functions include developmental or morphogenic processes [anatomical structure morphogenesis (GO:0009653, p value =0), developmental process (GO:0032502, p value = 0), multicellular organismal development (GO:0007275, p value = 0), developmental of an organ (GO:0048513, p value = 0), system development (GO:0048731, p value = 0) and development of an anatomical structure (GO:0048856, p value = 0)].

Validation using Sequenom® EpiTYPER™

To validate the microarray findings, 63 regions from chromosomes 13, 18 and 21 and an additional 26 regions from other autosomes were selected for confirmation by a different technology. Sequenom EpiTYPER™ technology was used to quantitatively measure DNA methylation in maternal and placental samples. For an explanation of the EpiTYPER™ methods, see Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D (2005) Quantitative high-throughput analysis of DNA methylation patterns by base specific cleavage and mass spectrometry. PNAS USA 102:15785-15790). For each individual CpG site in a target region the average methylation value across all maternal DNA samples and across all placenta samples was calculated. The difference between average maternal and placenta methylation was then compared to the microarray results. The results from the two technologies were in good concordance (see FIG. 144). For 85 target regions the quantitative results confirm the

microarray results (95% confirmation rate). For 4 target regions, all located on chromosome 18, the results could not be confirmed. The reason for this discrepancy is currently unclear.

In contrast to microarrays, which focus on identification of methylation differences, the quantitative measurement of DNA methylation allowed analysis of absolute methylation values. In the validation set of 85 confirmed differentially methylated regions, a subset of 26 regions is more methylated in the maternal DNA sample and 59 regions are more methylated in the placental sample (see Table 3A). Interestingly, genes that are hypomethylated in the placental samples tend to show larger methylation differences than genes that are hypermethylated in the placental sample (median methylation difference for hypomethylated genes = 39%, for hypermethylated genes = 20%).

Example 12

Example 12 describes a non-invasive approach for detecting the amount of fetal nucleic acid present in a maternal sample (herein referred to as the "Fetal Quantifier Method"), which may be used to detect or confirm fetal traits (e.g., fetal sex of RhD compatibility), or diagnose chromosomal abnormalities such as Trisomy 21 (both of which are herein referred to as the "Methylation-Based Fetal Diagnostic Method"). FIG. 147 shows one embodiment of the Fetal Quantifier Method, and FIG. 148 shows one embodiment of the Methylation-Based Fetal Diagnostic Method. Both processes use fetal DNA obtained from a maternal sample. The sample comprises maternal and fetal nucleic acid that is differentially methylated. For example, the sample may be maternal plasma or serum. Fetal DNA comprises approximately 2-30% of the total DNA in maternal plasma. The actual amount of fetal contribution to the total nucleic acid present in a sample varies from pregnancy to pregnancy and can change based on a number of factors, including, but not limited to, gestational age, the mother's health and the fetus' health.

As described herein, the technical challenge posed by analysis of fetal DNA in maternal plasma lies in the need to be able to discriminate the fetal DNA from the co-existing background maternal DNA. The methods of the present technology exploit such differences, for example, the differential methylation that is observed between fetal and maternal DNA, as a means to enrich for the relatively small percentage of fetal DNA present in a sample from the mother. The non-invasive nature of the approach provides a major advantage over conventional methods of prenatal diagnosis such as, amniocentesis, chronic villus sampling and cordocentesis, which are associated

with a small but finite risk of fetal loss. Also, because the method is not dependent on fetal cells being in any particular cell phase, the method provides a rapid detection means to determine the presence and also the nature of the chromosomal abnormality. Further, the approach is sex-independent (i.e., does not require the presence of a Y-chromosome) and polymorphic-independent (i.e., an allelic ratio is not determined). Thus, the compositions and methods of the technology herein represent improved universal, noninvasive approaches for accurately determining the amount of fetal nucleic acid present in a maternal sample.

Assay design and advantages

There is a need for accurate detection and quantification of fetal DNA isolated noninvasively from a maternal sample. The present technology takes advantage of the presence of circulating, cell free fetal nucleic acid (ccfDNA) in maternal plasma or serum. In order to be commercially and clinically practical, the methods of the technology herein should only consume a small portion of the limited available fetal DNA. For example, less than 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5% or less of the sample. Further, the approach should preferably be developed in a multiplex assay format in which one or more (preferably all) of the following assays are included:

- Assays for the detection of total amount of genomic equivalents present in the sample, i.e., assays recognizing both maternal and fetal DNA species;
- Assays for the detection of fetal DNA isolated from a male pregnancy, i.e., sequences specific for chromosome Y;
- Assays specific for regions identified as differentially methylated between the fetus and mother; or
- Assays specific for regions known to be hypomethylated in all tissues to be investigated, which can serve as a control for restriction efficiency.

Other features of the assay may include one or more of the following:

- For each assay, a target-specific, competitor oligonucleotide that is identical, or substantially identical, to the target sequence apart from a distinguishable feature of the competitor, such as a difference in one or more nucleotides relative to the target sequence. This oligonucleotide when added into the PCR reaction will be co-amplified with the target and a ratio

obtained between these two PCR amplicons will indicate the number of target specific DNA sequences (e.g., fetal DNA from a specific locus) present in the maternal sample.

- The amplicon lengths should preferably be of similar length in order not to skew the amplification towards the shorter fragments. However, as long as the amplification efficiency is about equal, different lengths may be used.
- Differentially methylated targets can be selected from Tables 3A-3C or from any other targets known to be differentially methylated between mother and fetus. These targets can be hypomethylated in DNA isolated from non-pregnant women and hypermethylated in samples obtained from fetal samples. These assays will serve as controls for the restriction efficiency.
- The results obtained from the different assays can be used to quantify one or more of the following:
    - Total number of amplifiable genomes present in the sample (total amount of genomic equivalents);
    - The fetal fraction of the amplifiable genomes (fetal concentration or percentage); or
    - Differences in copy number between fetally-derived DNA sequences (for example, between fetal chromosome 21 and a reference chromosome such as chromosome 3).

Examples of assays used in the test

Below is an outline of the reaction steps used to perform a method of the technology herein, for example, as provided in FIG. 147. This outline is not intended to limit the scope of the technology herein. Rather it provides one embodiment of the technology herein using the Sequenom® MassARRAY® technology.

1) DNA isolation from plasma samples.

2) Digestion of the DNA targets using methylation sensitive restriction enzymes (for example, HhaI and HpaII).

For each reaction the available DNA was mixed with water to a final volume of 25 ul.
10 ul of a reaction mix consisting of 10 units HhaI, 10 units HpaII and a reaction buffer were added. The sample was incubated at an optimal temperature for the restriction enzymes. HhaI and HpaII

digest non-methylated DNA (and will not digest hemi- or completely methylated DNA). Following digestion, the enzymes were denatured using a heating step.

3) Genomic Amplification- PCR was performed in a total volume of 50 ul by adding PCR reagents (Buffer, dNTPs, primers and polymerase). Exemplary PCR and extend primers are provided below. In addition, synthetic competitor oligonucleotide was added at known concentrations.

4) Replicates (optional) - Following PCR the 50 ul reaction was split into 5 ul parallel reactions (replicates) in order to minimize variation introduced during the post PCR steps of the test. Post PCR steps include SAP, primer extension (MassEXTEND® technology), resin treatment, dispensing of spectrochip and MassARRAY.

5) Quantification of the Amplifiable Genomes – Sequenom MassARRAY® technology was used to determine the amount of amplification product for each assay. Following PCR, a single base extension assay was used to interrogate the amplified regions (including the competitor oligonucleotides introduced in step 3). Specific extend primers designed to hybridize directly adjacent to the site of interest were introduced. See extend primers provided below. These DNA oligonucleotides are referred to as iPLEX® MassEXTEND® primers. In the extension reaction, the iPLEX primers were hybridized to the complementary DNA templates and extended with a DNA polymerase. Special termination mixtures that contain different combinations of deoxy- and dideoxynucleotide triphosphates along with enzyme and buffer, directed limited extension of the iPLEX primers. Primer extension occurs until a complementary dideoxynucleotide is incorporated.

The extension reaction generated primer products of varying length, each with a unique molecular weight. As a result, the primer extension products can be simultaneously separated and detected using Matrix Assisted Laser Desorption/Ionization, Time-Of-Flight (MALDI-TOF) mass spectrometry on the MassARRAY® Analyzer Compact. Following this separation and detection, SEQUENOM's proprietary software automatically analyzes the data.

6) Calculating the amount and concentration of fetal nucleic acid – Methods for calculating the total amount of genomic equivalents present in the sample, the amount (and concentration) of fetal nucleic acid isolated from a male pregnancy, and the amount (and

concentration) of fetal nucleic based on differentially methylated targets are provided below and in FIGS. 155 and 156.

The above protocol can be used to perform one or more of the assays described below. In addition to the sequences provided immediately below, a multiplex scheme that interrogates multiple targets is provided in Table X below.

1) Assay for the quantification of the total number of amplifiable genomic equivalents in the sample.

Targets were selected in housekeeping genes not located on the chromosomes 13, 18, 21, X or Y. The targets should be in a single copy gene and not contain any recognition sites for the methylation sensitive restriction enzymes.

Underlined sequences are PCR primer sites, italic is the site for the single base extend primer and bold letter (C) is the nucleotide extended on human DNA.

ApoE Chromosome 19:45409835-45409922 DNA target sequence with interrogated nucleotide C in bold. All of the chromosome positions provided in this section are from the February 2009 UCSC Genome Build.

GATTGACAGTTTCTCCTTCCCCAGACTGGCCAATCACAG*GCAGGAAGATGAAGGTT***C**T GTGGGCTGCGTTGCTGGTCACATTCCTGGC (SEQ ID NO: 262)

ApoE Forward Primer: 5'-ACGTTGGATG-TTGACAGTTTCTCCTTCCCC (SEQ ID NO: 263) (Primer contains a 5' 10 bp MassTag separated by a dash)

ApoE Reverse Primer: 5'-ACGTTGGATG-GAATGTGACCAGCAACGCAG (SEQ ID NO: 264) (Primer contains a 5' 10 bp MassTag separated by a dash)

ApoE Extension Primer: 5'-GCAGGAAGATGAAGGTT [C/T] (SEQ ID NO: 265) Primer extends C on human DNA targets and T on synthetic DNA targets

ApoE synthetic competitor oligonucleotide: 5'-GATTGACAGTTTCTCCTTCCCCAGACTGGCCAATCACAGGCAGGAAGATGAAGGTTTTGTGGGCTGCGTTGCTGGTCACATTCCTGGC (SEQ ID NO: 266) (Bold T at position 57 is different from human DNA)

2) *Assay for the quantification of the total number of chromosome Y sequences in the sample.*

Targets specific for the Y-chromosome were selected, with no similar or paralog sequences elsewhere in the genome. The targets should preferably be in a single copy gene and not contain any recognition sites for the methylation sensitive restriction enzyme(s).

Underlined sequences are PCR primer sites, and italic nucleotide(s) is the site for the single-base extend primer and bold letter (C) is the nucleotide extended on human DNA.

SRY on chrY:2655628-2655717 (reverse complement) GAGTTTTGGATAGTAAAATAAGTTTCGAACTCTGGCACC*TTTCAATTTTGTCGCACT***C**TCCTTGTTTTTGACAATGCAATCATATGCTTC (SEQ ID NO: 267)

SRY Forward Primer: 5'-ACG-TGGATAGTAAAATAAGTTTCGAACTCTG (SEQ ID NO: 268) (Primer contains a 5' 3 bp MassTag separated by a dash)

SRY Reverse Primer: 5'- GAAGCATATGATTGCATTGTCAAAAAC (SEQ ID NO: 269)

SRY Extension Primer: 5'-aTTTCAATTTTGTCGCACT [C/T] (SEQ ID NO: 270) Primer extends C on human DNA targets and T on synthetic DNA targets. 5' Lower case "a" is a non-complementary nucleotide

SRY synthetic competitor oligonucleotide: 5'-GAGTTTTGGATAGTAAAATAAGTTTCGAACTCTGGCACCTTTCAATTTTGTCGCACT**T**TCCTTGTTTTTGACAATGCAATCATATGCTTC (SEQ ID NO: 271)

3) *Assay for the quantification of fetal methylated DNA sequences present in the sample.*

Targets were selected in regions known to be differentially methylated between maternal and fetal DNA. Sequences were selected to contain several restriction sites for methylation sensitive enzymes. For this study the HhaI (GCGC) and HpaII (CCGG) enzymes were used.

Underlined sequences are PCR primer sites, italic is the site for the single base extend primer and bold letter (C) is the nucleotide extended on human DNA, lower case letter are recognition sites for the methylation sensitive restriction enzymes.

TBX3 on chr12:115124905-115125001
GAACTCCTCTTTGTCTCTGCGTGCccggcgcgcCC*CCCTCccggTGGGTGATAAA***C**CCACT CTGgcgccggCCATgcgcTGGGTGATTAATTTGCGA (SEQ ID NO: 272)

TBX3 Forward Primer: 5'- ACGTTGGATG-TCTTTGTCTCTGCGTGCCC (SEQ ID NO: 273) (Primer contains a 5' 10 bp MassTag separated by a dash)

TBX3 Reverse Primer: 5'- ACGTTGGATG-TTAATCACCCAGCGCATGGC (SEQ ID NO: 274) (Primer contains a 5' 10 bp MassTag separated by a dash)

TBX3 Extension Primer: 5'- CCCCTCCCGGTGGGTGATAAA [C/T] (SEQ ID NO: 275)
Primer extends C on human DNA targets and T on synthetic DNA targets. 5' Lower case "a" is a non-complementary nucleotide

TBX3 synthetic competitor oligonucleotide: 5'-
GAACTCCTCTTTGTCTCTGCGTGCCCGGCGCGCCCCCCTCCCGGTGGGTGATAAA**T**C CACTCTGGCGCCGGCCATGCGCTGGGTGATTAATTTGCGA (SEQ ID NO: 276)

4) Control Assay for the enzyme restriction efficiency.

Targets were selected in regions known not to be methylated in any tissue to be investigated. Sequences were selected to contain no more than one site for each restriction enzyme to be used.

Underlined sequences are PCR primer sites, italic nucleotide(s) represent the site for the single-base extend primer and bold letter (G) is the reverse nucleotide extended on human DNA, lower case letter are recognition sites for the methylation sensitive restriction enzymes.

CACNA1G chr17:48637892-48637977 (reverse complement)
CCATTGGCCGTCCGCCGTGGCAGTGCGGGCGGGAgcgcAG**G***GAGAGAACCACAGCTG*
*GAAT*CCGATTCCCACCCCAAAACCCAGGA (SEQ ID NO: 277)

HhaI Forward Primer: 5'- ACGTTGGATG-CCATTGGCCGTCCGCCGTG (SEQ ID NO: 278) (Primer contains a 5' 10 bp MassTag separated by a dash)

HhaI Reverse Primer: 5'- ACGTTGGATG-TCCTGGGTTTTGGGGTGGGAA (SEQ ID NO: 279) (Primer contains a 5' 10 bp MassTag separated by a dash)

HhaI Extension Primer: 5'- TTCCAGCTGTGGTTCTCTC (SEQ ID NO: 280)

HhaI synthetic competitor oligonucleotide: 5'-
CCATTGGCCGTCCGCCGTGGCAGTGCGGGCGGGAGCGCAG**A***GAGAGAACCACAGCT*
*GGAAT*CCGATTCCCACCCCAAAACCCAGGA (SEQ ID NO: 281)

Validation experiments

The sensitivity and accuracy of the present technology was measured using both a model system and clinical samples. In the different samples, a multiplex assay was run that contains 2 assays for total copy number quantification, 3 assays for methylation quantification, 1 assay specific for chromosome Y and 1 digestion control assay. See Table X. Another multiplex scheme with additional assays is provided in Table Y.

TABLE X: PCR Primers and Extend Primers

Gene ID	*	First Primer (SEQ ID NOS 282-288, respectively, in order of appearance)	Second Primer (SEQ ID NOS 289-295, respectively, in order of appearance)	Extend Primer (SEQ ID NOS 296-302, respectively, in order of appearance)
SOX14	M	ACGTTGGATGACATGGTCGGCCCCACGGAAT	ACGTTGGATGCTCCTTCCTAGTGTGAGAACCG	CAGGTTCCGGGGCTTGGG
HhaI_CTRL	D	ACGTTGGATGACCCATTGGCCGTCCGCCGT	ACGTTGGATGTTTTGGGGTGGGAATCGGATT	CGCAGGGAGAGAACCACAG
TBX3	M	ACGTTGGATGGAACTCCTCTTTGTCTCTGCG	ACGTTGGATGTGGCATGGCCGGCGCCAGA	CCCCTCCCGGTGGGTGATAAA
SRY	Y	ACGTTGGATGCGCAGCAACGGGACCGCTACA	ACGTTGGCATCTAGGTAGGTCTTTGTAGCCAA	AAAGCTGTAGGACAATCGGGT
ALB	T	ACGTTGCGTAGCAACCTGTTACATATTAA	ACGTTGGATCTGAGCAAAGGCAATCAACACCC	CATTTTTCTACATCCTTTGTTT
EDG6	M	ACGTTGGATGCATAGAGGCCCATGATGGTGG	ACGTTGGATGACCTTCTGCCCCTCTACTCCAA	agAAGATCACCAGGCAGAAGAGG
RNaseP	T	ACGTTGGATGGTGTGGTCAGCTCTTCCCTTCAT	ACGTTGGCCCACATGTAATGTGTTGAAAAAGCA	ACTTGGAGAACAAAGGACACCGTTA

TABLE X: Competitor Oligonucleotide Sequence

Gene ID	*	Competitor Oligonucleotide Sequence (SEQ ID NOS 303-309, respectively, in order of appearance)
SOX14	M	GGTCGGCCCCACGGAATCCCGGCTCTGTGTGCGCCCAGGTTCCGGGGCTTGGGTGTTGCCGGTTCTCACACTAGGAAGGAG
HhaI_CTRL	D	CCATTGGCCGTCCGCCGTGGCAGTGCGGGCGGGAGCGCAGAGAGAGAACCACAGCTGGAATCCGATTCCCACCCCAAAA
TBX3	M	GAACTCCTCTTTGTCTCTGCGTGCCCGGCGCGCCCCCCTCCCGGTGGGTGATAAATCCACTCTGGCGCCGGCCATGC
SRY	Y	GCAGCAACGGGACCGCTACAGCCACTGGACAAAGCCGTAGGACAATCGGGTAACATTGGCTACAAAGACCTACCTAGATGC
ALB	T	GCGTAGCAACCTGTTACATATTAAAGTTTTATTATACTACATTTTTCTACATCCTTTGTTTCAGAGTGTTGATTGCCTTTGCTCAGTATCTTCAG
EDG6	M	CCTTCTGCCCCTCTACTCCAAGCGCTACACCCTCTTCTGCCTGGTGATCTTTGCCGGCGTCCTGGCCACCATCATGGGCCTCTATG
RNaseP	T	GTGTGGTCAGCTCTTCCCTTCATCACATACTTGGAGAACAAAGGACACCGTTATCCATGCTTTTTCAACACATTACATGTGGG

TABLE Y: PCR Primers and Extend Primers

Gene ID	*	First Primer (SEQ ID NOS 310-311, 263, 312, 268, 313, 273 and 314-316, respectively, in order of appearance)	Second Primer (SEQ ID NOS 317-318, 264, 289, 269, 319, 274 and 320-322, respectively, in order of appearance)	Extend Primer (SEQ ID NOS 323-331 and 300, respectively, in order of appearance)
EDG6	M	ACGTTGGATGTTCTGCCCCTCTACTCCAAG	ACGTTGGATGCATAGAGGCCCATGATGGTG	TTCTGCCTGGTGATCTT
RNAseP	T	ACGTTGGATGTCAGCTCTTCCCTTCATCAC	ACGTTGGATGCCTACCTCCCACATGTAATGT	AACAAAGGACACCGTTA
ApoE	T	ACGTTGGATGTTGACAGTTTCTCCTTCCCC	ACGTTGGATGGAATGTGACCAGCAACGCAG	GCAGGAAGATGAAGGTT
SOX14	M	ACGTTGGATGCGGTCGGCCCCACGGAAT	ACGTTGGATGCTCCTTCCTAGTGTGAGAACCG	aAGGTTCCGGGGCTTGGG
SRY no2	Y	ACGTGGATAGTAAAATAAGTTTCGAACTCTG	GAAGCATATGATTGCATTGTCAAAAAC	aTTTCAATTTTGTCGCACT
SRY no1	Y	ACGTTGGATGCACAGCTCACCGCAGCAACG	ACGTTGGATGCTAGGTAGGTCTTTGTAGCCAA	AGCTGTAGGACAATCGGGT
TBX3	M	ACGTTGGATGTCTTTGTCTCTGCGTGCCC	ACGTTGGATGTTAATCACCCAGCGCATGGC	CCCTCCCGGTGGGTGATAAA
CACNA1G dig CTRL 1	D	ACGTTGGATGGACTGAGCCCCAGAACTCG	ACGTTGGATGGTGGGTTTGTGCTTTCCACG	AGGGCCGGGGTCTGCGCGTG
DAPK1 dig CTRL 2	D	ACGTTGGATGAAGCCAAGTTTCCCTCCGC	ACGTTGGATGCTTTTGCTTTCCCAGCCAGG	GAGGCACTGCCCGGACAAACC
ALB	T	ACGTTAGCGTAGCAACCTGTTACATATTAA	ACGTTGGATGCTGAGCAAAGGCAATCAACA	CATTTTTCTACATCCTTTGTTT

TABLE Y: Competitor Oligonucleotide Sequence

Gene ID	*	Competitor (SEQ ID NOS 308, 332, 266, 333, 271, 334, 276 and , respectively, in order of appearance)
EDG6	M	CCTTCTGCCCCTCTACTCCAAGCGCTACACCCTCTTCTGCCTGGTGATCTTTGCCGGCGTCCTGGCCACCATCATGGGCCTCTATG
RNAseP	T	GTGTGGTCAGCTCTTCCCTTCATCACATACTTGGAGAACAAAGGACACCGTTATCCATGCTTTTTCAACACATTACATGTGGGAGGTAGG
ApoE	T	GATTGACAGTTTCTCCTTCCCCAGACTGGCCAATCACAGGCAGGAAGATGAAGGTTTTGTGGGCTGCGTTGCTGGTCACATTCCTGGC
SOX14	M	AAAACCAGAGATTCGCGGTCGGCCCCACGGAATCCCGGCTCTGTGTGCGCCCAGGTTCCGGGGCTTGGGTGTTGCCGGTTCTCACACTAGGAAGGAGC

SRY no2	Y	GAGTTTTGGATAGTAAAATAAGTTTCGAACTCTGGCACCTTTCAATTTTGTCGCACTTTCCTTGTTTTTGACAATGCAATCATATGCTTC
SRY no1	Y	GCAGCCAGCTCACCGCAGCAACGGGACCGCTACAGCCACTGGACAAAGCTGTAGGACAATCGGGTGACATTGGCTACAAAGACCTACCTAGATGC
TBX3	M	GAACTCCTCTTTGTCTCTGCGTGCCCGGCGCGCCCCCCTCCCGGTGGGTGATAAATCCACTCTGGCGCCGGCCATGCGCTGGGTGATTAATTTGCG A
CACNA1G dig CTRL 1	D	GTGGGTTTGTGCTTTCCACGCGTGCACACACACGCGCAGACCCCGGCCCTTGCCCCGCCTACCTCCCCGAGTTCTGGGGCTCAGTC
DAPK1 dig CTRL 2	D	GCGCCAGCTTTTGCTTTCCCAGCCAGGGCGCGGTGAGGTTTGTCCGGGCAGTGCCTCGAGCAACTGGGAAGGCCAAGGCGGAGGGAAAC
ALB	T	GCGTAGCAACCTGTTACATATTAAAGTTTTATTATACTACATTTTTCTACATCCTTTGTTTTAGGGTGTTGATTGCCTTTGCTCAGTATCTTCAGC

T=Assay for Total Amount

M=Assay for Methylation quantification

Y= Y-Chromosome Specific Assay

D=Digestion control

Model system using genomic DNA

In order to determine the sensitivity and accuracy of the method when determining the total number of amplifiable genomic copies in a sample, a subset of different DNA samples isolated from the blood of non-pregnant women was tested. Each sample was diluted to contain approximately 2500, 1250, 625 or 313 copies per reaction. The total number of amplifiable genomic copies was obtained by taking the mean DNA/competitor ratio obtained from the three total copy number assays. The results from the four different samples are shown in FIG. 149.

To optimize the reaction, a model system was developed to simulate DNA samples isolated from plasma. These samples contained a constant number of maternal non-methylated DNA and were spiked with different amounts of male placental methylated DNA. The samples were spiked with amounts ranging from approximately 0 to 25% relative to the maternal non-methylated DNA. The results are shown in FIGS. 150A and 150B. The fraction of placental DNA was calculated using the ratios obtained from the methylation assays (FIG. 150A), the SRY markers (FIG. 150B) and the total copy number assays. The primer sequences for the methylation assays (TBX), Y-chromosome assays (SRY) and total copy number (APOE) are provided above. The model system demonstrated that the methylation-based method performed equal to the Y-chromosome method (SRY markers), thus validating the methylation-based method as a sex-independent fetal quantifier.

Plasma samples

To investigate the sensitivity and accuracy of the methods in clinical samples, 33 plasma samples obtained from women pregnant with a male fetus were investigated using the multiplex scheme from Table X. For each reaction, a quarter of the DNA obtained from a 4ml extraction was used in order to meet the important requirement that only a portion of the total sample is used.

Total copy number quantification

The results from the total copy number quantification can be seen in FIGS. 151A and 151B. In FIG. 151A, the copy number for each sample is shown. Two samples (nos. 25 and 26) have a significantly higher total copy number than all the other samples. In general, a mean of

approximately 1300 amplifiable copies/ml plasma was obtained (range 766-2055). FIG. 151B shows a box-and-whisker plot of the given values, summarizing the results.

Correlation between results obtained from the methylation markers and the Y-chromosome marker

In FIGS. 152A and 152B, the numbers of fetal copies for each sample are plotted. As all samples were from male pregnancies. The copy numbers obtained can be calculated using either the methylation or the Y-chromosome-specific markers. As can be seen in FIG. 152B, the box-and-whisker plot of the given values indicated minimal difference between the two different measurements.

The results showing the correlation between results obtained from the methylation markers and the Y-chromosome marker (SRY) is shown in FIG. 153. Again, the methylation-based method performed equal to the Y-chromosome method (SRY markers), further validating the methylation-based method as a sex-independent and polymorphism-independent fetal quantifier. The multiplexed assays disclosed in Table X were used to determine the amount fetal nucleic.

Finally, the digestion efficiency was determined by using the ratio of digestion for the control versus the competitor and comparing this value to the mean total copy number assays. See FIG. 154. Apart from sample 26 all reactions indicate the efficiency to be above 99%.

Data Analysis

Mass spectra analysis was done using Typer 4 (a Sequenom software product). The peak height (signal over noise) for each individual DNA analyte and competitor assay was determined and exported for further analysis.

The total number of molecules present for each amplicon was calculated by dividing the DNA specific peak by the competitor specific peak to give a ratio. (The “DNA” Peak in FIGS. 155 and 156 can be thought of as the analyte peak for a given assay). Since the number of competitor molecules added into the reaction is known, the total number of DNA molecules can be determined by multiplying the ratio by the number of added competitor molecules.

The fetal DNA fraction (or concentration) in each sample was calculated using the Y-chromosome-specific markers for male pregnancies and the mean of the methylated fraction for all pregnancies. In brief, for chromosome Y, the ratio was obtained by dividing the analyte (DNA) peak by the competitor peak and multiplying this ratio by the number of competitor molecules added into the reaction. This value was divided by a similar ratio obtained from the total number of amplifiable genome equivalents determination (using the Assay(s) for Total Amount). See FIG. 155. Since the total amount of nucleic acid present in a sample is a sum of maternal and fetal nucleic acid, the fetal contribution can be considered to be a fraction of the larger, background maternal contribution. Therefore, translating this into the equation shown in FIG. 155, the fetal fraction (k) of the total nucleic acid present in the sample is equal to the equation: $k=2xR/(1-2R)$, where R is the ratio between the Y-chromosome amount and the total amount. Since the Y-chromosome is haploid and Assays for the Total Amount are determined using diploid targets, this calculation is limited to a fetal fraction smaller than 50% of the maternal fraction.

In FIG. 156, a similar calculation for the fetal concentration is shown by using the methylation specific markers (see Assays for Methylation Quantification). In contrast to Y-chromosome specific markers, these markers are from diploid targets, therefore, the limitations stated for the Y-Chromosome Specific Assay can be omitted. Thus, the fetal fraction (k) can be determined using the equation: $k=R(1-R)$, where R is the ratio between the methylation assay and the total assay.

Simulation

A first simple power calculation was performed that assumes a measurement system that uses 20 markers from chromosome 21, and 20 markers from one or more other autosomes. Starting with 100 copies of fetal DNA, a measurement standard deviation of 25 copies and the probability for a type I error to be lower than 0.001, it was found that the methods of the technology herein will be able to differentiate a diploid from a triploid chromosome set in 99.5% of all cases. The practical implementation of such an approach could for example be achieved using mass spectrometry, a system that uses a competitive PCR approach for absolute copy number measurements. The method can run 20 assays in a single reaction and has been shown to have a standard deviation in repeated measurements of around 3 to 5%. This method was used in combination with known methods for differentiating methylated and non-methylated nucleic acid, for example, using methyl-binding agents to separate nucleic acid or using methylation-sensitive enzymes to digest maternal nucleic acid. FIG. 145 shows the effectiveness of MBD-FC protein (a methyl-binding agent) for

capturing and thereby separating methylated DNA in the presence of an excess of unmethylated DNA (see FIG. 145).

A second statistical power analysis was performed to assess the predictive power of an embodiment of the Methylation-Based Fetal Diagnostic Method described herein. The simulation was designed to demonstrate the likelihood of differentiating a group of trisomic chromosome 21 specific markers from a group of reference markers (for example, autosomes excluding chromosome 21). Many parameters influence the ability to discriminate the two populations of markers reliably. For the present simulation, values were chosen for each parameter that have been shown to be the most likely to occur based on experimentation. The following parameters and respective values were used:

Copy Numbers

Maternal copy numbers = 2000

Fetal copy numbers for chromosomes other than 21, X and Y = 200

Fetal copy numbers for chromosome 21 in case of euploid fetus = 200

Fetal copy numbers for chromosome 21 in case of aneuploid T21 fetus = 300

Percent fetal DNA (before methylation-based enrichment) = 10% (see above)

Methylation Frequency

Average methylation percentage in a target region for maternal DNA = 10%

Average methylation percentage in a target region for fetal DNA = 80%

Average percentage of non-methylated and non-digested maternal DNA (i.e., a function of restriction efficiency (among other things) = 5%

Number of assays targeting chromosome 21 = 10

Number of assays targeting chromosomes other than 21, X and Y = 10

The results are displayed in FIG. 157. Shown is the relationship between the coefficient of variation (CV) on the x-axis and the power to discriminate the assay populations using a simple t-test (y-axis). The data indicates that in 99% of all cases, one can discriminate the two population (euploid vs. aneuploid) on a significance level of 0.001 provided a CV of 5% or less. Based on this simulation, the method represents a powerful noninvasive diagnostic method for the prenatal detection of fetal aneuploidy that is sex-independent and will work in all ethnicities (i.e., no allelic bias).

Example 13 – Additional Differentially-Methylated Targets

Differentially-methylated targets not located on chromosome 21

Additional differentially-methylated targets were selected for further analysis based upon previous microarray analysis. See Example 11 for a description of the microarray analysis. During the microarray screen, differentially methylated regions (DMRs) were defined between placenta tissue and PBMC. Regions were selected for EpiTYPER confirmation based upon being hypermethylated in placenta relative to PBMC. After directionality of the change was selected for, regions were chosen based upon statistical significance with regions designed beginning with the most significant and working downward in terms of significance. These studies were performed in eight paired samples of PBMC and placenta. Additional non-chromosome 21 targets are provided in Table 3B, along with a representative genomic sequence from each target in Table 6B.

Differentially-methylated targets located on chromosome 21

The microarray screen uncovered only a subset of DMRs located on chromosome 21. The coverage of chromosome 21 by the microarray, however, was insufficient. Therefore a further analysis was completed to examine all 356 CpG islands on chromosome 21 using the standard settings of the UCSC genome browser. As shown in Table 3C below, some of these targets overlapped with those already examined in Table 5A. More specifically, CpG sites located on chromosome 21 including ~1000bp upstream and downstream of each CpG was investigated using Sequenom's EpiTYPER® technology. See Example 11, "*Validation using Sequenom® EpiTYPER™*" for a description of Sequenom's EpiTYPER® technology. These studies were

performed in eight paired samples of PBMC and placenta. In addition, since DMRs may also be located outside of defined CpG islands, data mining was performed on publicly available microarray data to identify potential candidate regions with the following characteristics: hypermethylated in placenta relative to maternal blood, not located in a defined CpG island, contained greater than 4 CpG dinucleotides, and contained a recognition sequence for methylation sensitive restriction enzymes. Regions that met these criteria were then examined using Sequenom's EpiTYPER® technology on eight paired PBMC and placenta samples. Additional chromosome 21 targets are provided in Table 3C, along with a representative genomic sequence from each target in Table 6C.

Tables 3B and 3C provide a description of the different targets, including their location and whether they were analyzed during the different phases of analysis, namely microarray analysis, EpiTYPER 8 analysis and EpiTYPER 73 analysis. A "YES" indicates it was analyzed and a "NO" indicates it was not analyzed. The definition of each column in Table 3B and 3C is listed below.

- Region Name: Each region is named by the gene(s) residing within the area defined or nearby. Regions where no gene name is listed but rather only contain a locus have no refseq genes in near proximity.

- Gene Region: For those regions contained either in close proximity to or within a gene, the gene region further explains the relationship of this region to the nearby gene.

- Chrom: The chromosome on which the DMR is located using the hg18 build of the UCSC genome browser.

- Start: The starting position of the DMR as designated by the hg18 build of the UCSC genome browser.

- End: The ending position of the DMR as designated by the hg18 build of the UCSC genome browser.

- Microarray Analysis: Describes whether this region was also/initially determined to be differentially methylated by microarray analysis. The methylated fraction of ten paired placenta and PBMC samples was isolated using the MBD-Fc protein. The two tissue fractions were then

labeled with either Alexa Fluor 555–aha–dCTP (PBMC) or Alexa Fluor 647–aha–dCTP (placental) using the BioPrime Total Genomic Labeling System™ and hybridized to Agilent® CpG Island microarrays. Many regions examined in these studies were not contained on the initial microarray.

- EpiTYPER 8 Samples: Describes whether this region was analyzed and determined to be differentially methylated in eight paired samples of placenta and peripheral blood mononuclear cells (PBMC) using EpiTYPER technology. Regions that were chosen for examination were based on multiple criteria. First, regions were selected based on data from the Microarray Analysis. Secondly, a comprehensive examination of all CpG islands located on chromosome 21 was undertaken. Finally, selected regions on chromosome 21 which had lower CpG frequency than those located in CpG islands were examined.

- EpiTYPER 73 Samples: Describes whether this region was subsequently analyzed using EpiTYPER technology in a sample cohort consisting of 73 paired samples of placenta and PBMC. All regions selected for analysis in this second sample cohort were selected based on the results from the experimentation described in the EpiTYPER 8 column. More specifically, the regions in this additional cohort exhibited a methylation profile similar to that determined in the EpiTYPER 8 Samples analysis. For example, all of the regions listed in Tables 3B-3C exhibit different levels of DNA methylation in a significant portion of the examined CpG dinucleotides within the defined region. Differential DNA methylation of CpG sites was determined using a paired T Test with those sites considered differentially methylated if the p-value (when comparing placental tissue to PBMC) is $p<0.05$.

- Previously Validated EpiTYPER: Describes whether this region or a portion of this region was validated using EpiTYPER during previous experimentation. (See Examples 1 and 2).

- Relative Methylation Placenta to Maternal: Describes the direction of differential methylation. Regions labeled as “hypermethylation” are more methylated within the designated region in placenta samples relative to PBMC and “hypomethylation” are more methylated within the designated region in PBMC samples.

TABLE 3A

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
chr13 group00016	chr13	19773745	19774050	chr13:19773518-19774214	0.19	0.22	0.32	0.1	HYPERMETHYLATION
chr13 group00005	chr13	19290394	19290768	:-	-0.89	0.94	0.35	-0.59	HYPOMETHYLATION
CRYL1	chr13	19887090	19887336	chr13:19887007-19887836	-0.63	0.74	0.21	-0.53	HYPOMETHYLATION
IL17D	chr13	20193675	20193897	chr13:20193611-20194438	-1.01	0.53	0.13	-0.39	HYPOMETHYLATION
CENPJ	chr13	24404023	24404359	:-	0.57	0.17	0.49	0.32	HYPERMETHYLATION
ATP8A2	chr13	25484475	25484614	chr13:25484287-25484761	0.81	0.16	0.43	0.27	HYPERMETHYLATION
GSH1	chr13	27265542	27265834	chr13:27264549-27266505	0.57	0.13	0.19	0.05	HYPERMETHYLATION
PDX1	chr13	27393789	27393979	chr13:27392001-27394099	0.55	0.06	0.2	0.14	HYPERMETHYLATION
PDX1	chr13	27400459	27401165	chr13:27400362-27400744; chr13:27401057-27401374	0.73	0.12	0.26	0.14	HYPERMETHYLATION
MAB21L1	chr13	34947737	34948062	chr13:34947570-34948159	0.66	0.11	0.17	0.06	HYPERMETHYLATION
RB1	chr13	47790983	47791646	chr13:47790636-47791858	0.18	0.45	0.48	0.03	HYPERMETHYLATION
PCDH17	chr13	57104856	57106841	chr13:57104527-57106931	0.46	0.15	0.21	0.06	HYPERMETHYLATION
KLHL1	chr13	69579933	69580146	chr13:69579733-69580220	0.79	0.09	0.28	0.2	HYPERMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
POU4F1	chr13	78079515	78081073	chr13:78079328-78079615; chr13:78080860-78081881	0.66	0.12	0.23	0.11	HYPERMETHYLATION
GPC6	chr13	92677402	92678666	chr13:92677246-92678878	0.66	0.06	0.19	0.13	HYPERMETHYLATION
SOX21	chr13	94152286	94153047	chr13:94152190-94153185	0.94	0.16	0.4	0.25	HYPERMETHYLATION
ZIC2	chr13	99439660	99440858	chr13:99439335-99440189; chr13:99440775-99441095	0.89	0.13	0.35	0.22	HYPERMETHYLATION
IRS2	chr13	109232856	109235065	chr13:109232467-109238181	-0.17	0.73	0.38	-0.35	HYPOMETHYLATION
chr13 group00350	chr13	109716455	109716604	chr13:109716325-109716726	-0.37	0.77	0.41	-0.36	HYPOMETHYLATION
chr13 group00385	chr13	111595578	111595955	chr13:111595459-111596131	0.87	0.06	0.2	0.14	HYPERMETHYLATION
chr13 group00390	chr13	111756337	111756593	chr13:111755805-111756697	0.71	0.12	0.34	0.22	HYPERMETHYLATION
chr13 group00391	chr13	111759856	111760045	chr13:111757885-111760666	0.86	0.11	0.36	0.25	HYPERMETHYLATION
chr13 group00395	chr13	111808255	111808962	chr13:111806599-111808492; chr13:111808866-111809114	0.96	0.13	0.35	0.22	HYPERMETHYLATION
chr13 group00399	chr13	112033503	112033685	chr13:112032967-112033734	0.38	0.26	0.43	0.18	HYPERMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
MCF2L	chr13	112724910	112725742	chr13:112724782-112725121; chr13:112725628-112725837	-0.47	0.91	0.33	-0.58	HYPOMETHYLATION
F7	chr13	112799123	112799379	chr13:112798487-112799566	-0.05	0.97	0.55	-0.41	HYPOMETHYLATION
PROZ	chr13	112855566	112855745	chr13:112855289-112855866	0.29	0.15	0.3	0.16	HYPERMETHYLATION
chr18 group00039	chr18	6919797	6919981	chr18:6919450-6920088	-0.38	0.88	0.39	-0.49	HYPOMETHYLATION
CIDEA	chr18	12244327	12244696	chr18:12244147-12245089	0.23	0.14	0.23	0.1	HYPERMETHYLATION
chr18 group00091	chr18	12901467	12901643	chr18:12901024-12902704	0.16	0.15	0.43	0.29	HYPERMETHYLATION
chr18 group00094	chr18	13126819	13126986	chr18:13126596-13127564	0.41	0.07	0.34	0.27	HYPERMETHYLATION
C18orf1	chr18	13377536	13377654	chr18:13377385-13377686	-0.12	0.95	0.69	-0.26	HYPOMETHYLATION
KLHL14	chr18	28603978	28605183	chr18:28603688-28606300	0.83	0.07	0.19	0.12	HYPERMETHYLATION
CD33L3	chr18	41671477	41673011	chr18:41671386-41673101	-0.34	0.49	0.44	-0.05	HYPOMETHYLATION
ST8SIA3	chr18	53171265	53171309	chr18:53170705-53172603	1.02	0.09	0.25	0.16	HYPERMETHYLATION
ONECUT2	chr18	53254808	53259810	chr18:53254152-53259851	0.74	0.09	0.23	0.14	HYPERMETHYLATION
RAX	chr18	55086286	55086436	chr18:55085813-55087807	0.88	0.11	0.26	0.16	HYPERMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
chr18 group00277	chr18	57151972	57152311	chr18:57151663-57152672	0.58	0.08	0.21	0.13	HYPERMETHYLATION
TNFRSF11A	chr18	58203013	58203282	chr18:58202849-58203367	-0.33	0.88	0.28	-0.6	HYPOMETHYLATION
NETO1	chr18	68685099	68687060	chr18:68684945-68687851	0.65	0.09	0.22	0.13	HYPERMETHYLATION
chr18 group00304	chr18	70133945	70134397	chr18:70133732-70134724	0.12	0.93	0.92	-0.01	NOT CONFIRMED
TSHZ1	chr18	71128742	71128974	chr18:71128638-71129076	0.23	0.95	0.92	-0.03	NOT CONFIRMED
ZNF236	chr18	72664454	72664736	chr18:72662797-72664893	-0.62	0.17	0.1	-0.07	HYPOMETHYLATION
MBP	chr18	72953150	72953464	chr18:72953137-72953402	0.6	0.44	0.72	0.28	HYPERMETHYLATION
chr18 group00342	chr18	74170347	74170489	chr18:74170210-74170687	-0.2	0.78	0.48	-0.3	HYPOMETHYLATION
NFATC1	chr18	75385424	75386008	chr18:75385279-75386532	0.23	0.14	0.84	0.7	HYPERMETHYLATION
CTDP1	chr18	75596358	75596579	chr18:75596009-75596899	0.07	0.97	0.96	-0.01	NOT CONFIRMED
chr18 group00430	chr18	75653272	75653621	:-	0.52	0.24	0.62	0.39	HYPERMETHYLATION
KCNG2	chr18	75760343	75760820	chr18:75759900-75760988	0.01	0.84	0.75	-0.09	NOT CONFIRMED
OLIG2	chr21	33317673	33321183	chr21:33316998-33322115	0.66	0.11	0.2	0.09	HYPERMETHYLATION
OLIG2	chr21	33327593	33328334	chr21:33327447-33328408	-0.75	0.77	0.28	-0.49	HYPOMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
RUNX1	chr21	35180938	35185436	chr21:35180822-35181342; chr21:35182320-35185557	-0.68	0.14	0.07	-0.07	HYPOMETHYLATION
SIM2	chr21	36994965	36995298	chr21:36990063-36995761	0.83	0.08	0.26	0.18	HYPERMETHYLATION
SIM2	chr21	36999025	36999410	chr21:36998632-36999555	0.87	0.06	0.24	0.18	HYPERMETHYLATION
DSCR6	chr21	37300407	37300512	chr21:37299807-37301307	0.22	0.04	0.14	0.11	HYPERMETHYLATION
DSCAM	chr21	41135559	41135706	chr21:41135380-41135816	1.03	0.06	0.29	0.23	HYPERMETHYLATION
chr21 group00165	chr21	43643421	43643786	chr21:43643322-43643874	1.14	0.16	0.81	0.65	HYPERMETHYLATION
AIRE	chr21	44529935	44530388	chr21:44529856-44530472	-0.55	0.62	0.27	-0.35	HYPOMETHYLATION
SUMO3	chr21	45061293	45061853	chr21:45061154-45063386	-0.41	0.55	0.46	-0.09	HYPOMETHYLATION
C21orf70	chr21	45202815	45202972	chr21:45202706-45203073	-0.46	0.96	0.51	-0.46	HYPOMETHYLATION
C21orf123	chr21	45671984	45672098	chr21:45671933-45672201	-0.63	0.92	0.43	-0.49	HYPOMETHYLATION
COL18A1	chr21	45754383	45754487	chr21:45753653-45754639	-0.18	0.97	0.72	-0.25	HYPOMETHYLATION
PRMT2	chr21	46911967	46912385	chr21:46911628-46912534	1.08	0.04	0.25	0.21	HYPERMETHYLATION
SIX2	chr2	45081223	45082129	chr2:45081148-45082287	1.15	0.08	0.36	0.28	HYPERMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
SIX2	chr2	45084851	45085711	chr2:45084715-45084986; chr2:45085285-45086054	1.21	0.07	0.35	0.28	HYPERMETHYLATION
SOX14	chr3	138971870	138972322	chr3:138971738-138972096; chr3:138972281-138973691	1.35	0.08	0.33	0.25	HYPERMETHYLATION
TLX3	chr5	170674439	170676431	chr5:170674208-170675356; chr5:170675783-170676712	0.91	0.11	0.35	0.24	HYPERMETHYLATION
FOXP4	chr6	41623666	41624114	chr6:41621630-41624167	1.1	0.07	0.27	0.2	HYPERMETHYLATION
FOXP4	chr6	41636384	41636779	chr6:41636244-41636878	1.32	0.04	0.33	0.29	HYPERMETHYLATION
chr7 group00267	chr7	12576755	12577246	chr7:12576690-12577359	0.94	0.08	0.26	0.17	HYPERMETHYLATION
NPY	chr7	24290224	24291508	chr7:24290083-24291605	0.93	0.09	0.3	0.21	HYPERMETHYLATION
SHH	chr7	155291537	155292091	chr7:155288453-155292175	0.98	0.19	0.52	0.33	HYPERMETHYLATION
OSR2	chr8	100029764	100030536	chr8:100029673-100030614	1.21	0.08	0.43	0.35	HYPERMETHYLATION
GLIS3	chr9	4288283	4289645	chr9:4287817-4290182	1.24	0.06	0.24	0.18	HYPERMETHYLATION
PRMT8	chr12	3472714	3473190	chr12:3470227-3473269	0.86	0.07	0.23	0.16	HYPERMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
TBX3	chr12	113609153	113609453	chr12:113609112-113609535	1.45	0.09	0.56	0.48	HYPERMETHYLATION
chr12 group00801	chr12	118516189	118517435	chr12:118515877-118517595	1.1	0.06	0.25	0.19	HYPERMETHYLATION
PAX9	chr14	36201402	36202386	chr14:36200932-36202536	0.89	0.11	0.32	0.21	HYPERMETHYLATION
SIX1	chr14	60178801	60179346	chr14:60178707-60179539	0.95	0.1	0.33	0.22	HYPERMETHYLATION
ISL2	chr15	74420013	74421546	chr15:74419317-74422570	1.08	0.08	0.27	0.19	HYPERMETHYLATION
DLX4	chr17	45397228	45397930	chr17:45396281-45398063	1.25	0.1	0.32	0.22	HYPERMETHYLATION
CBX4	chr17	75428613	75431793	chr17:75427586-75433676	1	0.07	0.27	0.21	HYPERMETHYLATION
EDG6	chr19	3129836	3130874	chr19:3129741-3130986	1.35	0.04	0.87	0.83	HYPERMETHYLATION
PRRT3	chr3	9963364	9964023	chr3:9962895-9964619	-0.85	0.9	0.09	-0.81	HYPOMETHYLATION
MGC29506	chr5	138757911	138758724	chr5:138755609-138758810	-0.63	0.93	0.17	-0.76	HYPOMETHYLATION
TEAD3	chr6	35561812	35562252	chr6:35561754-35562413	-1.17	0.92	0.13	-0.8	HYPOMETHYLATION
chr12 group00022	chr12	1642456	1642708	chr12:1642195-1642774	-1.33	0.66	0.09	-0.57	HYPOMETHYLATION
CENTG1	chr12	56406249	56407788	chr12:56406176-56407818	-1.07	0.95	0.19	-0.77	HYPOMETHYLATION

GENE NAME	CHROM	START	END	CpG ISLAND	MEAN LOG RATIO MICRO-ARRAY	MEAN MATERNAL METHY-LATION EPITYPER	MEAN PLACENTA METHY-LATION EPITYPER	METHY-LATION DIFFERENCE PLACENTA-MATERNAL	RELATIVE METHYLATION PLACENTA TO MATERNAL
CENTG1	chr12	56416146	56418794	chr12:56416095-56416628; chr12:56418745-56419001	-0.94	0.85	0.16	-0.69	HYPOMETHYLATION

Information in Table 5A based on the March 2006 human reference sequence (NCBI Build 36.1), which was produced by the International Human Genome Sequencing Consortium.

Table 3B: Non-Chromosome 21 differentially methylated regions

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	EpiTYPER 8 Samples	EpiTYPER 73 Samples	Previously Validated EpiTYPER	Relative Methylation Placenta to Maternal
TFAP2E	Intron	chr1	35815000	35816200	YES	YES	NO	NO	Hypermethylation
LRRC8D	Intron/Exon	chr1	90081350	90082250	YES	YES	NO	NO	Hypermethylation
TBX15	Promoter	chr1	119333500	119333700	YES	YES	NO	NO	Hypermethylation
C1orf51	Upstream	chr1	148520900	148521300	YES	YES	NO	NO	Hypermethylation
chr1:179553900-179554600	Intergenic	chr1	179553900	179554600	YES	YES	NO	NO	Hypermethylation
ZFP36L2	Exon	chr2	43304900	43305100	YES	YES	NO	NO	Hypermethylation
SIX2	Downstream	chr2	45081000	45086000	YES	YES	NO	YES	Hypermethylation
chr2:137238500-137240000	Intergenic	chr2	137238500	137240000	YES	YES	NO	NO	Hypermethylation
MAP1D	Intron/Exon	chr2	172652800	172653600	YES	YES	NO	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	EpiTYPER 8 Samples	EpiTYPER 73 Samples	Previously Validated EpiTYPER	Relative Methylation Placenta to Maternal
WNT6	Intron	chr2	219444250	219444290	YES	YES	NO	NO	Hypermethylation
INPP5D	Promoter	chr2	233633200	233633700	YES	YES	YES	NO	Hypermethylation
chr2:241211100-241211600	Intergenic	chr2	241211100	241211600	YES	YES	YES	NO	Hypermethylation
WNT5A	Intron	chr3	55492550	55492850	YES	YES	NO	NO	Hypermethylation
chr3:138971600-138972200	Intergenic	chr3	138971600	138972200	YES	YES	YES	YES	Hypermethylation
ZIC4	Intron	chr3	148598200	148599000	YES	YES	NO	NO	Hypermethylation
FGF12	Intron/Exon	chr3	193608500	193610500	YES	YES	NO	NO	Hypermethylation
GP5	Exon	chr3	195598400	195599200	YES	YES	NO	NO	Hypermethylation
MSX1	Upstream	chr4	4910550	4911100	YES	YES	NO	NO	Hypermethylation
NKX3-2	Intron/Exon	chr4	13152500	13154500	YES	YES	NO	NO	Hypermethylation
chr4:111752000-111753000	Intergenic	chr4	111752000	111753000	YES	YES	YES	NO	Hypermethylation
SFRP2	Promoter	chr4	154928800	154930100	YES	YES	NO	NO	Hypermethylation
chr4:174664300-174664800	Intergenic	chr4	174664300	174664800	YES	YES	NO	NO	Hypermethylation
chr4:174676300-174676800	Intergenic	chr4	174676300	174676800	YES	YES	NO	NO	Hypermethylation
SORBS2	Intron	chr4	186796900	186797500	YES	YES	NO	NO	Hypermethylation
chr5:42986900-42988200	Intergenic	chr5	42986900	42988200	YES	YES	NO	NO	Hypermethylation
chr5:72712000-72714100	Intergenic	chr5	72712000	72714100	YES	YES	NO	NO	Hypermethylation
chr5:72767550-72767800	Intergenic	chr5	72767550	72767800	YES	YES	NO	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	EpiTYPER 8 Samples	EpiTYPER 73 Samples	Previously Validated EpiTYPER	Relative Methylation Placenta to Maternal
NR2F1	Intron/Exon	chr5	92955000	92955250	YES	YES	NO	NO	Hypermethylation
PCDHGA1	Intron	chr5	140850500	140852500	YES	YES	YES	NO	Hypermethylation
chr6:10489100-10490200	Intergenic	chr6	10489100	10490200	YES	YES	YES	NO	Hypermethylation
FOXP4	Intron	chr6	41636200	41637000	YES	YES	NO	YES	Hypermethylation
chr7:19118400-19118700	Intergenic	chr7	19118400	19118700	YES	YES	NO	NO	Hypermethylation
chr7:27258000-27258400	Intergenic	chr7	27258000	27258400	YES	YES	NO	NO	Hypermethylation
TBX20	Upstream	chr7	35267500	35268300	YES	YES	NO	NO	Hypermethylation
AGBL3	Promoter	chr7	134321300	134322300	YES	YES	NO	NO	Hypermethylation
XPO7	Downstream	chr8	21924000	21924300	YES	YES	NO	NO	Hypermethylation
chr8:41543400-41544000	Intergenic	chr8	41543400	41544000	YES	YES	NO	NO	Hypermethylation
GDF6	Exon	chr8	97225400	97227100	YES	YES	NO	NO	Hypermethylation
OSR2	Intron/Exon	chr8	100029000	100031000	YES	YES	YES	YES	Hypermethylation
GLIS3	Intron/Exon	chr9	4288000	4290000	YES	YES	NO	YES	Hypermethylation
NOTCH1	Intron	chr9	138547600	138548400	YES	YES	YES	NO	Hypermethylation
EGFL7	Upstream	chr9	138672350	138672850	YES	YES	NO	NO	Hypermethylation
CELF2	Intron/Exon	chr10	11246700	11247900	YES	YES	NO	NO	Hypermethylation
HHEX	Intron	chr10	94441000	94441800	YES	YES	NO	NO	Hypermethylation
DOCK1/FAM196A	Intron/Exon	chr10	128883000	128883500	YES	YES	NO	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	EpiTYPER 8 Samples	EpiTYPER 73 Samples	Previously Validated EpiTYPER	Relative Methylation Placenta to Maternal
PAX6	Intron	chr11	31782400	31783500	YES	YES	NO	NO	Hypermethylation
FERMT3	Intron/Exon	chr11	63731200	63731700	YES	YES	YES	NO	Hypermethylation
PKNOX2	Intron	chr11	124541200	124541800	YES	YES	NO	NO	Hypermethylation
KIRREL3	Intron	chr11	126375150	126375300	YES	YES	NO	NO	Hypermethylation
BCAT1	Intron	chr12	24946700	24947600	YES	YES	NO	NO	Hypermethylation
HOXC13	Intron/Exon	chr12	52625000	52625600	YES	YES	NO	NO	Hypermethylation
TBX5	Promoter	chr12	113330500	113332000	YES	YES	NO	NO	Hypermethylation
TBX3	Upstream	chr12	113609000	113609500	YES	YES	NO	YES	Hypermethylation
chr12:113622100-113623000	Intergenic	chr12	113622100	113623000	YES	YES	YES	NO	Hypermethylation
chr12:113657800-113658300	Intergenic	chr12	113657800	113658300	YES	YES	NO	NO	Hypermethylation
THEM233	Promoter	chr12	118515500	118517500	YES	YES	NO	YES	Hypermethylation
NCOR2	Intron/Exon	chr12	123516200	123516800	YES	YES	YES	NO	Hypermethylation
THEM132C	Intron	chr12	127416300	127416700	YES	YES	NO	NO	Hypermethylation
PTGDR	Promoter	chr14	51804000	51805200	YES	YES	NO	NO	Hypermethylation
ISL2	Intron/Exon	chr15	74420000	74422000	YES	YES	NO	YES	Hypermethylation
chr15:87750000-87751000	Intergenic	chr15	87750000	87751000	YES	YES	NO	NO	Hypermethylation
chr15:87753000-87754100	Intergenic	chr15	87753000	87754100	YES	YES	NO	NO	Hypermethylation
NR2F2	Upstream	chr15	94666000	94667500	YES	YES	YES	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	EpiTYPER 8 Samples	EpiTYPER 73 Samples	Previously Validated EpiTYPER	Relative Methylation Placenta to Maternal
chr16:11234300-11234900	Intergenic	chr16	11234300	11234900	YES	YES	NO	NO	Hypermethylation
SPN	Exon	chr16	29582800	29583500	YES	YES	YES	NO	Hypermethylation
chr16:85469900-85470200	Intergenic	chr16	85469900	85470200	YES	YES	NO	NO	Hypermethylation
SLFN11	Promoter	chr17	30725100	30725600	YES	YES	NO	NO	Hypermethylation
DLX4	Upstream	chr17	45396800	45397800	YES	YES	NO	YES	Hypermethylation
SLC38A10 (MGC15523)	Intron	chr17	76873800	76874300	YES	YES	YES	NO	Hypermethylation
S1PR4	Exon	chr19	3129900	3131100	YES	YES	YES	YES	Hypermethylation
MAP2K2	Intron	chr19	4059700	4060300	YES	YES	YES	NO	Hypermethylation
UHRF1	Intron	chr19	4867300	4867800	YES	YES	YES	NO	Hypermethylation
DEDD2	Exon	chr19	47395300	47395900	YES	YES	YES	NO	Hypermethylation
CDC42EP1	Exon	chr22	36292300	36292800	YES	YES	YES	NO	Hypermethylation

Table 3C: Chromosome 21 differentially methylated regions

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
chr21:9906600-9906800	Intergenic	chr21	9906600	9906800	NO	YES	NO	NO	Hypomethylation
chr21:9907000-9907400	Intergenic	chr21	9907000	9907400	NO	YES	NO	NO	Hypomethylation
chr21:9917800-9918450	Intergenic	chr21	9917800	9918450	NO	YES	NO	NO	Hypomethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
TPTE	Promoter	chr21	10010000	10015000	NO	YES	NO	NO	Hypomethylation
chr21:13974500-13976000	Intergenic	chr21	13974500	13976000	NO	YES	NO	NO	Hypomethylation
chr21:13989500-13992000	Intergenic	chr21	13989500	13992000	NO	YES	NO	NO	Hypomethylation
chr21:13998500-14000100	Intergenic	chr21	13998500	14000100	NO	YES	NO	NO	Hypomethylation
chr21:14017000-14018500	Intergenic	chr21	14017000	14018500	NO	YES	NO	NO	Hypomethylation
chr21:14056400-14058100	Intergenic	chr21	14056400	14058100	NO	YES	NO	NO	Hypomethylation
chr21:14070250-14070550	Intergenic	chr21	14070250	14070550	NO	YES	NO	NO	Hypomethylation
chr21:14119800-14120400	Intergenic	chr21	14119800	14120400	NO	YES	NO	NO	Hypomethylation
chr21:14304800-14306100	Intergenic	chr21	14304800	14306100	NO	YES	NO	NO	Hypomethylation
chr21:15649340-15649450	Intergenic	chr21	15649340	15649450	NO	YES	YES	NO	Hypermethylation
C21orf34	Intron	chr21	16881500	16883000	NO	YES	NO	NO	Hypomethylation
BTG3	Intron	chr21	17905300	17905500	NO	YES	NO	NO	Hypomethylation
CHODL	Promoter	chr21	18539000	18539800	NO	YES	YES	NO	Hypermethylation
NCAM2	Upstream	chr21	21291500	21292100	NO	YES	NO	NO	Hypermethylation
chr21:23574000-23574600	Intergenic	chr21	23574000	23574600	NO	YES	NO	NO	Hypomethylation
chr21:24366920-24367060	Intergenic	chr21	24366920	24367060	NO	YES	NO	NO	Hypomethylation
chr21:25656000-25656900	Intergenic	chr21	25656000	25656900	NO	YES	NO	NO	Hypomethylation
MIR155HG	Promoter	chr21	25855800	25857200	NO	YES	YES	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
CYYR1	Intron	chr21	26830750	26830950	NO	YES	NO	NO	Hypomethylation
chr21:26938800-26939200	Intergenic	chr21	26938800	26939200	NO	YES	NO	NO	Hypomethylation
GRIK1	Intron	chr21	30176500	30176750	NO	YES	NO	NO	Hypomethylation
chr21:30741350-30741600	Intergenic	chr21	30741350	30741600	NO	YES	NO	NO	Hypermethylation
TIAM1	Intron	chr21	31426800	31427300	NO	YES	YES	NO	Hypermethylation
TIAM1	Intron	chr21	31475300	31475450	NO	YES	NO	NO	Hypermethylation
TIAM1	Intron	chr21	31621050	31621350	NO	YES	YES	NO	Hypermethylation
SOD1	Intron	chr21	31955000	31955300	NO	YES	NO	NO	Hypomethylation
HUNK	Intron/Exon	chr21	32268700	32269100	NO	YES	YES	NO	Hypermethylation
chr21:33272200-33273300	Intergenic	chr21	33272200	33273300	NO	YES	NO	NO	Hypomethylation
OLIG2	Promoter	chr21	33314000	33324000	YES	YES	NO	YES	Hypermethylation
OLIG2	Downstream	chr21	33328000	33328500	YES	YES	NO	NO	Hypomethylation
RUNX1	Intron	chr21	35185000	35186000	NO	YES	NO	NO	Hypomethylation
RUNX1	Intron	chr21	35320300	35320400	NO	YES	NO	NO	Hypermethylation
RUNX1	Intron	chr21	35321200	35321600	NO	YES	NO	NO	Hypermethylation
RUNX1	Intron/Exon	chr21	35340000	35345000	NO	YES	YES	NO	Hypermethylation
chr21:35499200-35499700	Intergenic	chr21	35499200	35499700	NO	YES	YES	NO	Hypermethylation
chr21:35822800-35823500	Intergenic	chr21	35822800	35823500	NO	YES	YES	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
CBR1	Promoter	chr21	36364000	36364500	NO	YES	NO	NO	Hypermethylation
DOPEY2	Downstream	chr21	36589000	36590500	NO	YES	NO	NO	Hypomethylation
SIM2	Promoter	chr21	36988000	37005000	YES	YES	YES	YES	Hypermethylation
HLCS	Intron	chr21	37274000	37275500	YES	YES	YES	NO	Hypermethylation
DSCR6	Upstream	chr21	37300200	37300400	YES	YES	NO	YES	Hypermethylation
DSCR3	Intron	chr21	37551000	37553000	YES	YES	YES	NO	Hypermethylation
chr21:37841100-37841800	Intergenic	chr21	37841100	37841800	NO	YES	YES	NO	Hypermethylation
ERG	Intron	chr21	38791400	38792000	NO	YES	YES	NO	Hypermethylation
chr21:39278700-39279800	Intergenic	chr21	39278700	39279800	NO	YES	YES	NO	Hypermethylation
C21orf129	Exon	chr21	42006000	42006250	NO	YES	YES	NO	Hypermethylation
C2CD2	Intron	chr21	42188900	42189500	NO	YES	YES	NO	Hypermethylation
UMODL1	Upstream	chr21	42355500	42357500	NO	YES	YES	NO	Hypermethylation
UMODL1/C21orf128	Intron	chr21	42399200	42399900	NO	YES	NO	NO	Hypomethylation
ABCG1	Intron	chr21	42528400	42528600	YES	YES	NO	NO	Hypomethylation
chr21:42598300-42599600	Intergenic	chr21	42598300	42599600	YES	YES	NO	NO	Hypomethylation
chr21:42910000-42911000	Intergenic	chr21	42910000	42911000	NO	YES	NO	NO	Hypomethylation
PDE9A	Upstream	chr21	42945500	42946000	NO	YES	NO	NO	Hypomethylation
PDE9A	Intron	chr21	42961400	42962700	NO	YES	NO	NO	Hypomethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
PDE9A	Intron	chr21	42977400	42977600	NO	YES	NO	NO	Hypermethylation
PDE9A	Intron/Exon	chr21	42978200	42979800	YES	YES	NO	NO	Hypomethylation
PDE9A	Intron	chr21	43039800	43040200	NO	YES	YES	NO	Hypermethylation
chr21:43130800-43131500	Intergenic	chr21	43130800	43131500	NO	YES	NO	NO	Hypomethylation
U2AF1	Intron	chr21	43395500	43395800	NO	YES	NO	NO	Hypermethylation
U2AF1	Intron	chr21	43398000	43398450	NO	YES	YES	NO	Hypermethylation
chr21:43446600-43447600	Intergenic	chr21	43446600	43447600	NO	YES	NO	NO	Hypomethylation
CRYAA	Intron/Exon	chr21	43463000	43466100	NO	YES	NO	NO	Hypomethylation
chr21:43545000-43546000	Intergenic	chr21	43545000	43546000	YES	YES	NO	NO	Hypomethylation
chr21:43606000-43606500	Intergenic	chr21	43606000	43606500	NO	YES	NO	NO	Hypomethylation
chr21:43643000-43644300	Intergenic	chr21	43643000	43644300	YES	YES	YES	YES	Hypermethylation
C21orf125	Upstream	chr21	43689100	43689300	NO	YES	NO	NO	Hypermethylation
C21orf125	Downstream	chr21	43700700	43701700	NO	YES	NO	NO	Hypermethylation
HSF2BP	Intron/Exon	chr21	43902500	43903800	YES	YES	NO	NO	Hypomethylation
AGPAT3	Intron	chr21	44161100	44161400	NO	YES	YES	NO	Hypermethylation
chr21:44446500-44447500	Intergenic	chr21	44446500	44447500	NO	YES	NO	NO	Hypomethylation
TRPM2	Intron	chr21	44614500	44615000	NO	YES	NO	NO	Hypomethylation
C21orf29	Intron	chr21	44750400	44751000	NO	YES	NO	NO	Hypomethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
C21orf29	Intron	chr21	44950000	44955000	NO	YES	YES	NO	Hypermethylation
ITGB2	Intron/Exon	chr21	45145500	45146100	NO	YES	NO	NO	Hypomethylation
POFUT2	Downstream	chr21	45501000	45503000	NO	YES	NO	NO	Hypomethylation
chr21:45571500-45573700	Intergenic	chr21	45571500	45573700	NO	YES	NO	NO	Hypomethylation
chr21:45609000-45610600	Intergenic	chr21	45609000	45610600	NO	YES	NO	NO	Hypomethylation
COL18A1	Intron	chr21	45670000	45677000	YES	YES	NO	YES	Hypomethylation
COL18A1	Intron/Exon	chr21	45700500	45702000	NO	YES	NO	NO	Hypomethylation
COL18A1	Intron/Exon	chr21	45753000	45755000	YES	YES	NO	YES	Hypomethylation
chr21:45885000-45887000	Intergenic	chr21	45885000	45887000	NO	YES	NO	NO	Hypomethylation
PCBP3	Intron	chr21	46111000	46114000	NO	YES	NO	NO	Hypomethylation
PCBP3	Intron/Exon	chr21	46142000	46144500	NO	YES	NO	NO	Hypomethylation
COL6A1	Intron/Exon	chr21	46227000	46233000	NO	YES	NO	NO	Hypomethylation
COL6A1	Intron/Exon	chr21	46245000	46252000	NO	YES	NO	NO	Hypomethylation
chr21:46280500-46283000	Intergenic	chr21	46280500	46283000	NO	YES	NO	NO	Hypomethylation
COL6A2	Intron	chr21	46343500	46344200	NO	YES	NO	NO	Hypomethylation
COL6A2	Intron/Exon	chr21	46368000	46378000	NO	YES	NO	NO	Hypomethylation
C21orf56	Intron/Exon	chr21	46426700	46427500	NO	YES	NO	NO	Hypomethylation
C21orf57	Intron	chr21	46541568	46541861	NO	YES	NO	NO	Hypermethylation

Region Name	Gene Region	Chrom	Start	End	Microarray Analysis	Epi TYPER 8 Samples	Epi TYPER 73 Samples	Previously Validated Epi TYPER	Relative Methylation Placenta to Maternal
C21orf57	Exon	chr21	46541872	46542346	NO	YES	NO	NO	Hypermethylation
C21orf57	Downstream	chr21	46542319	46542665	NO	YES	NO	NO	Hypermethylation
C21orf58	Intron	chr21	46546914	46547404	NO	YES	NO	NO	Hypomethylation
PRMT2	Downstream	chr21	46911000	46913000	YES	YES	NO	YES	Hypermethylation
ITGB2	Intron	chr21	45170700	45171100	NO	YES	YES	NO	Hypermethylation

TABLE 4

GENE NAME	**CHROM**	**START**	**END**	**SNPs**
chr13 group00016	chr13	19773745	19774050	rs7996310; rs12870878
chr13 group00005	chr13	19290394	19290768	rs11304938
CENPJ	chr13	24404023	24404359	rs7326661
ATP8A2	chr13	25484475	25484614	rs61947088
PDX1	chr13	27400459	27401165	rs58173592; rs55836809; rs61944011
RB1	chr13	47790983	47791646	rs2804094; rs4151432; rs4151433; rs4151434; rs4151435
PCDH17	chr13	57104856	57106841	rs35287822; rs34642962; rs41292834; rs45500496; rs45571031; rs41292836; rs28374395; rs41292838
KLHL1	chr13	69579933	69580146	rs3751429
POU4F1	chr13	78079515	78081073	rs11620410; rs35794447; rs2765065
GPC6	chr13	92677402	92678666	rs35689696; rs11839555; rs55695812; rs35259892

GENE NAME	CHROM	START	END	SNPs
SOX21	chr13	94152286	94153047	rs41277652; rs41277654; rs35276096; rs5805873; rs35109406
ZIC2	chr13	99439660	99440858	rs9585309; rs35501321; rs9585310; rs7991728; rs1368511
IRS2	chr13	109232856	109235065	rs61747993; rs1805097; rs9583424; rs35927012; rs1056077; rs1056078; rs34889228; rs1056080; rs1056081; rs12853546; rs4773092; rs35223808; rs35894564; rs3742210; rs34412495; rs61962699; rs45545638; rs61743905
chr13 group00395	chr13	111808255	111808962	rs930346
MCF2L	chr13	112724910	112725742	rs35661110; rs2993304; rs1320519; rs7320418; rs58416100
F7	chr13	112799123	112799379	rs2480951; rs2476320
CIDEA	chr18	12244327	12244696	rs60132277
chr18 group00091	chr18	12901467	12901643	rs34568924; rs8094284; rs8094285
C18orf1	chr18	13377536	13377654	rs9957861
KLHL14	chr18	28603978	28605183	rs61737323; rs61737324; rs12960414
CD33L3	chr18	41671477	41673011	rs62095363; rs2919643
ONECUT2	chr18	53254808	53259810	rs35685953; rs61735644; rs8084084; rs35937482; rs35427632; rs7232930; rs3786486; rs34286480; rs3786485; rs28655657; rs4940717; rs4940719; rs3786484; rs34040569; rs35542747; rs33946478; rs35848049; rs7231349; rs7231354; rs34481218; rs12962172; rs3911641
RAX	chr18	55086286	55086436	rs58797899; rs45501496
chr18 group00277	chr18	57151972	57152311	rs17062547
TNFRSF11A	chr18	58203013	58203282	rs35114461
NETO1	chr18	68685099	68687060	rs4433898; rs34497518; rs35135773; rs6566677; rs57425572; rs36026929; rs34666288; rs10627137; rs35943684; rs9964226; rs4892054; rs9964397; rs4606820; rs12966677; rs8095606

GENE NAME	CHROM	START	END	SNPs
chr18 group00304	chr18	70133945	70134397	rs8086706; rs8086587; rs8090367; rs999332; rs17806420; rs58811193
TSHZ1	chr18	71128742	71128974	rs61732783; rs3744910; rs1802180
chr18 group00342	chr18	74170347	74170489	rs7226678
NFATC1	chr18	75385424	75386008	rs28446281; rs56384153; rs4531815; rs3894049
chr18 group00430	chr18	75653272	75653621	rs34967079; rs35465647
KCNG2	chr18	75760343	75760820	rs3744887; rs3744886
OLIG2	chr21	33317673	33321183	rs2236618; rs11908971; rs9975039; rs6517135; rs2009130; rs1005573; rs1122807; rs10653491; rs10653077; rs35086972; rs28588289; rs7509766; rs62216114; rs35561747; rs7509885; rs11547332
OLIG2	chr21	33327593	33328334	rs7276788; rs7275842; rs7275962; rs7276232; rs16990069; rs13051692; rs56231743; rs35931056
RUNX1	chr21	35180938	35185436	rs2843956; rs55941652; rs56020428; rs56251824; rs13051109; rs13051111; rs3833348; rs7510136; rs743289; rs5843690; rs33915227; rs11402829; rs2843723; rs8128138; rs8131386; rs2843957; rs57537540; rs13048584; rs7281361; rs2843965; rs2843958
SIM2	chr21	36994965	36995298	rs2252821
SIM2	chr21	36999025	36999410	rs58347144; rs737380
DSCAM	chr21	41135559	41135706	rs35298822
AIRE	chr21	44529935	44530388	rs35110251; rs751032; rs9978641
SUMO3	chr21	45061293	45061853	rs9979741; rs235337; rs7282882
C21orf70	chr21	45202815	45202972	rs61103857; rs9979028; rs881318; rs881317
COL18A1	chr21	45754383	45754487	rs35102708; rs9980939
PRMT2	chr21	46911967	46912385	rs35481242; rs61743122; rs8131044; rs2839379

GENE NAME	CHROM	START	END	SNPs
SIX2	chr2	45081223	45082129	rs62130902
SIX2	chr2	45084851	45085711	rs35417092; rs57340219
SOX14	chr3	138971870	138972322	rs57343003
TLX3	chr5	170674439	170676431	rs11134682; rs35704956; rs2964533; rs35601828
FOXP4	chr6	41623666	41624114	rs12203107; rs1325690
FOXP4	chr6	41636384	41636779	rs56835416
chr7 group00267	chr7	12576755	12577246	rs56752985; rs17149965; rs6948573; rs2240572
NPY	chr7	24290224	24291508	rs2390965; rs2390966; rs2390967; rs2390968; rs3025123; rs16146; rs16145; rs16144; rs13235842; rs13235935; rs13235938; rs13235940; rs13235944; rs36083509; rs3025122; rs16143; rs16478; rs16142; rs16141; rs16140; rs16139; rs2229966; rs1042552; rs5571; rs5572
SHH	chr7	155291537	155292091	rs9333622; rs1233554; rs9333620; rs1233555
GLIS3	chr9	4288283	4289645	rs56728573; rs12340657; rs12350099; rs35338539; rs10974444; rs7852293
PRMT8	chr12	3472714	3473190	rs12172776
TBX3	chr12	113609153	113609453	rs60114979
chr12 group00801	chr12	118516189	118517435	rs966246; rs17407022; rs970095; rs2711748
PAX9	chr14	36201402	36202386	rs17104893; rs12883298; rs17104895; rs35510737; rs12882923; rs12883049; rs28933970; rs28933972; rs28933971; rs28933373; rs61734510
SIX1	chr14	60178801	60179346	rs761555
ISL2	chr15	74420013	74421546	rs34173230; rs11854453
DLX4	chr17	45397228	45397930	rs62059964; rs57481357; rs56888011; rs17638215; rs59056690; rs34601685; rs17551082
CBX4	chr17	75428613	75431793	rs1285243; rs35035500; rs12949177; rs3764374; rs62075212; rs62075213; rs3764373; rs3764372; rs55973291

GENE NAME	CHROM	START	END	SNPs
EDG6	chr19	3129836	3130874	rs34728133; rs34573539; rs3826936; rs34914134; rs61731111; rs34205484
MGC29506	chr5	138757911	138758724	rs11748963; rs7447765; rs35262202
CENTG1	chr12	56406249	56407788	rs61935742; rs12318065; rs238519; rs238520; rs238521; rs808930; rs2640595; rs2640596; rs2640597; rs2640598; rs34772922
CENTG1	chr12	56416146	56418794	rs11830475; rs34482618; rs2650057; rs2518686; rs12829991

400

TABLE 5

GENE NAME	RELATIVE METHYLATION PLACENTA TO MATERNAL	PRC2 TARGET
CRYL1	HYPOMETHYLATION	TRUE
IL17D	HYPOMETHYLATION	TRUE
GSH1	HYPERMETHYLATION	TRUE
MAB21L1	HYPERMETHYLATION	TRUE
PCDH17	HYPERMETHYLATION	TRUE
KLHL1	HYPERMETHYLATION	TRUE
POU4F1	HYPERMETHYLATION	TRUE
SOX21	HYPERMETHYLATION	TRUE
ZIC2	HYPERMETHYLATION	TRUE
CIDEA	HYPERMETHYLATION	TRUE
KLHL14	HYPERMETHYLATION	TRUE
ONECUT2	HYPERMETHYLATION	TRUE
RAX	HYPERMETHYLATION	TRUE
TNFRSF11A	HYPOMETHYLATION	TRUE
OLIG2	HYPERMETHYLATION	TRUE
OLIG2	HYPOMETHYLATION	TRUE
SIM2	HYPERMETHYLATION	TRUE
SIM2	HYPERMETHYLATION	TRUE
SIX2	HYPERMETHYLATION	TRUE
SIX2	HYPERMETHYLATION	TRUE
SOX14	HYPERMETHYLATION	TRUE
TLX3	HYPERMETHYLATION	TRUE

GENE NAME	RELATIVE METHYLATION PLACENTA TO MATERNAL	PRC2 TARGET
SHH	HYPERMETHYLATION	TRUE
OSR2	HYPERMETHYLATION	TRUE
TBX3	HYPERMETHYLATION	TRUE
PAX9	HYPERMETHYLATION	TRUE
SIX1	HYPERMETHYLATION	TRUE
ISL2	HYPERMETHYLATION	TRUE
DLX4	HYPERMETHYLATION	TRUE
CBX4	HYPERMETHYLATION	TRUE
CENTG1	HYPOMETHYLATION	TRUE
CENTG1	HYPOMETHYLATION	TRUE

TABLE 6A

SEQ ID NO	GENE NAME	SEQUENCE
1	chr13 group-00016	CAGCAGGCGCGCTCCCGGCGAATCTGCCTGAATCGCCGTGAATGCGGTGGGGTGCAGGGCAGGGGCTGGTTTTCTCAGCCGGTCTTGG CTTTTCTCTTTCTCTCCTGCTCCACCAGCAGCCCCTCCGCGGGTCCCATGGGCTCCGCGCTCAGAACAGCCCGGAACCAGGCGCCGCTC GCCGCTCGCTGGGGGCCACCCGCCTCTCCCCGGAACAGCCTCCCGCGGGCCTCTTGGCCTCGCACTGGCGCCCTCACCCACACATCGT CCCTTTATCCGCTCAGACGCTGCAAAGGGCCTTCTGTCTC
2	CENPJ	GCTTTGGATTTATCCTCATTGGCTAAATCCCTCCTGAAACATGAAACTGAAACAAAGCCCTGAACCCCCTCAGGCTGAAAAGACAAACCCC GCCTGAGGCCGGGTCCCGCTCCCCACCTGGAGGGACCCAATTCTGGGCGCCTTCTGGCGACGGTCCCTGCTAGGGACGCTGCGCTCTC CGAGTGCGAGTTTTCGCCAAACTGATAAAGCACGCAGAACCGCAATCCCCAAACTAACACTGAACCCGGACCCGCGATCCCCAAACTGAC AAGGGACCCGGAACAGCGACCCCCAAACCGACACGGGACTCGGGAACCGCTATCTCCAAAGGGCAGC
3	ATP8A 2	TTTCCACAACAGGGAGCCAGCATTGAGGCGCCCAGATGGCATCTGCTGGAAATCACGGGCCGCTGGTGAAGCACCACGCCTTACCCGAC GTGGGGAGGTGATCCCCCCACCTCATCCCACCCCCTTCTGTCTGTCTCCTT
4	GSH1	GCTGGACAAGGAGCGCTCACTGTAGCTCTGCTGTGGATTGTGTTGGGGCGAAGAGATGGGTAAGAGGTCAAAGTCGTAGGATTCTGGCG ACCGCCTACCAAGGGATTGGGTCCACAGCACAGAGGTCTGATCGCTTCCTTCTCTGCTCTGCCACCTCCAGACAGCAGCTCTAACCAGCT GCCCAGCAGCAAGAGGATGCGCACGGCTTTCACCAGCACGCAGCTGCTAGAGCTGGAGCGCGAGTTCGCTTCTAATATGTACCTGTCCC GCCTACGTCGCATCGAGATCGCGA
5	PDX1	TGCCTGACACTGACCCCAGGCGCAGCCAGGAGGGGCTTTGTGCGGGAGAGGGAGGGGACCCCAGCTTGCCTGGGGTCCACGGGACT CTCTTCTTCCTAGTTCACTTTCTTGCTAAGGCGAAGGTCCTGAGGCAGGACGAGGGCTGAACTGCGCTGCAATCGTCCCCACCTCCAGCG AAACCCAGTTGAC
6	PDX1	TCGGCGGAGAGACCTCGAGGAGAGTATGGGGAAAGGAATGAATGCTGCGGAGCGCCCCTCTGGGCTCCACCCAAGCCTCGGAGGCGG GACGGTGGGCTCCGTCCCGACCCCTTAGGCAGCTGGACCGATACCTCCTGGATCAGACCCCACAGGAAGACTCGCGTGGGGCCCGATA TGTGTACTTCAAACTCTGAGCGGCCACCCTCAGCCAACTGGCCAGTGGATGCGAATCCGTGGGCCCTGAGGGGCGAGGGCGCTCGGAAC TGCATGCCTGTGCACGGTGCCGGGCTCTCCCAGAGTGAGGGGGGCCGTAAGGAGATCTCCAAGGAAGCCGAAAAAAGCAGCCAGTTGGGC TTCGGGAAAGACTTTTCTGCAAAGGAAGTGATCTGGTCCCAGAACTCCAGGGTTGACCCCAGTACCTGACTTCTCCGGGAGCTGTCAGCT CTCCTCTGTTCTTCGGGCTTGGCGCGCTCCTTTCATAATGGACAGACACCAGTGGCCTTCAAAAGGTCTGGGGTGGGGGAACGGAGGAA GTGGCCTTGGGTGCAGAGGAAGAGCAGAGCTCCTGCCAAAGCTGAACGCAGTTAGCCCTACCCAAGTGCGCGCTGGCTCGGCATATGC GCTCCAGAGCCGGCAGGACAGCCCGGCCCTGCTCACCCCGAGGAGAAATCCAACAGCGCAGCCTCCTGCACCTCCTTGCCCCCAGAGAC
7	MAB21 L1	AGATCCCGGTGCATTTAAAGGCCGGCGTGATCTGCACCACGTACCTATCTCGGATTCTCAGTTTCACTTCGCTGGTGTCTGCCACCATCTT TACCACATCCCGGTAGCTACATTTGTCTACCGCTTGAGCCACCAGCGTCTGAAACCTGGACCGGATTTTGCGCGCCGAGAGGTAGCCGG AGGCGGTAATGAATTCCACCCAGAGGGACATGCTCCCTCTTGCGCCCCGTCGCTCAACTTCAGCACCGCGCAGCCGGGGCAGTGAGCCATCG TCCACGAAGTTGAACACCCCCCATTTGGTTGAGATAAAGCACCACTTCAAATTCGGT
8	RB1	ACTATGCCTTGAGGGTCAAAACGTCTGGATTTCCTGATCGATGCTGTCGTCGCTGTCCACGGAGCTACTGTCGCCGTCAGAGCGGGAAG GCACGTTCAGGGAGTAGAAGCGTGGGCTTGCAGAAAGGGACCTGTTGCTGCCTTACATGGGGGCCGGCAGGGTAGTCTTGGAAATGCC CAAGATTGCTTCCGCGCGCGTCAGTTCAGCGGACGTGTCTGCCTGGCACGAGGACCGTTCTACAAACTCGTTCCTGGAAGCCGGGCTCG CTGGAGGCGGAGCTTTGGTTTCCTTCGGGAGCTTGTGGGGAATGGTCAGCGTCTAGGCACCCCCGGGCAAGGGTCTGTGGCCTTGGTGG

SEQ ID NO	GENE NAME	SEQUENCE
		CCACTGGCTTCCTCTAGCTGGGTGTTTTCCTGTGGGTCTCGCGCAAGGCACTTTTTTTGTGGCGCTGCTTGTGCTGTGTGCTGCGGGGTCAGGC GTCCTCTCTCCTCCCCGGCGCTGGGCCCTCTGGGGCAGGTCCCCCGTTGGCCTCCTTGCCGTGTTTGCCGCAGCTAGTACACCTGGATGGCC TCCTCAGTGCCGTCGTTGCTGCTGGAGTCTGACGCCTCGGGCGCCTGCCGCCGCACTTGTGACTTGCTTTCCCCTTCTCAGGGCGCCAGC GCTCCTCTTGACCCCCGCTTTTATTCTGTGGTGCTTCTGAAG
9	PCDH17	GCAAGTCGGGTAGCTACCGGGTGCTGGAGAACTCCGCACCGCACCTGCTGGACGTGGACGCAGACAGCGGGCTCCTCTACACCAAGCA GCGCATCGACCGCGAGTCCCTGTGCCGCCACAATGCCAAGTGCCAGCTGTCCCTCGAGGTGTTCGCCAACGACACAAGGAGATCTGCATGA TCAAGGTAGAGATCCAGGACATCAACGACAACGCGCCCTCCTTCTCCTCGGACCAGATCGAAATGGACATCTCGGAGAACGCTGCTCCG GGCACCCGCTTCCCCCTCACCAGCGCACATGACCCCGACGCCGGCGAGAATGGGCTCCGCACCTACCTGCTCACGCGCGACGATCACG GCCTCTTTGGACTGGACGTTAAGTCCCGCGGCGACGGCACCAAGTTCCCAGAACTGGTCATCCAGAAGGCTCTGGACCGCGAGCAACAG AATCACCATACGCTCGTGCTGACTGCCCTGGACGGTGGCGAGCCTCCACGTTCCGCCACCGTACAGATCAACGTGAAGGTGATTGACTC CAACGACAACAGCCCGGTCTTCGAGGCGCCCATCCTACTTGGTGGAACTGCCCGAGAACGCTCCGCTGGGTACAGTGGTCATCGATCTGA ACGCCACCGACGCCGATGAAGGTCCCAATGGTGAAGTGCTCTACTCTTTCAGCAGCTACGTGCCTGACCGCGTGCGGGAGCTCTTCTCC ATCGACCCCAAGACCGGCCTAATCCGTGTGAAGGGCAATCTGGACTATGAGGAAAACGGGATGCTGGAGATTGACCGTGCAGGCCCGAGA CCTGGGGCCTAACCCTATCCCAGCCCACTGCAAAGTCACGGTCAAGCTCATCGACCGCAACGACACAATGCGCCGTCCATCGGTTTCGTCTC CGTGCGCCAGGGGGCGCTGAGCGAGGCCGCCCCTCCCGGCACCGTCATCGCCCTGGTGCGGGTCACTGACCGGGACTCTGGCAAGAA CGGACAGCTGCAGTGTCGGGTCCTAGGCGGAGGAGGGACGGGCGGCGGGGGCCTGGGCGGGGGTTCCGTCCCCTTCA AGCTTGAGGAGAACTACGACAAGCTTCTACACGGTGGTGACTGACCGCCCGCTGGACCGCGAGACACAAGACGAGTACAACGTGACCATC GTGGCGCGGGACGGGGGCTCTCCTCCCCTCAACTCCACCAAGTCGTTCGCGATCAAGATTCTAGACGAGAACGACAACCCGCCTCGGTT CACCAAAGGGCTCTACGGTGCTTCAGGTGCACGAGAACAACATCCCGGGAGAGTACCTGGGCTCTGTGCTCGCCCAGGATCCCGACCTGG GCCAGAACGGCACCGTATCCTACTCTATCCTCGCCCTCGCACATCGGCGACGTGTCTATCTACACCTATGTGTCTGTGAATCCCACGAACG GGGCCATCTACGCCCTGCGCTCCTTTAACTTCGAGCAGACCAAGGCTTTTGAGTTCAAGGTGCTTGCTAAGGACTCGGGGGCGCCCGCG CACTTGGAGAGCAACGCCACGGTGAGGGTGACAGTGCTAGACGTGAATGACAACGCGCCAGTGATCGTGCTCCCCACGCTGCAGAACGA CACCGCGGAGCTGCAGGTGCCGCGCAACGCTGGCCTGGGCTATCTGGTGAGCACTGTGCGCGCCCTAGACAGCGACTTCGGCGAGAGC GGGCGTCTCACCTACGAGATCGTGGACGGCAACGACGACCACCTGTTTGAGATCGACCCGTCCAGCGGCGAGATCCCGCACGCTGCACC CTTTCTGGGAGGACGTGACGCCCGTGGTGGAGCTGGTGGTGAAGGTGACCGACCACGGCAAGCCTACCCTGTCCGCAGTGGCCAAGCT CATCATCCGCTCGGTGAGCGGATCCCTTCCCGAGGGGGTACCACGGGTGAATGGCGAGCAGCACCACTGGGACATGTCGCTGCCGCTC ATCGTGACTCTGAGCACTATCTCCATCATCCTCCTA
10	KLHL1	ATGCGCCCTCTGCACCCCTAGAGCCAGAAGACGCTAGGTGGGCTGCGCGCTCTGCCAGGCGAAGGCTGGAGCGCAGACGGCAAAGCC GCGCGTTTCAGCCGTGGTCGGGTCCGCAGGACCTGGGCGTGGGGACACCACCAGGCAGAGGCAGAGGCAGGACTGGGACGCCAAAAG CTGAGAATCCTCGATGCCCGCGCGAGAGCCCCCGTGTTAT
11	POU4F1	TTCTGGAAACCGGGCCCCACTTGCAGGCCCGGCCACCTTGGGTTCTGGTGGCCGAAGCCGGAGCTGTGTTTCTCGCAGACTCGGGGAG CTACATTGTGCGTAGGCAATTGTTTAGTTTGAAAGGAGGCACATTTCACCACGCAGCCAGCGCCCTGCATGCAGGAGAAGCCCCCAGGG CCCAGGGTCGGCTGGCTTTAGAGGCCACTTAGGTTGTTTTAAGCACATGTGAAAGGGCAGACAGCAGGGGAGCAGGATATGGGTAAGAT CTTCGGGTCTCAGAACAGGGGCTGCCCTTGGGCTGTCCCGGCGCCCTGGGCTCTGACACTGAAGGGTGGAATGGAGGAAGGAATGGAG AAAGGACGGTGGAACTTTCGCTTCCCCTCTGGGCCGCCTTCCCAGGGTCATGCCTGAGCTGCTTTGATCCCAGTGTCGCGCATCTTGGTC CGCTACCTCCCAGGCGATAGCTACTGGGCTCCTCGCTGGCCTCACTGGGGGCCATCCCGGGCAGTGGCCTGCCCTCCGAGGCCCGCGG GACCCAGCCCAGAGCTGAGGTTGGAGTTCTCCGGGCCACGTTCCGGGTCGCTTAGGCTCGGAGATTTCCCGGAGACCGTCGTCCTCCCT TTCTGCTTGGCACTGCGGAGCTCCCTCGGCCTCTCTCCTCCTCTGGTCCCTAAGGCCCCGGAGTGGTTGGCGGTACTGGGGCCCCGTCGTC

SEQ ID NO	GENE NAME	SEQUENCE
		ATCTCTGCTTCTAAGGCATTCAGACTGGGCTCCAGCTGGGACCGGCAGAGGAGGTTCTCAAGGAAACTGGTGGGAAATATAGTTTTCTTT CGTCTGGTCGTTTAATTTAAATGCAACTTCCCTTGGGGACATTTTCCTGGACGTTAACCAGACCACCTTGAGATGTCGTTGATGACCTAGA GACCCAGATGATGCGTCCCAGGAAAGTTCACTGCTGACTATTGTCACTCTTGGCGTTATATCTATAGATATAGACCTATGTACATATCTCCA CCCTGATCTCTCCGTGGACATGAAACCCACCTACCTTGTGAAAGCCCTACGGGTGACACATGACTACTACGTCTCTGTCCCAACAGGGGC TGGGCCTCCCCTGCCTAATAGTTGCCAGGAGTTTCGCAGCCCAAGTGAATAATGTCTTATGGCTGAACGTGGCCAAGGACTCCTGTGATT TAGGTCCCAGGAGGAGCAGAGACGTCCCCGCCCCGCCTGGGCCCTGCCGCATTCAAAGCTGGAAGAAGGCGCTGATCAGAGAAGGGGC TTCCAGGTCCTGGGTTAGAACAACAACAAACGAAACTCCACACAAGACACAGCCTGCCCATGACCCCACGCAAGGACATAGGAAGTT CTGTCGCCTTCCTGCTCCGCGGATAGCCGCCTGCCGTCTGCTGCCACCAGAACGCAGCGCTCGGGGTGGAGGTAGTCAATGGGCA GCAGGGGACCCCCAGCCCCCACAAGCGCGGCTCCGAGGACCTGGAAGCGGGTGCCTGTCGCTCTCCGCAGGCTCCGCTCTGCCTCCA GGAGCAAGATCCCCAAAAGGGTCTGGAAGCTGTGGAGAAAAC
12	GPC6	TTTTTTAAACACTTCTTTTCCTTCTCTTCCTCGTTTTTGATTGCACCGTTTCCATCTGGGGGCTAGAGGAGCAAGGCAGCAGCCTTCCCAGCC AGCCCTTGTTGGCTTGCCATCGTCCATCTGGCTTATAAAAGTTTGCTGAGCGCAGTCCAGAGGGCTGCGCTGCTCCCTCGGCTGGC AGAAGGGGGGTGACGCTGGGCAGCGGCGAGGAGCGCGCCGCTGCCTCTGGCGGGGCTTTCGGCTTGAGGGGCAAGGTGAAGAGCGCACC GGCCGTGGGGTTTACCGAGCTGGATTTGTATGTTGCACCATGCCTTCTTGATTCGGGGCTGTGATTCTTCCCCTCTTGGGGCTGCTGCTC TCCCTCCCCCGCCCGGGGCGGATGTGAAGGCTCGGAGAGAGGTCCCGCCAGGCGTACGGTGCCAAGGGATTCAGCCTGGCGGAC ATCCCCTACCAGGAGATCGCAGGTAAGCGCGGGCGCGCTGCAGGGGGCAGGCTGCAGCCCTCGGCTGCCGCACGTCCCACTGGCCGCC CGGCGTCCCCTTCCTCCCCCTGTTGCTGAGTTGGTGCTCACTTTCTGCCACCGCTATGGGACTCCGCGTCTCCGTGTTGGGCGGCGGA TGCTCCTGCGGCTTCTTCGGCGGGGGAAGGTGTGCGTCTCCGCCGCCTCATTGTGTGCACACGCGGGAGCACCCTGGCTCCCGCCTCC CGCTGCTCTCGCGCCCTTCTACCCCTTAGTTGATGGCTCAGGCCCGGCTGGCCAGGGAGCCCGGGTCACTCCGGGGCGGCTGCAAGGC GCAGACGGAGAGCCGAGCCGGGCGCTCACTCCGCGTTCTGGTTCGGGCAAACTTGGAAGAACTGCGACCGCAGTTTGCCCAGCGCCAC AGTCTGAGTGGCGCCTTCTCCACTCCCGCCCTTGCGCCGGCAGGGGCGGTGAGAGACGCGGAGGGCTCCCCCAGCCCCTCTCTCCCC TATCCGTCCTTCGGGCGACAGAGCGCCCGGCGCTCGGGCCGGGGCGGGCAAGGCTGGGAGGGACCCTCGCCGGGGACCTGGCCTC TGGACGCCGGCGTTTCAAGGCTGGTTTGGGGACTTCACGGGCTGCCTGTTTCAGATGTGGGGCCGGCTCGTCTGCCAGCGAACAGCCAA GCTTCCCCAGTTGCTGTTGGCCACTCAGGGCCCGGGGACACCCTGCCACCCGGTCTGGAGCCGGCCTCGTCTGCCAGCGAACAGCCAA CTTTAGCGGGTGGCTCAGCTGGGGATT
13	SOX21	CACTCAGTGTGTGCATATGAGAGCGGAGAGACAGCGACCTGGAGGCCATGGGTGGGGGCGGGTGGTGAAGCTGCCGAAGCCTACACAT ACACTTAGCTTTGACACTTCTCGTAGGTTCCAAAGACGAAGACACGGTGGCTTCAGGGAGACAAGTCGCAAGGGCGACTTTTCCAAGCGG GAGATGGTGAAGTCTTTGGACGTGTAGTGGGTAGGTGATGATCCCCGCAGCCGCCTGTAGGCCCGCAGACTTCAGAAAACAAGGGCCTT CTGTGAGCGCTGTGTCCTCCCGGAATCCGCGGCTTAACACATTCTTTCCAGCTGCGGGGCCAGGATCTCCACCCCGCGCATCCGTGGA CACACTTAGGGTCGCCTTTGTTTTGCGCAGTGATTCAAGTTGGGTAACCCTTGCTCAACACTTGGGAAATGGGGAGAATCTCCCCCCACCC GCAACCTCCCGCACCCCAGGTTCCCAAAATCTGAATCTGTATCCTAGAGTGGAGGCAGCGTCTAGAAAGCAAAGAAACGGTGTCCAAAGA CCCCGGAGAGTTGAGTGAGCGCAGATCCGTGACGCCTGCGGTACGCTAGGGCATCCAGGCTAGGGTGTGTGTGGTGCGGGTCGGGGGG CGCACAGAGACCGCGCTGGTTTAGGTGGACCCGCAGTCCCGCCCGCATCTGGAACGAGCTGCTTCGCAGTTCCGGCTCCCGGCGCCCC AGAGAAGTTCGGGGAGCGGTGAGCCTAGCCGCCGCGCGCTCATGTTTATT
14	ZIC2	AGTCACTCCAGGATCAGAGGCCGCGTCGGTTCTGCTTGGGGCATGGGCAGAGGGAGGCTGCTGGGGCCAAGCCCCGGCTGGACGCGA GGGAAGAAACTCGTCCCCAGGACCCGCACGCCCATACCTGGCTGTCCCAGAGCTCTTCCCTAGGCCGGCACCTTCGCTCTTCCTCTTCCC CACCCCCTAGCCCTTTTGTCTCTTTTTCAGACGGATGTTTTTCAGTCTCAAGTGGTTTTATTTTCCGCACACAAAACCCTGAGATCAAGGGCAGA TCACAGACTGTACCGGAGGCTCGGGTTTCCCTGGACTCTGTGCTGTTCTGCGTCCCAGGGTTGGCTAGGAAGGAAGGCCTGGGCCGGC

SEQ ID NO	GENE NAME	SEQUENCE
		GAGGTGACGGGTCTCCCGCCCAGGTCGGCAGGACGGGGGGAGGTGTGTCCCGGTAGGTCCCTGGTGAGCTCACCCGTGGCATCGGGG ACCCGCCGGGAACCCACCGGGCGCCCACTAGAGACTCGGGTCCTACCCTCCCCCACACTACTCCACCGAAATGATCGGAAGGGCGCGCT AGGCCTGCTTCCAAGGGCTCAGTGATAAAGGCCTCAAAATCACACTCCATCAAGACTTGGTTGAAGCTTTGGGTAGGTTTGTTGTTGTGT TGTTGTTGTTTGTTTTAGCAGACACGTCCTGGAAAGAGGTCCTCAGAACCCAAAGGTTCAATAATGATTTGTGGATGGATTGAT TATAGTCTGATATCGCTCTGGTTCCACAGAAACCCGGAGCTCCTTGGCCCACTGTTACCCCAGCAGACCTAAATGGACGGTTTCTGTTTTT CACTGGCAGCTCAGAACTGGACCGGAAGAAGTTCCCCTCCACTTCCCCCCTCCCGACACCAGATCATTGCTGGGTTTTTATTTTCGGGGG AAAAACAACAACAACAACAAAAAAACACTAGGTCCTTCCAGACTGGATCAGGTGATCGGGCAAAAACCCTCAGGCTAGTCCGGCTG GGTGCCCGAGCATGAAAAGGCCTCCGTGGCCGTTTGAACAGGGTGTTGCAAATGAGAACTTTTGTAAGCCATAACCAGGGCATCCTGAG GGTCTGAGTTCACGGTCAAGGCTGTGGGCTACTAGGTCCAGCGAGTCCAGGCCTCGCCCCGCCCCCGAGCTGCCACAGCCAAGATCTTC GGCAGGGAATTCGAGACCAGGGTCCTCCCACTCCT
15	chr13 group-00385	TTTCGTGCCCGCTGTTTTCAATGCGCTAACGAGGCACGTTATTCTTAGCCCGCGTCCGGGAGGGGATCACATTCCTGCGCAGTTGCGCTGCT GGCGGAAGTGACTTGTTTTCTAACGACCCTCGTGACAGCCAGAGAATGTCCGTTTCTCGGAGCGCAGCACAGCCTGTCCCATCGAGAAG CCTCGGGTGAGGGGCCCGGTGGGCGCCCGGAGGCCGCTGGAGGGGCTGTGGGAGGGACGGTGGCTCCCCACTCCCGTGGCGAAGGGC AGGCAAACCAGAAGCCTCTTTTGAGAGCCGTTTGGGATTGAGACGAGTAAGCCACAGCGAGTGGTTAGAAGTAGGTTAGGAAGAAGGGG AGGTAAGAAAGCCGAGTAGGGTT
16	chr13 group-00390	GTTCGGTGGACAAGGGGGGCAGCGCCCACAGCAAGCCGGAAAGAGGGAGGCGCGGGGCCGCGCTTGGGGCCTGCCGCTGCACGCCAG CCTGGGCAAAGAGCTGCCACCTTCTGCGGGCGAAGCGGGTCGGGACGCAGGAGACGGCAGCGGGGCTGGAGGCAGCTACGTGGGTCCAC ACCCCCATGCCCTGCCAAGGCTCCTTGGCCCTGCTTCTCCTCTGTCTCGGCGGGAGAGGAGCAGCCTCGGTTTTACAGAATTTC
17	chr13 group-00391	TGTGCCATTTAGTGAGAGGTGTTTTGGGCAAAGAATCAATTTAACTGTGACTGACCGACGGGCTTGACTGTATTAATTCTGCTACCGAAAA AAAAAAAAAAAAAAAAAAGCAATGAGCCGCAAGCCTTGGACTCGCAGAGCTGCCGGTGCCCGTCCGAGAGCCCCACCAGCGCGGCTCACGC CTCAGTCTC
18	chr13 group-00395	AGAGTCCCAGTTCTGCAGGCCGCTCCAGGGCTAGGGGTAGAGATGGTGGCAGGTGGTGCCGTCAACTCTCTAGGGAAGAGGAACTTGCAT TACAAAGACTTGTCTTTCTGAGCTGAAGTCAAAACGGGGGGCGTCAAGCGCGCTCCGTTTGGCGGCGGTGGAGGGGCCGCGCGCCCGCG CTGTCCCAGCCGGAGCTGCCCTGGCTGGTGATTGGAGGTTTAACGTCCGGAATTCAGGCGCTTCTGCAGCTCAGATTTGCCGGCCAAGG GGCCTCAGTTGCAACTTTTCAAAATGGTGTTTCTGGAAAATAACAAATTCAGACTCAACTGGTGACAGCTTTTGGCTATAGAGAATGAAACT GCTTCCCTTTGGCGGTGGAACTCTTAAACTTCGAAGAGTGAAAGAAATACAATGAAATAAAATGCCATAAGATCACTGGATTTTTCAGAAAAA GGAAGACCCCAAATTACTCCCAAAATGAGGCTTTGTAAATTCTTGTTAAAAATCTTTAAATCTCGAATTTCCCCCTACAACATCTGATGAGTG CTTTAAGAGCAAACGAGCAAATCCCACCTCGAGAATCAACAAACCCAAGGCTCTGGCCAAGGCTCTCCCCGCGTTTTCTTCTCGTGACCTG GGGAATGTCCCGCCCCATCGCTCACCTGGCTCTTGTCATCTCGCTCATCTTGAAGTGACCCGTGGACAATGCTG
19	chr13 group-00399	AGCTGCCCTCTGTGGCCATGAGCGGGTGTCCAGCCCCTTCCAAGGCTGCACCGGGGAGACGCTGGTTTTCTGCTCGCTGTGACCGAACA AAGCCCCTAAGAGTCAGTGCGCGGAACAGAAGAGCCGGACCCCGACGGGCCGAGTCCCAACGTGAGGCACCCGGCAGAGAAAACACGT TCACG
20	PROZ	CCTCGGCAGCACCGGCATGGCTGGAGGCCAGTACGGCCAGGTGTGGCGGAGGAGCGCCGTCTGGCTTGGGTCGTCCATCCTGACA GGACGCTGCAAGGGCAGGAGCCCCGCGCCCCCGTGTCCTGCGCCCCCCGCTCGAGGACAAGCCCCAGCCGCCGGTCTCCGCTGGGTTCC GACAG

SEQ ID NO	GENE NAME	SEQUENCE
21	CIDEA	CTTTAAGAGGCTGTGCAGGCAGACAGACCTCCAGGCCCGCTAGGGGATCCGCCGCCATGGAGGCCGCCCGGGACTATGCAGGAGCCCTC ATCAGGCCGAGTGCCCCCGCGTCCCCCTGATTGCCGTGCGCTTCCAATCGCCTTGCGTTCGGTGGCCTCATATTCCCCTGTGCGCCTCTAGT ACCGTACCCCCGCTCCCTTCAGCCCCCCTGCTCCCCGCATTCTCTTGCGCTCCGCGACCCCGCGCACACACCCATCCGCCCCCACTGGTGCC CAAGCCGTCCAGCCGCGCCCGCGGCAGAGCCCAATCCCGTCCCGCGCCTCCTCACCCTCTTGCAGCTGGGCACAGGTACCAGGTGTG GCTCTTGCGAGGTG
22	chr18 group-00091	AGACTTGCAGAACTCGGGCCCCCTGGAGGAGACCTAACCGCCACGGTCTTGGGGAGGTTCCGGAGGGCCTCGGTTGTCTGCACTCCCA ACACCAAGAAACCCCTGAGACGCGAAGCTGCCAGCGTGCTGCCCTCAGAGCAGGGCGACGCAAAGCCAGCGGACCCCGGGGTGGCGG G
23	chr18 group-00094	TGCTCGGCTGGGGGGCTCGCTCCGCACTTTCGGTGCCAGAAAATGCCCAGAGGAGCGGGGCGCCCCAGAGCCTCCTTTCGGGGCGC GAGGCCCGGCGCGTGTGTACGGAGTCCAGTCCCCCCCAGGGAGTGGGGTGCCCGCACCTTCCCCTCCGCGCTCGGAGCCAC
24	KLHL14	TCTTGCACACCTGCTTGTAGTTCTGCACCGAGATCTGGTCGTTGAGGAACTGCACGCAGAGCTTGGTGACCTGGGGGATGTGCAGGATCT TGCTGACCGACACAGCACCTCCTCCACCGTGTCCAGGGACAGGGTCACGTTGGCCGTGTAGAGGTACTCGAGCACCAGGGCGCAGCCCGAT GGACGAGCAGCCCTGCAGCACCAGGTTGTTGATGGCCCCGGGGGCTGGTCAGCAGCTTGTCGTCGGGGGAGGAAGAAGGAGTCCCGGG CTCCTCCTGCGGCGGCGGCTGCTGCTGCTGACGGCTGCTGCTGCCGGCGGCTGCTGCTGGTCCTTGGGGGCCCCCCAGGCCGTCCTG GCCGCCGACCCCCTCCCCCGAGAGGGGGGTGGCTGGAGAAGAGCGATCGGAAGTACTGCGAGCAGGAGGCCAGCACGGGCCTTGTGGCA ATGGAACTGCTGGCCCTGGGCCGTCAGGGTCACGTCGCAAAACAGCTGCTTCCTCCACAGCAGGTTGAGGCCGTGCAGCAGGTTGTCGC TGTGGCTGGGGTCGAAGGTGGAGGTCCTGTCCCCGGATCTGGCATGGCGAGCTGACTCGGTGCACCTGGCTTTAAACCCTCCTCCAAC CTGGCAGACAGGGGTGGGGGATGGGAGGGGAGCAGGGTGGTGGAGCGGGGTGGGGTGTGGTCGGGGGTGGGGAAGGGTGTGGA GGGGAGGGGAGGGGCGAAGAACAAGAATCAAGGCTCAGCTTGACTCCCTCCTGGCGCGCTCCGGACCCCGACCCTAGGAGGAAAGTCCG AAGACGCTGGATCCGTGAGCGCCACCAGAAGGGCCCTGTCTGGGGTCCCGGCGCCGGTTCTGCGCCCTGCGGCTCCTCTCGCCACCTC CCACACACTTCGTCCCTCACTTTCCTAAAACCAACCACCTCAGCTCGGCTGTTGGCAGCAACAGCAGTGGCAGCGACGGCAAAGTG GCGGCTGAGGCCGAGGCACCTCGTGGGCTCGTGTCCATGCCGGGGCCAGATGAAGGGAAAGGCCGGGAAGTGGGGAGCCGGGGGTGC CCTGAAAGCTCAGAGGCGACCGACGGCGAAGGTTCCAGGTCAACTTGTGCCCGAAGCTTTGCTTTTCGCAGTTGGCCCAGTTTGGGGGA GGGGGTAGGAACAGGGGGCCCGACCAGCGTGCGGGGGTGTGCGAATCTTAGCTCTCCCAAAAGCTG
25	ST8SIA3	CCTCTGTGTTAGTGCCCTCGGGAATTTGGTTGATGGGGTGTTTG
26	ONECUT2	TGATGTCGCACCTGAACGGCCTGCACCACCCGGGCCACACTCAGTCTCACGGGCCGGTGCTGGCACCCAGTCGCGAGCGGCCACCCTC GTCCTCATCGGGCTCGCAGGTGGCCACGTCGGGCCAGCTGGAAGAAATCAACACCAAAGAGGTGGCCCAGCGCATCACAGCGGAGCTG AAGCGCTACAGTATCCCCCAGGCGATCTTTGCGCAGAGGGTGCTGTGCCGGTCTCAGGGGACTCTCTCCGACCTGCTCCGGAATCCAAA ACCGTGGAGTAAACTCAAATCTGGCAGGGAGACCTTCCGCAGGATGTGGAAGTGGCTTCAGGAGCCCGAGTTCCAGCGCATGTCCGCCT TACGCCTGGCAGGTAAGGCCGGGGCTAGCCAGGGGCCAGGCTGCTGGGAAGAGGGCTCCGGGTCCGGTGCTTGTGGCCCAAGTCTGC GCGCCGAGTCACTTCTTGATTCTTTCCTTCTTTCCTATACACGTCCTCTTTCTTCGTTTTTATTTCTTCTTCCATTTTCTCTTTCTCTT CCGCTCTTCCCCCTACTTTCCCTTTTTCTTTTTCTTTCCTTGTCCCTGAGCTTTCATTGACCGACCCCCCCCCCCCATTTTC ATTCGCCCTCCCCCTCAATGTGCCAACCTTTGCCCTATTTCCGATCTTCCCAGGTACTGGGAGGCGGGATGGGGGTGTGCGTTTTCCTCTA GGAGCCCTGTCTTTCCAAGACCCACAGAAACCAGGACCTGCCCTTATTCAAAACCCCATGCACTTCAAGTCTCTTTTAGACAACACATTTC

SEQ ID NO	GENE NAME	SEQUENCE
		AATTTTCCGGGCTGACTAGTCTCCCTGTGCAGAGGCAGTTGAGAGGCTTTGCTCTGCAGAGGGAAAAGAGCTCTCTACTCTCCCACCCAC
		CATATAGGCAAACTTATTTGGTCATTGGCTGAAGGCACAGCCTTGCCCCCCGCGGGGAACCGGCGGCCAGGATACAACAGCGCTCCTGGA
		GCCCATCTCTGGCCTTGGCGTTGGCGCAGGGACTTTCTGACCGGGCTTGAGGGGCTCCGGGCCAGCTCCAATGTCACTACCTACAGCGAG
		GGCAGGGTGTAAGGTTGAGAAGGTCACATTCACCGCTTTGGGAGGACGTGGGAGAAGAGACTGAGGTGGAAAGCGCTTTGCCTTGCTCA
		CCGGCCGTCCTTGCCCCGGTCCCAGCGTTTGCTGGGATTTGCCAGGATTTGCCGGGGCTCCGGGAGACCCTGAGCACTCGCAGGAAGA
		GGTGCTGAGAAATTAAAAATTCAGGTTAGTTAATGCATCCCTGCCGCCGGCTGCAGGCTCCGCCTTTGCATTAAGCGGGCGCTGATTGTG
		CGCGCCTGGCGACCGCGGGGAGGAGGGGACGGGTAGAGGCGCGGGTTACATTGTTCTGGAGCCGGCTCGG
		CTCTTTGTGCCTCCTCTAGCGGCCAAGCTGCGAGGTACAGCCCTCTATTGTTCTAGGAGCACAGAAACCTCCTGTGTGGGCGGCGGGTG
		CGCGAGCTAGAGGGAAAGATGCAGTAGTTACTGCGACTGGCACGCAGTTGCGCGCTTTTGTGCGCACGGACCCCGCGCGGTGTGCGTG
		GCGACTGCGCTGCCCCTAGGAGCAAGCCACGGGCCCAGAGGGGCAAAATGTCCAGGTCCCCCGCTGGGAAGGACACACTATACCCTAT
		GGCAAGCCAGGGTGGGCGACTTCCCATGGATCGGGTGGAGGGGGGGTATCTTTCAGGATCGGCGGGCGGTCTAGGGGAACAATTCGTGG
		TGGCGATGATTTGCATAGCGCGGGTCTTGGGATGCGCGCGGTTCCGAGCCAGCCTCGCACAGCTCGCTTCCGGAGCTGCGAGCTCAGG
		TTTCCACCCCCGATCCCCCGGGCTTTCCTCGCACCGCTGAGCCCAGCTTGTGGGGTGCACTCGACCAACGCCCGACAGGGCTGGGGAA
		TGTGACAGGCAGCAGGTTCACCCGGGCTTGGGGAGGGGGGAGTTTCCGCTTTGACAGCATTTTCCTTTGCCGTCTGCTGGTGGATTCCTAT
		TCCCAGTCGGTAATCGCCCCGCAGTGTTGATCTAAGAAGGTAAAGAAAACTAGGTTTCCCTGCAAAGAGCCTCCCCCAAATCGGCGGACT
		CCGGATACTTTGAGTGGATTTAGAAATTTATGTAATCTTTCTCCTTTAGTTTATTTTTCATCCTCTCCTACAGTTTTCTCTGATTTGCTGTTGG
		TTCGGGGCAAGATAAAGCAGCCAGTAGAGAGCGATAATAATAGCGGCGGGAAATGAACTGGAGACTGGCTGACAGTTCTTAACATTTTGT
		CATAGATCCCCCCCGAATGTCCCAGGCTGTCTCTGGTGGGTTTTAGTACCCGCCGGCTTCTTGGGCACCGGGGACCAGAAGGAACTTGGC
		AGCTGGTCTTAGGGGTACAGTTAAAGGCAGGATGACAGCTATTCTCCTGCTCATCTCAGAGCGCTGCCGCCCCCTCATGCCGGTCGCGC
		AAAGAACACAGCTTTTAAAAAACACGTGCCTTCTGCCCATATAGGTCTGAAAGTGATGAGGAAAGTAATGCTTCGCCTATTAGCGAGTTTCA
		GCTTTTAAAATGATCCCAAGCGTTGCTGAGATGAGAAAGCGGTGGCATCCCGGGGGTCCTCAGCCCCACCCGCGCCCATGGTGCAAGTCT
		GCAGGGACAGGCCCGGGACAGCACTGCCCACGCTGCTAGATTTTCCGCAGAGGATCGCTGAAGCTGCCTTCGTGGGAGACAGAATGCC
		TCCTCCAGCGAGTGGAAAAGGCCTGCTGAGGACCCCGCTTTGCTCGAGCATTCAAATGTGTGTCTGTTTTATTACCCTGGGTTGAAAAGG
		GACAAGAGCTTTAGCCTTTTTATCTGGCCATTTTATCAGCAACTACAAGTGTGTTGAGTGGTTATTATTACATAGGAGGCTTTTCAGTTTGG
		GGTCAGTAGATCAGTCTCTTCAGACACTGATGCAGAAGCTGGGACTGGTAAGTAGGTATTATGTGCTCGGAGCGCTAGGGGACAGGAGC
		AAATGGAGAAGAAAAGCGGAGGCTTTCTCCGCCCGGAGTATCGATCGGAATCCCCGCCGGTACGCCGCAGAGGGCCCTCGCCGTTGGG
		CCCCGGGGGGTTTAACAAGCCCAGCCGCTCCGCAGGCGGCTCGGCCGGACTCTCAGACCGGTGCCTGGAAGACACCGTCCCTGCCCCCC
		TCCCGCCAAACCTGCCTCTTCTCTTTCTCTCATAGGTTATAGGTTCCCTTTCTCTCATTTTGGCCCCCGCCCCGGGTCCTGCCAAACAG
		CCAAGCAGGCCGGGGTTTAGGGGGCTCAGAATGAAGAGGTCTGATTTGGCCAGCGCCGGCAAAGCTCACCCTTAGGCGAGGTCACAAC
		AGAGGCAGGTCCTTCCTGCCCAGCCTGCCGGTGTAGTCACAGCCAAGGGTGGCACTTGAAAGGAAAAGGGAGAAAACTTCGGAGAAATT
		TAGATTGCCCCAACGTTAGATTTCAGAGAAATTGACTCCAAATGCACGGATTCGTTCGGAAAGGGCGGCTAAGTGGCAGGTGGTTGCAAC
		CCCGCCCCGGTCGGGCCTTCGCAGAGGTTCCCCAAGACCAGCCCTTGCAGGGCGGTTTTCAGCAACCTGACAAGAGGCGGCCAAGACAA
		ATTTCTGCGGGTTCGAGCACACACTCTCGGGCGTTGGGCCCCAGAGACCTCTAAACCAAGCACAAACAAGAAGGGAGTGAGAGAACCCA
		GGCTAGAACTTGCACGGGCATCCCACTGAGGAAAAGCGAGGCCTCGGTGGCAGGCATGTTTTCTTCCGACGCCCGAAAATCGAGCCGAG
		CGCCCGACTACATTTACTGCAGAGGTTTCCGCCTCCAGTGAGCCCGGATCCCCCCAGCGGCCTGCCCGGAGCTGGTCTCCAGTCCCCGCC
		GTAGTCCGACGCACGGCCCTCTCCTGGCAGCAAGCTCCCAGCGGCCAGTCTGAAGCCAATTCTGTTCAGGCGGCCGAGGGCCCTTAGC
		CAACCCACCATGATGTCGCCTGGGCCACCCTGATGCCCGCAGCGGCGGGACACGGCCCGGGCAGTGCCGCAGTGGCTCCTGCTAGGGGC
		ACCGCGTGCGTGCTTGTCTCCCGCTGCGCCGGGGACGTCCTTGGGTGACACGGGCCGCTGGGCACCTCCCAAGCCGAGGAAACGGAC
		CCCCTTCGCAGAGTCTCGCGCCCACCCCCCAACCTCCCACCTCGTTTCTCGCTGCTAGGGCTCCCGACTCAGCCCACCTCTCCTGGCGG
		TTTAGTTAGGGATCAGAGCTGGAGAGGCTGAACGCAACCCCGTGCCCAGTACGGAACAGACGATATGTTTGCCTGCTAGCTGCTTGGATGAA

SEQ ID NO	GENE NAME	SEQUENCE
		TAATTGAAAAGTTCGCTGCAGTCTGTGCTTCGTCAAGTCCCGGGTGCCGGGAGAACACCTTCCCAACACGCATCAGGGTGGGCGGGAGC GGGCAGAGGAGGCGGGACCCGAGGGAGAGTGAACCCGAGCAGGAAGCAGCCCAGGCAGCCCAGGCGCCCTCGATGCGAGAGG CTGGGCATTTATTTTATTCCAGGCTTTCCACTGTGTGGTTATGTCACTTTCTCAAACAAATGTGTATATGGAGGGAGATCGATGCTGATAA TGTTTAGAAGATTAAAAGAGCATTAATGCTGGCAACAATAACGTGTGTGGACCCAGATTTCATTGATCTGGAACTTGATCCGGCGCG TTTCCAGTAAGCCCGACGGCGCGCTCTTCCCAGCAGAGCGCTCACCAGCGCCACGGCCCCGCGGTTTTCCAGCGGTGCCGCTTCGCCA GCTCTGCGCGGGTTCTCCCGTCTGACCGCAGCTCCTCCCCCGCGAGGCCCCAGCCCGCCTTACTTCCCCGAGGTTTTCTCCTCCTCTCG CGGGGCTCTCTGCCCTCTGCACCCCCTCCCCCGACCTCTGCACCACCCGCCCCTGTGCGCACACACCGCTACTTGCGCTTCCGGCGATC CGCCTG
27	RAX	AACCGGAGATCTGCTTGGTGAACTGAGAGGAGTCCTTAGGAGAGCGGGGACCGCCAGGGGCCGGGGGGACACTTCGCTCTCGCCCTAGGG AAGGTGGTCTTGACGCTTTCTATTGAAGTCAAACTTGAAAATATCAGCTGCCGCTGGACTAT
28	chr18 group-00277	CGTGAGCAGAACGCCCCGCCCTGGAGCAGTTAGGACCGAAGGTCTCCGGAGAGTCGCCGGCGGTGCCAGGTAACGCAGAGGGCTCGGG TCGGGCCCCGCTTCTGGGGCTTGGGACTCCGGGCGCGCGGAGCCAGCCCTCTGGGGCGAAATCCCCGGGCGGCGTGCGCGGTCCCTC TCCGCGCTGTGCTCTCCCAGCAACTCCCTGCCACCTCGACGAGCCTACCGGCCGCTCCGAGTTCGACTTCCTCGGACTTAGTGGGAGAA GGGGTTGGAAATGGGCTGCCCGGGACTGGGGGAGCTGCTCTCTCTGGAAGCAGGGAAGCTGGGGCGCACCGGGGCAGGT
29	NETO1	TAGAAGAGGAAGACTCCTCTGGCCCCACTAGGTATCATCCGCGCTCTCCCGCTTTCCACCTGCGCCCTCGCTTGGGCCAATCTCTGCCGC ACGTGTCCATCCCTGAACTGCACGCTATCCTCCACCCCCGGGGGGGTTCCTGCGCACTGAAAGACCGTTCTCCGGCAGGTTTTGGGATCC GGCGACGGCTGACCGCGCCGCCCCCACGCCCGGTTCCACGATGCTGCAATACAGAAAGTTTACGTCGGCCCCGACCCGCGCGGGAC TGCAGGGTCCGCCGGAGCGCGGCGCAGAGGCTTTTCCTGCGCGTTCGGCCCCGGGAAAGGGGCGCGGAGGGCTGGCTCCGGGAGCGCC ACGGGCGCGGCGGGGAGGGTACTCACTGTGAAGCACGCTGCGCCCATGGATCATGTCTGTGCGTTACACCAGAGGCTCCGGGCTCCAC TAATTCCATTTAGAGACGGGAAGACTTCCAGTGGCGGGGGGAGGACAGGGTCGAGAGGTGTTAAAGACGCAAAGCAAGAAGGAAATAAA GGGGGGCCGAGAGGGAAGGGAGACCGAGAGGAAGGGGGAGCTCCGAGCCCACGCTGCAGCCAGATCCGGATGAGTCCGTCCTCCGCCCCCGG GCGGGCTCTCGCTCTCGCTGGCCCTCAGCGCCGCGCAGCCAGCATCCCCACCGTGACGCTCGCATCACACCCGGGCGCCGGCCGCCG CCACCATCCGCGCCGCCGCCGTCAGGACCCTCCTCCCGGGCATCGTCGCCGCCGCGGGGTCGGGAGGACGCGGCGCGCGGGAGGCG GCGGTCGCAGGGCGAGCCCCGGAGCCCCGAGCCGGGGCCGGGGAGAGGGCGCAGCGAGGTGGGGGCCAGTCCAGACC GACGGCAGCGACGGAGCGGCGGCGCCGGCGGGTGGCTCAGTCCCCAGTCTCAGACGCGCCGCGCAGCA GGTCGGAGCAGCCTCCCCCGGGAGGATGTCCAGCGGCAGCGCTCCTCGCTCCAGCCCTTGGGGATCTTCCGCTGAGGCATTGAAGGCAG GAAGAAGGGGTCCGTCATCGGCTCGCCGGCTGCGCGCCACCTCTGCTATCTTGCGGAAAGAGGAGCGGGTGGGTGGCGTCTGGGA GGCGGGCTGGAGGGCGGTGCAGGGGAGCCGGGCGCCGGGGGGGGCCGGGGGGCGGGAAGGGAGGAGAGGAGATGCAGCCG GAAGAGGGCAGAGTTACCAAATGGGCTCCTTAGTCATGGCTTGGGCTCCACGACCCTCCTGGAAGCCCGGAGCCTGGGTGGGATAGC GAGGCTGCGCGCGCCGCCCCGGGGCTGGTGCGCGGCGGCGCAGCAAGGACATCCCAGCCGCGCG GATCTGGGGGAGGGGCGGGAGGGGGTGAGGACCCGGCTGGGATCCGCGGCTCGGCCCGCCAGGGCGCAGAGAGAGGATGCAGCCG CAAATCCCGAGCCGGATCCTCGTGCCGGACGGAAGGCGTGGAAGCGGGAGGGGCCTTCGTGTGAAAATCCCTTGTGGGGTTTGGTGTTT CACTTTTAAAGGTTAGACCTTGCGGGCTCTCTGCCTCCCACCCCTTCTTTTCCATCCGCGTAAAGGAACTGGGCGCCCCCTCTCCCTCCC TCCCTGGGGGCGCAGGTTTCGCCGCGGACTCCGCGCTCAGCTTGGGAGACACGGCAGGGGCGCGCCCCCAGGGAAAGGCGGCCGTAAAA GTTTCGCGGTTGAGCACTGGGCCTGATGTCCAGTCCCCCCCACCAAATTACTCCTGCAAAGACGCGGGCTTCTTGCAATTGAGCCCCCCAC CTCGAGGTATTTAAAACCACCCCAAGGCACACACGGACCCCCCGTTCCCCCCGCGCCACTTCCCTCCTACAGGCTCGCGGCGGCGTTAAAG TCTGGGAGACACGAGTTGCGGGGAAACAGCACCGGAAG

SEQ ID NO	GENE NAME	SEQUENCE
30	MBP	AAGAAACAGCTCATTTCGGAGCTGAGGACAAGGCGTGGGAAGAAGACGCGTTTGGTTTCACCCAGGCGGGTGGCGGCAAAGCTGTGGG ATGCGCGCTGCACACTCCTTCCGTCATCCCGTTCCCACCTTCCACACACCTGCGGGAGGTCGGACATGTCCTGATTGCGTGTTCATCA CGATGGCAAACCGAACATGAGGAGAACGCCACTGACGCTGGGTGCGCCGGCTTTCCCAGCCCTCGTGCATAACGGGGAGGGAGATGCA GAAGTTTTTTCCAACATCGGTGCAAAGGGGAAGCTGAGGTTTTCCTAT
31	NFATC1	TCTGTCAGCTGCTGCCATGGGGCAGCGGGAAGGCCCTGGAGGGTGCCTGGGCTGTGTCTGGTCCCGGCCACGCGTCCCTGCAGCGTCT GAGACCTTGTGGAACACACTTGACCCGGCGCTGGGACGGGGTCGGCCCACACGCACCGCCAGGCCCGCAGGAGTGAGGTGCAGGCTGC CGCTGGCTCCTTAGGCCTCGACAGCTCTCTTGAGGTCGGCCCTCCCTCCCCGAGAGCTCAGCAGCCGCAGACCCAGGCAGAGAGA GCAAAGGAGGCTGTGGTGGCCCCCGACGGGAACCTGGGTGGCCGGGGGACACACCGAGGAACTTTCCGCCCCCCCGACGGGCTCTCCC ACCGAGGCTCAGGTGCTCGTGGGCAGCAAGGGGAAGCCCCATGGCCATGCCGCTTCCCTTTCACCCTCAGCGACGGCGCCCTCCTGTGC CCGCGGGGAACAAGACGGGCTCTCGGCGGCCATGCAGGCGGCCTGTCCCACGAACACGATGGAGACCTCAGACGCCGTCCCCACCCTGT CACTGTCACCATCACCCATCCTGTCCCCTCACGCCTCCCCCACATCCCATCATTACTAC
32	chr18 group-00430	GAAGTAGAATCACAGTAAATGAGGAGTTAGGGAATTTAGGGTAGAGATTAAAGTAATGAACAGAGGAGGAGGCCTGAGACAGCTGCAGAG AGACCCTGTGTTCCCTGTGAGGTGAAGCGTCTGCTGTCAAAGCCGGTTGGCGCTGAGAAGAGGTACCGGGGGCAGCACCCGCCTCCTG GGAGAGGGATGGGCCTGCGGGCACCTGGGGGAACCGCACGGACACAGACGACACTATAAACGCGGGCGAGACATCAGGGACCGGGAA ACAGAAGGACGCGCGTTTCGAGCAGCTGCCCAGTGGGCCACAAGCCCCGCCACGCCACAGCCTCTTCCCCTCAGCACGCAGAGA
33	OLIG2	TACTCCGGGCGACGGGAGGAGGATGTTGAGGGAAGCCTGCCAGGTGAAGAAGGGGCCAGCAGCACACAGAGCTTCCGACTTTGCCTTCCAG GCTCTAGACTCGCGCCATGCCAAGACGGGCCCCTCGACTTTCACCCCTGACTCCCAACTCCAGCCACTGGACCGAGCGCGCAAAGAACC TGAGACCGCTTGCTCTCACCGCCGCAAGTCGGTCGCAGGACAGACACCAGTGGGCAGCAACAAAAAAAGAAACCGGGTTCCGGGACAC GTGCCGGCGGCTGGACTAACCTCAGCGGCTGCAACCAAGGAGCGCGCACGTTGCGCCTGCTGGTGTTTATTAGCTACACTGGCAGGCG CACAACTCCGCGCCCCCGACTGGTGGCCCCACACAGCGCGCACCACACATGGCCTCGCTGCTGTTGGCGGGGGTAGGCCCGAAGGAGGCATC TACAAATGCCCGAGCCCTTTCTGATCCCCACCCCCCCGCTCCCTGCGTCGTCCGAGTGACAGATTCTACTAATTGAACGGTTATGGGTCA TCCTTGTAACCGTTGGACGACATAACACCACGCTTCAGTTCTTCATGTTTTAAATACATATTTAACGGATGGCTGCAGAGCCAGCTGGGAA ACACGCGGATTGAAAAATAATGCTCCAGAAGGCACGAGACTGGGGCGAAGGCGAGAGCGGGCTGGGCTTCTAGCGGAGACCGCAGAGG GAGACATATCTCAGAACTAGGGGCAATAACGTGGGTTTCTCTTTGTATTTGTTTATTTTGTAACTTTGCTACTTGAAGACCAATTATTTACTA TGCTAATTTGTTTGCTTGTTTTTAAAACCGTACTTGCACAGTAAAAGTTCCCCAACAACGGAAGTAACCCGACGTTCCTCACACTCCCTAGG AGACTGTGTGCGTGTGTGCCCGCGCGTGCTCACAGTGTCAAGTGCTAGCATCCGAGATCTGCAGAAACAAATGTCTGAATTCGAAATG TATGGGTGTGAGAAATTCAGCTCGGGGAAGAGATTAGGGACTGGGGGAGACAGGTGGCTGCCTGTACTATAAGGAACCGCCAACGCCAG CATCTGTAGTCCAAGCAGGGCTGCTCTGTAAAGGCTTAGCAATTTTTTCTGTAGGCTTGCTGCACACGGTCTCTGGCTTTTCCCATCTGTA AAATGGGTGAATGCATCCGTACCTCAGCTACCTCCGTGAGGTGCTTCTCCAGTTCGGGCTTAATTCCTCATCGTCAAGAGTTTTCAGGTTT CAGAGCCAGCCTGCAATCGGTAAAACATGTCCCAACGCGGTCGCGAGTGGTTCCATCTCGCTGTCTGGCCCACAGCGTGGAGAAGCCTT GCCCAGGCCTGAAACTTCTCTTTGCAGTTCCAGAAAGCAGGCGACTGGGACGGAAGGCTCTTTGCTAACCTTTTACAGCGGAGCCCTGCT TGGACTACAGATGCCAGCGTTGCCCCTGCCCCAAGGCGTGTGGTGATCACAAAGACGACACTGAAAATACTTACTATCATCCGGCTCCCC TGCTAATAAATGGAGGGGGTGTTTAACTACAGGCACGACCCTGCCCTTGTGCTAGCGCGGTTACCGTGCGGAAATAACTCGTCCCTGTACC CACACCATCCTCAACCTAAAGGAGAGTTGTGAATTCTTTCAAACACACTCTTCTGGAGTCCGTCCCCTCCCTCCTGCCCGCCCTCTACCCC TCAAGTCCCTGCCCCCAGCTGGGGGCGCTACCGGCTGCCGTCGGAGCTGCAGCCACGGCCATCTCCTAGACGCGCGAGTAGAGCACCA AGATAGTGGGGACTTTGTGCCTGGGCATCGTTTACATTTGGGGCGCCAAATGCCCCAGCCCCGGATCCTGGGTCGCCAGGGTTTTCCGCGCGCATCCCAAAAG GGCGGGAAACCAGACAGAGGAAGAGCTAGGGAGGAGACCCCAGCCCCGGATCCTGGGTCGCCAGGGTTTTCCGCGCGCATCCCAAAAG GTGCGGCTGCGTGGGGCATCAGGTTAGTTTGTTAGACTCTGCAGAGTCTCCAAACCATCCCATCCCCCAACCTGACTCTGTGGTGGCCGT

SEQ ID NO	GENE NAME	SEQUENCE
		ATTTTTTTACAGAAATTTGACCACGTTCCCTTTCTCCCTTGGTCCCAAGCGCGCTCAGCCCTCCCTCCATCCCCCTTGAGCCGCCCTTCTCC TCCCCCTCGCCTCCTCGGGTCCCTCCTCCCCAAGAATCTCCCGGCCACGGGCGCCCATTGGTTGTGCGCAGGGAGGAGG CGTGTGCCCGGGCCTGGCGAGTTTCATTGAGCGGAATTAGCCCGGATGACATCAGCTTCCCAGCCCCCCGGCGGGCCCAGCTCATTGGC GAGGCAGCCCCCTCCAGGACACGCACATTGTTCCCCGCCCCCGCCCCACCGCTGCCCGCCTGCCACCGGGCTATAAAA ACCGGCCGAGCCCCTAAAGGTGCGGATGCTTATTATAGATCGACGCGACACCAGCGCCCGGTGCCAGGTTCTCCCCTGAGGCTTTTCGG AGCGAGCTCCTCAAATCGCATCCAGAGTAAGTGTCCCCGCCCCACAGCAGCCGCAGCCTAGATCCCAGGGACAGACTCTCCTCAACTCG GCTGTGACCCAGAATGCTCCGATACAGGGGGGTCTGGATCCCTACTCTGCGGGCCATTTCTCCAGAGCGACTTTGCTCTTCTGTCCTCCCC ACACTCACCGCTGCATCTCCCTCACCAAAAGCGAGAAGTCGGAGCGACAACAGCTCTTTCTGCCCAAGCCCCCAGTCAGCTGGTGAGCTC CCCGTGGTCTCCAGATGCAGCACATGGACTCTGGGCCCCGCGCCGGCTCTGGGTGCATGTGCGTGTGTTTGCTGCGTGGTGT CGATGGAGATAAGGTGGATCCCGTTTGAGGAACCAAATCATTAGTTCTCTATCTAGATCTCCATTCTCCCCAAAGAAAGGCCCTCACTTCCC ACTCGTTTATTCCAGCCCGGGGGCTCAGTTTTCCCACACCTAACTGAAAGCCCGAAGCCTCTAGAATGCCACCCCGCACCCCGAGGGTCAC CAACGCTCCCCTGAAATAACCTGTTGCATGAGAGCAGAGGGGAGATAGAGAGCTTAATTATAGGTACCCGCGTGCAGCTAAAAGGAGG GCCAGAGATAGTAGCGAGGGGGACGAGGAGCCACCGGGCCACCCTGTGCCGGGACCCCCGCGCTGTGGTACTGCGGTGCAGGCGGGAGCA GCTTTTCTGTCTCTCACTGACTCACTCTCTCTCTCCCTCTCTCTCTCATTCTCTCTCCTCTCCTCTTTTCTCCTCCTGGAAGTTTTCG GGTCCGAGGGAAGGAGGACCCTGCGAAAGCTGCGACGACTATCTTCCCCTGGGGCCATGGACTCGGACGCCAGCCTGGTGTCCAGCCG CCCGTCGTCGCCAGAGCCCGATGACCTTTTTCTGCCGGCCCGGAGTAAGGGCAGCAGCGGCAGCGCCTTCACTGGGGGCACCGTGTCC TCGTCCACCCCGAGTGACTGCCC
34	SIM2	TTAATTCGAAAATGGCAGACAGAGCTGAGCGCTGCCGTTCTTTTCAGGATTGAAAATGTGCCAGTGGGCCAGGGGCGCTGGGACCCGCG GTGCGGAAGACTCGGAACAGGAAGAAATAGTGGCGCGCTGGGTGGGCTGCCCCGCCACGCCGGTTGCCGCTGGTGACAGTGGC TGCCCGGCCAGGCACCTCCGAGCAGCAGGTCTGAGCGTTTTTGGCGTCCCAAGCGTTCCGGGCCGCGTCTTCCAGAGCCTCTGCTCCCA GCGGGGTCGCTGCGGCCTGGCCCGAAGGATTTGACTCTTTGCTGGGAGGCGCGCTGCTCAGGGTTCTG
35	SIM2	CCGGTCCCCAGTTTGGAAAAAGGCGCAAGAAGCGGGCTTTTCAGGGACCCCGGGGAGAACACGAGGGCTCCGACGCGGGAGAAGGAAGGATT GAAGCGTGCAGAGGCGCCCCAAATTGCGACAATTTACTGGGATCCTTTTGTGGGGAAAGGAGGCTTAGAGGCTCAAGCTATAGGCTGTC CTAGAGCAACTAGGCGAGAACCTGGCCCCAAACTCCCTCCTTACGCCCTGGCACAGGTTCCCGGCGACTGGTGTTCCCAAGGGAGCCCC CTGAGCCTACCGCCCTTGCAGGGGGTCGTGCTGCGGCTTCTGGGTCATAAACGCCGAGGTCGGGGGTGGCGGAGCTGTAGAGGCTGCC CGCGCAGAAAGCTCCAGGATCCCAATATGTG
36	DSCR6	GCGCAGGTCCCCCCCAGTCCCCGAGGGAGTGCGCCCGACGGAAACGCCCCTAGCCCGCGGGCCTCGCTTTCCTCTCCCGGGTTCCTGG GTCACTTCCCGCTGTCTC
37	DSCAM	TTCCCTCGCGGCTTTGGAAAGGGGGGTGCAAATGCACCCTTCTGCGGGCCCGCTACCCGCTGCAACACCTGTGTTTCCTTTCTGGGCACCT TCTAGGTTTCTAGATATTGCTGTGAATACGGTCCTCCGCTGTACAGTTGAAAACAAA
38	chr21 group-00165	TGGGAATTTAGGTCGGGCACTGCCGATATGTCGCCTTCCACAAGGCGGGCCCGGGCCTCTGCTGACCGTGCACCGGTCCTGGGGCTGG GTAATTCTGCAGCAGCAGCGCAGCCCATGCCGGGGAATTTGCGGGCAGAGGAGACAGTGAGGCCCCGCGTTCTGTGCGGGAACTCCCGA GCTCACAGAGCCCAAGACCACACGGCTGCATCTGCTTGGCTGACTGGGCCAGGCCCACGCGTAGTAACCCGGACGTCTCTCTCTCACAG TCCCCTTGCGTCTGGCCAGGGAGCTGCCAGGCTGCACCCCGCGGTGGGGATCGGGAGAGGGGCAGTGTCGCCCATCCCCGGAAGGCT GAGCCTGGTGCAG

SEQ ID NO	GENE NAME	SEQUENCE
39	PRMT2	CGGTTTTCTCCTGGAGGACTGTGTTCAGACAGATACTGGTTTCCTTATCCGCAGGTGTGCGCGGCGCTCGCAAGTGGTCAGCATAACGCC GGGCGAATTCGGAAAGCCCGTGCGTCCGTGGACGACCCACTTGGAAGGAGTTGGGAGAAGTCCTTGTTCCCACGCGCGGACGCTTCCC TCCGTGTGTCCTTCGAGCCACAAAAAGCCCAGACCCTAACCCGCTCCTTTCTCCCCGCCGCGTCCATGCAGAACTCCGCCGTTCCTGGGA GGGGAAGCCCCGCGAGGCGTCGGGAGAGGCACGTCCTCCGTGAGCAAAGAGCTCCTCCGAGCGCGCGGCGGGGACGCTGGGCCCGACA GGGGACCGCGGGGGCAGGGCGGAGAGGACCCGCCCTCGAGTCGGCCCAGCCCTAACACTCAGGAC
40	SIX2	AGGGAATCGGGCTGACCAGTCCTAAGGTCCCACGCTCCCCTGACCTCAGGGCCCAGAGCCTCGCATTACCCCGAGCAGTGCGTTGGTTA CTCTCCCTGGAAAGCCGCCCCCGCCGGGGCAAGTGGGAGTTGCTGCACTGCGGTCTTTGGAGGCCTAGGTCGCCCAGAGTAGGCGGAG CCCTGTATCCCTCCTCCTGGAGCCGGCCTGCGGTGAGGTCGGTACCCAGTACTTAGGGAGGGAGGACGCGCTTGGTGCTCAGGGTAGGCTG GGCCGCTGCTAGCTCTTGATTTAGTCTCATGTCCGCCTTTGTGCCGGCCTCTCCGATTTGTGGGTCCTTCCAAGAAAGAGTCCTCTAGGG CAGCTAGGGTCGTCTCTCTTGGGTCTGGCGAGGCGCGGCAGGCCTTCTTCGGACCTATCCCCAGAGGTGTAACGGAGACTTTCTCCACTGCAG GGCGGCCTGGGGCGGGCATCTGCCAGGGCGAGGGAGCTGCCCTGCCGCCGAGATTGTGGGGAAACGGCGTGGAAGACACCCCATCGGA GGGCACCCAATCTGCCTCTGCACTCGATTCCATCCTGCAACCCAGGAGAAACCATTTCCGAGTTCCAGCCGCAGAGGCACCCGCGGAGT TGCCAAAAGAGACTCCCCGCGAGGTCGCTCGGAACCTTGACCCTGACACCTGGACCGCGAGGTCTTTCAGGACCAGTCTCGGCTCGGTAGC CTGGTCCCCCGACACCACCGCGACCAGGAGTTCCTTCTTCCCTTCCCTGCTCACCAGCCGGCCGGCAGCGGCTCCAGGAAGGAGCACCA ACCCGCGCTGGGGGCGGAGGTTCAGGCGGCAGGAATGGAGAGGCTGATCCTCCTCTAGCCCCGGCGCATTCACTTAGGTGCGGGAGCC CTGAGGTTCAGCCTGACTTTC
41	SIX2	CACTACGGATCTGCCTGGACTGGTTCAGATGCGTCGTTTAAAGGGGGGGGGCTGGCACTCCAGAGAGGAGGGGGCGCTGCAGGTTAATT GATAGCCACGGAAGCACCTAGGCGCCCCATGCGCGGAGCCGGAGCCGCCAGCTCAGTCTGACCCCTGTCTTTTCTCTCCTCTTCCCTCT CCCACCCCTCACTCCGGGAAAGCGAGGGCCGAGGTAGGGGCAGATAGATCACCAGACAGGCGGAGAAGGACAGGAGTACAGATGGAG GGACCAGGACACAGAATGCAAAAGACTGGCAGGTGAGAAGAAGGGAGAAACAGAGGGAGAGAAAGGGAGAAACAGAGCAGAGGCGG CCGCCGGCCCGGCCGCCCTGAGTCCGATTTCCCTCCTTCCCTGACCCTTCAGTTTCACTGCAAATCCACAGAAGCAGGTTTGCGAGCTCG AATACCTTTGCTCCACTGCCACACGCAGCACCGGGACTGGGCGTCTGGAGCTTAAGTCTGGGGGGTCTGAGCCTGGGACCGGCAAATCCG CGCAGCGCATCGCGCCCAGTCTCGGAGACTGCAACCACCACCGCCAAGGAGTACGCGCGGCAGGAAACTTCTGCGGCCCAATTTCTTCCCCA GCTTTGGCATCTCCGAAGGCACGTACCCGCCCTCGGCACACAAGCTCTCTCGTCTTCCACTTCGACCTCGAGGTGGAGAAAGAGGCTGGCA AGGGCTGTGCGCGTCGCTGGTGTGGGGAGGGCAGCAGGCTGCCCCCTCCCCGCTTCTGCAGCGAGTTTTCCCAGCCAGGAAAAGGGAGG GAGCTGTTTCAGGAATTTCAGTGCCTTCACCTAGCGACTGACACAAGTCGTGTGTATAGGAAG
42	SOX14	GGAGCCTGAAGTCAGAAAAGATGGGGCCTCGTTACTCACTTTCTAGCCCAGCCCCTGGCCCTGGGTCCCGCAGAGCCGTCATCGCAGGC TCCTGCCCAGCCTCTGGGGTCGGGTGAGCAAGGTGTTCTCTTCGGAAGCGGGAAGGGCTGCGGGTCGGGACGTCCCTTGGCTGCCAC CCCTGATTCTGCATCCTTTTCGCTCGAATCCCTGCGCTAGGCATCCTCCCCCGATCCCCCAAAGCCCAAGCACTGGGTCTGGGTTGAGGA AGGGAACGGGTGCCCAGGCCGGACAGAGGCTGAAAGGAGAGGCCTCAAGGTTCCTCTTTGCTACAAAGTGGAGAAGTTGCTCTACTCTGGA GGGCAGTGGCCTTTTCCAAACTTTTCCACTTAGGTCCGTAAGAAAAGCAATTCATACACGATCAGCGCTTTCGGTGCGAGGATGGAAAGAA ACTTC
43	TLX3	TTTTCCTGTTACAGAGCTGAGCCCACTCATGTGGTGCCAAGTAGCGACTATCTCTCGGCCACCTCCACCCAGAGCAATGTGGGCGCCCCC AGCGGGTGGGAGCGATTGCCGAGCGGCGCAAGGGCGTTTAACGCCTAACCCCCTCCTCCTGGGTTGCCAAGCCGCTAGGTCGCCGTTT CCAACGTGGCTGCGCGGGACTGAAGTCCGACGACTCCTCGTCCTCAGTAGGAGACACACCTCCCACTGCCCCCAGCCACGCGAGCTATG GGCAGAATCGGGGCAACGGTAATATCTGGATGGGGCAGGCTCCCCCTGAGGCTGTGCTTAAGAAAAAAGGAATCTGGAGTAGCCTGAGGG GCCCCACGAGGGGGCCTCCTTTGCGATCGTCTCCCAGCCTTAGGCCAAGGCTACGGAGGCAGGCGGCCGAGTGTTGGCGCCCAGCCC

SEQ ID NO	GENE NAME	SEQUENCE
		GGCCGAGGACTGGATGGAGGACGAGAAGCAGCCTGCCTCTGGGCGACAGCTGCGGACGCAGCCTCGCCGCCTCGCCGCCTCAGCCTC GGTCCCCAGCGTCTCTAAAGCCCGCGCCCCATTTTACAGATGCAGGGGCAGGGGAGACAAGAGGCATCTCCGGGGGCCGAGTAGAATGATGGC GCGGGTTCTCCCCGGCGCCCCTGATTTCGAGGCTGCGCCCCGGGGCCCTACATGCAGGCGGGGAGGCCTGGGCCGAAGGCGTCTGCAAGG AGGGGCGAGTCTGCCCGGTCCGGGCAGGGAGTGAGGCCACAGTCAGTTCTCCCTAGGAGGCCCGCGCAGCGGGTAGGGTATGGGACTG GGGGACGCAACGGGGGACCTGGCCGAATCAGAGCCCTCAGCAGAGAACGCCGAAAACTCTGGGGCCGGCCGCTCGCTTCCCGCTAGTG GGAATGGTTTTCCGGTCATCCGTTCCCAGTCCAGCCCCCGGGTAGGGAGCTCTGATTTGCAATGCACAGCACTTGCGAGGTTCGAATGCCC CCGCAATTTGCAGATGGAAATACTAAGCCTAGGCCGGGGCGTGGTGGCTCAAGCCTATCATCTCAGCCCTTTGGGAGGCCAAGCCGGGAG GATTGTTTGAGCCCAAGAATTCAAAACCAGCCTGAGCAACATAGCGACCCCGTCTCTACAAAATAAAATAAAATTATCCGGGCGTGG TGGCACGCGCCTGTGGTTCCAGCTACTCCGGAGGTGGGAGGATCGCTTGAGTCCGGGAGGTCGAGGCTACAGTGAGCCGTGA TCGCACCACTGCACTCCAGCCTGGGCGACAGAGTGAGACCTTGTCTCAAAAAAGGAAAAAAGAAAAAGAAAGTAAGCTTCAAAGAAGCT CTGATAATAGTTCTGGGTCGTGCAGCGGTGGCGGCCCCGCGCTCTCGCCCCTAAAGCAAGCGCTCTTTGTACTGGGTGGAGGAGCTTTG AGTAGTGAGGGTGGAGATGCAGCTTCGGGGTGGCGCAGCCACCCTGACACTAGGCCCGGGGTCGCAGTGGGACAGAAGAGTCTGCCG CTCTGACTTGGGCTCTGAGTTCCAAGGGGCGCCCGGCACTTCTAGCCTCCCAGGCTTGCGCGCTGGCGCCTTTGCCATCCGTGCCGAAGT GGGGAGACCTAGCCGCGACCACCACGAGCGGCAGCGGTGACACCCAGAGGTCCCACCGGGCCCCCTGGGCAGGGTAACCTTAGCCTGTC CGCTTCGGCAGCTTTGCGAAGAGTGGGCGCGCAGCTAGGGCTGAGGCTCTTGCGGACCTGCGGTCGAAGCAGGCGGCTGAGCCAGTTCG ATCGCCAAGGCCTGGGCTGCCGACAGTGGTGCCGCGCTCTGTTCCGCCGGGCCGGGCCAGGCGCTCTGGAATAGCGATGGGGGACA CGGCCTCCAACTTTCTGCAGAGACCATCGGGCAGCTCCGGGCCTAAGCAGCGACCTCACCGAAGGTTCCTGGGAACCTTTGCCAAAATC CCAGCCTCTGCCTCGGTCCAGCTAAACCGTGTGTAAACAAGTGCACCAAG
44	FOXP4	ATAAAGGACCGGGTAATTTCGCGGAATGCGGATTTTGAGACAGGCCCAGACGGCGGCGGATTCCCTGTGTCCCCCAACTGGGGCGATCT CGTGAACACACCTGCGTCCCACCCCGATCCTAGGTTGGGGGGAAAGGGTATGGGAACCCTGAGCCCAGAGCGCGCCCCGCTCTTTCCTT TGCTCCCCGGCTTCCCTGGCCAGCCCCCTCCCGGCTGGTTTCCTCGCTCACTCGGCGCCTGGCGTTTCGGGCGTCTGGAGATCACCGC GTGTCTGGCACCCCCAACGTCTAGTCTCCCCGCAGGTTGACCGCGGCGCCTGGAGCCGGGAATAGGGGTGGGGAGTCCGGAGAACCAAA CCCGAGCCTGAAGTTGCCATTCGGGTGACTCCCGAGAAAGCCCGGGAGCATTTTGGCCAATGCGGGTTTTTACCTGAACTTCAGCATCTT CACC
45	FOXP4	AATTGGAAAACCCTGGTATTGTGCCTGTTTGGGGGAAGAAAACGTCAATAAAAATTAATTGATGAGTTGGCAGGGCGGGGCGGTGCGGGTT CGCGGCGAGGCGCAGGGTGTCATGGCAAATGTTACGGCTCAGATTAAGCGATTGTTAATTAAAAAGCGACGGTAATTAATACTCGCTACG CCATATGGGCCCGTGAAAAGGCACAAAAGGTTTCTCCGCATGTGGGGTTCCCCTTCTCTTTTCTCCTTCACAAAAGCACCCCCAGCCCGT GGGTCCCCCCCTTTGGCCCCAAGGTAGGTGGAACTCGTCACTTCCGGCCAGGGAGGGGATGGGGCGGTCTCCGGCGAGTTCCAAGGGC GTCCCTCGTTGCGCACTCGCCCCGCCCAGGTTCTTTGAA
46	chr7 group-00267	GGGAAGCGATCGTCTCCTCTGTCAACTCGCGCCTGGGCACTTAGCCCCTCCCGTTTCAGGGCGCCGCCTCCCCGGATGGCAAACACTAT AAAGTGGCGGCGAATAAGGTTCCTCCTGCTGCTCTCGGTTTAGTCCAAGATCAGCGATATCACGCGTCCCCCCGGAGCATCGCGTGCAGG AGCCATGGCGCGGGAGCTATACCACGAAGAGTTCGCCCGGGCGGGCAAGCAGGCGGGGGCTGCAGGTCTGGAGGATTGAGAAGCTGGA GCTGGTGCCCGTGCCCCAGAGCGCTCACGGCGACTTCTACGTCGGGGATGCCTACCTGGTGCTGCACACGGCCAAGACGAGCCGAGGC TTCACCTACCACCTGCACTTCTGGCTCGGTAAGGGACGGCGGCGGGACCCCGACGCACCAAGGCCGGCGAGGGGAGGGCGTAG GGGTCTGAGATTTGCAGGCGTGGGAGTAAAGGGGACCGCAAACTGAGCTAG
47	NPY	CTCAGGGGCGGGAAGTGGCGGGTGGGAGTCACCCAAGCGTGACTGCCCGAGGCCCCTCCTGCCGCGGCGAGGAAGCTCCATAAAAGC CCTGTCGCGACCCGCTCTCTGCACCCCATCCGCTGGCTCTCACCCCTCGGAGACGCTCGCCCGACAGCATAGTACTTGCCGCCCAGCCA

SEQ ID NO	GENE NAME	SEQUENCE
		CGCCCGCGCCGCCAGCCACCGTGAGTGCTACGACCCGTCTGTCTAGGGGTGGGAGCGAACGGGGCGCCCCGCGAACTTGCTAGAGACGC AGCCTCCCCGCTCTGTGGAGCCCTGGGGCCCTGGGATGATCGCGCTCCCACTCCCCAGCGGACTATGCCGGCTCCGCGCCCCCGACGCCGGA CCAGCCCTCTTGGCGGCTAAATTCCACTTGTTCCTCTGCTCCCCTCTGATTGTCCACGGCCCTTCTCCCGGGCCCTTCCCGCTGGGCGGT TCTTCTGAGTTACCTTTTAGCAGATATGGAGGGAGAACCCGGGACCGCTATCCCAAGGCAGCTGGCGGTCTCCCTGCGGGTCGCCGCCT TGAGGCCCAGGAAGCGGTGCGCGGTAGGAAGGTTTCCCCGGCAGCCATCGAGTGAGGAATCCCTGGAGCTCTAGAGCCCCGCGCCC TGCCACCTCCCTGGATTCTTTGGGCTCCAAATCTCTTTGGAGCAATTCTGGCCCAGGGAGCAATTCTCTCTTTCCCCTTCCCCACCGCAGTCGT CACCCCGAGGGTGATCTCTGCTGTCAGCGTTGATCCCCCTGAAGCTAGGCAGACCAGAAGTAACAGAGAAGAAACTTTTCTTCCCAGACAAG AGTTTGGGCAAGAAGGGAGAAAAGTGACCCAGCAGGAAGAACTTCCAATTCGGTTTTGAATGCTAAACTGGCGGGGCCCCCACCTTGCAC TCTCGCCGCGCGCTTCTTGGTCCCTGAGACTTCGAACGAAGTTTTCAGGTGGAGCAGAGGGGCAGGTCCCGGACCGGAC GGCGCCCGGAGCCCGCAAGGTGGTGCTAGCCACTCCTGGGTTCTCTCTGCGGGACTGGGACGAGCGGATTGGGGGTCGCGTGTGG TAGCAGGAGGAGGAGCGCGGGGGGCAGAGGAGGGAGGTGCTGCGCGTGGGTGCTCTGAATCCCCAAGCCCGTCCGTTGAGCCTTCTG TGCCTGCAGATGCTAGGTAACAAGCGACTGGGGCTGTCCGGACTGACCCTCGCCCTGTCCCTCGTGTGCCTGGGTGCGCTGGCCG AGGCGTACCCCTCCAAGCCGGACAACCCGGGCGAGGACGCACCAG
48	SHH	TGGAGAACCTTGGGCTCTGTGGCCTCAAAGGTAGGGGTGATTTCGAGGGGCCGGCACCTCACAGGGCAGGTTCCACCGCGGAAACGCA GTCATCGCCCAGCGACCCTGCTCCTGGCCCTCAGCCTCCCCCCAGGTTTCTTTTTCTCTTGAATCAAGCCGAGGTGCGCCAATGGCCTTC CTTGGGTCGGATCCGGGGGGCCAGGGCCAGCTTACCTGCTTTCACCGAGCAGTGGATATGTGCCTTGGACTCGTAGTACACCCAGTCGA AGCCGGCCTCCACCGCCAGGCGGGCCAGCATGCCGTACTTGCTGCGGTCGCGGTCAGACGTGGTGATGTCCACTGCGCGGCCCTCGTA GTGCAGAGACTCCTCTGAGTGGTGGCCATCTTCGTCCCAGCCCTCGGTCACCCGCAGTTTCACTCCTGGCCACTGGTTCATCACCGAGAT GGCCAAAGCGTTCAACTTGTCCTTACACCTCTGCGAAGACAAGGGGACCCCCACCGACGGACACGTTAGCCTGGGCAACCGCCACCCCT CCCGGCCCCTCCATCAGCCT
49	OSR2	TCTCACGACCCATCCGTTAACCCACCGTTCCCAGGAGCTCCGAGGCGCAGCGGCGACAGAGGTTCGCCCCGGCCTGCTAGCATTGGCAT TGCGGTTGACTGAGCTTCGCCTAACAGGCTTGGGGAGGGGTGGGCTGGGCTGGGCTGGGCTGCCCGGCTGTCCCGCC TTTCGTTTTCCTGGGACCGAGGAGTCTTCCGCTCCGTATCTGCCTAGAGTCTGAATCCGACTTTCTTTCCTTTGGGCACGCGCTCGCCAGT GGAGCACTTCTTGTTCTGGCCCCGGGCTGATCTGCACGCGGACTTGAGCAGGTGCCAAGGTGCCACGCAGTCCCCTCACGGCTTTCGGG GGGTCTTGGAGTCGGGTGGGGAGGGAGACTTAGGTGTGGTAACCTGCGCAGGTGCCAAAGGGCAGAAGGAGCAGCCTTGGATTATAGT CACGGTCTCTCCCCTCTCTTCCCTGCCATTTTTAGGGCTTTCTCTACGTGCTGTTGTCTCACTGGGTTTTTGTCGGAGCCCCCACGCCCTCCG GCCTCTGATTCCTGGAAGAAAGGGTTGGTCCCCCTCAGCACCCCCCAGCATCCCGGAAAATGGGGAGCAAGGCTCTGCCAGCGCCCATCCC GCTCCACCCCGTCGCTGCAGCTCACCAATTACTCCTTCCTGCAGGCCGTGAACACCTTCCCGGCCACGGTGGACCACACCTGCAGGGCCTGT ACGGTCTCAGCGCGGTACAGACCATGCACATGAACCACTGGACGCTGGGGTATCCCAAT
50	GLIS3	TGGTTTCCTTTCGCTTCTCGCCTCCCAAACACCTCCAGCAAGTCGGAGAGGCGCGAACGCGGAGCCAGAAACCCTTCCCCAAAGTTTCTCC CGCCAGGTACCTAATTGAATCATCCATAGGATGACAAATCAGCCAGGGCCAAGATTTCCAGACACTTGAGTGACTTCCCGGTCCCCGAGG TGACTTGTCAGCTCCAGTGAGTAACTTGGAACTGTCGCTCGGGGCAAGGTGTGTGTCTAGGAGAGAGCCGCGGCGGCTCACTCACGCTTTC CAGAGAGCGACCCGGGCCGACTTCAAATACACACAGGGTCATTTATAGGGACTGGAGCCGCGCGCAGGACAACGTCTCCGAGACTGAG ACATTTTCCAAACAGTGCTGACATTTTGTCGGGCCCCATAAAAAATGTAAACGCGAGGTGACGAACCCGGCGGGGAGGGTTCGTGTCTGG CTGTGTCTGCGTCCTGGCGGCGTGGGAGGTTTATAGTTCCAGACCTGGCGGCTGCGGATCGCCGGGCCGGTACCCGCGAGGAGTGTAGG TACCCTCAGCCCGACCACCTCCCGCAATCATGGGGACACCGGCTTGGATGAGACACAGGCGTGGAAAACAGCCTTCGTGAAACTCCACA AACACGTGGAACTTGAAAAGACAACTACAGCCCCCGCGTGTGCGCGAGAGACCTCACGTCACCCCATCAGTTCCCACTTCGCCAAAGTTTC CCTTCAGTGGGGACTCCCAGAGTGGTGCGCCCCATGCCCCGTGCGTCCTGTAACGGTGCCCTGATTGTGTACCCCTCTGCCCCGCTCTACTTG

SEQ ID NO	GENE NAME	SEQUENCE
		AAATGAAAACACAAAAACTGTTCCGAATTAGCGCAACTTTAAAGCCCCGTTATCTGTCTTCTACACTGGGCGCTCTTAGGCCACTGACAGA AACATGGTTTGAACCCTAATTGTTGCTATCAGTCTCAGTCAGCGCAGGTCTCTCAGTGACCTGTGACGCCGGGAGTTGAGGTGCGCGTAT CCTTAAACCCCGCGCGAACGCCACCGGCTCAGCGTAGAAAACTATTTGTAATCCCTAGTTTGCGTCTCTGAGCTTTAACTCCCCCACACTCT CAAGCGCCCCGGTTTCTCCTCGTCTCTCGCCTGCGAGCAAAGTTCCTATGGCATCCACTTACCAGGTAACCGGGATTTCCACACAACAAAGCC CGGCGTGCGGGTCCCTTCCCCCGGCCGCGCGAGTGACAGCGGGCGGCCGGCTGGCGAGGAGTAACTTGGGGCTCCAGCCC CTTCAGAGCGCTCCGCGGGCTGTGCCTCCTTCGGAAATGAAAACCCCCATCCAAACGGGGGGGACGGAGCGCGGAAACCCGGCCCAAGT GCCGTGTGTGCGCGCGCGCTCTG
51	PRMT8	GAAAGCCATCCTTACCATTCCCCTCACCCTCCGCCCTCTGATCGCCCACCCGCCGAAAGGGTTTCTAAAAATAGCCCAGGGCTTCAAGGC CGCGCTTCTGTGAAGTGTGGAGCGAGCGGGCACGTAGCGGTCTCTGCCAGGGTGGCTGGAGCCCTGGAAGCGAGAAGGCGCTTCCTCCC TGCATTTCCACCTCACCCCACACCCCCGGCTCATTTTTTCTAAGAAAAAAGTTTTTTGCGGTTCCCTTTGCCTCCTACCCCCGCTGCCGCGCGGG GTCTGGGTGCAGACCCCTGCCAGGTTCCGCAGTGTGCAGCGGCGGCTGCTGCGCTCTCCCAGCCTCGGCGAGGGTTAAAGGCGTCCGG AGCAGGCAGAGCGCCGCGCGCCAGTCTATTTTTACTTGCTTCCCCCGCGCTCCGCGCTCCCCCTTCTCAGCAGTTGCACATGCCAGCT CTGCTGAAGGCATCAATGAAAACAGCAGTAG
52	TBX3	ATCGAAAATGTCGACATCTTGCTAATGGTCTGCAAACTTCCGCCAATTATGACTGACCTCCCAGACTCGGCCCCAGGAGGCTCGTATTAGG CAGGGAGGCCGCCCGTAATTCTGGGATCAAAAGCGGGAAGGTGCGAACTCCTCTTTGTCTCTGCGTGCCCGGCGCGCCCCCCTCCCGGT GGGTGATAAACCCACTCTGGCGCCGGCCATGCGCTGGGTGATTAATTTGCGAACACAAACAAAAGCGGCCTGGTGGCCACTGCATTCGGGT TAAACATTGGCCAGCGTGTTCCGAAGGCTTGT
53	chr12 group-00801	ATCAACATCGTGGCTTTGGCTTTTCCATCATGGTGAGTGAATCACGGCCAGAGGCAGCCTGGGAGGAGAGACCCGGGCGGCTTTGAGC CCCTGCAGGGGAGTCCGCGCGCTCTCTGCGGCTCCCTTCCTCACGGCCCGGCCCGCGCTAGGTGTTCTTTGTCCTCGCACCTCCTCCTC ACCTTTCTCGGGCTCTCAGAGCTCTCCCCGCAATCATCAGCACCTCCTCTGCACTCCTCGTGGTACTCAGAGCCCTGATCAAGCTTCCCC CAGGCTAGCTTTCCTCTTCTTTCCAGCTCCCAGGGTGCGTTTCCTCTCCAACCCGGGGAAGTTCTTCCGTGGACTTTGCTGACTCCTCTGA CCTTCCTAGGCACTTGCCCGGGGCTTCTCAACCCTCTTTTCTAGAGCCCCAGTGCGCGCCACCCTAGCGAGCGCAGTAAGCTCATACCCC GAGCATGCAGGCTCTACGTTCCTTTCCCTGCCGCTCCGGGGGCTCCTGCTCTCTCAGCGCCCAGGACTGTCTCTATCTCAGCCTGTGCTC CCTTCTCTCTTTGCTGCGCCCAAGGGCACCGCTTCCGCCACTCTCCGGGGGGGTCCCCAGGCGATTCCTGATGCCCCCTCCTTGATCCCG TTTCCGCGCTTTGGCACGGCACGCTCTGTCCAGGCAACAGTTTCCTCTCGCTTCTTCCTACACCCAACTTCCTCTCCTTGCCTCCCTCCGG CGCCCCCTTTTTAACGCGCCCGAGGCTGGCTCACACCCACTACCTCTTTAGGCCTTTCTTAGGCTCCCCCCTCACCAGCAA AGTGGGTGCGCCTCTCTTACTCTTTCTACCCAGCGCGTCGTAGTTCCTCCCCGTTTGCTGCGCACTGGCCCTAACCTCTCTTCTCTTGGTG TCCCCCAGAGCTCCCAGGCGCCCCTCCACCGCTCTGTCCTGCGCCCGGGGCTCTCCCGGGAATGAACTAGGGGATTCCACGCAACGTG CGGCTCCGCCCCGCCCTCTGCGCTCAGACCTCCCGAGCTGCCCGCCTCTCTAGGAGTGGCCGCTGGGGCCTCTAGTCCGCCCTTCCGGA GCTCAGCTCCCTAGCCCTCTTCAACCCTGGTAGGAACACCCGAGCGAACCCCACCAGGAGGGCGACGAGCGCCTGCTAGGCCCTCGCC TTATTGACTGCAGCAGCTGGCCCGGGGGTGGCGGCGGGGTGAGGTTCGTACCGGCACTGTCCCGGGACAACCCTTGCAGTTGC
54	PAX9	ACAAATAAAACACCCTCTAGCTTCCCCTAGACTTTGTTTAACTGGCCGGGTCTCCAGAAGGAACGCTGGGGATGGGTGGAGAGAG GGAGCGGCTCAAGGACTTTAGTGAGGAGCAGGCGAGAAGGAGCACGTTCAGGCGTCAAGACCGATTTCTCCCCCTGCTTCGGGAGACTT TTGAACGCTCGGAGAGGCCCGGCATCTCACCACTTTACTTGGCCGTAGGGGCCTCCGGCACGGCAGGAATGAGGGAGGGGGTCCGATT GGACAGTGACGGTTTGGGCCGTTCGGCTATGTTCAGGGACCATATGGTTTGGGGACAGCCCCAGTAGTTAGTAGGGGACGGGTGCGTT CGCCCAGTCCCCGGATGCGTAGGGAGGCCCAGTGGCAGGCAGCTGTCCCAAGCAGCGGGTGCGCGTCCCTGCGCGCTGTGTGTTCATT TTGCAGAGCCAGCCTTCGGGGAGGTGAACCAGCTGGGAGGAGTGTTCGTGAACGGGAGGCCGCTGCCCAACGCCATCCGGCTTCGCAT

SEQ ID NO	GENE NAME	SEQUENCE
		CGTGGAACTGGCCCAACTGGGCATCCGACCGTGTGACATCAGCCGCCAGCTACGGGTCTCGCACGGCTGCGTCAGCAAGATCCTGGCG CGATACAACGAGACGGGCTCGATCTTGCCAGGAGCCATCGGGGGCAGCAAGCCCCCGGGTCACTACCCCCACCGTGGTGAAACACATCC GGACCTACAAGCAGAGAGACCCCGGCATCTTCGCCTGGGAGATCCGGGACCGCCTGCTGGCGGACGGCGTGTGCGACAAGTACAATGT GCCCTCCGTGAGCTCCATCAGCCGCATTCTGCGCAACAAGATCGGCAACTTGGCCCAGCAGGGTCATTACGACTCATACAAGCAGCACC AGCCGACGCCGCAGCCAGCGCTGCCCTACAACCACATCTACTCGTACCCCAGCCCTATCACGGCGGCCGGCCGCCAAGGTGCCCCACGCC ACCCGGGGTGC
55	SIX1	AGGAGGCGCAACGCGCTGCCAGGGCGGCTTTATCCTGCCGCCACAGGGCGGGGACCAGCCCGGCAGCCGGGTGTCCAGCGCCGCTCA CGTGCCTCGCCTGGAGCTTAGCTCTCAGACTCCGAAGAGGGCGACTGAGACTTGGGCCTGGGAGTTGGCTTCGGGGTACCCAAGGCGA CGACAGCTGAGTTGTACCACGAAGCTCAGGCCGAGGCCTCCTCCCTTGTCTGGCCTTCGAATCCATACTGGCAGCCTCTCCTCTCAGGCA CTCCGCGGGCCGGGCCACTAGGCCCCCTGCTCCTGGAGCTGCGCTATGATCCGGGTCTTGAGATGCGCGCGATTCTCTCTGAACCGGT GGAGAGGAGGCTCTGCCCCCGCGCGGAGGACAGCGGCGCCCGAGCTTCCCGCGCCTCTCCAGGGCCCAATGGCAAGAACAGCCT CCGAAGTGCGCGGATGACAGGAAAAGATCTTCAGTTCTTCTGCCGCTAGAGAAGTGCGGGATACAAGCCTCTATTGGATCCACAACCTGG AGTCCTGCCTTCGGA
56	ISL2	ATCTGCGTGCCCTTTTCTGGGCGAGCCCTGGGAGATCCAGGGAGAACTGGGCGCTCCAGATGGTGTATGTCTGTACCTTCACAGCAAGG CTTCCCTTGGATTTGAGGCTTCCTATTTTGTCTGGGATCGGGGTTTCTCCTTGTCCCAGTGGCAGCCCCGCGTTGCGGGTTCCGGGCGCT GCGCGGAGCCCAAGGCTGCATGGCAGTGTGCAGCGCCCGCCAGTCGGGCTGGTGGGTTGTGCACTCCGTCGGCAGCTGCAGAAAGGT GGGAGTGCAGGTCTTGCCTTTCCTCACCGGGCGGTTGGCTTCCAGCACCGAGGCTGACCTATCGTGGCAAGTTTGCGGCCCCCGCAGAT CCCCAGTGGAGAAAGAGGGCTCTTCCGATGCGATCGAGTGTGCGCCTCCCCCGCAAAGCAATGCAGACCCTAAATCACTCAAGGCCTGGA GCTCCAGTCTCAAAGGTGGCAGAAAAGGCCAGACCTAACTCGAGCACCTACTGCCTTCTGCTTGCCCCGCAGAGCCTTCAGGGACTGAC TGGGACGCCCCCTGGTGGCGGGCAGTCCCATCCGCCATGAGAACGCGCGTGCAGGGCAGCGCAGTGGAGGTGCAGACGTACCAGCCGCC GTGGAAGGCGCTCAGCGAGTTTGCCCTCCAGAGCGACCTGGACCAACCCGCCTTCCAACAGCTGGTGAGGCCCTGCCCTACCCGCCCC GACCTCGGGACTCTGCGGGTTGGGGATTTAGCCACTTAGCCTGGCAGAGAGGGGAGGGTGGCCTTGGGCTGAGGGGCTGGGTACA GCCCTAGGCGGTGGGGGAGGGGGAACAGTGGCGGGCTCTGAAACCTCACCTCGGCCCATTACGCGCCCTAAACCAGGTCTCCCTGGAT TAAAGTGCTCACAAGAGAGGTCGCAGGATTAACCAACCCGCTCCCCCGCCCTAATCCCCCCCCTCGTGCGCCTGGGGACCTGGCCTCCTT CTCCGCAGGGCTTGCTCTCAGCTGGCGGCCGGTCCCCAAGGGACACTTTCCGACTCGGAGCACGCGGCCCTGGAGCACCAGCTCGCGT GCCTCTTCACCTGCCTCTTCCGGTGTTTCCGCCGCCCCAGGTCTCCTTCTCCGAGTCCGGCTCCCTAGGCAACTCCTCCGGGCAGCGAC GTGACCTCCCTGTCCTCGCAGCTCCCGGACACCCCCAACAGTATGGTGCCGAGTCCCGTGGAGACGTGAGGGGGACCCCCTCCCTGCCA GCCCGCGGACCTCGCATGCTCCCTGCATGAGACTCACCCATGCTCAGGCCATTCCAGTTCCGAAAGCTCTCTCGCCTTCGTAATTATTCT ATTGTTATTTATGAGAGAGTACCGAGACACGGTCTGGACAGCCCAAGGCGCCAGGATGCAACCTGCTTTCACCAGACTGCAGACCCCT GCTCCGAGGACTCTTAGTTTTTCAAAACCAGAATCTGGGACTTACCAGGGTTAGCTCTGCCCTCTCCTCTCTACGTGGCCGCCGCT CTGTCTCTCCACGCCCCCACCTGTGT

SEQ ID NO	GENE NAME	SEQUENCE
57	DLX4	AGGTCTCTTCAGACTGCCCATTCTCCGGGCCTCGCTGAATGCGGGGGGCTCTATCCACACGCGCGCGGGGCCGAGCTCAGGCAGGCTGGG GCGAAGATCTGATTCTTTCCTTCCCGCCGCCAAACCGAATTAATCAGTTTCTTCAACCTGAGTTACTAAGAAAGAAAGGTCCTTCCAAATAA AACTGAAAATCACTGCGAATGACAATACTATACTACAAGTTCGTTTTGGGGCCGGTGGGTGGGATGGAGGAGAAAGGGCACGGATAATCC CGGAGGGCCCGCGGAGTGAGGAGGACTATGGTCGCGGTGGAATCTCTGTTCCGCTGGCACATCCGCGCAGGTGCGGCTCTGAGTGCTGG CTCGGGGTTACAGACCTCGGCATCCGGCTGCAGGGGCAGACAGAGACCTCCTCTGCTAGGGCGTGCGGTAGGCATCGTATGGAGCCCA GAGACTGCCGAGAGCACTGCGCACTCACCAAGTGTTAGGGGTGCCCGTGATAGACCGCCAGGGAAGGGGCTGGTTCGGAGGGAATTCC CGCTACCGGGAAGGTCGGAACTCGGGGTGATCAAACAAGGAATGCATCTCACCTCCGTGGGTGCTTGTGCTGCGCAAGGAATTATTACC GGAGCGGTTGCGATGGCCTTTGCCCGGCGACCCAAGAAGAGTAAGCAAACTACCGTCCACCCAGCGGATCAGGTCCAAT
58	CBX4	GATGTCCTGTTTCTAGCAGCCTCCAGAGCCAAGCTAGGCGAGAGGCGTAGGAGGCAGAGAGAGCGGGCGCGGCGAGGCCAGGGTCCGC CTGGGGGGCCTGAGGGGACTTCGTGGGGTCCCGGGAGTGGCCTAGAAACAGGGAGCTGGGAGGGCCGGGAAGAGCTTGAGGCTGAGCG GGGGACGAACGGGCAGCGCAAAGGGGAGATGAACGGAATGGCCGAGGAGCCACGCATTCGCCTTGTGTCCGCGGACCCTTGTTCCCGA CAGGCGACCAAGCCAAGGCCCTCCGGACTGACGCGGCCTGAGCAGCAGCGAGTGTGAAGTTTGGCACCTCCGGCGGCGAGACGGCGC GTTCTGGCGCGCCGGCTCCTGCGTCCGGCTGGTGGAGCTGCTGCGCCCTATGCGGCCTGCCGAGGGCGCCGCGAGGGCCCGCGAGCT CCGTGGGGGTCGGGGTGGGGGGACCCGGGAGCGCGACAGCGCGGCCCGAGGGGCAGGGGCGCGCCTGGCCTGGGGTGTGTC TGGGCCCCCGGCTCCGGGCTCTTGAAGGACCGCGAGCAGGAGGCTTGCGCAATCCCTTGGCTGAGCGTCCACGGAGAAAGAAAAAGAGC AAAAGCAGAGCGAGAGTGGAGCGAGGGATGGGGGGCGGGCAAAGAGCCATCCGGGTCTCCACCACCGCCCTGACACGCGACCCGGCTG TCTGTTGGGGACCGCACGGGGGCTCGGGCGAGCAGGGGAGGGAGCCTGCGCGGGGGCTCGTGTTCGCCCAGGAATCCCGGAGAA GCTCGAAGACGGTCTGGTGTTGAACGCACACGTGGACTCCATTTCATTACCACCTTGCAGCTCTTGCGCCACGGAGGCTGCTGCTGCCC GGCGGCTGCTACCCACCGAGACCCACGTGGCCCCTCCCCAGGGGGTGTAGGGGTGACGGTTGTCTTCTGGTGACAGCAGAGGTGTTGGG TTTGCGACTGATCTCTAACGAGCTTGAGGCGCAAACCTAGGATTCCCTGAGTGTTGGGGTGCGGCGGGGGGCAAGCAAGGTGGGACG ACGCCTGCCTGGTTTCCCTGACTAGTTGCGGGGGGTGGGGGCCGGCTCTCAGGGGCCACCAGAAGCTGGGTGGGTGTACAGGAAAATA TTTTTCTCCTGCCGTGTTTGGCTTTTTCCTGGCATTTTTGCCCAGGGCGAAGAACTGTCGCGCGGGGCAGCTCCACCGCGGAGGGAGAG GGGTCGCGAGGCTGGCGCGGGAAGCGCTGTAGGTGGCAGTCATCCGTCCACGCCGCACAGGCCGTCTGCGCCGTCGGACCATCGGGA GGTCTGCAGCAACTTTGTCCCGGCCAGTCCCCTTGTCCGGGAAGGGGCTGAGCTTCCCGACACTCTACCCTCCCCTCTTGAAAATCCCCC TGGAAAATCTGTTTGCAATGGGTGTTTCCGCGGCGTCCAGGTCTGGGCTGCCGGGGGGAGGCCGAGCGGCTGCTGCAGCCTCCCTGCTG CCAGGGGGCGTCGGACTCCGCTTCGCTCACTACGCCCAGGCCCCTCAGGGGCCCACGCTCAGGACTTCGGGGCCACACAGCAGGACCC GGTGCCCCGACGACGAGTTTGCGCAGGACCCGGGCTGGGCCAGCGCGGAGCTGGGGAGGAAGGGGCGGGGGTCGGTGCAGCGGAT CTTTTCTGTTGCTGCCTGTGCGGCGGCAGGAAGCGTCTTGAGGCTCCCCAAGACTACCTGAGGGGCCGCCCAAGCACTTCAGAAGCCCA AGGAGCCCCCGGCCACCCCCGCTCCTGGCCTTTTTGCCAACGACTTTGAAAGTGAAATGCACAAGCACCAGCAATTGACTTCCCTTCCGT GGTTATTTATTTTGTCTTTGTGGATGGTGGGCAGATGGGGAGAGAGGCCCCTACCTAACCTCGGTGGCTGGTCCCTAGACCACCCCCTGCC AGCCGGTGTGGGGAGGAGCTCAGGTCCGCGGGAGAGCGAATGGGCGCCAGGAGGTGGGACAGAATCCTGGGAAGGTACAGCGGACGC CCTGGAAGCTCCCCCTGATGCCCCAGAGGGCCCTTCCTGGGAAACCTCCCGGGGGGGGTGCCCCATACCATCCCACCCGGCTGTCTTGGC CCCTCCCAGGGAGCCGCAGGAGAAACTAGCCCTACACCTGGGATTCCCAGAGCCTTCTGCTGGGGCTCCTGCCCCCGACTTCGGATAAC CAGCTCCGCACAGGTCCCCGAGAAGGGCCGCTGGCCTGCTTATTTGATACTGCCCCCCTCCCAGACAGGGGCTGGTCGAGCCCCCTGGTTC TGCTGCCAGACTGAAGCCTTCCAGACGCCACCTCGGTTTGGGCCCCCAGGGCCCTCAGGGAGAGGAGGAGCTGCTATCTAG CTCAGCCACAGGCTCGCTCCTGGTGGGGGCCAGGCTGAAGGAGTGGACCCTGGAGAGGTCGGGAACCTTTTAACAGCCGTGGGCTGGA GGGTGGCTACTAAGTGTTCGGTCTGGGAAGAGGCATGACCCGCACCATCCCGGGGAAATAAACGACTTCTTAAGGGAATCTTCTCGCTGA GCGGGTGCTCTCTGGGCCAGGAGATTGCCACCGCCAGCCCACGGAACCCAGATTTGGGCTCTGCCTTGAGCGGGCCGCCTGTGGCTTCCC GGGTCGCTCCCCCGACTCAGAAAGCTCTCAAGTTGGTATCGTTTTCCCGGCCCTCGGAGGTGGATTGCAGATCACCGAGAGGGGATTTA

SEQ ID NO	GENE NAME	SEQUENCE
		CCAGTAACCACTACAGAATCTACCCGGGCTTTAACAAGCGCTCATTTCTCTCCCTTGTCCTTAGAAAAACTTCGCGCTGGCGTTGATCATAT CGTACTTGTAGCGGCAGCTTAGGGGCAGCGGAACTGGTGGGGTTGTGCGTGCAGGGGGAGGCTGTGAGGGAGCCCTGCACTCCGCCC CTCCACCCTTCTGGAGGAGTGGCTTTGTTTCTAAGGGTGCCCCCCCAACCCCCGGGTCCCCACTTCAATGTTTCTGCTCTTTGTCCCACC GCCCGTGAAAGCTCGGCTTTCATTTGGTCGGCGAAGCCTCCGACGCCCCCCGAGTCCCACCCTAGCGGGCCGCGCGGCACTGCAGCCGG GGGTTCCTGCGGACTGGCCCGACAGGGTGCGCGGACGGGGACGCGGGCCCCGAGCACCGCGACGCCAGGGTCCTTTGGCAGGGCCC AAGCACCCCT
59	EDG6	TGGCGGCCGGCGGGCACAGCCGGCTCATTGTTCTGCACTACAACCACTCGGGCCGGCTGGCCGGGGGGCCGGAGGATGGC GGCCTGGGGGCCCTGCGGGGGCTGTCGGTGGCCGCCAGCTGCCTGGTGGTGCTGGAGAACTTGCTGCTGGTGCTGGCGGCCATCACCAGC CACATGCGGTCGCGACGCTGGGTCTACTATTGCCTGGTGAACATCACGCTGAGTGACCTGCTCACGGGGCGCGGCCTACCTGGCCAACGT GCTGCTGTCGGGGGGCCCGCACCTTCCGTCTGGCGCCCGCCCAGTGGTTCCTACGGGAGGGCCTGCTCTTCACCGCCCTGGCCGCCTCC ACCTTCAGCCTGCTCTTCACTGCAGGGGAGCGCTTTGCCACCATGGTGCGGCCGGTGGCCGAGAGCGGGGCCACCAAGACCAGCCGCG TCTACGGCTTCATCGGCCTCTGCTGGCCGCGCTGCTGGGGATGCTGCCTTTGCTGGGCTGGAACTGCCTGTGCGCCTTTGAC CGCTGCTCCAGCCTTCTGCCCCTCTACTCCAAGCGCTACATCCTCTTCTGCCTGGTGATCTTCGCCGGCGTCCTGGCCACCATCATGGGC CTCTATGGGGCCATCTTCCGCCTGGTGCAGGCCAGCGGGCAGAAGGCCCCACGCCCAGCGGCCCGCCGCAAGGCCCGCCGCCTGCTG AAGACGGTGCTGATGATCCTGCTGGCCTTCGTGGTGTGCTGGGGCCCACTCTTCGGGCTGCTGCTGGCCGACGTCTTTGGCTCCAACCT CTGGGCCCAGGAGTACCTGCGGGGGCATGGACTGGATCCCTGGCCCTGGCCGTCCTCAACTCGGCGGTCAACCCCATCATCTACTCCTTCC GCAGCAGGGAGGTGTGCAGAGCCGTGCTCAGCTTCCTCTGCTGCGGGTGTCTCCGGCTGGGCATGCGAGGGCCCGGGGACTGCCTGG CCCGGGCCGTCGAGGCTCACTCCGGAGCTTCCACCACCGACAGCTCTCTGAGGCCAAGGGACAGCTTTC
60	chr13 group-00005	TAGTAAGGCACCGAGGGGTGGCTCCTCTCCCTGCAGCGGCTGTCGCTTACCATCCTGTAGACCGTGACCTCCTCACACAGCGCCAGGAC GAGGATCGCGGTGAGCCAGCAGGTGACTGCGATCCTGGAGCTGGTCGCAGCAGGCCATCCTGCACGCGGTGGAGGCGCCCCCTGCAG GCCGCAGCGCATCCCCAGCTTCTGGACGCACTGTGAGCGGTTATGCAGCAGCACGCTCATATGAGATGCCCCGCAGGGTGCTATGCAGG CCCACGTCCCCACAAAGCCCATGGCAGGCGCCCGGGTGCCGGAGCACGCACTTGGCCCCATGGATCTCTGTGCCCAGGGCTCAGCCAG GCATCTGGCCGCTAAAGGTTT
61	CRYL1	TCTCATCTGAGCGCTGTCTTTCACCAGAGCTCTGTAGGACTGAGGCAGTAGCGCTGGCCCGCCTGCGAGAGCCCGACCGTGGACGATGC GTCGCGCCCTTCCCATCGCGGCCTGGGCGGGCCCGCCTGCCCTCGGCTGAGCCCGGTTTCCCTACCCCGGGGCACCTCCCCTCGCCC GCACCCGGCCCCCAGTCCCTCCCAGGCTTGCGGGTAGAGCCTGTCTTTGCCCAGAAGGCCGTCTCCAAGCT
62	IL17D	CAGTCCCCGAGGCCCTCCCCGGTGACTCTAACCAGGGATTTCAGCGCGCGGCGGGCTGCCCCCAGGCGTGACCTCACCCGTGCTC TCTCCCTGCAGAATCTCCTACGACCCGGCGAGGTACCCCAGGTACCTGCCTGAAGCCTACTGCCTGTGCCGGGGCTGCCTGACCGGGCT GTTCGGCGAGGAGGACGTGCGCTTCCGCAGCGCCCCTGTCTACAT
63	IRS2	AGAGAGACATTTTCCACGGAGGCCGAGTTGTGGCGCTTGGGGTTGTGGGCGAAGGACGGGGACACGGGGGTGACCGTCGTGGTGGAG GAGAAGGTCTCGGAACTGTGGCGGCGGCCCCCCTGCGGGTCTGCGCGGATGACCTTGGCGCCGCGGTGGGGGTCCGGGGGCTG GCTGGCCTGCAGGAAGGCCTCGACTCCCGACACCTGCTCCATGAGGCTCAGCCTCTTCACGCCCCGACGTCGGGCTGGCCACGCGGGCA GCTTCTGGCTTCGGGGGGCGCGATAGGTTGCGGCGGGGTGGCGGCCACACCAAAAGCCATCTCGGTGTAGTCACCATTGTCCCCGGG TGTCCGAGGACAACGATGAGGCGGCGCCCTGGGCGGTGGCAACGGCCGAGGCGGGGGCAACGGCGGTACAGCTCCCCCGGG GCCGGCGGCGGTGGCGGCGGCTGCAGAGACGACGGGGACGCGGACGCGGGGGCAACGGCGGATACGGGGAGGAGGCCT CGGGGGACAGGAGGCCGTCCAAGGAGCCCACGGGGTGGCCGCTCGGGGCGCCCGGCTTAGGAGACTTGGGGGAGCTGAAGTCGAGG

SEQ ID NO	GENE NAME	SEQUENCE
		TTCATGTAGTCGGAGAGCGGAGACCGCTGCCGGCTGTCGCTGCTGGTGCCCGGGGTGCCTGAGCCCAGCGACGAGGCCGGGCTGCTG GCGGACAAGAGCGAGGAGGACGAGGCCGCCGACCGCAGCAGGGGAGGCCGGGGCGGCGACAGGCGGGCCCCGGGCTCGCCAAAGT CGATGTTGATGTACTCGCCGGGGCTCTTGGGCTCCGGTGGCAGTGGGTACTCGTGCATGCTGGGCAGGCTGGGCAGCCCCTCCAGGGA CAGGCGCGTGGGCCTCACCGCCCGGCCGCGCTGGCCCAAGAAGCCCTCCGGGCGGCCGCCGCTAGGCCGCACGGGCGAAGGCACTA CAGGGGTGAGGGGGGCTGCGTGGGGCCGGCCCCGAAGGCGCTGGCCTGGCGTGGCCTGAGGCTCCAGACGCTCC TCCTCCAGGATGCCGCCCCACGGGGGAGCTCATGAGCACGTACTGGTCGCTGTCCCCGCCACAGGTGTAGGGGGCCTTGTAGGAGCGGG GCAAGGAGCTGTAGCAGCAGCCGGGAACGCCCCTGAGCGGCTCCCCGCCGGGGTGCAGGGCTGCGGGAGAAGAAGTCGGGCGGGGTG CCCGTGGTGACCGCGTCGCTGGGGGACACGTTGAGGTAGTCCCCGTTGGGCAGCAGCTTGCCATCTGCATGCTCCATGGACAGCTTGG AACCGCACCACATGCGCATGTACCCACTGTCCTCGGGGGGAGCTCTCGGCGGGGCGAGCTGGCCTTGTAGCCGCCCCCGCTCGCCGGGAA TGTCCTGCCCGCCGCAGAGGTGGGTGCTGGCCCCCGCAGGCCCCGCAGAAGGCACGGGCGGCGCGGCCGGCGGCCCTGGGCT GCAAGATCTGCTTGGGGGCGGACACGCTGGCGGGGGCTCATGGGCATGTAGTCGTCGCTCCTGCAGCTGCCGCTCCCACTGCCCGCGAG GGCCGCGCCGGGCGTCATGGGCATGTCGTCTCTGCCCCCAGGTTGCTGCTGTGGGAGCCGATCTCGATGTCTCCG TAGTCCTCTGGGTAGGGGTGGTAGGCCACCTTGGGAGAGGACGCGGGGGCAGGACGGGCAGAGGCGGCCCGCGCTGCCCGGAGAAGGTG GCCCGCATCAGGGTGTATTCATCCAGCGAGGAGGGCTGGGGCACCGGCCGCTGCCGGGCTGGTCAGGGAGTAGGTC CTCTTGCGCAGCCCTCGGTCCAGGTCCTGGGCCGCGTCCCCCGAGACCCGGCGGTAGGAGCGGCCACAGTGGCTCAGGGGCCTGTCC ATGGTCATGTACCCGTAGAACTCACCGCCGCCGCCGTCTCGGGCCGGGGGCGTCTCCGCGATGGACTCGGGCGTGTTGCTTCGGT GGCTGCAGAAGGCGCGCAGGTCGCCTGGGCTGGAGCCGTACTCGTCCAGGGACATGAAGCCGGGGTCGCTGGGGGAGCCCCGAGGCG GAGGCGCTGCCGCTGGAGGGCCGCTGGGCCGGGGCCGTGGTGCAGCGGATGCGGCAGAGGCGGGTGCGGGCCGGCGGCGCGGGT AGGAGCCCGAGCCGTGGCCGCTGCTGGACGACAGGGAGC
64	chr13 group-00350	TAACCTAAAGAATGAAGTCATGCCCCGGCCTGCACCCGGGAAACTGCACACAGCGAAAGATCGCCACTGAGATAAAGAGCTGAAAGCTAT TCCCCAATTCAGCTGTTTCAGCCGTGCGGTCTCACAATGGGCTCACAGACGGCAGCATC
65	MCF2L	GTTTCCACAATCCACCTCGTAGCTGGGGCGTGCCGCTTGCCTCGGCTTGTCCCGGCAGAACACTCTTACCTTTAATGGCGACTGAAAAGT TGCCACGAGTTCCTGATCATTGTGGTAGGTGCTGCGTGAAGCTGAGACGTGCGTGAGCCACATCCCAGGGGGCTTTGAGCCCCCACCGC GGCGGCGGCTGAGGGGGAGGCTTGTCGTACTCGCACACGGAGGACACAGGGCTGCAGTGTTCACTCCAGGGCCTCTTATCATTGGGATCT GAGGAATTTTCCGAGAGGGAAGTGCGAATTAACAATGATGAAAGGTTTGTGAGTGAGTGACAGGCACGTTCTATTGAGCACTGCATGGGGC ATTATGTGCCACCAGAGACGGGGGCAGAGGTCAAGAGCCCTCGAGGGCTGGGAGAGTTCGGAGGATAGAAGTCATCAGAGCACACAATGAA GCCAGACCCTGCAGCCGCCTTCCCCTTCGGGGGCTTCCTTAGAATGCAGCATTGCCGGGGACTGAAGCGTCCCAGGTGAAGGGGGGCCG TCACGGTGTGTGGACGCCCCTCGGCTCAGCCCTCTAAGAGCTCGGCAGCCAGGATGGGCTCAAGGCATGAGCCCTCAAAGGAGGGTTA GGAAGGAGCGAGGGAGAAAAGATATGCTTGTGTGACGTCCCTGGCCGAAGTGAGAACAATTGTATCAGATAATGAGTCATGTCCCATTGAG GGGTGCCGACAAGGACTCGGGAGGAGGCCACGGAGCCCTGTACTGAGGAGACGCCCACAGGGAGCCTCGGGGGCCCAGCGTCCCGG GATCACTGGATGGTAAAGCCGCCCTGCCTGGCGT
66	F7	TCCAGCTGCAGCGAGGGCGGCCAGGCCCCCTTCTCCGACCTGCAGGGGTAGCGCGGCCTCGGCGCCGGAGACCCGCGCGCTGTCTGG GGCTGCGGTGGCGTGGGGAGGGCGCGGCCCCCGGACGCCCCGAGGAAGGGGCACCTCACCGCCCCCACCCAGAGCGCCTGGCCGTG CGGGCTGCAGAGGACCCCTCCGGGGGCAGAGGCAGGTTCCACGGAAGACCCCCGGCCCCGCTGGGGCTTCCCCCGGAGACTCCAGAG

SEQ ID NO	GENE NAME	SEQUENCE
67	chr18 group-00039	ACTTACTGCTTCCAAAAGCGCTGGGCACAGCCTTATATGACTGACCCCGCCCCCGAGTCCCAGGCCGCCCCATGCAACCGCCCAACCGC CCAACCGCCACTCCAAAGGTCACCAACCACTGCTCCAGGCCACGGGCTGCCTCTCCCCACGGCTCTAGGGCCCTTCCCCTCCACCGCAG GCTGAC
68	C18orf1	TGCCACACCCAGGTACCGCCCGCCGCGAGAGCCGGGCAGGTGGGCCGCGGATGCTCCCAGAGGCCGGCCCAGCAGAGCGATGG ACTTGGACAGGCTAAGATGGAAGTGACCTGAG
69	CD33L3	TCGCCAGCGCAGCGCTGGTCCATGCAGGTGCCACCCGAGGTGAGCGCGGAGGCAGGCGACGCGGCAGTGCTGCCCTGCACCTTCACG CACCCGCACCGCCACTACGACGGGCCGCTGACGGCCATCTGGCGCGCGGGCGAGCCCTATGCGGGGCCCGCAGGTGTTCCGCTGCGCT GCGGCGCGGGGCAGCGAGCTCTGCCAGACGGCGCTGAGCCTGCACGGCCGCTTCCGGCTGCTGGGCAACCCGCGCCGCAACGACCTC TCGCTGCGCGTCGAGCGCCTCGGCCCTGGCTGACGACCGCCGCGCTACTTCTGCCGCGTCGAGTTCGCCGGCGACGTCCATGACCGCTACG AGAGCCGCCACGGCGTCCGGCTGCACGTGACAGGCGAGGCGGCGTGGGAGCGGGTCCCCGGCCTCCCTTCCCGCCCTCCCGCCTGCC CCGCCCCAAGGGCTACGTGGGTGCCAGGCGCTGTGCTGAGCCAGGAAGGGCAACGAGACCCAGCCCTCTCCTCTACCCCAGGGATCTC ACACCTGGGGGTAGTTTAGGACCACCTGGGAGCTTGACACAAATGCAGAATCCAGGTCCCAGGAAGGGCTGAGGTGGGCCCGGGAATA GGCATTGCCGTGACTCTCGTAGAGTGACTGTCCCCAGTGGCTCTCAGACGAAGAGGCGAGAAAGACAAGTGAATGGCAATCCTAAATATG CCAAGAGGTGCAATGTGGTGTGTGCTACCAGCCCCGGAAAGACACTCGCAGCCCCCTCTACCCAGGGGTGCACAGACAGCCCACCAAGTAG TGCCTAGCACTTTGCCAGACCCTGATATACAAAGATGCCTGAACCAGGGTCCCGTCCCTAGAGCAGTGGCTCTCCACTCTAGCCCCCACC CTGCTCTGCGACAATAATGGCCACTTAGCATTTGCTAGGGAGCCCGGGACCTAGTCCAAGCACCCACAAGCATGAATTTGCCAAATCTTTTC AGCAACCTCTTAAGGCAACTGCTATCATGATCCTCACTTTACACATGGAGAAGCAGAAGCAGAGATGATAGAATCTTTCGCCCAAGGCCAC ATCTGTATTGGGACGGGGGCAGCCTGGCACCCAAGTGCCCATTCCTCCCTTCTGACCAGCCCCCCACCCCTCCGGCTCTGGCGTCCAAAG GGCTAAGGGGAGGGGTGCCCTTGTGACAGTCACCCGCCTTCTCCCCTGCAGCCGCGCCGGATCGTCAACATCTCGGTGCTGCCCAG TCCGGCTCACGCCTTCCGCGCTCTGCACTGCCGAAGGGGAGCCGCGCCCGCCCTCGCCTGGTCCGGCCCGGCCCTGGGCAACAG CTTGGCAGCCGTGCGGAGCCCGCGTGAGGGTCACGGCCACCTAGTGACCGCCGCAACTGCCCGCACTGACCCATGACGGCCGCTACACG TGTACGGCCGCCAACAGCCTGGGCCGCTCCGAGGCCAGCGTCTACCTGTTCCGCTTCCATGGCGCCACGGGCGGGCCTCGACGGTCGCCC TCCTGCTCGGCGCTCTCGGCTTCAAGGCGCT
70	TNFRSF11A	ATGAACTTCAAGGGCGACATCATCGTGGTCTACGGTCAGCCAGACCTCGCAGGAGGGCGCGGCGGCTGCGGAGCCCATGGGCCGCC CGGTGCAGGAGGAGACCCTGGCGCGCCGAGACTCCTTCGCGGGGAACGGCCCCGCGCTTCCCGGACCCGTGCGGCGGCCCCGAGGGG CTGCGGGAGCCGGAGAAGGCCTCGAGGCCGGTGCAGGAGCAAGGCGGGGCCAAGGCTTGAGCGCCCCCCCATGGCTGGGAGCCCCGAA GCTCGGAGC
71	ZNF236	TCAGTGTTATGTGGGGAGCGCTAGATCGTGCACACAGTAGGCGTCAGGAAGTGTTTTCCCCAGTAATTTATTCTCCATGGTACTTTGCTAA AGTCATGAAATAACTCAGATTTTGTTTTCCAAGGAAGGAGAAAGGCCCAGAATTTAAGAGCAGGCAGACACACAACCGGGCACCCCCAGA CCCTGGCCCTTCCAGCAGTCAGGAATTGACTTGCCTTCCAAAGCCCCCAGCCCGGAGCTTGAGGAACGGACTTTCCTGCGCAGGGGGATC GGGGCGCACTCG
72	chr18 group-00342	GTGGAAACACAACCTGCCTTCCATTGTCTGCGCCTCCAAAACACACCCCCCGCGCATCCGTGAAGCTGTGTGTTTCTGTGTTACTACAGG GGCCGGCTGTGGAAATCCCCACGCTCCAGACCGCGTGCCGGGGCAGGCCCCAGCC
73	OLIG2	TCCACACCTCGGGCAGTCACTAGGAAAAGGGTCGCCAACTGAAAGGCCTGCAGGAACCAGGATGATACCTGCGTCAGTCCCGCGGCTGC TGCGAGTGCGCGCTCTCCTGCCAGGGGGGACCTCAGACCCTCCTTTACAGCACACCGAGGGCCCTGCAGACACGCGAGCGGGCCTTCAG

SEQ ID NO	GENE NAME	SEQUENCE
		TTTGCAAACCCTGAAAGCGGGCGCGGTCCACCAGGACGATCTGGCAGGGCTCTGGGTGAGGAGGCCGCGTCTTTATTTGGGGTCCTCG
		GGCAGCCACGTTGCAGCTCTGGGGGAAGACTGCTTAAGGAACCCGCTCTGAACTGCGCGCTGGTGTCCTCTCCGGCCCTCGCTTCCCCG
		ACCCCGCACAGGCTAACGGGAGACGCGCAGGCCCACCCCACCGGCTGGAGACCCCGGCACGGCCCGCATCCGCCAGGATTGAAGCAG
		CTGGCTTGGACGCGCCGCAGTTTTCCTTTGGCGACATTGCAGCGTCGGTGCGGCCACAATCCGTCCACTGGTTGTGGGAACGGTTGGAGG
		TCCCCCAAGAAGGAGACACGCAGAGCTCTCCAGAACCGCCTACATGCGCATGGGGCCCAAACAGCCTCCCAAGGAGCACCCAGGTCCAT
		GCACCCGAGCCCCAAAATCACAGACCCGCTACGGGCTTTTGCACATCAGCTCCAAACACCTGAGTCCACGTGCACACAGGCTCTCGCACAGG
		GGACTCACGCACCTGAGTTCGCGCTCACAGATC
74	RUNX1	CTGCCCTCGCGGATCTCCCCCGGCCTCGCCGGCCTGTCCTCCCACCACCCTCTCCGGGCCAGTACCTTGAAAGCGATGGGCA
		GGGTCTTGTTGCAGCGCCAGTGCGTAGGCAGCACGGAGCAGAGGAAGTTGGGGGCTGTCGGTGCGCACCAGCTCGCCCGGGTGGTCGG
		CCAGCACCTCCACCATGCTGCGGTCGCCGCTCCTCAGCTTGCCGGCCAGGGCAGCGCCGGCGTCCGGGGGCGCCCAGCGGCAACGCCT
		CGCTCATCTTGCCTGGGCTCAGCGCGGTGGAAGGCGGCGTGAAGCGGCGCTCGTGCTGGCATCTACGGGGATACCGCATCACACAAGAAG
		CCGATTGAGTTAGGACCCTGCAAACAGCTCCTACCAGACGGCGACAGGGGGCGCGGATCTTCAGCAAGCAGCTCCCGGGAGACCAACATA
		CACGTTCAGGGGCCTTTATTACTGCGGGGGGGTGGGGGGGGCGGGGGGTGGTTAGGGGAGGAGACTAAGTTACTAACAGTCCAGG
		AGGGGAAAACGTTCTGGTTCTGCGGATCGGCCTCTGACCCAGGATGGGCTCCTAGCAACCGATTGCTTAGTGCATTAAAAAGTGGAGACT
		ATCTTCCACGAATCTTGCTTGCAGAGGTTAAGTTCTGTCTTTGGCTGTTAGAAAAGTTCCTGAAGGCAAAATTCTCATACACTTCCTAAAATA
		TTTATGCGAAGAGTAAAACGATCAGCAAACACATTATTTGGAAGTTCCAGTAGTTAATGCCTGTCAGTTTTTTGCAGGTGAGTTTTGTCTAA
		AGTCCCAACAGAACACAATTATCTCCCGTAACAAGGCCACTTTTATCATGCAAAACTGGCTTCAGTCCCGAAAAGCAAGAGCTGAGACTTC
		CAAAGGTAGTGCTACTAATGTATGTGCACGTATATATATAAATATATACATATGCTCTACTTCATAAAATATTTACAATACAATCTGTGGAGAATT
		TAAACACAACAGAAATCCATTAATGTACGCTGCAGATTTTTTTTAAGTAGCCTTGAAAATCAGCTTCAGTAGTTGGAGCAGTGCTGAGCTAGA
		AGTACTTGTCATGTTCTCTGTTCTCAATGAATTCTGTCAAAACGCTCAGTGCAGAAAATTCAGCGTTTCAGAGATCTTCAGCTAATCTTAA
		AACAACAATCATAAGAAGGCCCAGTCGATGACACTCAGGGTTCTACAGCTCTCCCACATCTGTGAACTCGGGTTTGGGGATGTTGGTTAA
		GTTTGTGGCTGGTCCTCTGGTTTGTTGGGAGTTGAGCAGCCGCAGAGTCACACACATGCAAACACGCACTCTTCGGAAGGCAGCCACTGT
		CTACATCAGCTGGGTGACTCAGCCCTGACTCGGGCAGCAGCGAGACGATACTCCTCCACCGTCGCCCAGCACCCGCCGGTTAGCTGCTC
		CGAGGCACGAACACCCACGAGCGCCGCGTAACCGCAGCAGGTGGAGCGGGCCTTGAGGGAGGGCTCCGCGGCGCAGATCGAAACAGA
		TCGGGCGGCTCGGGTTACACACGCACGCACATCCTGCCACGCACACTGCCACGCAACTTCACGGCTCGCCTCGGACCACAGA
		GCACTTTCTCCCCCTGTTGTAAAAGGAAAACAATTGGGGAAAAGTTCGCAGCCAGGAAAGAAGTTGAAAACATCCAGCCAAGAAGCCAGT
		TAATTCAAAAGGAAGAAAAGGGGAAAAACACAACAAAAAAAAGGAAGGTCCAACGCAGGCCAAGGAGAAGCAGCAGAGGTTGAC
		TTCCTTCTGGCGTCCCTAGGAGCCCCGGAAAGAAGTGCCTGGCGGCGCAGGGCCGGGCAGCGTGGTGCCCTGGCTGGGTCCGGCCGC
		GGGGCGCCCGTCCCGCGCCCGCTGGCTCTATGAATGAGAGTGCCTGGAAATGAACGTGCTTTTACTGTAAGCCCGGCCGGAGGA
		ATTCCATTCCCTCAGCTCGTTTGCATAGGGGCGGCCGGCCAATCACAGGCCTTTCCGGTATCAGCCAGGGCGCGGCTCGCCGCCG
		CCGGCTCCTGGAATTGGCCCGCGCCCCCGCCGCGCCGCGCTACTGTACGCAGCCCGGGCGGGGAGTCGGAGGCCACCCCC
		CGCGCCCCGCATCCAAGCCTGCATGCTGGCCCGGGGCCCCCGCCCGCGTGCGGACCCCTTTCCGCAGCCACACGCAGGCTTGTGCGGC
		TCCGCGAGTGGCCACGGTCCGGAGACCTGGAAAAAGAAAGCAGGCCCCCGCCGGCCCGAGGAGGACCCGGCCGCGCCGCACCCG
		GAGAGGCCCGGCCCCCGCGAGCCGCTGCAGGCAGGCGCAGTGGCCGCCACGAGGCTCCCGAACCGGGCTGCAGCCCGCGGAGACGGCCC
		CAGATCCTGCGCGGCCAGGCCAGGGCCAGGCCTCCGCTTCCAGGGCGGGGGGTGCGATTTGGCCGCGGGGCCCGGGGGAGCCACTCCG
		CGCTCCTGCACCGTCCGGCTGGCAGCTGCGGCGAAGCGGCGCTGATTCCTTGCATGAGGCCGGACGGCGTCCGCGCGTGCCGTTTGCT
		CTCAGCGTCTTCCCTTGGGTCGGTTTCTGTAATGGGTGTTTTTTTACCGCTGCGCCCGGGCCGCGGCTCGATCCCTCCGCGCGTCTCACTT
		GCTGCGTGCGTCAGCGGCCAGCGAAGAGTTTCCTAGTCAGGAAAGACCCCAAGAACGCGCGGCTGGAAGGAAAGTTGAAAGCAGCCAC
		GCGGCTTGCTCCCGGGCCTTGTAGCGCCGGCACCCGCAGCAGCCGGACAGCCTGCCCGGGCCCCCGCGTCTCCCCTCCGGCTCCCCGG

SEQ ID NO	GENE NAME	SEQUENCE
		AAGCGGCCCCCGCTCCTCTCCCCGCCCCCGTGCGCTCGAGCGGCCCCAGGTGCGGAACCCACCCCGGCTTCGCGTGCGGGCGGCCGC TTCCCCCCTGCGCCCGGTCCCCCGCGGTGCTGCGGGCATTTTTCGCGGAGCTCGGAGGGCCCCCGCCCCCGGTCCCGGCGTGCGCTGCCAACT CCGACCCCCGCCCCGGCGGGGCTCCCTCCCAGCGGAGGCTGCTCCCCGTCACCATGAGTCCCTCCACGCCCTCCCTGCCGGGCCCTGCAC CTCCCGGGGCCTCTCATCCACCCCCGGGGCTGCAACCCCAGTCCCCGGATCCCCGCCCCGTTCCACCGCGGGCTGCTTTGTGGTCCCCGC GGAGCCCCCTCAATTAAGCTCCCCGGCGCGGGGGTCCCTCGCCGACCTCACGGGGCCCCCTGACGCCCCGCTCCTCCCCTCCCCCAGGGCTA GGGTGCTGTGGCCGCTGCCGCGCAGGGACTGTCCCCGGGCGTTGCCGCGGGCCCGGACGCAGGAGGGGGCCGGGGTTGACTGGCGT GGAGGCCTTTCCCGGGGCGGGCCCGGACTGCGCGGAGCGTCGGGCGCCGCGGGCTCTGGCGGACGCCAGGGGGCAGCAGCCGC CCTCCCTGGACGCCGCGCAGTCCCCGGAGCTCCCGGAACGCCCCCCGACGGCGCGGGGCTGTGCGGCCCGCCTCGTGGCCTTCGG GTCGCCCGGGAAGAACTAGCGTTCGAGGATAAAAGACAGGAAGCCGCCCCAGAGCCCACTTGAGCTGAAACGGCCAAGGCGCGTTTCC GAGGTTCCAATATAGAGTCCAGCAGCCGGCCAGGTGGGGACTCTCGGAACCAGGCCTCCCCGCTGTGCGGCCCGGTCGGGGTCTCTTCCCG AAGCCCCTGTTCCTGGGGCTTGACTCGGGCCGCTCTTGGCTATCTGTGCTTCAGGAGCCCGGGCTTCCGGGGGGCTAAGGCGGGGCGGC CCGCGGCCTCAACCCTCTCCGCCTCCCCCTGGGCACTGCCAGCACCCGAGTTCAGTTTTGTTTTAATGGACCTGGGGTCTCGGA AAGAAAACTTACTACATTTTTCTTTTAAAATGATTTTTTTTAAGCCTAATTCCAGTTGTAAATCCCCCCCCTCCCCCGCCCAAACGTCCACTTT CTAACTCTGTCCCTGAGAAGAGTGCATCGCGCGCCCGCCCGCCGCAGGGGCCGCAGCGCCTTTGCCTGCGGGTTCGGACGCGGC CCGCTCTAGAGGCAAGTTCTGGGCAAGGGAAACCTTTTCGCCTGGTCTCCAATGCATTTCCCCGAGATCCCACCCAGGGCTCCTGGGGC CACCCCCACGTGCATCCCCCGGAACCCCCGAGATGCGGGAGGGAGCACGAGGGTGTGGCGGCTCCAAAAGTAGGCTTTTGACTCCAGG GGAAATAGCAGACTCGGGTGATTTGCCCCTCGGAAAGGTCCAGGGAGGGCTCCTCTCGGGTCTCGGGCCGCTTGCCTAAAACCCTAAACCC CGCGACGGGGGGCTGCGAGTCGGACTCGGGCTGCGGTCTCCCAGGAGGGAGTCAAGTTCCTTTATCGAGTAAGGAAAGTTGGTCCCAGC CTTGCATGCACCGAGTTTAGCCGTCAGAGGCAGCGTCGTGGGAGCTGCTCAGCTAGGAGTTTCAACCGATAAA
75	AIRE	TTCGGAAGTGAGAGTTCTCTGAGTCCCGCACAGAGCGAGTCTCTGTCCCCAGCCCCCAAGGCAGCTGCCCTGGTGGGTGAGTCAGGCCA GGCCCGGAGACTTCCCGAGAGCGAGGGAGGGACAGCAGCGCCTCCATCACAGGGAAGTGTCCCTGCGGGAGGCCCTGGCCCTGATTG GGCGCCGGGGGCGGAGCGGCCTTTGCTCTTTGCGTGGTCGCGGGGGTATAACAGCGGCGCGTGGCTCGCAGACCGGGGAGACGGG CGGGCGCACAGCCGGCGCGGAGGCCCCACAGCCCGCCGGGACCCGAGGCCAAGCGAGGGGCTGCCAGTGTCCCGGGACCCCACCGC GTCCGCCCCAGCCCCGGGTCCCCGCCCACCCCATGGCGACGGAGCGGCGCTACGCCGGCTTCTGAGGCTGCACCGCACCGGAGAT CGCGGTGGCCGTGGACAG
76	SUMO3	ACGCACACACTGGGGGGTGTGATGGAAAGGGGGGACGCGATGGATAGGGGTGGGCGCACACTGGGGGGACGCGACGGGGAGGGGTGAGCAC ACACTGGGGGGTGTGATGGAGAGGGGCGACGCAATAGGGAGGGGTGGGCCGCACACCAGGGGACGCGATGATGGGGACGGGGTGGGCGCAC ACCAGGGTGGCATGATGGGGAGGAGTGGGTACACACCATGGGGGGGCGTGATGGGGAGGGCGTGGGCGTACACCGGGGGGGCGCGATGGG GAGGGGGTGGGCGCACACCGGGGGACGCGATGGAGGCGGTGGGTGCACACGGGGGCGCGATGGGTGTAGGGTGCACACTGAGGGGC ACGATTGGGGAGACACGAAGGAGAGGGGTGGGCGCACACTGGGGGACGCGATGGCCGGGACACGATGCGGAGAAGTGGGTGAATACC GGGGTCGCGATGGGCGCCCTGGAAGGACGGCAGTGCTGCTCACAGGGGGCCAGGCCCCTCAGAGCGCGCCCCTTGGGGGTAACCCCAG ACGCTTGTTCCCGAGCCGACTCCGTGCACTCGACACAGGATC
77	C21orf70	CCACAGGGTGGGGTGCGCCCACCTGCCCTGTCCATGTGGCCTTGGGCCTGCGGGGGAGAGGGAATCAGGACCCACAGGGCGAGCCCC CTCCGTAGCCCGCGGCACCGACTGGATCTCAGTGAACACCCGTCAGCCCATCCAGAGGCTAGAAGGGGGA
78	C21orf123	TTGAGGTCTCTGTGCATGCTTGTGCGTACCCTGGACTTTGCCGTGAGGGGTGGCCAGTGCTCTGGGTGCCTTTGCCAGACAACTGGTCT GCCGGGCCGAGCATTCATGCTGGTC

SEQ ID NO	GENE NAME	SEQUENCE
79	COL18A1	TGACGCGCCCCTCTCCCCGCAGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCAGGCGGCATGCGGGGGCCGACTTCCAGTGCTTCCAGCAGG
80	PRRT3	AACACACTGTCTCGCACTAGGTGCTCGCGGAAGAGCGCGGCGTCGATGCTGCGGCTCAGGTTGATGGGCGATGGCGGCCGCAGATCCAGCTCGCTCAGCGATGGCGCCGGTCCCACACCGTTGCGGGACAGTCCCGGGCCACCCTGGGGTCCGCGACCCAACGACGCAGCCGAGCCCCAGGCGCCTGAACTGGGCGTGGCCAGCTGCCCACTCTCCGCCGGGTTGCGGATGAGGCTCTTGCTGATGTCCAAGCTGCCTGCACCAACGTTGCTTATTGGGTCGCTCCGGCACCTCGCTCTTTCCTGACGGCGCCGGGCACGCCAGACGCCATCAGCTTAGCCCAGCAAGCGTGCTCCGTGGGCGGCCTGGGTCTCGCGGCAGCCACCGCGGCCAACGCCAGGGCGAGCGCCCATGTCAGCTCCAGGAGGCGCAGCCAGAAGTGGACACCCCCACCACGAGAAGCGGCCCACGCGGCCTGGGCCCGGGTACAGCCAGAGCGCAGCCGCCAGCTGCAAGCCGCTAGCCAGCGCCCCAGCGCGCCCGCCACAGCCAACACAGCCGAGGGCCCGGGCTGGCATCCCAGCCCCGTGGGCCGTCCAGGCAGGCGGCGACAGGCAGAGCGTGCCCCAGAGCCAC
81	MGC29506	GTCTGCACGAAGCCCGCGGCGGCCTGCAGGGGGCCCAGCGACTCGTCCAGGGAACCGGTGCGCAGGAGCAGCCGGGGGGCGCGGCGCGCCGGCGCCCTTGGGGGACTCTGGGCGCAGCTCGATCTGACGCTTGGGCACTGTCCGGGGCCTGGCGGGCGCGCCCTCCTCCAGAGCCACCTCCACACACTCGAACTGCGCTGGGGCGGCAGGACTTGGCCCACGGGGCCGCAGCTCTAGGTAGGTGGCCCAGCGCGGAGCCACCATCGGGGACCTGGGACTGGCGTGGGACCGCGGCGGCAAGGGGCTGATGAAGGCCGGACTCCGTGAACTGTTGTTGCGCCTCGCGATCGTCTGCGCCCGGAGCAGCCGAACAGGGGTCCGACGCCGAAGATGACTTCCATCTCCCCCGACGGCAGCGTGCGACTGGGGCTGGGCCGTGGCGCCCCGGCCCCGCCCAGGAGCCACCCAGGGCCAATTCGCTGCTGGGCCTTTCGCGTCCGGCCAACGTCCGGGAGAACCTGGAGCCGTGTAGTAGGAGCCTGACGAACCGGAGGAGTCCTGGCGCCGCGCGGGGGCCGTGGCAGCTGCCTCGGGATCCCAGGCAGGGCTGGCGGGGCGAGCGCGGTCAGCATGGTGGGGCCCGGACGCCGTGCACTATCTCCCTCGCATTCGCCTCCGCTGGTGGCGC
82	TEAD3	CTGGAGAGAACTATACGGGCTGTGGGAGTCACCGGGGCGACTATCACCGGGCCTCCTTTCCACATCCTCCTCCGGGAAGGGACCCCGTTCCGGGCCTCGACCGGCGCAGACTGGGCTGACCCACTTTCTTGGGCCCACTGAGTCACCTCGAAACCTCCAGGCCGGTAGCGGGGAGGAGAGGAGCAGGCGGGGGTGCCAAGGTGTGGGCTGCGCCCTGGTTAGGGGGCGAGCCCGGCTTGTTTATGAGGAGGAGCGCGGAGGAGGATCCACACACAGGCTTGCGCGCCCAGACTCCGCCCGGCCAGCGGCTGGCGGCCTCCGACGTCACCAAACCGGTTGGGTGAGAGGGCAGGGGGAAGGGCCGCAGTCCCGCCCCGCGCCCCCGGCACGCACCGTACATCTTGCCCTCGTCTGACAGGATGATCTTCCCG
83	chr12group-00022	GAGTGCGGAGTGAAGGGGTGCACTGGGCACTCAGCGCGGCCCTTGGGAGGCAGGGCCGCCCCAGCCTGCCCTCCTGTCTCTGGGAAGGCCGTCCAGAAGCAGGAGCCCCGGGGAAAACAACTGGCTGGACGGGGGCGGCCTTCAGTGTCTCTCCCAGCCTGAGAGTCGCTTCCCACCACCCTGGGCACGAACCTGCTCTGCGATCTCCGGCAAGTTCCTGCGCCTCCTGTCGGTAAAATGCAGATCGTGGCGTCTT
84	CENTG1	TCTTCTTTCCGCCCCCTAGGGGGCACAAGCGGGCATGTCCAAGCGCCTAGGAGCCCGTACCGCTGCCCTCCTGTCTCTGGGAAGGCGAGCGGGTAGACCCAGAGTCCGAGTGTGGAAACAATGGAGAGGGGGCGTGTTGAGCTGGGGTCTCCATGCCTCGTTGGGGAGAGGAGGTGAGTTTGTGTCTTCTGGAAGGCGTGGGGGCTGTGCCCTCGTGGGGGTAGGAAGTGCTCCCGTGGGGATGGGGGTGTCTCTCCCGCTGGGCAACTATACCAGCGCAACCGGGGCGTCCGCCCGGTGTGGTCGTGCCCGGCCGCTGCTCCGGCGGCAGCCGTGCTGGAGAAGGATGTCGGCGCACAGCTGGCGTGGTGGCCGCGCTGGGCGTGGCCGCTGGCCTGGGCGTAGAACAGCGCCGTGCCGGCCCTGGGCGTCACGGGCCCGCCACGTCCGCGCCGTACTAGAGGGCGGAAA

SEQ ID NO	GENE NAME	SEQUENCE
		CGGCCGCGTGACCGCGCCGTCCCCAGGGGCGCCCACACCCGGCGCCGCCTCCCCCACACATGGCCAAGCCTACTTCCGGGGTCCCTCTGGG
		AATTTCGGGCTTTCCCCGCGCCAGGCGTTTTTCCGAGATGAAGCCTCAAAGACCCCCTTTCCTCCCCCCAGCTCACGTACCCACAGCAGCAG
		TTGCGTGATGACGACGTGGGCGAGCTCGGCCGCCAGGTGGAGTGGGGAGCGCAGCTGTGGGTCCTCTACGCTGGTGTCGAGCGGCCC
		GTGTCGCGCATGGGCCAAAAGCAGGAGAACGGTAGCCCACGTCCTGGGCCTGCACGGCGCGGCCCACAGCTGGCGGCCCAGCGGCTCCTC
		CGAGGTGCTCAGCGGCGCCAGGAACAGTAGCTGCTCGTACTTGGCGCGAATCCACGACTCGCGCTCCTCCCTGCAAGACCAGGGATCAA
		CGGAAAAGGCTCTAGGGACCCCCAGCCAGGACTTCTGCCCCTACCCACGGGACCGTCTCAGGTTCGCACACCCTCAGCAACCCTCCCCC
		CGCTCTGTTCCCTCACGCTTACCGCGAAGAGTCCCGCGAGGGCTTGGCACGGCCTCGCGTGTCGCTTTCCCACACGCGGTTGGCCGTGT
		CGTTGCCAATAGCCGTCAGCACCAGGGTCAGCTCCCGTGGCCAGTCGTCCAAGTCCAGCGAGCGAACGCGGGACAGGTGTGTGCCCAG
		GTTGCGGTGGATGCCAGAACACTCGATGCAGATGAGGGCGCCCAGGTTCAAGCTGGCCCACGTGGGGTCTGCGGAAGGAGCGTAGAGG
		TCGGCTCCCAGCCGGGCAGCACAGGCACCCCGGCATTCACTACACTCCCTAGCCCCTCCGCTGCCTCCTGGCACTCACTGGGGGCCCC
		GCAGTCCACGCAGATTGAATTCCCCTTGGCGTTCCGGATCGCCTGGAT
85	CENTG1	AGCCAGGTCCAGCCCCCCGCGCCTGACACCGGCCGGACGTTCCCGGGGCGCCGCAGCTGCGGCGGGAACTCTGGGATCCGGAGCCATC
		TGCTCCCACCCGCTCCGGAGCCAAACCCCGGGGGCCGCCTCCGCTCCCGGACCCGCCTCCTCTCCCGGGAGTGTGAGCCGAACCAAGA
		GTCTCCTGCCTATCTCCTCCAGTAGGAAAATAGTAATAATAGACACCCTGCCCCCCGTAAAAAACACTACCTTCCCCGTACCGCCTCCC
		AAGTCTCCCCGGGGTACGGATTGCCTTTGCAGCAGTTCCGCCCCCACCTGACTCACTCCAGGGTCAGCCCCCGGGTGGGTTTCAATGCGGCT
		CTGGGGAGGGGGGTGGGCAGTGGGGGAAGTGAGGGCTTCCTATCCGCCCCCTCTCACTTCACATTTAAATATTCTGCACGTTCCAGCCCCC
		GCGGACTCGCGTACCGTACCGCCCAATCCGCCTTCACCGCACGAAAAACATCACTAGCCTGCTCTCAGCCCAGGGGACGACTAGTCCCTGGCGA
		GAAGCTGCCTGCAAGGTCACTGTCATGCCACCTGCCCCAAGTGCTCAGGGGAAACTGAGGCTTCCTCATCCCCTTCACCTTCAACGTCGC
		TCTAAACACGGCAAAGCCCCGTTTCCATGCTCCCAGAGTTCAGCTGAGGCTGGAAGTGGGGTCCTGGGCTTCTCTGGGAGCAATTTTCTA
		GTCACTCTGATCAAGGACGTTACTTTCCCAGAAAGCTCTGAGGCTGAGTCCCTCTTTCTCCTGTCGCACAATGTAGCT
		ACTCGCCCCGCTTCAGGACTCCTATTCTTTGCCCCAATCCTTGACAGAGGGGTGAGCTTGGTTCATCCGCCCACCCCAGAGAAAAGCTTC
		CCTAGTTTCCTGGACCTCGCTCCTCCACCCCAAGCTGAGCATTCCAGGTACCCTTCCCTCCCTGTTCTCAAGCCCTGACTCAACTCACTAG
		GGGAAGCGCGGAGCTCGGCGCCCAGCAGCTCCCTGGACCCGCTGCCAGAAGACAGGCTGGGGGGTCCGGGAAGGGGCCCCGGAGCCA
		GGAGGCCCTCCTGTGCTCTTGGTGAAGATGCCGCTGATAAACTTGAGCATCTTGCGGTCACGAGTGGATGCTCGGCCCCCCTCCCGGCC
		CCGTTTCAGCCCCCGGAGCTGGAGGCTCCAGAGTGATTGGAGGTGCAGGCCCCGGGGGGCTGCGCGGAAGCAGCGGTGACAGCAGTGGC
		TGGACTCGGAGTTGGTGGGAGGGTTAGCGGAGGAGGAGCCGGCAGGGCGGTCCCGGATGCAAGTCACTGTTGTCCAAGGTCTTACTC
		TTGCCTTTCCGAGGGGACAACTTCCCTCGGGCTCCAGCCCCAGACCCCGACCCCACCAGAGGTCGAAGCTGTAGAGCCCCCTCCCCCGG
		CGGCGGCGGCGGTGGCGGCAGAGACCGAAGCTCCAGTCCCGGCGCTGCTCTTTGACCCTTGACCCTGGGCTTGCCCTCGCTTTC
		GGGCCATGACAGGCGGCTACCCGCGCCCTTGCCCCCGCCGGCTTTGGCTCCACTCGTGGTCACGGTCTTGCAAGGCTTGGGAGCCGGC
		GGAGGAGGCGCCACCTTGAGCCTCCGGCTGCCGGTGCCAGGGTGCGGAGAGGATGAGCCAGGGATGCCGCCGCCCGGCCTTCG
		GGCTCCGGGCCGCCCCAGCTCGGGCTGCTGAGCAGGGGGCGCCGGGAGGAGGTGGGGGCGCCCCCAGGCTTGGGGTCGGGGCTCAG
		TCCCCCGGAGAGCGGGGGTCCCGGAGGGACGGCCCAGAGGGAGAGGCGGCGGCCGGAGCGGGGGAGACTGGGCGGGCCCGGACTG
		GCCGGAGCCGGGGACAGGGCTGGGGGCTCCGCGCCCCCGGTGCCCGCGCTGCTCGTGCTGATCCACAGCGCATCCTGCCGGTGGAAG
		AGACGTTCGTGCCGCTTCTTGCCCGGCTCCTCCGCGCCTCGGGGGCTGCCAGGATCCCCCAGTCTCGGAGCCTCTGGCACCGGGCGGCGC
		CGGCCGCGGCCGCAGACGGAGAAGGCGGCGGAGGCACCGACTCGAGCTTAACCAGGGTCAGCGAGATGAGGTAGGTCGTTGTCC
		GGCGCTGAAGCGCCGCCCGGCTCATGGGGCCCGGAGACCCCCGAGCTGGGGAGGGGAGGGACTCCCCCCGGACTGCCTCA
		GGGGGCCCCGGCCATGGGGCCGCCCTGCTCGCTGCCCCCAGCCCCCGGACCCCGCTGAGCCCCCGGCCCGGCTCCGCTGTCGCCGC
		CGCCTCCGCCGCCTCCGCTTGCGCCCCCCCTCCCATCACATGGGGCGCCCCCTCCCCATGCTCCCCGCCCTGCGCCCCCACCCTCTTGG
		AGCCCCCGGGACCTTGGTGCTGCTCCAGGGAGGCGCCGCCGACCGTCCACACCCCGGCCTGGGTGGGGGCGCTGAGATGGGGTGGGGGAG

SEQ ID NO	GENE NAME	SEQUENCE
		GGCGGGGAGGACAGTAGTGGGGGCAAATGGGGGAGAGAGGAAAAGGGAGCAGAAAAGGGGACCGGAGGCTAGGGGAAACGAACCT GTGCGGGGGGAGGCAGGGGCGGGGAATTGGGACTCAAGGGACAGGGGGCCCGCGGATGCCGGTCGGAAAGAGGGGTCTAGAGGAGGGTGGG AAGCTAGTGG
86	chr18 group-00304	AGGAGCGCAAGGCTTGCAGGGCATGCTGGGAGAGCGCAGGGAACGCTGGGAGAGCGCGGGAAATACTGGGATTGGCTCCCGAGGGCT GTGAGGAGGGGCACGAAGGGGACACTCCGATGAAGGCAGGGCACGCGGGGCGAGCCGGGAGCGTCTCCTGAGGGCAGCGAGGAGGGAG CTGAGGCACGCGGGGCTCTCAATCGACGCCCCACAGAGACCAAGAGGCCTGGCCTTGGGGGGCAGCTGCTTGAAGGAGGCAGAGCGGA AGCGAGGGAGACTGCTGGAGGCCCTGCCGCCCACCCGCCCTTTCCTCCCCCTGAGGAGACGCCTGACGCATCTGCAGTGCAGGAGGCC GTGGGCGTTAGAAGTGTTGCTTTTTCCAGTTTGTAAGACCATTTTCCTGATTCTCTTCCCCACGGTTGCGGAGGAGCAGGTCAGGGCCGCC ATGAGGGCAGGATC
87	TSHZ1	TCGACCGCTACTATTATGAAAACAGCGACCAGCCCATTGACTTAACCAAGTCCAAGAACAAGCCGCTGGTGTCCAGCGTGGCTGATTCGG TGGCATCACCTCTGCGGGAGAGCGCACTCATGGACATCTCCGACATGGTGAAAAACCTCACAGGCCGCCTGACGCCCAAGTCCTCCACG CCCTCCACAGTTTCAGAGAAGTCCGATGCTGATGGCAGCAGCTTTGAGGAGGC
88	CTDP1	TGTGCCGTCGCACACAGACGCCCTCAACGTCGGAGAGCTGTGAGCGGGGCCGTGCTCTTGGGATGGGAGCCCCCGGGAGAGCTGCCC GCCAACACCACTCCGACGTGATCCATGCTGGACATAAAGTGCTCTTCCCTCCGCTAGTCATCGGCCGAGCGGGCCCCTCGCTCCTGGGT GTAAGTTCTTTCTGTGCGTCCTTCTCCCATCTCCGTGCAGTTCAG
89	KCNG2	CCATGCGCCGCTGCGCGCGCGAGTTCGGGCTGCTGCTGTTCCTCTGCGTGGCCATGGCGCTCTTCGCGCCACTGGTGCACCTGGC CGAGCGCGAGCTGGGCGCGCCGCGACTTCTCCAGCGTGCCCGCCAGCTATTGGTGGGCCCGTCATCTCCATGACCACCGTGGGCTAC GGCGACATGGTCCCGCCGCAGCCTGCCCGGGCAGGTGGTGGCGCTCAGCAGCATCCTCAGCGGCATCCTGCTCATGGCCTTCCCGGTCA CCTCCATCTTCCACACCTTTTCGCGCTCCTACTCCGAGCTCAAGGAGCAGCAGCGCGCGGCCAGCCCCGAGCCGGCCCTGCAGGA GGACAGCACGCACTCGGCCACAGCCACCGAGGACAGCTCGCAGGGCCCCCGACACAGCGCGGGCCTGGCCGACGACTCCGCGGATGCGCT GTGGGTGCGGGCAGGGCGCTGACGCCTGCCGCCGCCCAC

Table 6B

SEQ ID NO	GENE NAME	SEQUENCE
90	TFAP2E	GTCCTAACATCCCAGGTGGCGGCGCTGGCTCCCTGGAGCGGGGCGGGACGCGGCCGCGCGGACTCACGTGCACAACCGCGCGGGA CGGGGCCACCGCGGACTCACGTGCACAACCGCGGGACCCCAGCGCCAGCGGGACCCCAGCGCCAGCGGAGGACCCCAGCGCCAGCGGGAC CCCAGCGCCAGCGGGACCCCAGCGCCAGCGGGACCCCAGCGCCAGCGGGTCTGTGGCCCAGTGGAGCGAG TGGAGCGCTGGCGACCTGAGCGGAGACTGCGCCCTGGACGCCCCAGCCTAGACGTCAAGTTACAGCCCGCGCAGCAGCAGCAAAGGGG AAGGGGCAGGAGCCGGGCACAGTTGGATCCGGAGGTCGTGACCCAGGGGAAAGCGTGGGCGGTCGACCCAGGGCAGCTGCGGCGGCG AGGCAGGTGGGCTCCTTGCTCCCTGGAGCCGCCCCTCCCCACACCTGCCCTCGGCGCCCCCCAGCAGTTTTCACCTTGGCCCTCCGCGGT CACTGCGGGATTCGGCGTTGCCGCCAGCCCAGTGGGGAGTGAATTAGCGCCCTCCTTCGTCCTCGGCCCTTCCGACGGCACGAGGAACT

SEQ ID NO	GENE NAME	SEQUENCE
		CCTGTCCTGCCCCACAGACCTTCGGCCTCCGCCGAGTGCGGTACTGGAGCCTGCCCCGCCAGGGCCCTGGAATCAGAGAAAGTCGCTCT TTGGCCACCTGAAGCGTCGGATCCCTACAGTGCCTCCCAGCCTGGGCGGGAGCGGCGGCTGCGTCGAAGGTTGGGGTCCTTGGTGC GAAAGGGAGGCAGCTGCAGCCTCAGCCCCACCCCAGAAGCGGCCTTCGCATCGCTGCGGTGGGCGTTCTCGGGCTTCGACTTCGCCAGC GCCGCGGGGCAGAGGCACCTGGAGCTCGCAGGGCCCAGACCTGGGTTGGAAAAGCTTCGCTGACTGCAGGCAAGCGTCCGGGAGGGGC GGCCAGGCGAAGCCCCGGCGCTTTACCACACACTTCCGGGTCCCATGCCAGTTGCATCCGCGGTATTGGGCAGGAAATGGCAGGGCTGA GGCCGACCCTAGGAGTATAAGGGAGCCCTCCATTTCCTGCCCACATTTGTCACCTCCAGTTTTGCAACCTATCCCAGACACACAGAAAGCA AGCAGGACTGGTGGGGAGACGGAGCTTAACAGGAATATTTTCCAGCAGTGA
91	LRRC8D	CACCTTCCCCGAGGTAATTATTTTCTGGGGGGTAGGGGTGGGGGTTGGGAGGGTGAAGAAAGGAAGAAAAAGAAGGCCGATCACACTGG GCACCGGCGGAGGAAGCGTGGAGTCCATTGATCTAGGTACTTGTGGGGAGGGGAGAACCCGAGCAGCTGCAAACGGAAGGGCTGTG AGCGAGCGGGCGGGCGGGTGGCTGGCAGCGAGGCCACCAGCAGGGGGGGCCCGGGCCGAGGCCGCGCCACCTCGGCACCACGCGGG CAGCCGGTGCCGGCGGGGTCGCCACGGCCAGGGGAGCGCTGGGTGCCCACCATGGCAGTTATGCAAGCGGTGACCCCCTGGTCTTGCCT CCCCGCCGCCCTGCACTCCTTCCTCCCCGCTGCCGACACTTGGATCTCTCTAGCTCTTTCTCTCCCCTGTGTTTTCAAACAGGAAGTGCAC GGCTGTCTATAACGTGCTGCCGGGTCTCAGGATGGAGGAGTGAAGTCTCCTGTCGCCGTGGTTCCAGCCTCCGGAGCTCGCCCAAGCCG CGTCCCCAGAGAGCGCGCCCTGAGAGAACAGGGTGGCCGCTTGGTCCAGGTGCGCGGGGTCGGGTCTGGGTCCAGGGAGCGGGTCGGGAA GTCTGCGGCACGGAGCACTGCTAGTGTCGGATCTGCATCTCCAGCTCTGTGCTGCAGCTTCACTTGCCCGCCCCCCACCACTGGCTTCTC ACCCGGGGTCTCTGCCAAACTCTGGCTGCTGCCGCCCTGGGTTCGGGCCGGCGGAAGGCCCTGGGCGTGCGCTGCGGAGCCGCCTGCG AGGACTCCACTAGGGCGCTTTCCAGGCTGGACTGCCCCCGGGCTGCGCTGGAGCTGCCAGTGCTCGGGGAGTCTTCCTGGAGTCCCCAGC TGCCCTCTCCACC
92	TBX15	CTCTTCCCAAGTTACGCCACCGGTCGAGGACGGCAGGAGACCCCCGAGTGCAGAGAAAGCTCAAACCGGCAGCGAAGTCGGTCCTAGCC AAGCTGAAAAAACGTCTCGGATTTCGCGGACACAGCGGCCTAGACACAGCCCGATCTTCCAGTCCTAGTGCCCTGGTCGAGACGGTTCTATCC TTTTGCAAAGAAGCCGGAAA
93	C1orf51	TCTCGGTTGCAATCCCCACCCTCCTCACCCAGCAGGGCAGGAGGCACCCAACTTGGAGGAGAAAGGGGTGGGGGAGGTGAAACAGAGAC CGGAGAGTCACGAGAGGGCTGGGCCGCCGAGAGCAGGAGAATATACCGTGTCACACACCTCCATTCTCTCACACACGTTGCAGACACAAATC ACTGACGGTTTCCACGTGCTCGCTGAGCGGAGGTGTTCAAAGAGGGGGCAGATGAGTTACTTCCCGAGACGGAACCGGGGGTCCC ACGTCCGCCGCCTTCAGTAGCACAACCAATCTCTGAACACTCAAACCGCGCATCTCTGGCGCATCACCATCCTATTTAAGGCCACGGGCTC CGCCCTTTTCCTCCCCTCCCTTCTTTTCCACTCTTTTTCCA
94	chr1:17955390-179554600	CTGCCAGAGATGTGTCTGTCTTGCGCCCCGCATGCACTGCCTGCGCTGCACTCCCCGGCGGCGCCACGGGTCTGGCCCCC GCGCTTCTACGTGTTGGGGGGATGCATGGACCTTGGAGATCCGTAGTTGGCCCTAACCTTCTCGGAATCTCCCTCTGCACGCGCTGCCTGTT CCTCCTCTGCACGCTCTGTCCGTTCCTTTGCAACTTCTGTGGGAATTGTCCTGGCGTGGGAAACGCCCCCGCGCTCTTTGGCACTTAGGGT GTGAGTGTTGCGCCCCCTTGCCGCAGCGCTCAGGGCAGCATCCCGCTCGAGGATGCAGGGTTCTCACCAAGCAGTGAGGGGGACTCACGC GCCGCCGGGGAGCGGAGCCAGGCTCCGAGAAGGGAGCAGGCTCGAGCCGCTGGGTTTTCGCAAGCCTTGGGGCCTCTGGCCGCCCTTC CATGCCTCCGGGCGCGGGCGGCTCAGCAGGTCCCCGGCTTCGGGAAGTTTTGTGCGCGGATCGCTGGTGGGGAGGGCGCGCGGGCCA GTGGCTGAGCTTGCAGCGAAGTTTCCGTGAAGGAAACTGCATGTGCCTTTGGAGGCGACTCGGGACTGCTGTAGGGTGGACTGGGTGTCT ATGGAGTTGCGGGTCAGAGCGAGTAGGGTGGGTCCTTTCCTGGGACAGGACTGGGAATTGGGGCTCGAAGTAGGGG

SEQ ID NO	GENE NAME	SEQUENCE
95	ZFP36L2	AGGGGTGTCCTCCAACATCTCTGAACCGCCTTCCCTTCCTCCTCACTGGCGCCCTCTTGCCTCAGTCGTCGGAGATGGAGAGGCGGCTGA AGATTGGCAGGCGGCGGCCAGGGTCGAGGCTGGGAGACTCAGAGCCGCTGAGGCTGCCGGAGCTCAGGGAGCCGCTTAGGTAGCTGTC GCGGTCCGACAGCGAGTCCGGG
96	SIX2	TCTGACTCTCGGGCTGGAGCAGCCGAGACAGCGCTCCCCAGCGGGACTACAGAATCCCGGGTGTCGGCCTGGGGGCCCTGGATTGGCA GTGGTGGAGTCTTCTGAGCCTAACAGCTACTAGGAATGACAGAGTTGCAGATGGCTTTGTCGCCCGCGGGGCGGCTCAAGCGTCCTGGGT CCCAGGCCTCTGTCCTACGGCCAGGCCGCCGGCTCAACGGGCCGAAGGGAATCGGGCTGACCAGTCCTAAGGTCCCACGCTCCCCTGAC CTCAGGGCCCAGAGCCTCGCATTACCCCGAGCAGTGCGTTGGTTACTCTCCCTGGAAAGCCGCCCCCCGCCGGGGCAAGTGGGAGTTGCT GCACTGCGGTCTTTGGAGGCCTAGGTCGCCCAGAGTAGGCGGAGCCCTGTATCCCTCCTCGGAGCCGGCCTGCGGTGAGGTCGGTACCCA GTACTTAGGGAGGGAGGACGCGCTTGGTGCTCAGGGTAGGCTGGGCCGCTGCTAGCTCTTGATTTAGTCTCATGTCCGCCTTTGTGCCGG CCTCTCCGATTTGTGGGTCCTTCCAAGAAAGAGTCCTCTAGGGCAGCTAGGGTCGTCTCTTGGGTCTGGCGAGGCGGCAGGCCTTCTTCG GACCTATCCCCAGAGGTGTAACGGAGACTTTCTCCACTGCAGGGCGGCCTGGGGCGGGCATCTGCCAGGCGAGGGAGCTGCCCTGCCGC CGAGATTGTGGGGAAACGGCGTGGAAGACACCCCATCGGAGGGCACACCAATCTGCCTCTGCACTCGATTCCATCCTGCAACCCAGGAGAA ACCATTTCCGAGTTCCAGCCGCAGAGGCACCCGCGGAGTTGCCAAAAGAGACTCCCGCGAGGTCGCTCGGAACCTTGACCCTGACACCTG GACGCGAGGTCTTTCAGGACCAGTCTCGGCTCGGTAGCCTGGTCCCCGACCACCGCGACCAGGAGTTCCTTCTTCCCTGCTCACCA GCCGGCCGCCGGCAGCGGCTCCAGGAAGGAGCACCAACCCGCGCTGGGGGCGGAGGTTCAGGCGGCAGGAATGGAGAGGCTGATCCT CCTCTAGCCCCGGCGCATTCACTTAGGTGCGGGAGCCCTGAGGTTCAGCCTGACTTTCCCGACTCCGCCGGGCGCTTGGTGGGCTCCTG GGCTTCTGGGCTCACCCTTACACCTGTGTACTAAAGGGCTGCTACCCTCCCGAGGTGTACGTCCGCCGCCTCGGCGCTCATCGGGGTGTT TTTTCACCCTCTCGCGGTGCACGCTTTTTCTCTCACGTCAGCTCACATCTTTCAGTACACAGCCACTGGGTCTCCCTGCCCCTCCAGCCTTT CCTAGGCAGCTTTGAGGGCCCAGACGACTGAAGTCTTACTGCTAGGATGGGAACACGATGAAAAAGGAAGGGGCCCAGTCAAAAGTCCTC TCCTCTTCGGTTTTTCTTCAACTGTCCTTCACAAAAACATTTATTTCTGTCCCAGCGCCCTGGCGGATTTCGGCAGATGGGCCCTAGGGGGT TGTGGAGGCCAAATTCCCAGGATGCTGGTCCTGCCTTTTTCATTGGCCAAAACTGTATTTCCTACAACGACTAAAGATAACCAAGAACTGAG TAGACCCTGTTCTCTCACCAGATCTCCCTGGCTCTGTTTAACTTTTCCTGGTGCAATGCGATGGCACCACCAGCTCCCCAGGCAGGCACCA CTCCCTCAAGATACCATTTGGGGTAGGGATTTGAGTCCTGGAGAGGGTCAGCGGGGCGCCGGGGTGGGAAGGAGACTGACAG GGACACACCGCGAGCTCCGCATACTCTCCTCTGCCCCCTGTAGCCCGGGGCTTTAATGACCCCAAGCAGATTTCCTGTCTCTGGTCTAGCC AGCTGCCCCTAGGGCTGGATTTTATTTCTTCATGGGGTTTCACCCTAAAGGGCCCCCTGGTCATGGGACCTGGTTGGGAACAAATGAAAGA TGTCTTGTAGCAAATGCTTTCAGGGGAGCAGAAAAGAAGATTGGGCACTTCCAGTCACTTGGTCACTTTAGGTGGCTGGAACAAAACTGGT GACTTTCACGACTGCTACAGGGTGAGGGGGTGGCAGAGAGGTGACAAGCCACTGGGAATCCTATTCAGTGGGGATGCCGACA GGGAGTGGCTGTAATCAACTGAGCAACATCTGTGTGAATGTTATTCACAGGTCAGGACAGCAGCTTGGTCTTCCCAGGTGAGGAACTGAGG ACTGGCCTGCATAGATTTGTGCAGTAGGTGAGTAGCTTCCAAATTTATTTTCAGAACTTCCATGTAGTACCTGCCTCTCCATTTAAATATTTTT TAAAATTTTATTTATTTAAATATTTTCTTGGTTAGCTTTCCAAGAGGGAGGAAAAGAGGGGAGTTGCAACAAGTAGTGCCCCTATGCTGGGAT TCATTTTCCAGAGTAAAGCCTGGGACTGGCACCCTGACCCCTACCGGCAGGTGAAAACTCCAGGCAAACTGCTGAGATCCCACCTGGGCT GGCTGAGATAGTGCCTGGGGTGCATCCCTCAGCAGCTGCCACCTGGGCCCTGGGCCATCTCTTTCTCTGGCATCAAGCAGCCAGGTGTC AAGGCCTTCCCAGCAATCCATGCTGCATGGCTGGGTCTTGTTCTAGCAGGTCGATGGGCAGGGACTGGTAGCTTAGCCAGGGCACCAGTG CGTGGCTGTGGGTTTGTGTGCTTCTGTGGAGAAGCATGATGTGTATGTGTGTGTGGGCACAGGCATGAGGAAGGGTTCATTTGTGCAG GTATCTCCCATGTATATCAGTGTGGGAGAGGATGTGTTTGTGTGTCTGAAAATGGGCGGAGGGTCTGTTGTGCTAATGTGTGC AGGGGTGAACATGTGTGTGACAGTCTGTGTTTCCCTGAGTGGCTGCGTGAGAGGGTGAGGGGATTTGGTGTTGTCTACCATGCCC GGCACATAGCAGGCTCTTAATAATCTTGAATTTAATTAATGTTAAATGTGTATGTTCCCATCCTTGTGGAAGTTGGTATAGAGCCTGTTTTCCT GTGATTGTGAGACTGGAAAATGGGGGACGGGCGAGACAGGATACAGAGGCTACTGTTTTCTTCCTCCCTAGAAGTAAGTACATA GAAGAGTGGGCTCTGGCACCTCACGGGACATCACCAAGTCCTGTGTGGCTGGCTAGGCTGTCCCAAGGTGGCTTCAGGCATCACTTGAAT

SEQ ID NO	GENE NAME	SEQUENCE
		CTTTTGAGACCTTCAGGCAGTAGCCTGCCATTCACCCTGTCAGTCAGCAGAAGTTGGGCCCACACAGGCCATAGAAACACAGAGCAGTTCC
		CGGGAGGACCTGAGCTGTCCCTGAGAGCAGAGCTTCCAGGAGAGGCCGCAGGAACTGCCTTGACCGGAATTCCTCTTGGGGTGCAAAGG
		TGGAGGGACACATGGTGCGACCCCAGGCAGAGGACTGCAGCCACTCCGTGCAGTCCCAGCCTCTGGGGTAGCCCCTTGACCTCCAGGCC
		TGCACAGATCCAAGGCCGAGGTCCAGCGCCAAATTAGCTGGCCTAGCAGCCTGCAGCCGCTCTAATCTCAACTAGGAAGGAAT
		CCTTGCGCTTAGAAAGTCCAAGCGAAAGGGTATTCTGATTTTATCCCGGTTTTACCAGAAAATGCTGAAAGGAAAAGCCCCGAGAGGACAC
		AGTGCTCTAGGAACTCGGGGCGCCACGAGCGCCTCATCCCCTCCCTTCCGCCCGGCCGGTGCCCTGGTCGCTGAGGGACGCGGTCA
		GTACCTACCGCCACTGCGACCCCGAGAAGGGAAAGCCTCAACTTCTTCCTCTCGGAGTCCTGCCCACTACGGATCTGCCTGGACTGGTTCA
		GATGCGTCGTTTAAAGGGGGGGCTGGCACTCCAGAGAGGAGGGGGCGCTGCAGGTTAATTGATAGCCACGGAAGCACCTAGGCGCCCC
		ATGCGCGGAGCCGGAGCCGCCAGCTCAGTCTGACCCCTGTCTTTTCTCTCCTCTTCCCTCACTCCGGGAAAGCGAGGGC
		CGAGGTAGGGGCAGATAGATCACCAGACAGGCGGAGAAGGACAGGAGTACAGATGGAGGGACCAGGACACAGAATGCAAAAGACTGGCA
		GGTGAGAAGAAGGGAGAAACAGAGGGAGAAAGGAGGAAACAGAGCACGAGGGCGGCCGGCCGGCCCTGAGTCCGATTTC
		CCTCCTTCCCTGACCCTTCAGTTTCACTGCAAATCCACAGAAGCAGGTTTGCGAGCTCGAATACCTTTGCTCCACTGCCACACGCAGCACC
		GGGACTGGGCGTCTGGAGCTTAAGTCTGGGGGTCTGAGCCTGGGACCGGCAAATCCGCGCGCAGCGCATCGCGCCCAGTCTCGGAGACTGC
		AACCACCGCCAAGGAGTACGCGCGGCAGGAAACTTCTGCGGCCCAATTTCTTCCCCAGCTTTGGCATCTCCGAAGGCACGTACCCGCCCT
		CGGCACAAGCTCTCTCGTCTTCCACTTCGACCTCGAGGTGGAGAAAGAGGCTGGCAAGGGCTGTGCGCGTCGCTGGTGTGGGGAGGGCA
		GCAGGCTGCCCCTCCCCGCTTCTGCAGCGAGTTTTCCCAGCCAGGAAAAGGGAGGGAGCTGTTTCAGGAATTTCAGTGCCTTCACCTAGC
		GACTGACACAAGTCGTGTGTATAGGAAGGCGTCTGGCTGTTTCGGGACTCACCAGAGAGCATCGCCAACCAGAACGGCCCACCCGGGGT
		GTCGAGTCTTGGTAGGGAAATCAGACACAGCTGCACTCCCGGCCCCGCGGGCCTTGTGGCATATAACCATTTATATATTTATGATTTCTAATT
		TTATTATAAAATAAAAGCAGAAATATTTCCCGAAGAACATTCACATGAGGGCATTACGGGGAGACGGCAAGTCGGCGGCTCGGGGGGCGC
		GCTCAGCCGGGAGCGCTGTAGTCACAGTCCCGGGAGGAAGAGCGCG
		TGGAACAAGTGTCAGAGAGTAAGCAAACGACTTTCTGAGCTGTGACTCTGCTCCTCGACTGCCCACGTGCTCTCCGCTGTCTGCACTCCTG
		CCTCACCTGGGCTGACTCGGACTCTCCACCTCCTTTGCTGCTTCCGGCATGAGCTACCCCAGGAGCCTAAGGCGCTCCTTCCCCGCAACTCC
		GGTCCCCCGCGCCCCCGGGACTGCAAATCCTTTAAACAGAGGCCCCAGAGCTAGGGGTTTTCCCAGGCTCTGGTGGGCGTGGGCTGACAGT
		CGCTGGGAGCCCCCGCAACAGGGGGGATGTCCAGGCAGGTATGCACCCAGCTCCCGGCGTTTCCCGGAGTCACCACAATGTTTCCCTTTCT
		CTCTCCCCCACGTATGCTGCTAGGGGTACTCCCCAGATAGGATTTTCTTTGTCTTTTCTCCTAGTAACACCGAAGCCCTCTCGTGCCCGGG
		GACTGCAGAGGAACGCCAGACCATCCGGACCTTGCGGGATGTGGCTCGGTGTGTGTTTACTGTGTGTCGGAGTGTCGCGCATGTGTGCG
97	chr2:13723850-137240000	TGTTGGGGCGCGTTATCAACAGGGGCCTAGGGCACCCCCACTCTTTCTTGCTCTCTTCCCCCATCACTTCATGGACCTCCGAGGCGCAAAG
		CGCTCGACCCTCTCCTGGCTCAGTGGCTTGGGTACTCCGGGCTGAGCTCAGCTGGGGAGTCCCCTTACCCAGCCCGCACCGGCACCCC
		GAAGCTTCAAAGTTGCGGCAAACAGTTGCGGGGAGCAGAGGAACTGAGGTCCAGGCCAGCGCGCCCGCTCGCTTCGGCCTTGGGGAGC
		AGGCTGAGCCGAGGGTCGTGCGGGTGCGCGGCAGAGGCGGTAGGAGGCGGAGGGGGGAGAAAGAGGGGGCGGTGGGGAACA
		GCTGCCGGGGTAGGCGAGGCGCAAGGTGGCTCCCCGCGGCCCCGCGCTCTCGGACGCACCAGGCAGCCAATGGCTGCGC
		AGAGGTGTACAGCAGATGGCGTCTGACTGCGCCGTTCCTCCTCCTCCTCCTCTTCTCTTCCTCCTCTTCTTCCTCCTC
		CTCCTCCTTCAGTGCTGAGGAGCCAGAGTCGCCGCCGGGTTGCCAGACGCTGGAATGGGTGGTCTTCCGACACACCACCATCTTTCTT
		GCGCTCGGGAAGCTCGGGGCTCAGCGGCTCCCAGAGGTTACGGCGGCGGCTCTGGCGAGACGGGTGAGTGCAAGCACGCGGAGCCCC
		GAGTCGGGGATGCCGGGCCCCCTGGCCGGCCGACTGGGGCGCGGGGTGGCAGCGCCGGGGAAGGGGGCGCGCTGCCGGCGCAGACT
		TTGCTCTTTCCTCGCCGGACAGCCATCGTCGCCCCTTCTCCCAGCCAGACGCGGGAACTTGGAAGCGGATCTTCTCGGACGCCTCTGGCT
		TGGGGCTGCGGGGAAGCGTGGGCTGCCCGGGGGCGCAGTGTGCGGAGACCCTCTAGGCGGGCGGGGACGCCCCAC

SEQ ID NO	GENE NAME	SEQUENCE
98	MAP1D	GTTATTATCCACGGGGTCCTAATTAAAGCTTGATTAAAATGCCCTTCTTTCTCTAAAAAATTACGAACTAGGCAACTTCATACATTTTGAATGG CGCAGTGTTTCCTCTTCCAACTGTTTAGTTTGTAGTATACTATGTAAGCAACATCAATTATCAACCCTTGCAAGATGACAACATGAGCCTGTG GGGGAAGCACTTGAGGGGAGGGAGAAACTTCTCTTTTTTTAATAATCAGCCGGAAACAATGTTTAACAAGAATCTGATGAGGTCACTGC AGTAAATATTTTTCCTCTTACAGAGCCAATCATCACGGAGGGATCCCCTGAATTTAAAGTCCTGGAGGATGCATGGACTGTGGTCTCCCTAG ACAATCAAAGGTGTTTGCTTTCTGCTCTGTTGCTTTTAAATTGTATGGGAAAGGAAGATTGGTCCGACGGCGCGCTTGTGGCCCGGCCGGA GCTTGCGTGCGCGTTCTGACGGCTGGGTGCTGTGTTACAGGTCGGCGCAGTTCGAGCACACGGTTCTGATCACGTCGAGGGGCGCGCAG ATCCTGACCAAACTACCCCATGAGGCCTGAGGAGCCCGCCCGAAGGTCGCGGTGACCTGGTGCCTTTTTAAATAAATTGCTGAAATTTGGCT GGAGAACTTTTAGAAGAAACAGGGAAATGACCGGTGGTGCGGTAACCTGCGTGGCTCCTGATAGCGTTTGGAAGAACGCGGGGGGAGACTG AAGAGCAACTGGGAACTCGGATCTGAAGCCCTGCTGGGGTCGCCGCGGCTTTGGAAAAACAAATCCTGGC
99	WNT6	TCCCTGCTGTGGGACCCGAGGAGAGGAGAACTGGTTCGCT
100	INPP5D	TCTCTCTCTCTCTTGCTTGGTTTCTGTAATGAGGAAGTTCTCCGCAGCTCAGTTTCCTTCCCTCACTGAGCGCCTGAAACAGGAAGTCA GTCAGTTAAGCTGGTGGCAGCAGCCGAGGCCACCAAGAGGCAACGGGCGCGGCAGGTTGCAGTGGAGGGGCCTCCGCTCCCCTCGGTGGT GTGTGGGTCCTGGGGGTGCCTGCCGGCCCGGCCGAGGAGGCCCACGCCCACCATGGTCCCCTGCTGGAACCATGGCAACATCACCCGC TCCAAGGCGGAGGAGCTGCTTTCCAGGACAGGCAAGGACGGGAGCTTCCTCGTGCGTGCCAGCGAGTCCATCTCCCGGGCATACGCGCT CTGCGTGCTGTGAGTACAACCTGCTCCCTCCCCGGGCACAGATATGACAGAGGGGGCTTAGAGGGGGCCCAGCTTTGAGATGGGTTGTTCT TATGTCACAGGACAGAGTGATCTGACATGCACACTTCCCCGCCACCCTGTCAT
101	chr2:24121110 0-241211600	TGTCCTCGAAGAAGGGCCTGAGCAGCAGCAGAGGACCCCAGGCGACCGTGCCTGAGCCGGGCGCCGACGACGACTGAGCACCTGATAT GTCCCCGGCACTCGCAGCCCCCGCGGCCGGAGTCGCTGTGGGTGAGCGGTCGTCGAGCTTCACAGAGGCCGGGCTCTGTGCCAGGGCCC CGACAGGGCAGGAAGCAGATAGAGTCCCACAAGCACAAGCCCAGTGCGCAGAAAGGGTTACTTAAAAAATAAGTTCTGTGATAAAATCAAA CAGGGTGAAGGGCTGGAAACAGGTCATGAGGGCGCAAACAGGTCGTGAGGGCGCAAACAGGTCGTGAGG GCGCAAACAGGTCGTGAGGGCGCAAACAGATCGTGAGGGCGCAAACAGGTCGTGAGGGCGCAAACAGGT CGTGAGGGTGCAAACAGGTCGTGAGGGCGCAAACAGGTCGTGAGGGTGCAAACAGGT
102	WNT5A	AAATGAGACCTCTGGGGAGACTGTCAACCCCAGGGGTAAAACAAAAATTCTGATCAGAAACTGAGTTTCCCAAAGAAGGGGCTAAATGTTTT CCAACACTTTCGGGGCTCAGGGAAGATGACTCTGTAAGGACACTGAGAATCTTCCTCGCGTGCCACGGGGAGGAGGACTGGGGGCGTTTG AGGGGCTCAGCGCACCAGAGGAGTGAGGTGGAGGAGGGCGTTCCCGCGTCCTCCTCTTCAATCCAGAGGCAGCTCAACGACGTGGCTCCC TTTCTATGTATCCCTCAAAGCCTTCGCGT
103	chr3:13897160 0-138972200	TAGGCTCTAGTGGACCTAGCAGTGGGAGAGCTACTTGGGCTGGTTTCTTTCCTGACGCTGCAGGGATGGGCATCGGCCTGGAACCAGAAG CGCAGGAGCTGGGCCACGGCAGAGTAATTAAGAAAATAATGAAATTGATGGCGGATGGGGGCGCTAGAAATCCTGGGGCGTCTACTTAAA ACCAGAGATTCGCGGTCGGCCCCACGGAATCCCGGCTCTGTGTGCGCCCAGGTTCCGGGGCTTGGGCGTTCTCACACTAGGA AGGAGCCTGAAGTCAGAAAAGATGGGGCCTCGTTACTCACTTTCTAGCCCAGCCCCTGGCCCTGGGTCCCGCAGAGCCGTCATCGCAGGC TCCTGCCCAGCCTCTGGGGTCGGGTGAGCAAGGTGTTCTCTTCGGAAGCGGGAAGGGCTGCGGGGTCGGGGACGTCCCTTGGCTGCCACC CCTGATTCTGCATCCTTTTCGCTCGAATCCCTGCGCTAGGCATCCTCCCCGATCCCCCAAAGCCCAAGCACTGGGTCTGGGTTGAGGAAG GGAACGGGTGCCCAGGCCGGACAGAGGCTGAAAGGAGGCCTCAAGGTTCCTCTTTGCTACA
104	ZIC4	GAGGTTGCTGACTCAGGAGCCAGGAGCTGAGAAACTCCTAGGCTAGCAGCCGTTGAGCCTAATTTTATTTTCTGGCTTTCTCCGAAATGTCT CGTTTCCCTCATCTTTCTGGTCCTTTTCGTCTCTCTTATTTTCCCCAAAACGTCTACCTCACTTCGTCTTCCTCTCCCCTCCCCTCTC

SEQ ID NO	GENE NAME	SEQUENCE
		TTTCCTCTATACTCTCTTCCCATTTAGCCTTGCAGGCCCCTCCTCCCCGGTGTTGGAGAGCTCAAAGACGCGCGAAACTCAAGGATCTGGC
		CCTGACCAGGGACGGGATTAGGCGGGAAGTGGTGACGGCCTGAAAAGGCTGGGCTCGAACCCGTGCCTTCCTGAAAGGACTCTCCCCGC
		CACAAGTCACACCCACCCGCAGGCCTGCTGGCCAAAGAAACAAAGGAGTCGGGCGTGGATCCAGGAGAAACAGGTTTTCGCTCTCGGATC
		TCCCTGGGCAAATCAGGGATCCTGAGCGCTATACCCCGCAGTCGTACGGAGCCTCTGGGAAAGGGGATTTAAGGGTGACTTCCACTTTCA
		GCTTCGGCTACTTGTTGCCTGCGGTCCAAGCCTTCTCTGCTTCCTCCTACCTCGTCTTAGGCCTCTGTAGAAAGTGCACGCCGCGTTTCCC
		CTTCCAGGCTCTGAGAGGGCCCGTGGCCGCCTCCGACAAGATGCCTTCCAGTGCTAGGGGGGCCACTTTGGCGGGATGGGG
		GTCGGTTGGTTAAAAAAAACTTAAGTTCTGGCTCAGTCGAGTGTGGCAAAAGCCGAGGGTCGGGGGGTTGGGGGG
105	FGF12	TACTGACCTGGTCTCCGCCTCACCGGCCTCTTGCGGCCGCTGCAGAAGCGCACTTTGCTGAACACCCCGAGGACGTGCCTCTCGCACAGG
		GAGCGCCCGTCTTTGCTGGGGCTGGAGCCGGCGCTTGGAGGCCGACACTCGGTCGCTGTTGGACTCCCTCGCCTGCCGCTTCTGCCGGAT
		CAAGGAGCTGGCTATCGCCGCAGCCATAGCTGCTCAGCGAGGGCCTCAGGCCCCAGCCTCTACTGCGCCCTCCGGCTTGCGCTCCGCCG
		GGGCGAGGGCAGGACCTGGGCGGCCAGGGAAAGGGCAGTCGCGGGGAGGCAGTGCTAAAATTTGAGGAGGCTGCAGTATCGAAAACCC
		GGCGCTCACAAGGTTAGTCAAGTCTGGGCAGTGGCGACAAAATGTGTGAAAATCCAGATGTAAACTTCCCCAACCTCTGGCGGCCGGGG
		GGCGGGGCGGGGCGGTCCCAGGCCCTCTTGCGAAGTAGACGTTTGCACCCCAAACTTGCACCCCAAGGCGATCGGCGTCCAAGGGGCA
		GTGGGGAGTTTAGTCACACTGCGTTCGGGGTACCAAGTGGAAGGGGAAGAACGATGCCCAAAATAACAAGACGTGCCTCTGTTGGAGAGG
		CGCAAGCGTTGTAAGGTGTCCAAAGTATACCTACACATACATAGAAAACCCGTTTACAAAGCAGAGTCTGGACCCAGGCGGGTAGCG
		CGCCCCCGGTAGAAAATACTAAAAAGTGAATAAAACGTTCCTTTAGAAAACAAGCCACCACCGCACGAGAGAAGGAGAGGAAGGCAGCAA
		TTTAACTCCCTGCGGCCCGCGGTTCTGAAGATTAGGAGGTCCGTCCCAGCAGGGTGAGGTCTACAGAATGCATCGCGCCGGCTGCGGCTT
		TCCAGGGGCCGGCCACCCGAGTTCTGGAATTCCGAGAGGCGCGAAGTGGGAGCGGTTACCCGGAGTCTGGGTAGGGGCGCGGGGCGG
		GGGCAGCTGTTTCCAGCTGCGGTGAGAGCAACTCCCGGCCAGCACTGCAAAGAGAGCGGGAGGCGAGGGAGGGGGAGGGGCGCG
		AGGGAGGGAGGGAGATCCTCGAGGGCCAAGCACCCCTCGGGGAGAAACCAGCGAGAGGCGATCTGCGGGGTCCCAAGAGTGGGGCGCTC
		TTTCTCTTTCCGCTTGCTTTCCGGCACGAGACGGGCACAGTTGGTGATTATTTAGGGAATCCTAAATCTGGAATGACTCAGTAGTTTAAATAA
		GCCCCCTCAAAAGGCAGCGATGCCGAAGGTGTCCTCTCCAGCTCGGCGCCCCACACGCCTTTAACTGGAGCTCCCCGCCATGGTCCACCC
		GGGGCCGCCGCACCGAGCTGGTCTCCGCACAGGCTCAGAGGGAGCGAGGGAAGGAAGGGGCGCCCTGGCGGGCTCGGG
		ATCAGGTCATCGCCGCGCTGCTGCCCGTGCCCCCTAGGCTCGCGCGCCCCGGCAGTCAGCAGCTCACAGGCAGCAGATCAGATGGGGAT
		TACCCGCCGGACGCAAGGCCGATCACTCAGTCCCGCGCCGCCCATCCCCGGCCGAGGAAGGAAGTGACCCGCGCTGCGAATACCCGC
		GCGTCCGCTCGGGTGGGGCGGGGCTGGCTGCAGGCGATGTTGGCTCGCGGCGGCTGAGGCTCCTGGCCGGAGCTGCCCACCATGGT
		CTGGCGCCAGGGGCGCAGGCGGGGCCCCTAGGCCTCCTGGGGCTACCTCGCGAGGCAGCCGAGGCGCAACCCGGGCGCTTGGGGCC
		GGAGGCGGAATCAGGGGCCGGGGCCAGGAGGCAGGTGCAGGCGGCTGCCAACTCGCCCAACTTGCTGCGCGGGTGGCCGCTCAGAGC
		CGCGGGCTTGCGGGGGCGCCCCCCGCCGCGCCGCTCCCCAGGCCCGGGAGGGGGCGCTCAGGGTGGAGTCCCATTCATGGG
		CTGAGGCTCTGGGCGCGAGCCGCCGCCCCTCCGGCTGGCTCA

SEQ ID NO	GENE NAME	SEQUENCE
106	GP5	GGGGGACACAGAGAGGAGGAGGGGTTGCGGGGCCTGTGAGAATGAAGAGCACAGAGCGGAGAGGGGGGAGGAGGAAAGGAAGGGCGTGG CAGTGAGAGAGGAAGAGGAAGAAGAGGAGGAGTGGGGAGAGGGGAGAGGAGCAAGACAGCAGCCGGGTCTGGATTCCCCTCCGAGCCAC ATCTGGTCAGGTTCTAAGTAATTAGAAGATTTTCCCATTGGTTTACCCAAGGGCTCTCTCTCTGATTAATTTTCGAAAGAGTTGGCCAATTTTA ATCATAGCAAACACGATGATCACGGTGATCATGGCCTGAACAGCTAAAAGCAGAAAATAAAACCCCCCAGAACGGACTATGATCTTGACCTTT GCCCCGTGGTCACCGGCTGGGCCCACACCCAGGGTTCTGAGCTGTTGGGAGCCAAGGCTGGGTGGACAGGGGCTTCCGAGGAGCTGTCC GCAGCGGGGCGGGGAGGCGGGCCCCGGGGGCCCGGGCACTCCGCGTCACCCCCCCGGCAGGGCCCAGAGCCGGCAGGCCGGCGTGCGC CCCAGGGCCTGCGCACCGTGGGGGCTCTTCCCCGCCCACGAGGCCTAGGTGCTGCCGCAGCCACCCCCAGGAAGGGCCCCAGGCCACAG TCGCAGCGCCAGGAGTTGTGCCCCCAACAGGACCTCCGTCAGCCGGGGCAGAGCCCCCAAACACGTCGCCAGGCAGGGTCTCCAGCTGGTT GTGGTCGAGCTGGACGCTCTCCAGGCTGCTGAGATTGCGGAAGAGGGCACGGGGGCAGGGCGCGCAGCCTGTTGCGGCGCAGGGACACC
107	MSX1	GCCCCGGTGCACCGCGCCGTCCAGCCGGCCCAACTCGAGCTAGAAGCCCCAACCACTGCCCAGTGCCTGAGTTGCAGTCTTGGGTCCTTTA GAAACCTGGAGATGTGCGTAAAATTCAGATGCCGGTATTCCCGAACTTCCCCAGGCCTCAGCATATCTCGGCGGCCTGTGGACAGATGGG AGGCTACCAATCGCTCCGGCGTCCGCAGCCCGACCCCTGCCGCCAGACCCCGGACGTCTTCCGGATAATAAAGTTCCCGCTCTCTAATTCAT TTTCCCTAATCTGGACGCCCCTAATCTACAGCTTTTATTGCGCCCCAGTTAAAAGTCGAGGGAATTCGCTGTCCCTCCGCGCTCGGATAATTA CCCCTAAATGGCCACGGCAGCCCCCTTGTGTTTCCTGGAGATTAGAACCCCGCAGTCATCAATGGCAGGGCCGAGTGAGCCGCCAATCACC TCCGCTCACTCCCCTGAGAGCCGCTGGCCTGGGCCGCAGGAGGAGAGGCCATAAAGCGACAGGCGCAGAAAATGGCCAAGCCCCGACCC CGCTTCAGGC

SEQ ID NO	GENE NAME	SEQUENCE
108	NKX3-2	AGGGTGCCTCTGTTCAAATTAGAAAAAGGCGCCCCCTCAGGGCAGACTCAGCCCAGCTGCCAGGGGACAAGTCCTGGCTAACGGGAGCT GGAGCTGGGTTTCACCTCCAGGTGCCTCCTTGGCGGGGCGCCCCCGTGCAGGCTACAGCCTACAGCTGTCAGCGCCGGTCCGGAGCCGG AGCGCGGGAATCACTCGCTGCCTCAGCCCAAGCGGGTTCACTGGGTGCCTGCGGCAGCTGCGCAGGTGGAGAGCGCCCAGCCTGGGAG GCAGTAGTACGGGTAATAGTAGGAGGGCTGCAGTGGCAGAAGCGAGGGTGGCCGCAGCACTTCGCCGGGCAGGTATTGTCTCTGGTCGT CGCGCACCAGCACCTTTACGGCCACCTTCTTGGCGGCGGGCGCCGAGGCCAGCAGGTCGGCTGCCATCTGCCGGCGCTTTGTCTTGTAG CGACGGTTCTGGAACCAGATTTTCACCTGCGTCTCGGTGAGCTTCAGCGACGCGGCCAGGTCTGCGCGCTCGGGCCCGGACAGGTAGCG CTGGTGGTTAAAGCGGCGCTCCAGCTCGAAGACCTGCGCGTGGGAGAAAGCGGCCCGCGAGCCGCTTCTTGCGTGGCTTGGGCGCCGCC GGCTCCTCCTCCTCCTCCGCGACGCCTGCCGGCCCCCGCCGCCCGGCCCCGCTGCACAGCGCGGACACCGTGTGCACCTC TGGGGCCAACACCGTCGTCCTTGGGCTGCGGTCGCCTGCGGACCCCGGTGGGAACAGAAACAAGAGACTGTCAGCGCCACAG ACGAGGTGAGGGCCGGGCCTCAACTGCAGGGAGTGGGGCGGAAATACACTTTGATCCCACTCAAGCGGAGCGGAGGTCTGG GAGGCCCTGGGCCCGGGAGACCAGTCTTAGACTCTTGCCCCACTGGGTATCCCATCTAGGCCTCTTCTGGGGAGGGCGGCAGACTCAGC CGCTGTGTCAACGCTGTGTTGTCGAGACCAGCTCCCCACCCTCTCTGGGCCCCAGGCTCCCCTCAGTAACTTGGGGCACTCGACCCGAGC ATCCGCGAAAGCCCTCCCGGCTCTCAGCGTTGAGCATTGGGATTCTAGACTGCATTTCCGTCTCTCTGCTTGGGTTCACGCGCCTCTCCAC ACTTAGTTCACACGCACACACGCGCGCGTCCTCGCAGCACACACTTGTCTGGTGCAGGTAAGGGAGGTGGAGGCGGATCCTGGGGCCA AAGGTATTTAGAATCTTTCACCCTCAGCCGCCTGGGATTGCTGTGAGAGACATGGAAACAGGCTGAGCCGAGGCCTTAGATGAGAGGATG GACTGGAGAGTAAAGAGGGAGGGTTGCCCCTGCATCGAGTTTTTGGACCCTGATCCCACACCAGCTTCTCGGTCTCGTACCCGCCCTTCC GAAGAACTCCAGCAGAAAGGTCCCCTGTGCTTGAGGCCTACAGAAGCTTGTACCCAACTAGGGCAGGCACCCGGGTCTTCCA GACCACAGGACAGGACAGGCCACGGCTGAGGAGGCCTCTCTCCTGCCTCCAGGATGAACTAAAGACCCAATCCGGGATCTTCGGCCTAG GGCTGCTCTCCCAGACCTGGGGTCTGAGAAAGCCAAACCAGCCCTTTCCCCAAAGCTCTAGTTCTGCAGATTCTCAGCTCTGGCCCACTCG GAGGTGTTCTTCACCACCTATCCACTACTGTGGGGCCCGGCCCTGGGACCTTGAACTGGCAGGTCTCTGGTCCAGAGCTAGGTCACTGG CTACCTGAGGTCTCTGAACCCCTCACTTTTCCGCTTCCCTGATTTTGGGGATTTGGGGACAGACACGGCAGAAAGCACTGGCGACGAACTC AAAAACTCCCGAACGCAAGGGGCAGCGGTTCTCCCAACCCAGTCTAATGCACATTGGCCCAGGATGTCTCAGGCCTCACCCCAGGACGTA GGGCTCTGAGGAGCTACTCCGGTCTCTCGCGGGCT
109	chr4:111752000-111753000	GAGAAGGGATGTGGCGGGGGGCTCCTCCGGCCCTGGACTCCCTGGGTGGACTAGAAAAGGGCAAAGAAGTGGTCACATCTGTGGGCCAG ACTGGTGCGCGATCTTTGGAGGCGCAGCAAGGCCGCGCCAGGGCTGAGCCCAGACCGCCCACGAGGAGGCCCCGCCAGGCCCGGA GCAGCGGCGCGTGGGCTGCCGAGCGCAGGCTCTAGGGCCCCCTGCTTCGCCCCAGCTGGACCCCGCGGGCGGTCGGTGCAGCT CGAGCGTGTGGGCTGCGATGCCCTGAGACTTCGGGGCTAGGGATGCGGGGCGGGGAAGTGGGGGTGCGGCGGCAGCTGCAGATTAGA TTCCTTTTTTTTTTTGGCCGGAGGGACGTGCAAACTTCTAGTGCCCGGGCCAAGAGGGCGACCCCGGAGGTGCGTAGGTGGCCCTCCGGGT TCCCGCTTCTCCTAGTGCCTCTGAAAATACCGTCAGGGTAAAGGGAGACAGGCAGTAAGTCTTACCACCGCCCTTTCCCCATGTCATT GGCCAAAAACTGAACATTAAGATAAAGCAGCTGTTTCAGTCAATGGAAAGCGGTAGGGCGAGGTTGTACCCAAAACCCGGTTTAGACGGCC AATGAAGTCCTAGGAAAAGCCGCCCCGGGGGCACGTTCAGGTGGAGCGGCTGCACCTCGGGTCGTTCTAAGGGATGGGCTGCGTGGTAC CCACGGAATTCATGGGTCCAAAAGGTCCTGGTCACCTGTCCAAACATCCATCCCCTGGCGCATGGCGGTTGACAAGATGGCCCGGCCACC CAGAGGAAGGAGGATCCGGGACGGGGAAGCTTCGCGCCGGGAAGCTGTAGCCCAGAGCTGCAGCTCAGCATTCGCAAGAGATTCATCTTTT TTTTCTCTCGTGTTCGGAGAAACAGATAAACAAGACACCGCCTCATCAGATAAGAACGTCTCCTTCGATGTCACGGATTTCAAGAGGTAGCT GGAGAAACTGACGTCA
110	SFRP2	CAGGTCAGGCAGAACTTCTGCCCTTCCCGCTACTGGCACCCCAAGCAGGGATGCACTGGGATGCGTGGCAGGGGGCGGGATCTCCTGGGA GCGTCTCAGCCCAGCAGGGAGTGGGGAAGCAAGAGGGAAGGCTTACCTTCCTCGGTGGCTGGCAGGAGAGGTGGTCGCTGCTAGCGAGGG GGATGCAAAGGTCGTTGTCCTGGGGGAAACGGTCGCACTCAAGCATGTCGGGCCAGGGGAAGCCGAAGGCGGACATGACCGGGGGCGCA

SEQ ID NO	GENE NAME	SEQUENCE
		GCGGTCCTTCACCTGCACGCAGAGCGAGTGGCATGGCTGGATGGTCTCGTCTAGGTCATCGAGGCAGACGGGGGGCGAAGAGCGAGCACA GGAACTTCTTGGTGTCCGGGTGGCACTGCTTCATGACCAGCGGCGGATCCAAGCGCCGGCCTGCTCCAGCACCTCCTTCATGGTCTCGTGGC CCAGCAGGTTGGGCAGCCGCATGTTCTGGTATTCGATGCCGTGGCACAGCTGCAGGTTGGCAGGGATGGGCTTGCAATTGCTGCGCTTGT AGGAGAAGTCGGGCTGGCCAAAGAGGAAGAGCCCGCGCCGCCGAGCCCAGGCAGTGCGAGGCGAGGAAGAGCAGCAGCGAGC CAGGGCCCTGCAGCATCGTGGGCGCGCGACCCCGAGGGGGCGAGGAGCCGGGAAGGGCGAGGCCGGCCGGAGTTCGAGC TTGTCCCGGGCCCGCTCTCTTCCGCTGGGTGCGACTCGGGGCCCCGAAAAGCTGGCAGCCGGCGCTGGGGCGCGGAGAAGCGGGACAC CGGGAGGACAGCGCGGGCGAGGCGCTGCAAGCCCGCCGCGCAGCTCCGGGGGGGCTCCGACCCGGGGGAGCAGAATGAGCCGTTGCTGG GGCACAGCCAGAGTTTTCTTGGCCTTTTTTATGCAAATCTGGAGGGTGGGGGGAGCAAGGGAGGAGCCAATGAAGGGTAATCCGAGGAGG GCTGGTCACTACTTTCTGGGTCTGGTTTTGCGTTGAGAATGCCCCTCACGCGCTTGCTGGAAGGGAATTCTGGCTGCGCCCCCTCCCCTAG ATGCCGCCGCTCGCCCGCCCTAGGATTTCTTTAAACAACAAACAGAGAAGCCTGGCCGCTGCGCCCCCACACAGTGAGCGAGCAGGGCGCG GGCTGCCGGGAGTGGGGGGGCACGCAGGGCACCCCGCGAGCGGCCTCGCGACCAGGTACTGGCGGGAACGCGCCTAGCCCCGCGTGCCG CCGGGGCCCGGGCTTGTTTTGCCCCAGTCCGAAGTTTCTGCTGGGTTGCCAGGCATGAGTG
111	chr4:174664300-174664800	TGCGATCATTAAAATCAGTTCCTTCCCTCCTGTCCTGAGGGTAGGGGCGGGCAGATTTTATTACTTCTCTTTTCCTGATAGCAGAACTGAGG CGGGGTTGTGGAGGAGCGACGGAGGACCACCTCTAACTTCCCTTCACTTCCTGGATTTGAAGCCTCAGGGCCACCGGCCTCAGTCCTGTT ACGGTGGCGGACTCGCGAGGTTTTCCAGCAGCTCATTCCGGGACGGCGGTGTCTAGTCCAGTCCAGGGTAACTGGGCTCTCTGAGAGTCC GACCTCCATCGGTCTGGGAGCGAGTGGTTCGAGTTCAGATGCTGGGAACCGTCGCTTCTCCCCGGCCGGGCTCGCTGTTTTCTCCTCCGC TCGCCGTCATCAAGCCCCGGCTATGAGCAGGGCTTTAAATCCTCCCTCCCTCACCCGCAGGTTTACCGAGCAGCCCCGGAGCTCTCAGACA TGCTGCGCTGCCGGCGCCAGAGGAGGGGGTGGGGGGCATTGCCCTCTGCA
112	chr4:174676300-174676800	GGGCTTGGGCCGCAGGCTTCCCTGGACTTCCGCAGTCCCCCTTCTCCCCATTCCAGAACCTGCCGAGCCCCTGCTGCATCTGGGACCCGC CTTCACCGTTTCCCAATCCCAGCGGTTAGCCCCTGCGCCCCCTTTTTGGTCTCCACTTTGCCGTTCGAAAATGCCTAGGTTGGTGGATCGA CCCTCCGCGGAGCAAAGACGGATGGCTGGCAGGAGGCAGGTTCAGGAGCTGGGCCAAGGTATTCTCTGCTTCCGCCTTTGTGTCCGCCCC CCCGCCCCCTGCTCCCCGCTTCCCGCCAGCATCTCTCCTTTTCTGCTCAGGAGTGTTTGGCCCGGCGTCCACCCCGGCTTCCCGAGATA CGCTAGAGTTGCCCCCCACGTCCTGTCCGCCGCGCCCCTACCCACCGGGTTGCCTTCGGGGCCCTTCGGTGCTGTGTAGTCGGCGTGGCG CTGTGAGCTAGGCGAACAGGAACCCCCAGGCCCGCCACGTCTACGCTATTA
113	SORBS2	TTCTGGGGCCTGGATGGGTGCGAGCGGGGACCCGGGGGGAGTGGGAGTCGCCAGGCTCTGAGCAAGCAAGGGCTGCACCTGCACCTCTGC CGGGCATGAAGAAAGGTAAGGAAGGAGCTCACCCGGGTGGGAGACAGAGCCGGGGCGCGCGAGCTTGGTGTGGGGGCGCCACTC CGGGGCGGAGGGGAGGGGCTACCAGTGACTTCTCCGAGTCGGGAGCTAGAAAGAGGCTTCCGGCCAGGTTCCCTTGGAACAGGTGTCG GAGTTGTTGGGAGAGGGGGGCTGCAAGAAAGAGGGGTGCAGAAACTGGTTCATTAGATGGAGGCTCTGGGCGGAACCGCGAGGACACCCT GGCAGCGCGCTGTGCCTGCGTTAGGCCGGGAGGGGAGAGGCCTCCGGACGGCGAAGTGTCCCTAGGGACCCAGACGCCTCGGGAGCG ATCCGGGGCCGCTGCGAAGCCCTGCCCACCAGGAGTGGATCCCCAGGATTCACCTCCCGGCTGCCTGCTCTGAGCTGAGAAGGGGATCTG GTTCTTCACAATACCGTGGATGGCGGGGAAGGGGAGGGGAGCCTGGGGTAAAATCCCATCTTGGTTTCCTCG
114	chr5:42986900-42988200	TGTCACAGAAACCCCAGCAGCGCAGCCACCGGACTGGGTTCTGGAGGCCGAGCCGCAGTCCGTGCGGCGGCGCTGGGAAGAGAAGGCG CCCCGGCAGCTCCCCTGCCACCGGCCCCGAGGAGCGGCTGGCTCCCCCAGCCCAGCGCCGCCGCTCCCGGTAACTCCAGGCGCAACT GGGCGCAACTGGGGCAGCTGCGACACCGAATCCCTCACATCTGCAACCTGGGTGCTGCGGCCACTGAGAAAATGGAGGCGCAGACCAAC GAGCGGTGCCCGCGACCGAGAGACCTCGGCTGGCGAAATGGTGGTGCCGGGAGCCTGCGAGTGACGCCAGCCGGCGGGGTTGTCAAGGA CAACATTCGTTTTGACGCAGCCAATGGCGCCGTCACCAAGAAACCATCGACTCTGAGAAAAAAGAGAGGTTCGGCCACCCGAGAAACTCCGT ACGACAAGTGCTGTGGCAGAAAAACCCGCCTACTCCGCGCCACAGGCAAAACAGCCCAATGGAAACCCCAGGTGCTGCGACCGTGACACCG

SEQ ID NO	GENE NAME	SEQUENCE
		GCACTAGAGGGTCTCGGATGGAGAAAGCGGCGCACGGAGACCAGGAAACTATGTGTAGCACAACTAGCAGAAAACCGTCTGGTCGGCCAT
		CCGGGAGAAAGCGCCGGATCAGAAACAAGCGACTTCGATGCAGGGAACCGCGCAGCCACTGAAGAAAGTGACCCACGTGGCAGTGGTGCC
		AGCGAAACACTGCAGTTTGGACGGCAGCTGTGGGGATGCCACAGAGAAACATGCACTGCCACTGAAGTACATCCAGCTCCGCGGAGCTAG
		TGTTCATATGATCAAGAAACCGCCAGTTGGGCTCTGCTAGAAACTTTTAGTCCTCCCTTAACGGCTATCCTACCCACAACAGACAATGCCTTT
		ACCCAGCACCTAGCGGTGCTGAGACCCGCCTGGGCCAGCACAGAGCGCAGAGGCAGTACGGGTACGGAGAAACGCCGGACTCAGTGAAAC
		CAGCCTTGCCTCCAGCGGATTCCCCGGCTTCGCCGGACGCCACAGGCAGAGTGCCCGCGGGGAAACCTCTGGCTCCCTAAACCGATTAGA
		TTGTGGGAGTGGGGGGGACACTCACAAGTTGTGTGGAAGGGAACCAGCGGCAATGGGACCCGGCGAGCACTTGCCCGCAGCAAATGCCT
		GCGCTGCTGCAAAAAAAACAACTTTTGGCGCAAAGAATGTTGCGGCCAGAGAGCATCCGCTGTCGCTGACAAAGGAGTAGCAATGGCAAT
		GAGAAACCGCCGGCGCCACGGCCGACCGCGGCGGCTCACGCCTATGAT
115	chr5:72712000-72714100	CAAACGCTGAGAGACAAAAAGACACCAACACCCACCAGGACTGCGTCCTGCCAGCTCTTCACTCCGCTGACCTGACCTTCCACGCCCCTA
		GTCCTCGAGCCGGACTTGACCTGTGGGGGAGTACCGAACCGTCCCCATGAGGCCCTCCCAAGCGGCCAGGTGGCCTCCGCCACTCTCTCCA
		CCCCCACCTCCTCCACCCCCCCAGCCCATCGGTCCATCTTCGATCTGCAAAACACGCCGGGTCAGCGACATCGGTCCCAGGCTTGTGAC
		CACCTCTTTCTCTGTTACTTGGGGAGCCAGGCCCACCGCTCAGGATCACAGTGAGGAGAAAAAAGACACAAACGCCAGGACAGGGCGGCT
		GGGGAAGGAAACTGCTAGGGACCGCTCATTGTCAGCCTGGCGTGTCCCACGGATCGCAGGACCCGTCGAGGCTTTGCTCTCTGCGACCC
		GAATACTCCTGGGCCTCTCGACCTCCTCCTCGGACTCAGGCGTCCGCGTCTCCGGTCATCACGGGAGACCAATTGGTTTACAAATAGTGAT
		GATAAACCTGGGACCGACCTTGGGGCTGTGTAAAAGTCTACTGACAGATGTAATGGAGGGTTGTTAGCAGTCACAAAGCCTGTCGGACCCG
		TAGCATTAGTTCAAGAGACTATTTTCGTGTCGCACCAAAATTACTGCGCGTGTAAACCAATTTCCCCGACGGAAGAATAAACAGAGATTCGTT
		TGAAGCGCGAGATGAAAACAGATGGGGTATCGCAAACAGTTCCCCAAAATACAACAGACTTCTGGGCCAATTACACGTGGTTAGCTCTGAA
		TGGCAGAGGAAATAGTTTTCTTTGCTGCTAAATGTCACAAAAGTCACCTAAAGGCACAGAGGAGGCCGCTCTGTTTTTGCGAAACTTGCTAA
		AATTAATCTGCGCTGGGCCACTTGCAGAAAGCAGAACCACCTCCCGCCCCCACCTCGCCTCCAGCCGCGGGGTTCAGGCGTTTGTGAAA
		GACAGAACCTTTGGGCTAGGGACCCGGGCACTGGTGCTTCGAAGTCCGAATCCGGCCGAGAAAACGACAAGAGAAAGAAATCCAGC
		GGGCGCTCTCTCCAGCGCCAGGCCGGTGTAGGAGGGCGCTGGGGCTCGGCCTGCCACACCCTACCCGACATTGGGAAGCAGCCCCCTGCG
		CTCCCGCGGCCGCCTCAGCCTCCGGTCCCCCGCCCCGAGGTGCGCGTTCCTCCTCCCGCATGCCCCGTCTCGGGCCCCCACGGAGCAAGAAG
		ATAGACGATGACGAGGCGCGCCCATCCATCCGGGCCGACGAGGTCAGGCCCGCGCGCCACAGGCAAAAATTGCGCAAGCCCCGGCCGCAGG
		GATTTCGCCGGGCGCCTGGGTCCCAGGTGCGCGGCCCGAAATCCTCAGGGGAAAATCCCGAGGGGCCAACGGTCTAGGCCACAGGGCTGCT
		GGGCCCGGGCCTGGCTCAGAGCGCATTCGGGCGGGGAGGGCCGCACGCCGCACCCGGGCCTCTCCTCCGAGCCCGAGGCAGGCACTGA
		GCTCCGGGCCAGCCAGGTGCCTCCCGGCTGGTGCGAGACCCCGGGCCTGCTGGGAGGGCGTGGGCAGGGCAGGGCTGAACCCC
		AGCGACTGAATCTCGAAGGCAGGAGGCCTCGGAGGTCATCGGCCCAGCTCGCCTGAAACTGTCCCTGCTCGTGCCAGGGCGCGGGCAGA
		GGAGAAAGGACAGGGCGGAGCAAGCCCACTGCAGAACTGCGGTCGGTGGCTGCGAAGGGTCCGGGTCACCGCGCTCCCGGACGCCGGA
		AGCCGCGCTGGCGGGGCCGCGGGAGGAGGGCTGGGTACCGGGGCCGTCCGGCCGAGGAAGCGGCTCCGGCGCTGTCCGCGC
		TTGGGAGCCGCGTGCAGGGTTCAGCCGTGTTTCAGTTGCCCTCTGACCTGACCCCGGGCGCACAAAGGCCTCCCGGGTGCGCCGCCATG
		GCCCAGTCTTCCAGTCGCTGCCAAATTAATGAGCCCACGTCAGGTTGGGTTTACAGCTCGGCCGGGAAGCAGCCGAGTGGAAAATGAGCT
		CGGGGCCGCTCCAGAGGCTCCCGCACAACTGCAGAGGCTGCCCGCG
116	chr5:72767550-72767800	TTTCCAAGACAGAAGGAGGGAACTAGGCGCCTTTTTTCCACTCCGCTGACCCCAACGTCTGGGCTGTGCGTTGTAACGCAGTTGGCGGGG
		CCTTCAGCTTGGGATGAGGGCGAAGGGGCTCGGGATGGGTGGGAAAGCAAGGACCGGGCAACAGGTGGGGAGGGTGGCGGACTTTTGTC
		TCGGGGAAGGAAATCGGCTGTGCTGAAAGGGCGGAAAGCAGTAGCGCACAGAACTAGTGTCTGCGGGGGTCCC

SEQ ID NO	GENE NAME	SEQUENCE
117	NR2F1	CCCTCCTGTGGCTGCTTGGGCAGACGCCTGTGGCCTGTCGGATGCGGCCCACATCGAGAGGCCTGCAGGAGAAGTCGCAGTGCGCACTGG AGGAGTACGTGAGGAGCCAGTACCCCAACCAGCCCAGCCGTTTTGGCAAACTGCTGCTGCGACTGCCCTCGCTGCGCACCGTGTCCTCCT CCGTCATCGAGCAGCTCTTCTTCGTCCGTTTGGTAGGTAAAACCCCCATCGAAACTCTCATCCCGCGATATG
118	PCDHG A1	TCCTCCTTTGTGTATGTCAACCCAGAGGATGGACGGATCTTTGCCCAGCGTACCTTTGACTATGAATTGCTGCAGATGCTGCAGATTGTGGT GGGGGTTCGAGACTCCGGCTCTCCCCCCATTGCATGCCAACACATCTCTGCATGTGTTTGTCCTAGACGAGAATGATAATGCCCCCAGCTGTG CTGCACCCACGGCCAGACTGGGAACACTCAGCCCCCCCAGCGTCTCCCTCGCTCTGCTCCTCCTGGCTCCTTGGTCACCAAGGTGACAGCC GTGGATGCTGATGCAGGCCACAATGCGTGGCTCTCCTACTCACTGTTGCCACAGTCCACAGCCCCAGGACTGTTCCTCGTGTCTACACACA CTGGTGAGGTGCGCACAGCCCGGGCCTTACTGGAGGATGACTCTGACACCCAGCAGGTGGTGGTCCTGGTGAGGGACAATGGTGACCCT TCACTCTCCTCCACAGCCACAGTGCTGCTGGTTCTGGAGGATGAGGACCCTGAGGAAATGCCCAAATCCAGTGACTTCCTCATACACCCTC CTGAGCGTTCAGACCTTACCCTTTACCTCATTGTGGCTCTAGCGACCGTCAGTCTCTTATCCCTAGTCACCTTCACCTTTCTGTCAGCGAAG TGCCTTCAGGGAAACGCAGACGGGGACGGGGGTGGAGGGCAGTGCTGCAGGCGCCAGGACTCACCCTCCCCGGACTTCTATAAGCAGTC CAGCCCCAACCTGCAGGTGAGCTCGGACGGCACGCTCAAGTACATGGAGGTGACGCTGCGGCCCACAGACTCGCAGAGCCACTGCTACA GGACGTGCTTTTCACCGGCCTCGGACGGCAGTGACTTCACTTTTCTAAGACCCCTCAGCGTTCAGCAGCCCACACAGCTCTGGCGCTGGAGC CTGACGCCATCCGGTCCCGCTCTAATACGCTGCGGGAGCGGAGCCAGGTGAGGGGCTCGGCGCCGCCCGGGCGACCCCTGGGGGGCG GCACTGGAGAAGCCGCCGTCCTCATAAGGGATTGAACTTGCATCCACTCCTCTCCGGCCGGCTTGGTCGCTGGCTGCGCTCCACCCGAT TCTCGGGATCATTGGACCGTTTGCGCGAAACCAGAGTGGCCGATTAAGGGATGGGGCTCCGAGCACCGGGGGTGGTGGCGACTGTGGGC GAGGGGAGGTGGGACCGACCCCCACCCCTACACTCAAAAAAGGCCGGGGCCTCCTTCGAGCTTCCGGTGAATTTCGGGCGATTTCCGCG GGTGTCGGGGGTCCCGGGAGGAGGCAGTCACAGATCCACCCCTGCAGCCAGCCTCCTAGGCGCCGGCTCCGGCACGCTTCGCCGGTCT GTAGATTTCCTCTTCGATTTCTCCCCAGCTCCCAGCATCTGTGACTTCACTGTTACCCTCCCTATCCCCGCATCACCCAACCGCACCTGTCT GCGGGACTTAGGTGTGCGCCGGGGCTCATGCGTGTCCTCCCTGCTGGCCACCCCCACGGCCCACACAAGTTGCACGGGCTCGCCACGC CCCGCCAACACGTGCCGCGGACGCACGCACTCCTCGCACGTGGGCTTACGCGAATACCAGCTTTCACTGCCACTCGCTCCGCGGCCA GATTCACAGGCCTGTTCCGGTCCACTCGCAGCTCCCCTCTGCCGCTCCCTCCGCCGGGCTCAGGAGTACTCGTAGCTGATTGTGCGCGCC TGAGGGTCCCAGATCGCGGCCGCCCAGGACCAGGCGAGGACTCCGGAGCCTCCTCTCACCTCTCCCACCTGCGCCCCGGGCTGGGCCG GGTCGCCTGGGGGGCGGCCTGAGCGAGGCGCGGGGCCAGGAGCGCTGGAGCGACTGCCGCTCTAAGTGCCGGGCGGGCAGGACTCTA CGATCCTTGGGCCAGAGGTCCCGGATGGTCCCGGGACTCCGTCTCAAGGGTCGGCGACCCCTCAACCCAGAAGCCTCGAGCAGGCGGACA GGCAGAGCTGCCCAGTGGCCGAGGCGCGG

SEQ ID NO	GENE NAME	SEQUENCE
119	chr6:10489100-10490200	ATTTTGTCGTTTGTGCCACTTGCTGCCACTGTTGTTCTTGTCCAGGGAAACACCGGTGGCCAACCCAGATCGGATACAATGGTGCGGCTCTGGA CTGAGCCTCCAACCACACATTAGCCATGGGCAGCATTGTTGCTGCCGCTGCTGTTATTTTAATTATGATTGTACGTTAACCACCACCTTCCTTCC TCTGCCTCCCTTCAGCTGCAATGATGTATGTTACTTTTTGGTAACTGGATTTCATTAACATTTATGAACTCTCATAAAGTAGTAGAAAAAGCAA TTTGTGTGGAAGAATTTTCCACCTCATTAAACAGTGTTCTTTTGGGGGTCAAGCTGATATTTTTTTTTTGTTGTTAGATTTTTTTATAGGTCCTTT GTCCTTCCCTAAGCCCCTGGGGGATGAAAGGAGACCGTCCACACCCAGCGAGGGGCTTGTGTGCCCTAGAGGGCGCTGGGCCCCGCGCGC TTTCCTGGCTGTCCCCGCCGGCTTTCCACCCTCCCCAAAGCCCCAGGTGCCCACCGTGGGTCGCTGCGGCCTTTCCCCTTCTTGGCCAAAT CCGATTACTTCGCAGCCTGCAGATGGCATCGCCGGCTAAGGGCAGCCTGCGGCAGGTCCCCGAGCCTGAGCACTCCTCCTATCTGGGGC CTGAGAGGACGCTCTGGGCTTTTTCCCAGGCCCAGGGTGCGCGGCCTGCTAGCGCCTTTCGAGGCACAGTCCCAAGATAGGCTCTTGTCC TTCGACGCCCCCCTTGGCACAAGCGCACTGGCGCCCTCCGCTCAACCCACCTTGCCTTTGGGGCGGGCTTCAACCCTGGGAAGACAGGCC TGGGGGAAGCGAGAGGAGAGGCCCGAATAGAGGTTCCGGCTCAATCTTTCCCAGACGGAGGCCTGGTGTTTCCAGCTCAGTTGCATCTTC CAGCCGCGGGGCTCCTGGCCCAAACAGAATGTGTTTGCTTTCACACCGGGACGGCAAGCGGAGTCCGCCTCAGTGAGCAGCGAGCTGCGC AGTCCGGACGGGGTGTCGCCCCCCAGAGACTCGCCAGCCGCCCCCAGACACTCGCCAGCCGTCCCCATCTCTAATCCACCGTCCAGGCCCG GGCCCTGGGAAGA
120	FOXP4	CCGTGTCTCCCTTAAGAACTGGGGCCTCATCTCCACTCCAGCTGCGCGTGCACGTGTGCTCCCGGCAGGACGCGCGCCCAGGAGCGCGC TGGGGGCTGCCCCGCCCCTCTCTCCCTCCCCGCGGGTAAACTCCGGGCATCCATCAGTCTGTTAATTGCACTAATTAGAGATCGCAGAG GTGTTAATTGGAAAACCCTGGTATTGTGCCTGTTTGGGGGAAGAAAACGTCAATAAAAATTAATTGATGAGTTGGCAGGGCGGGCGGTGCG GGTTCGCGGCGAGGGCGCAGGGTGTCATGGCAAATGTTACGGCTCAGATTAAGCGATTGTTAATTAAAAAGCGACGGTAATTAATACTCGCT ACGCCATATGGGCCCGTGAAAAGGCACAAAAGGTTTCTCCGCATGTGGGGTTCCCCTTCTCTTTTCTCCTTCCACAAAAGCACCCCCAGCCC GTGGGTCCCCCCCTTTGGCCCCAAGGTAGGTGGAACTCGTCACTTCCGGCCAGGGAGGGGATGGGGCGGTCTCCGGCGAGTTCCAAGGG CGTCCCTCGTTGCGCACTCGCCCGCCCAGGTTCTTTGAAGAGCCAGGAGCCTCCGGGGAAGTGGGAGCCCCCCAGCGGCCCGCAGACTGC CTCAGAGCGGAAGAGGCAGCCGCGGCTTTGACCCAGCTTCCTTCCGACGGCATCTGCAGGAGCCTCTAGGCCTGACATAGGCTCCGAGG TGCCCTGGCTCCCCCCACGGGGAATGCTGAGGGGTTGGGCCACTAGGTCCTGCCTAAGTGCAGGACCTGAGCCTCAGACAAATC
121	chr7:19118400-19118700	GGGATTGCCGGCTTTGAGAAAATATGAAGAAACCGATTTCTCCTTCCACTTTGCCAGTGCACTTTCCTTCCACTTTCACTGGTGCTGGGGGC GGCGCACTCTTTACGACATATAAGCGGAAAATTCTGCAAAAGTGGCCCCCGGGGATCCCCCGCCCGACCCCTGTCTGTCGCTAATGTGGGC CTGTCTCCGGAAATTCGAGGTTGGGCCTTTGCCTGAATCTGTTGCTATTGCTCCCCTTGCTACCGCTGACACTTGGCACCGCCGCCTCCTA GCAGCGGCCAGACGCGGGGCTGGGGGC
122	chr7:27258000-27258400	GTTGCGAGCGCGGGCACAGGTTGCTGGTAGCTTCTGGACTCTGGAGGCTTGGCCTTCCTTCTAAGCCGATGGCGGGGAAAGAACCTCGTTT CCACAGCTTCCCCGACCCCCGCGCTTGCCATTTGGGGACGGGAAGCGCGCCCGGGTCGCTTCACGTCCCTCTGGGCCGGAGCCCTTTC CATGGCTGGCTCCTCTGGGGGCCCTTGGGCCTGTGAGCAGCGTCTACTTCCCTCAGAGAAGAATCCTTTCCTTCCCCCATCGAAGTGTCCC TTTCTGTATCCTGAAATAACCCCTCCTGGGTGAGGCCAGTTCCCCTCTGTCGCCCTCCTCCCGCAGGCGTCCGGGAGCCTCGTGAGGACC CCGTGCAGTTGAGTCCAGGCGACAGGTGCCTCCCCAGGTG
123	TBX20	CAGTGCGCCCCTTACCGGAGCACCCATGGCCTCCCCGCGTTACCCCAAATTTTGTAGGCAGACTGTCAGAGTTCGAAGCCAGCTGTGTCCT CTGCGGGCCGTGTGACCCTAGGCTATCTGGGCTGCTCGGAGCCTTAGTTTCCCTAGTTGTGAAGAGGGAGGGTGTGACCATGGCCCCGGA GCTCTCCGAAAGGCTGTGCGGATTGCTCGGTGGCGGGATGTGGAGCGCGTCTTCTATGATGCCCAGGTGCTGGCCAAGCGCTCGATGCAG GCTGCTCCAGTTAGGTCGATGCGATGGCGGGAAGCACTTTCCTCTGCAATGGAGAGACGCCGACACCCCGAGCCCGAAGGCTTGCAAGG CGCGCTCTCGCCACTGGGGTCGGGGATCCGTGGGTTCTCTATCCCGCTTACCCACTCCATCCTTAGCAGCTGTCGTCGGTCCCAGACCTC TACCTTGGAGAGACCAAGGCGGCCCAGAGGCCCAGAGACTACTGCGCGGTACGCCAGGATCCAGAAGTGGATTCTGACTTCTAAAGACCC

SEQ ID NO	GENE NAME	SEQUENCE
		CTCCCAAGCCAACGCTATCAGGGTCCCTGCAAGCGGTTGACTGTGGCGGAGGCAGAACCAAAACCTTTGCTCTGCCCGCGGCGCTCCAGC CTCTCACCCAGGACAGTGCTCTGGGCTCCAGCCGCTGCAGTGGGGTCGGGACACAGACGCCGAGTTAGAAGCCCCGCCGCTGCAGGTCC CTGCTTGGTCGGCGCGGTGACGGTGTCGCTGGCGGCGGGGGCCTTCCTTTGGCTGCCCGGCCATTTAATCAGAGCTATTAT
124	AGBL3	TTTAGTATTTAAGGAGAAAAGCCTCATTTTCCAGAATCGAATAAGCGAATTAATCGCACAATTGTGTAGAATGGAACTCAGTCTGTAAAAAAT CAAGACCAACGTACTTTTTTAATATTCTAACATCTCCAAGTAGTAGTTACAAGTATTGTACCCATGAAGTCCAGGTAATTAATTTGTTCAATGTC ACACTGTTAAAAGTCAGGTGGGCTCCAAAGCACAGTCCTAACCAGCATGCTCTACTGCCTCCTCTGAGGCAACAGCCGAAGTGCAGACCAC TGGGAATAAATAGCTGCCCGGTCTTCCCCACTCCTAAATTCTCCCGACAGACCCCAAAGCCTCTCTGAGAGCCTCTCTGACCGCCCTGCGG CCCACCCCGAGTTCCCGGCATCCTCTGGGATCCCTCTTCCTGGAGCCAAAACCTACGCAGGCTCCTTTCCTCCGAGCTGGTTGCTAGGTG ATCTCCGAAGGCTGTCCGAAGTCTCGCGAGGGCGGACCCGTTGCCTGATGACGAGAGTTGGGAGTGTGGCTGGGGCTGCGGATCTCCAG CAGTGGCGTTACTTCTAGCGGCTGGATACCGGGTTCTCCGCGAGATCGCGAGATCCCGAGATATTCTCCCCGCACGGAAGCGACGACTGG CCTGGCCAGAGGACTCGCGTGGGAGCGAGGTGCCGGCCCCGACAGGACGGTGAGGTATGCAGAAGTAAGGCGGGGCGCCCCCTGCGG GAAGCGAGCGCGCCCCGGAAAATGAGCGCCTCCCCACACCAAGGTGTCCAGGAGTGAGTGCGGGAAGGAACTCGGCCGCCCGGAGTTG TGGCCTCATCGTGCTTCCCGCCAAAAACGCCTTGGTACTGTCGGGACGCGGCTAAGCGTGGACGCGCCCGCATCTGCCCCTCCTCCGCA GTGGTGGAAGACACCCGCGGAGCGCCGGTGGATAAGGGCCGTTTCCTGAGACCAGAGCTGTATCCGCAGCAGGTCAGCACTTCGTGCGC CCTGTGTGC
125	XPO7	AGCGGCGCTGTTCCCGGGCTGGGTGCAGCTGCTAAGGACAAGGCCCCTGCTCCGAAGAACGCGGTGGCTCGGGGATACCCTGAAAGGG ACGGCCATGGCGCACATGGGATGCCCTAGGGTTCGTGGGAGGGCATGCAGGCGCAGCCCCCGCAGGGGTTGGCCTGCCAGAGAAGGCA GGGGAGAGCACTCGGGGCTGCACAAATGGTGTGGCCGGAGGGAAGGTGCAGCCTTGTGTGTCTGGATGAGGGCTGGGCATAGGAGC TTGGTATTTGATCCTGAAAGCTCTGCGTTTCCAAAG
126	chr8:41543400-41544000	GAGTCATACTTGTAGTCACATCCTTTTCCTTTCTCCAACCCACTGGTTAATCATGAAAGGCTCTTCTGATTGGCTGCCTCCTGGCAGTAGTGC CTCAGCGCGACGGTTCGGGAGCAAATAAATAATTCCCGCTGGGAAGCTGTTTCTCAGACAGGAGCAGCGACACCCCTGCCACGCCTGCCG CCTGGAGTTGAGTGGGGTAAGCACGCCGGCCTCCAGGAATCGACGGTGCCACGTGGTTCTTCTTGCACTTCTCTTCTCCAGTTTCAGG GGACACCGTGGGGTGTGCGAGCCCCGGGGGAGCGCAGGGAAGGGCGGGTTGGGCTGCAGGTGGGAATGTGCGGTCCTTCTGCGCCCTCA ACAGAGCTTCCTTCCTTTTTGCCAAGGTCCCCGTGCCGCCTTCAGCGCGCCTCCTTATGCACCTCTACCTCTGCTGCAGCGTACCTCTTTCC GCAGCCCTAGCGGCCTCCCCGAGGGGCGCCGGCCTCGGCTGTCCCTCCCCTGCCTGGCACGACCACCTGACCCCCAGCGACCCAAG AAGCAAGTTGTGTTTGCAGACGCAAAGGGGCTGTCGTTGGTATCGGTGCACTGGTTTGA
127	GDF6	ACACTTTCTGTGTGGGAGGGCACAAGACATGGGCTATGACATGGCCAGAGACCCCACCTTCTTTACACATGTAAAAACCAACCAAATCAAG ATGCGTCAACGGTGATTCTTCCTCCCACATTGTTTCCCTTTTTAAACTGTTATTTTTTTCAATCCATGGAGCAGTTGAGAAACGGGTATGCATC TCTCCTCCCCCTCCCCTTCTATCAAAGCCTGTAAGACACATAAGGAAATCCAAAGCCACACAGTAATAGAGAGAGAGAGAGAGAGAGA GAGAGAGAGAGAGAGAAACACAGAACAAAAGAAATCCTCCTTGGCTTGTTTTTCCAGGGTGGCCAGGCAAGGTGTGAAAATCCATATT TCCCTCTGGGCTGGCAGGTAGAAGTTACTGGGAAGGCTGCGCTCCCTTCTCTCCCACCGGCTCTCACATCCAGGCTGTTCCCTCACCCTCA GCCTCCCCCCAGCGCCAGCTTCCTCCTCCGCCTCTCTGCAGCCAGGCCTCCCCCTGCAAGGCGGACCTTGGCCCACCTTGGTTCCGGGCCA AGGCGGCGGGGAAAGGCACCGCTACCTGCAGCCGCACGACTCCACCACCATGTCCTCGTACTGCTTGTAGACCACATTATTGCCCGCGTCG ATGTATAGAATGCTGATGGGAGTCAATTTGGTGGGCACGCAGCAGCTGGGCGGGGTGGAGCCGGGGTCCATGGAGTTCATCAGCGTCTG GATGATGGCGTGGTTGGTGGCTCCAGGTGCGAGCGCAGCGGGAAGTCGCATACACCCTCGCAGTGATAGGCCTCGTACTCCAGGGGCG CGATAATCCAGTCGTCCCAGCCCAGCTCCTTGAAGTTCACGTGCAGGGGCTTCTTGCTGCAGCGTAGCCTGGACTTCTTGCCGTGCCGCTT GCCATGGCGACTGGCGAAGGCCGTGCGCCGCCGGCGAGGGCAGCCAAGGCCTGGCATCCGGGGGCGCCCGACGGCG

SEQ ID NO	GENE NAME	SEQUENCE
		GCGGCCACGACCCCTCGGCGCCCGCGCCCGGGCCCGCAGCCTCGGCCGAGCCCAGCTGCTCGCGCATCTCTGCGAACAGGTTCTTGCG
		CTGGGATCTGGTGAATACCACCAGCAGGGCCCGCTCCTGGGGAGGCCCGCACCCTCCGGCCGAAGCCCAGACTCCGCAGGTCCGGGGGC
		GGCGGTTGCTGGGGTCCCCGCGCGCCTCGGCCTCCCCGGCGTCCAGCTCGCCCCCATGCGGCCCCGCAGCTCCAAGCACAGCTGCT
		TCCAGGGCTGGTGGCGCAGGCCCTGCCACACGTCGAAGACTTCCCAGCCGGCGCGGCCCCCTGCGGGTCCAGGGTCCGCGCGTC
		CAGCAGTAGGGGCGAAGGGCAAGGGAAGAGCTGCACCGTGGAGCGGCCCCGGCTGGTGGCCCCCAGGGCGCTGAGGGCGCCTGGCGAAA
		GAGCCGCAGCTCCGCGCCCCACCAGCTCTTCTTTGTCTGAGAGCATGGACACATCAAACAAATACTTCTGTCTCCGGAGAGGAGGTGTGCGA
		GAGATCGTCTGCCGAGATAAAAAATAATTACAGTCAGTTTCACTTAAGGGGGAGATCAGCCCGGTGCTCTTCGGCCGCCCCGGGAGGAAAA
		GGGCGGGGAGTGGGGGCAGGTCGGCCGGGCAGTCCAGCTTGCCCGGCCCAGGGCCTGACCACCCCGGCTCCCCATCTGGCTGGTGCAT
		GG
128	OSR2	GCCCGCTGTGAATGTAGGTGAGGTGATCCCGGGAACCTGGGTCTGAAATCAGACCTGTGTTGCCATTGGGAGCACGGAGAGAGGGAAG
		CGCCCTGCTTAGGCCCAGGCCGGGCGTCCTGGTGGTGGGACCGCAGCCGCACTCACCTCCAGGCCAACGGACAAGGTTCCTGCAAGCCA
		GCAGGGCCACTCTGTGCTTGGCCTACTGCAGCTCCCCTGCAGCTCCTTTCCTCTCCCTCCCGGAGCGCTCTCCTCTCTCCCCTCT
		CTTCTCTCTCCTCTCTCGTCTCCTGGGGCATCCCGGGTGGAGGGATGTAGGGGTCGCTCCTCGGTGCCAGGCCGGGAAGCAGCTCAGGC
		CTCCCAAGAGCTTGGCGCTCAGTCTGGGAAAAGGGGTTCCTCTGGCCTCAGGGACGTTCTCCGCCCCCCACCCCCTGGGAGCCTG
		AACCATCTGGAAGGGATCTTAGTCGGGGGTTGGGAGGAGAGCCCGTGGATAGGAGGAGGGGGCGATTCTAGGCCGAATCCAGCCCCTGA
		GGTGTCACTTTTCTTTCCTGCGGCCCGTCACCGCTGATAGATGGGGCTGAGGGCAGAGGAAGGAAAAAGAAACCTCCGAGGTCAGTGCG
		GGGCGAGGTGAGCCCCTCTGGCCCAGGAGGATGAAGCGCGCCGGCTTCGCTCTTGCACGCCGGCTTGCCATCCGGGT
		AAGCGCGGGAAAGGCGGCCACAGGGCGCGGCAGCGCGTGGGATCTCACGACCCATCCGTTAACCCACCGTTCCCAGGAGC
		TCCGAGGCGCAGCGGCGACAGAGGTTCGCCCCCGGCCTGCTAGCATTGGCATTGGCGGTTGACTGAGCTTCGCCTAACAGGCTTGGGGAGG
		GTGGGCTGGGCTGGGCTGGGCTGCCCGGCTGTCCGCCTTTCGTTTTCCTGGGACCGAGGAGTCTTCCGCTCCGTATC
		TGCCTAGAGTCTGAATCCGACTTTCTTTCCTTTGGGCACGCGCTCGCCAGTGGAGCACTTCTTGTTCTGGCCCCCGGGCTGATCTGCACGCG
		GACTTGAGCAGGTGCCAAGGTGCCACGCAGTCCCCTCACGGCTTTCGGGGGGTCTTGGAGTCGGGTGGGGAGGAGACTTAGGTGTGGT
		AACCTGCGCAGGTGCCAAAGGGCAGAAGGAGCAGCCTTGGATTATAGTCACGGTCTCTCCCTCTTCCCTGCCATTTTTAGGGCTTTCTC
		TACGTGCTGTTGTCTCACTGGGTTTTTGTCGGAGCCCCCACGCCCTCCGGCCTCTGATTCCTGGAAGAAAGGGTTGGTCCCCTCAGCACCCC
		CAGCATCCCGGAAAATGGGGAGCAAGGCTCTGCCAGCGCCCATCCCGCTCCACCCGTCGCTGCAGCTCACCAATTACTCCTTCCTGCAGG
		CCGTGAACACCTTCCCGGGCCACGGTGGACCACCTGCAGGGCCTGTACGGTCTCAGCGCGGTACAGACCATGCACATGAACCACTGGACG
		CTGGGGTATCCCAATGTGCACGAGATCACCCGCTCCACCATCACGGAGATGGCGGGCGGGCAGGGCCTCGTGGACGCGCGCTTCCCCTT
		CCCGGCCCTGCCTTTTACCACCCCACCCTATTCCACCCCAAGCAGGGGGCCATTGCCCACGTCCTCCCAGCCCTGCACAAGGACCGGCCCCG
		TTTTGACTTTGCCAATTTGGCGGTGGCTGCCACGCAAGAGGATCCGCCTAAGATGGGAGACCTGAGCAAGCTGAGCCCAGGACTGGGTAG
		CCCCATCTCGGGCCTCAGTAAATTGACTCCGGACAGAAAGCCCTCTCGAGGAAGGTTGCCCTCCAAAACGAAAAAAGAGTTTATCTGCAAG
		TTTTTGCGGCAGACACTTTACCAAATCCTCAATTTGCTCATCCATGAGAGGACCCACACGGACGAGAGGCCGTACACGTGTGACATCTGCC
		ACAAGGCCTTCCGGAGGCAAGATCACCT
129	GLIS3	CACTCCCCCGCCGCCTCCGCCCCTAACCCTCGGCCCCGTGCGCGAGCGAGCGAGGGAGCGAACGCAGCGCAACAAAACAAACTAGTGCC
		GGCTTCCTGTTGTGCAACTCGCTCCTGAGTGAGTCGGGGGGCCGAAAGGGTGCTGCGGCTGGGAAGCCCCGGGGCGCCGGGGACCTGCGCG
		CGCTGCCCCGGCCTGGCCGGAGCCTGTAGCCCCGGGGGCGCCACGGCCGGGCTCGCAGTCCCCCCCACGCCGGCCCCCCGGTCCCCGCCG
		AGCCAGTGTCCTCACCCTGTGGTTTCCTTTCGCTTCTCGCCTCCCAAACACCTCCAGCAAGTCGGAGGGGCGCGAACGCGGAGCCCAGAAAC
		CCTTCCCCCAAAGTTTCTCCCCGCCAGGTACCTAATTGAATCATCCATAGGATGACAAATCAGCCAGGGCCAAGATTTCCAGACACTTGAGTGA
		CTTCCCCGGTCCCCCGAGGTGACTTGTCAGCTCCAGTGAGTAACTTGGAACTGTCGCTCGGGGCAAGGTGTGTGTCTAGGAGAGAGCCGGCG

SEQ ID NO	GENE NAME	SEQUENCE
		GCTCACTCACGCTTTCCAGAGAGCGACCCGGGCCGACTTCAAAATACACACAGGGTCATTTATAGGGACTGGAGCCGCGCGCAGGACAAC
		GTCTCCGAGACTGAGACATTTTTCCAAACAGTGCTGACATTTTTGTCGGGCCCCATAAAAAATGTAAACGCGAGGTGACGAACCCGGCGGGGA
		GGGTTCGTGTCTGGCTGTCTGCGTCCTGGCGGCGTGGGAGGTTATAGTTCCAGACCTGGCGGCTGCGGATCGCCCGGGCCGGTACCCG
		CGAGGAGTGTAGGTACCCTCAGCCCGACCACCTCCCGCAATCATGGGGACACCGGCTTGGATGAGACACAGGCGTGGAAAACAGCCTTC
		GTGAAACTCCACAAACACGTGGAACTTGAAAAGACAACTACAGCCCCGCGTGTGCGCGAGAGACCTCACGTCACCCCATCAGTTCCCACTT
		CGCCAAAGTTTCCCTTCAGTGGGGACTCCAGAGTGGTGCGCCCCATGCCCGTGCGTCCTGTAACGTGCCCTGATTGTGTACCCCTCTGCC
		CGCTCTACTTGAAATGAAAACACAAAAACTGTTCCGAATTAGCGCAACTTTAAAGCCCCGTTATCTGTCTTCTACACTGGGCGCTCTTAGGC
		CACTGACAGAAACATGGTTTGAACCCTAATTGTTGCTATCAGTCTCAGTCAGCGCAGGTCTCTCAGTGACCTGTGACGCCGGGAGTTGAGG
		TGCGCGTATCCTTAAACCCGCGCGAACGCCACCGGCTCAGCGTAGAAAACTATTTGTAATCCCTAGTTTGCGTCTCTGAGCTTTAACTCCCC
		CACACTCTCAAGCGCCCGGTTTCTCTCGTCTCTCGCCTGCGAGCAAAGTTCCTATGGCATCCACTTACCAGGTAACCGGGATTTCCACAA
		CAAAGCCCGGCGTGCGGGTCCCTTCCCCCGGCCGGCCAGCGCGGAGTGACAGCGGGCGGCCGGCGCTGGCGAGGAGTAACTTGGGGCT
		CCAGCCCTTCAGAGCGCTCCGCGGGCTGTGCCTCCTTCGGAAATGAAAACCCCCATCCAAACGGGGGGACGGAGCGCGGAAACCCGGCC
		CAAGTGCCGTGTGTGCGCGCGGCAGCGGCGGCAGGGGGAGGAGGGCAGAGGCGGGGGTGGCTGGACCCTCGGCAT
		CAGCTCATTCTCCCCTGCTACACACACATACACAAATAATGTTTCTAAAAAGTTCAGTTGCGACTTTGTGCCTGTCCTGTTCATCC
		TCGTCCTGGGCCGGGGAATGCTTCTGGGGGCCGACCCCGGGATGCTGGCTAATTGCTGCCGGCGGGTTCCGTCGCCGGTGTGACCCTG
		GACGGCGCGGACGGCGTACAGGGGGTCCCGGGAGGGGCAGTGGCCGCGCGGCACTCGCCGCCGGTGCCCGTGCGCGCCGCTCTGGG
		CTGCCCGGGCGGCGCAGTGTGGACGCGG
130	NOTCH 1	CTGAAAAGCCGTCAGGGAAACCACACACATGTTCAACCCCTGGCGGCTCCCCCAAACCTCTCATTTCCAGTAACTGTGTGTTTCCGCTCGTCA
		ACAGCTGAAACCGAGCGGAACTTGGGGGGGCCCCACCACGCGGCCCTGCTGTGCGGCACGGGGCTCATCTGTCCCCCGGCTGCGGGGAG
		TCAGCTCTCACCGCCCACCTCCTTCCCAGATAGTCTCTGTGCCCACTCGACGGCCCGGCAAGCCCCAGCCCCTGCCTGCCACGGCCACAGC
		AGCCTCAGAGAGCTGCCCTCTCTGGCCAGGGTCAGGGCCTGAGCTGCTCCTCCGCAGGGTCGAGGGCAGGACACTTGTCTGAGGCTT
		GGGTGGGGCAATGGCACCTCCTCAGGGCCTCAGCCCCCGGGGCAGGCTCGGTGACCATGGGCCTACAGCAGGGAAAATTCTGGGCCAAAA
		GCTCCAGCCTCCTACTAGGGCATCTGTCTGCAAATGCACCTTAACCTGACCGCTTGGGCTGTGGGGGAGCCTGTTTCAGGGAAAGTGAGG
		GACGCGCCAGTTTCCTCCTTTGGACTTGAGGCACGAACGCATCTCTAATAAAGCCAGGTCTCCCCGCCGTGGCTCCCTGGGCGGGTG
		CCTGTGGCTCGGGCCATGAGTCACGCTGGGTAACCCCACTACGGGGGAAGAGGGCAGGAAGCTGGGAGCCACCGCCTCTGTGCCCGGTTG
		TCATCTCGGCACGAGGGCGACCGTCGGCTTCGTCCTGCCCTCATGGCTGAGGGCTTTTGGGATGTGGCGGGAGACGGGGGAGTC
131	EGFL7	AAATCATCAGAATGGCTAAAATGAAAAAGACAGACAACAGCAAGTGCTGACAAGGGTGTGGGGCGGCCAAATGCTCCTGCACTGCTGGCA
		GGGGACCTGAGAACTGCAGGGCATTCCCTGGCTTCCTGCCCCTCCTGGGACTGGGGACCCCCCAGGGACAGCCTAAGGGAACTGCATTT
		ATCTTCACGTCTGCCAAAAGATAACACGAAGATGTTCAAAGCTAAGCCCCCAGGCTGGTAAGAGCTCCAAGGCACCAGCAGTGTGTGCAGA
		ACTGGGGGGGAGTCTGTTCTCCCAGGGATGCTCCCATCACCTGCTGCCAGCAGTGGGGCATGCCGGTCCCCTGGGGTGTGGCCAAGGGGC
		TGTGTCTCCTGCCCGGGCTGCCGGCCCCTCTCAGGTTCACTTTCCCATCTCTAAGCCCACGTCTCGCTGCAGTTTCAAGTTTGCCAGGCCAC
		CAACGGGTGACACGCCCGGGCGCAGTGGGGGACTCCGCACTTTCTGCGCAC
132	CELF2	ACCCTTTGTGCCTGGGTCCCATAAACAATGTGCTTTTTAAAGGGGAGCCCCCTCCCAGCTCCGGCCTTTTTCTCCAGCGTGGGCAGCCAAT
		CAGCTGCGCAGAGCTGCATAGCTGGACCGCTTTCCATTCTGAGTAGCAACAACGTACTAATTTGATGCACACATGGATGCCTCGCGCACTC
		TGCAAATTCATCACCCGCATCTTGCATTAGTCATCTGACGGACTGCCAAGTGTTTCATTTTCTTTCCATGTGACTTTATTATTACCACCTCTCT
		CCTCTCTTCCAAAAACCTCCCAAAAAGGGCGGTGGGGCGGGGGCAGGGAGAGGGAGAAATCCAGCAGACATCTAGCTCTGC
		CTTTCTTTCCCAGCCACAGCCAGGGTAGGGCTGATAAGGCGCTGATGGCAGCCTTGCAGAGCTAGACCTGCACTTAACTTGCA
		GCTGCCTCCCGAGCCTCCAAGATGTCCACGCCCTGGGTGACAGGCGGCAGGGCGCTGCCCCCGGCTCTGCTCGACAGCA

SEQ ID NO	GENE NAME	SEQUENCE
		GCACGCAGTGAGAGCCTCGCCGCCGAGGAGCAACTCATGGTGCCTCCGCTTTGTTTTAGTTCATCAAATTTCTACGACTCATTAGGCA CTTTGCCACTGCTCTTCTTCCTCCTTCCGCCTCCCCCGCTCCCCCACCCCCCACTATTTTTTCTTCCCTGTCCCTCATCGTGCCGCCCTAAC TCTGGCTCCCGGTTCCGTTTTTGACAGTAACGGCACAGCCAACAAGATGAACGGAGCTTTGGATCACTCAGACCAACCAGACCCAGATGCC ATTAAGATGTTTGTCGGACAGATCCCCCGGTCATGGTCGGAAAAGGAGCTGAAAGAACTTTTTGAGCCTTACGGAGCCGTCTACCAGATCA ACGTCCTCCGGGACCGGAGTCAGAACCCTCCGCAGAGTAAAGGTACAGAGCGCGGGGCGGGGTCGCCAGGCGTCCAGGTGGGCGTCG CGGGGCACTGGGGCTGTCCGAGCCCCCAGCCTGCAGGAGGAAGGGCCGGTAGGCAGGAGGGCTGGAAGCAGCCGGTGCTGGCCGGCCC CTGTGCTCCAGGGGCTGCTCCCGACTCCTCCCCCGCACCCCCGCCTGCCCGCCGGGACAGGTTGGAGGCGGGAGAGAGGGACCGA GGCAGGGCCGGGAGCGCAGAGGCTCGGTC
133	HHEX	TAACAAATAAGCCCGCCCCGTGGTCCGCGCTGTGGGTGACCCTTGGCGCCTTCGAGGTCTGGAGCCCTAGGGTAAATAAGGAAACGGGGCG CCTCTAGAGTTTTAAATGAACTCTGTTATTGGAAGCTTCAGTAGGGACCCTGAAAACAATTAACGTCTTAATTAGCATTTTAATGTCTCCATTA TTACGGCGCGGGCTCTAGCTCAGCCCTTTACCTTCTCACCGTTAACAGGGGAGGGGGATTGTATTTTTAGTTCATCTTTTTATGTTTT TGAGTTGTTATCCTGTCTGTCTGATTCCAGCCTCGAGGGTTTGATGATGCGGCCCGAGCCTGGCTGTGGTCGCCTGTCGGGGCTGGAGCG GGACCCTCAGCCGGGCCGGGGCCTGGGGGCTAACGTTTTCACAGTGCGCCCTGAGTTTCCTTGGGTTACTGCTGGGACCGCGCAGGAGGA AGCAAAGAGTTTTTTCGAGCTAGACCAACAGGAAACACATTGACGGAAATGTTGCCATAGCCCATGGGGGTGGCTTTAACTGGCCGCCCCCGC GGGCTGGGTGTGAAATCAGAGGAGGCCGCGGCTCCCCCGGCCAGGATTGGAGGCTCCCTCGCGCAACCTAATGCGGGTGTCCGGGCCCG AGCGCTTCCCGCGCAGCCAGGCCTTGTCGGTGCAGCAGCCCCGCTCCTCCCCAACACGCACACACCCGGTGTTCGCAAGTGCGGCTCAC CAAGGGAGATCCAAGGGGGGCAAAAAGTTATGTATAAATCCGAGAGCCACTGGGGAAAGAGGGTCGTGGTATTGTAAG
134	DOCK1 /FAM19 6A	CTACCCTGTGCTATCCTGAGCTGTAGTCTTCTGAAATGATCGTTTGGCTTCCCAGCCAAGGCAGGGCTCCCCCAAAGTTCATTCCCACTCTT GCAGTTTCACCTCGGGATGCTTCCGCAGAATTTCAGCGCCTAAGCAGACAAGGTCAAAGTAAACCGCTTCACCGCTGCTTCTGGCGCAGG GGCCCAGAGCGCGTGCAGCTCCCCAGCACAGACCAACAGCAGGAGAGGGGTCCGGGCGGGAGCCCTGGGCTGTAGATAAGGCAAAACGC ACCCATTTTCTCTCCTCATTTACTCCAGAGGCACCTCTCCTCCCCCACTCCTGGCATCTCTTTATCACTGGCTCCCTCTCCCTGTGGCATATTT TTGGGTAGTAGAATGCTGAGGTCACAGGGAGCGGCTCTTTATCCAAGCAGTGGGGACATCAGCCTGAGCCCTGAGCATGAACCAGCAAG ATGCAGACTCTCGCTCTTGACTTTGGGCTCCAGGAGCTGCCCCGACC
135	PAX6	CAGTGCTCCGCTCCGGGAAATTGCATCGTCACGACAAACGGGACCGTGATAAAACGACCCTTTCCGTCCTTATTTGTAGATCACTCAGACG AGATTGAACTGCACTTGTTTCCCCTTCGAGGGGAGCCGCGTTTTCAGGGTAGCCGAAGGCTTGGGGCTGAGGGGGGGCCCTCACCAAGG CGCGGGTGGGGGGCCGGAGCCTCAACTCGATGAGAAGTGACAGGCGTTTGGGGGATCTGGGCTCCGGCCGGGACCAGCGCAAGCAGGGA CTTTGCGGGGACACCGCTTCTCCAACAGAGCAAGGCCTGGCCCACGTTTCCGTTTCCTCCTAACTTCCTTTTATTGCCTTCCTTTGCTTCGC AAGTTCCATCTACCCCTCCAGCTACAGAGCCCCCACCTCTAGGCACAGGAAGCTTCCCGGAAAAAGAAAGGCTGTCCCAGAAAGAGACCGA GAGAGACTTTTCCAAACTTCGGGCATAGCCACGGCAATTCCCAGTCTGCTAATGCCAAGGCGGGCGCGTAAGGCCGCCTAAATCTAGACCT CCCTCCTCACTCATTTCAAAAAATAACAACGTGCCAGCCACCTCCGCAGATACCGCCGGCTGGTGCTTGCCCAGGAGACGCCAGGGCCAG AGCGCCACTCCCAGCATCGAAATGGCAGAGAGAAAGCGCAGCTCCAAATTCCCCTTCAGAGGTTAAGCCTCAATCATTGTGTCCCTTCCCT AGGGACTGCTGGCGCTCTCGCCCACTGGCGATGATTATGCGCCTAGAACTCGACCGCGAAGCAACTAATAGGAAAACATATGGTGTCAATT TGGATGCTCCCGCCTCGCGCACACCCGGGAACGAGCGGCACAAAGCCCTGCCGGCCCGGCCCCGCGAGAGACAGCCTGCGGCTGGGGAGTCTGAG CCAGCCGGGCCGCAGCGACAAACGCTCAGGGCTGCCGCGCCCTGGCTGGGGCCCGCCCGAGAGACAGCCTGCGGCTGGGGAGTCTGAG CTCCAAGGGGAGAGCCCAGCCGCGAAGGCGAGCCTACCGGCCAAGCCCTGGGGTCCGGCAGGTTCTGCACAACTACTCCCGCAAAGCT CGCCACCTTTGTGCCCTTTCCTCAG

SEQ ID NO	GENE NAME	SEQUENCE
136	FERMT 3	GGGCCCTCGCGGCTCAAGCGCCAGCGCTGGAGAGAGAGTCTGAGGGTACCACGGGCGTGCTGGCCTGGGTGCTCACTCCCGCCCTCCT TCATGAGCGGCTTTCCTCTCTGGGTGTGTCCAGGGCATCACAGAGCTCTTCTGCCCAAACCCGGAGGCCTACCAGGGCCTGCCCACCTTGCC TCCTTCCACACTCTCTGTAGCAGCAGCCGCAGCCATGGCGGGGGATGAAGACAGCCTCCGGGGACTACATCGACTCGTCATGGGAGCTGCG GGTGTTTGTGGGAGAGGAGGACCCAGAGGCCGAGTCGGTCACCCTGCGGGTCACTGGGGAGTCGCACATCGGCGGGGTGCTCCTGAAG ATTGTGGAGCAGATCAGTGAGTGTCCGCTGCCCGCTTGCTGAACTCGGCACCATGGGCGGCCGCCACGGGTGTCTCTGGGCACTTCCGG GCCATCCCCTGCTGCTCAGCTCCCGATAATGGTGTCACGGTGACTCAGGCATTAGC
137	PKNOX 2	TGTTTACGGAATCGGGATCGAGGGGCCGATAAGTAGTTTACACGCCGGCCAGAGCAGAGGGCTGGAGGTCGGAGTTGGGGGCTGGAGGA ACGGGTGGCGTTTTTAGGATTCAGTAACAGGATCACAGCTTTTTCTTGTGGTGGAAGCTATTGGAATTTGGGAGGGTAGCACGAGGGGTC CTGCAGCTCCGCGTGTGAAAAAGCGTTTAGGTAGGCGATGAAAGTAGTTGATCTGAGCCATGGCAGGCGAGCCCCGAATTTTTGCTGCTTC CCCCTGAAAGTGTTTCTTTAGGAGAGGACTTGGGCCACACAGGACCCGGTCCTAAGAGAGCGATTCCGGGAAGCGGACAGATCGAAG AGACCTTCTGGGCGAAGCGGCAGGGCAGCCTCGCGGGGGCTGGGAGTCTGAGGTCCCGACCCAGGCGGCTCGGAGTGCTCCAGGA GCCACCTGGGTCTGCGGGCGCAGCGCGGTGGCCCGCAGGGGCCGCGGCCTGCGATGAAGGCCGGGGGGGCAGC GCTAGCAGCGAGGTGCCACAGTGGGCCGAGGAGTCTGGGCTGTGGCCCAGGGTAGGACCGGCTCA
138	KIRREL 3	ACCTAAACCAAGCTCTCCCTCCCTGCCGTCTCCTTCCCTGGCCTGGGTCTGAAGGAGAGGAGGTGCCCAGAAGTTCAGAGCGGCATAACC ACAGAGATACTACCTAATTAACATACCAGAAGCATAAAGAACTCATTTGCATTGGAGAGT
139	BCAT1	ATAACTACGGGGGGTGGGGGGTGGGGAAGGAAGAGATCCAAGGAGAGGCAGAAGGCTGCGGTCAAAATATTTTGGGGTGGCAGAGTCACGTAG GATGTGGCTGTGGGTTCTGGCAGCCCAGAGATTCAGCTCCCGCCTCCTCCCTCAGAGCGAGTCCATAGCTACCCTCACGTCCCCCGTGGC GGTCCTCGCCACGCTCCGGAGCGGGTTACCCATGAGGGTGCTAGACCTGGGCAGCGGGAACCTCGAAGAGGTGGAGATTGCAGGCTGG GACTCCAGATTTCGGGCAGGGATGCGGGGAAGGAAGACGCCTCGCTGGAGGCCGGAATGGAGGGCAAGGCGAAGGAGGATGGTGCAGG AAACGGCGACAAGGCCCCGGCCAGGCCCGCGAGCTACCGAGACCCGGGTTCCAATCCTCCCCCCTTCCGCAAACGCCCGGGTTCGAG GTACCTGGCGGGCAAGGGCCGCAGCGGAAGCGGGCTGGCCATGGGGAGGCTGCGGGGACGCGGGGCTGCAGAGAGCGGCAGT GGCACGGAGCGCGCGGCTGGAAGCGAAAGCAGGCGGTGTGGCCAAGCCCCGGCGCACGGCCCATAGGGCGCTGGGTACCACGACCTG GGGCCCGCGCCCAGGGCCAGGGTACGACGCAACCCCTCCAGCATCCCTTGGGGAGGAGCCTCCAACCCGTCTCGTCCCAGTCT GTCTGCAGTCGCTAAAACCGAAGCGGTTGTCCCTGTCACCGGGGTCGCTTGCGGAGGCCCGAGAATGCGCGCCACGAACGAGCGCCTTT CCAAGCGCAGATATTTCGCGAGCATCCTTGTTTATTAAACAACCCTCTAGGTGAATGGCCGGGAAGCGCCCCTCGGTCAAGGCTAAGGAAAC CTCGGAGAAACTACAT
140	HOXC1 3	CAGTCCAGCCGCTTGCCTCACTTCTTCCCGCTTGCCTTATCTCCCCGCAGACGTGGTTCCCCTGCAGCCCGAGGTGAGCAGCTACCGGCG CGGGCGCAAGAAACGCGTGCCCTACACTAAGGTGCAGCTGAAGGAGCTAGAGAAGGAATACGCGGCTAGCAAGTTCATCACCAAAGAGAA GCGCCGGCGCATCTCCGCCACCACGAACCTCTCTGAGCGCCAGGTAACCATCTGGTTCCAGAACCGGCGGGTCAAAGAGAAGAAGGTGG TCAGCAAATCGAAAGCGCCTCATCTCCACTCCACCTGACCACCCACCCGCTGCTTGCCCCATCTATTTATGTCTCCGCTTTGTACCATAACC GAACCCACGGAAAGAAGACGCTGCGCGGGTGCAGAAGAGTATTTAATGTTAAGGAAAGAAGAACCGCGCCGCCCGGAGGCAGAGAGGCTC CATGGCCGTGCTGCTGGGCCATCCCCAACTCCCTATCCCATCCCCAGCCTCCACCCCCATCCAGATGGGACTCACGTGGCTTCAACAGCT TTGGAAATGGGTCCCGAGTGGGCCGTGCGAGGAAGGCTGTCGACCTCTACTCCCTCCTTGC
141	TBX5	CAAGATCGACTTTCTTAGGAAGGGGGAGAGGGAGGGAACTCTTCACGAAGGGAGGTGGGAGTCCACCTCAGACCTCTATTGGAAGGAAATC GAGTTGTTCCGGGGGGACTGAGGTCTCTTGCATAAGGCATGGGATCCTTATTATTATTATTTTTAAATCCCCCGCGGAGGAGCTCTG GGCAAATGAATACCGAGGCGCCGCTCTAGCTGGTTAGGCTTGGGATGCGATAACTCAGTGCCCTCTTGCAGACTTGCATAGAAATAATTAC

SEQ ID NO	GENE NAME	SEQUENCE
		TGGGTTGTCGTGGAGGGGACACGAGACAGAGGGAGTTCTCCGTAATGTGCCTTGCGGAGAGAAAGGTCCAAGAATGCAATTCGTCCCAGA GTGGCCCCGGCAGGGGCGGGGTGCGAGTGGGTGGTGGAGTAGGGGTGGGAGTGGAGAGGTGGTTTCTGTAGAGAATAATTATTGTACC AGGGCCCCGCCGAGGCACGAGGCACTCTATTTTGTTTTGTAATCACGACGACTATTATTTTTAGTCTGATCAATGGGCACAATTTCTAAGCAG CGCAGTGGTGGATGCTCGCAAACTTTTGCGCACCGCTGGAAACCCCACTAGGTTGAGTTGCAAAACGTACCGCGTAGACGCCCCTGGTGGC GCCGAGAGAAGAGCTAGGCCTGCCCAGCACAGAGCCGGAGAGCGTCGGGCCTTCCGGAAGGGTAAGTTCTCCGCCAAGGGGTCCCGAG GGAGCTGGACGTCTGAATCTGGACTTGCCCCCAGCTTCGGGGTTCGATTCTGGGTTTTGCGCGTCCCCAACCCCCAGGGCTTTCCGAAGC ATGGCCTGGCTCCAGGCCCGGTCCTGTAAGGACTGGAACGGCAGCAAAATGTGCAGGGAGGCAGTCGGCCGGCAGAGCTGCGGCCGGGA GCCAAGGTCAGGCCCCGCGGGAGAGCGGGCAGCTTCCAGCGCCGGCCACACAAGCTCCCAGGCCAGCTGGGCCGCAGACCCCTTTGCTTC CAGAGAGCACAACCCGCGTCCTTTCTCTCAGCCAGGCTGGCTGCCCCGAGCTTCGCTTTCGTTTCCCAAGCTGTTAATAACGATAT GTCCCCAAATCCGAGGCTCGTGTTTGCTCCCAGATGCCAAGAAACGCAACCCGAAATCCTTCTCCCAAACCCTAGGTCGACGAGATGAGTTC CTACTTGACCTCTGAGCCGAGGTGGGCCGGAAACCGAGGCCTAGGCCCCGCCGGGGCTGCAAGGAAAAGGGGAAACTCCGAGCGTAGC GTCTTTTCCTTGTGGTTCCTTTCTCCGGCATCCCGGACTGCGGGCCCTGCAGCCACCTGGACCGGCATTCAAAGGATTCTGCAAGTCCAGC TTCACAGACTGGCTTTCCCAGACGCTCCGAAGCCCGCACCACGAACAGAATAAAGGAGAGACGAGAGATCGCAACTAGATTTGAGAATCCT CGTTCTTTTCCCCAATCGTTCGGGCAGTAAACTCCGGAGCCGGCTACAGCGCGCATCCTC
142	TBX3	ACTGTCCTCCTCCCTCAATTGCCTATTTTTTGCCCATAGCTCTAACTTAACCCTGTGATCACCCCAGATCGCTACTTCTGACCCCCATCTCCT CTCCCACACCAACCTCCAGCGCGCGAAGCAGAGAGGAAAGTTTGCGGGGTTCGAATCGAAAATGTCGACATCTTGCTAATGGTCT GCAAACTTCCGCCAATTATGACTGACCTCCCAGACTCGGCCCCAGGAGGCTCGTATTAGGCAGGGAGGCCGCCGTAATTCTGGGATCAAA AGCGGGAAGGTGCGAACTCCCTCTTTGTCTCTGCGTGCCCGGCGCGCCCCCCTCCCGGTGGGTGATAAACCCACTCTGGCGCCGGCCATG CGCTGGGTGATTAATTTGCCGAACAAACAAAAGCGGCCTGGTGGCCACTGCATTCGGGTTAAACATTGGCCAGCGTGTTCCGAAGGCTTGTG CTGGGCCTGGCCTCCAGGAGAACCCACGAGGCCAGCGCTCCCCGGA
143	chr12:136221000-113623000	CTCAGGGAATCACATGTCCGCCTGGCCTGGTACCAAATGTTTATAGACAGGACGAGGGTCGCTGGAATCGCCTCGCTCCTTTCAG CTTGGCGCTAAGGCGCGAATCTCGATCCTCCTAGTATTTCTCTGGCGTCTGTCTCTATCTCAGTCTCTGCTTTTGTCTCTTTCTCCCTCCCTC CGCCCCAGTCTTTCCGTCTCTTTTTCCTCGAATGCACGTGGAATTCGGAATTGAAAATTGAGGTCAGAATCTCCCTTTTTCTTCCAGTTATCC GCGCCGCTGCCCCACGCCTAGCGGCTTGGATCTGCATAGACATCTATCTACCCGCAACAAGATCCGAGCTGCAGAAGCAAACCTAATCTGT CTCCGCACCATCCCCCTGCTCTGTAGACCCACTGCCCCATCCCACGCCACATCCTTGAGGTTCAAGTAGCGACTCCAGCGGATGATTCGGA GAATGCCCTGCTTTCCAAAGGCCCCAACCCGTGTTTTTATTTTCTTTTTCCTTTGCCCGCTTGACCAACTTTGGTTTCTTTCAGGGCCCGGAG GTGCCTGCGCCGCGCTTGGCTTTGCTTTCCGCCGCCCCAGGAGACCCGGGACTGTGGTTTCCGCTCGCCACATCCCAGCCTGGTGCGCA CACAAGAGCCTGGCGAGCTTCCCTCGCGCGCTTACAGTCAACTACTTTGGGCCTCGGTTTCCCTGCTCCTTGTAGATCAGAGAAGGGACG GGCGAAATGCCTGCGAGGGAGGGTTGGCGAATGGGTTGGTTGGTGGCAAGACTGCAGTTCTTGTACATGGACGGGGGTTGGGGGGTCAA CACTGGAAGAACTCCTGCCTGACGCCAAGAGCCACCCGCTTTCCAGCTCGTCCCACTCCGCGGATGTTTACCCACCTTCATG
144	chr12:136578000-113658300	TTTGGGGCACCCAACCCTTCCCAAGCCTCGGTTTTCCCGATCTTGTGGGATCCTTGCGGCGCGAATGGGGTTGGAAGCACCTTGGAAGCT ACAGAGTACCGGGTCGGGACAATTTCCGGCACTGCCCCAGTTCAGTGGTTTATAGAAAATTTCTTTCTCTCTCAGGTCCACTAAGACCGA GAGAGAGAGAAGTCGACTCTGGCACACCCGGGCGAGGGGCTGCCGGGATTCGGGAGCTGGCGCGGTTGATTTTTTCCGAGAATCCTC CACTTGGGGTGACGTCGGGCAGCGCGCGGCCGTGAGGTTAATGCCCAGGCTTTTCTCTAAAGCGTCCGGGAATGATCCGGCGAATA AAACGGGTGTCTGCAAAGTTAATGAATTGTACAAGGAGGCTGAGGGTGGGGACTTCGACCCGGGGAGCCAGAGGCGGTTCTGGTGGACG CTTCCCCGTGCGCCTAGGGGTGCGCTGGGCTTTCCCAGCCGAGGTCTGCAG

SEQ ID NO	GENE NAME	SEQUENCE
145	THEM2 33	CCAGACAGTTAAGGTAAAACGTTGAAGTCAAGAGGAAGTAGTGAGTCTGTTGCCAACTGGATAGGGTTGGTCCTGTCCCATCTAAATGTATT AGAATTAAGTGGCTTTTAAAAATGAGCTGGTCATCTTCAGCCCACGGGCTGGCCAATTTGGAACTTAATGGGCCTTTGCGTCCTCCTTCCCT GAGCCTCCTTTTATTCCAGACTTCTCAGTGTGAGTCTGTGCGTCCCTCCGACGATCTCAGGGAGTGGGGTGCCTTCATCTGCCTGTTCCCT GTTCCTCAGGCTGACGCTCCCGCTGTCCTCCCCGCCTCCCTCACTCCTTTCTCCCTCCCTTCCTCCTTGTGGGGAGGCTCTTGGCCAGG GTCCCTGAGCCCGGGCGGGTGCTGGCAGAGGACGCAGAAGGGGTGAGGTCACGTCTCCCTTGAGCCCCGAGCCGCTGGCTTTTCAGAG CCTCGCCACAAGCCGGCGGCCAGAGCCCCAGACCACACAGACCGTGCGCTCCTCCGCCCTCCCGGGCGCCGGCCTCGCCCATGTCT CAGTACGCCCCCTAGCCCCGGACTTTCAAGAGGGGCTTTGGACAGCAGTCCCGAGGCCAACACTGAAGATGACAAGACCGAGGAGGACGTGCC CATGCCCAAGAACTACCTGTGGCTCACCATCGTCTCGTGTTTTTGCCCTGCGTACCCCATCAACATCGTGGCTTTGGTCTTTTCCATCATGG TGAGTGAATCACGGCCAGAGGCAGCCTGGGAGGAGAGACCCGGGCGGCTTTGAGCCCCTGCAGGGGAGTCCCGCGCGCTCTCTGCGGCT CCCTTCCTCACGGCCCGGCCCGCGCTAGGTGTTCTTTGTCCTCGCACCTCCTCCTCACCTTTCTCGGGCTCTCAGAGCTCTCCCCGCAATC ATCAGCACCTCCTCTCGTGCACTCCTCGTGGTACTCAGAGCCCTGATCAAGCTTCCCCCAGGCTAGCTTTCCTCTTCTTTCCAGCTCCCAGGGT GCGTTTCCTCTCTCCAACCCGGGGAAGTTCTTCCGTGGACTTTGCTGACTCCTCTGACCTTCCTAGGCACTTGCCCGGGGCTTCTCAACCCTC TTTTCTAGAGCCCCAGTGCCGCGCCACCCTAGCGAGCGCAGTAAGCTCATACCCCGAGCATGCAGGCTCTACGTTCCTTTCCCTGCCGCTC CGGGGGCTCCTGCTCTCCAGCGCCCAGGACTGTGCTCCCTTCTCTCTTTGGCTGCGCCCAAGGGCACCGCTTCCG CCACTCTCCGGGGGGGTCCCCAGGCGATTCCTGATGCCCCCTCCTTGATCCCGTTTCCGCGCTTTGGCACGGCACGCTCTGTCCAGGCAAC AGTTTCCTCTCGCTTCTTCCTACACCCAACTTCCTCTCCTTGCCTCCCTCCGGCGCCCCCTTTTTAACGCGCCCGAGGCTGGCTCACACCC ACTACCTCTTTAGGCCTTTCTTAGGCTCCCCGTGTGCCCCCCTCACCAGCAAAGTGGGTGCGCCTCTCTTACTCTTTCTACCCAGCGCGTC GTAGTTCCTCCCCGTTTGCTGCGCACTGGCCCTAACCTCTCTTCTCTTGGTGTCCCCCAGAGCTCCCAGGCGCCCCTCCACCGCTCTGTCC TGCGCCCGGGGGCTCTCCCGGGAATGAACTAGGGGATTCCACGCAACGTGCGGCTCCGCCCGCCCTCTGCGCTCAGACCTCCCGAGCTGC CCGCCTCTCTAGGAGTGGCCGCTGGGGCCTCTAGTCCGCCCTTCCGGAGCTCAGCTCCCTAGCCCTCTTCAACCCTGGTAGGAACACCCCG AGCGAACCCCACCAGGAGGGCGACGAGCGCCTGCTAGGCCCTCGCCTTATTGACTGCAGCAGCTGGCCCGGGGGTGGCGGCGGGGTGA GGTTCGTACCGGCACTGTCCCGGGACAACCCTTGCAGTTGCGCTCCCTCCCCCACCGGCTCACCTCGCCTGCAGCTGGGCCACGGAACT CCCCGGCCACACAGACGCA
146	NCOR2	CTCTCTGGGCCTTAGGAAATGGAAATGACACCTGTACCTGCCCTTCCAGGACTGACAGGAGGGGGCTGCTCCATGAAACCTCACTGCTGC GGTCATAATGTCATTATCTTTGCCTTAAAGGGATTTCTTCTGCACCAGCACCTAAAGTGGCAGCCCCTTACCCTTGGCCATCAGCTGGACC CTGGTGCTCTCCTGGAGCCCAAAACCTCTGTTTTGTGTTGCATCCTGCTGACCAGCCACAGTCCACACCCATCTGAGTGTCTGAGCAGAAC AGCCCAGAGGCCACACCAGGATGGCTTTCCACCGGTCACCTTCCCCCACCCACTCATAAACCCTGCGTCTCTGGGGGAGAGGGTGGCGA GGTCCCCTCCCCACATAGATGGAAACACTGAGGCCTGATTCATGGTGCCCCCTGTGAAGCGCCTCATGGCCAGCACCGGGGGGCAGCAG GCCAGGGCGGGGACACATACCCGGTTCTCGTCGTAGATGATCTCTGCACCAGGCTGCGGTGCTTCGACTCGATGGGCGGCGGTGACACGGG CTTCTCAGGCTCGGGCGGCTTGGCAGCCTCCTCCTCCAGCTGTTGCTGTGGGGAGAGGCA
147	THEM1 32C	CTTGAAAACTCCCAGCCCCCTTTGTCCAGATGGGGATGGAGGTGGCCAGGCTGCCCCGTTGATTGTGTGCCGAGGAGCCCTCCCCGGGA AGGCTGTGATTTATACGCGCAGGCTTGTCACGGGGGTGAAAGGAAGGGCCACTTTTTCATTTTGATCCAATGTTAGGTTTGAAAGCCACCCAC TGCTGTAAACTCAGCTGGATCCGCGGGCCGTGATTAAACACATTGCCCGCTTTGTTGCCGAGATGGTGTTTCGGAAGGCGCTGTGAATGCA CTTCCCTTTGCGGGGCTCACACAGACAAGATGTGTGTTGCAAGGATGAGGCGCCTGCTCGGCCTCCAGCCCAGGGCCGGGAAGGGAGAA GGTGCTGTGCGTCGCTGCCTGTGTCGCCCGCGGCTCTCC
148	PTGDR	CGCGTCAGGGCCGAGCTCTTCACTGGCCTGCTCCGCGCTCTTCAATGCCAGCGCCAGGCGCTCACCCTGCAGAGCGTCCCGCCTCTCAA AGAGGGGTGTGACCCGCGAGTTTAGATAGGAGGTTCCTGCCGTGGGGAACACCCCGCCGCCCTCGGAGCTTTTTCTGTGGCGCAGCTTCT CCGCCCGAGCCGCGCGCGGAGCTGCCGGGGGCTCCTTAGCACCCGGGCGCCCTCGCCCTTCCGCAGCCTTCACTCCAGCCCT

SEQ ID NO	GENE NAME	SEQUENCE
		CTGCTCCCGCACGCCATGAAGTCGCCGTTCTACCGCTGCCAGAACACCACCTCTGTGGAAAAAGGCAACTCGGCGGTGATGGGCGGGGT
		GCTCTTCAGCACCCGGCCTCCTGGGCAACCTGCTGGCCCTGGGGCTGCTGGCGCGCTCGGGGCTGGGGTGGTGCTCGCGGCGTCCACTG
		CGCCCGCTGCCCTCGGTCTTCTACATGCTGGTGTGTGGCCTGACGGTCACCGACTTGCTGGGCAAGTGCCTCCTAAGCCCGGTGGTGCTG
		GCTGCCTACGCTCAGAACCGGAGTCTGCGGGTGCTTGCGCCCGCATTGGACAACTCGTTGTGCCAAGCCTTCGCCTTCTTCATGTCCTTCT
		TTGGGCTCTCCTCGACACTGCAACTCCTGGCCATGGCACTGGAGTGCTGGCTCTCCCTAGGGCACCCTTTCTTCTACCGACGGCACATCAC
		CCTGCGCCTGGGCGCACTGGTGGCCCCGGTGGTGAGCGCCTTCTCCCTGGCTTTCTGCGCCGCTACCTTTCATGGGCTTCGGGAAGTTCGT
		GCAGTACTGCCCCGGCACCTGGTGCTTTATCCAGATGGTCCACGAGGAGGGCTCGCTGTCGGTGCTGGGGTACTCTGTGCTCTACTCCAG
		CCTCATGGCGCTGCTGGTCCTCGCCACCGTGCTGTGCAACCTCGGCGCCATGCGCAACCTCTATGCGATGCACCGGCGGCTGCAGCGGC
		ACCCGCGCTCCTGCACCAGGGACTGTGCCGAGCCGCGCGGACGGGAGGGAAGCGTCCCCTCAGCCCCTGGAGGAGCTGGATCACCT
		CCTGCTGCTGGCGCTGATGACCGTGCTCTTCACTATGTGTTCTCTGCCCGTAATTGTGAGTCCCCGGGCCCCGAGGCAGCAGGGCACTGA
		GACTGTCCGGCCGCGGATGCGGGGCGGGAAGGGGTGGA
149	ISL2	CTTCCGCCGCGGTATCTGCGTGCCCTTTTCTGGGCGAGCCCTGGGAGATCCAGGGAGAACTGGGCGCTCCAGATGGTGTATGTCTGTACC
		TTCACAGCAAGGCTTCCCTTGGATTTGAGGCTTCCTATTTTGTCTGGGATCGGGGTTTCTCCTTGTCCCAGTGGCAGCCCCGCGTTGCGGG
		TTCCGGGCGCTGCGCGGAGCCCAAGGCTGCATGGCAGTGTGCAGCGCCCGCCAGTCGGGCTGGTGGGTTGTGCACTCCGTCGGCAGCT
		GCAGAAAGGTGGGAGTGCAGGTCTTGCCTTTCCTCACCGGGCGGTTGGCTTCCAGCACCGAGGCTGACCTATCGTGGCAAGTTTGCGGCC
		CCCGCAGATCCCCCAGTGGAGAAAGAGGGCTCTTCCGATGCGATCGAGTGTGCGCCTCCCCCGCAAAGCAATGCAGACCCTAAATCACTCAA
		GGCCTGGAGCTCCAGTCTCAAAGGTGGCAGAAAAGGCCAGACCTAACTCGAGCACCTACTGCCTTCTGCTTGCCCCGCAGAGCCTTCAGG
		GACTGACTGGGACGCCCCCTGGTGGCGGGCAGTCCCATCCGCCATGAGAAACGCCGTGCAGGGCAGCGCAGTGGAGGTGCAGACGTACCA
		GCCGCCGTGGAAGGCGCTCAGCGAGTTTGCCCTCCAGAGCGACCTGGACCAACCCGCCTTCCAACAGCTGGTGAGGCCCTGCCCTACCC
		GCCCCGACCTCGGGACTCTGCGGGTTGGGGATTTAGCCACTTAGCCTGGCAGAGAGGGGAGGGGGTGGCCTTGGGCTGAGGGGCTGGG
		TACAGCCCTAGGCGGTGGGGGAGGGGGAACAGTGGCGGGCTCTGAAACCTCACCTCGGCCCATTACGCGCCCCTAAACCAGGTCTCCCTG
		GATTAAAGTGCTCACAAGAGAGGTCGCAGGATTAACCAACCCGCTCCCCCCGCCCTAATCCCCCCCTCGTGCGCCTGGGGACCTGGCCTCC
		TTCTCCGCAGGGCTTGCTCTCAGCTGGCGGCCGGTCCCCAAGGGACACTTTCCGACTCGGAGCACGCGGCCCTGGAGCACCAGCTCGCG
		TGCCTCTTCACCTGCCTCTTCCGCGTGTTTCCGCGCCCCAGGTCTCCTTCTCCGAGTCCGGCTCCCTAGGCAACTCCTCCTCGGCAGCGAC
		GTGACCTCCCTGTCCTCGCAGCTCCCGGACACCCCCAACAGTATGGTGCCGAGTCCCGTGGAGACGTGAGGGGGACCCCCTCCCTGCCAG
		CCCGCGGACCTCGCATGCTCCCTGCATGAGACTCACCCATGCTCAGGCCATTCCAGTTCCGAAAGCTCTCTCGCCTTCGTAATTATTCTATT
		GTTATTTATGAGAGAGTACCGAGAGACACGGTCTGGACAGCCCAAGGGCGCCAGGATGCAACCTGCTTTCACCAGAGACTGCAGACCCCTGCT
		CCGAGGACTCTTAGTTTTTCAAAACCAGAATCTGGGACTTACCAGGGTTAGCTCTGCCCTCTCCTCTCTACGTGGCCGCCGCTCTGT
		CTCTCCACGCCCCCACCTGTGTCCCCATCTCGGCCGGCCCGGAGCTCGCCACGCGGACCCCCGCCCTGCCCCAGCTCAGCGCTCCCTGG
		CGGCTTCGCCCCGGGCTCCTAGCGGGGAAAAGGAAGGGGATAACTCAGAGGAACAGACACTCAAACTCCCAAAGCGCATGATTGCTGGGAA
		ACAGTAGAAACCAGACTTGCCTTGAAAGTGTTTAAGTTATTCGACGGAGGACAGAGTATGTGAGCCTTTGCCGAACAAACAAACGTAAGTTA
		TTGTTATTTATTGTGAGAACAGCCAGTTCATAGTGGGACTTGTATTTTGATCTTAATAAAAAATAATAACCCGGGGGCGACGCCACTCCTCTGT
		GCTGTTGGCGCGGCGGGAGGGCCGGCGGAGGCCAGTTCAGGGGTCAGGCTGGCGTCGGCTGCCGGGGGCTCCGCGTGCTGCGGGGCGG
		GGCGGGCCCGGTGGGGATTGGGCGC

SEQ ID NO	GENE NAME	SEQUENCE
150	chr15:87750000-87751000	AGTTTGGGGAGCCTTTTCTCCATTTGAGAAAAAACAAACTTACAGCGAGGGGTGAGGGGTTAGGGTTTGGGATTGGGGAAAATGTGGGTGG GGAGCCCCCCCCAAGGAAGTGAGGAGGGGGCTGCAAGGATTACACCTGGGCATACGTTTCCCTAGAAATCACATTCATTGTATTTTTATAATT TATTCTAAATCTTTCATGCGAAGAAAGTCAGTAGTGAGTGTTAGTACTGGTGGCCCTCCTGATCACACTTGCATCTCTTGAGTGTGCCTTAAA GGTCTTGGGAATGGAAAAATATAAAAACTGCTTCGTGATGCGTCATCTTTATCCCCCACTCCCCCACCCATTCCAATATATTTTCTACTTCCAG CCTAAATTCGGGGGCCCCCCTACCGAGGCCGGCCATGATCTTGAGGGCGGCATAGGGGAGGCCGCCGCTCTGTCCACCCCCAGCCTGGTGATG CCGTTCGCTTCTTGTGCCCGGTATTGTGGGCTACATGCCTTTCCGGCGTACGGAGCTGAGCGTCCAGGCCAGTGCCCCCTCAACCTCTCAG TAATGTTTACCCGAGGCCGTCGTGCAATGAGACTATTCGCATGGCATTGTCAACGCGGCCGGCGCGCGTCTCGGCCCTCCGCGGCTTGC CAGACTGTCCTGCAAACCACCTCACCCGTCTCTTTGGCGCAGGAGACTCAGGCTGTAACCGGAGAAAACACTTCACCCTGGAACCCTAACT CAGGTCCTGGCAAAAGATGCGAGAGGAAGACTTGCTCTCTTAATAAATCTCGGCCGCCCGCACATCTGGCCCCTAGACCTGCTCGGTAGA GGACTGGCTGGTGGATGCGCGGTCCAGGCCGTGGGCACTCGACCCACCTCTATTTTCCTTCCTTCCCGAGGCGCCCCCTGGATTACCACTTTCGG TTTTGCGCTTACATCCGGGATGTCGAATTTCCCAGGGAATCATAATTATTTTATCTATAATTTATTCTAACCCCAAGGTTCCAAGAAAATCT
151	chr15:87753000-87754100	ACATTCCTTCTAAAATGTGGGCTTTCTGTGTACATGGGCGCGCATTCCCAGGACTCGGTTCCCTGGGTGGAATTCACCCAGGAATACAATC GATTTTCTGAACCTGCGTAAGGCCACAGGCAGCTCTGAAAATGAAAGCGTTTGCTAAGTGGGGGAGATCTCACCGATCGAACGTTTAAAAA TGGCTTTGTCTTCATTCAGCTCTCCCGATTTATTCTGTGTTTTACAAATAGAAGCTCAGAGCTTCTGTCGCCCAGTCCTTGCATGACTCATGG CGGTGGCCACACGGGTTTCAGGGATAACGGGATGTTTAGAAAATCGCTGCATATCGGAGTTTCCTAGCACGTTCCATTTATACTGAACGCA GGCGGCCGCTGAAAATCCAGCCTCGACTCTTGCTAATGACTGGGTAGGACCCTCGGGGTCCTGCGACGGTGCTGGAGAGGGTGTTCCCGGC TCCGATGTGGGGAGGCCTGCGCGGGGACTAGGTTCTCGAGAGGCGAGCGGGCGCGCCAGAGAACCCGAGACTGCTGCGGGGGCCGGAT GCGGGATCCCCTGGGCTGCGGTTCTACGCAGAAACGCCAATGGCCATGCCTCCTCCCCAGCCCCCAGTCACTAGGCCGGCGCCGCCT GGCCCGGAGATCCTCCCCAGAGCCCTGGCGGTGCCATCATGCCGGAGAAGACAAGCTCGGCCCCCGCTGGAATTCGCTCCACAACACAGATG CTCATTTTTGGAATATTCTAGAAAAATAACAAGATCTTGTTTGTCGTTATGATTCACGGGAGGTAACTGATGGGAGGGCCATTTACATGAGGG CAGACACTGTGGGGCGAAGGTGACTTCTGGACGTAGGCTTTAAAGTAGGAACGGCTCCAAATTCCCAATATCTCCGGCCTTACCGGTTGCA AATCGGACCCCCTGCGGGAAAACCAGACACTTCTGTTTCGTGGCTTCGGGCTGCCTCCAGCCCACGCAGGCTCGTTTAGTCCCCGTGGAG TCAGCCCCGAGCCTTCCTAGTCCTGGAACAAGGGCTCCAGGTCGCGGCCGCGGGAAGCCCGCCAAGAGGGCGGGGAGTAGGGATTCCCT CCAGCTCCGCAGGGCATC
152	NR2F2	TCCTCCTCGGCCTCAGATGTCGTCCCACCTGCCCACGAGCAGGGAACCTGGAACCCACTCTCCCGGCAGTCCCCAGCGGGTTCCGCCAC CCGGCGGCCGCCCCTGACACCGAGTGGGTGGGAGGAAGAGGCAGCTGGCGGGGATGGGCCATTGAGACCTCTTGAAAAATATTAAAAGA CAGGATGGGTAGAGATTTCTCCGGGAGAAAGTTCGAGGGTGCATCGGGTCGCGGCTGGGAGGAGTACCCGAAATGCCAGCAGGAGAAAT GCAACCTGTTTAGGCCACACCTTCAATCCCCGAGGCTGTCTGGAGAAGACTGCGTGCGGGGGACTTGCCGGCGTTCCCACACCGCGCCTG CAATCCACTCCCGCGGCTGCCTCTGCCACTCGCGGCTTGAAGCCAGTGGCTCTCAAGCCCTCGGCCCCGCGGCGCCCGCGCAG CCTTCACCCGGCGCCGGCACCACGAAGCCTGGCCGCAGTGGACTCCCCCGCAGCTCGCTGCGCCCTGGCGTCTCCCGTCGAGGAGGGAG GGACGGAGGCCTGAGCCGGGAGCTCCCTGGCGGTGGTCGGGCCGCCCCCCTTGAGGCCTGCTCCCCCCCTCTCGGCCTCGCCAAATCCC TGAAAGCCCAGTCCCCCTTCGTCACCCCGGGGGGCTTCTAATCACTCGGTATCGATTTCCCTAACTCTCTTTCATCCTGTTGAAGACACATCTT AAAACACTCCAGCCCGGAGTGTGCTCTGGGCTTTATCCCACACTAATAAAATGATTTACCCTTCTCTCCGCGCTCTCCTCACAGAGGAAAATC GTTCGAGCCCCCGGCTATTTGTGTGATCAGTAAATATTTAGTGCGCTGACATCCTTAGCTGGGCTTCGGATCGATTCGGGGGCCCACCGGG AGGTGCGCACGGTCCGGGCGGGCCGCGAGCTCGCCGAGGGGGGCTCCTCCCGCCCTCGCCGGCGGCTGATTTACGGCCCCCT GCAACCAGCTAAGGGGGGCGAAAGCGCGCCTGGAAAATTGGCTTTTCAACCTTTTACTTTTGACATTCAGCCACTTCCCCAGGCTCTAATTC TCGCCCGCACTCCTCCCTCCCGCCCTACTAAGGGTTGCCCTGTGCGCCCTGCGAGCCCTTCCAGCAGCAACGCGCGGCGCTCGCGCCCCC CTCGGCCCGGGGGACCACCTATCACAGCCCTGAGCCGCGACGCGGGGAGGCCCCCGGCCCCCTGCTATGGGGGTCGCCTCCTTCGAGGAGA

SEQ ID NO	GENE NAME	SEQUENCE
		GATGCTCTCCGCCCGCCCACACCTCTGAGGGAGGAGAGGGGGTGGAGAAGCCCAGAGCTGCATCTGCTGGATGACGAGCCGCTCTCCCT GCTACCCTTTCTCCGACCCGTCGGCCTTTCTCCTACTCTGGAGACTGATCCTCGACGTCCATCGGGCCGGATGGCGTCGGGTGGAAGCGT TACTTTCCTCGCAGAAAAACTCCTCCTCTTTCCTAAGATCAGAAAAAGCGCTTAGCTTGGAATTGTTAG
153	chr16:112343000-112349900	CCTAGGCATTCTCAGCCCGTTTTGCTGGAGGGGGCATTTGAGGCCTGGCCAGCTTAGCCAGCCTACAAGGAGTGTTACTGGGGTGAAAAC AGCCAGCGGGGACCAGTCTGCTTGTGGCCCGCCAGGTGCCTGGGATGGGGAAGCAGCAAATGCCCACCTTCCTGCCCAACCCCCTCCTC CCTCTTCATGGGGGGAACTGGGGGTGGCAGCGGCTGCCGGGTGCGAGCGGGCTCAGGCCTGTGGCCCTGCCTGACGTTGGTCCCCATC AAGCCATGTGACGAGACCAGGCCACACAAGAAAGAGGTTTCAACAAGCGTTATCGTTTCCTGGAACTCCAACTCGGCGACTTCCCCGAAGACC GGCTGTGCCTGGCGGGCGGCTGCGCACAGCGGGGACAAGGCTGCCCCCTTCCTCCTCCGCTGCCTCCGCGGCCGCGTCTATCTCAGT CTGACTACCTGGAAGCAGCACTCCACCCTCCAGCCCGCCCTCGGCTCAGCTGCCAGGTCACCGGCAACCCCCGGGAGCGGTGGGGG CAGGGGCTGCTCCGCCAGCCTCTGTGATGTTCAGGCCGGGCTGCACCAGCCCGGGACCCCTAGGTG
154	SPN	GCACTGGTTCCCCTTTACCTGAGCCAACAACCTACCAGGAAGTTTCCATCAAGATGTCATCAGTGCCCCAGGAAACCCCTCATGCAACCAG TCATCCTGCTGTTCCCATAACAGCAAACTCTCTAGGATCCCACACCGTGACAGGTGGAACCATAACAACGAACTCTCCAGAAACCTCCAGTA GGACCAGTGGAGCCCCTGTTACCACGGCAGCTAGCTCTCTGGAGACCTCCAGAGGCACCTCTGGACCCCCTCTTACCATGGCAACTGTCT CTCTGGAGACTTCCAAAGGCACCTCTGGACCCCCTGTTACCATGGCAACTGACTCTCTGGAGACCTCCACTGGGACCACTGGACCCCCTGT TACCATGACAACTGGCTCTCTGGAGCCCTCCAGCGGGGCCAGTGGACCCCAGGTCTCTAGCGTAAAACTATCTACAATGATGTCTCCAACG ACCTCCACCAACGCAAGCACTGTGCCCTTCCGGAACCCAGATGAGAACTCACGAGGCATGCTGCCAGTGGCTGTGCTTGTGGCCCTGCTG GCGGTCATAGTCCTCGTGGCTCTGCTCCTGCTGTGGCGCCGGCAGAAGCGGCGGACTGGGGCCCTCGTGCTGAGCAGAGGCGGCA AGCGTAACGGGGTGGTGGACGCCTGGGCTGGGCCAGCCCAGGTCCCTGAGGAGGGGGCCGTGACAGT
155	chr16:85469900-85470200	TGTCCGACAGGCACACAGAGCGCCGCCAGGCACGGCCCTCATTCTTCACCCCGAGCTCCCGCAAGGTCGGCGAGGAGGCTGGAGCAGC GGGTAGGAAGCGGGCCGAGGCTCCCCCGACGCTGGGCCGCAACTGTCATCGCAGATCCCTGAAAAACGAGCTCTGTAATCGTTGCCGTC AGCGGGTGTACAATTGCAGCCTTATGTTTCCTGCCGCTGTTTACCTTCCTGAGCGGCGCCCAGAGATGCACACACGCTGCCCTGAAGCGG GACGTGACCTCTGGGCACCTGTGAGGTCCTGGG
156	SLFN11	GTCGGCTCCTGCGCTCCCAACGGGGTGGCCCGTTTCCTTCCTCGCACCCTCTTCTCTCCCGGTGCCTGCGGTCCCACCTTCCAGATACCCC TCGGAGAGTCCAGCTGAGCTCTCGCCAGAGCTTTCCCCTTCCAACCCGCTCGACTTGCCCAGATCCCAAGCTGGGCTTCTCTCTCCATCGC CCCAGAAAGTGGGTCTTGGAGACCGAGGCAAGAATTTGGGCCTCCCGCTTCTGTTCCAGACCCCGGACCCCTTGCCAAATGCGGCAGATG TGCAGATTGGGCCGCGCTTGGTTCCTGGCTGGGTTTATGGAGCCTGCGGCTGAGGCAGGCTCCGCAGACCCCGAGCCAGAGTGGGATTT AACGGCGGCCCGGTGCGCTGTGCTTGGTCAACCCCGGTAACCGTCACGCTGCTAGTGATATGAAAAAAACCTGCCAGCGTTCTGCTTTTCTG CCCCGCTGCAGTCTTTAGCACCCGCCAGGATTCTGTCCGAGTGTTTGGA
157	DLX4	TTTAGTGTGTGCATAAAACATCCCAGCTAATCTCAAATAGACTTTTCCTGAGCAGAGGCTGAAATTTGCAAGTAATGCAAAGAAGACTCCGG GAGAGCGTCGCCGATGGTGGAGCGGGAGACGGGCGTGGGGAGCCCCCACTGCAGTGCTGGGATCGAAGTGGTGCTGACCCCAAGACCTC TCCCCTCCTCCTCCCCCGGGAGCTTCTCCAGGGTTATTTGGGAAATGAGGGGGGAACTCCAATCCCTGAGAAAGCGCTCAGGGGCTTGCTG AGGTGAGCGCAAATGGAAGCACAAGGCCGGGCTGGCCGTGGGCTCAGTAACCAGTCGGCTGCCCGGCTTGCGCCAGCACTAAATGCTCG ATCAGAAAGAGAAAAAGAGGCGCAATAATTCCAAATTTCAGGAAAAGTCAAATCGGAGAGGGGGGACGCAGGTCTCTTCAGACTGCCCATT CTCCGGGCCTCGCTGAATGCGGGGGCTCTATCCACAGCGCGCGGGGCCGAGCTCAGGCAGGCTGGGGCGAAGATCTGATTCTTTCCTTC CCGCCGCCAAACCGAATTAATCAGTTTCTTCAACCTGAGTTACTAAGAAAGAAAGGTCCTTCCAAATAAAACTGAAAATCACTGCGAATGACA

SEQ ID NO	GENE NAME	SEQUENCE
		ATACTATACTACAAGTTCGTTTTTGGGGCCGGTGGGTGGGATGGAGGAGAAAGGGCACGGATAATCCCGGAGGGCCGCGGAGTGAGGAGG ACTATGGTCGCGGTGGAATCTCTGTTCCGCTGGCACATCCGCCGCAGGTGCCGGCTCTGAGTGCTGGCTCGGGGTTACAGACCTCGGCATCC GGCTGCAGGGGCAGACAGAGACCTCCTCTGCTAGGGCGTGCCGGTAGGCATCGTATGGAGCCCAGAGACTGCCGAGAGCACTGCGCACTC ACCAAGTGTTAGGGGGTGCCCGTGATAGACCGCCAGGGAAGGGGCTGGTTCGGAGGGAATTCCCGCTACCGGGAAGGTCGGAACTCGGG GTGATCAAACAA
158	SLC38A10	CATGGTGCTTCAGGAAGGGAGGGACGAGAGCCCTGGGCTTGTGGTGTCCACGTGGACAGCTAATGAGGAGCCTTGCCGATGAGGAGCA TGCGTTCCCGACGGGGCGGCCGAATGCGGAAGGAGCCGCCATTCTCTCCGCCCTGACCGCGGGATTCTCTGCAGCAGATGAGAAACGGC GCTGACTCAGCAGGGTCCCTCCCAGGCCCCGAGCGGTCATCTGGTGACCCCCCGCGCTTCCCCCACGGCCCAGCCGGAGAAGGGCAAAG GGAAGTCCCGGCTCCAAGGCGCACCCAGAGATGCGGTGCATGTGGCAGGATGGCCCAGCCCCCGTCGGCAGCCCCCAGCTTCCTGCCCCT GGTTTCCTTCCTCCCACGGGCTACAGGCCTCTGATGAGCTTTGGAAAGCAGGAAACACACAGGCTAGTAACTATGAATGGGTCCAAAAAAC ACTCCTTATTACTTTAAACTACTTAGGAAGAAGCACAGCGTTGCCCAAACGCCAGA
159	S1PR4	GCGCGGGGGGGCCGGAGGATGGCGGCCTGGGGGCCCTGCGGGGGGCTGTCGGTGGCCGCCAGCTGCCTGGTGGTGCTGGAGAACTTGCT GGTGCTGGCGGCCATCACCAGCCACATGCGGTCGCGACGCTGGGTCTACTATTGCCTGGTGAACATCACGCTGAGTGACCTGCTCACGG GCGCGGCCTACCTGGCCAACGTGCTGCTGTCGGGGGCCCGCACCTTCCGTCTGGCGCCCGCCCAGTGGTTCCTACGGGAGGGCCTGCT CTTCACCGCCCTGGCCGCCTCCACCTTCAGCCTGCTCTTCACTGCAGGGGAGCGCTTTGCCACCATGGTGCGGCCGGTGGCCGAGAGCG GGGCCACCAAGACCAGCCGCGTCTACGGCTTCATCGGCCTCTGCTGGCTGCTGGCCGCGCTGCTGGGGATGCTGCCTTTGCTGGGCTGG AACTGCCTGTGCGCCTTTGACCGCTGCTCCAGCCTTCTGCCCCTCTACTCCAAGCGCTACATCCTCTTCTGCCTGGTGATCTTCGCCGGCG TCCTGGCCACCATCATGGGCCTCTATGGGCCATCTTCCGCCTGGTGCAGGCCAGCGGGCAGAAGGCCCCACGCCCAGCGGCCCGCCG CAAGGCCCGCCGCCTGCTGAAGACGGTGCTGATGATCCTGCTGGCCTTCCTGGTGTGCTGGGGCCCACTCTTCGGGCTGCTGCTGGCCG ACGTCTTTGGCTCCAACCTCTGGGCCCAGGAGTACCTGCGGGGCATGGACTGGATCCTGGCCCTCCTCAACTCGGCGGTCAAC CCCATCATCTACTCCTTCCGCAGCAGGGAGGTGTGCAGAGCCGTGCTCAGCTTCCTCTGCTGCGGGTGTCTCCGGCTGGGCATGCGAGG GCCCGGGGACTGCCTGGCCCGGGCCGTCGAGGCTCACTCCGGAGCTTCCACCACCGACAGCTCTCTGAGGCCAAGGGACAGCTTTCGCG GCTCCCGCTCGCTCAGCTTTCGGATGCGGGAGCCCCCTGTCCAGCATCTCCAGCGTGCGGAGCATCTGAAGTTGCAGTCTTGCGTGTGGAT GGTGCAGCCACCGGGTGCGTGCCAGGCAGGCCCTCCTGGGGTACAGGAAGCTGTGTGCACGCAGCCTCGCCTGTATGGGGAGGCAGGGA ACGGGACAGGCCCCCCATGGTCTTCCCGGTGGCCTCTCGGGGCTTC
160	MAP2K2	GGGCGGGTTGCCACACTGTCCCCTTTCTGCATGGGAGGAAGGGGGCTCGAGAACTGAGTCAGCCACACAAAACGAGGATGGACAGAACT CCTGAGTAGCGAGGGTGCCTGCCGGGCGCGAGGAGGAGGGGAAGACGAGGAGGAGGAATAGGGAGCACCACACATGACA GAGGGGCTGCCTCAGACCACAAAGCGCTTCCTCATCCTTTCCTCGCCCTTTGATGCCGCGCCGGCAACGTGACTCTGCGAGCAGCGGGGCAG ACGCCAGGTCTCCCTCGCAGGCGGGAAAGGGGCTCCAAGGCGGGTGCTGCCTTGCTCGGGTCACATGGCTACGTGGGGGCCTTGCTCAA ATTCACTTCCTGCCTTCATTACAAAACTGTCAAAGGGGATCGCACGTTTGCAGGGTGTCACCCAAGCATTCTGGTTTTTGCAAACGACGCTGT GCGGCAGGCGGTCTGATACCTGATGAGCTCGGTGTGGCGGGGTCGGCAGCATTTCCTCCGGGGTTTTGAGCTCTGGCCACTTCTCCTTTT GTTCCACCCAATCTCACCCACTTCTGGGCTTCGAGGCCAGAGTGTCTTAACAAGGGGGGCACGT

SEQ ID NO	GENE NAME	SEQUENCE
161	UHRF1	GAGCGAGACTTTGTCTCAAAAAAAAAAAAAAACCAAATAAATTGAAAGCTGAGAAATTCAGAGCACAAGAAGACAAGCGCGCCCCCTCTTTTA GCTGTCAACATGGCGGAGCCCGTCCCTGGTGACGCAGCCTCCAAAGGCCTCCCTGTGCCCTCCTGAGACCGCAAGAGGGAAAGTGGCAGC GACAGTGATCGTGGTGTCTTTGTGGCGGTTGTGTTGACCTCACTGACCCCCGAAGTGCCGCTCTAGGGTCTGTCCTCAGCGGTGACCCGG CCGGGTCGAAGGGCAGAGTTCCGCTGTCACTAGCCCTCCACCCGTCCTGTGTGCTGGGATGCCCTCGCGGCGCCGTCCACGCCACCGCC GCCCCCTCTTGTGGGTTCTGTCTCCTCCGGTGTCTAGGATCCTCCTGCATCCGTTTTTCCTTCCTCCCTTCTCTCCCTCCGTCTGTCTTGCCC GCACCTGAGGTTGTCGCAGAGGCGCTGAGACGGGCCAGCAGGAGCTGT
162	DEDD2	TGCTGTCCCGGTCCTGTCGCAGTCCTCAAAGATGCTAGAGTGACAGTCCTCTAGGGGTAGAGATGGTCGTCCTCCCAGGAGAAGGTGGCC CGGAGACTTGGAGGTGGGATCAATCCTGCCAGTCCTGGATCAGGAGGCCTCTGTCGGGCGCCGCCCCCTTCCTCCTCCATCAGCAACAG GCGGCGCCCGGCCAGCCTCATAGTCAGCCTCATCCACACTGACCAGCAGGCGAACAGCCTCCCGGCCCACAGCCTCTCGCAGGGCCTCAG TCAGGAACACGCCCCCGCAGGGCCTGCAGCAGGGCGCCACTCAGGTAGTCGCCCCCAGAAGGCGTCCAGATAGGAGAGCTCTGAGAACTTG ATGTCACAAACCACAGAGCCCAGGTCCCTTGAGCGCAGCACTGCGGTGGCCTGCCCAAACACGTCCAGCTGCCGCGCCAGCGCCTGGGG CCGCCGGGATGCCACGCCCTGCTCCAAGGCTGGCCCATGCTCGCTCGAACCCGGAGCCGGATGTCTGCAGGGGAAGGAG GGATTTGTCAGGGAGGGGGCCAACACTAGACACACTTATGGGGAACGCCACCCTTCCTCCCTCC
163	CDC42 EP1	TGATGCCCGGCCCCCAGGGGGGCAGAGGCGCCGCCACCATGAGCCTCTCGCCTGTGGGCTGGGTGTCCAGTTCACAGGG AAAGAGGCGGCTGACTGCAGACATGATCAGCCACCCACTCGGGGACTTCCGCCACACCATGCATGTGGGCCGTGGCGGGGATGTCTTCG GGGACACGTCCTTCCTCAGCAACCACGGTGGCAGCTCCGGGAGCACCCATCGCTCACCCCGGCAAGAAGCTGCAGCTG GTGCGGAGGGTGGGGGCGCCCCCCGGAGGATGGCATCTCCCCCTGCACCCTCCCCGGCTCCACCGGCCATCTCCCCCATCATCAAGAA CGCCATCTCCCCTGCCCCAGCTCAACCAGGCCGCCTACGACAGCCTCGTGGTTGGCAAGCTCAGCTTCGACAGCAGCCCCCACCAGCTCCAC GGACGGCCACTCCAGCTACGGTGAGGGCCTGGGCCATCTTGGCCCACTTTTCAGA

Table 6C

SEQ ID NO	GENE NAME	SEQUENCE
164	chr21:9906600-9906800	GGCCGGGCAAAAAGCCGCCGCAACAAAAAGCTGCGCTGACGGGCGGCGGAAAAAGCCGCGGCGGAGCCAAAAAGCCGGGGGCGGCAAAA AGCCACGGTGGCGGGCGCAAACAGCCGCAAAAAGCCGCGGTGGTGGGGGCAAAATCAGTGGGAGCAGGGGCAAAAAAACACAAAAAGC CGCGGCGGCGGGGGGCAAAAAGCCA
165	chr21:9907000-9907400	TGGCTTTGCTGGAGTGTGATGTGATAGGAAATGTGCAGCCAAAGACAAAAGAAGATGTAAGTAGGCTTGACTCATTGCAGCTAAGAACCCA GATGTTACCTTGAGGGTATTAACTAATAAGCAGTTTAAATCAGAATGGCACATTCTGATTTGTTTTTTGTATGTTCACATTTGGCAGGCATAGA TACTGTTTGAAAAGAGAAAAGTCAGTACATAGAGGTAACAAGCTTAAATATGTGCCAAGTCTAGAAACAAGAGACTAGGGGGATAAGGACCT TTCGAAATTAAATGCAAGATTTGAAAACTGATTGGCTGGGGGATGAGGCAAAGGCAGGTCTTTAAGGTCAATCCCTGTTTTTGCTTTAAGTTG TTAGCGGGTGGTTTTATCATATATTGTAGAA

SEQ ID NO	GENE NAME	SEQUENCE
166	chr21:9917800-9918450	TTCCTGGGAATGTCAGCTAACCTGAGCCTAGGGGCCTGAGCCCAAGGGCAGACTGAGGCTCCCCCAGCACAGGGAGGTGCTGCCTGTGA CAAGGGGTAGTGCTGGCACAGTGCAGGCTACTCCCTAGAAAGATCAGCTTGAATATGCAGGAAGAGCAGGACCCTCGGGCTGAGGCAGA GGTGGAATGGGAAGTGCATGGTGGTAATTTAGTTCTCCAGAGGCCAGAAGTAGGAGGAGCGGTTGGAATGCTGATGGCCCAAAGGGAAAC CCTGGACTACCCTGGCCTCCCACAGGACTCTCATAGTAATTGCGGCTCCCTGCAGTGGTGAGGCCAGAAGGAGTGTTGCCCAATGCTGTC ATCATCCAGTCCACCCCCCACCCACCATCAACAGATGAGTATGGTCATGAGTGTGGTCACCTCATCAGTCATTTGCTCAGTTGTGAAAAGA AATTGTTCAGAGAAGAAGAGCAAAGTGTTTTTCCATGAGCCAAAGGTCAGCCAAGTTATGCTAATGAGGAGGACTGGAGACAGCGTGTCACAGA CACCGAGAAGGAGCACTGGGCAAGGGCACTTCTCCCAGGGCAGAGCCCACAAGAAGCGTCCTGGCACCAGACACTCAGGGAACTGAAGG CTGGCAGGGGCCCGCCCAGT
167	TPTE	TCCCCCCAGCTGGGTATAAGCAAACTTTCCTGTCTATGGGCCGCAGAGACCACCATCTAGTTCCCCCGCCAAAACTTTACATGATTTTAATT CTCCTGATGAAGATGAGAGGATAACAGCCAACAGAGAGGGCAGAGGATGGGATGGGACTCCCTTGCTCAGAGACCTCACCTCTAGGTCTT TACCTCCTATTGAGAATAAGTCAGTTCTGTAGTAAGAACTCTGTGTCCACGGCAACCCCAAACAGAATCCTAGCGCTCTTGTGATTCTTGTA GAATGGGGAATAGAACGAGCTTGGCCCAAGACTGCACAGACTTAAAAACATACTATTCTTTGAAAATGGCAATCATTAAAAAGTCAGGAAAC AACAGGTGCTGGAGAGGATGTGGAGAAATAGGAACACTTTTACACTGTTGGTGGGACTGTAAACTAGTTCAACCATGGTGGAAGTCAGTGT GGCGATTCCTCAGGGATCTAGAACTAGAAATACCATTGACCCAGCCATCCCATTACTGGGTATATACCCAAAGGACTATAAATCATGCTGC TATACAGACACATGCACACGTATGTTTACTGCAGCACTATTCACAATAGCAAAGACTTGGAACCAACCCAAATGTCCAACAATGATAGACTG GATTAAGAAAATGTGGCACATATACACCATGGAATACTATGCAGCCATAAAAAATGATGAGTTCATGTCCTTTGTAGGGACATGGATGAAATT GGAAATCATTCTCAGTAAACTATCGCAAGAACAAAAAACCAAACACTGCATATTCTCACTCATAGGTGGGAACTGAACAATGAGAACACGTG GACCCAGGAAGGGGAACATCACACTCTGGGGACTGTTGTGGGGGAGGGGGAGGATAGCATTGGGAGATATACCAAATGCTA GATGAGGAGTTTGTGGGTGCAGCGCACCAGCATGTCACACGTTTACATATGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAA AGTATAATAAAAAAAAATACTGTTCTGCCATACATACAGATACTCATTAAAGATGAGGGAGAAGGGCATGGGGTGGGGGAGAATGTACCAAAA CCAAAGACCACAGGATAATAACCTCAGAGCAGAGACTATCTCTCTAGTTATTTTTTCTTTTGTATGTAATGGAGAGGATTATTATTTACTCTGA TGAAGAAGTTTACATCAAGTGTTCAGCTTCCTTTGTGGGTTACAGAGAATAACCAGAGGGCTCAGTTATGCTCTCTGAATAACTATGTTTGCT TAGTGTTTTCTAAACAATATTAAATTTCACTAAAATAGACAAGGTTGATAGGACTTGGGGGCATAACTCATTGACTCAAGCTATCATTTTATAG GATTGTGAGAAACAAATAGATGAACATTTAAAATACACTCATATTCTCGCTAGAAAAGAGGATTTTGAATATTCTTACATCAAAGACATGGTA AATGTTTAAGGCAATGAATATGCTAATTACCATGATTTGATCATTATGCAATGTAAAATGTACTGAAACATCACATTGTACCTCATAAATATGTA CAATTTATTATGTGCGAATTAAAATTTTGAGTATAAGAAAAAATAAACTTCAATTGTAAGAAAACAACCCAACTTTTAAAAAACGGGCAAAATA CGTGAACAGATACTTCACTAATAGAGATTTGCAACTGGCAAATAAGCAAATGAAAAACTGGTCATCATCACTATCTATTAGAGAAATGCAGAT TAAAACTACAATAAGAAACAATGCTGCCCGTCCAGACGCATTGTTTTGACCGTTTCCAACTTGTCCCAGCCCTTCCCGGGGCATCGCTGGG GACCCTACGCCGACGTCCCCCCTCCGCCCGCGCCCCAAGGGCCGACTGGGCAAATTGGGAGACCCGCCCGCGGGGGCGACCCAACTTT TCGGAACAGCACACCCCACACCGCCACACCCCCGGACCCCCCGCTCCCGGCGGGAGACTCAGGGAACCCCGCACCCCAAGCCCTT CTAAATCGTGCAGCGTGAGTGTGACGGCCAAGAGCGGATGCAGCCCCGGGATCGCCCGCACCTTCCCGTGGGCGGAAGCGCAGGAGCCA GCTGGGGAGGGGGCGCCCTAGAGGAGCGGCTAGAAAGCAGACACGGGGAACTCAGGTCATCCTGGGGGGGGGACAAGACAACGAGAGCC GGGCGCCTCGGGGGCGCGGGAGCCTCCGCAGGACCGGGCGGCCCCGGCTGGCGCGGGGGGGCGCGCCCCCTTTACCT GCGGCTCCGGCTCCTAGGCCATTTCCTCACGCGGCGGCCGGGACTGAGCTAACACCACTCAGGCCGGCCGGGTTTGAATGAGGAGG AGCGGGGCGCGGAGAGGGAGGGGACGGGGAGGGAGGGAGGCGTCGCGGAGTTTTTCTCGGCCTTTTGTGCGGACACCT CCCGGATTCCGCGCCGCACCCGGCCCCCCAAAAGACACGGGGAGCCGCGGGCGAGGGGTTCAGCCATCCGCCGAGGCGCCTAGTGCC TTCGCGCCTCCAAGACCCCCCCCCCAACAAAAAGGAGCGTCCCCCACCCCCTACCCCCGCCCGGAGGACTTAGGGCCTGGGCTCACCTCGG GCGCGGAGCTAAGTGTAGGCGCCGGGGGTCCCTAGAGCCGCGGGGCGCAGCGAGTCCGGCGCTGGGTAACTGTTGGGTCAGAAACTG TTCAGGTAGCAGCTGTTGTGCCCTCCCTTGGCCCCGCCGCTCGGAGACGCCCCCGCCCCTGCCTTGAACGGCCGCCCGGCCCCGCCCCA

SEQ ID NO	GENE NAME	SEQUENCE
		GCGCCCACGTGACTAGCATAGGCGCGCCCCCGTTCCGCCCGCCGCGCAGACTCCGCCTCCGGGACGCGAGCGAGCGGCGAGCGGCGC
		GCACTACCAGTTCTTGCTCGGCGACTCCCGCGCACGCGCGCCGTGCCACCCTCCCCGCACCCCTCCTCCCGCCATCCGGCTTAACGT
		GGCGGGCGCGCGCCGCGCAGTAGCCGTGACAGGTACCCGGCGGGGCGGGGGGAGGGGGTTGGCCCCGCGAGGGTGTGCGCAGGC
		ACAGACCCGGGTCCTGTCCCCGCCCCCTCCTCTGCAAGGTGTGCCTGGGCGAGGGGAGGGGCCCCGCGCCCGAACCCCTGGGTCA
		CCCCCGAATTACAAACAAAAACCTTAACGCCATTGCTCGCGGGTTAGAAGGCAGCTGTGCGTGCTCAGGAAAAGAAGCCACGCACAAGAG
		ACCGCACGCGGCGTGGATACAGTGACACGAAACACCCAAAATCTCTTTTGAAAGGGAAACCAGGCACAGTGGCTCATGCCTATAATCCCAG
		CACTTTCGGGGGCCAAGGCGCTCACCTAAACCCGAGAGTTCAAGACCAGCCTGGGCAATACAGCGAAACCCTGTCTCTACGAAAAATATAA
		AAATTAGCTGGGCATAGGGCTGGGCACGGTGGCTCACGCCTGTAATCCCAGCATTTTGGAGGCCGAGGCGGGCGGATCACGAGGTCAGG
		AGTTCCAGACCATCCTGGCTAACACAGTGAAACCTTCTCTCTACTAAAAATACAAAAAAATTAGCCGGGCGTGGTGGCAGGTGCCTGTAGT
		CCTAGCTACTTGGGAGGCTTGAGGCAGGAGAATGGCATGAACCCAGGGAGCGGAGGCTGCAGTGAGCTGAGATTGCGCCACTGCACTCCAG
		CCTGGGGGACAGAGTGAGACTCCGTCTCAAAAAAAAAAAATAATTAGCTGGGCATGGTGGCTGGCACACATGGTCCCAGCTACTCAGGA
		GGCTGAGGTGGAAGGATCTCTTGATCCCGGGGAGGTCAAGGCTGCAGTGAGCCAAGATGGCATCACCGCACTCCAGCCTGGGCCACAGA
		CCCTGTCTCAAAAAAAAAAAGAAAGTGGGGAAGAAAATGTAATACAAATTAATATACCAACAGCAATTAGTGAGTACTTTTTCCATGGAGCT
		GGGAGAGGGAATAAATGTTTGTAAAATTAAAATGTTCTACGCTAGAAATCAACTTTCCTTCTATGCTTTCTTTACTTCACCCCTTATAGCTACT
		TAGTAAATCTCACAAATCCTATCCTTCTGATCTCTCTGAAATGTATGTACCCTTTCCCTTCTATTCTCACCACCCATGTTTCTTTGTTTCCTTCT
		AGCCTGTGTAATAATCTCATAATCGCACCTCCTGTACCTGCCTTCTTTCTAGTCCAGAATACGTTTTCCTAAATTCCACCAATAACCATCCTG
		CTACTGCTTTGTGTGAAATTCTCCAAAAAAAAATTTTACTTTTCCAAAATAAGTCAGGCTCCCTCTCTTAGGATACAAAACCACACCATGGTCCC
		AGCCAATCTTTCAGCCTGATTCACTCAGTATATATTTATTGACCTCTCCTTTCTCCCAAGCACTTGGCTAGATAATAATTAAAGAGTGCGGCA
		CAAAACAAATTGGATTCCTCCCCTCATGGAGCTTGTATTTTCACAGGAAGCACAGACATTAAATAAATTAAAACACAAAAAATAGACAAGCA
		TATAATTACAGTATGTATCCTAGAGAAATATCACTCATGCAGAAAGCATACACAAGGATGCAGCACTGTTTCCAATAGCGAAAAGCTAGAAAC
		AACCTACATGTTCACCAAAAGAAAATGGCCACATAAACTATACCATATCCAAATTATCCAAATTTTAGAATATAGACAACAGGTTGGGCGCGG
		TGGCTCACACCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGGTGGATCACAAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTG
		AAACCCCGTCTCCTCTAAAAAAAACAAAAAAATCAGCTGGGCACTGTGGCAGGAGCCTGTAATCCCAGCTACTGAGGAGACTGAGGCAGGAG
		AATCGCTTGAACCCTGGAGGCAGAGGTTGCAGTGAGCCAAGATCGCGCCACTGCACTCTAGCCTGGGTGACAGAGCAAGACTCCATCTCA
		G
168	chr21:13974500-13976000	TGTAGGAGTCCTCCGGTGCTGGAGTCCAGAGCACAGTGAGGCTGGGTCCTCCCGTGCCATAGTGTAGGGCATGGCGGGACAGGGATCCT
		GCCCTGCGATAGTCCAGTGCTTGAGTCCGCAGTAAGGCAATGGTCCTCCAATGCTGGAGTTCACGGCGTTGTGGGGTCGGGGTCCTTTGG
		TGACTTAGTCCAGGGCGTACCAGGGCGGGGGTCCACAGTTGCCATAGTGAGGATCTTGGAGGAAGGTGGTTCCTGCCTTGCTGTAGTCCG
		GGGAGCAGGGGGCAGGGGTCCTCTCTTGTCAGAGTCTCTGGCGCGGGGTGGGGGTGGAGGTGGGGGTTTTCCTATGCGATAGCCCACG
		GGTCGGTGAAGCCGGGTCCTCCCGTGCCTTTGTCCAGGGCGCAGGGGGGCGAGGGTCTTCGGTGGTGGAGTCCGCGGAGCGGCAGGAC
		GGGGGTCCTCCAGTGCCATATTCCAGGGCGCGGCGGAGTGGGGGACCTGTCCTGCAGTGGTCCAGGGCATGTGGGAGTGGTGGTCCTG
		CTGTGCCTCAGTCCAGTGCGCGGTGGGACGGCGGTCCTGCTGTGCTGTAGTGCAGGACGCGGTGGCGCAGGGGTAGTCCAGAGAGCGC
		CGTGGCAGGGGGGTCCTCCAGTGCTGGAATCCAGTGCAAGGCGGGTCAGGGGTCTTACCGTGCCGAAGTCGGTGGCAAGGGTCCTCCCGT
		GCCATAGTCTAGGGGGGCGACGGGGGCAGGGTTCTCTAGTGCAGGTGTCCAGGGTGTGGCAGGGGCAGGAGTCCTCTTGTGCAGGAGTCCAG
		GACGTAGCCGAGGAGTCCTCCAATGTCAGAGTCCAGGGCTCTGCGGGGCCGGGTTCCCCCATGCCAGAGTGTAGGGCGCGTTCAGGGTGA
		GGGTCTTGGCGTGCAGTAATCCAGGGTGGGGCAGGGGTAGTCCAGACCTCCATGGCGGGCGTCCCTCTGTGCAGGAGCCCAGT
		GCCTGGCGGATCGGGGGTCCTTCTGTGCTGTAGTCCAGGGCACCGCAAGGTGTGGGTCCTCTGGTGCCCTAGTCCAGGGGGCGGCGAGT
		CAGAGGTTCTCCCGTGTCTCAGTCTAGGGCCTGGTAGGACTGGGGTCCTGGAGTCCACGTGGTAGCCCAAGTTGCCGCAGGACCAGGTA
		CTCTGGAACCACAGTCCAGGGCGCTGAGGGGCAGGAGTAGTTCAGGGCGAGCCGGGGCCCAGGTCCCTCGGGAGCCAGAGTCCAGGGTG

SEQ ID NO	GENE NAME	SEQUENCE
		TGGAGGGGTGGGGGTTCTGCAGTGGCACAGTCCAGGACACCGCGGGGGCGGGACAGGGCGGGGATCCTCCCGTGCCTTAGTCCAGGGCT GAGCCGCGGGAGAGGTCCTTCAGTAGCACAGTCTAGCGCACGGCGTTGCAGGTGTCCTCCAGTGCCTGAGGCCACGGCAGGTCGCGGGT CCCACTGTGCTCTAGTTCAGGGCGGAGTGGGTCTGAGGTCTTCTCCTGCCTCAGTCTAGGGCGCTGGAGAGCGGGGATCCT
169	chr21:13989500-13992000	GGGTTGGTCCTAGAAAGCGTGAGGATCGCCGAGTGCACTGCCCTCCAGCCTAGGGTCCACTCTTCCTTGGCCCGAGCCCAGAGCTCGG GGTTTCAGGCGCTGGGCCCTGTGCAGCTGCCCAGAATAGGCTGAGCGGCAGGTTCCCGCCCTGGCAAGGGATCCAGCAGTGGAATCCTC ACTGCTGTTGGCTGCGGGCAAGGTCAGCGGGGTTTCCATCGCTGCTGGTGGGAGCCACCTGGCGGTGGTAGCTGCAAGTGAGCGCGTGG CAGAGACTGGCAGGGCTGGTCCCAGACACCCTGAGGGTCTCTGGGTGCATCGCCCTACCACCCTAGGGTCTCGCTCTTCCTTAGCCTGCTC CCAGGACGCGGTGTACGAGGGCTAGACTCTGAGCAGCCTCCAGGATGGGGCTGAGCAGCGGATTCCTGCCCTGCTGCAGCTACAGTCTG AATTAGGCGCCACCGCAGTATCTGGCCCTGGGGTACGTGCTACTGGGTGGCATGGACAGAGATGGGGGCTGCCACAGCTGCTATGGGGC TGAGCAGCCGATTCTCGCCCTGCTGCAGCGGGGCGACCGCTGCAATCCCCAGCGCTATGGGACCGACCACCTGACTTAGATGCCTTGGAG GCATCCGGTCCTGGGGTCTTGCTGCTGGTGTCTGCGGGCAGGGTCACGGCTGCCACTACTACTGCTGTGCGCCATGGGCAGGTGCCAGC TGCAGCTGAGTCCGAGGCAGATGCTGTCTGAGGTTGCCTAAGGGTGGCTGAGTGCACCACGCTTCCACCCCAGGGTCCG TTATTCCTAGGCCGGCTCCCAGATTGCAGGGTTGTGGGCGTTGGACACTGTGCAGCCATGAGGATCTGGTTGGGTGCAGATTCCCGCCCT CCTGCAGCTGAGAAGCCAATCTCATAACAGGCGCTGCAGTGACCTCTGGCTCTGCGGTCCGCGCTGCTGCTGGAGCTGGCAGAGAACAGA GCTGCCACCGCTGCTGCTTCCAGGAGTGTGCAGCTGGCAGCTGAGCCCGTGGCGGAGGCTGGAAGGCCTTATTCCAGAAGCCT TGAGGGTCCCCGAATGCACCGCCCTCCCACCCTAAGGTCCAGTCTTCCTTGCCCGCGCCCAGAGAGTTGGATTGCAGGCGCTGAGCACAG TGCAGGTGCTGGGATGGGGCTAAGCTGAAAGTTTCCGCCCTCTGGCTGCTGCGGGGCCGACAGCCTGAGTTATGCGCCGCGGCGGCTTT TGGTCATGGGATCCGCACTGCCGGTGGCTTGCACAGGGTCGGGGGCTGCCACAGCTGCTATAGTTCACCGTGTGCACGTGGCAGCCGCC CCTGAGCCCACCGCTGAGGCTGGTCCGGTCCCAGACGGCCTGAGGGCCATTTGCCCGCGCCCAGATCCGGGTGGCTGCGC TGGGCACTGTGCAGCCTCCCGGAATCCGCTGAAGGGCACGTTCCCGCTCTCCTACAGCTGTGGGCCGACTGCCTGATTTTGGCCACTAGG TGGAGTCTGGCTCTAGGGTTTCGAGGCCGCTGGTGTTGGTGGGCGGAGTCCGGGTTTGCCACCGCTGCGCTCCATGAGCAGGTAGCAGC TGCAGCGGAGCTTTAGACCGAGGCTGGCAGGGCTGGCCCCAGACGGCCTGAGGGTCAGGGAGTGCAGGGTCCTCCCACCCTAGGTCCG CTCTTCCTTTCCCCTTACCCAGAGCGGGTTGTGCGGGCTCTGGGCTCTCTGTGCCGGCGCTGGGCTCTGTGCAGCCGCCGAGATGGGGCTG AGCAGCGGATTTCCTCCCTGCTGCAGCTGGAGGACGATTACCTGCACTAGCCGCTGAGGCGGCATCTGGCCCTGGGTTACTGCAGCTGGT GACGCGGGCAGGGTCAGGGTTGGTTGCAGGTGGCAGCTGCTGCTAAACCCATTGCGAGCCTCAGGGTCACCAAGTTCACCGTCCTTTCAT CATAGTATCTGATCTTTGGCCCGCGCCCAGAGTGCGGACTGGCCTGCGCTGGGGACTGCATAGCTTCTGGGGGCCGGTCAGCGCCAGTTT CACGTCCTCCTGCAGCTGCGTGGCCTAAGGTCTTAGGCGCCGCGGCGCTATCTGGCCCTGCTGTCGACGCTGCTGGTGGTGGGGACAGG GTCAAGGGTTGCCACTGCTGCTCCCGTGCGCCATCGGCAGGTGGCAGTTGCAGATGAGCCCACAATTGAGGCTGTTGGGGCTGCTCCCA GGTTGTTAGAGGGTCGCCGAGTTCACCGACATGCCACCCTAGGTTACGCTCTTGGCCCGCACCCAGAGCGCCGGGTTACGGGTCCTGGG CCCTGTGCAGCCACGGGGATGGTGCTGAGTGCAGGTTCCCGTCTTCCTGAGATGCGGGGGCGACCACTGGAATTAGCCTCTGTGGTGGTAT CTGACCCTAGGGTCCGAGCTGCTGGTGGCGCGGGGTCGAAGTCGCCTCTGTTGCTGCGGCGTGCCATTTGCACCGTCCTCTGGTA C
170	chr21:13985000-14000100	AAATACTCTACTGAAAAAACAGAAATAGTAAATGAATACAGTAAAGTTTTAGAATACAAAATACAGCATAGAAAAATCAGTCGCATTTCTATACC CAACAGCATACCATCTGAAAAAGGAATCAAGAAACCAATCCCATTTAAAATAGCTATAAAAAAATGCCTGGGAATAAACTAAGCCAAATAAAT ATGTCTAAAATGAAAACTATAAAAACATTGATAAAAAATCAATTGAAAAAGATACAAATAAAGGGAAAGTTATCCCATTTTTATGAATTAGAAGTAT TAATACTGTTAAAATGACCATCATACTCAAATCAGTCTATAGGTCCAATACAAATTTCCAATGTAATTCTTCAGAGATGTTTAAAA AAGGTTTTAAAAATCGTTCTGCGGATGTTAAAAGGATTTTTAAAACGCTTTTTTCGTTCTGCAGGCGAAGGCTGTGGCCGTGCTCCCGCCGG CCAGTTCCCAGCAGCAGCGCATTGCCCCTGCTCCACGCCTTCGCTCCAGGGGCGCAGCCCCGCGGGAATCAGCACTGAGCC GGTCCCGCCGCCGCCCCCAGTGTCCGGGCTGCGACTGCGGGGAGCCGATCGCCCAGCGATTGGAGGAGGGCGACGAGGCCTTCCGCCA

SEQ ID NO	GENE NAME	SEQUENCE
		GAGCGAGTACCAGAAAGCAGCCGGGCTCTTCCGCTCCACGCTGGCCCGGCTGGCGCAGCCCGACCGCGGTCAGTGCCTGAGGCTGGGG
		AACGCGCTGGCCCCGCGCCGCCGACCGCCTCCCGGGTGGCCCTGGGCGCGTTCTGTGTCGCCCTGCGGCTCGAGGCGCTGCGGCCGGAGGAG
		CTGGGAGAGCTGGCAGAGCTGGCGGGCGGCCTGGTGTGCCCCGGCCTGCGCGAACGGCCACTGTTCACGGGGAAGCCGGGCGGCGAG
		CTTGAGGCGCCAGGCTAGGGAGGCCCGGCCCTGGAGCCCGGCGCCCCCGCGACCTGCTCGGCTGCCCGCGCTGCTGCACAAGCCG
		GTGACACTGCCCTGCCGGGCTCACGGTCTGCAAGCGCTGCGTGGAGCCGGGGCCGAGCGGCCACAGGCGCTGCGCGTGAACGTGGTGCT
		GAGCCGCAAGCTGGAGAGGTGCTTCCCGGCCAAGTGCCCGCTGCTCAGGCTGGAGGGTCAGGCGCGGAGCCTGCAGCGCCAGCAGCAG
		CCCGAGGCCGCGCTGCTCTCAGGTGCGACCAGGCCCTGTAGCTGTGACTTGGCTGTGGGGCTGGCCCGCCTCCCTGACCCCTGTCAGGCG
		GAGCAGCTGGAGCTGACCCACGGGCCTGGGCTTTCGAGCGCTTTGTCCAGGCGCTAATGATGGGAAGGTGAAAGGTGGGGGTGGCCACA
		CCCTGCAGTCAGGGTGGCAGGTGTCAGAGGCCACATGCAACCCACTGGTTTTGTCTTTTCCAGGATGCTGATAAGTTTCCCGCGGCCCCC
		GGAGCAGCTCTGTAAGGCCCTGTAATTGCCTTTCGTTCCCTTCTGCTCTATTGAGGAGTGGGAAGATGACAAAGTGTTTTTGCTCAACCCGA
		AGGAAAATGCACATGGGAGGACACACCGGGTTACTATTTGAGTAGCCCAGACAGGAGAGCAGCGGTCTGCT
171	chr21:140170000-140185000	TGGGTGGATTGCTTGAGCCCAGGAGTTCGAGACCAGCCTGGACAAAATGGCAGAAACTCCATGTCTACAAAAAATACAAAAATTAGCCGGG
		CATGATGTTCTGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGTGGGAGGATCGCTTGAGCCCAGGAGGCGGAGTTTGCAGTGAGCT
		GAGATGTCACTGCATTCCAGCCTGGGAGACAGAGCCAGACTCTGTCTCAAAAGAAAAAAAAGAAAAAAAAAAAAGAAAAAACGAAATT
		GTATTCTGAATACATCATCTTCTAAAACACTACATTTACTTGCACTATATTAAACTGGTTTTATCCTGACCACAATTGCAGGTGAAAGATACCACTG
		TTGTTCTATTTTTTCTGGTAAGTAGAGTGAGCCATGTCTTCCCCAGGGAAAGACGCCTCCTAAAAATTTGTAGGACCACCTTTGGTTTTCTTCC
		AGATATTTTTTTTTGTCATCGCTTTTCCTGCGCCCAATTCCCATCTGTCTAGCCCTTCTGCCTCCGCTGGTCTTTTTCGGCGAGCCTCTCCCCAG
		CCGCAGGTATTCGTCTGGGCTGCAGCCCCTCCCATCTCCTGGGGCGTGACCACCTGTCCAGGCCCCGCCCCCGTCCAACCCGCGGAGAC
		CCGCCCCCTTCCCCGGACACCGGGTTCAGCGCCCGAGCGTGCGAGCGCGTCCCCGCTCGTCGCCCGGCTCGGCGTCGGGAGCGCGCTC
		TGTGTGGTCGCTGCTGCAGTGTTGTTGTGGCTGTGAAGAAGGCGGCGGCGGAGCAGCAGCCGGACCAGACTCCCTAGTAGCTCA
		GGCGCTGCCCTGCCGCCGGCCCTGGTGAGATGGTGGAGGAGGAGGCTGTGCCGTGGCTTGCTGTGTCCTGCT
		GCCTGGTTAGAACCCCATCCCCGTCCCCCGTCTCCTCCGGGGGGGTGAGGAGGAGCTGGAAGAGGGGCCGGCCTCTGTCCGGCCCGGCC
		AGGCGGCAGTCACCCTCTGAGGAGGCAGCGCCCGGGGAGAGGGGCCTCCCAGGCGGCCGCCGCCGCCAGGGGGAGGCGCTGGGAGTGG
		GAGTGGGAGCGGGACCTCAGCTCGGCCCGGACCCTAGGTGCGGGGGAGGCGGGGTCCCGGGCTCGGGCTGCCTGCCCGG
		ACCTGGCGGGGATGGCCCGTGCGGCTCCGGGTGTGGGACGTACCCTCAGAGCGCCCGGGGTTATTCCCACTGACTCCAGGGAGGGTGA
		GTGTGCGCCCCTTCGCTCCCTGCCGTGTCTGTGAGGGTCCATCGTTGCCGGAGACTGGAGGTCGGGGGCCATGGGAGCCCCGGGGCGAA
		CGGTGCGGACATGGGCCTTGTGGAAAGGAGGAGTGACCGCCTGAGCGTGCAGCAGGACATCTTCCTGACCTGGTAATAATTAGGTGAGAA
		GGATGGTTGGGGGGCGGTCGGCGTAACTCAGGGAACACTGGTCAGGCTGCTCCCCAAACGATTACGGT
172	chr21:140564000-140581000	GTCTCTAGGACACCCTAAGATGGCGGCGAGGGAGACGGTGAAGGTTGGCTCCCGCCTGTCTGGGCTCTGATCCTCTGTCTCCCCCTCCCC
		CTGCGGCCGGCTCATGGCCTGGCGGAGGCCCGAACCAAAGACCTCCGCACCGCCGTGTACAACGCCGCCCGTGACGGCAAGGGGGCAG
		CTGCTCCAGAAGCTGCTCAGCAGCCGGAGCCGGGAGGAACTGGACGAGCTGACTGGCTAGGTGGCCGGCGGGGGGACGCCGCTGCTCA
		TCGCCGCCTGCTACGGCCACCTGGACGTGGTGGAGTACCTGGTGGACCCGTGCGGCGCGAGCGTGGAGGCCGGTGGCTCGGTGCACTT
		CGATGGCGAGACCATGGAGGGTGCGCCGCCTGTGGGCGCGGACCACCTGGACGTGGTGCGGAGCCTGCTGCGCCGCGGGGCCTCG
		GTGAACTGCACCACGCGCACCAACTCCACGCCCCTCCGCGCCGCCTGCTTCGAGGGCCTCCTGGAGGTGGTGCGCTACCTGGTCGGCGA
		GCACCAGGCCAACCTGGAGGTGGCCAACCGGCACGGCCACATGTGCCTCATGATCTCGTGCTACAAGGGCCACCGTGAGATCGCCCGCT
		ACCTGCTGGAGCAGGGCGCCCAGGTGAACTGGCGCAGCGCCAAGGGCAACACGGCCCTGCACAACTGTGCCGAGACCAGCAGCCTGGA
		GATCCTGCAGCTGCTGCTGGGGTGCAAGGCCAGCATGGAACGTGATAGCTACGGCATGACCCCGTTGCTCCCGGCCAGCGTGACGGGCC
		ACACCAACATCGTGGAGTACCTCATCCAGGAGCAGCCCGGCCAGGAGCAGCTCATAGGGGTAGAGGCTCAGCTTAGGCTGCCCCAAGAA
		GGCTCCTCCACCAGCCAGGGGTGTGCGCAGCCTCAGGGGGCTCCGTGCTGCATCTTCTCCCCTGAGGTACTGAACGGGGGAATCTTACCAA

SEQ ID NO	GENE NAME	SEQUENCE
		AGCTGCTGTCCCACCAGCCGGGAAGCCTTGGAATTGCTGGGATCTACCTATGTGGATAAGAAACGAGATCTGCTTGGG GCCCTTAAACACTGGAGGCGGGCCATGGAGCTGCGTCACCAGGGGGGTGAGTACCTGCCCAAACTGGAGCCCCCACAGCTGGTCCTGGC CTATGACTATTCCAGGGAGGTCAACACCACCGAGGAGCTGGAGGCGCTGATCACCGACGCCGATGAGATGCGTATGCAGGCCTTGTTGAT CCGGGAGCGCATCCTCAGTCCCTCGCACCCCGACACTTCCTATTGTATCCGTTACAGGGGCGCAGTGTACGCCGACTCGGGGAATATCGA GTGCTACATCCGCTTGTGGAAGTACGCCCTGGACATGCAACAGAGCAACCTGGAGCCTCTGAGCCCCATGAGCGCCAGCAGCTTCCTCTC CTTCGCCGAACTCTTCTCCTACGTGCTGCAGGACCCGGCTGCCAAAGGCAGCCTGGGCACCCAGATCGGCTTTGCAGACCTCATGGGGGT CCTCACCAAAGGGGTCCGGGAAGTGGAATGGGCCCTGCAGCTGCTCAGGGAGCCTAGAGACTCGGCCCAGTTCAACAAGGCGCTGGCCA TCATCCTCCACCTGCTCTACCTGCTGGAGAAAGTGGAGTGCACCCCCAGCCAGGAGCACCTGAAGCACCAGACCATCTATCGCCTGCTCA AGTGCGC
173	chr21:14070250-140705050	TAAAAATAAATTGTAATAAATATGCCGGCGGATGGTAGAGATGCCGACCCTACCGAGGAGCAGATGGCAGAAACAGAGAGAAACGACGAG GAGCAGTTCGAATGCCAGGAACGGCTCAAGTGCCAGGTGCAGGTGGGGGCCCCCGAGGAGGAGGAGCGCGGGCCTGGTGGCC AAGGCCGAGGCCGTGGCTGCAGGCTGGATGCTCGATTTCCTCCGCTTCTCTCTTTGCCGAGCTTTCCGCGACGGCCGCTCGGAGGACTTC TGCAGGATCCGCAACAGGGCAGAGGCTATTATT
174	chr21:14119800-14120400	CGCCACCACGTGCGGGTAGCGCCGCATCGCCCCAGCCGTGTTCCTTGGTCTCCGTCTCCGCCGCGCCCGCCTGGTGAACTGGAGCACAG GGACCATAGTTCTGGAAATTTATCCTTTTTCTCTCCATGGATTCAGCAGCAGTGTCTAAAAGAAAAAAAATTCATCAATCATTTATGTATATTTTA ATATAAAGGGTAAAACACTGCGAACCAGTGGAACCGGATAGAAAGTAATTCAGTTTTTACAGAACACAACTGTTTTTCAGGCTCTTTTATTAAAT ATAAAAGAGCCATATATATTTCTGTGGAATTCCCCTTTTACTTAAGAATTCATTATCAGCGAATTAGTTTAAGGAGGCTGTTTTTGTTAGAGGCT GTGGTTTGCATTCAAAAATTGGAATAGGAACAATGACTTGTAAAAATTCAACATTTTATTTTTATTTTTGAGATGGAGTCTCGCTCTGTCGCCCAG GCTGTAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCTCAGCCTCCCGGGTTTAAGGAATTCTCTGCTTCAGCCTCCTGAATAGCTGGGA TTACAGGCGCATGCCACCACGCCCAGCTAATTTTTTTTTGTATTT
175	chr21:14304800-14306100	CCCTGAACAGTCAGAGTTTACTGCCCACTTTTGCTGGAGGAGAAGCTCCCTGAACAACTAGAGAGACTGTGGTTCCCAAAGAGCAGCCTGTA GGCCTGAGGACTGCTCTATGACCGGCGTCAGTCCCTGCCTCCCTCCCGTCCCTCCTTCCCTCCCAGGCCTTCTCTGACTA CCAGATCCAGCAGATGACGGCCAACTTTGTGGATCAGTTTGGCTTCAATGATGAGGAGTTTGCAGACCATGACAACAACATCAAGTGAGTC CACTTGGATGCCCCCCTGCACGAGGCACGACTCCCCCTCCTCGCTGCTGAAGTCCCATGGGGGCAGCTCCCTTAGTCCTTGCCGGGAGATA ACAGGTGTTTCCAGTTGCATGAGGGTGCTGAGGCCCCCCAGTGAGAACCAGGGGAGGAGCACTGAGGCCTCAGATGAGCACCGGGGGGAGG AGCCCTGAGGCCCCAGATGAGCACCAGGGGAGGAGCACTGAGGCCCCAGATGAGCACCGGGGGAGGAGCGTTGAAGCCCCAGATGAGC ACCAGAGGAGGAGCTGAGGCCCCAGATGAGCCCCGGGGAGGAGCTCTGAGGCCCCAGACGAGCACCGGGGGAGGAGCGCCGAGG CCCCAGATGAGCACCGGGGGAGGAGCGCCGAGGCCCCAGATGAGCAGTGGGGGAGGAGCCCCCGAGGCCCCCAGATGAGCAGTGGGCG GGGCAGGGAGCGCCGAGGCCATCCCCCCTTGCTCTTGCAGCGCCCCCATTTGACAGGATCGCGGAGATCAACTTCAACATCGACACTGACGA GGACAGTGTGAGCGAGCGGGGCTGTGCGGGGTCATGCAGGCACCCTGTTCCCAGGCCAGCTCAGGCCCGCCCATGGCTCGGTCTGTGG TGGGCCTGTGCGGTGGGGCTGGGAGAGGCCCCCTCTGTGGAGCTAGGAACAGTCGCTTTTCTTGACCCTCCCCATCATGCCCTCCAGCCCA TGGCGCCCCACATCCTGAACTAAGCCCCCTCTGGGAGCCCTGTGGGGAGAGCGCCTCCTGTCTCCCCCCAGACCCTCTGGAAACTGACCTTGG CGTTTTACTCTGCAGCCCAGCGCGGCTCTGAGGCCTGCTGCAGCGACCGCATCCAGCACTTTGATGAGAACGAGGACATCTCGGAGGACA GCGACACTTGCTGTGCTGCCCAGGTGAAGGCCAGAGCCAGGTGCGGGGCCTGCCCATCCCCCCAAAGCCTCTGCCGAGGAGGTGCAGC CCCCAGAACACCCGTCAGATGCCCCAGACGCCCTGCTGTTTGTTATGCCGG

SEQ ID NO	GENE NAME	SEQUENCE
176	chr21:15649340-15649450	TTTGGGCCACGAGGCAAGTTCAAAGCGGGAGACTTTTGTTTTATAAAATGATGGTGAGCAGCTCCGGTTTTATGTCAAACATCAGGGTTTCG TGCAGGATATAAACATTT
177	C21orf34	ATTGCCGTACTTTGCTTCCCTTTGTATGTATTTCTTGTATGCTGCCGAGTCACTGATGGCTAGCTCTGTCTGGCAAGTAATTCAAAAATGCTG TTTATGTAGAAAGGAAAGGTAGGGACTTTACCACACACTCTGTCATTAAAGGGAGCAATTGAAGAACAAAGGAACTGAGTAAATACCTATATATT GCCTTTTGTGTTGCGAAACACTGTAGCACAAACACATTTGTGTTCAGCCAAATGTTTTACTTCCTTTTGTAATAACGCATATAGTAGGTTGTCT CCACATATGTACAAGAATCCATATTTTATTTAAACGTATATAGTCAATTGTTCATATTTATAGGCTGCAAACATTTCTCAATCTCAAAGACTTTT ACATATCCACTCCCACACAGCTATTTGTTATTATTTTAAAAGTTCTTAAATTAAAAAAAAAAAATAAAATATACTAATATCTCTGTTGGTTGATTTT ATTAAGCAACTTAGGATTTCAACACAGTTTAAATCATATTGATGACTCAGATCCTGGCAGGTCTTACAATTCCTGTGAAATGAGAGCACAGCT AATAAAAATATTAAGCAATTACTTTTTATTAAAATCATAGGGTTTTTTTTCATTATCACATAGAAATGATTGATCTATACAGATTGGTCTCACTCAT GTGTCTTTTGGGCTGCTTGGGAGCTTCATGTAGAAGTGGAAAGTCCCCTTTGCTCTTCCTCGACCAAGGTGGGGAAAATGAAGGCATAGA ATACAATCTAGGGCTATTAAAGAATTGCTGGCATTACTTCTCTCTATCACGTGTGAGCCTGGCTGCCTGCTTCCTGAGGTAGGGGATCCAGG ATGAGACTGTGCCGGAGCCTGTTTCCACAACTGCATTTGGAGATCCGTCTTATTGATTAGCGGGGGGAAAGGGGTGGGGATCAGGAGTGTG AGGTGAGGGGAGGACCAACTGACGACTGGCTCAATGAAGCACAAGACATTTTCTTCCGGAAAGATGTCAAACAACTGAGAAACAGCCAGAG TGCTTTGCCCTCGCCTCTCGCTGTGTATCTGTATGCCTGTGTGCGCGCGCGTGCTCGCTCGGGCATCCGTGTCTAGCCGAGGGGAGGGGGT AGGAAGTAGAAAGGTGGAAAAATGAGGAGACCCTGGAAGAAATGAAGGCATTTCCTATGAGACAGCCTTGGGGCTTTTTTCTTTTCTTTCTT GGCGTGTGAGTGCGTGGAGGGTAAAAGCCAGTCAGTCAGTGAGAAGCAAAGGTACGTTGGAGAGCAACTAAAATCTGACTGATTTTCCATCT TTTTTTTGCTTCCATCATCTGACCTGCAAAGGCTAGAGTGACAGCGTCATGCAAATGCTGCAGTCCAGCAGGTCTGGGAGAGAGGGTGGATGC TTGGAGCATCAGATGTATTCCC TAGACTGTGAGTTAATGTTAATGATGAGCGCAGTGAAAATACCAGCCGCTGCCACCCCCTGCTCACAGAAGCGCTCTGAGTCAGCATCAGA	
178	BTG3	GCAGCCTCCTCCTGAAAAATGTAAGCCATTTCCACTTTGTAAAGCTACGTTTATATTCCACCACGATACGATGGAAAAGAAAACCCAAGGCA ATTTAATATACGGGTTGGGAAGAAAGTTTTGCTGATGGAACTACATTAGCCTCCACTCCAGCAAAGCAAACAAGGAACCACACTAAAGAAAT GTACTGAATCTTTTAA
179	CHODL	TGCCTGAGCGCGCAGAGCGGCTGCTGCTGATCCAGGACCAGGGCGCACCGGCTCAGCCTCTCACTTGTCAGAGGCCGGGGAAGAGA AGCAAAGCGCAACGGTGTGGTCCAAGCCGGGGCTTCTGCTTCGCCTCTAGGACATACACGGGACCCCCCTAACTTCAGTCCCCCAAACGCG CACCCTCGAAGTCTTGAACTCCAGCCCCCGCACATCCACGCGCGGCAGGGAGGCAGGGAGGCTGCAGAGTCAGAGTCGCGGGCTGC CAGGGGGTCGGGCAGCTGGGCTCGGGCGGCGGGAGTAGGGCCCGGGCAGGGAGGCAGGGAGGCTGCAGAGTCAGAGTCGCGGGCTGC GCCCTGGGCAGAGGCCGCCCTCGCTCCACGCAACACCTGCTGCTGCCACCGCGCCGCGATGAGCCGCGTGGTCTCGCTGCTGCTGGGC GCCGCGCTGCTCTGCGGCCACGGAGCCTTCTGCCGCCGCGTGGTCAGCGGTGAGTCAGGGGGCCGTCTCCCCGAAGAACGAGCGGGGAG AGGGGACCACCGGGGCGCGGCGGGCAGCCTGTTCTCGGGCGGAGGCTCTCCGGGGGCGTTGGAAACCTGCATGGTGTAAGGACCCGGGAG GAGGCGGGGAGAAATTGATTGTGCTGTTCTCCTCCCTCTCTTCTCTAACACACACGCAGAAAAGTTTAAATTTTTGTGAAGCCGCTTGCTTAC GTAGCTGCGGAGCGAGCCTCTGCTTCATTACGAGCGGCATAGCCTTTTTCAGGAGTGATTTCCACTTTCTTTGTGAGAGAGTTGACCACAC	
180	NCAM2	TTCAATTTACACTCGCACACGCGGGTACGTGGGTGTTCGGGGTAGGGCACTGATCTGGGGAAGGTCTCCCCCCCCGCGACCCAACTCATCT TTGCACATTTGCAGTCCTCCCTCGGTGCACTCCTGGCGGGGATCTGGCCAGTGCAGCGCACTGGGACCGAGGGCAGAGGCCCGCGGAGTG AGGCCAGGAGAGACTTCAGGCCTCTAAGGACACAGCTGAGGCTAAGGCTGAGTTGAACGCAGCCCCTCCCGCGGCTCGTCCCCTCTCCA

SEQ ID NO	GENE NAME	SEQUENCE
		GTGTCTCTCCCGTAAGGTGCCGCTCCCAACAGCAATGGGTCGAGATGTAGAGGAAACACTCTGTACGTTATTTTTCCGCCCACCCTTTAGC GCCTGAGGAGACAGACAGTGTAGACTTTAGGGTACAATTGCTTCCCCTCTGTCGCGGCGGGGTGGGAAGCGTGGGAAGGGGACAGCCGC GCAAGGGGCCAGCCTGCTCCAGGTTTGAGCGAGAGAGGAAGGAGGTCCACGGAGAGACAAGAATCTCCCTCCTCCCACGCCCAAAA GGAATAAGCTGCGGGGCACACCGCCCGCCTCCAGATCCCCCATTCACGTTGAGCCGGGGCGCG
181	chr21:23574000-23574600	TCATTATCCGATTGATTTTCCTGGTATCACATCACTTAAGTTTAAGTAGCTCTTATGTTACTTAGTAATGACTGCAAAACACGAGTTGTGATGC GGGCAATTTGGATACAACAAAAAGAAGCCATTAAGTTTGTTCGTTAGTTAACAGGTGAAAGCTCTCAAGTTATTAAGGATAAAAATGCTAGTA TATATATATGGTTTGGAACTATACTGCGGATTTTGGATCATATCCGCCATGGATAAGGGAGGAATACTATAATCAGGTTTGTTTTAAATTCC ATGTCTAATGACTTCGTTATCTAGATCACCTGTAGAGCTGTTTTTATTGTAGGAGTTTTCCTTGGTTTTAATCTTTTGATTTGTTTTTCATGTTA ATACTGAAAATTTTTAAAAATTGCATATTGTACTTCCTATATGAAAATTTTACTATGTATTTTTATTTTTCCTTTTCCTTTAGGAAGAATTAG TTTGTTCCCTGACAGAGTTAGAGTAAGGGCAAATTACTTGTCTCTATAAACAACTCAGATGTTTTGAGCCGGTGTTGTAGGGGTTATCTTTTT CTGGTTTTGCATTTTATTATAGGACATAGTGCTT
182	chr21:24366920-24367060	AGAAAGAAGAAATCCGGTAAAAGGATGTGTTATTGAGTTTGCAGTTTGGTGTTTGATCTTGCACAGATTTTCTCAGGGGCCTTAAGACCGGTG CCTTGGAACTGCCATCTGGGCATAGACAGAAGGGAGCATTTATACGCC
183	chr21:25656000-25656900	CGAAGATGGCGGAGGTGCAGGTCCTGGTGCTCGATGGTCGAGGCCATCTCCTGGTCCGCCTGGCGGCCATCGTGGCTAAACAGGTACTG CTGGGCCGGAAAGTGGTGGTCGTACGCTGCGAAGGCATCAACATTTCTGGCAATTTCTACAGAAACAAGTTGAAGTACCTGGGTTTCCTCC GCAAGCGGATGAACACCCACCTTTCCCGAGGTCCCTACCACTTCCGGGCCCCCCCAGCCGCATCTTCTGGCGGACCGTGCGAGGTATGCC GCCCCACAAGACCAAGCGAGGCCAGGCTTCTCTGGACCGCCTCAAGGTGTTTGACCGCATCCCCACCGCCCTACGACAAGAAAAAGCGGAT GGTGTTCCTGCTCCCTCAAGGTTGTGCGTCTGAAGCCTACAAGAAAGTTTGCCTATCTGGGCGCCTGGCTCACGAGGTTGGCTGGAAGT ACCAGGCAGTGACAGCCACCCTGGAGGAGAAGAGGAAAGAGAAAGCCAAGATCCACTACCGGAAGAAGAAACAGCTCATGAGGCTACGG AAACAGGCCGAGAAGAACATGGAGAAGAAAATTGACAAATACACACAGAGGTCCTCAAGACCCACAGACTCCTGGTCTGAGCCCAATAAAGAC TGTTAATTCCTCATGCGTGGCCTGCCCTTCCTCCATCGTCGCCCTGGAATGTACGGGACCCAGGGGCAGCAGCAGTCCAGGCGCCACAGG CAGCCTCGGACACAGGAAGCTGGGAGCAAGGAAAGGGTCTTAGTCACTGCCTCCCGAAGTTGCTTGAAAGCACTCGGAGAACTGTGCAGG TGTCATTTATCTATGACCAATAGGAAGAGCAACCAGTTACTATTAGTGAAAGGGAGCCAGAAGACTGATTGGAGGGCCCTATCTTGTGAGC
184	MIR155HG	GCCTGAAGACCATTTCTTCCTCTCTTAGGGACCTGCTGGTCTCCAGCTGATTCGGTCCAGGAGGAAAAACCTCCCACTTGCTCCTCTCGGG CTCCCTGCAAGGAGAGAGTAGAGACACTCCTGCCACCCAGTTGCAAGAAGTCGCCACTTCCCCCTCCAGCCGACTGAAAGTTCGGGCGAC GTCTGGGCCGTCATTTGAAGGCGTTTCCTTTTCTTTAAGAACAAAGGTTGGAGCCCAAGCCTTGCGGCGCGGTGCAGGAAAGTACACGGC GTGTGTTGAGAGAAAAAAAATACACACACGCAATGACCCACGAGAAAGGGAAAGGGAAAACACCAACTACCCGGGCGCTGGGCTTTTTTC GACTTTTCCTTTAAAAGAAAAAAGTTTTTTCAAGCTGTAGGTTCCAAGAACAGGCAGGAGAGGGGGGAGAAGGGGGGGGTTGCAGAAAA GGCGCCTGGTCGGTTATGAGTCACAAGTGAGTTATAAAAGGGTCGCACGTTCGCAGGCGCGGGCTTCCTGTGCGCGGCCGAGCCCGGGC CCAGCGCCGCCTGCAGCCTCGGGAAGGGAGCGGATAGCGGAGCCCCGAGCCGCCCGCAGAGCAAGCGCGGGGAACCAAGGAGACGCT CCTGGCACTGCAGGTACGCCGACTTCAGTCTCGCGCTCCCGCCCTTTCCTCTCTTGAACGTGGCAGGGACGCCGGGGGACTTCGGT GCGAGGGTCACCGCCGCCGGGTTAACTGGCGAGGCAAGGCGGGGGCAGCGCGCACGTGGCCGTGGAGCCCGGCCTGGTCCCGCGCGCCGCC TGCGGGTGCCCCCCTGGGGACTCAGTGGTGTCGCCTCGCCCCGGGACCAGAGATTGCGCTGGATGGATTCCCGCGGGCAGAGGCAGGGGG AAGGAGGGGGTGTTCGAAACCTAATACTTGAGCTTCTTTGCAAAGTTTCCTTGGATGGTTGGGGACGTACCTGTATAATGGCCCTGGACCAG

SEQ ID NO	GENE NAME	SEQUENCE
		CTTCCCTGTTGGAGTGGCCAGAGAAGTGTGTAAAACACACTAGAGGGGCAGGGTGGAAAAAGAGACTGCCTTCAAAACTTGTATCTTTTCG ATTTCATTTTGAAAAATAACTACAAATCTATTTTAATTTTACAAAGTTAGACTCATAGCATTTTAGATATCAATGTCTTCATTTAACAGAAGTGAA GATGGAGCAAACGCTCAATCAGCGTCTGTATTTATTCGCTCCTGTTGTGCCAGGGTGCGTTTTTTGCCGAGCGGTTGCCTTTCTTTACTCACA AAACCCCCTTGATGTCTGTCCTCCACGTTTTACGAGGGAGAGCCGGATCTTTTTGAAGTTTGTATCATCTAAAGCAGGTATATTGGGATGACT ATGGATAGAATTTAACCTGAAAACACTGAAGTTGACAGCTGACAAAG
185	CYYR1	CATAACAAGAGTCATTCTAATGTGATTATAAAGGACCCGAAGCTTTGCTTTTAAAATTCAATACTTAGGTAGAAAGAAAATGATAACTTTTTCC CTTTGATTTTTTATTCACTATTTTTATAACACTAGCAGCCCTGAGACACCGGATTGGAAATATCTATGCCTCTTGATGTTACCTGGGCACCACT GCATCACAGTCCT
186	chr21:26938800-26939200	AATAGTAATTGCCAACAGTCAAGATATGTACTACCACCAAATTCCGTGTTATTTGTGATCAAAAGATATACACAGATACTTGAAAACTGATTTC TACGTTGCATATGGGAAAAATACCTCATTTTTCTCAGCTGTCCATTATTTTTGAGATATTATGTGCAGTGATAGTAAGAACAAGCAGATTTGGA ACACATCAGCAATAATTTTTTCAATCAGAGTCCTGCCAAAATGAAAGAATTTGACAGTATCCGGCACCCTGTACTCATGCTTGGCTTCTGTAG AAACTGTGGCTTGCAAAAGGGCAGCTGGGTACTGTGTTTTGGTACCTCATTCTTTAAACGTATAATGGGAATCTGGTTGGTTCAGGAAAACC CTTGCCTACTTATTATTACTCTGTTTT
187	GRIK1	GGCCCATACTTAATGTATTTTTAAACGTTTTAACATTTACTAATATAGAACCTTCTATTGCCTATTTCCTTCTGGTTTATTCCCTTTCCTTCTGT CATTGAAGAAATGGTTCTAGTGGTAGAAATACTCCACGATTGAGAAGAATGTGGGAAGAAAGGAGGGCTGGTGGGTAAGAATTGCTCATGA TGTCTCCCTCTGAATTCTGTGCTCTCACAATGACACTCCAATGTGTGGTTTGACGCCTGGAAGA
188	chr21:30741350-30741600	TGCTTCAACCGGAAATGTGGTTGAATTACCCTTACAGTGAACCTGATCAGTGGTAACAGGAGATGCTAGAACAGGAAAAGACAAGTTTCCCC TTTCCTCCCTATCCCATCAATTACTTTGAGGTGTATTTTTTCTTTGCAACCCCTCCAGAGAAGTCGGCAATGTTTAACGAGCATGCCTGCCAA GTGGCTTGCCTTATACCTCATTATGAAGTGATACTCAGGGCCACTAACACATCGCACACAGCATTGC
189	TIAM1	TATGATTCCCCTCGATTTCCCTCAATCTTAACCATTGTGGATCACAGCAGGAGGGCCAGAAAGTGAGCTTCAGCCTGGCACCGGGACCTCAG CCTCTCCCTTAAACTTTCCCTAATCCTCGGAGCTAGTGTTACTCAAGTGACTCCACAGTGTTGCCCGATCCCTTCAGACATGGCCTTGATGA TCTCCAAAACTCATGCTACCTTTGCCAGCCTAAAGCATCCACTCTGTGCCCCAAAACGTGAATGTCAAATACCCTTCAAGGCAGAAGGCTAT TTCTATTTTTGTTTTGTTTCTGTTTAAGGCAACAATCACCAACATTTGGTACACATGAGCCATCCTGTGAAACATCAAGGCGCTTCGTTGGCAG CAAGTCAACTTCGGTTTCAGAAGAAAGCTGCACTATTTCCTGAGGTTAGAGGTTTAAACCAAAACAAGACAACCACACATTTTAACCCCAAATCT GCCGACTGAGGGTAACCATGATCCTTCCTTCACAGCACC
190	TIAM1	TACTAAATCAACCCAAACCCGAGAACCCGGTCATGGAGAAATAAATGATAGTAATCTATGCTGTTCATCTGTTCCATCACTCACTCACTCTCT TGCTGAACAAGAAAGGGCCACCCATGTAGCAAACCACACATGTAAAGAGCCGGGAAGAC
191	TIAM1	TATTATTTTGTTCAAAGTAGACGGGTATACTAACATCTGTGGGCAAGTTTACCACACGCCACTTAAAACAGGCTAACAGGGTCATATGCCAAA ACGTTCAGGTTTGCATTTTTGAAAAGCTCAGAGATCTGACAGATGTGTTCCGGCCGCGATTTAACATGCGGCTCCAGTGAGAAGGAAGCAG ATATGACAAATGGTTCACTTATTTCAGAACTAAAACCCCAGAGGAGCAGCCTGAGCCAAAAAGGGAAGTGATCAATGGAAAAGACGGTCGA ATCTGCTCACAGGCAAGGCAAGGGG

SEQ ID NO	GENE NAME	SEQUENCE
192	SOD1	AAGACCTGGAGTTTCCATTACACCGAATTGGCACTTAATAACTGTTGTCGGAGCATTTCTTCTTAAGCCACATTTTCGTAAAGTGGCTTTAAAATT GCTCTGCCAGTAGGCAGGTTGCTAAGATGGTCAGAGACAAACTTCTGAACGACTCTTGTAAAATATACAGAAATATTTTCAGAACTTTTATCA GTAAAATTACAAAACGTGTTGCAAGGAAGGTGCTTGTGATAACACTGTCCCCAGAACCTTAGTGAAGTTACCAACTGGTGGAAAATTTTCTCT TGCACTCGGCTTAAAAATCAT
193	HUNK	GCAGGGGTGACTGGTCCTCTCTCTGCACCTCGCAGGATTTCTCTGGAAGATCTGAGCCCGAGCGTCGTGCTGCACATGACCGAGAAGC TGGGTTACAAGAACAGCGACGTGATCAACACTGTGCTCTCCAACCGCGCCTGCCACATCCTGGCCATCTACTTCCTCTTAAACAAGAAACT GGAGCGCTATTTGTCAGGGGTAAGTGCGACCCTAGAGGCGATCGTCTCTGCTGTCTGTGGAAAAAAGAGCTCCTACACCCAAAGTGCTTCT CAGTTGCTGACACTTGATCCAAGCTGCTAATTTAATCTAATGTGAGGCTGAGTTTTCTGAATGTGGGATAAAGTCGTAGCTAAACCTGCTTCT CAGGGAGTGCCTTTTTATCTGCAAATGTTTTTTCAAAT
194	chr21:32722000-33273300	AAGTAACGGGATCAAATTAATTATTATTTTGGTGGCCGCCTCTCTTCTCCACCCCAAGCCAGGCAAGACTCACCCTCGGCCCTGCCCGCCC CAGCATTTCAAATGGAATACCTAGGTGGCCCAGGGGGACCCCCTGACCCCCTATATCCTGTTTCTTTCTGCCTGCTTTGCTACTTTTCTCCTTG ATAAAAGGAGAGAGTGAGAGATAATTAACAAAAAAACATGGCCCCCAGGACAATGAAACAACTGGCCTTGGCCGGCCAGAAATGTATCCTGGT TTTCTAGGTGAACTTTCTCCCATCAATCTTTCCTTTAACCTCTCTGTTTAGTGGAAGCAATAGGAACACCCCTCCCCTCCCCTGAGCAAATGCT TTCTTTTGACTGGAAACAAAACAGGGGCTCGGCGAAGGGCTGAGGTGAAATCTGGGTGGCATGGGCGCCGCACAATGGGGCCGCTGTTCCC CGGCCCGGGCTTGTGTTTACAACAGGGGAGGGGCGGGCGTGAATGGTCTGATGATTGGAACAATCCCCCCGATTCAGGCCTACAAACGC ATCTTCTGTTCCACACCGAGGGGACAGAAAGGAGAAAAGTGACAAAGAACGCCGGGGCGGGGGAATTAAAACAAAATGCCGCTCGACTAAA AAATCTCTCATATCCTGCATATTCCAGAAAGCGGCTCTATGGAGAGAGGCCTTCAGGAGGCCTCAGCCATATCTGAATGGCTTTCTCTGGCCT CTGATTTATTGATGAAGCTGAAGCGACTTGCTGGAGAAAGGCCTGGAGCCTTCTTTGTCTCCGAGATGAAGTACAATAGGCCACAGGGCGG AGATCTCTTGTGATGCTCTCGGGTCCTGCCTTTCTCTTGCCCTCTCCTCCCTGCAAATACCAGCAGCGGTGACAAACGATTGGTGGTGTGC CTGGGAGAGCCGGTGACAAGACTGGGCCACTTGAGGTCTCCTTAAGAGGGTATTATGGCCAGGGCGACGTTTGTGCTGTGAAGATGGCAC ACTCCATTTTGTCAATGGCTCTCATCGGCCCAGATAATCGCCCCCTGCCTGTCAGGGGCGCAGCCGGCCGATTCATGGCGCCCTCG GAGAAAGTA
195	OLIG2	GTCTTTCCCGCCCCCTTGTCTAAACTCAAAACCGAGTCCGGGCGCGCCTTGCAGGGCGCCCGAGCTCTGCAGCGGCGTTGCGGGCTGAA CCCATCCGGCACAAACTGCGGGCCACTGGCCCCTCACACCTGGGAGTTTGCGGCGCTGGCCTGCAGCCCGGGGCCCACGTGGCGGAAG CTTTCCCGGGCGCGCGCTGCGCAGCCCCGCGGGGCCGGGGAGACACCGCTCGGGAGTCCTCCGGCTCGGCTGCAGAATCTTTATCAGCT GCACTTTACCGCAGCCCTGGCTAGGACGCTAGGCGGTGGAGCGCCCTATCCAGGTGCGCCGCCGCACCATGGATCACCGGCGCCCGGTCC CGCAGTCCCGCCATGGCCTGGGGAGGCCCGAAGCCCGGGGACAGTGGCCGGCCCATCTCCGGCTCCGCGGACCCCCGGCTCAGGCGG GAGGGCAGGCGGGTCCCTGCAGGCCCCCAGGGAGCCCCGGGAGCCTCTCTCTGGCGTCATTCAGTCCCGGGGCAACCTGAAGCGCGGTA GATATTGGAGAGGGGGGCGTCTGTTGGGGGGACCTGGCGTCATTACTGATGGCTAGCAGGGAGGAGGGAACGGGTTGTCACCTCGGCCTC ATAAGGCCGTGAGTGAGTAGTCCAGGGCCTCTTCAGGCATTTTTGAAACTGGATTAACTAGGGGGGGAAATTGTAGCACTGAAGCCACCGTG ACTGTCTTTTGCGCTGTGTGGAAACTCCGGTAAAACTCTTTGGGCAACACAGTCTTATCACCCAGCTCTTCAACGTGTGCAGCCCTTCTGGTCCT GTCCCTGTTCTGGGCCCCAGGAATGCAAAGCAGGTCCAGGCACTGTGAAGACCCTGGCCGGTGGAGGAAGAGGCTTCCCGGCTGTGGAGG AAGCCAGACCCTTACAACACAAGACGAGAACCAGACCTGCGTGGGGGAGCTCTGGATGCTACAGGGGCTCAAGGAGAGGGGTGGAGGGGCC TTCCCAGGCCAACCCCTGAACGGCTTGGACAAGATGCTCAGATGGACGGGAGGAACGGCGTGTGGGATGGGGGAGCTGGAGGCGGGTG GGTGGGGGGGGGAGGATGGGGAAAGCGCTGGCCCACCCAGTGTGGGAGGGGTAGAGGAAAAGCCCGCAGGGGCCAGGTTGGGACCCC GTAGGCCGGGTTAGAGGGGCTTGGACTTGATCCTGACAGGCGACAGGGAGACATATTGCTACTTATTATGTGCACAGTGGCCAGATCTCTAA AGAAAACACCATCCCCCCACACCCCCCATATAGTAAACCAGGTGGTCCGCCCAGTGCTCCCAGGGAGGTGATGGGAAATCCCACTCC ATACCCTGCGGTGAGGGGTTCCATGCCCTCCACGTGTGCAACTACTCCGGGCCCCAGGGAAACACTGGGCCCCCATCCGGTAACCCCCGGC

SEQ ID NO	GENE NAME	SEQUENCE
		CCAGTCGGGTTTCCCAGTTCACATTATAACCAAACGGTCTTGCCAGCTAGACAGACACCCCTGACCTGTTTACCCTGATCCTCTGCTC
		TCAGGATTAATCACAACTTGTCGAAGGGGGTGGCTTCCAGTGGGGTGGACCGCTCTGTCAATGCCAGCGTGTGTCTAGCATCTCCTGGGG
		TGGGGGTGTGGGGAAGGGAGGTGTAGGATGAAGCCCTAGAAGCCTCAGGCAATTGTGATCCGGTGGGCTGGATACTGAAGCCCACCCCT
		GCCTTGACCTCAATTTTCAGTATCTTCATCTGTAAAATGGGAACAACCTGCCTTCCCTCCTAGCCCTAAAGGGGCTGCTGTCAAGATTGGCTG
		AGATAGCTGTTTGCAAGCTGAGCTCAATGAAAGTTCATTGTGTCCCCCTCAGTCCTATCCCAATATCGTCTCACTGCAAAGGTGGGGGGCA
		GCTTAACTTCAAGGGCACTTCAAGGATAGCCAGGTGGCTGTCAGCCCAGCTTTCCAGGATGGGAGCAGGATCTTGACAGAAGGGTTGACT
		GGGAGGGGCAGTTGCTGGTTTGGGCTTCGTTAGGTTGCATTTTTGTTTGTTCCTTTCATTTCCCTGGGGCAGCACACCCCTTCCTGCAAGCT
		CCAGGCCTTCCTCTGGAATGCTCCTAGAGCCCCAACCTCTGCTGGTGCCTGAGCTTAAGCCAGGCCAGCTAAGGGGATCCTGGATTCACAC
		GGCCTCACAGTCACTCAGATTGTTAGCAGAAGACAAAAATTACAAGGGGAGGGCGTCATGTGATTCTTACACACCCTCCAAATCCAGCAGA
		CACCTTGGAAGCCACAGGTAGCTTCAAGAAACCCATTTTACGGATGAGAAGGAGACAACTGGAGATCTCTGAGTCTCT
		GAGCCCACACTCCCTACCTCCCTGCACCTCTGCTGGCAGGATCTTGGGCAAATGCCCACAGCTCTCTGAGAGTCAGTTTTCC
		TGTCTGTAAAATGGGAGTCATACCTTCCTCCTATGGCCGGTGAGAGACTAAATTAAACTATGTCTGTCAAGACACCTGAAACTCCTGGCACA
		ATTTAGGTTGCCTTCAAGTGGTCACAGTTGTCATTAGGTGGAAGTCAACACCCCAATCATTGTAAAGGTGCCCATATACCCCAAGATCCAGA
		TTACAGCTCTCACAGTTTATATACAGCGAAAAAACACATAACACACCTTTGCCCACATTTACATGTATTTTACGGACCATGTTTCACATCA
		GTCCGCATGCACATCTGCACGGTGTGTGCATTCGGCAGTATTTACCAAGCACCTGCCAAGTGCCAGGGCCTGTCCTCCGCACCCGGCGTGA
		ACTGTCCTGGACCAGTCCCGGGAGCCGCGGTTCTGACCAGCCGTGCTGACCCTGGACGACTCCATGAGCTGTTTTGTGAGAAAGACACGC
		CATTTGTTTGCAGAGTTCTGACTTCTGAGGGGTCATGTAGCACATGTTTGGTAGCCAAACGCTGTCATTCACGACCAGGAGCGATGGCTGC
		AATGCCTTTTTCTTTGCTTTCCGGTGCCGGAGCCTTGCCTCCCGCCGCCACCCCTGGTCAGCTCTGCGCAAGAACGTCGTTCTGT
		TTGGCAGCCAGGCCGAGACGCAGCCTGAATGTGAGCAGGAACTCGGAGAAGGGAAGGGAGAGAATCAGAAAGAAGGCCCGGGAGGGAC
		CCGGGAAGCAGTGGGAGGTCTGCGCCCTGGAGCCCCGCGAGAGCCCGGTTTGGCACGGGCTCCTCCCGGGCCGCCCGGCGGTCC
		AACAAGGCCGGCCCCGACACGCACCCGGTCTTTTGTGGGAGAGAAACACAAAGAAGAGGGGAAAAACACGGAGAGGAGGCCAACAGCACCA
		GGACGCGGGGGGCCAACCAGGAACTCCCGGAGCCGGGGCCCATTAGCCTCTGCAAATGAGCACTCCATTCCCCAGGAAGGGGCCCCAGCT
		GCGCGCGCTGGTGGGAACCGCAGTGCCTGGGACCCGCCCAGGTCGCCCACCCCGGGCGCCGGGCGCAGGACCCGGACAAGTCCTGGG
		GACGCCTCCAGGACGCACCAGGGCAAGCTTGGGCACCGGGATCTAATTTCTAGTTATTCCTGGGACGGGGTGGGGAGGCATAGGAGACA
		CACCGAGAGGTACTCAGCATCCGATTGGCACCAGGGCCAAGGGAGCCCAGGGGCGACACAGACCTCCCCGACCTCCCAAGCTACTCCGG
		CGACGGGAGGATGTTGAGGGAAGCCTGCCAGGTGAAGAAGGGGCCAGCAGCACAGAGCTTCCGACTTTGCCTTCCAGGCTCTAGAC
		TCGCGCCATGCCAAGACGGGCCCCTCGACTTTCACCCCTGACTCCCAACTCCAGCCACTGGACCGAGCGCGCAAAGAACCTGAGACCGCT
		TGCTCTCACCGCCGCGAAGTCGGTCGCAGGACAGACACCAGTGGGCAGCAACAAAAAAAGAAACCGGGTTCCGGGACACGTGCCGGGCGGC
		TGGACTAACCTCAGCGGCTGCAACCAAGGAGCGCGCACGTTGCGCCTGCTGGTGTTTATTAGCTACACTGGCAGGCGCACAACTCCGCGC
		CCCGACTGGTGGCCCCACAGCGCGCACCACACATGGCCTCGCTGCTGTTGGCGGGGTAGGCCCGAAGGAGGCATCTACAAATGCCCGAG
		CCCTTTCTGATCCCCACCCCCCGCTCCCTGCGTCCGAGTGACAGATTCTACTAATTGAACGGTTATGGGTCATCCTTGTAACCGTTG
		GACGACATAACACCACGCTTCAGTTCTTCATGTTTTAAATACATATTTAACGGATGGCTGCAGAGCCAGCTGGGAAACACGCGGATTGAAAA
		ATAATGCTCCAGAAGGCACGAGACTGGGGCGAAGGCGAGAGCGGGCTGGGCTTCCTAGCGGAGACCGCAGAGGGAGACATATCTCAGAAC
		TAGGGGCAATAACGTGGGTTTCTCTTTGTATTTGTTTATTTTGTAACTTTGCTACTTGAAGACCAATTATTTACTATGCTAATTTGTTTGCTTGT
		TTTTAAAACCGTACTTGCACAGTAAAAGTTCCCCAACAACGGAAGTAACCCGACGTTCCTCACACTCCCTAGGAGACTGTGTGCGTGTGTGC
		CCGCGCGTGCGCTCACAGTGTCAAGTGCTAGCATCCGAGATCTGCAGAAACAAATGTCTGAATTCGAAATGTATGGGTGTGAGAAATTCAG
		CTCGGGGAAGAGATTAGGGACTGGGGGAGACAGGTGGCTGCCTGTACTATAAGGAACCGCCAACGCCAGCATCTGTAGTCCAAGCAGGG
		CTGCTCTGTAAAGGCTTAGCAATTTTTTTCTGTAGGCTTGCTGCACACGGTCTCTGGCTTTTCCCATCTGTAAAATGGGTGAATGCATCCGTA
		CCTCAGCTACCTCCGTGAGGTGCTTCTCCAGTTCGGGCTTAATTCCTCATCGTCAAGAGTTTTCAGGTTTCAGAGCCAGCCTGCAATCGGTA
		AAACATGTCCCAACGCGGTCGCGAGTGGTTCCATCTCGCTGTCTGGCCCACAGCGTGGAGAAGCCTTGCCCAGGCCTGAAACTTCTCTTT

SEQ ID NO	GENE NAME	SEQUENCE
		GCAGTTCCAGAAAGCAGGCGACTGGGACGGAAGGCTCTTTGCTAACCTTTTACAGCGGAGCCCTGCTTGGACTACAGATGCCAGCGTTGC
		CCCTGCCCCAAGGCGTGTGGTGATCACAAAGACGACACTGAAAATACTTACTATCATCCGGCTCCCCCTGCTAATAAATGGAGGGGTGTTTA
		ACTACAGGCACGACCCTGCCCTTGTGCTAGCGCGGTTACCGTGCGGAAATAACTCGTCCCTGTACCCACACCATCCTCAACCTAAAGGAGA
		GTTGTGAATTCTTTCAAACACTCTCTTCTGGAGTCCGTCCCCTCCCTTGCCCGCCCTCTACCCCTCAAGTCCCTGCCCCCAGCTGGGGG
		CGCTACCGGCTGCCGTCGGAGCTGCAGCCACGGCCATCTCCTAGACGCGCGAGTAGAGCACCAAGATAGTGGGGACTTTGTGCCTGGGC
		ATCGTTTACATTTGGGGCGCCAAATGCCCACGTGTTGATGAAACCAGTGAGATGGGAACAGGCGGCGGGAAACCAGACAGAGGAAGAGCT
		AGGGAGGAGACCCCAGCCCCGGATCCTGGGTCGCCAGGGTTTTCCGCGCGCATCCCAAAGGTGCGGCTGCGTGGGGCATCAGGTTAGT
		TTGTTAGACTCTGCAGAGTCTCCAAACCATCCCCATCCCCCAACCTGACTCTGTGGTGGCCGTATTTTTTACAGAAATTTGACCACGTTCCCTT
		TCTCCCTTGGTCCCAAGCGCGCTCAGCCCTCCCTCCATCCCCCTTCGAGCCGCCCTTCTCCTCCCCCCTCGCCTCCTCGGGTCCCTCCTCCA
		GTCCCTCCCCCAAGAATCTCCCGGCCACGGGCGCCCATTGGTTGTGCGCAGGGAGGAGGCGTGTGCCCGGCCTGGCGAGTTTCATTGAGC
		GGAATTAGCCCGGATGACATCAGCTTCCCAGCCCCCCGGCGGGCCCAGCTCATTGGCGAGGCAGCCCCCTCCAGGACACGCACATTGTTC
		CCCGCCCCCCCGCCCCCCGCCACCGCTCGCCGCCTGCCACCGGGCTATAAAAACCGGCCGAGCCCCCTAAAGGTGCGGATGCTTAT
		TATAGATCGACGCGACACCAGCGCCCGGTGCCAGGTTCTCCCCCTGAGGCTTTTCGGAGCGAGCTCCTCAAATCGCATCCAGAGTAAGTGT
		CCCCGCCCCCACAGCAGCCGCAGCCTAGATCCCAGGGACAGACTCTCCTCAACTCGGCTGTGACCCAGAATGCTCCGATACAGGGGGTCT
		GGATCCCTACTCTGCGGGGCCATTTCTCCAGAGCGACTTTGCTCTTCTGTCCTCCCCACACTCACCGCTGCATCTCCCTCACCAAAAGCGAG
		AAGTCGGAGCGACAACAGCTCTTTCTGCCCAAGCCCCCAGTCAGCTGGTGAGCTCCCCGTGGTCTCCAGATGCAGCACATGGACTCTGGGC
		CCCGCGCCGGCTCTGGGTGCATGTGCGTGTGCTTGCTGCGTGGTCGATGGAGATAAGGTGGATCCGTTTGAGGAACCAAATC
		ATTAGTTCTCTATCTAGATCTCCATTCTCCCCAAAGAAAGGCCCTCACTTCCCACTCGTTTATTCCAGCCCCGGGGGCTCAGTTTTCCCCACAC
		CTAACTGAAAGCCCCGAAGCCTCTAGAATGCCCACCCCGCACCCCGAGGGTCACCAACGCTCCCTGAAATAACCTGTTGCATGAGAGCAGAGG
		GGAGATAGAGAGAGCTTAATTATAGGTACCCGCGCTGCAGCTAAAAGGAGGGCCAGAGATAGTAGCGAGGGGGACGAGGAGCCACGGGCC
		ACCTGTGCCGGGACCCCCGCGCTGTGGTACTGCGGTGCAGGCGGGAGCAGCTTTTCTGTCTCTCACTGACTCACTCTCTCTCTCCCTC
		TCTCTCTCTCATTCTCTCTCTTTTCTCCTCTCCTGGAAGTTTTCGGGTCCGAGGGAAGGAGGACCCTGCGAAAGCTGCGACGACTAT
		CTTCCCCCTGGGGCCATGGACTCGGACGCCAGCCTGGTGTCCAGCCGCCCGTCGTCGCCAGAGCCCGATGACCTTTTTCTGCCGGCCCGG
		AGTAAGGGCAGCAGCGGCAGCGCCTTCACTGGGGGCACCGTGTCCTCGTCCACCCCGAGTGACTGCCCGCCGGAGCTGAGCGCCGAGC
		TGCGCGGCGCTATGGGCTCTGCGGGCGCATCCTGGGGACAAGCTAGGAGGCAGTGGCTTCAAGTCATCCTCGTCCAGCACCTCGTCG
		TCTACGTCGTCGGCGGCTGCGTCTCCACCAAGAAGGACAAGAAGCAAATGACAGAGCCGGAGCTGCAGCAGCTGCGTCTCAAGATCAAC
		AGCCGCGAGCGCAAGCGCATGCACGACCTCAACATCGCCATGGATGGCCTCCGCGAGGTCATGCCGTACGCACACGGCCCTTCGGTGCG
		CAAGCTTTCCAAGATCGCCACGCTGCTGGCGCGCAACTACATCCTCATGCTCACCAACTCGCTGGAGGAGATGAAGCGACTGGTGAG
		CGAGATCTACGGGGGGCCACCACGCTGGCTTCCACCCGTCGGCCTGGCGCACTCCGCGCCCTGCCGCGCCGCCACCGCG
		CACCCGGCAGCAGCAGCGCACGCCGCACATCACCCCGCGGTGCACCACCCCATCCTGCCGCCCGCGCAGCGGCTGCTGCCGCCG
		CTGCAGCCGCGGCTGTGTCCAGCGCCTCTCTGCCCGGATCCGGGCTGCCGTCGGTCGGCTCCATCCGTCCACCGCACGGCCTACTCAAG
		TCTCCGTCTGCTGCCGCGGCCCGCCCCGCTGGGGGGGCGGGCGGCAGTGGGGCGAGCGGGGGCTTCCAGCACTGGGGCGGCATG
		CCCTGCCCCTGCAGCATGTGCCAGGTGCCGCCGCGCACCACCACGTGTCGGCTATGGGCGCCGGCAGCCTGCCGCGCCTCACCTCCG
		ACGCCAAGTGAGCCGACTGGCGCCGGCGCGTTCTGGCGACAGGGGAGCCAGGGGCCGCGGGGAAGCGAGGACTGGCCTGCGCTGGGC
		TCGGGAGCTCTGTCGCGAGGAGGGGCGCAGGACCATGGACTGGGGGTGGGGCATGGTGGGGATTCCAGCATCTGCGAACCCAAGCAAT
		GGGGGCGCCCACAGAGCAGTGGGGAGTGAGGGGATGTTCTCTCCGGGACCTGATCGAGCGCTGTCTGGCTTTAACCTGAGCTGGTCCAG
		TAGACATCGTTTTATGAAAAGGTACCGCTGTGTGCATTCCTCACTAGAACTCATCCGACCCCCGACCCCCACCTCCGGGAAAAGATTCTAAA
		AACTTCTTTCCCTGAGAGCGTGGCCTGACTTGCAGACTCGGCTTGGGCAGCACTTCGGGGGGGGGAGGGGGTGTTATGGGAGGGGGACAC
		ATTGGGGCCTTGCTCCTCTTCCTCTTTCTTGGCGGGTGGGAGACTCCGGGTAGCCGCACTGCAGAAGCAACAGCCCGACCGCGCCCTCC
		AGGGTCGTCCCCTGGCCCAAGGCCAGGGGCCACAAGTTAGTTGGAAGCCGGCGTTCGGTATCAGAAGCGCTGATGGTCATATCCAATCTCA

SEQ ID NO	GENE NAME	SEQUENCE
		ATATCTGGGTCAATCCACACCCTCTTAGAACTGTGGCCGTTCCTCCCTGTCTCTCGTTGATTTGGGAGAATATGGTTTTCTAATAAATCTGTG GATGTTCCTTCTTCAACAGTATGAGCAAGTTTATAGACATTCAGAGTAGAACCACTTGTGGATTGGAATAACCCAAAACTGCCGATTTCAGG GGCGGGTGCATTGTAGTTATTATTTTAAAATAGAAACTACCCCACCGACTCATCTTTCCTTCTCTAAGCACAAAGTGATTTGGTTATTTTGGTA CCTGAGAACGTAACAGAATTAAAAGGCAGTTGCTGTGGAAACAGTTTGGGTTATTTGGGGGTTCTGTTGGCTTTTTAAAATTTTCTTTTTTGG ATGTGTAAATTTATCAATGATGAGGTAAGTGCGCAATGCTAAGCTGTTTGCTCACGTGACTGCCAGCCCCATCGGAGTCTAAGCCGGCTTTC CTCTATTTTGGTTTATTTTTGCCACGTTTAACACAAATGGTAAACTCCTCCACGTGCTTCCTGCGTTCCGTGCAAGCCGCCTCGGCGCTGCC TGCGTTGCAAACTGGGCTTTGTAGCGTCTGCCGTGTAACACCCTTCCTCTGATCGCACCGCCCCTCGCAGAGAGTGTATCATCTGTTTTATT TTTGTAAAAACAAAGTGCTAAATAATATTTATTACTTGTTTGGTTGCAAAAACGGAATAAATGACTGAGTGTTGAGATTTTAAATAAAATTTAAA GTAAAGTCGGGGGATTTCCATCCGTGTGCCACCCCGAAAAGGGGTTCAGGACGCGATACCTTGGGACCGGATTTGGGGATCGTTCCCCCA GTTTGGCACTAGAGACACACATGCATTATCTTTCAAACATGTTCCGGGCAAATCCTCCGGGTCTTTTTCACAACTTGCTTGTCCTTATTTTTAT TTTCTGACGCCTAACCCGGAACTGCCTTTCTCTTCAGTTGAGTATTGAGCTCCTTTATAAGCAGACATTTCCTTCCCGGAGCATCGGACTTTG GGACTTGCAGGGTGAGGGCTGCGCCTTTGGCTGGGGGTCTGGGCTCTCAGGAGTCCTCTACTGCTCGATTTTTAGATTTTTATTTCCTTTCT GCTCAGAGGCGGTCTCCCGTCACCACCTTCCCCCCTGCGGGTTTCCTTGGCTTCAGCTGCGGACCTGGATTCTGCGGAGCCGTAGCGTTCC CAGCAAAGCGCTTGGGGAGTGCTTGGTGCAGAATCTACTAACCCTTCCATTCCTTTTCAGCCATCTCCACTACCCTCCCCCAGCGGCCACC CCCGCCTTGAGCTGCAAAGGATCAGGTGCTCCGCACCTCTGGAGGAGCACTGGCAGCGCTTTGGCCTCTGTGCTCTTTCCT
196	OLIG2	CCGGCACGGCCCGCATCCGCCAGGATTGAAGCAGCTGGCTTGGACGCGCGCAGTTTTCCTTTGGCGACATTGCAGCGTCGGTGCGGCCA CAATCCGTCCACTGGTTGTGGGAACGGTTGGAGGTCCCCCAAGAAGGAGACACGCAGAGCTCTCCAGAACCGCCTACATGCGCATGGGG CCCAAACAGCCTCCCAAGGAGCACCCAGGTCCATGCACCCGAGCCCAAAATCACAGACCCGCTACGGGCTTTTGCACATCAGCTCCCAAAC ACCTGAGTCCACGTGCACAGGCTCTCGCACAGGGGACTCACGCACCTGAGTTCGCGCTCACAGATCCACGCACACCGGTGCTTGCACACCG CAAGGGCCTAGAACTGCAAAGCAGCGGCCTCTCTGGACCGCCTCCCTCCGGCCCTCCTGAGCCCTACTGAGCCCTGCTGAGTCCTGGAG GCCCTGTGACCCGGTGTCCTTGGACCGCAAGCATCCTGGTTTACCATCCCTAC
197	RUNX1	GGACGCGGCCCGCTCTAGAGGCAAGTTCTGGGCAAGGGAAACCTTTTCGCCTGGTCTCCAATGCATTTCCCCGAGATCCCACCCAGGGCT CCTGGGGCCACCCCCACGTGCATCCCCCCGGAACCCCCCGAGATGCGGGAGGGAGCACGAGGGTGTGGCGGCTCCAAAAGTAGGCTTTTGA CTCCAGGGGAAATAGCAGACTCGGGTGATTTGCCCCCTCGGAAAGGTCCAGGGAGGCTCCTCTGGGTCTCGGGCCGCTTGCCTAAAACCCT AAACCCCCGCGACGGGGGCTGCGAGTCGGACTCGGGCTGCGGTCTCCCAGGAGGGAGTCAAGTTCCTTTATCGAGTAAGGAAAGTTGGTC CCAGCCTTGCATGCACCGAGTTTAGCCGTCAGAGGCAGCGGTCGTGGGAGCTGCTCAGCTAGGAGTTTCAACCGATAAACCCCGAGTTTGA AGCCCGACAAAAAGCTGATAGCAATCACAGCTTTTGCTCCTTGACTCGATGGGATCGCGGGGACATTTGGGTTTCCCCGGAGCGGCGCAGG CTGTTAACTGCGCAGCGCGGTGCCCTCTTGAAAGAAGAAACAGACCAACCTCTGCCCTTCCTTACTGAGGATCTAAAATGAATGGAAAGA GGCAGGGGCTCCGGGGAAAGGGAACCCCTTAGTCGGCCGGGCATTTTTACGGAGCCTGCACTTTCAAGGACAGCCACAGCGTGTACGAAG TGAGGAATTCCTTTCCACCAAGAGCGCTCATTTTAGCGACAATACAGAATTCCCCTTCCTTTGCCTAAGGGAGAAAGGAAAGGAAACATTAC CAGGTTCATTCCCAGTGTTTCCCTGGAGTAATGCTAGAATTTACTTTTGTCATAATGCAAAATTAAAAAAAAAAAAATACAACGAAGCGATAC GTTGGGCGGATGCTACGTGACAGATTTTTTCCAAATTTTGTTGCGGGGAGAGGGAGGGAGAATTGAAAACGGCTCACAACAGGAATGA AATGTA
198	RUNX1	TTTTTAATGCTCAGAGAAGTTCGTATTACTGATTCGGGAACACTGAGTTTTTCAGCTCCTGTAAAACTATTTTCAGGTTTATTTTCAAGTACAT TCTTTA
199	RUNX1	CACCCTAGAGGCAAGGACGGGGTCTGTGTCAAGAGGCTTCCCAGAGAAGTGAAAACTCTGCAGGTGCAGCCGCTGGGAGAGCATCAAGA AGGGCAGGGTGGAGGGGCAGGGGGCGAAGGGAGGGGGTGAAGCCCGCACCCTACCCCCACATGAAACTGATTCCACTACCCCATCTCTG

SEQ ID NO	GENE NAME	SEQUENCE
		CAAGCGTCCAGAGGCAGAGAGGCCAACATTTCGGGGACAGCTTGGAGGCGGGAGATTTAGGCAGGGCTCCTTAAACTTTTATGTGCATGA
		AAATCAGGCCAATCACGGGGCTCTTGAGCAAATGGGGACGATGATTCAGCAGGTCTGGGCTGAGGCCTCAGATTCTGCACTTCTAACAAGT
		TCCCAGGTGGTAGTGATGCTGCCAGTCCAAAGACCACACTG
200	RUNX1	TGCTTCAGTGGGGTAAACTTGAACCGCTGAGAAGACAAGCAGGGAGTCGGTCTCGCTGAGATTTTTACCTGTGGTTCTAGGAACGCAGAGG
		CATGTGAGTGTTCAGGCTTTGCATAGACCACTAAGCCACTTCTAAGAACAAGGCTACCTGAGCCATTTTGCAAAAATATGTACGTGCCGAGG
		CTTTTCCTCCCCACACCTACCTCAACTCTTTCTGCCGACACACTGCACTTTTCAAGGGAACCCAAGTTTGGGTTCGGCAAGAATTGTACGTT
		GCACACCGTGTGTGATAATTCCAGGGAATTTCAATCGCATCTTGTCTTCCTTCCTAAGCAAATTCGGTGGGAACCTGGTGTGGTGATAGA
		AAAAGCCCCGAGTTCTCTGTGGTAGACCACATCAATTTCATGTGCCAGTCTCTCAGACTCCGGCTTGCCTCTCTCAAGGAAGGGAACAATG
		GTTTGCTTGGCTTCACTCCTCTCTTTCCCCCCAATTTCCACATGGGTATCTGGCTAAAAATGAGTTACAGGTTTCCTTCTGTGAGAATTGCAT
		GGACTGATAAAGTACCATCCCAGGAAGAAAACAAAGATGCTGTCTTCCCTTTCGGCTCACAGTTGCCGTTGGGGAGGGAACACACGCTGTA
		AATTATAGGCAGCCAGAAGTGACCGCATTGACCACTGCGAGTGGCCCAGCTATGGCAACAGGCTGAGAACTCTGGGGGGAGAGCCATTTGT
		TGGCAGGGATGGTGATTCTTCTAGCATCAAGCTCTAAGATGATGACCAAACGGTATCAAAAGAAATGATATTTTGCTACCTCTCCGGCTTGG
		GTGAATGATGTGGACAGTTAACCTGGACAATTTAAACCTTTATGTTGATGGATCACTTGGATGAAATTAACCAGGAAATTGCCAAGATTTCAC
		TTGGCCCTCTGACATCAAATCTCAATATTATACCAAATTAGAGATTCTAAAGAACCCTGAGTTCCTTTCACTGAAAGGAAGGAGTGGAAA
		AACCTTTCCAGATGATCCCTTTTGAGTCTTGGTGCGAGCTCAGGCCCTCCCTACACTGCCTCCGTGAAAGCTAACCGACCCTTGTTCCTAAC
		CTAGCGCAGGTCAGCTGAGTGTCCATCGGGCACAGGAGCCCTGGGCTTGTCCGGGAGATAGCCAGACTCCCTGCTATTTCCTGATGTCTGC
		ATAGCTCAGCGTGTCCCTCACCATCTTTGCCGTTGGCCAGTAAGGAGAGCCCCAGGGGCCAGCACTGCACACTGAAACCCAACCTATTGCT
		CAATGGAATGCTTAAAAATTTCCTGAATCTGCCTTCCTGAGTTGATAAAATAGGAAACAATACACGTTCTGAGGGGGTACTGAAAGCAGAGT
		AAAGCCAGGAAGATCTTTTTTTTTCTGTTATTCTATACAAATATTGCTTCCTCTGCTTGTTAGCAGCCCAGAGGAAATGCAGCCAGGGAGCCGT
		TTGCAGCTTTTCACCAGTGGCCGGTGTCTCTGTGTTACCAACCAAACGACGCTGCAAGACTAGTGACTAACGCACGTCTGCATGATTCAACT
		TCACTAAAATTCCCTCTCTGCTGCCAGTAAAGAAGCACTTGAAAACTCTTTAATTTGAAACTTGAGCTTGGTTAATGACTTGTTTTCTTCTTTC
		TCTTTAACTTCTCTCTTGCCATCTCCAACACACACACACACACACACACACACACACACACTCTCTCTCTCTCTCTC
		TCTCTCTCTCTCTCTCATCAAGTTTTTTTAATTTCAGGGACCCGGAAACATACAGCCCCGTGCATTCACAATAGCATTTGCTGTGATAAAGT
		GGCCGGCAAGCCCTCTGCATTCCCCTGCTCACTTAGCTGTATGAATAAATAATGAGTCACAGATACAATTTGGGTGCTCAAGAGAGTTTGTA
		GCCAGAAAATTAATTATTCTCCCATCCCAGCCCACTCCATCTCAGCTTTGCCAAACCATCAAGATACACTTTGCAGGCACTGGTCAGAGTGC
		GTGCCCCGACGCACACGGCAATGCCTTTGAGACATTTTATGTTATTATTTTTGTTTGTTAAGCACAGCCCTCTTTTACCACGAAAGATACAC
		AAGACGCACATGCACACACACATACTCACACACTCACAGCTCAACCACAGCTTTGTCCATTTCAAGAGGCTGGTTTCAAAAATGGAGACAGGTT
		TTCCACCCTGGCTGTTCCTATTCATAAGCCTGTAATCTAACGACTTAAGCTGCGAGAATGCTTAACTCGGGAAACTTCTCTATTGCCCTTTTC
		CAGAGAGACCTCGGTATGCCACAATTTGCTTCCTTTCTCTCTTGAAAGATGCTGGTTGTCTCTTTGCATTGAGGCTACAAGGAAAAACACAG
		CACAGCCCCATGCTGATGATTTTAACCTAACCAAGTCTGTCAGTCTCCTGTACTCTCTGCCTTATAGAGACAGCTGCCTTGCCACTTTGGCC
		CTGAAGTCCCCCAGGCTGGTGCAAGGCTATCTGAGAGCCTCCCGCCTCCTGCCCCCACACTGGCACCAGCCCTCCTGGCTGGCTCTGTGCATG
		TGCCTGCTAAGCCCCCAGGGCAGGCTGCATTCTGGGCCACACAGCATGCCGAGTTAAGGATAACTCAGACACAGGCATTCCGGGGCAAGGGA
		CAGCAAAATAAAACCCAGGGAGCTTCGTGCAAGCTTCATAATCTCTAAGCCTTTAAACAAGACCAGCACAACTTACTCGCACTTGACAAAGT
		TCTCACGCACCGACTGAACACTCCAACAGCATAACTAAGTATTTATTAAAACATTTCTGAAGAGCTTCCATCTGATTAGTAAGTAATCCAATA
		GACTTGTAATCATATGCCTCAGTTTGAATTCCTCTCACAAACAAGACAGGGAACTGGCAGGCACCGAGGCATCTCTGCACCGAGGTGAAAC
		AAGCTGCCATTTCATTACAGGCAAAGCTGAGCAAAAGTAGATATTACAAGACCAGCATGTACTCACCTCTCATGAAGCACTGTGGGTACGAA
		GGAAATGACTCAAATATGCTGTCTGAAGCCATCGCTTCCTCCTGAAAATGCACCCTCTTCTGAAGGCGGGGGGACTCAATGATTTCTTTTACC
		TTCGGAGCGAAAACCAAGACAGGTCACTGTTTCAGCCTCACCCCTCTAGCCCTACATCTCTCTTTCTTCCCCTCTGCTGGATACCTCTGG
		GACTCCCCAAGCCCTATTAAAAAATGCACCTTTGTAAAAACAAATATTCAAATTGTTAAAGATTAAAAAAAAAAAAAAGCCAGCGCCGCCTT

SEQ ID NO	GENE NAME	SEQUENCE
		GGCTGTGGGTTGGTGATGCTCACCACGCTGCGAAACCCTGTGGTTTGCATTCAGTGTGATTCGTCCTGCCTGCTGACCACTATGCTGGGTT
		CAGACTTCTGACACTGCCAGGCTACCCAACTTGTGGTTCTGTGGTTGTTTATGAGGCCCAAAGAAGTTTTCACACAACCCAAATTACAAATTT
		AACTGTTCCCCTTTCCACAGCCCATCTCAATTGGTTCTTGCCAATCATGTGACTTAAGTGATGTCAATTTTTTTTTTTTCTTTTCTGAGCAATGC
		CCTTCCTTCCCTCCACCTGCCCTCCCCCAGGCTGTGCAAGAAAATAGCCGAGTAGACTTTGCAAGAGGGGGGGGATGTAGAAAAAAGTGACT
		CAGTCACTTATTATATCTCAATGGTCTTTGCTGATTTAGTACAACTCGGCTCCCTGTTGTTATTTGTGGTTTTTGGAACTACTGATTATTTTGATA
		AAGATTTCATTGCTGCTTATTCAATAGTAATTCAACGCTGGCATCAAGCCGCTGCTCCGACAGGATGTGGATCCCATCATTTAAAATGCTAG
		GCATCAGCTCCGGGAGAGTTAAGTCCTTGGTAACGTCTATCATGGCATAAGTGAAACTATAAAAGGGAAAAATAAATAAAAAGAAATGTTTTG
		GTGAGAGTCTGACCCCCTACAACGGGCTGGCAACTCACAGGTATTTTTAAAGCCTGGGAAAGGGAAAGAATTTTTACTTTTGAAATAAAAGGACT
		GTTTTAATGAAACCAAAATTATGTGGTTTTATTCCCCCTAAATGGACAACTTTAGTATGTATCTCTTTCAGTAAAGAGATAAAATCATAGTACA
		GTCTTAACACACACACACACACACACACACACACAAATTAGGAAGCTAAAGGAAAACAAAGCAGAGAGAATTTCTGTATT
		TGGGACAAAGCAGTGGTTACTCTGCAGATGTTTTATTTGTCACTTGGGAAAGCTCCCTGTATTGCCTTTCTCTAGTTCAATTCAAATCA
		ATAGGCTAATTTACACCTGTAGGTAAAACTACACTTTGAGCACATGAGGATGCCACAATAGAAGGGGAACCAGGAGGAGACACTTCTCCTG
		GGGCTGACTAATGAATATTATAGCGCGTCCTCTACCTTAGAAAGACATGCCTGTTTGAAGATGCTAAAAACAGGATAATTTTGTAAGTGGG
		CAAACCACTGTGGTCACACGTATTTCATTTTCCGGCCCCACTGGCTTTACCTGCTGACAACTAAAACGTCATTTTGTTTTGTAGTTCCAAGAT
		GAAGAAAGGCTTATTTTCCTGATTTACTACCTTATTCATTTGGCTCTGCTCTGCCTACATCCGCCATAGCACTCTGCGCACGTGAAATTTCGA
		CACATAGGGTCAAGAGAACCTGTGTGATGATGGGTTGTAAATGCCAGTCCTGGATTCTAAGCTGCAGTAGCCAGCACAGGCACTTCAGAAA
		GGCTGAACTCCCCACAACACTCCCTCGGTTTTCCCTCATCCACTTAATTTCACACACACAAAGACCCCACAACGATAGTAGCTTCCATGGCACA
		AGTCTTTCAAAAGGAACAGACACAATTTTTACTTACTCCTGTTTTGACTAAAGCAGGAATTGAAACTCAACAGACCGCTTTCTCTTACACTTGT
		GAGAAGTTAGCTGGCCCACATGT
201	chr21:35499200-35499700	AGGGAAAAGAGATAACGAAAGAAAGAAAAAAAAAAAGGGCCGGCAATTTCATGTACATTTGTTTTTGGCATTCGCTGAATTCTAGAGATG
		AAAACAATCTCCTGCTTTTAATTCAGTCCACGTGCAACAAAGTTGTACGTTGGGAGATCTGGCTTTTAATAAGAACGATTAACAAGCGTTTTTT
		GATCACAGGAAGTTGAGAAGAGTCGCTGCTTCTAAGAATACAATAAACATTGACTAGCAGTTAGACGGTCCATCTTTCTCTATCAGCCGTTTA
		GCAGCCTCTACTTTGATTTGGGGCAAATGCGAGATGGGACCAGGAGAGAGCTCCCCCACACCCCCACCACGTGGGCAGTGGTTCTGTT
		CCAGAGCGCCTTCCTTCCTGTCCAGGGAGGCAGGCTGCTGAGGCCGTTTCTGGGCAAGAGGCCATTGTCGGGATATTTGCTTTAGATAGC
		TTGCAGCTGGGCTGAGTGGGTGTTTCATTCAGACTCAACACA
202	chr21:35822800-35823500	AGCCTGGCGCACCCGCCCTAATTTGAGTCAGGGACCCTAGGCGCCTGCAGCTCCGGTTCGGGTTGAGTGCCTCCTGTCAGGATGTGAAGC
		TGCTGTCCCCCCCCGGGGGGCCTCCAGCACTGCTGAGGACTCAGCAGTCAGCCTCTCCTCCCACTTGGGCTCATTTACAGAGAGCATCTCCA
		GGAATCAGTCATGGGGAAAGGGGAAACGCGGAGTGACAACACAACACGTAGAAAGTTCTCTGCCGCCTTGGTCAGGCTTGTCAGCCTCAC
		AGCCCATCCTGCTCCTGCGGGAGGAAAAGTGAGCAGAACTCAGCCCGGAGATGAGCCGCAGGCCGGCAGCCCCTGCCTCTGCCCTGCTT
		GTTGTGACTGCAATGCAAGGCTCTCTGTAGGTGCGGGGGGATTCGGGTTAAATGGGTCTCCAGTGGTCCAGCGCTCCCAGCAAAGGCCGAC
		CACAAGAATTAGCGGGCTAGTTATTTACCATAACCATATACAAAACCACACAAGCATCAGCGTTCCCTCAAATACATCCGAGACGCTGTATATCT
		CTTTATTAAAGCCTGTCAGGGTTTGTTATTGCACAGCTTGGCCTTGAACCCCAACTAAACCAGGCTGCTTGAGCAAAGAACCAAGCAATGCA
		AGCATTCAGGCAGGACCATTATAACCCTGAGGCCAAAGGCAGAAGCAGGGAGAGGAGACGTCTTCC

SEQ ID NO	GENE NAME	SEQUENCE
203	CBR1	AGACCAGCCTCGGTCTTCGGCCTGCGGGTTCTGCAAAGTCAGGCTAGCTGGCTCTCCGCCTGCTCCGCACCCCGGCGAGGTTCCGGTGG GGAGGGGTAGGGATGGTTCAGCCCCCGCCCCGCTAGGGCGGGGCCTGCGCCTGCGCGCTCAGCGGCCGGGCGTGTAACCCACGGGTGC GCGCCCACGACCGCCAGACTCGAGCAGTCTCTGGAACACGCTGCGGGGGCTCCCGGGCCTGAGCCAGGTCTGTTCTCCACGCAGGTGTTC CGCGCGCCCCGTTCAGCCATGTCGTCCGGCATCCATGTAGCGCTGGTGACTGGAGGCAACAAGGGCATCGGCTTGGCCATCGTGCGCGA CCTGTGCCGGCTGTTCTCGGGGGACGTGGTGCTCACGGCGCGGGACGTGACGCGGGGCCAGGCGGCCGTACAGCAGCTGCAGGCGGA GGGCCTGAGCCCGCGCTTCCACCAGCTGGACATCGACGATCTGCAGAGCATCCGCGCCC
204	DOPEY 2	AAACGTTTAAAATATATTTCTAAACAGAATGGGCCAATTCAGTCACAGTAACTGTTGATCTCCATAGCAGAGCAACCCACAAAGACAGAACTG ATTTTTTTCCCATAATCAGGGGTGAAAAATATACAACTTGTTTCTGAACCAAAACCACACAATTTCTGCAGTTTAAAATGTTTCACTGCTAATATG GCCCTGGTAGAAATTATGTAGTTTCTTTCTTCTTTAAAAAAAAAAAAAATTAAAAAAATTTCCTAAGACACTAAATGCTCCATCTGGAATGTAG ATTCTGATCACAAAGCAGCTCAGTTAACCTAAAAAATAAAAAATTCCCATCACCTGTCTCAGTAGGGCCTGAGAGTAGTGTGGGGAACCCCA GCTTTGGTATGGAGAGTCATGGCCCCTTGAACCAGATAGAGACCTTGAATAGCCATAGCTGGTGCTTCTCTCAGGATAAACTCTGATGTAG GAAGTATCACCCTCATGAGAGTGGAATTTGGTCATCCAGTTGACGCAGGGCATATTCCATGTCTTCTTTTCTGAGACACCCAACCATCCCCA CTCCATCCTTCTGCACATCCGTGTAACAGGCATCCCCAGCTTCTCGCGGTGTGATCCTTCAGGTCCTGCCAGCTGCCTGATGGAAGAAGTCC ATTTCTTCCATAAATAGCATCCTCTCGCATCTCGAGGGTCCTCGAAGCGCACGGAGGCGAAGGGCACAAGGCCGTACCGGCTCTTGAGCTC GATCTCGCGGATGCGGCTGTACTTGTAGAACAGGTCCTGCGGCTCCTTCTCGCGCACGTGGGTCGGAAGGTTTCCCCACGTAGATGCACC CGTCGCCCTCCCAGCCGCGCTCGTGTCCGCCCAGCCGGACAACCGCACCGCCCGACGCTGCTGGCCAGCCGCAGCCCGCATCCGCCCG TATCGCCGCCGCTGCCGCCTCAGCACGGCTGCCCCCGCAGCGTCTGTTTTGTTTTATTCTAACAGGGTCTCTCTCTGTCGCCCAGGCTGGA GTGCAGTGGCGTGATCTTGGCTCCCTGCCTCCCGGGTTCAAGCGATTCACCTGCCTCAGCCTCCCAAGTAGTGGGCATTATA GGTGCCAGCTAACCATGGCCGGCTAATTTTTTT

SEQ ID NO	GENE NAME	SEQUENCE
		ACCAACCTAAAAGAAAAGACTCTCCCAGGGAGTAACTCCCAGTGAAATAATTTATTAAAAAAAGCAAAAAAGAGACATAAATTTCTCTCTACT
		ACTTGAGGAAACAGCAAACAGAACGAATTAGGGTCTTGGCCTCTGCAGGAATAAATTATTTCCGACTTGGTCTGGATACCTGTAATTATTTGT
		AAGCTGTGGGTAGTAATACTGTAATTGTCCCCCGGTCCTTTCTGGAAGTAGCAATGACCCCAAGGACAATTGGTGACGTCTCCACAGGGTT
		TACACATGGAAAGGAGTGAAAAATCGAGGAATTCTTTCAGATAGCCCAGACCAAAAATCCTCTCTCAGCCATGAAAAGGTCATATATGTGATGC
		TGGGCCAAGCGGACTTTTCTGGAGTAACCATATCATAACTGATTGCGGATGTAGACAAGAGCGTATAAACCAAATAGGCTTGAATCAACGCA
		GTCCTGGATTTTCTGTTGCCTCTGCTTGCTGGGGCAGTGGAAGTTCTTAAACTCCACTTCAGAGGTTGGAAATTCTTCCCCCTCCCCCACCT
		CCTTAGTGACAAGGTCTCTGATCTCCTGCTGCCACTGCAATAGCCTCTCCCATCCCGCGGGGAACGGCCGGAGTTCTTCCCTTGATCTCTC
		CCGAGTCGGCTTCCGCTGGGGATGGATCGCAGGTAGGCGCCGGCGCGGCCTGGGGAAGAACAGTTGCGGAGCATCTGAAGCGGAAAAT
		CCAAGCAGATGTGAGGCGATCCGGGCCCGCCTCGTTCCTCTTGGGGCCTGAATTTCTTCCAGATAAGTTTCCTAATGGAACATTTCTAAGA
		GGTGGGGTACGAGGCGGCTTGCTCGCACGCGCAGTGGGACAGACTGCGGGTGGGGACGTACTGAGAGGTCCGGACCTCAATGCGTCCG
		ACCCGTCTCCACACCGCCCTTTTCCAGCCCCCAGTCTCCTTTCATTCCCTACTCTTCAGGCTCCTTTGGGGCCAGTGGGTGAACCGCCATT
		TAGAACGGTGCCTCGGACTCGGGGGTCGTGCGCTCCATCTCTGCCTCCCCCCTGGGGCCCGCGAGGCTGGTCCGGGCTTTCTGAGCTGG
		GCGTTCGGCTTTAGGCCCAATACCTGGACCAGGAATTTCTTCTCCCCGCGCCAGAAGGGAAAGACATAGGAGGTGTCCCAATCTGCGGTC
		ACCGCCGATGCTCCTGACCACTCTAGTGAGCACCTGCCCGGTACTTTTCCATTCCAACAGAGCTTCCAGCTTCATACTAACTATCCCACATA
		CGGCCTGTGGGTATTAGCTCTAAGTGTCCTTTTCCGAGGGCCCGAGGCTCCCCCCTCCAGCAGGGAGAGCTCCGGGACGGCCCCCACCAA
		GGGTTGGGTTTCTTCCTTCACAATTCCACAGAGGCATCCCTGTCCTTCCTACCTGGGAAACCTCGAGGTGCGGTGCCCGTGTACTTCTGGT
		ACTTTGCGTGGTGCCATCAGGGACCCCAGAGCCACAGCTGCGTGTGTGTGTGGATGTGTGTGTGCGCCGCGCGCGTGTACGG
		CGAAGGGATGTGCTTGGGGGAGCCGAGTACACAACGTCTGCTTGGGCAGCTGCTGGGCAGGCGTTGGGCCTGGAGGTATCTCACACCCA
		CGTATCTTCCAGTCTTCAAACACGGCATTGCTCTGCCTCCCGTAGCGCGCTTCGAACCTGCCTCGCGGACACGTGAACAGAGGCTGTCCCT
		GGGAAGATAAGTGCGCTTTCCCGTAAAATCCGGGAAATTTGCCTTGAGGAAAGTTTCCGTTCTTGTTACTTGTCGGGTTTCTCCCACTTCCA
		CTTAGCCATGTTTCTGCGATCTGGGTAATCCCTTTCAAGCCCAGGAGGAATTCTCCCGGGTCCATAATTGAGGGTCCGAAGCCGTGGGGGT
		GAGAAACGCATTAAATCCTCCCGAAGCCCAGGAGGTGCCAGAGCGGGCTCAGGGGGCCGCCTGCGGAAGCTGCGGCAGGGGCTGGGTC
		CGTAGCCTCTAACCCCTTGGAGCTCCTTCTCCCAGAGGCCCGGAGCCGGCAGCTGTCAGCGCAGCCAGGAGCGGGATCCTGGGCGCGGA
		GGTGGGTCCGACTCGCCAGGCTTGGGCATTGGAGACCCGCGCCGCTAGCCCATGGCCCTCTGCTCAAGCCGCTGCAACAGGAAAGCGCT
		CCTGGATCCGAAACCCCAAAGGAAAGCGCTGTTACTCTGTGCGTCCGGCTCGCGTGGCGTTTCGGAGCACCAAGCCTGCGAGC
		CCTGGCCACGATGTGGACTCCGCAAGGGGCTAGGGACAGGCAGGGGGAGAGCCCGGGTTTGCGCACACCTTCCAGCCCCTGGAGGGAG
		CCTGCTCGGCTTCGAACGCCTTCGAACTTTTGACCTTCAAAGGAGTCCCTGGAAAAGGTCAGGAGCGCCTGCTGCAGGCACGGTTGCCGA
		AGGCCAGGCCTTCCTGGCGCAGGGGAGGGCCAGGGGAGGGAAGCGGATACTCAGTCGCTGTCCGACGGCGAGTTTTCGGAGCAGCAGG
		CTCATGATCCCGGGCCAGTGGCGAGAGCAGTGACACCGAGAACCCAAATCTCCGCGCCCCCATCCGCGGCCCGGTGTCCTCCCGGCCCC
		TGCTGACCTCCAGGTCACGCACCCCACTGCTCCACGGCTCTGCAGCCTGTGGCACACGGCCGAGAGTCCCCACATGATCTCGACGCCAAG
		GTAAGGAATTGCCCTGCGTCCTCTGAGCCTGTCTCTGGCCTGGGGGGCCGGGAAAGCTGCACTCCTGGAAGAGGTGGGGTTATGTGACC
		GCCGCTGCAGGGGTGCGCGGAGGACTCCTGGGCCGCACACCCATTTCCAGGCTGCGGGAGCCGGACAGGGGAGGGCAGAGGGGGGAC
		AAAAGGACTCTTTAGGTCCAAAATGACCCTGAAGGAGAGTCCAGAATGCCCAGTGGCCGCGTCTGCAACGGAGTCTTCTTTCTCCAATTGC
		CTTCTGCCCCATCACCATGGGCCCCACCTGCGCCACCCTGTGACCCTGGCTCAGCGACCTTGGCCCTTAATCGCCCCAACG
		CCGATTCCTCAAAATTCCGGCTGCGCTGAATCGGGCTGCTTTTGCCGCCGCCCCGGCAGTTGGGCCCTGTTTCCGCCGGCGCCCTGGGA
		GAGGCCTCACCACTCGGCTGGGCTCCCCTCCCTTCCCCTGGCCTGAGCGCCCCTGCGGCCTCCCGCTCCTCCTGAGAAGGCGA
		CAATCTCTTTGCACCTTAGTGTTTCGAGGACAGAAAGGGCAGAAGGGTCACTTCGGAGCCACTCGCGCCGTTTTCACGTGTGTGTGTAATG
		GGGGGAGGGGGGCTCCCGGCTTTCCCCTTTTCAGCTCTTGGACCTGCAACACCGGGAGGGCGAGGACGCGGGACCAGCGCACCCTCGG
		AAGGCTCGATCCTCCCCCGGCAGGGCGCCTGGCCAACGAGTCGCGCCGCCTCCTCTCGGCCGCGCCTGCTGGTGACCTTCCCGAGAGCCA
		CAGGGGCGGCCTCGGCACCCCTCCTTCCCTCGCCCTCCCTGCCGCCCATCCTAGCTCCGGGGTCCGGCGACCGGCGCTCAGGAGCGGG

SEQ ID NO	GENE NAME	SEQUENCE
		TCCCCGCGGCGCGCCGTGTGCACTCACCGCGACTTCCCCGAACCCGGGAGCGCGCGGGTCTCTCCCGGGAGAGTCCCTGGAGGCAGCG
		ACGCGGAGGCGCGCCTGTGACTCCAGGGCCCGCGGCGGGTCGGAGGCAAGATTCGGCCGCCCCCGCCCCGCGCGGTCCCTCCCCCCC
		TCCCGCTCCCCCCTCCGGGACCCAGGCGGCCAGTGCTCCGCCCGAAGGCGGGTCTGCCATAAACAAACGCGGCTCGGCCGCACGTGGA
		CAGCGGAGGTGCTGCCCTAGCCACACATCGCGGGCTCCGGCGCTGCGTCTCCAGGCACAGGGAGCCGCCAGGAAGGGCAGGAGAGCCG
		CGCCCGGGCCAGGGCCCGGCCCCAGCCGCCTGCGACTCGCTCCCCCTCCGCTGGGCTCCCGCTCCATGGCTCCGCGGCCACCGCCGCCC
		CTGTCGCCCTCCCGGTCCGGAGGGGCCTTGCCGCAGCCGGTTCGAGCACTCGACGAAGGAGTAAGCAGCGCCTCCGCCTCCGCGCCGGC
		CGCCCCCACCCCCCCAGGAAGGCCGAGGCAGGAGAGGCAGGAGCGAGCAGGAACGGGGCTCCGGTTGCTGCAGG
		ACGGTCCAGCCCGGAGGAGGCTGCGCTCCGGGCAGCGGCGGCGCCGCCGGGTTGCTCGGAGCTCAGGCCCGGCGGCTGCGGG
		GAGGCGTCTCGGAACCCCGGGAGGCCCCCCGCACCTGCCCGCGGCCCACTCACCTGGCTCCCGGCTCCCCCTTCCCCAT
		CCCCGCCGCCCGCAGCCCGAGCGGGGCTCCGCGGCGCCGGGTCTAATATGCCCGGAGCCGAGGCGCGATGAAGGAGA
		AGTCCAAGAATGCGGCCAAGACCAGGAGGGAGAAGGAAATGGCGAGTTTTACGAGCTTGCCAAGCTGCTCCCGCTGCCGGTCGGCCATCA
		CTTCGCAGCTGGACAAAGCGTCCATCATCCGCCTCACCACGAGCTACCTGAAGATGCGCGCGCCGTCTTCCCCGAAGGTGAGGCCTCAGGTG
		GGCGGCCGGGGACGCTGGGGAGCCCGGGCGGCCCCAGGGCGGGAAGCGCAAGCCCAGCCCGCCCAGAGGGGTTGCCGCGGCCTG
		GCGTCCAGAGCTGGGGCGTCTGAGGGAGGTTGCGTGAGGGTCTTCGGCTTCGGCGCTGGCTTGGGGCGAGGGGCCAGGGCCTTGGCGG
		CCCAGGCGACCAAACCCTCTCCTGGTCCAGGGGCTGGGTGAGGGGCGAATTACGAATTGTTCCAGGGGGCAGGCAGTCCCCCCAGCCCCGCACG
		GCCAGCGAGTTCTTTCTGGTTTTGTTCTTTCTCCCTTTCCTCCTTCCTTCGCCAGTGCATTCTGGTTTGGTTTGGATTTTTTTCTCTCTT
		TCTTTCCTTTCTTTCTTTCTCTTTCTTTTCTTTCTCTTTCCTCTTTCTTTCATTCTCCCCCTTCCTTCCTTCGGCCCCCTCTCTCCCCT
		CCCTCCTTCCTTCCTTCCTTTGCCAATGCATTGGTTTGTTTTCTTTCCTTTTTCTGCTTTCCTTTCTTTGGAAGTTTCACTCTGGTTTTTGCTT
		TCTTTCTTTCCCCATCCCTTCTTTTCTTTATCCCTCCTTCCCTCCTTTTCTTTCTACGATTCCCTTTATTTTCCTTCATTCCTCCCTCTTT
		TTGTCTCTTCTGGAGGAGGTGAAGGAGGGTCAGCTTCAGGCGCTGCGAGTCAGCGGGGATCACGGTGAGGCCCAAGCACTGCAGGCTGA
		GGCCACAGAGCGAACACTTGTGCTGAGCCGGGCCCTCTCGTGAGGCTGGGGTGCCGGGCAGGAGAGACCCCGCCCCCGCCG
		TTGCTGAGCTGAGACCCGGCTGAAAGAGAGGGGTCCGATTAATTCGAAAATGGCAGACAGAGCTGAGCGCTGCCGTTCTTTTCAGGATTGA
		AAATGTGCCAGTGGGCCAGGGGCGCTGGGACCCGCGGTGCGGAAGACTCGGAACAGGAAGAAATAGTGGCGCGCTGGGTGGGCTGCCC
		CGCCGCCCACGCCGGTTGCCGGCTGGTGACAGTGGCTGCCCGGCCAGGCACCTCCGAGCAGCAGGTCTGAGCGTTTTTGGCGTCCCAAGC
		GTTCCGGGCCCGCGTCTTCCAGAGCCTCTGCTCCCAGCGGGGTCGCTGCGGCCTGGCCCGAAGGATTTGACTCTTTGCTGGGAGGCGCGC
		TGCTCAGGGTTCTGGTGGGTCCTCTGGGCCCAGGAGCTGGGAGGGCTGCGCCGGCCTCTGGAGCCCCGGGAGCCAGTGCCGAGGTAGG
		GAGACAACTTCCGCCGCAGGGCGCCGGACGGTCGGGGCAGAGCAGGCGACAGGTGTCCCTAGGCCGCAGGGCGCTTCCATAGCGCCAT
		CCCCACCAGGCACTCTACTCGAAATCGGAAAGCTCGACCTTTTGCGTTCGCCTCTGCCAAGCCTGTTATTTGTGCTGGCCGCTGGGTCTGG
		AGCTGCGCTTCTCGGCCCCTCCCCGGTGGAGCGCAGAGGGCTGGTCTGCAAGCGCGGCCTCCAGCCCCGCGGCTCCCCGGCCCAGGAG
		CCAGGCGCGGGCTGACCCGGGAGCACCCGGCAGCGGAGGGGCTGGAAGCGGGACCCTAGGCCTCTCCTGTGCCACCCGGCCCTACCG
		CGCGGCCGCGGGGCGCTCTCCTCTCGGGCGCAGCGGTCCTTCAGCCCAGGGCAGGTTCCTCCCTTTCCTACTCGGAACGTGGCAAAGAT
		ACCCCAGTCCCAGCCCCTCCAGCTGAGAGCTGTTGCCCAAGGTCGTCGCTACTTGTCCGCTCAATGGTGACCCCTTGGCAGAGAACTAGG
		GATGATTCCACTCCGGTTGATGTTTTAGGGGAAATTAAAGAACATTCGGTTTTCTGAGTCTCCTTCCGGGGAGGCGTGGTGGTAACTGGTT
		TGCTGGGAAGAGCCGTTCCTTAACCGCATGCAACAAAGCAGGTGTGGAATCCGGACGAGAGGGCACTCACTGCCTTCTGCCCCCTTTGGA
		AATAGAAAAAGCCTTCGAAGCAGCAATCCAAAGATCAAATGATTTGCGGTCAATGATTTCAATTAAACCAGAAATTAGTAAGGGAGGGCCGA
		GAAGACACGGCTGCTCAGAAGCTGTTCGCTGTTTGAGGGATTTCCCGGAGAGCCTGTTAAAAGATGCGAAGTGGTGGGTGTACCGCTCAG
		CCACCTTTAAACCGGCTCTGTGCGTTCTGGCTCTGGAAAGCAAGTCTCCAGGCATTTGGGCTCAGAATTGCTGGGCCCCGAGTTTGGGCG
		GGGGTGGTCCTTCTGGGGGTCAGGCCTTGAGCAGCTTGCACTGGTGGCAGGTTTGGGAGCAGTTGAGGGGCCCTGAACCTGTTCGGTCCCCCTCGG
		GGGGGTGACCCTGGAAGTTGGCACTCTGGAAGGGAGCTGTTTGGCCCTAGAGTTTTGGAAAGGGCCCTGAACCTGTTCGGTCCCCCTCGG
		AAAGGGAAGGGAGCAGTGGCTTAGTCCCTCCCTCCTCCATTCGTGCAATGCCCTGGGGTAGGGGTAGACCTGGAGCCGGTGGACTCATATC

SEQ ID NO	GENE NAME	SEQUENCE
		CTTGGAATTCGTCAGGACAGCTGCTCCGGGGCCTTGGCCCTCAGTCAGTCTGGGGCTGAGGAGTAGGGAAGCTGGGAACTTGGGGCAGA
		GGAAGAAGATGCGTTTAGAAAGACCTCCATTATGCAAACTGGAGTCCATTTATGCAAACTGGTCACCCTTCCAGTAGCTCCAAAGAGTGGCA
		GTGGAGTGGCATCTTGATTGATTTAACCTCTTCTCAGGGGACCTGGGTCTGCGAGGGAGGATATGGCTGCGGGGTTGGAATAGGATCTGT
		CTGAGCTGCCAGGGTCAGGGTGGTGGCCCTAGGGAGGTTTTAGGGCCAGGGTGGTCCCGGGCTGTGGCAGGGGCTCTCAGATCGCCTC
		GGGCTCTCAGCTGCAAGGTGAAAAATACCATGAGGAATTGATCTGCCAAGGGCGGTCTTGTCTCAAAGCAAGTGGATTGCTGGGGTAAAGA
		ATCTAGAGACCAGCTTAGGACTCTGGGAGGAAGAAAAAAAAAAAAAAGAATAGCATAGTCCTAAGGAACTGCAAGGATCACCAGATTAACCCT
		TCATACCTGGGGAAATTAAGGCCAGACATGACACAGGCCTTTCCCAAGGCTCTGTAGCAAGGGCAATAGCAGGCCAGTTGCTGCCACTGC
		GGTCCTGTGGGGCATGTTCTCACTCCACTGCACCCAGGAGGCTGCCAGCCTCTGTTCCTTTTAACATAGATCTCCTCAGTTGTTAAGACAGA
		AAGAGGAACTCAGAGGGGTCCCTGTGTGCAAGGCAGAGGGAGACCACCAGAACCAGGGTAAGCACCCCACTTGGTAGCCAGTTCAAGGA
		CTTGGGGATGTTTTTCAACATTTACAGCGAGGTTTGAGGCCCCATTGTCATGCAGCGCTACTCGGCCTTGGTCTCCTTATCTGTAAAATGGGC
		CCATTAGCAATGCACAGGGTTGCTGTGATGAAGGGTGAGGTCCCACAAGCAAAAGCTGTGCAGTGAGGGGGGAATCCTAAGCATTGTTCC
		TATGCCATTCACCCCTTCCTGTGAGCTCCCCATATTCCCTGGCTCAAAGGAGTCTTGAATGGCAGGGATGGAGGAGACTCACTGCCTGGACTT
		TGAAGACCCCTGCTTTCTGGGTGACCACCTTTTCTTCCCTTTGACAGTGAACTAATACATTGGAGGTAGATAGTGCTGGGAAGAGGACAGG
		AGACCACGGCTGACTTTGGACATGGGCTCGAAATTGATAACTTGATGAGTCTTGGAGGGGTGGTTAAGATAAGCTCGGGGCTGGGGCAGCG
		CTGAGGTCTGATGGTCAGCCAGCCCTCCCCAAAGTGTGGCCCTCCGTTCTGGAGATAGGGGCTTTGGAAACTGCAAAAGCCGTCCTGGCAG
		GCCAGCTCTGGTTGCTCCCTGGCCATAGCTGCTCTGACTACAGGCAGCAGGACGCAGGTCGGCCTCTGCCCATCGGAGGTCAGAGGCAG
		GGCCTCCAGCACCAGACTCAGCAGTGCCACTGCAAACCTGGCACAACAGGCTGGTCCCAGGACTCAGCTCAGCAGTGAAGTTGGAAACCA
		AGGTTGAGTCTCCCCATCTCCCTTTCCCCAACCCGAAAGACCCAAGATGGGTGTGGGTGAAAGAGGGAGAAAGAATTGCTACTCCAGAAAC
		TGTCATTTGCCCACACGAAACGAGGTGGGGTTCAAGGTCTGAACTCTTCCAGTGCCTGGGTGCCTTTGGGTTTAAATTCAGCTGCAGGTGC
		CCCCATCACCACTTCCACCTGAGCACACCACGAGAAGCCAGGTTATCTTAGAAACTGTTTCCCGGAATCAAAGCGACTTGATTTGGAGAGTT
		GGGTGAGGAGAAACTCACCCCTATACCCCTCAGGGCGTCAGAGATGTGAGGCAATTCTCTACCTCCGCTGGAAAAAATGCAGATTTATTAA
		AGGTCGACTGTTTAGCAGAACAACGTAGATTTTTTTACAACGCTTTCCCCGTCTCTGCTTTGAAGCCTGCCAGGCTGGGGATCCAG
		GAGGGAAAGCCCCGCAGGCGCAGAGGGGACAATCCGGGAAGTGGTAAAGGGGACACCCGGGCCTGTGCTTTCGTTGCAGGCG
		AGGAAGTGGAGCGCGCGCTGCAGATTCAGCGCGGGGCTAGAGGAGGGGACCTGGATCCCTGAACCCCGGGGCGGAAAGGGAGCCTCCG
		GGCGGCTGTGGGTGCCGCGCTCCTCGGAGCCAGCTGCTGGGGCGGCGTCCGAACTCCCCAGGTCTGCGCACGGGCAATGGGGGCAC
		CGGGCCTTCTGTCTGTCCTCAGAATACGTAGGATACCCGCGGGCGACAAGCCGGGCCAGGCTAGGAGCCTCCTTCCCTGCCCCTCCCCAT
		CGGCCGCGGGAGGCTTTCTTGGGGCGTCCCCACGACCACCCCCTTCTCACCCGGTCCCCAGTTTGGAAAAAGGCGCAAGAAGCGGGCTT
		TTCAGGGACCCCGGGGAGAACACGAGGGCTCCGACGCGGGAGAAGGATTGAAGCGTGCAGAGGCGCCCCAAATTGCGACAATTTACTGG
		GATCCTTTTGTGGGGAAAGGAGGCTTAGAGGCTCAAGCTATAGGCTGTCCTAGAGCAACTAGGCGAGAACCTGGCCCCAAACTCCCTCCTT
		ACGCCCTGGCACAGGTTCCCGGCGACTGGTGTTCCCAAGGGAGCCCCCTGAGCCTACCGCCCTTGCAGGGGGTCGTGCTGCGGCTTCTG
		GGTCATAAACGCCGAGGTCGGGGGTGGCGGAGCTGTAGAGGCTGCCCGCGCAGAAAGCTCCAGGATCCCAATATGTGCTTGCGTGGAGC
		AGGGAGCGGAAGAGGCAGCCGGTCCTCACCCTCCTCTCCCGCCACGCACATATCCTTCTTGACTTCGAAGTGGTTTGCAATCCGAAAGTG
		AGACCTTGAGTCCTCAGATGGCCGGCAACGCGCCGAGGTCACGCTCCCCAGAAACACCCCTCTCCCCTCCCCAGCTCCCCCCTGG
		GGCGGGTGGTAATTGGGGGAGGAGAGGCCGCAGGCAGGGAAGGGGTGGGAAAGCCAGAGAGGGAGGCACAAAGTGATGGCAGCCCGG
		CAAACACTGGGGCTTCGGGCTGGGCCGCGCTCGTTTAATCCCACAAAAATCCCATTTTGGAGGTGAGAAATAGAGGTTAGAGGTCGGGCC
		CTTCTGGAGATCAGACCGAGGAGACGGGCCCAGCTGGCGTCTTAAAGCAAGGAGGGGGAGTCGGGAGGAGGTGAGACCCCCTGCACCCA
		GGTGGGGCTCCCAAACCGTTCTGGATTTACCACACTCCCAGGTCCGATTTTCCATGGAGGGCTGGGGTTAGGGACTGGCACCTTCTTGTTG
		TTAACCGCATTTGATATTCACAAGAACCCTGTGAGGAGACTTTGTCACCGTTTTTAGATGCCTGAGGTTGCCGGAGGGGCAGTGAGAGAAT
		CGTCTAACCTGGTGTTCCTACCACAGTCCAGGCCCTGTGTCCTGGGCTGGACCCACAGCCCCTGCCACCACCCAGAGGAAGGCGCGAAG
		CTGGCTGCCTCCTTTACGGGTCTCCCTTAGGTGCCCTCATGAAGGGGGACGGCCACCTCACAGTGCAGGAACTATCTCCCCGTTTGCTCC

SEQ ID NO	GENE NAME	SEQUENCE
		CAAATAGTCTTCTTGGTGTGGTGCTGTCTATGGTCTGTGACCTGCATCTGGAGTTACCCCCAGGACCAGCTTCGGAAGAGGAGGGATCGCT
		TGGAGGCCGTGCAGTGTGAGGAACGGCAGGCAGGGTGTGGGACCAACATGCACACACTCGCAGGTGCTGGGGCCAGGGAGGAATGAGG
		CGCTGGCTCCCTTTCCCTCCATTTCTCCCTGGGGGTCCCAGCAACCTGGCCATCCCTGACTTCCAACAGCACAGCGTCCCCACAGGTCCT
		GCAGTGCTCTGCAGGGGTGCAGGGAGCTCCCCCTCCCCCAGCCGCAACCTCACCTTCCTCACCCCCACCCCTCCGGCAGGAAACCACAG
		GCTGGGTTGGGGACCCCCTGGTGCTCCAAGAGAGCAGTGAGTGCTGGGAGCCGCTAACCCCGAGGCGCCTAGCACAGACTCTTCTCACCC
		CTTATTTCTGAAATAAAGCCCTTCCTTAGGTCCAGATGAGGACCACGTGCTCAGTGCCTCACTTTCGTGGGAGTGTATATCACTTTACAGTAT
		CAAGACAATTTTCTTTCGTTACAAATCTTTATTTAGTCTCTGCGTTTAGACCAAAGTAGATTTTTATGGGCTGAGTGAAAAAACCTCGCCCGCA
		TTGGTTTCTGATGGAACAGCTGGCAGCGCCACGGCCCCGGGTGGGGTGGCCTAGAGGCAGGGGTGCTTGGGAGGAACATCTAGCACCCG
		ACCACCTCCACCAGGTGGGAAAGGGACGTTTGCACCAAATCTCCGCCGGCAAAGCAGAGGCTTTGGGGAATTACAGAAAAACTATAATGAT
		CTAAAAGAGAACAAGTTATCTTGAACTGTGCGGGTATTTGAATCATACAGAAAATTGTCCTGTGTGCCCAATGCACTTTTGCATGTAGAGCCA
		GGGCCTTCGAGGAAGCTTTCAGGAGATCCCGGGCAGCGGAGTCTGGTCTGGAGTTTCATTTCCGTAGGTGCAGATTTCTCCCCAAGTCTTC
		CCGCCATGGGCTTTGCAAGAAGCCAGGGCCCAGAGGCCACGCTCACCGTTAACACTGCACAGGGCAAAGGTGGCTCCAGGACAACTGCC
		CAACCCCAGGAACGACCCAGCAGCAGAGAAAGGGACAGCTGCCAGGGTGCCTTTGTCGCTTTTTGGAAATCAGAATTCCTGGGTCCTTAGT
		TAAGTCTTACTTCACCAAATCCCAGGACCTTCACATTTTGGTTCTTGCCATTGCTAACAGTTGTAAATGCTGCCGCCACGAGGCCTGGGAGG
		AAGGACCCGCTGGTGAGAGCACAGGGAGTGCTGCTGTGATCACGGTGGTGATGCGGGGGTGAGCCGCGATTTCCCGGGATTAAAAGCCAC
		CGCTGCCCCCCGTGGTGGAGGCTGGGGGCCCCCGAATAATGAGCTGTGATTGTATTCCCCGGGATCGTGTATGTGGAAATTAGCCACCTCCT
		CAGCCAGGATAAGCCCCTAATTCCTTGAGCCCAGGAGGAGAAATTAAAGGTCATCCCTTTTTAAATTGAGGAATAGTGGTTTTTTTTAACTTT
		TTTTTTTTTAGGTTTTTAGTTGCCGAATAGGGAAGGGTTTGCGAAGCCGCTGCCCTGGGCCGAGGTGCATTTTACGCTTCCAGAGGTCGAG
		GCCTCCAGAGACCGCGATGCCCAGGGCCGTTCCCGGGGAGGCTGAGAGACCCAGGGTGCTCTGGGTGACTGCACGGCGACTCCTCGGGA
		ACCCACTCGTGGCTGCCCGCTTGGAAGGGCTTTGCGGCCCCGGGAACGATCTCCAGGATCTCCACGGCTGGTCAGGTTCCCCGTCCCTC
		GTATCCCGCGCTGCCCGGGGGCTCCTGCCTTTGGTTCAGTGCTCGCGGCACCACCGCACTCAGGACGGCAGTGGGGGGCTGGGGCTGG
		GGCTGGGCCTGGCCCAGCGTGGGTTGGGGCGGGGACGCGCCGCAGCAGCGCCCGCAGCTCGCTCCGCAGGCCAGGGGTCGAGGGTCG
		GGAGCTAGGCTCGTGGGCCGGGAGACGCCGGGCGCGTTGTCCTCCGGGGAGGGTTGGGGTGCAGGCGGTGCACCGACCCTCGCCATCTG
		GCGCTGCAGCCACCACGCCACGGCCGCTTAGTGGAGGGTCTGCCGGCCAGGCTCCCGGCGGAAAGATTCCGGGGAGGGCTCGGGGGTTGTC
		CCAGCCCGCGCTAAGCGCCGCAGCCTCGCCCGGCTTTCCTGCTTCCTCGGACTGTGCAGGGGAAGCCTGGGGTCTCGCGGGGCGCAGC
		AGTCAGGTCGAGGGTGCAGCAGGAGGGGAGTCCTGACGGGCAGGTCCCTCTTTCCCCTGGTGCGCAACACTGGTTGGTAGCTTTTGCGG
		AGGTGGTGAAGAAGGGCAGGAGGCCTGTTGAGCGGAGGAGTCCGGGGATCCCTAATTATGTGACAGGAGACCCTTTCCAGTTCGGCCTGT
		GGCCCATCCCTCTCTCACCGCCCGGCAGATTGGAGTCTGCTCTCGGGGAGCCCCCAGGTAAACCCCTCACAGGGAGAAGGTTTCGGATTGG
		AAGGAGGACCGCGCTCGTGGGGCGCCTGTGAGAGCTGGGAAGCCCAAGGGGTAGCGTGTAGGGGGTTTTTTATGCGGGAGGAGCTGCC
		TCCTGGGCGGCGGGGACTTTCTGTCTCAGCCTGTCTGCCTTTGGGAAAACAAGGAGTTGCCGGAGAAGCAGGGAAAGAAAGGAGGAGG
		GAAGGAGGGTCCTTGGGGGAATATTTGCGGGTCAAATCGATATCCCCGTTTGGCCACGAGAATGGCGATTTCAAAGCAGATTAGATTACTT
		TGTGGCATTTCAAATAAACGGCAATTTCAGGGCCATGAGCACGTGGGCGACCCGCGGGAGCTGTGGGCCTGGCAGGCTCGGCACAGGCG
		CCCGGGCTGCCGGCCGCTGCGGGGGATTTCTCCCCCAGCCTTTTCTTTTTAACAGAGGGGCAAAGGGGCGACGGCGAGAGCACAGATGGCG
		GCTGCGGAGCCGGGGAGGCGGCGGGAGACGCGCGGGACTCGTGGGGAGGGCTGGCAGGGTGCAGGGGTTCCGCGTGACCTGCCCG
		GCTCCCAGGCATCGGGCTGGGCGCTGCAGTTTACCGATTTGCTTTCGTCCCTCGTCCAGGTTTAGGAGACGCGTGTGGGGACAGCCGAGCC
		GCGCCGGGCCCCCTGGACGGCGTCGCCAAGGAGCTGGGATCGCACTTGCTGCAGGTAGAGCGGCCTCGCCGGGGGAGGAGCGCAGCCG
		CCGCAGGCTCCCTTCCCACCCCGCCACCCCAGCCTCCAGGCGTCCCTTCCCCAGGAGCGCCAGGCAGATCCAGAGGCTGCCGGGGGCT
		GGGGATGGGGTGGTCCCCACTGCGGAGGGATGGACGCTTAGCATGTCGGATGCGGCCTGCGGCCAACCCTACCCTAACCCTACGTCTGC
		CCCCACACCCCGCCGAAGGCCCCAGGACTCCCCAGGCCACCTGAGACCTACGCCAGGGGCGCCTCCCGAGCGTGGTCAAGTGCTTTCCA
		ATCTCACTTCCCTCAGCAGGTTCCACCCAGCGCTTGCTCTGTGCCAGGCGCCAGGGCTGGAGGCAGCAGAAATGATTGGGCTGCTCTGAGC

SEQ ID NO	GENE NAME	SEQUENCE
		TCTGAAGCATTCGGCCGCTGTGTGTGCAAGGGGCGCAAGGACGGAGAGACAGCATCAATAATACAATATTAACAGGAGCACTTGTCCAG AGCTTACTGCAAGCCACATTCAGTTCCGGACCTTATTGACTTCCCCCTCCCATCTAGAGTGGATTCTGGTTTTTTCAATTTGTTTTGTTTTGTTT TTTGTTTGTTTGTTTTTGAGACGGAGTCTCACTCTGTGGCCCAGGCTAGAGTGCAATGGCGCGATCTCGGCTCACTCCAACCTCCGCC TCCCGGGTTCAAGCGATTCTCCCGCCTCCCGAGTAGCCAGGATTGCAGGCACCCGCCATCATGCCTGGCTAATTTTTGTAGAGAC AGGGTTTCACCCAGGCTGGTCTCGATCTCCTGACCTCGCCCACCTCAGCCTTCCAAAGTGTTGGGATTACAGGCGTGAGCCA ACGCGTCCTGCCTTGATTCTGTTTTTAACTCCATTTTTTAGAGGAGGAAATTGAGGCACAGAGAGGTTAAATAACATGTCTAAGGTCACACAG CAAGGGGTGGAGCGGAGTTAGCCCACTGGCCTAGCTCTAGAGCCCACCCGGATAACCAGAACTTGGTGAGGCCTCCGGGCTCTTGCTTG GTTTGGAGCCAGGTGCTTAGCGCCCCGAGCCCGGGGCCATTCACCCTGCAGGAGCTGCACGCGCCCCTGACCTCGGCTTTTCCCTGGCA GCAGAGGGGCTTTGCGGGTCGGCCGGGTAGCCCTGAGCACAGCTCGCCACTTCCAGGTGGGCTGTTGGCGCTGGCTGGGGACACATCC CGATCTTTCAAATGCCCTTTACAGAGCCTCATCAACGACCCGATTCATTCCCCCCTCCTGTCATTTGTCTCTGCCATCGAAAAATGCCTACC GAGAGCTGCTCTGCATTTCCGCCCTCTATTTTTGTGTTTTACTTTAAAATAATAATAAAAAAAATGTTGGCTGCAGGACGCCATGACTTAGGTC AGCGAGTCAGCCGCTAGCTCTGCATTTCCAAAAAGCAGATCTTTTCACAACTCTCTTGCCCCAAGTGCCCTGGTGTGGTTTATTTTTTAAAAT GCATGCCTGCGGAAGAGAAGACCCGGGGAATATTCGAAACCCCGAGCTTTTACAACATAAAGCGCATGGTGTGGCCGCGGCGAGTAATGG CGCT
206	HLCS	CAAATCACTTGAACTCAAGTTCAAGACCAGCCTGGGCAACATGGTGAAACCACATCTCTACAAAAGTAAAGAAAATTAGCCAGGCATGGTGC TGTGTGCCTGTAGTTCCAGCTACTCCTGGGGAGGTCGAGGCTGCAGTGAGCCGCAATCACGCCACTTGTACTCCAGCCTGGGCGACAGAG CAAGTCCCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGCTGGGTGTGGTCCCAGATACTCAGAGGCTGAAAAGGGAGGATTG CTTGAGCCCAGGAGTTCAAGGCTGCAGTGAGCTGCGATCACACTCCATCAATGCACTCCAGCCTGAGCAATGGAGTGAGACCCTGACTATAT TTAAAAAAAAAAAAAATAGGAAGAAACAACTCAACCACAGGGCTAGTATGTTACTCGGTTATAAAATGATAAAGCCCTAAACAGAGAATTAGC CCGTTTCCAGAAGAGGCCAAGAACAGATGATACAGCTGAACTGAACTCCTGCCTGTACAGCTCGTTTTCTACAAGATTCCAGACCTGGAAG ATGATGGCATCCAGCCCCCCATTGAAGCACCTCGAACAAGAAAAACGCCGAGTCCGAAGAGCCAGGCCTTGAACACACGATTCCTGTCTATA AATAACTCCCCCTGGGGAATAAAAAGCAGGATCCAAGGCAGGAAACCCGAGCCGTGGAATCTGGTAAGTTCTTAGGAAACCCACTCACGG GCCTGAGTCCCCCGTGGAAGCGGCGACTTCGGCACCTGGACACCCGAGTCCCCAGAGCCCCGGGCCGCGCTCCCTACCTGCAGG CCTGATACCGGCCGCGGAGCGCTCCTGGCCCCGCTCCCGCCAGGCTCCGGGACCGCTGAAACGCACACCCAGGGGGGTGAAGGCGTAGTC GCCAAGGACAGCAGCGCAGATGGCAGCGGAGGCATGGGAGCCGGAACCTACCGTGGCAAAGGGCCAGGTCGGGACGCCCCTCGGCGCAGC CCCAAATCCTGCCCGCGCCCCCAGCCCCGCTCAGGCCGCGCCCCTGCCACCTCTGGCCACACGGGCTGAGACGGTCTGGCTCCTGCACAGC GCACTTCCCGCTGCCCTTCTCCACTGGCTGCTCAGGCCCTGCCTCGCCAGCACGGCATCCGCGGGGGATCCCTACCTGTCCTTTAGGGCT TGCCTCATAGGTCAAACGTCACCTCCCAGGGAGGTATGGCCTGCCCCCTGGCCAGGTGGGCCCCTTCCACGCTCGCCTGCAACACCACCC ACCCACCTTGATAACTGCTTGTAAAGGTTGTACTGCTTTCCCCCTTGAGACTGCAAACCTTCAAGGGCAGGAAATGGGTCTGTTTTCCTGGC AAAATAATGAAGTTGGCTTAAGGTTTTTGCTGAATAAAATGAGTGACAGACAAAAGTAGCCAAATTTGGCACTCCTGATGGGTTATTTGATGAA GGAGGTGCAATGTATGGGCTTAACTAGTTATTCTGGATTTCTTTCCCCCATGTTA
207	DSCR6	CAAGGCCGGTGCACGCGGACCCGAGGATTCGGTAGATGTCCCCGAAGACCCGCTGCCGCTCTAAGGCGGTGGAAGCGAGATTCTCCGGA AACCCAGGGAATCCGATGCTCGCACAGGACCAAAGCCCGAGGCCGCGGGGACCACAGAGGGACGGAGAAGCCGGGACTCCTCACATCC CACATCCGGCAGGGGAAGCCCAG
208	DSCR3	CTGATAATAAAGTTTTACCATTTTATAATTTAAAAATGTAAATATGGAGTTGGGCATGGTGGTTGGGAGGCTGAGACCAGAAGATCGCTTGAG CCCAGGGGTTTGAGACCAGCCTGGGCAACATGCAGAAACCCTGTCTCTACAAATAAAAAATTAGCCAAGCGTGGTAGCACGCACCTGTAAT CCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAGCCTGGGAGGTGGAGGCTGCAGTGAGCTGAGACTGTACCACTGCACTCCA GCCTGGGTGACAGAGTGAGGCTCTGTCTCAAAAAAAACAAAACACAAAAAAAACAAACAAAAAAAGCAAATATATGTAAAAATAGGAAGTGCG

SEQ ID NO	GENE NAME	SEQUENCE
		GTTTCCCAAAATGAGGTCTGTAAACAACTGATCTAGAAAATGTTCTGGAAAAAGTAAAAAAGGATCAGGATCTGAGGTCAACTGACCTCTCC CTGCGCTCTGGACAGGCAAACAGGCAAGGTTCCCTCTGAGGCCGTAGCGGCTTCTCGTGGGCGAGTCCCTGTTCGCAGGTGACGTGTGG ACCACGCTCTTCCGAAGCGTCTGTGCTCTCGGGGAGGGGACGCAGGTCAGCCCACCTAGCCGATGGCTAACAAGTCAGTTTGT TTTCTGAACGGAAGCTTAAACCTAGAAAAGTAACTGGGTTGGGGTGGGGTGTAGCCACATGCAGTAAAAGCACTGCCTGTCTGTATAACA ACGACCTGATGAAAAAAGGAACGCGTGAAATGGGGAGTGTTAGGGCGTCACAAACTCCAGTGTGGTTGAAATGAAAGCAGAAAGCAAATG GCAAGCTGGCTTCCCCTTCCAGCTTTTCACAACCCTGCCTTGCTCATGGTCAGCCCCAAGCACGGGCGGAAGAAAGGACTGGAGGGGAGG GAAAGGGGTGGGGAGCGAGGGTACCAGAGGCGTGGGAGGACGGGGACAAAGGGGCAGCAAGGGACCGGCGGAAAGGAAAGTCGGCGT TAGCTGGATTGGAAACAGTCCAGACAGAACGATGGGCTCTGCTGCCTCCGGGTGGGGCACCAAGCGGGAGCGGGGCCACGAGGCAGG GGACAGTGAAGCACCATGCAGCGCCCACCAGCCGGCAGCGCCCACCAGCCTGCGCTGCGCTGCACATGGTACCCGCGGCCCCAGCTGG CCAGTGTGTGGCGGAGATGAGACCCTCGTGAAGAGACTAAGCGGCCACAGCAGGGGGAAGGGTTGCTCACATAACCCCATACTGCTCACA CTACGAGGTTAACTGCCGTGAGATCTGCCTGCAGCCAGCAGAAACCCGTTCTAGGAAAACGTTGCCCAGTGACTTCAGTGAGTGCCACTGA CCCGGGCGCCTCCGCCCCGGCGTCCGGCAGCAGCACCGATTGCGCAGGAGGCACCTTGCAAACAACCTTTCCTGATCCGCGCTGCAGTT CCCAGGCCGGTTGCAGCCGTTTCACAGAGACTGCGCACACAAAGCGTCTCCGTGCCCTGCCATTCACCTTTCGACACAGCCGCAACCCCT CTTTTCAGTGTTAAAACCTGGCGCCAAAAGGAACATGCGATGTGACGTGTTACCTCTGCGCATGCGCCGGGCATTCCCAGCGCCCCGAAC CTGATGAACGCGCGCGGTGGGGACCCCAGGCTTCCGTGCTTTCGTTTTCCTGGAAGCTACGTGTCCTCAGTCTACATATTGTTACCTGGAAAA TAAAGTTTTCTCCTTTTTTCTTCCTTTGTTAACAGGCAGAAGGTGTAGGCTGCAGGTTTCGGGCCTAAGAGAGGGCATGGCTGGCGACACG GAGTAGACTCCTAGATGACATAACGGAGGCGAGTCTGCACCGGGGACTCCGGCATTAGGAGGAGGCAGAGGAAAAGCCCACCACCGTGGC CGAGGGAGATCTAGCAAGCAGCTTGCAGGGGGGTGAAGTGTGTGCAAAGCAGGCTGAGACCTGTCCAGTATCGAAACACGCCCGCGGTGGT CAAGCAGGCTTTACCATGCT
209	chr21:37841100-37841800	TGAGGCTCAAAACAGGTGTCTGTGAGCTTCACAGGCGGTAAGGCCGTGTCTACATGGCCGGGACATGCATCCCGGGGCTGCCCCTGCCG TGCTGCCCGAGTGCACGGGGGATGAGGACCTGACAAGGCCATTGATCTTGCGGGAGCTTCCTGAACTACTCCAGCGTGAAAATCTTCCAG AAGGATTCTCCACAGGGCAATGAGGCAAGAAATTTACAGCTTAGCCTGATTAATGGGCCAGGCAGTTAAGAGTTCTTTGCCAAGCTATGAG CATAATTTATAGTCATCACGGCAGGAGGAAAGGCCACATAACTCACATCCTTAAAGGGCCCTTAGAACAAGAGACACGCCGGATCATTGAAA ACGTCTCCACTCCTGGCGCCAAAAGAGATCGGCACGTTTCTGGGTATTCTGGTCAAAGAACAGGGAGTCTGGATTAATATACACGGCAGAA AAAAGCGAAGAAAAGACACACAGGTCATATATTTCTGACTGATATTCCGTTTGTTGTTTCGGAGGGACTTGGTATTTATTTAACCACATTCT CACTTGACACGCCCCCTCCCCACACCTTGTAAATGCCTTCCTCTTTAGCCGAGTCATTTTTCATCACATAGAATTGAAATGTTGCCAGGAAG GCGGTTTATGAGATTGTAGAAATGGCACTAGAGAAAGCAGTGTGAAAAGAGGCCTAGAACGT
210	ERG	TCTCTACATGCTATCTACTAAAAACTTAGGCAAGGAAATGCATCAGACCAAACACCCCACAGCACAGAGAACCGACCGGCCATTGCTTTCCA ATCTCCGCAAACCTAACCATTGCTGGAAGAAATCTTACTCACAGTGCACAGACAGTAGGTATTTTATTGAAGATAAACATATAGTGGAACAAA CCAAATTACCCCCATTTGAGTTACGTGAGCACTCAGTTCTCAGCGTGGATGTCCCACAAATCAAGTCAACATTTGCGTCCCATTACCAGCAG CCACTTGCCGAGTATCTCTTCGCTTCCACTGGGACTGCCTGGCATCCCTGATGCTAAGGAGCCACTGAAGAGCCTCCAAATGTCTGACATT CACAAACGCATCTTTTGCTTTGACCCGACCCTTCAACCTCTCCGAGTCTGCTGCCTTTTCTCAGACACACATCCAGGCACCGTTAGGGATAG TTAGAGAATCTGAAAATTCAGAAGCGCTCCGAAAAGCCTTTCCAAAAGTAATCCACAGCACTCAACAGTGAATTTAGAAACCCCAATTTTTTT CTGAGTTTGAAGTTTTTAAGCCTTGCGGATGGTTGGAGTAGGAAAAA

SEQ ID NO	GENE NAME	SEQUENCE
211	chr21:39279870-39279800	TCAGACAAGCTCTGTGCAGTCGGAATTTTTTAAAGATGCACTGTCACTTGAGGAAGACAGGTGATCTTCCTGCGGCACAAATAGAAGCAAAG AGATTTCTCTTCTTCTCTGTAGAGCAACACAATTGATAAATGGCCGATAATCTCCACCAAATTGGCAGCAGTAGGCTGCCCGAAGGCAGCAG GCATATTCGTCTTTGTGAATTGTTTTACTATGATGCTGTCACATTTCCAGGAATAAGACGGTTAAAATGATATATTGTTGTGGTTTGGCATTTG CAGCTTTGCTCTGACTTCCCTGGTAACTGCCAACATCTGCAAATTATTATGTGCTTAAAAAAAAAATCAACCGCCACCGCAGGCTGCCCCCA CGGTCCCTGGCTGGGCCAGGCCTCCTGCCAGGGCACAGAGTTCTTGGACCAGGAGGCAGCAGGGTCAAAACCCAGGTTGCCTAG GAAGCCCCCAAAGACAGTTATGGATAGAGCTGGGAGCCCGAAACACATGCGGCAGTCTCTCAGTTTCCAGGTACCGGTTCTCACATCATCC ATGCATGTGTTTGAGGAAAAACAAAAAAAAATTGATGGTTGCCAAAAACAAAAATGCTTCCATATCAAAGTTTATCAGTGTCAATGTCAAGAG ACTTCTGGTTCGTAGACTCATTTTGGCTTGAGGCCACCAGAAGTGAACTCTGGTTTCTAAATGCAGAAGCAGAGGCACTGGCCGATCATGG AAGATGCAGGGAACTGTTCAAGAGGCCCAAGCCTGGTGCTCAGAAACTTGGCAGGATCAAGCATCTCGCCCAGGAATTCATCCCCTGCTTG TCTAAGCCGGCTGGCTCTCGTGACTCGGAACAACAGAGCAGATGTTTGCGTGGGAGGCAAGCCTCACCCAACATCTGTCCTGCGGC GGGAAGGCCTGGGTGTTCACAGATAGAGCTGGAGTTCCCCGGTGGGTGGCACAGACAATTAGCTGGGGCTGCCTCACATGTAATCTAATT ACAGGGGAAACAGGCTCAAACACCGGGTGATAAGCAGCGCAACTGTTTCGGGTGACTCTGTAATTTTTCCTCCATTAATTTTCTCCATAACG CAC
212	C21orf129	GTTGCCTGGGATATGCTTATATCAAAAACTTACGTGTCACTTACCTAGCATTTGCATTTCACTGGGCCTCCTAAATTCTGTGTGGTAACCGAC TGCCACCGGACATGCTGTTTACTTCTCTATCCTCACGCAGCCAGTTGCCACATTCAACATAACACTGCAAATATTGCCGGTGGATCCTGACT TCCTCGTGGACCCTACTGTGTCGGGAAAAACAAACGAACCCTGGAAGGAAACACCATGAGT
213	C2CD2	TCATAAATATTTCCAAATGTATTCCTATTTGTCTCTACAGAGTCTAACAGACATAAATAGCGAATTGAAGGTTCTGTCTTAAAACCCAGCAGAA AGAAAAACAATGACCAGAAAAAAAAAACAATTGTCTTTGGCTTCCCAAGAACAGCATCGGATTTCAACTGGAACCACAGATGGTCCGTTGAT AGAAGCGACTACTTTTTAGCTCTGGAGGACGACAAAGGAACCAGCTTCTTCCTGTGGGTGTCACAGCGAGGTCGCCTGGCCACATCAGGT ACCAGAGCGAGCGCCCTCACCTGATAGGCCCTGTACAACCTCAGCCACAGCACTGTCAGGAGGAACACGCGGAACTAGCAACCTAGGAG GGTAAAGGCGGAGTTGGGAGGGAACACGAGGCAGGTCGGCTGGCTGAGCTACAGGCTGCACTCCTAGGACGTCTACGTGTAA TTGAGAAAAATAAGACAAAAATAACTTACTGTGCAGGCAATTAATTCTGGTTGGCATAGCGATCCTCTCTTAAGTTAAAGGGAATGAGCATGAGA TGAAGAGAAGTAAGAGGCAGAAAGAATTATGCAAGAGCAACATCAGAGTGGA
214	UMODL1	ACGCCGAGCCGCCTCTGCAGGGGAAACCGAAGCAGATGTGGTGAGATAATACATCCAACCCTGAGTGCTACTCTAACCTGCCAGAGGCGG AGGGTTCTCAGTGAGATGAAAGCATTACAGATGCGTTAGATCTAAGGGAGAGGGGCCTGCAGATGCGCAGCTGGCAGAGAAACCAGGGAGG GGCTGAACTGTCAGTCGCGACCACCAGGGATCTGAATCAGTTCACCGACACAGCCTTGGGGACATTCACCTTGGGCTCCACAACCTGTCAGA AATGCCCCCCAAGCCCAAAGGCGTCGAGAGAATGGCCAGGTTGTTTCAGATTGACACATATCCTAATGTACAAGTCAGCCCACACACCCCAC GTGCACTGAGCGTCTCTTGTTGTTCACCCCAAATAAACTCTGCCGGAACTGGGGCGGGACTCGCAGGGGCGGAGAAGGGGGGGAGACGGG CAGAGGGCAGAAGTGGATGGTGAGAAGAGCCAATGGAGAGGGGCCCCGTGAGAGTGAGCAAGGCTGCACACCCCTAACCGACGTCCTGGGGC TACTGTACAAACAAAGAACCACAGGCTGGGAGGCTGAACAACAGACCTGCACTCTCTCGCAGCTCGGAGAGGCTGCAGGTCTGAAATCGAGG GGCTGACAGCGCTGGTTTCCTCTGGAGGCTGCGAGAGGGAGAAACCGTCCCCTGCCTCTCCCAGGCTCTGGGGTGAGCCCTTCCTGGCATC CCGGGCTCATTGTAGATGGATCACTCCAATCTCCATGGCTTCTCAGGGCTTCCCTCCATGCACCTCAAATCTCTCTCTCCTTCCTTTTGTAA GGATGCCAGTCATTGGATTTAGGTTCACCTTAAATCCAGGATGATCTCATCTAAATTACATCTGCAAAAAGACCCTTTTTCCAAGTAAGTTGA CATTCACAGGTACCTGGGGTTAGGATTGGACATATCTTTTGCAGGGGGTGCAGGGGGCTGCCACTGAGCCCGCTGCACAGGGTGACCTGGG CCAAGGGCCCTTCACTTTCCTCATTGGCAAGCTGCCCTGTGTTTGGACTGGGTCGAGGCTGTCAACCTTGCTGCCCCTCGGAGTCCC CCCCTGGTGTCCCCCAAACAGATTCTAAGCTGCTTTCCTGGGGCTGGAGGCCAGGCATTGGGATTTTTTAAAGAGCTTCCCAGCAGGTGAG CAGCCTTTCATGGGTATCAGGAGACCTTCCTGGCAAATGTGGTGAAGGTCCTTCCTCCTGAGCGATGCCTTAGACCCAGGAGCCCAGGGA GGCTGCTCACCTGATCGTTAGGACAGGAGCAGTGGAAACCTCTGGCCTCAGACCCCCTGGAGGAATCCCTCCCTCTAAGACTCTGGGACT

SEQ ID NO	GENE NAME	SEQUENCE
		GGTGCACGCAAGGAGCTATCGTGAACATTGCTCCCAACTGGCCGCTTGCTTGTCCCCGGCTCCCCTTGGCCCCAGTGGCGGCTTTGCCT GAATTAGAGGGCGTGAGAGCCACCTGTGTCTCAGCACTGCAATTAAAGCAGGAAGCCCTTTCGGAAGCAGCCGTGTGCACCAGCCTCCCA TGGGTGGAGCAGAGCAAACCACCCACTTCTGCCCTCTGCCCTTTTCTCGACACCCTGCGGCCCCCCAGTTTCAGCAGAGTTTA TTTGGGGTGAAAAACAAGAGATGCTCAGCGCCTGTGGGATGTGTGGGCTGACTCGTACATTAGGATGTGTGTCAATCTGAAATAACCTGGC CGTTATATGGATGCCTTGGGGGCTTTCTGGCAGTCTGTCGAGCCCGAGGTGAATGTCCCCAAGGCTGCTGGTGAATCAGATCC CTGGCGTTCTCCGTTGGCAGTTCAGCCCAACAGTTTCTCTGCCGGCCGTGCCTCTGCAGGTCCCTCCTCTGATCTGATTGGATTAATATTTG AATCAATAGACTGAGTCAAGCAGAATGTGGGTGGGCCCTCATGCAATCAGCTGAAGCCCTGAAAAGAGCAAAAGGGCTGCCCCTTCCCCCG AGGAGGAGAGAAC
215	UMODL1/C21orf128	CACATTTCAGAGCTGAGGTGCTGGTGCGGGCAGGTCTCCTGAGCTGGGGGGGTCAGCTGTGTGGCCAGTGATGGTGACGCCTCAGGCCGT GCATGGCCGGGGAGGCGGCCCTGCCTCTGCACTCTTTTGACTCCATGACTACTGGTGTCTTCGGACGCCAGAGTCGGGGGAGCAACCAT GGGGCACCGCCCCCTGCCTGGGGAGGCAGCACGAGGCCTGAGCCCAGCTTACAGGGGGGACATCCACCCCCGCTGAGAGCCCCCACCTTCA CGGCGAGGATCTGTAGAAGAAGACATTTGATATTACTCGGCAAAAAAAACAAGAAACGAAAACACAAAAAGAGCTCCTCTGAAGAAGAAAAG GTATTTGCGCTGTGGTCCACCTAGAAATAATGTTGTTGGCACAACTAGAGCATTCCTCAGTCATTCAGGAGCACTCCCTGCCGGTGCGTCC ACATGTCCCAACCCCGATAGATGAGGCGCTGTTCGCCCGTGGAGGGGTCAGGTTGTCGTGACCTTATCTTTACCCTTAGGCCGTCCATCCC GGGGCCTGGGGTTTCCTGCGCCAGTCACGGTGGGCTGTGTAGGTGGCCATGTGTTCGGTCTTTCCCCAGGAGGTACGTACCATGTGCTG GGAGGCCTGGAGGCTGAGCCGCCCCCCGCGCCCTATGAGTTGCACCCTCACAGCGGCGGCCAAACCTCCTGC
216	ABCG1	CAGGCTTGAGCGGTGACTGGGAGACCCCGGGAATGGAAATGGCGCTCAAATGCTGGTGTGGTGTCCGCAGGGGAACGGCCCCGCGGGTG TGTGGAGTCTGCCGCCCCCTGTGGCTTCAGCTGCGTCGGGGGACTGCGGGAATCTTCCAGACTCCAGTTTAAATCAGAGAGGGTGTGTCCACG AAAAGAGTCAAACTAAAACATT
217	chr21:42598300-42599600	AACGAGACAGTGCAAAAAGCCGCTGCCTGGTGACCTGGCATGCAGACTCGGCCCTCCCACTTGCACGGTGATCCACTGAAGACAACAGCT GCCTCTGTACTCACGCTCCCCCCACACTCCCCTCCTTCCTGCCCTGGTTTCTCCATCCCTAGATGCCATCCCCAAACCATCCGCCA AGCACAATAACCTCGCCCCCACCCACCCCATGAGGTCACTCGAGTTGACAACCAGATAACAGTTTTTGTTTTGTTTTGTTTTGTTTTG TTTGTTTTTGAGACGGGGTCTCGCTCTGTTGCCCAGGCTGGAGTGCAATGACGTTATCTCGGCTCACCACAACCTCCGCCTCCCGGGTTCA AGAGATTCTTCTGCCTCAGCTGCCTGAGTAGCTGGGACTACAGGCGCGTGCCACCATTCTCAGCTAACTTTTGTATTTTTAGTAGAGACAGG GTTTCATTATATTGGCCAGGCTGGTCTCGAACTCCTGACCTCTTGATCCGCCCACCTCAGCCTCTCAAAGTGCAGGGATTACAGGCGTGAG CCACCGCGCCCAATAGCAATTTGATGACCCATCCCCTCCACTGCTGGGAAAAGGCTGGGCACCGCCCACACTCCATGCAGCTCTCTTTCCC TGGCTCGGAATCGCTGCAGGCGCCACAGACCAGACGCGCACTGTTCCCCACTCCTGCTTATCGGCCGCGCGGCATCCCCTTGTCGCAGC ACTCCAGCATCCATGCAGCCGCGCGGCACCCCGTCTTCGGAGCACTCCAGAATCCATGCAGAGCGCAGCACCCCACATCCAGAGCGCTC CAGAATCCATGAAGCACGCGGCACCCCCTCGTCAGAGTGCTCCAGAATCCATGAAGTGCGCAGCACCCCTTAATCGGAGCGCTCTAGAAC CCGTGCAGCGAGCAGCACCCCACACCCGGAGCGCTCCAGAATCCATGAAGCCAGCAGCACCCCACACCCGGAGTGCTCCAGAATCCACG CAGCACGTGGCATCTCCTCGTCATAGCGTTCTAGAATCCATGCAGCGAGCAGTACCCCACACCGGGAGCGCTCCAGAATCCACGCAGCGT CTGGCACATCTTTATCAGAGCGCTCCAGAGTCCATGCAGCCACAGTCCTCCAACGGACCCTGAGATTGTTTCTGCAAAAGGCCATGCCTTC ATAAATCTGAAAATTTGGAAAACATCCTTCTACTTATATCCTTACAACCCACCATTCAAGCTGTAGAAGCCTTTCTGGAACCCCAAGCAGAAG GATATCCAAAATGTAAAAACGGTGGGGCCT

SEQ ID NO	GENE NAME	SEQUENCE
218	chr21:42910000-42911000	ATAGTGCGACTGTTCCGAAGTCTTTATCACAGTTACTGGTGATGCTTTTTTTCCAGATGTCCTCGACGTGCACCCATGAAGGGCTCCACCTGA GAGTGCCAGGGTCCTCCGTGGGATGGGGCTGGAGGGGGTGCTCTTGCCGTCCTGGGCTCCCAAGCAGCCATAGGAACAATAGGGTGATG GGGTCCCAGAGATAGAGGCCAGTGACAGCAGCGCTTTGAACCCCTCACACGGGCACGGGCCCTCTGGCAGGGATGGGCGTCCCGGTCAC ACGGAGATGGGGGCTGCTGCCTGCAGGTAGAGGAAGGGACGTGTTTGGCAGTCCTGTGACCCCTGGGCACCTCGCCTCCCCCACGG CCGGCTCTGCTTGTAAACAGACAAGTGCACAAGCGCAGCCCGGTGAAGGCACAGCGGTCCCAGGAGGCATCTGGGCTGCACCCCAGCGA GCCGCCCATACACGTGGAGATGCCGGCCAAGGCCCTGCAGCACACGGCAGAGGAAGGCGCGATGGGAGCCATGCTGGGCCCGGAAGGT GCCGCCGCCCCGGAGCTGTAGCCATCACTCCAGCTCTTCTTTTAAGTGTTCCCAGAAATTGTGACCCACCAAAATCTGAGAGCACCCGACAG TAAGCCAGAGGACCTTGATGTGAGATCCCAGCACGGTGTGGGGGCGGACTGTGGTGGGTGCTGTCTCGGCCCCCCACCCCTTCCACAGGT CGGTGTGCACATCCCCACGGCGCCTGCTAAGCTGCAGTCTTCTCCAAAGGGGTCACTCTCCGTGGGAAGGGAGCCACCCGCCCCCGGGTG ATGTCCCCAGTCAGTGACTGACGACAGTCCCCAGCCGAGGTGAGGGACCAGCTCCTGCATCCCCTCACTCCGGGGGCTTGCCTGTGGGCCA GGGTGGGGGGCGAGCCTCAGCAGAGACCGCGTCCCCCTTGCCTGTCCTGCCCTCCCCCTGCCTCCCCCGCGCCTCTGCTGAGCACGC CCAGAGGGAGCTGCTTG
219	PDE9A	CACTTGAAAAGCACAACTCATGGTGCCAAAGCTCTGACACGGACTCCACTGGAGCTGTGGGCAGGGGGTGCCAAGGTACCGAGTTCCAAG CCGTTGTTATTTGAGAGCGTGCCCCCCGCCATGAGAGCAGGTGGGGGGACATAAAGTGACACAGGATGGACTGGCCAAAGGCTGAGGAC GATCACTTACCTCACAGGATGATGCCACCCCCACGGACAGGCAAGGAGCTCTCACCTTCCCCAGGACCCCAGCTGCCACCAGAGCTCCAG ATGGCCCTGGGGGTGTCTGTAAAGCCTGTGACCGTCCACCAGGTGGAGACCAGGCTGGCCAGGGGAGGGAGGAAGTGACCACTGGC CCTGGCACTGGCTGGCCGGCTCCAGCAGGCCCGAAGGGGAGGGAGGCCTGGGTGCACCAGACTCTCTCAATAAGCAGCACCCAGACA CTTAACAGATGGAAAGCGGTGGCTTGGAACTCACTTCCAACGAAACAATAGCAC
220	PDE9A	AGCACCTCCTACCCCACCCTCCCCATTCCTGCCATCCCCAGGGTCCAGGGAGCCCAGATTCCAGGGAAGGGTTGCATTAGCTCCCACTCG GAGTCCTGATGCAGCAGAGACAGACAGAGGCCCTGGGAGAAGTGAGCATGAATTATTAAGACAAGACAAGGACAAGGGTGAGGCCCCAGAGAGGG GGTGGCGGAAGGGTCATGTTCATGCAGCGAGAGTTGCTTCGAGCTTGAACCGCGTATCCAGGAGTCAAGCAGATTGCAACTGGCGAGAGG CCTTCAGAAATGCCCCCGTGAGAGTCCTGTGTGCAGAGCTCCATCTCAGCACACTTCCTGTTCTTTTGGTTCGTCGATTTTTGCATTTTTCAGTC CCCTGTGATCCATTATTTATAACAGTGGAGATTGGCCTCAGACACTAGCAGTGAGGAAAACAAAAGCGAAGCTACGCAGAAAAATGACAAGA GTGATGAGCACAGCAGTCATGACAAATGAGCCCTGTGCGGAGGCCCGGGATCCGCGCGCAGATGCCGGCGGGGGAAATGGGCCCTGAA ATCCCACCGTCAGGCCAGGCAGCTCTGAGCGTGACCTGGAGGGCTGTTCAGACGGTCTGGGTAGCCGTGTCCTGCGCATGAACATCCTCC GTCGGGAGAGGAATTCCCCACGGATTATCAGAGCTGCTCCCTCACCCCCCGCCACGTCCCACACATCAACTCCCTCTGCAGC CTCTGGCCAGCGGCTGAGCCCTCCGTGTCTCCCCTCGTTAATGCCTCCTTCACCATCCCCTCCTGAAGTTTCCCCCATTGCATACACGCGC TGAGGCCCACCCGGTATCAAGGACTCCCATTGCTTGCGAAAAAGATTCCACCCCTCTTAGAACAGAGACCAGGGCCGCTGTAGCAAATGG CCATAAATGCCACAGCTTAAAACAACAGAAACGGATTATCTCGCAGCTCTGGAGGATGGAGTCCAAAATCTGAATCGCTGGGCTGAAATCC AGGTGTGGGCAGGGCCGCGCTCCCTCTAGAGGCTCCCCCGGAGATTCCCTTCCTTCCTTCCAGCTGCTGGTGGCTGCCAGCAGTTTG GGAATTGCGGCCGCATCACACCACCTTTCTGTTTGTTGACATCCCCGCCTCCCCTGCCTGCGGGGTCTTAGATGTCTCTCTCCTTCCC ACTGAGTTTCACTCCACATTTGAATTGGATTAACTCATGCCATGTTAGGCAAACGTGCCCCTCAAATCCTTCCACTTAACAGACATTTATTGA AGGTTCCTGTGTGCGGGGCCCAAGAAGAAGGGA
221	PDE9A	GAATGTTCAAAGAAAGAGCCCTCCTTGCCTTCCTCTTCCACCCCTGCCCTCTGCAGACTGGGGTTCTGTAGACCCCCCAAAGTAAGTCC GCCACACCGGAAGGAAGTGAGTTACACAGGGGCCCACATGGGAACCGCTTTTTGTCCTGTCTTGGTGGGAAAATGGCCACGACCCCAGCC CAGGCTCTGCCACGCCACA

SEQ ID NO	GENE NAME	SEQUENCE
222	PDE9A	CCATCTTCCTAGGCCTGCGTTTCCCCCACACCGGGGACTTGTGCTGGAAAGAAAAGCTGCGTTGGCAGCCAGGAGCCGGGGAAACTGTCC AGGGAGGCATCCTCTCTGCGATGAAGGCGGGGCCTCGGCGTGGCCCGTTCCGCGCTCTGTCCAGCCCTGGAGAAGCCCCACCCTCACCGA GCTCGAAATACCCCCTCCCTGAGAGCCGAGACTCATGGCCGGGACCCCTTGGACAGAAGATGCGGATGCTAACCCGGCGCTTCCACCACA GCCCCGGCGGCACTGGGAGCGAGCGCGGCCATCCCGCGCGTAGGTGGTGTTTCTCTGCAGGCGCCAGTTTCACCGCGGGCGCCCAGG ATCCTCAACGGTTCTGTTGTGATGTGATTCCCCTCTTCGACTTCGTCATTCAGCCTCAGTCCCTCAGTCCCCAAATACCGAAAGGCAGTCTT TTTTTTTTTTTTTTTGAGACGGAGTTTCACTCTTGTTGCCCAGGCTGGAGTGCAATGGTGCGATCTCGGTTCACTGCAACCTCCGTCTCCCTG GCTCAAGCGATTCTCCCGGCTCAGCCTCCCGAGTAGCTGGGATTACAGGCACCTGCCACCACGCCCGGCTAATTTTTTGTATTTTTAGTAG AGACGGGGTTTCACCATGTTGGCCAGGATGGTCTGGAACTCCTGATCTCAGGTGATCCACCCGCCTCTGCCTCCCAAAGTGCTGGGATTAC AGGCGTGAGCCACCGCGCCCGGCCTTTTTTTCTTTTTTCTTTTGAAGTTAATGAACTTGAATTTTATTTTACAGAATAGCCCCCATGAGA TACTTGAAGACCCGGTGCCAAGCGACAGTGTTGACCCCAGGTGGTCAGTCCTGCCTGGCCCCTTCCGAGGGATGCGCCTTCACCATAACC ATGTCACGGACAGGCGTGTGGGCAAGGGGGCATCGCTGTATTTTTTCACAACTCTTTCCACTGAACACGACAATGACATTTTTCACCACCCGT ATGCATCAACCAAATGAAAAGATGAGCCTGTGACATTCCCGTGCGTAGAGTTACAGCTTTTCTTTTCAAAACGAACCTTCAGTTTGGAGCCG AAGCGGAAGCACGTGGCGTCTGACGTCTCCAGGGAGACCCCGCCGCCCTCGCTGCCGCCTCACCGCGCTTCTGTTTTGCAGGTAATCTTCA GCAAGTACTGCAACTCCAGCGACATCATGGACCTGTTCTGCATCGCCACCGGCCTGCCTCGGTGAGTGCGCGCTGCGGGCTCTGCCCGG TGACGCCACGCCGGCCTCCTCGCCTTTTCGGGATGGCTGGGAGGGGCGGGAAGAGGCGCTGAAGGGCCCGAGGCACCGGCCTTCTACAA GGGGCTCTTCGAAATCAATCAATGCGCAGAATCCCGAGGGAGGCTCAGCCCGCCCTCCGGGCCTCTCTGCCTCCACAGGTGATGGCTGTGT CCACAAGGAGGAAACCGTCGGGCTGAATTAAACAGAACCGCCCTCCTAAGAGTGTGGGTTTTTCTGCCGGGCGTGGTGTCTCACACCTGT AATCCCAACACTTTGAGAGGCCGAGGTGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGC
223	PDE9A	AGGCAGCAGGGTTAGGACTTCAACATACAACTTTTGGGGGGAGATGTACTTCAGCCCATAACACACCACGTGGGAGGATAACACCGATTTC AGAGCTTGCAGAGGAAGCCGCCAGGAACTCCAGTGAGACATCAGCCCCCAGGTGCCTGTCAGGCACGCCGGGCTGTGGGGGGCACCTG GGCCCATCTGAGTAACGGAGGCGCATCCGCACTTCCCCCCAGGAGTACATTTTTAGAACCCACAGCGCCATAAACCAAAGACAAGGAGACTT CCTGGTGCCCCGTCAGCTTCTGGAGGCGACGTTCTCGGCTGACAGCTCTGGCAGCCTCCCCCTGTAGGTGAGAGACAGGTAAATGGGACTC TTGCTTCCAAAACGGAACAGGGTAAAAATTCTCAAGCGTT
224	chr21:43130800-43131500	TGCTGCACCCCCGCTGCCCTCCCTCCGCTGGCCGGCAGCACCTTCTCCACCCGGGCCCCTCTGCTCACAGCGCTCCCCGCCCCCGTCT CCCCGAGGGGCGGGGAGCCAGGACATGGCCCTGAAAGCCTAGCCCTGGCCTTGACCTCCCCAGAGCGCCCTCCCCACCCTCCGCCCCTCT GCCAACCCTGGCCCCCTGCCCTGGCCCCGTCCTTGTCCTCTGCTGGCCTTGGGGTCGCGCCCCCGCAGACTGGGCTGTGCGTGGGGGT CCTGGCGGCCTGTGCCGTCCCACGCCTACGGGGATGGGCGAGGTCCTTCTTGGGGCTTCTCTTACCCACTCTCCAGTCACCTGAGGGCG CTGCTTCCCTGCGGCCACCCCAGGTTTCTGTGCAGCCGAAGCCTCTGCCTCTGCGGCCGGGTGATCCCAAGACCCCGGGGTCCAGGGAG GCACGGGATCTGCTCCCCCGGTCCCAAATGCACCGGCTGCGCCTTAGGAGGGACGGCCTCCACCCATGGCGCTGGCGCCCAGGGGCCG CTCCTCGGACTACAGCACTTGCTCGTCGCCCTGCGCCCTGTTTAGTTCTCATCACCAGCAGCCTGGACTAGGGCCCTGGTCCTTCTGGCCT CCTTCCACAGCCCGCTGCACATCTCACCCACTTCCCCGAGGTGCTGTCATTGTTTAGCTGGGCCCCTCAGCCTCCG
225	U2AF1	TTAAAGGGGAGTGGTTGTATGAAGAGTTCCTCAGTCAAAGGTGTGCAGCTGGGAAGCCCACCCCACCCTAAGAGGGGAGGTCTGACAAACTG TCCACACTGAACCACTCAGACCTGCATCAGGGCCCCCGTTTCTTCCATAAGCCGCCAAGTACAGCCCTGAGTCAACTGAACTCAGGCCTGGG AGGCTTCCCAAAGCTGACTTGACTCAGCTTTGAACTGAAATGACCGTACCATGACAACCCTGATGAAAAGCTAAACTGAGCCCAATTATTCA ACAGTAAAATTCAGTTGGTCTCACTCA
226	U2AF1	TGCTACCAGCTGCTTGGGCTTGGGCAAGTCACCCTAGCTCTCAGATGTCATCTGTAAATGATGACAATGCCAATGTGGCACTGTTCTGAGA GTCAGACAGAACGTATGTGTGCTTCACATATGGTGCTCATGAAGTGCTATCATTATCTAAGGAAAACAGAAAACGAAGTTCAGAGTCTCTCT

SEQ ID NO	GENE NAME	SEQUENCE
		AAACGCATGACACCAGACCAACAGGGAGTTTCAAAAAATAGGTCTGAAGTAAATCAATTCTCCTGGTCTCAATACACTGAAAACAAACTATTA GGGGACTGACCGAACCCACCTTAGGAACCACCTTCACGTCACCTTCTGTCTCTACTGCAAAACCCTCCCTTAATACTGTTCAAATACGCTGAC AATCCAGATCCATATCCAATGGAACCAGCAATCATGCCTGTGTGCCAGCAATGTCAGGGAGGGAAGCCGATCTCTGATGAAT
227	chr21:34466000-43447600	CAGGTGCCGGCCACCACACCCGGCTAATTTTTTGTGTTTTTAGTGGAGACAGGGTTTCGCCATGTTGGCCGGGCTGGTCTCAAACTCCTGAC CTCATGTGATCCACCCGCCTCGGCCTTCCAAAGTGCTGGGATTACAAGTGTAAGCCACTGCGCCCGGCCAAGAGTGAAGTTCTGATAGCTG GGGTAAGAAAGGCCGTGGGAACAGCCGGTTTCAGACACGCTGGGTCTAAGACGGCTGCGTCTGGCGCTGCTCGGCATCCAATGGGAGCCG TGGAGAAGCCAGGCGAGTGCGTAGGGCGGAGCCAGCGCACAGGAAATAGGAGACGTGATGAGGTCAACCGGCTGGTCCAAGTGTGGACGG AAGTAGAGGATGCAAGCACCGAGCCCCCGGGGCCCCCCAGCATTGGCGGGGGAGGAGCTCGCGGTGCGGGAGAAGCAGGGGACCGCGCAT CCTGGAGACCAGGTGGAGCCAGTGCGCCCGGAAGGGGCGTGGCCCGCTGACAGCCGCCCAGGAGGCCGGGGGAGGCCTGGAGCCGAG GGCCGCGCGTGGCAATGTGGAGAGACATTTTGGTGGAGTCATGGGGCCACAGCCTGATTGGTGAGAACAGGAAGGGAAATTGCAGATGG GCCTGGGCCCCCTGGCTCCCGCATACTCCAGGACCAGGGCTGAGTCATCGTTCACCGTGTGTGACCAGGGCCCCGTGTGGCCGGCTGTC ACTCGGTATCCAGTTACCCTGGGCAGACCACTGGCGGCACCCCCCAGCCAGAGGCCGCAGCAACACACGCCTGCAGGCGACCAGGCC GGACTGCATGCCCCGTGGGGGAACTGAGGGCGTTTCAGTAACAGAGTGTTAGGGGACACGGGTTGGGTGGCTTGGAAAGGGCCTAAGGT GGGGTTTGTTTTAGATTGGGGTGGTGAGGGCGCAGGGGCCCGGTAGGATTCTCTAACAGGGCAGCAGCCACTCATTTAGCAACAGGAGAG GCGTCCAGCGTTTCGTGGGCT
228	CRYAA	ACCCAACCACAGGCCTCCTCTCTGAGCCACGGGTGAGCGGTGCAGGTTCTGCTGTTCTGGAGGGCCTGAGTCCCACCCAGCACCTCATAA ACAGGGTCCTCCCCAGGGCTGCTGCAGTAGGCATCAACGCCAGGGTGCAAAATGCCTCAGGGAGCCAAGGCTGAGCCAGGGGAGTGAGA AGGAGCATGTGGAAGTGCGTTTTGGAGAGGCAGCTGTCGCAGGCTCCGGCCGCTTCTATAGACAGCATGACACCAAGGG CAGTGACCTCATTCCACAGGCTGAGTCCAGCCAAGCATCACCAGCCAGACGATTGACCCTAACGGACCAACCAACCCGTAAC GACCCCTCCTACCATAACCAGTAGCCAGCCAGCCCATAACCAGCCAACTTATCTATAACCAGCCACCTGACCATAGCCAAACAACCAGCCG GCCCACCAGTAGCATTCAGCCCCTCAGCTGGCCCTGAGGGTTTGGAGACAGGTCGAGGGTCATGCCTGTCTGTCCAGGAGACAGTCACAG GCCCCCGAAAGCTCTGCCCCACTTGGTGTGTGGGAGAAGAGGCCGGCAGGTGACCGAAGCATCTCTGTTCTGATAACCGGGACCCGCCC TGTCTCTGCCAACCCCAGCAGGGACGGCACCCTCTGGGCAGCTCCACATGGCACGTTTGGATTTCAGGTTCGATCCGACCGGGACAAGTT CGTCATCTTCCTCGATGTGAAGCACTTCTCCCCGGAGGACCTCACCGTGAAGGTGCAGGACGACTTTGTGGAGATCCACGGAAAGCACAA CGAGCGCCAGGTGAGCCCAGGCACTGAGAGGTGGGAGAGGGGGGCGAGTTGGGCGCGAGGACAAGGGGGTCACGGCGGGCACGACCG GGCCTGCACACCTGCACCATGCCTTCAACCCTGGGAGAGGGACGCTCTCCAGGGGACCCCGAATCAGGCCTGGCTTTTCCCCAAGGGAG GGGCCGTGCCCACCTGAGCACAGCCAGCCCCTCCCGGTGACAGAGGTCACCATTCCCGAGCTAATGTGGCTCAGGGATCCAGGTTAGGG TCCCTTCCCGGGCTGCACCCAGCCGTCGCCAGCTCCATCCCTGTCACCTGGATGCCAGGGTGGTCTTAGAAAGAACCCCAGGAAGTGGGA GTGCCCCGGGTGGCCGCCTCCTAGCCAGTGTACATCTTCACATGAACCCTACCTGAGGAAGCCAGTCCCCGACGGCATAGCTGCATCCGC TTGGAATGCTTTACAGGCATTGACACCTTCGCCTCACAGCAGCACTTTGGAACCAGTGTCCTCATTATTCCAGGGCACGGCTGGGGAACAA GGGGGTCCTCAGCCTGCTGGGTCCCACAGCTAGTACCGGGCAGGTGGACGGGAGCTTCTCCCCCACAGTCACCCTGATGCCCCGCTCTTG CTCGGCTGGAGGCCTCGGATCTCCGTGGTGTTGAGGGAGCCGGGGCACTGGAGCCCTGGTGACCTGCATCTCCTGGCGGAGCCGGGAA GAGCTCATGGACTGTCACAGATGGACAGTGCCCCGCGGGGGCTGGAGAGCAGAGTGGGGCTGGAAGGTGGAACTCTTAGCCAAAGTCTT GGTTTCTTTTGGCCAGGGTCCTCTTTCAATGGCTGGAGAAGGTGGTGCTGGGGGGGTGAACGCTGACCTCCTCATGTGCTGCCCCTCCCTC GCCTGGGCCCGGTAAAGCCCCCACGTAGCCCCAGCCAGCCTGGAACATGCTTCCTGAGCTCCCAGCTCTTGGTCTTTGCACCCAGTGGAG GAGGAGGTCAGCCCAGGGAGCTGAGTCTGCGGTTTAGGGCGTCCAGGGGACGTGGAAGCATGTGGGTCGTCTGGCCACATTAGGTAGGG CTGCAGAGACCTGGGCTAGAGCAGTCCTGCGGGGTCTGGAAGGGGAAGACTGGCTGAGGTGCGGGGGCCTGGTCTGGAATGATCCTGCGA TTTTGGAGTGAAGCCATGGAGCGGGAAGAGACAACCCCCCGCGGGGAATAGCCCGGCAAGTGGCCACGAGGCCAGGCTGAGGTCCAGA GAAGCAGGGGCATGAATCCCATAAATCCCAGGGGGCCTGGCCATGGGATGTGCTGGCTGCACCCGGCCCCCTGTGAGAGCCCCCGCAGGCT

SEQ ID NO	GENE NAME	SEQUENCE
		GGCCCCCTTCTGCAGTCAGTGGGGCTGGGGCAGCTTCTCTGGCATGGGGCGAGGCAGCCGCCTGCACAGTGGCCCCCCTGACTGTGCG
		CCCCCACCCTCTCCAGGACGACCACGGCTACATTTCCCGTGAGTTCCACCGCCTGCCGTCCAACGTGGACCAGTCGGCCCT
		CTCTTGCTCCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGGCCTGGATGCCACCCACGCCGAGCGAGCCAT
		CCCCGTGTCGCCGGGAGGAGAAGCCCACCTCGGCTCCCCTCGTCCTAAGCAGGCATTGCCTCGGCTGGCTCCCCCTGCAGCCCTGGCCCATC
		ATGGGGGGAGCACCCTGAGGGCGGGGTGTCTGTCTTCCTTTGCTTCCCTTTTTTCCTTTCCACCTTCTCACATGGAATGAGGGTTTGAGAG
		AGCAGCCAGGAGAGCTTAGGGTCTCAGGGTGTCCCAGACCCCGACACCGGCCAGTGGCGGAAGTGACCGCACCTCACACTCCTTTAGATA
		GCAGCCTGGCTCCCCTGGGGTGCAGGCGCCTCAACTCTGCTGAGGGTCCAGAAGGAGGGGGTGACCTCCGGCCAGGTGCCTCCTGACA
		CACCTGCAGCCTCCCTCCGCGGCGGCCCTGCCCACACCTCCTGGGGCGCGTGAGGCCCGTGGGGCCGGGGCTTCTGTGCACCTGGGC
		TCTCGCGGCCTCTTCTCTCAGACCGTCTTCCTCCAACCCCCTCTATGTAGTGCCGCTCTTGGGGACATGGGTCGCCCATGAGAGCGCAGCC
		CGCGGCAATCAATAAACAGCAGGTGATACAAGCAACCCGCCGTCTGCTGGTGCTGTCTCCATCAGGGGCGCGAGGGGCAGGAGGGCGGC
		GCCGGGAGGGAGGACACGCGGGGTCTCCTGCTCGCGTTGGACCCGGTGGCCTCGGAACGATGG
229	chr21:35450000-43546000	TTTTTGTGTTTTTAGTAGAGATGGGATTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGGCCTCATGCAATCCTCCTGCCTCAGTAGTA
		GTAGTTGGGATTACAGGTGTGAGCTGCCATGCCCAGCTGCAGGTGCGGAAGCTGGGGGCCTCAGAGACTGTGGACTCCTGGCCGGTGAG
		GAGCGGCATGGGCCGGGAGAGCTGACTCTTCAGCGGGACTGAGGTGGCTGGAGCGTGACCCTTTCCTGAGGGCAAACAGGGAGGGCCT
		TGGAGCCCGGCGCTCAGGACAGGCCCCTGCTGGCCCGGCAGCCTGAGCTTCCACACTTTTCCAGGGCGTCTCGAGTTCGCCCACAGAGC
		TGTTGTTTCAGGATAAAAATGCCCTTGTATTCCACGTTCCAGTTCAGAGGCCCGTCTGTTCCCAAGAGCGGAGGCGTCAGCCGCATGAGT
		CCCACCGGAAGCCGGGTTGCCGGGTCCCCGTCCCTGCCCTGCAGACGACGCATTCCGGAGCCCCCTTGGGAAGCTGCCTGGCTCTCCCA
		GGCCTGGCTGCCTTCGCACGAGGGCTCCGAGGCATGCTCATCCTACGTGACTGCCCGAGTGTGCACACGCCTGGCCGTGTGTGGGCGTG
		TGCCTGGGGCCCGAGCTCAGGAGCAAGGCCTGCGTGGACCTGTTGTCTGAAACAAGCCAGTAGACAGCTGCGTCAATGCAGGCAAGCTG
		AACAGGGCTGCTTTTTCAGCCTGACAACCCCAGGGGCTGAACAGGAGCTGGGGGAGGAGCAAGGGGCCGTTCCCCTGCCCCACAGCACA
		GCACACGACCCCCGCCTTGGAACCTGGGGCCCGGGGTGAATCGAGGGTCCTGGAGCAAGAGGGGCTGCTCCACAGGAGAGCCTGTCCCG
		CCACCCCTCAGCCACCAGATTCGGGGCTGCTGGACTTGTTCTCAAACCTGCACAGTGAGTGACAGCTGCTGAGACGGAGGTCTCAGGCAG
		TGCAGGTGAATCAGCAT
230	chr21:36060000-43606500	TCCTTATTTTTTAGTTCTCAAGCCCTGTAGGGTGTTTTCGGTCGCAGTTGTTTGGGCTGTGGTCCTGACCCTCCTGAGTTCCAGTGGCTCTG
		TTCAGGAGAGCTGCCTGGGGCCGGGACTTCTGAAACACACACTGAGCCACAGGCCGGCCCGGCGCTTGGGTTCACCGCCGCCTCTTTTG
		TGTGTGATGTCCTGGGATAGGCCCCGTGCACGTTCAGATGACACTGTACATATAAATAACTTGTAGCCGAGAACAGGATGGGGCGGGGAGGG
		AGGGGAGGGCAGAACGTACCACAGCAGCAGAAGTCACTGTGGATGCCTTCGTAAGTTGCATGGAAGGTTTTTAAACCTAGCCCTGCCGAG
		CAGCCCTCTCCTGGTCCGGGAGAACGATGGGGAGAGAGCTGGCGTTCAGCTTTCATCACTGGAGCCGTTCCTTCTTCCGGCCCCCCGAGG
		GCCTGTCCATGATCACACTTTGTCTTGTTTCGGGGGTGGCCCCTGTGAC
231	chr21:36430000-43644300	CAAGCCTGTGGTAGGGACCAGGTCAGAGTAAACAGGAAGACAGCTTTCGGCCAGGCGGTGCACCTCGGTGCCGGTGAGTGTGAGCGTGT
		GTGCGTGTGCACGTGTGCAGATGTGTGTGGACGCTCCCTTCTCCGCAGCAGCTCCTGACCCCCCTGCAGGTGACCCTCAGCCAGCCCCAG
		GGCTGCCCCCCACTCTCCCCTGTGGACACCTACCTCATTTGGGGTGAAGTGGGGGGGACTGGGGTGTGAGGGGTGCTTTGGGGGGCACACT
		TCGACCCCCTCTCTCTGCAGGCCAAGTCCTGAGGCTCAGTTTCCTCCTCTGTGCCCCCGGCGACGTGGTGCAGGCCTCGCGAGTGACGTGAG
		GGTTCATGACCCAGGTGTGGGCAGCCAGCCCTTCACGGGAGGCCACCCACCTGGCCACAGTGCCTGGGAATTTAGGTCGGGCACTGCCG
		ATATGTCGCCTTCCACAAGGCGGGCCCGGGCCTCTGCTGACCGTGCACCGGTCCTGGGGCTGGGTAATTCTGCAGCAGCGCAGCCCC
		ATGCCGGGGAATTTGCGGGCAGAGGAGACAGTGAGGCCCGCGTTCTGTGCGGGAACTCCCGAGCTCACAGAGCCCAAGACCACACGGCT
		GCATCTGCTTGGCTGACTGGGCCAGGCCCACGCGTAGTAACCCGGACGGTCTCTCTCTCACAGTCCCCTTGCGTCTGGCCAGGGAGCTGCC
		AGGCTGCACCCCGCGGTGGGGAGAGGGGCAGTGTCGCCCATCCCCGGAAGGCTGAGCCTGGTGCAGCCAGGGAGTGAGGGG

SEQ ID NO	GENE NAME	SEQUENCE
		GCGGGAAGCCGGGGTGCTGCCCTGAGGGTGCCCCGACACGCTCTCCTGGGGCCCTGAGCGGCTGCCACGTGCGTCCAGGGTTCTGGCC ACAGGGTGGGCAGGGGCCCTGTGCTCCTCACTGGAGGCCCCTGAGGCTCTGGAACTGAGACCATCCACCCGCCGGCCCCCTCTCGCCG GCTCCGGCACCCCTGCCTACTGTGACTTCCTGCCCCGGACTCGCTCTGCCAGCTTGGGGCAAACCACTTCCCTCTGGGGTTTTCACTTCCC TCTTTCCCAAGTGGGGAAAGACCACCTGTCCCCGACCCAGAAAGGGCCCCCTGCCCGAGGGCAGCAGTGCCAGGCTGGCATGTGAGG CTTGGGGCAGGCCCGGCCCCCAGAGGCACAGGGCGATGCTCTGTGGGACGCTGTGTCGTTTCTAAGTACAAGGTCAGGAGAGGAGCCCC CTGACCCCGGAGGGGAGGAGGCAGGGCAAGGAAACCGCCACCATCTCAGCCCA
232	C21orf125	GCCCACTGTGGGTGTGCCCGTGTGTGGCTGTGAGGCGTGAGTGCAGGCGTGAAGTGTCTGGGAGTGGGAGCGGGCATGAGTGTGTG CCACGGGCCTGCTGTTGGGTCCTTGGAGGCCACGGTTGCCCCTGAAGGGACTGCAAGCTCTTTTTTGATTTGTAGTTATTTGAGAAGTCTA TACAGGAAGAAAATTAAACCG
233	C21orf125	AGCGCCCAGCGCAGGGCCGGGACCCAGAGTGGACTCTACCGTGGGGCTGCCTCAAAGAAATCTCAGCAAACACAGGAAGCCAGCCCACC CGTGCAGCCATGGGGCCAGGAAGCCCGCCCTTTACCAAGTCATTTGGGCATTTTTTCTCTGTGCTAACAGCCCAGATGGAGCCATAGCCTC AACCTCTGTGTTCTGATAACACCAAGCTGGGACGCCGGAGCCATGCAGGGGACAGTGCCCCGGCCTGAGGCTGCAGCCTGGGTCTGGATG CCTTTCTAATTCAGGGCCTCCTCATGGCCTGGTTCCATAAATGGTCAAATGCAGCCTGACAGCGCAGCCTCCTATCAGCGCTGGGCTCCGT ACCGCCACACAGCCCACATACCCCGTTCCCCAGGAGACGCCCGCAGGTGGGCAGCGTCACTCCCACCCGCCGAGCACACGCTGTCCCCG TCTCGTGTCCCGAGGAGCCAGCTGCTTCCTCCCAGCCTGAAAGCTGCACCTCGGCTGCACTCGGCTCCCCGAACCCGCCCTC CGCTGCCCTGCAATTCGCCAAGGGAGCTACCCTTCCCATATAAAAATTTCACCTCCATTTCCTTGTAGAGAAGAAACATTTCTGACAGCAAG GAAGATTCTAATTTGAAAAGCAAGTGATTCATCTCCCGGTGCCAAACAGCAGACGCAGGCGTTACCAGTCTGGGTGGGGCGCCCGAGCTG GGGACCTGGGGTCCTCTGGGAGGGGCAAGAAGGCAGCGATGCTGGCCCCCGCCTCCATCTGCCCATCCCATCTGCTTCCACACACCGCC CTGCCGTAGCTGCTTGCAGCCCTTCTCTGTCAGTTTCTCCATCTTTTGGTTTGGTGATAAATGAGAGTTCCCATCGGGTGTGCCACCCTCTG TGTGACGGGGGAGCAGAGAAGACCCTGCGTCCAAGTCCTCCTGGGGGAAGAGCGAAGATGCTGGGACCAGCCCCAGCTGTCAGGGGGTCT CCAATCCCAG
234	HSF2BP	GGAACGGAGAGAGCCGCCAGGCCCAAACCTCCCAGAATTTGCGCAGTATTCTCGGCCTAGAGAGCGAGGAGTGGCCTTGGCGAGGTCCCTC TTTGGCTCTTCTGGCTTAGCCGGGGTTTTAAACTTGTTATCTGCAAAGCAGAAGGAAAGTCAGCCCCTGATGTAAGTGTCAAGTAAAATAAA TCGGATGGGTCCTTTCCTGTTTGGCGAGGAATGCTACACTAAGGGGGGACTGCGTTCAAATGGGCAGTCTTTGCTGGAAACCTCGCCTCCGC GCGCCTTCCCTCGCTCGGATTCAGGCGCTTTTACGTTAAGGGTTGAATTTTTTGTGTCAACAGGCACCTCGGGAGGTCGCCTAGACAACTGA GCGGAGCAACTGAGATAACCCCCGCTACGTGTGGAGTGACCTAGTCCATTAACTTGCCCCCAGCACGCCCGCTGAGTCCGCAAAATATAGG ATGGCCTCGGGTTTTAGATGAACCCAAAGCTAAGATTTCTTCCCTCTCTGGAATTAGCAAGCAGCCCCGCCCTGCCCAACTCCCCTGGAAGC GCGCGTGCTCGCCAGGCCTCGGGACGCCTGCGCGGGGCGCCCTTGCACTGGCACCAGGGCTCCGGGGGTAGGGGCGCACCGATCTGCCCA AGCCTCTGCAGGCACTGGAGGAAGGCGAGCCCTCCACCCGCTCAACAGGCCCCCAGTGCCGGCCTTTCCTTCCAGTCTCAACTCCACCCG GGGGCCCGGGGGGCTCCACAGTTAAAAACTCCACGCCACGGAGATCGCAGGTAAGCTGCTGGCTCAACGAGGTGTGCTAAATGGGATTAAA GATCCTGGACCGTGGCCAGGCGCGGCGGCTCAAGCCTGTAATCCCAGCGATCAGGGAGGCCGCCGCCGGGAGGATTGCTTGAGCCCAGG AGTTTGAGACCAGCTTGGGCAACATAGCGAGACACCGTCTCTACAAAAAAATAACAAATAGTGGGGCGTGATGGCGCGCCTGTAGTCTC AGCTACTTGGGCCGGTCGAGATGGGAGGATCGATCGAGTCTGGGAGGTCGAGGCTGCAGTGAGCCAGGATCACCGCCAAGATCGCGCCAC TGCATTCCAGCCTGGGCGACAGAGGGAGACCCTGTCTCAAAAAACAAACAAAAAATCCTAGACCGTTTACAAACAGCCTTCCGTCTCTTCCTG GTCAAGTCCTAACCCTGGCTAACCTCGCCGTCTACAGCCTGAATTTTGGCAACCGAAAGGCAGCGCCCGGCGCCACGTGCACACGGGCTGG GCCGCTCCGCCAGCTGCCAGGGCCACTGCCGCGCTCACT

SEQ ID NO	GENE NAME	SEQUENCE
235	AGPAT 3	CGCACACACAGCACAGACGCCTGCATCTTCCCATGCGTGGTTTCTGCTCTTGCCTCTCTGGGTTTTTGTTTCACTTCGGTCGAGTTTTTGGT GGTGTTGAGCGGATAGCCGGGGAAGTTGGAGTCTTGTTTGTGGCCGCCTCGTGCTCGTGTCTGTATCTAAGATCCTCAGGCTGCTCCTTTT TGGGTAAGGTCTGTTGCTTCTCTAGGAACAGTGACGGTGGCAGAGCCCGTGGCCCCTCTCTCCTGTCCCAGAGCCAAGCTGTTTCCTCTCC CCACTCCCGGGCACCCTGCGGGCAAG
236	chr21:44465000-44447500	CACAGCCCAGCTTCAAGCCTGGCCGACCAGGGGTTTGGCATGAAGACCCCGGCAGGGCTGGGGCTGTGCTGGAATCCACCCGGAAGTTT CCTGCCCCCTTGGGCTGCCCACCAGGTCCCCTTTCTGCTCTGATCAAGCTGGACAAAACGTCGTGGGGCCACAGCACAGGGGGCCAACGC AAGCTGGGATCGTCAGACGTTAGGAAATCCCAAGGAAGAAGAGGGGACACATTCGGGAGACGTCGGCACACGCTCGAAGCAGCGG ACAGGCACCTCTCTGTGGACAAGGCAGACTGGGCGGCCGAGATTCCGCATAGATGCCTGCTTCCTCCACGACCTCCACGTGTGGCTGGCC CAGTCCGGGTCCCCCTCACCTCCTCTGTCTGTCTTGGTGGCCTCACGCCGTGGGCTGTGATGCCGGCTACGCTGCTTGGGTGGCCAAGG GTCTGAGCTGCAAGACGCCCAGCCTGGGTCTCTCCCGAGCTCTCCCACGGTCCTGTCTGCTCCTCCTCGAGCTCCCGGTTGACTCTCACG ACTGCACCAGCCTCTCCCCCAGGAAGGCGTGGAAACAACCTCCTTCTCCCAGGCCCGCTCTGCCTCCTGCGTTTCAAGGCAAATCCGTTC CTCCAGGAGATGATGCAACCACATCCTGTTGGAGCCCAGAGAAGTGCGGATGCAGCCCGGGGCTCTTTCTTTCCTAGAACCCTGCCTGGG AGTGGCTTCCCTGAACTAAGGACAGAGACTTTGTCTTCGTTGCCTCTCGGCCTGTGGGCACTGAGCATACAGTAGGTGCTCAGTAAATGCT TGCAGGCCGATGCCCAGAGCCATTAGCCCTCATCATGGTGAGCTCGGCAGCCGGTGTTGGGGCTGGGCCTAGGTGTGCGTGGGG GCGGTGCTGGTCTGCTTTGCTGGGAGCCATGGACACCGGAGGAACAGGGCCCCATCAGTGCGGTCAGAGTGCAAACTCGGAGCGTCCTT CTCTGGAAAACGAAT
237	TRPM2	GGGAGGGGGCGTGGCCAGCAGGCAGCTGGGTGGGGCTGAGCCAGGGCGATCCGACCCCGAACCGGAGCTTTTAGCACTTTGAGTCCCT GTACTCAGAGGTCTCCTGCAGCCCGGGAATCCCCACTGTGCTGTGGTCCCTGGCAGCCAGCACCCACCCCCAGCTTCTCCGTCAAGGTTGAG GACGGAGCACTCCTGCCTCTGATTAACTGGACGCAGGAGAAGCAGTTGCTTTAATCCGGAGCCTTGAGTTGGGACAGATAATGAGTCATTC AACCAGATTTTCCAAGGACACACTAACTTTGGTATGATGCGTGTGTGCCCCTGAATCCACGTGGTCAGGAAAGCCCAGGGAACACTGGCCT GTGACTCACTGAGCAGGTTCCCTTGTTACCCCGAGGGGTGATTTACTCCTCTGACAGTGACACGGACACACTGTGCGTCCATTCCCCGGGCG GGCAGAGGACACTCCCAGATGCCCACGAGGGGCCCAGCAAGCACTGGCCA
238	C21orf29	CTGCAGGACCTGCTCGTTCACAGATGTTCTCCTAGAAGCAGAAGCTGTTTCTTGTTGCAAACAAATTTGCTGTGTCCTGTCTTAGGAGTCTC ACCTGAATTTACCAAGGATGCATCTGTGCTTGGGGATGGCTCGGTTTGAGGGGTCTGAGGAGCGGCTCCCCCTGGATCCTTTCCTCCCCAG GAGCCCACCTGCCGAGCTGTCAGCGTCAGCCCCACATCTCAAGATGAGGAAATGGAGGTCGAAGCCATGCACACGCAGGCGTCCTGCTG ACATGCAGGCCAGGCGGGTGCCTCTGTATTCAGCAGCCTCAGGGCTGTGGCCAGTTCAGGCAGCAGAGGGGCCTCATCCCGGTGCTTCC CTGCAGGCAGTTGTGGGGCCGGCCTGCAGCAGGGGCTCAGACAGGGCCTTGGGAGAGGGAGGGATCACAGAGGTGTCCAGTGACAGGC AGGGCGGGCAGAGCCCATGGGGCCTTGGCTCCTCACTCCTTCGGTCAGTCAGGGTGACATCTGGAGCCACCTCCATTAATGGTGGGTTA TGATTTGGTTCCCATGCAGCCCGTGCCAGCTCGCTGGGAGGAGGACGAGGACGCCCTGTGATC
239	C21orf29	AAGAGGAAATTCCCACCTAATAAATTTTGGTCAGACCGGTTGATCTCAAAACCCTGTCTCCTGATAAGATGTTATCAATGACAATGGTGCCC GAAACTTCATTAGCAATTTTAATTTCGCCTTGGAGCTGTGGTCCTGTGATCTCGCCCTGCCTCCACTGGCCTTGTGATATTCTATTACCCTGT TAAGTACTTGCTGTCTGTCACCCACACCTATTCGCACACTCCTTCCCCTTTTGAAACTCCCTAATAAAAACTTGCTGGTTTTTGCGGCTTGTG GGGCATCACAGATCCTACCAACGTGTGATGTCTCCCCCCGGACGCCCAGCTTTAAAATTTCTCTCTTTTGTACTCTGTCCCTTTATTTCTCAAG CCAGTCGATGCTTAGGAAAATAGAAAAGAACCTACGTGATTATCGGGGGCAGGTCCCCCCGATAACCCCCCAGCTGCAGATCGAGGCCTAGTG CGAGCACAGGTCCCCCCCAGACCCTTCCCAGTGCCCACCAACCGGCGGCCTAGGCCAGGTAGAACTGGCAGCGCCTCCCCCTGCTGCAACA CCAGGCTCTGGTAGAAACTTCAGAAAACATGCACCGGCAAAACCAAGGAAGGGTGGCTGCGTCCCGGGTTCTTCCGCGCAGCTGTGTGTA CACGCATGCACACACCCACACGCACACCCCCATGCACACGCACCCACTTGCACGCCCATGCACGCACACACGCGC

SEQ ID NO	GENE NAME	SEQUENCE
		GTGCACCCATGCGCACGCACCCATGCACACACACGCGCGCACACACCCACGTGCGCACCCACATGTACACACCCACGTGCACACACCCAC
		GCGTACACACCCACGCGCACACACCGCTGTCCCCAGCCGTGCAGAACGATCCTCCCCTGAGTCCCCGGCTCCGACCCACACGCAGCACTC
		GCTAAACGCTTCCCACGCAGTCGTTTTGCTGGGTTGCGCTTCACCCACTTCTCAGAGGGGGGCGGCCGAGGCAGAGGTGTCGGGGATCGA
		GCAGCTCCGGGCCTCAGGGGTCGCCCCGCCACCGTTTTCCTTTCCCAGATGCTGGGACGGGGGCAGGGAGGGGCTCCCCCAGGCTGAAC
		CCGACTAGGTCACCCTAGAAGCGAGGCGAGCTTCTCTTCTGTTTTTCTTCGGCGCCCCTGAGCCCCTGACAGTGCCCAAGCTGCCCATGG
		GATTGGATTCGCCAGAGCCTCCTACGCAGACCCCACCCAGGGCCAAAGCCAACCCCAAGCCCCCACCACCTTGGTGGTGTGGGATGAAAAG
		TGAGCCATCGAGAGATGGGGTCCCCCCACCCCCAACCCCTCCAAGGACAAAGGCGGGCTGGGAAGCACCCGCTTTCACGTCCGCCCCCTG
		CCCGGCTTTCCTAGCGGAATTGGCGCCCGGCATCAGTTGGGGGTTGTGGGATCAGTGAGGAATCCCGTGGGGTCGCCTCCATTTATCAGTT
		GTGTGGGGTTGGGCGAGCACCCCTAGCCCCAGCCCAGGCGATCAGGGCGCGAAGCCCCACTGGACGCGGAATTTGGGATTAGGACGGGGG
		TGACAGCCAGGAGGACCGCCACCTGCCCTCCCCACTCCTGCCGCTCCACCCCTGCCCCCACCGCAACACCAAGGTCTCCACCAGGAAGAT
		GGGGGTGGGGAAAGGACGCGGGGGTGGGGGGGGAGAAGAGGACACAGGGTCGGAAGGGGTGAGGGGGTAGTGGCAGAGGCGG
		AGGCCGAGGCCACCGCAGCTGCGGGGGCGCAGAGGAGGGGGCGTTCAGATGGGAACCTAGTCCAGACCCGTCGGGGCCCT
		CGTGTGCGGCTCGTTATCCTGGAACCAGAGAGGCTGGAGACCCTTGGCTTGTCTGGAGCGGAACCGTAGTGTCCAATAGAGTGTGTGGGG
		CTCAGCCCTAAAGCTAAACATTCTTTATTTCCTGATGACCATGGGGGGCGGAGCGGGGGAAAAGCCCTGGCCTTATAGTTTAGAATTTTATAA
		AAGGAAAGGCGTGGCCACTGACAATTTGCGCTTCAGGAGTCCCAGAGTGACCGCCTGGCTCGGAGCAGGGAATGAGGGGGTCCTTAACT
		CTGAGATTTGTTTTCTGAGAGACAAAGGTGATGGGTGAGGCGGCTAAGCCTCTGATTCTCTATAGGTGGCGGTCATTCATTTCAGAACATGA
		ATGGATTCAGTAAATAAACATGATAGAAAAATGCCACAAGCCCTAGGCCCATTGGAGTGGACTGGACAGTCTGTTCCCAGTGTGTCCCTCA
		GCCTCGGTCCCCCACCCTTCCCGGAGCCCTGGGGGTCACACACATCCCTCCTGGCTGCCTAGCCTGTGCCCCCCGATTCCCCCCCTCCC
		CGCCCCGCGCGTGCACACACACACACACACACACACACCACACGAGGCGACAGAGATATGAGAGAGAGCG
		AGCGAGAGAGGACGGGAGAGAGGGAGTGCAAGTGTGCGCTGGGGGTAACCCCGTGCATGCATTGGGGGTAACAGGCTGGAGCT
		CAGATCCCTCCCCCAGCCCCCAGCAGGGGGGACTGCAGGCTCCTGGTCTGAGTGGGGAGCTGGGCCCCCTGGACAGAGGACTGGGCTG
		CGGGGTCAGGAATGGGCACACTTCCTAACTGCAGGACACTCTAAGGGCTTTGGTCATGCACACGCAGCCAAGAGAAGGTGTCGCTGGCAC
		ACAGCCTTCCAGGAGCGGACTTGGAGACCTCGCCAAGGACCAGGACTCCCCAGCACTCACACTCCCTTAGGCGCTGAAGTCCAGAGGACA
		GAGGTTGAGGGCAGAGCTCCTGGGAGCACCAGTGGAAGTAGGAGGGCTGGGCTGGAAAACCTCCCCCAACCTCCTATTGCAAAGAGGCT
		CCAGCCAGCAGCCTCCACACCCCAGTGATCTTTTAAGATGCAAATCTGCGCCATCATTTATTTCCTCAGTGCCTTCTCCAGCTCCTGGGATG
		CACACTGCCCGTCCCCAGGCCCAGAGACCTGACCACCCTCATTCCTCCCTCAGCCCACCCTGGGGTCTCTCCACCAGCTGACAGCCTTCC
		TGCAGTCCCCTCCCCGAATGCTGCTCCCTGAGGCCCTCCTGGACACCTGCAGGGCAGGCACAGCCCGCGGGACCTCACAGCACTTGCTC
		CGGGCAGAGCTGCAGTTTGGCCAAGTTGCCAGCTCCGTGTGGGCAGGGGCCCTGGCCTGTGGCTGCCACATCCCGGGTGGGGGCACGG
		CCTTTCCTGGCGTGGATGCTGAGCAAACGTAGGGGGAAGGGGAGTGAATGAGGAGAGCCAGGTAGCTCAGGGGCTGAGGCCTCACTGAG
		CAGGGTCCCGCGTGACCGGTCCCCACCGCTGACGGTTCCTGGGGTAACACTCAGGACAGGGAGAGGCAATGGAAAGAGACGTGGCCGC
		CCTCGCATCCTGCAGCTCCCGCACTCCCAGCCTCCCACCCAGCCCCCCAGAGCCCACCAGTGACCCCGCCCACTGGGTCCT
		CAGATGGCTCCCACGGGATCTCCTGCCTTGATCTCCTGTCCACATGGAGGTGAAGTGGGTTGCTCTGAATGAGGGGTGCCGAGCCTAGGG
		CGCAGCCCACTCTCCTGGGTCCGCAGCATCACGCAGCCCGGACCACAGGCTCCTTACAAGAATCGGAAGGGTCCCTGCAATCGCCCTTCG
		CACTGAGGCTTCCTACTGTGTGGTGTAAAAACACAGGCTTGTCCTCCCTTGCTGCCCACGGGGCTGGAGCCGCCTGAAAATCCCAGCCCA
		CAACTTCCCCAAAGCCTGGCAGTCACTTGAATAGCCAAATGAGTCCTAGAAAGCGAGAGACGAGAGGGGAATGAGCGCCGAAAATCAAAG
		CAGGTTCCCCTCCTCCTGACAACTCCAGAGAAGGCGCATGGGCCCCGTGGCAGACCCGAACCCCCAGCCTCGCGACCGCCTGTGACCTGCGG
		GTCAACCACCCGCCGCGGCTCCACGCCGTGGGCACAGACTCAGGGAGCAGGATGAGAAAGCTGAGACGGCGCAGCCACGGCCCGGTGC
		CTTCACGCGCACAGCGACACAGCCCCAGCCAGCGGGGCCCACGCTAAGGCGGAATCCCACAGAAGCCTACAGAGCGAGCGCGCGCCTG
		TGCTTCCCAAAACGGAATGGAACCAAGGTGACTTCTACAGAACGATCTGAAGCCCTGGCTGGCCCTTATGCTAGTCTCTTGGGAGCGTTCC
		AAATGCAGCTCAATATTACTTACTTGACTTTTATCTTTCCTCCCTGGTTCGTGGTATTTATAACTGGGTCATCTTTTAACTATTTGCAACGTAG

SEQ ID NO	GENE NAME	SEQUENCE
		CTTCAGGGGAGAGGGGGAGGGCTTTATAAATAACCTGTATTATTATTATGCAGGTTGATTCTGTTCCCTGAGCTAAAGGGAACATGAAAATA CATGTCTGTGACTCATGCCCCCCCCCACCCCACTCCAGGGTGTGCTGAGGAGTCTCTCAGCTGCCCCGGGGTCCTCGAGCAGGGGAGGGA GAAAGGCTGGCGCTGCGCCTCCATCGCGTGAAGCCAGGGGATTTTGCTCTGCGACAAGCTGACTTGGCTCTCGTATTGTTTGCAGAATCA CCCAGTTCCAAGGCAGTCCCTGCGGGCAGGTGCAGCTGTGCGGAGCTTCAGTCCTGTCCCCAACACCCAGGCAGTAATGGTTCCAGCAC GGAAGGTCTACCTACCTCCCACTGCACAGCCCGAGGGCTGTCCTGGAGGCACAGCCATCCGTCCCTGGGTGGGCAGGCACGTTTATGAC CCCCACCCCCCACCCCCCCACGCGAGTCAGCACGTTCCATACTCGGGTGATCGTGCTCATCCCCCTGGTCATGTCATCGGGATCTGA GTGCCATCCGAGCAGAGAGCTGTGGCCCGGTGCCGGGGGTGGACTTCATCTATTCCAGGGAACCAAGGATGCATGATTTGCAAACAAAAC CAGAAGCGCAAGCCATCTCCTCGCCTCCCCTGATAGCCGTGCTGCGGAGCCTGAGTGCTGGAG
240	ITGB2	CAGGAACCACGGGACCTGCTGCCTAGCGGCCCTGTTCCACCCTTGGCCGCTCGCAAAATGTTTAGGCTTCATAAGGTTTGCCCAGGGTCA CAAATTTAACTCACAGCAAACAATGAAATCAGCGCATGATTTTCGAGCCCTCGTGGTCACCCTCCCTTCCTCCTGCCCTTTCCTGCATGGGC AGCAGCAGGGTGAGGAGCTGCTCTCCCCAGGCCCAGGCTGGAGTCCCTCAGACGACCTGCCGGCCAGGGTACCCCCCCTGCCCCCACACA GCGCCTGACAGAGCCCCCCCACACTGGGGGAACGTGGGGACCCAAGCAGGGGCAGCGGCCTCACCGGGCAGGCGGCGACCTGCATCATG GCGTCCAGCCCACCCTCGGGTGCATCCAGGTTTCCGGAAATCAGCTGCTTCCCGACCTCGGTCTGAAACTGGTTGGAGTTGTTGGTCAGC TTCAGCACGTGCCTGAAGGCAAACGGGGGGCTGGCACTCTTTCTCCTTGTTGGGGCATGGGTTTCGCAGCTTATCAGGGTGCGTGTTCACG AACGGCAGCACGGTCTTGTCCACGAAGGACCCGAAGCCTGCAGGGCACATGGAGGGGCTGG
241	ITGB2	TGCGTTTAGTGTAAAAATATCAGGTGTGGCTGCACGGAGTGAAAAATCACAGGCTCCACGGAGCCGGGAGGCCTGCTGCCCTGCCCTCTT GCTTTGATGAGGAAATGGCGACCGCAGAAGGAAATGTAGCAGCACCGGCAACCGGCATCCGTGGGGCCACCGCCGGGCTGCTTCCCAGGG CCCTCCAGCCAAGCAGCCACAGGAAAGAGTAGATGTTGATCCCAAGCTAGGACTGAGGAGTCCGTCCCTAAGAGCCGAGGGAGTCAGGTG GGCGAAACTGGCCGCATGTCTGGGTACAACTGCTCAGGGTTTCTCATCTGCTGAATCACCAAGCTAGGTTCTGAAGCCAGGCCGTGAGTGA GCAGGACTGGAGCAGGATTCTGGGAACAATCTTTTCCCTCC
242	POFUT 2	GCTGGGGAACTGAAGGAAGGGCTGTGGAGCCTGAAGCCTGGGCCTGTGCTGCGGCCGCACCGCTGGGTGATGCAGGAGCCACT CCACCTCCCTGGCACCCCAGCCTCATCCGGCAACCTGGGAGCGTGGGCCTCCTGCCCCTCCAGGGAGGCCCTGGCCGTGTCCTCATGGG GCCCCTCCAGGTCCTTGTGGCTCCAGGTCGGGACAGTGGCTGTGAGATCTGACCCTCCCGTTCCCCCTCCACCAAGTAGGAGAAACCCCG GAGCATGAGCCCTCGTCCTTCACCGTCCCGGGGACAGGGGGACCCCCAGATGCTGCACGGCTGACAGGCCAACGTGGCAGAAGCTCCAG CTTCACAGGAAGCCAGTGACCATGAGAGTCTGTAGCTGTAACGAAGCCACAGAGCTGTGGCTTTCTTTCCCCTTCAGCTCTAGGAAAGGTT ATCTGCCCTGCACAGATCTCCGGAGGCCTGGGCTCTGAGAGCATCAGACTGATTATCGTAAGAAAATAATCTCTGCAGACACATTCC TTGCTAGAAGCAGGGGACAAAGCCCAGCTTCAAAGACAATTCCACACACGCCCTCCCTGCCCTGCACAGCTGCCTGCGGGTGGGAGCAG AGCCCTTGCAGCCGGGCTCAGGGGCCTGGGCAGGGACAGCGTGTGGCAGGGGCACAGCTGAGACAGGAGCCTCAAAGCGACACCAACC CGACGTGAAGCTACAGTTGAGGAGACACAGCTGCCCCCATTCCCGGGCCTCATCTCCACAGTGAGACGCTGGACTCTCTCCCTGACCCAC CGTCTCTTAGAACCTCCCCCTCCATCCGGAGCAGTTCGGCAGCCCCAGGGCAGCCCAGGGGAACCCTGCCGAGTGCCTCTGGGCCGCCACA GACCGCAGAGCCCGCGGGAGCCTTGCTCACACAGCCTCAGGTCCACTGTGGTCTTGGGGGAAAGCCCCTGTCCTGGGACACAGGGGAGCCCG GGGGTCCTGGCCCTGGACCACCATCTGGGGACCACGTTGTCACGCCTGCAAAGCTCCCCTGCCCCCACCCCCCATGTGCCGGCTGGTGTTGA CACCTTTGTAGAGTGGGAACCTGCCTCCGACCCCCAGCCTGCAGCCACAGGGCAGGTTATAGACCAGGTGAGAGGGCGCCGCGCCCCAGAA CCAAGGAGCACAAGTCCGCAGTGCCCATGAGATCCTCATGCTGGCCGGCGCAGGAGCCATCCTCGGCCTCTCGTGCAGGTCCTCGTGGGAAA CCGCGGGGGCACGTGGGGCGGCTGCAGGGTCCGCAAAGCCGGCTGTTTGCGAAGGGCGCAGCTCCACCTGGAACAGCCGGAGGCCGCC CACGCGCTTCCCGCGGGATCAGAGCAGCCTCCACGGCTGTTGTCTCAGGCACCACGGGATGCCTTTCTTCGTTTCAATAGCTGTGGGAAA GCCTCAATCGGTCCTGAAAGAACCCAGATGTGCAGCAATGACAAGGCCTTCTCTGAGACTCTAGAACCTTCTGCCATCTCAGACAGGAGGG AGCCGTGAGGCAGGCGGGAGATTTGCAGTCAGCAAAGGACGGGCAGGTGGGGCAGCTGCACACCCAGGGCCCTCTCCACGGTCTTCCC

SEQ ID NO	GENE NAME	SEQUENCE
		GGGCCCACCCCTCCCGCGGTCCTGGGTCATCCACCTGCTGGCCTCACTCTGCCCACGCGGCCAGGTCCCACCGGCCCCTGAGCTCAACA
		GACCAAAGCTGGCCCGACCCCACCCCCAAGAAGAATGAAACAATTTTTTTTTTACCTCTTGCAGAAAAGTAAAAGATCATTTATTCATTCTGTT
		TCTAGATAGCAAAACTAAGTGTCAAAAGCACCTTCTGCACACACAGTCTGGCCGGTGGTCCTGTTCCCGCAAGGTTGAGCTGTG
		TTCCAGAGACATGGGTCCTCCGGGTGATGAGGAGCCGCTGGAGGGCCCTGAGCTGCACGTGCTAATGATTAACGCCCCGTCCGTGCTGG
		CCGGTTTCTCAAATGCCTCCTGACGATTGCGC
243	chr21:45571500-45573700	GGCCTGAGGAGTCAAACGGTGCAAACCCTGCCCCACTCTGTTTGGGAAGCACCTGCTGTGTGGCAGGCGCTGCGCTTGGTGCTGGGGAT
		AGACCATGGGGAAGAAACACACAGAACCTGCCCTGCTCTCAAGGAACAGGCCCTGGGGGCGGCCAGGGGCAGAGACCCAAGGCAGACAC
		CCACACAGTGGCGTAATGACAGTGCTTATGGTGGGGACCTGGCTGCACAGCAGGTCAGCAAGGGGATGTTCAGGTGACACTGGGGGCAC
		GGAGACCCAGGGGGAGAGTGGATTGACAGAGGGGACGCTGGGCAAATGTCCCGAGGCTGAGGTGGAGTTGCGGGAAGGAGGAGGCTGCC
		GGGCAGAGGCGCAGAGAGCTTTGCAGGTGTTGGCAGAGACCAGCAGGCCCTGCGAGGCCTGGGGTGTGTCCTCAGCTGGGAGGGCCAT
		AGAAGGATCTGGGCTTGCAGATGCTGGTGCAGACTGGAGGCCTGGGGTGTGAGAGTCCAGGCGGGGCTCCTGCCAACACCCAGGGGAGT
		GGGCCTGGGCCAGGTGGACCGGGAGCTGGCACGGTGGTCAGGTGCTTGGAGGCTGCGTGCCACGCTGGGGACCTGGAGGTGTGTGAG
		GAGGTGTCTGTTGCTCCTGGGGCTGCCGCCTGCAGGGCTGGGTGTGCAGCAGTGCGGGGGCAATGAAGTGGGCGGGTTCTGGGATGGTG
		GACGTTCCCTTTGTTGGGAACGTGTTGGTGCCAAGCTGCCATTTGAGTTTGGCTCTGAGGGGTCTGGGCAGGGGACACACAGGGAATCAC
		ACAGGATGGAGTGAGTTCCCAGGGACCCAGGGTGGCTTGGCCTGAGAACAGCTCCCACTCCCAGATGTGTGGGAAGCCCTCGGCACCAA
		GCCTCAGCCTCTCCATCTGTGAAATGGAGACAACGTCACTGGACTTGCAGGCTGTCCATGAGGGTGATGCGATCAGAAAGGGTGGAGTTC
		CTGAACGCCCCGGGGTCGGGGTCTCACAGCAGGAGCTTAGCTGGTGTCGGCATCTCCTGGACCCGTCCTCAGCTCCGAGCGCCCAGTCC
		TGCCACCTGTGTCCAAGTCTGCACTGTGCCCACGAGGCCCTCAAGGCCGCAGACAGCCCCACACTTCTCGGACGCCGCCCCAGCACGGT
		CCTTGTGTGAGGTGGACACTCCCTTCTGGACGCCGCCCCAGCACGGTCCTTGTGTGAGGTGGACACTCCTTCTGGACGCCGCCCCAGTACG
		GTCCTTGTGTGAGGTGGACACTCCTTCTAGGGAAGGAGTAGTAACTCTTGGGTGGTCGGGTAGTTGCCATGGAAAGGGGCAGTAATGCCC
		AGGTATTGCCGTGGCAACCGTAAACTGACATGGCGCACTGGAGGGCGTGCCTCATGGAAAGCTACCTGTGCCCCTGCCCTGTTAGCTA
		GGCCTCAATGTGGTCCAGTATCTGAGCACCGCCTCCTGCCTCAGATGTTCCCGTCTGTCACCCCATTACCAGGGCGGCACTTCGGGTCCTT
		TCCAGCCATCATTGTCCTGGCATTGCCACAGTGGACACTGCCACACAGGCTTGTGTGCTTGCGCGTACCCAGGTCCTCACCTCTCTGGGAT
		AAACCAGGCACGTGGCGGCGCCCCATTTTCCACCCGCCAGCGGTGGAGGAGTTGCCCAGCCTTGCAGGAAAACAGCTCTCATGCCAGC
		AGCGGAGCATCCTATTCAAGTTTTCTCAGGGCTGCCAGCACAAATGCTGCATGCCGGGGCGGCTTCCTCAGCAGACCGTTGTTTCTCTGCGT
		CCTGGAGGCTGGACGTCCCAGGTCCCCGGTGTGGCAGGCCCGGTTCCTCCCGCAGCCTCTCCTTGGCTTGTGGGCGGCGTCTCCTCCCTG
		GGTCCTCGCAGGGCCACACCCCTCCGGTGTGTCTGTCCTCCCCTCCCCTTATAAGGACCCCAGGCAGACTGGATCAGGGCCTGCCCTAAGGA
		CTGAATTTTACCTTAATCACCTCTTTAAAAGCTGTCTCTCCAAATACAGTCACCTTCTGGGGTCCTGGCTGTTAGGGCTTTGATGCATGGATTTG
		GGGGACACCGCTCAGCCCCCTAACAGCCCCCATCCTCTGCCTGCCTTTACCATGGGGCTGAGCCCAGCCCTGCAGGAGTCCCCTGGTTTGA
		TGTCTGCTGTGGCCACGGCGACCCTCAGGCTGCTCCAGCCGCACTTGTGCTT
244	chr21:45609000-45610600	GGGGAGTCTCCAGGGGCTGGGGCTGGAGCCGCATCAGAGAGGAAAGGGGTGTTTGAAAAGGGGCAGGGCCTGGGACCCAGGAAACTG
		TTCTTCCAGAGACACCCGTGAAGCTGAGCTTTGCCTCTCAGGGAAGCTGTGACCCCACGGGTGCTGCCCAGAGAGATCGGGCCAGGTGGA
		GCCAAGATGGACTGGAATTCCCCGACGGGGACAAGGGGCCGGACGAGGCTGACTTGCCCTGTCTGATGAATGGTCAGGTTTGCTTTTTCT
		CCTGAAAACACGAGGCAGTGATCCCGGCCAGCTAATTCCAGCAGACTGGAGACGGGATGGTGGAGAATGAGGCTGTGGGCGGGAAGAGC
		AGATGGGACTCCGCCAGCATCCTCACGGCAGGGCCGCGCTATTGCCCTCCCTCCCTACTCTCTGGGGTCCCAGGAGCCCCAGATACG
		CAATGCTGCCAGGCGATTTCTGGCGCCCCGCAGACCCCTGCCCCTGGAGTTGGGCCAGGTCCCGGCTGGAGCAAAGGGGGCTCCTTCAA
		GCCCGCTCCTCCCTGTCAACCCGAGGAGCCTGACAGGCGCAGCGTCACCAGCGTCACCGGGCCATAGTGAGCCGCCAAGCCAGCGTCA
		CCGGGCCATAGTGAGCGGCCAAGCCAGCGTCACCGGGCCATAGTGAGCCGCCAAGCCAGCGTCACCGGCCAAGCCAGCGTCACCGGCCAAGC
		CAGTGTCACCGGGCCATAGTGAGCGGCCAAGCCTTGGTCTGCCAGAGCCGGCCGCACCAGAAGGATTTCTGGGTCCCCAGTCCTGGAGG

SEQ ID NO	GENE NAME	SEQUENCE
		AGCACACGGTTTACACCAGGCCTTGGGAGGGGAAGAGGCAAGGCGTGGGCCCAGCCCTCACTCCCCAGGAGAAACCCTGTTTGAGCGGC
		AGAGGAGACTGGAGAGACCCCAGGGCGGGGATCCCTGAGAGGAGAAACCCGGAATTCATCCACGGAGGCGTTCACCCAGAGGAGACC
		CGGAGCTTCTCCAGGAGAGGCTGGATTGCTCCAACAGGGGCCCTGAGGAGCTGATGGCAAGAGCGGAAGGCAGCTCTGACTCGTGCGTC
		TGACTCCAGGTGTGGCCGTTGGGCTACAGTGGGACCAGCCTGTTGTCACTGAACCCACAAAGTGCCTCCGAGCGCGGGTGGAGAGAGG
		GGGACCTCCCACCGTCTGCTGGCCTTGAATCTAATTCCCGTCTGTGCTTTGATGGGAGAGGCACTGGGAGCGGGCGGCTTTTTC
		AGTTCCTTTTATCTTGAATGGCCTTTGGGGATTTTCACAGATTCTGAGTTCAAAGCCCAGGGAGGTGTGGGAACGTGACATTCCTCACCGC
		ATTCCTCACCGCATTCCTCTGTAAACCAGGCGGTGTTGGCACCCATGAGCCTGTGTCTTCTATGACATCAGGAGTTTTATCCCTCACGTCAG
		AAATCAGGGTTCCAGGCGCCTTGGTTTTTCTTGGCGCCAGCGGCTTGGCTATAGAAGAAAAACTGAAGGGGGCCAGGTGCGGTGGCTCACA
		CCTGTAATCCCAGCACTTTGGAAGGCCAAGGCGGGTGGATCACGAGGTCAGGGGTTCGAGACCAGCCAACATGGCAA
245	COL18A1	GCTCCTCAGGGGGAGGTTCGGGGCCTTTGGTCTCTGGACTTGGGCAGCAGAAAGGAAACATCCCTGGGGGCCTGTGGTGACCCCCATCC
		TCCCCAGGGTGGTCTGGCAGGGGACACTGTTTTCCAAAGCAAAGCCAGAGCGCCAAGGGCTCTCGGGATTCACGAGATCCACATTTATCC
		CAAGTTAGAACAGCACATCTGTGCGTGCAAACTTCATTCTGACTTCGGCCGGCTGTCCTTCTTGCCCAAAGCACCGTGAGGCCTCATCCCT
		GCATCCCTGTTGCTTCTTTCATGTGGGATGAGAACCCAGGAAGGGGCTGAGTGTGACTCCTCTGGTTTTTAGAGAGCACTGCCCCCGCCCC
		GCCCCCTCCTGCTTCCCCACCTTTTCACAGTTGCCTGGCTGGGGCGTAAGTGAATTGACAGCATTTAGTTTGAGTGACTTTCGAGTTACTTT
		TTTTCTTTTTTTTGAGACAGAGTCTCGCTCTGTCGCCCAGGGTGGACTGCAGTGGTGTAATCTTGGCTCACTGCAACCTCTACCTCCCGGGTT
		CAAGCGATTCTCACATCTCAGCCTCTGGAGTAGCTGGAATTACAGGCGCCCGCCACCACACCTGGCTAATTTTTGTGTTTTTAGTAGAGATG
		GGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGT
		GTGAGCCACCGAGCCTGGCCTGGAGTTATTTTGGGAGAGGGCAGCCCCTGGTTCAGCGTGGCGAGGCTGCGCTTGCTCTCCCGGGCGGG
		CGTCCACACCCTCCTCGCCGAGATGGAGAAGCCCAAACCCCTGCAGCGCTCCCCCATCACGTCCGGCCCTGGAAGCCCCCGGAAACCCT
		GCCACGCCCTGAGTGGGAGAGCGCAGGTCCCTTTCCGGCCCTGGAAGCCCCCAGAAACCCTTGGGTGCCAGGCCTGGCCGGGACAGCA
		GCGACACTGCATGCTCAGCCCTTGCGTGAGACCACGGGAGTGTCCGCCCTCTGCACGTGCTGCTGATTGCCCACTTCGTCCAGCAGGTTT
		GGGAGCTTGTGGCTGCATCCTCCTGCAGACACTTGCCCATTCTGGGGCCTCCTCTCTGTCTTTTCTCCTCTGTTGAGGGGTCTGGGAGGGA
		GGCCTTGGAGGGTACCCATGCTGCTGGGACTGATGCTCCCCGCGGTGGAAGGAGCTGCCTCTTGAACAGCAGGGGGCTGAGCAGAGGG
		GAGGGGATGCGGGGGTGCCGTGCACACAGGTGCTCTCAGGACGCAGGGGCTTCTCAGCCCTGCTGTCCCAGGGCTGCACTCCAGCAGG
		GCAGACTCCTGAGGTGCAGACACCCCAGCTTCACGCTCACACTTCTGGAAGGCGATGTCTGTGCGTTTGCTTTCTGCTGCAGTTTAAAAAG
		CCGGGCTCTCTCCGGAGCGTGTAGGGCCTGGTCACTGGAATATCTGGACTCAGTGTTAATGGCAGCCACGCTGGGGGCTGGGCCCAG
		CTTTCTGTTCTCCGTGTGGGTGCCATATCCACCTCCATCGCAGCCCTTTCTCTCTCGACCTTTTAAATCACAGTGTCACCTCCCCCTGCTGT
		CCTGCCAGTGGCCCCTGGAGGCTTCTCCCCACCCCTTTCTTCTGGGGCAATTCTTAAGGCTGGCATTGAATCAGGAGGCCAGATGTGGCC
		CCTAGTAACTCACCAGCAGTCCCTGAGGCTTCTGGCTCCCCTGGCCCACCAGCCTCCCATGTCTGCCTCAGGCCTCTTGACCCGCCTGGC
		ACTGACCAGACTGTGTGCCCGGGTGCCGTGCCCATGGGCTCCGCCTCCCCCAGGCAGGCCCCCCTCTTGCTCCGCGGCCACCCCTGCTCT
		TGACCTCACACCTCTGCGGTGTGTCTGGACACACCAGCACCACGGCGGGCGGGGAGCGGAATTCTCCAGGTGGGGTGGGCAGGCCGGC
		GGGTGTTGAGGTCTCTGTGCATGCTTGTGCCGTACCCTGGACTTTGCCGTGAGGGGTGGCCAGTGCTCTGGGTGCCTTTGCCAGACAACTG
		GTCTGCCCGGGCCCGAGCATTCATGCTGGTCGCCATCACGTGACTCCCATGCGCCCCTGGCCCTGGGGTTGGGTCTGCAGGACTGAGAACCA
		GCGGAAGGGGGGGCGAGGCCTCGGGAATGCCGCCGGCAACTGGCGATGAGCTCAGGCCTGACTAATGAGCCCAGGTGACTCATACACCCG
		GGGCCTGGATGAGTCTGACTGGGTCAGGACTTCCCTGCTTGTTCTGTCCTGGGAGATGTTGTCCCTGGCCCTGCAGAGCCGGGAGGACAC
		GAGGCCTCCTGGGTCACAGCCAACGCAGCCTACTCCTGCCCACTGCTCGCGCCGGCCAAGGCCCGTCGGCACCACCTCCTCCATGAAGC
		CTTCCTGACTGCCCCCATCCCTCTGTGGGCAGCTCGAGTGTGCATCTTGAGTGCTGTGCAGGTTGGGGTCCGGCGCTCCTGCAGGCAGGC
		GGCGTCTGGGCCTGGGGGCTCTCAGAGTTTGAGGAGCGTGTGGTGAGGGTGGCCTCGGGCCTCAAAGACGCAGCGCTGTGGGAACCGG
		GAGACTGGCTGAGCCCCGCTCTGAGGAAGGTGGGGCCAGGGGGCACCCTCAGCTGACCCGGCGTGCAGGGGGTGACCAGCCAGGCGTGGCC

SEQ ID NO	GENE NAME	SEQUENCE
		AAGGATGGGGTCTCTGGGATCAGGAGACTTCAGTAGCAGCCAGGACCGAGGCCACCAGTTTCCACCCTGGCATTTTCCATCTTTTGAAGGA
		CTGGAAACGATTGGATTCTTTAACTTTTTTTAAGTTGAGGTGAAATTCACAACGCATAAAATTAACCATCTTAAAGCGAACAATTCGGTGACATT
		TAGTACAGCCAGAAGGCTGTGCAGCCATCACCACTGCCCAACTCTAGAACATTCACACGCCGGAGAGAGGGAGCCCTGGGCCATCACGCA
		GCCACCGCCCCGGCCCCAAGAACCTGCGAGTCCACTTTCCACCCTCTGGATCGGCGGTTCTGGACGTTCATGCAGGTGGTTCCCGCAGTGCG
		AGGCCTTTTGTTTCGGGCTCCTCTCACAAGCCTCACGTTTCCAGGTACGTCGTGGTGTTGTGCAGACCCACAATTCATCCCTTTTCATGGGT
		GTGTAATAGTCCACCATAGATTCTCTACGTTTTTAAAGCATGTTTTATGTGCCTGAAATGTCTCTGCACTCGAGACTATAGCTTGCTTTCTTTCT
		TTTCTTTTTTTTTTTTTTTAATTTGAGACGGAGTCTTGCTCTGTTTTCAGGCTGGAGTGCAGTGGTGCGATCTCGGCTCACTATAACCTCTGCCT
		CCCAGGTTCAACTGATTCTTTTGCCTCAGCCTCCCGAGTAGCTGGGACTATAGGCGCGCCACCCCACCCGGCCAATTTTTTTTGTATTTTTAG
		TAGAGATGGGGTTTCATCATGTTGGCCAGGATGGTCTCGATCTTCCGACCTTGTGATCTGCCCGCCTCGGCCTCCCAAATTGTTGGGATTA
		CAGGCGTGAGCCACCGCGCCCAGCCGAGACTACAGCTTTCTTTAACTGCATCCCTGGAGGGATCTGAGAGTCTCTTTCCCTGTCTCCTTTC
		CTTTGGAAAACATTTCAGCCAGGGCTCCCCAAGATGAAAGGCCAGAGTCCCAGGCATGGGCGTTGCAGGTGCACAGTTGCCACGGGGAGC
		TGTGGGTGATGGTCGCTGTCAGCGATGGCTGCTGCAGGTCCCTGTGAGGAAGGGGGCAGTGCCACAGCAGGAGGAGAGGGAGTCAGCGG
		ACGTTGATTGGCAGTGCCCGCCCATTCCATCATTCAGTCACACCCAGCACCCAGGCTCGGCTGCATAGAACATGGCCCAGG
		AAGGCTCCACTTCCTGTCTCCTCTTCTCCCCTCTCCAGTCTCATGATGGGGCTGGAGGCATCTTCTAGTTTTGAGTTCTGAGCTAATGAACA
		TGCTCATGAGCAGGCAGGCGGCAGGATCCCCAGGGACGGTGGAGCTGGGAGCCTGACTGCGGGGTGACGGACAGGCTCTGGCAGCCCCCTGTCAGC
		ATCCTCTCCAGGGCATGTGAAAGCCAGTGTGTCCTCAGCTGCCAGTGCCCCCCTCCCCACCTCCTCTGGGCCCATGTGCACGGGACCTGGG
		CTCCCCCCAACCAAGCCTGCCCCGCCTTGGTTCAGCAGAACGGCTCCTGTCTCTACAGCGGTGCCAGGGCCAGGAGTGCTGTGTCTGTGAAGC
		GGGGTCATGGTTTTGGGGCCCTCATCTCCCTCGCGCCCTCTCATTGGGGACCCCCCCGTCTCCCTAGCGTTCTCTCGCCCTCTCCAGCATGTGAAGTGGGGTCATG
		TGCTGTGTCTGTGAAGCGGGGTCATGGTTTTGGGGCCCCCCCGTCTCCCTAGCGTTCTCTCGCCCTCTCCAGCATGTGAAGTGGGGTCATG
		GTTTGGGGGCCCCCCATCTCCCTAGCGCCCTCTCGTTGGGGACCCCCCCCGTCTCCCTAGCGCCCTCTCGCCCTCGCATGTGCTGTGTC
		CATGAAGTGGGGTCATGGTTTGGGGGCCCCCTATCTTTCTAGCACCCTCTCGCCCTCTCCTGTATGTGAAGTGGGGTCATGGTTTGGGGG
		CCGCCATCTTTCTAGCGCCCTCTCGCCTTCTCCTGAGCGGTGTGGAACTCTGTGGTGGTCAGAGCTAAGGTTCTGAATAGGTCGAAGCACCT
		CCCCGGGTGCCTCTCACCCTGAATGCTCTGGGAGGACACAGCCTTTTCATAGGCTACGACTGACATGGCAGGAGGGGCCTGCCTGCCACCCC
		GGGTCCTCTGCTGCCTGCTTGCTGGGGAGGGGGCTCGAGACTGGGATCCTGGGCTTCTGCTCCAGCTGTGCCCAAGGGAGCTGCTG
		AGGAGGGACCGGGTGGGGCATCCACTCTGGGCAGGTTCAGGGTCATTCTTGGTGACCCCGGGTCCGGTTACAAAGGCTGATGGAGCGCG
		TGGGTGGCTGCCTAAGTCTCTGGAAGCCCAAGAATGTGGAGATGGCGCGTCTCGGCCCGGGGTCTCGTGGCTGGTCTGGGAGAACTTGC
		CTTTATTTCTAGGCAGGAGGCTGCACTGCAAGGGAGCGTCAGTGGCCCGGCTGGCTTTCCCCGGCCCTCAGCCCGCACTCGTCCACCAAA
		GCAAGCTCCTTTGTGGGGCTGCCCTGGGAAGCCGGGATCACGAGGCTCTGCCGGCCGTGGTCACCCCATGAGGCAGGGTCAGCTCGGG
		AGCAAGGCGGATCAGATGGAACAGAACACGTAGACCACCTCGCCCGCCCTTAGTCAGCTGGGCCATTGAAAATCAAGTCCGTAGAAAGAC
		CTAGAAATAAGTCCCGGGGTGCCCTTGCCTGTTGACGGGCGGGCCGAGCAGGACTGTTCTCAGGCAGGCACTGGTCTCTTGGCTTCCAGG
		TGGTTTGTTTGCTGGTTTGAGGCTGGGGGTGACGCTCCTGTGCGGGAGGAGGTCGCATTCCATTCATAGCGGCTTATCTGGGCTGTCAGG
		CAGGCCTGGGAGGGGAGCCTGCCTCTGTGCTCTCCAAGGGTGGGCGACGGACAGACAGGGTGTCCCACCCCTTCTGGGCCAAGGACAGA
		GGGTCAGTGTTTGCAGAGACCTGGGGAGGCCCAGGTGACCTCCACCGAGCACCTGCTGTGTGCAGGGCCAGTGCTGGCTGCAGAGACAG
		CGGAGCGTGTGTGGACCCGGCGGCCCAGGGGAGGGGGGCAGGCAGGACCCGGCGGCCCAGGGGAGGGGGGCAGGCAGGACCCGGCG
		GCCCAGGGGAGGTGGGCAGGCAGGACCCGGCGGCCCAGGGGAGGGGGGCAGGCAGAGGCCCAGGCGGCCCTGGGGGTC
		GGACCCGGCGGCCCAGGGGAGGGGGGCAGGACTCGGCGGCCCAGGGGAGGGGGGCAGGCAGGACCAGGCGGCCCTGGGGGTC
		AGGGGTGGAGGCCAGGCCTAGACGGCCCACAGGAGGGTGGACTCATTCTGACCGATTCCTGGAAGCCCCCGGAAAGTGGTGATGTTCTG
		GAGGGCCCAGCAGACCCCAAGGCCCCCAAGACAATCCCAGCTGGCTCTCTGCGGCTCTCGGTGTCTGCCATTTGAGACAATTTGGGCACA
		GGCAGGGCAGGCCGTCGCGGACGGTCTAAGCCGCGCGCATTGGTGGGGGCAGCAGAGCCCCTGCTCTCAGCTCCTCGGGGTACAGCGG
		GGGTACCAGGCGGGGTGAGTGGGTGGGTCACTGCTCCTGCCAAGGGCAGCCCTGGTTTGGTTTGCACTTGCTGCCCTGGTGACGGCT

SEQ ID NO	GENE NAME	SEQUENCE
		GCTCTCATTCCTGCCCCATTGCTAACAAGGGTGTCATAAGCTACTTTCCCGGCCCACATCCTATTAAGCCCATGGAGACCCTCCCACAGCT
		GAGCCTGCTGTGGGCTGCAGGCCCTGGGCGGTGCCCACCTCGGTCCCCACTGGCCTCCTTCCAGCACTTTAGAGCAGACACAGGTTGGA
		GATAAGGAAAGTTCCAGAGCACAGACTGGAACAAGCCCCAGGCCTCTCCCTGCCCCAGCAGGGCCTCCCTGGATTTGGGGGACAGGTGC
		CCTCATGGGGGGTCCTGAAGGTCAGAGCTGGGGCTGGGCTGGCGGAGGTGGCCTTGGCGGAGGCCACATTCCAGGGTCTCAG
		TGAGAGTCTGTGGCAGGCAGCCTTGCAGATGCCGCTGAGGGACCCCCCACTTCATGTTGTGGGTGATGTCGTCCATTGATTGCCTCCAGG
		TTTAAATCAGGTGGATATTTACCTAGCGGCCTCCTCTCCCTCTGCACAGGGCCTGGAGTGGGATGACTGGGGTGCTCAGCTGGAGGCTC
		TGCAGACACAGCCCCCCTGGGCTATGCAGGCCCTGCTGGGAGCCACATTGCCATTTTTCATCACCCACTTTTTGGGTGAGAACCCCCTCGAG
		TCCTAACATCTGCCGCATCTCAGAGCCTGTGGCTCCAGTCAGAGCATCTGGACCATACTGCTGGGGGTCAGAGCGCGGCAGGACAATGGC
246	COL18A1	TGCCACCACCATCTTCAGGTAGAGCTTCTCTCTCCTCCTTGCTGGGCGGGGCCCCTCCCTGGGGAAGCCTGCAGGACCCAGACAGCCAAG
		GACTCTCGCCCCGCCGCAGCCAGCAGCTCCAACGCCCTGACGTCCGCCTGCGCCACGCCACTTCTGCACCCCCTGGTGATG
		GGCTCCCCTGGGCAAGCACGCGGCCCCCTCCGCCTTCTCCTCTGGGCTCCCGGGGCGCACTGTCTCAGGTCGCAGTCACCACTTTAACCAG
		GGACAGCGGTGCTTGGGTCTCCCACGTGGCTAACTCTGTGGGGCCGGGTCTTGCTAATAACTCTGCCCTGCTCGGGGCTGACCCCGAGG
		CCCCCGCCGGTCGCTGCCTGCCCCTGCCACCCTCCCTGCCAGTCTGCGGCCACCTGGGCATCTCACGCTTCTGGCTGCCCAACCACCTC
		CACCACGAGAGCGGCGAGCAGGTGCGGGCCGGGGGCACGGGCGTGGGGGGGCCTGCTGCAGACGCACTGCCACCCCTTCCTCGCCCTGG
		TTCTTCTGCCTGCTGCTGGTCCCCCCATGCGGCAGCGTCCCGCCGCCCCGCCACCCTGCTGCCAGTTCTGCGAGGCCCTGCAGGA
		TGCGTGTTGGAGCCGCCTGGGCGGGGCCGGCTGCCCGTCGCCTGTGCCTCCGCTCCCGACCCAGGAGGATGGGTACTGTGTGCTCATTG
		GGCCGGCTGCAGGTAACTGGCCGGCCCCGATCTCCCCACCCTTTCCTTTTTGCCTTGCCAGGTAAGTGTGGGCGGGGCTGACGTGAGCCT
		GGTACAGGTTCCCCCCCACATCGAATCTCTACGTTCAGGGGCCCGTGGCCCTCGGGAGGTGGGAGAGCTGGGAGTGAGGCCTCCTGTGTG
		GGGAGGAGGCCGGCGTCTGGACAGGAAGAGGGCTGGATGAACCGCAGCCGATGTGTCCAGGTGCCACCTGGGCCTGGAGCTCCCTGAG
		CATTTTAGCGCATTTAGTCCTCAGCACGGTCCCGAGATACCCTGCCATGCCCCGAGTCACAGAGGGGAAACTGAGGCGTGGGGCAGTGGC
		GTGACTCACCCCAGGGAGCCGAGATTCCCGCTCAGGTGTGGCTGCATCGACCTTGCTCCGGTCACTAAGCTGCACGGTTCGATGCGCTTC
		CTGGGAGCCCCAGCGTGCTCGGGCCAAGGGTGCTGCCGCGTGGGCAGTGCAGAGACCCTACCAGCGTGGGGACCAGGGAGGTCTGCAG
		GGCCCGTCCTGAGAGGGAGCCTTTCATGTCCCCCTCCCCATCCTGAAGCACACAGCCTCCCTGCCACAGTGGGGGCCGCTTCTGGGCCC
		AGGGGACGTTGCCCCATCACCGTGTGGCCTTGTTGCTGGCTGGACAGTTGGGGGCAGGAAGAGGAGGGGAAAGGGGACTCTTTA
		ACCTCCTGGGGGCAGGGGCAGCCCAGAAAGGACCCCAGCAGATCCCTCCTCTGTGTCCGGGAGTAGACGGGGCCCC
247	COL18A1	GGGCTCCACAGCGGCCTGTCTCCTCACAGGGTTCAGCCCAGTCTGCTCTCACTCATTTGCTGATTCATTCTTTCATTCAGCCAGTCAATAGT
		CATGGCCCCTCCTGTGTGCCGGGTGGCCATGGATATTGCCCTGGGTAACACACAGCCTGGCCCTGTGGAGCAGACAGTGGGGACAGCCA
		TGTGGACAGGGTGCAGGTGGATGGCAATGGCAGCTGGGTCAGGAGGGGCTGAGGGCCGTGGGGAAAGGTGCAGAATCAATAGGGGCAT
		CCGGACTGGGGTGCAGGCCTGGGGGCTGGGATTTCTAGGGTGGAGGTCACCTCTGAGGGAGACAGAGCAAGGCCCTGGGAGATTAGAA
		GGTCGAAGGTCGCCGTGTTGAGGTCAGGGGCCCCTGAATTGGAGCCGCGGCAAAGGAGAGGGCAGGTCAGGGCACGTGGTGAGTGATTG
		CTGCGGCTTCTGAGCACGGCTGGGTCTGTGGGGCCTGAGCAGAGGTGACCCGCGATCCGGCGCCACGGCAGGCAGGACTCCCCACCCT
		TGCTGCTGCCTACACCCCCAGGGCAGCCCCAGAGTCGGGGGGCGCAGCTCCCTGCTTGCCAGTTCAGAGCCCAGCCCCTCTCACCCAGCC
		CAGAGGAGGACACAGATGGAGGAGGGGCACCCGGAGGGTCCCCCCCGCCGACAGGCCCCACGTCTCCCACCTGCAGGACAATGAAGTGG
		CCGCCTTGCAGCCCCCCCGTGGTGCAGCTGCACGACAGCAGCAACCCCTACCCGCGGCGGGGAGCACCCCCACCCCGCGGCGGCCCTGGCG
		GGCAGATGACATCCTGGCCAGCCCCCCTCGCCTGCCCGAGCCCCAGCCCTACCCCGGAGCCCCGCACCACAGCTCCTACGTGCACCTGC
		GGCCGGCGCGACCCACAAGCCCACCCGCCCACAGCCACCGCGACTTCCAGCCGGTGGTGAGTGCCCCCCCCAAGTGGGCTTGGCTCCAT
		CTAGCCCCTCGGCTCTCGGCAGCAGAAGAGGGCCCAGCCCCTGCAGAGCTGCTGGGGGTCCCAGGCTTCGGCCATGGGTGGGGGTCTG
		GCGGCTCAGGGCCACTCAGGGCGGCTTGGCTGGCCCTGGGACTTGCCCCTCTGGTGGCCAAGCAGTGGTCATGAAAGTCCAGCCGCTGTC
		ACATCCTTGAGGAACCGGCGTACCTCCGCCTACAGCGGCAGCTGGGGGCACCCACGTGGCCCGGGGCTGCTCTGACCTGGCAGCGTATG

SEQ ID NO	GENE NAME	SEQUENCE
		GGGGCTGCTGCCTGGGCCCCTCAGTGTGTCACTTGCGCGCCTCCCGCTCAGCGCCCCTCGGCCGTGCCTGTCCACACAGGTGCGGGGC
		CGGGGTGGTGCGCCCCGGGGCCTGGGTGCAGGGGGCAGCGTGGGACACAGCCCGTGACGCGCCCCTCTCCCCGCAGCTCCACCTGGTT
		GCGCTCAACAGCCCCCCTGTCAGGCGGCATGCGGGGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGGCGCGGGCCGTGGGGCTGG
		CGGGCACCTTCCGCGCCTTCCTGTCCTCGCGCCTGCAGGACCTGTACAGCATCGTGCGCCGTGCCGACCGCGCAGCCGTGCCCATCGTC
		AACCTCAAGGTGGGTCAGTCCAGTCCTGAGGGCGCGGGCTCCTCGGCCCCCACTTGACCTCTGGGGTGAACTCCCAGCGGGGAGCTCCC
		CTCTAGGGCCTCTGGAGGCCACCATGTTACAGACACTGGCGCCTAGGCTGGCGACTTCAGGGCAGGCTCCGGGTGGGTCACACCCCTCC
		AGGCTCAGGCCAGGCCTCTGCATCCCTGGGCACTGCCACGTCCCCCAGGGCATCCCATGAGGCCCCCCCGTGGCCCCCTGACCCCCCGC
		TCCCCCGGCAGTGCCCCTCAGAGGGTCCCATGCTGCTGGACCAAGTGTCCACACAGGTGATAGGGCTCACATACAAGCCTGGAATCAGGA
		ACCGTCCTTTGGGCCTCTAGTGCCATGCGGGCTGGTGGCCCCTCTGCCA
248	chr21:45885000-45887000	GCCTGGAGTGTAGTCCTGCTGAAGGCCAGAGACCACACACTCCACCCAGACTCCGGATCTCCCTCCCCAGCAGGGGGATGGAGGCCCTG
		CCGCTGGGAGTGCTGGTGTTATGTGGAAGGGCTGGGCTTCTCCAGGGCTCCTGGGAGGCCTAAACATCTTGCAAGGTTTTGACGTTAATTA
		CTATTATGATTGCTTTCTGTGTTACTGTTTTCCCCACACTTTAGCCAGCTAATGTGGAGCTACAGAAGGCCCTCGCCCCTACCCCTCCAG
		ATGTCCCAGCCCATGACAAGCAGGAAGGCCGGGTGCTGGGAGACTTCCTGGGGCTGGATCTGACATCATTCCAAGCAGATGATAACCTGC
		CTTCCCGATTTCCAAACCCACAGCAAGACACCCTGGAGTTATTTATAAATGCGAGCCCCTGGGTGCACTTCTGACGGGACCAGCACCCTGA
		CGGCCATGAGAGGGTGGAGACAGCGCACCCCGAGCTCAGGGAGGCAGGAAACTCTGGACCTGGAGGCCGGGCACCATGAGGGACACGC
		TGCAGGCCCAGCTGCTGCCGCCTGGGGCGGGGCTGCCCTGCAGGCTCCGGGAAAACCCAGAACCAGGCCGGATCAGCGTGTGTCAAGA
		GGCGGGGCGTGAGAGATGAGCTGCTTTTTTTTCTTCACAGGGTTGGCAGGAACTGCAAATAATAGAAAGTCTTTAGGGTCTAACACGCTGCC
		CTGAAAACACTATCATTACTTTCCTAATGACTAACTGTGTCTTTCAGCCGGCGGGGCAGGCAGCTGAGGCCGCAGGCTCCCGCAGAGGAC
		CGGGGGAGGCTGGCAGCCTGTAATCTGGGGGCGCTGACAGTGCTCTGCCCAGACCCTCGCGCCAGCTCCAGCTCCAGCACAGCAGCCCT
		GGGTCCCTCTGGCCCCCTGCCCGCAGAGTCCAGGTGTGGCAGAGGCCGCCCAGTATCCCTTCTCCTCCTTTTCTAAAAACAGAGTCT
		CACGATGTTTCCCATGCGGGTCTCCAACGCCTGGGCTCAAGCGATCCTTCTGCCTCGGCCTCCCAAAGCGTTGGGATTAAGGGGCGAGCC
		ACCGCGCCCGGCCCACCTTCCCTTCTGGTTCATTTCCAGTAAGGTCCTGTCCACAGCGTCCTTCCCAGCATTCCCACCAGGCTGCAGGCCT
		TGGCCTCCCTCCCCCTCCCATTCTCATTCTCCCCGAAACCGCCAAGCGCGTCCAAAGCACGGGTTCGCCAAGCGCCCCCCCCGCCCCACTCC
		ACATTCCCTTCCCCCGCCGACTCAGCCTCCGTAGCTCGCGGACGGCCCCCTCCTCACGCCAGCCCAGGCTTTTTTTTTTTTTTTTTTTTTTCTTCTATTT
		TAAGGTTGTCTTTTAATGACACAAGCGACATTTGGAGACAAAAGGACACATCTCTTCCTGACCCACCTCCAACCCCAGCTGACGGCCGCCC
		TGAGCCTGGCGTAGACGGCCCGGAACGTTCCCTGCGTGGGTTCCGTCCATCCCGAACCCCTGTCCCCGCGCCGGCTCCGGGGGTGCTCG
		GGGGGCCGCGTGGGGTCTGTGACGTCGCCTCGAGGCTGCATCCCGGTGACCCGGCAGCCCCTGGCGCTCGCGGGAGGCGCGGGCGGGCG
		CGGACCCCAGGCTTTAGGGCGCGATTCCTGCAGCTGGCTGCCGGCCCGAGGTTCTGGGGTGTCTGAGGTCTCGGGCGGGGCGAGGACG
		TTTCTCCGGCTCAGCCCCCCCACCTCCTGCCCTGCCGCCCCCCACACCCAGCTCCCCACGGACGCCAAGAGGCGCCTCCCACCCCGGCG
		AGGACCCGCGGGGGAAACGGGGCCCAGGCGGCGGCGACTGCGGAGGACGCGCCTCGGCCCCAGCGCCCTGGTCCTCGGGGCGTCCGGCT
		GCCCTTGCCCGAGGCCGGGGCGGCGCTCAGCGCCGCGGAAGAAACGCCCCGGGCGGGGACGCACAGCGAGGCGGGCTCCCGCGGGAA
		GTACCGGGAAAACGGCGCGGAGCGGAACAG
249	PCBP3	TGGAGCAATCCCAGAGAGGCTGAGGTGTTCAGGCTGGCCCCAGATGCACACGAGCGTGAAGCCTGTTCAGAAGCCAGCTCCTCACACCCT
		CTCCCCTGCCAGAGGCTCCAGCACCCCCTCCTCCTCCCCTCCCTTCCCTGTGGTCCTCCTGCCCACCCCACCCCCGTCTGCATG
		TGCACCGTCACGGAGATGCGTGTACTAGGGCGGAGGTCGGGGACAGTCGTCAGAAGGAACACAGGAAAGAAGGGAACAGGAATCCCATAA
		CAGAACATTATCCGGCAGGAGTAATTAACACAGGCAGGACTGGAGGCTTTGTTTTGTTTTGCTTAAAAAACACAGTGGTATTTAAATTAATGGGC
		ATGGGAAGACTATTCAGTGAAAGACATCGGTCATTGAGGTATCTATTCAAAAACACGGTTTAGTACTCTGCCACACACCGAACGCAACGCCA
		CAGCAGCCATAGAAGCGTGTGTGGCTGTTTAACGTGGTCTTTTTGGGAGGGCATCCTAGGCAGAGCAGGCGTGGAAGGGAAGGCGGCGCG
		GACGGAACAAAACGCGGGCACGCAACGGCTGCTGCCGCCGGATCTGAGGCAGGGCCAGCCTGTGGGAGCAGCAACATCGCTCGCAGGAC

SEQ ID NO	GENE NAME	SEQUENCE
		AGCGATGGAGCCCCCACGAATCCGCGTGAAAGCAGCAACCACCTAGAAATGAACGTACAGCTGCTTAGAAACAGAATACGGATGACCCGA
		AAGACTTCCCGATGGTAGTCACCAGCATACAGGACCTGACACGGGCGTGCGGGCAGGGTGTGCCGCTACGGGGTCCCTGGCGCACCTGC
		TACCCCTGCTACCCGCATTCACCGCACGGAGGGTGCGGGCCGTGAAGGTTATACATGCAAATATCCTTCCACCAGCCAGTTCTCCTTCC
		AGGAATCTGCCACCCGACCCTTGTGTTGCACAGACATGGTCCAGGTGTTTGCGACGTGATTGTTTATCAGAGAGAGGAAGGGAAATCT
		CCAGGCTCGCTGTAGCTGCAGGAGCTCTGGGGGCTGCGCCCATCGTGGAGACGGATAGCTGTCTCTCATGAACACAGGACAGCAAGTCC
		GGCTGCGGCCACAGAAGACTCGCCCTCCTGGACGCAGCGTCTTCCTTCCTCAGCCCCACACTGGAGGTGGCCAGTGCCATCCACAGCAG
		AAGGGGCCAGCCGGGACCAGGCTCACGCCGTGGAATTCTGCTCTGTGGTAAGAGGAAGAGCGATAGCTGGAACCCAGCGCCGTCGCACA
		CACAGCGGGGAAGAGTCTCAGAAATGTTACTTTGAGTCAAAAAGCTGGACAAAAAAAGGCGCAAGCCAGATGGTGCTGAAGAGGCCACAG
		GAGGCTGGCAGCCAGGGGGTCTGGCACCTCACTCGGAGGCGCAGTGGGCCCGTCCGGAATTAGTGGCCATACGGCAAGTGCCGAGTGG
		ACATCAAACCGTCACTTCAGACTCCTGCGCTTCACTGCCTGTCGGTTATGCCTGGGTTTTGAAATCAAGTCACAGAACACCTGGAATGTGGT
		GTTTACGCAGAACAAAGCGGGTGCCTCGGAGGAGAGAGCCTAGGGACAGGGGCACCTCCCGGTGTGGGTGCCCAGGGTTGCAGGGTGG
		CTTCCTCTGTCTGCGCGGTTTTCAGAGCCCCAGGGTCCTGCCCTGCCCGGCTGCCTGGAGGCGGCGGCCCACATCCTGCTCTGCGCCGCCGAA
		TCTCAGCCTGAACAGCTTCGCTGGTGTTTGTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATAAAGGATTCCGATGCTGTTACAG
		TCAATAAAAGCCACAGGTCTGGGTGACCTACAAATGTGTGTCTGACTTTCTGCAGTTTAAATCGCCACTGAGCCTTAAGGCGTCTGGCCC
		GCGCATTGAGGAATCCACGTGGGTCTCGGGGTCCCCATGCCTGCCCAGCTCCCTGCTTCAGCCTGGGCGGGTCTGGCGGGCATTTCTGC
		GAGCCTGTCCCTGGGCCCGCCTCCTGGCCAGACTTCCAGAAACATTGTCCACATCCCCGTTGCACGTCCCCCCGTCACCGGAAACTGCAG
		CCCACAGCACTGGGAAGAACCCGGGAGGCAGGCGTTAGGACGGGGTGGCCGAGACAGGGAAGGGAGCCATGGCGGACGTCCTCACCCA
		AGCCAGGGCTTCCTGCCCCCTGTGGTACTGACAGGAGCCCCCGCAGGACGTGGGGTTGGCTTTGGGCAGCTCGGTGGACACTTCTCTTTCAG
		ATCCTGCCACAGCAAAGCTCACGAGACTCACTTCTTCCCATTGGAATTCACTAAGAACAAATTCAACAATTCAGACGCCCCAGCTGGAGGTT
		TATTTTATGGATTTTACCTGTGCGGTATTTAGGGTTGTGTTTATGAATAAAGGTGTGCGTTCTGGCAAGTAGAAATACAGAGCTTGTCTTTCA
		CCCAAGTATCTGTAACTTTCTCCAATGCAGACACTAAAATGCAAATAAAACAAACCCAAACCCATTAAACATGAATTAGATGAGGCAGGCTGAT
		GGGAGGTTGTGGGATTAACAGGCCGTCAGCGGATTGAAGCTGCGCACATCGCTGGGATGCTGCTGCGGGAGGATTCGGTCTAATCCGGG
		AGCATCTGGCTGGGCAGTGGGCAGCGTCTGCAGTCGTGGCTGCTTGAAGGTATGAAGGTTGTGGCCTTTGCTTCCCCCCATCAGGCTGCC
		CCACCCTGGACCCCACCCAGACCCCTCGGGCACCCTGGGGTCATCTTCAGCTCCCCCTTCTCTTCCTTCCGCCTGGGCCCCT
		ACTGTGACCCGAGGTCAGCAGAGGACCCTGGCAGGTGGCTGCTCCCTGGGACTCGACTGTGCAGGTGAGGCTTGGGGTGACCGCTGCTC
		CTGCTCCTGCTCCTCTCGCCGTCCCCACCCTCCTCCATCATGCTGTCAACATGCATGTGGGCTGCAGCCCTCAGCCTGCAGGACGCTGTC
		AGTGCAGCTCCTCAGTGGCCAGG
250	PCBP3	ATCTTGTCTTCCTTGTCCCAGTCCTGGAACCAGCCACTGCCCCAGCAGCTCCTGTGTGTGGTGGCATGTTCTGGAAGCCAGGATGCATGGT
		GCTCCTGGGCTGCTGTGGGTCCTGGGCTGCTGTGGGTCCCGAGCTGCTGTGGGTCCTGGGCTGCACCCCTGCAGAACACTTCCTTCCAT
		GTTCAGCTCCCTATATGGAACCCCAGTTCCAGCCCCCACAGCACACAGGGTCCCCCAGTTCTTCCTGCCTCAGGTGTGCACCACGAGGAATCCA
		ACTGCCAGTATCTGTGCGTGGCCTCCCGCCGGGAGGAGGCTGCCGGAGGCTCTGAGCTCTAGCCCCACAGCACTGGCACATCCTAGATTT
		CCGGGAAGACACGGCCTCCTCCCCAGGGGAAGGTGGTGGTGCCCACACCCAGAGCATTCATTCCTGCAGTGGAGACAGAGGGACCTGCC
		TCTCCAACTGTGGGTGTCAGGAGCCAAGGCCGCATGGTAAATGGGGCTCTCTGTGAGGCCAGGTGCACGGCCCCATCTCCAGCAGCAGCG
		GCCATGCCACCCAGCTGCACTCTGTGGGGGAGGTGCCATGATTGACGGGGGCCCCTCCCTGTGTCCAGTGTCCTCCTCCCTCCACGGGC
		CCCTCTGCACACCCGTCCTCACAGTCTCCCCTCTGCACACCCGTCCCTCTGCACACCATCCTCCTCATGGTCTGCACACC
		GTCCTCACAGCCTCCCTCTGCACACCGTCCTCCTGCACACCGTCCTCACACCCTCCCTCTGCACACCATCCTCCTCATGGTCT
		CCCTCTCCTTCCACAGACCCCTCTGCTCGCCATCCTGACGGCCTCCCTCCCTCCACGGACCCCCTCTACACACTGTCCTCCCAGCCTCCCC
		TCTACACGCCATCCTCCTCACAGCCTCCCTCTCCCCTCCACGGGCCCCCTCTACACACCGTCCTCACGGCCTCCCTCTCCCTCCACGGGCCCCCTCT
		GCACACCGTCCTCACAGCCTCCCTCTCCCTCCACGGGCCCCCTCTGCACGCCGTCCTCACGGCCTCCCTCTGCCTCCACGGGCCCCCTCTGC

SEQ ID NO	GENE NAME	SEQUENCE
		ACGCCGTCCTCACGGCCTCCCTCTGCCTCCACGGGCCCCTCTGCATGCCGTCCTCACGGCCTCCCTCTCTCCACGGGCCCCTCTGCAC
		GCCGTCCTCACGGCCTCCCTCTCTCCACGGGCCCCTCTGCACGCCGTCCTCACAGCCTTCCTCTTTTTCCACAGACCCCTCTGCACGCC
		GTCCTCACGGCCTCCCTCTCCCTCCACGGGCCCCTCTGCATGCCGTCCTCACAGCCTCACCGACGTCACCATTGCTGGCCCCGCTTCAGG
		TGACAGGCCACAGTAGCACCTGTCAGCTCTGTCCCGCTGCTGGACAGGGAGATACTGGGCCACTCAGCCCAGCGGGGAACGTGTGTCCC
		GAAACTGCCTTGGGCTCGCCATCAGAACTGTGGCAGCATCTTCCAGCGTTCCTTTTAACAGGCTGCCGTTGGAATAGGAGTCACGGAGCAA
		TTGCAGTGCTAAGTTTTCTTTAAGTCACACAATTGAAGGAGGCTTTATTTTTCACACATTTCTTCCAGAGTTTCCTGGTAGCCTGAGTGCATG
		GGTGATGCCCCCTGAGTTATTTATCAGGGGCAGCCAGCTGCCCTCCCCCGGGGCACTTACAGTCAGCCCATCTCTGTCCTGGTCAGGTGG
		GCGCCAAGGAAGACCCGGCTCTGTATGGGCAGCCTGGCTTGTACACACACCCCTCCCCACCAGCAGATTCTGAATTCTCCCT
		TCTTCATGCACACCGGGAAGGTCCCTTCTGCACTCATACCGGGAAGGTAGGCAGGTTTCGGTAGTGTCTGCCTCCAGTGTTTTCCTCCTCC
		TGCTCTATGACATCATCATTTCTGTGATTTTTTTTTTTTCTTGCAGGAAGTTGGAAGCATCATCGGGAAGGTAATTATTGATTGAATCTCTGCCTCT
		CCTGGGGTCTCTGTAAGGGGATGGTGAGGATGGCAGCCTCCCTGGGTACTAGGTGGCACCCAGTAGGTGCGCCTTTCCCAGTTGGTGGG
		TGGTCTGTGTTCCATGAAGACAGGACCCCAGAGGTGTCGCCTTTATGCTGTATGACATTGAAGCTGGTCCCTGGCTCTGCGTGGCCTGAGG
		GGAAGGGGTTCACTCCAGCTGGTCACCTGCCCCCCTGCCCGTGGCCTTGGTGGCCAGTCCTTCTTTCCCGGTTGAAGACCCCACGAA
		GAATGATTTCTCACGCCTTCTTCAGCCGGCTGTGTAGTCTGGGTGGTCTCCAGGAGTGCCAGTGGAGGCAGCAGCCCCCAGACAATTCCTT
		TCCAAATCAGGGCTGGCCCGGGGGAAGTAAGGCCCAGTTTGGAAGCCTGCTGCCCCGGGAGGCCGAGCAGTGAGGGCCACCTCCCTGTC
		TTCATCACATTTTCACCGCTTCCGGGGGTCCTTCCCCTCAGTCCCACCATGGGGGCGCC
251	COL6A1	GCTGGACACCTCTGAGAGCGTGGCCCTGAAGGCTGAAGCCCTACGGGGCCCTCGTGGACAAAGTCAAGTCCTTCACCAAGCGCTTCATCGA
		CAACCTGAGGGACAGGTAGGAGGGACGCCCCGTGACCTTCCTCCTCTGTGCTTCTGGGCCTCTTGGAGGGAGGGGTGGGGGCCCAGGGGA
		ACACGGGTGCGACGGCCTCAACCTCCTAAGGTTGGGCGAGCGTTGCCCTGACCGGGGCCCCTCCCGGCGCCCTCCAGAGTGAGGCCGG
		GGCCCTTTCCGGCGCCCTCCAGAGTGAGCTGGTCTGAGCCTCTCCCAGCGCCTTCCAGAGTGAGCTGGTTTGAGACCCTGCTCGCGGGG
		GTGGCACCTGTTCAGCAGGGCCGAGGTGACAGTGAGGCTGAGATGTAGGGAAGAGAGGCTCCCGCAGGCTGACCGAGAGGGCTCAGCG
		CACTGGCCCAGACACGCAGTCCTGCCTGGTGCGCGGGAGCCCCCTCACTAACCACCTGGACCCTGGTTTGTTCCGTGGGCAGTGAGAGCC
		TCTACCTGGGTCCTGGATCCCACGTTCTGAAGGTCCCCGACTCGGGAGCCAGGAGGGGTGTCGCTCTGCAGCCCCAGGGCCCCCAGGCT
		TGGTTCTGGGCTTGGGACACGGCACCCTCTGCTCCACGTTCCTCCATCTGTGCGTGGCTGAGGACAGACGGGGGGAGAGGGGAGTC
		GGTCCTGTGGGTGCACAGGGCCGCTGAGGGGGGGGCATGTAGAACGGGGCTCCCCCACTGAGACGGGTCCTGGCAGTGGGGACACAGC
		TTAGCCGGCGTAGGAACCCCCGTCCTCCTTGACCCTGCTGACTGGCCGCTGGGCCGGAGCCTCCCGCCACCAGAAGGGGCACAGTCAGA
		GGCTGCCGGTAACAGCAGGGTGGACCTTCCAGCCCACACCGTGCCCAGCAGGAGCCATTGGTACCAGGAACCCTGAGCTTAGTGGACAT
		GGCCAGGCCCGTGCGGCAGTGTTTGGGGGGGGGTCTGGCTGTGGATGGCACCGGGGAGGGGCGGCCGCGTGGCCCAGCGTCCCCCGA
		GTCGCCCTTGTTGCCTTTACTCAGTCTCCCCATGACTCAGTTTCCCACCTGTGAAATGGGGGCGGAGTCATCCCCATGTCGCTGCCACTGGA
		TTCCTGCAGGCGCCGTGGTCACTCTGCTGAATGGATGGGAGGGTGGGTGGGCAGAGGTGGGCCCACCCCAGGCTGGGGCAGAGCAGA
		CCCCTGAGAGCCTCAGGCTCAGGTGCTCAGAGGGCAGCGAGGGGGCTGCTCAGATCCCCGGGGTGCCTCCTTCCCCCACTGTCATGCTG
		CCCCACTGCAGGCCCAAGGACCCCCACCCCAGCAGGCCACACACTCAGGGCTCCTGGTCTGAGGGCCTGAGGGATCGGGGGCGCAGGTC
		GCTTGCTGGCCACACCCGCCTGCACAGCCTTCCAGGAGGGCCCGGCCTCAGGGCCACAGGGCAAGTCCAGCTGTGTGTCAGCCACGGCCA
		GGGTGGGGCAGCCTGTCCATCTGGGTGACGTCGCGCCCTGGGACCGGGTAGCGATGGCGCCAGGGGGCCGCCCGCCTCACGCCCGCCGT
		GCCCTGTTCCTGGCAGGTACTACCGCTGTGACCGAAACCCTGGTGTGGAACGCAGGCGCGCTGCACTACAGTGACGAGGTGGAGATCATCCA
		AGGCCTCACGCGCATGCCTGGCGGCCGCGACGCACTCAAAGCAGCGTGGACGCGGTCAAGTACTTTGGGAAGGGCACCTACACCGACT
		GCGCTATCAAGAAGGGGCTGGAGCAGCTCCTCGTGGGGTGAGTGGCCCCCCAGCCTCCTGCCCACGCCAGTTCTCACGCGTGGTACCCAG
		CCTGGGCTGGGGTTGGCCTGGGGTCCCTGTGCGGCTTCAGCTGCAGCCTCCCTGTTCTCTTGGAGGCTGCACGGCCTCCCTGACCCACTT
		TGTGGGCAGGAAAGAGACGGAGACAGACAGAGACAGAAACAGAAACAGAAACAGACACAGAGAGAGACAGAGACAGAGAGA

SEQ ID NO	GENE NAME	SEQUENCE
		GATAGAGACAGAGACAGAGAGACAGACAAAGAGTGACAGAGGGACCAAGACAGGCAGACAGAGACAAACAGAGACAGAGACAGAG
		ACACAGAGAGAGACACAGAGACGGGAACAGAGACAGGCAGACAGAGAGACAGAGAGACAGAAACAGAGACAGAGG
		GACAGAGACAGGCAGAGAGACAGAGACAGACAAACAGAGACAGAGAGACAGAAACAGGGACAGAGACAGAAAG
		AGAGAGACAGAGGGAAACAGAGACAGAGGCAGAGAGAAGCAGACAGAGAAACAAAGACAGT
		CAGAGACAGACAGAGACAGAAACAGAGACAGAGGGGCAGAGACAGGCAGACAGAGAGACAGAGAG
		ACAGCGAAACAGAGACAGAAACATACAGAGACAGAGAAGCAGACAGAGGCAGAGAGACAGAGAAGCAGAGA
		CAGGGACAGAGACAGAGACAGAAATAGAGAGATAGAGACAGAGGGACAGAGAGGGAGACAGAGAGATAG
		AAGCAGAGAGAGAGACAAAGACAGGCAGAGAGACAGAGAAGCACAGACAGAGACAGGGACAGACAGAGA
		CAGAGAGACCGGAAACAGAGGCAGAGAGACTGAGAGAGACGGGGGTGGTTTTCCCCACAGCATCAACACCAAGCAGGGCTAG
		GATCACTGAAACAGACTCATCAGACCCGAAGCATGCGCTTTCTCGGGGTTTTTCTGGACTGAGGGGTTTCCTCTCATCCCAGTGTCCAGCT
		GTGGGGACGCAGGGGGCCGCAAGCCCCGGAGTGTCCAGAGGGGAACGTGGCCTCCCCACACCCAGCCCTTCACGAGGCCTCAGGATCCC
		AGTGGGGGGTACCCGAGGCTGCCCTGTCCAGCCAGGCGGTGCGGGGGGTTTGGGGAGAGCCTCTCCCCGAGGTCGGTCTCAGAGGGCCA
		CATGGCCGGTGTGGGCCGGACATTCCCTTTCCAATGGTTGTGCCCACTTCCCTCCAGAGTTGGTGCCAAGCTGGGGACCTGGGGGACTTGG
		AGTCTCAGGAAGTCGTCCGCTGTCTGCAGGGGGGTGCATGGGGGATGTGGCCACACACGTCAGAGTGCGGCCCCCCTGTGGAAGCCACAGA
		CAGACACGACTCCCCCTAAATGAGCTCGCCCTTCTGGCCGAGATGCTCAGCGTCCCCAGCAGGCTGCCCGACTGCCCTGCGATACTGCCCT
		CCTTCCTGCTGCTCCCACTTTCCCTTTCGGGGGGGTTGGATTTGGGGCATTCAGGGATCGCCCTGTTGTTTGCTCATCACACCCATTTCCTGC
		AAGAGCCACGGTGACCGAGCAGCCTTGAGTTGAGGCAGCTTGTGGGTAGACGCGGCGGGCATCTCGGAGGGGCACGCTCCCCTGCCACC
		CTCAGCCTCCACTCACTGGTCAGGGGGCTTTGCCGCCCCAGGGCACCCCAGGAACCGAGCCTCCTTTGGGGTCATGGGTGCCTCTCCTGGG
		AGGGCGTGGATTTTCCAAAGCAGTTTAGAGAAATGAGACCCACAGGCCGTTATTTCCCATGGTGAGGTTCTTTTCAGTAACCCCCCACCGTATA
		GCCAGGATCAGCAAAGAGAGGCGGCTCCTCCCGGTGAGACAGGGACCAGCACCTCCCGGACAGGCTTGGGTCTCCCTCCAGTTCCCCCA
		CCTAGTCTCGAGGTCTCACGCTGCCCTCTCCTGTCCAGGGGCTCCCACCTGAAGGAGAATAAGTACCTGATTGTGGTGACCGACGGGCAC
		CCCCTGGAGGGCTACAAGGAACCCTGTGGGGGGCTGGAGGATGCTGTGAACGAGGCCAAGCACCTGGGCGTCAAAGTCTTCTCGGTGGC
		CATCACACCCGACCACCTGGTAGGCACCGGCCCCCCCCGGCAGATGCCCCCCAACCACACAGGGAGTGGCGGCTGCAAGGCCCCCGGCAGC
		TGGGACCGTCTTTTGGTCCTCGGGAGGGTGTGGGTTCTCCAGCCGGCCACCCTTGCCCCTGAGAGGCCAGCCCCTCCTGCTGAGGAGCC
		TGGAGCGCCCCAGCCCAGCCTCCCCTCTGGCCCTGTGGGAAGCGGCCCCGGCCGTCAGGGGTCCCAGCCCTGCTCAGCCCACCCTGAA
		CACTGCCCCCAGGAGCCGCGTCTGAGCATCATCGCCACGGACCACACGTACCGGCGCAACTTCACGGGCGGCTGACTGGGGCCAGAGCCG
		CGACGCAGAGGAGGCCATCAGCCAGACCATCGACATGATCGTGAGGCCCCTGCCCAGGAGACGGGGAGGCCCGCGG
		CGGCCGCAGGTGGAAAGTAATTCTGCGTTTCCATTTCTCTTTCCAGAAAAATAACGTGGAGCAAGTGGTAAGAGCCCTCCCCACCACCCCC
		AGCCGTGAGTCTGCACACGTCCACCCACACGTCCACCTGTGTGTTCAGGACGCATGTCCCTATGCATATCCGCCCATGTGCCCGGGACAC
		ATGTCCCCTGCGTGTCTGCCCGTGTGCCCGGGATGTGTGTCCCCCTGCGGTGTCCACCTGTGTGTCTGCCCATGTGCCTGGGACATGTGTC
		CGCCTGTGCGTCCATCCGTGTGTCCGTCTGCCCATGTGCCTGGGTCGCATGTCACCCTGTGTCCCAGCCGTATGTCCGTGGCTTTCCCAC
		TGACTCGTCTCCATGCTTTCCCCCCACAGTGCTGCTCCTTCGAATGCCAGGTGAGTGTGCCCCCCGACCCCTGACCCCGCGCCCTGCACC
		CTGGGAACCTGAGTCTGGGGTCCTGGCTGACCGTCCCCCTCTGCCTTGCAGCCTGCAAGAGGACCTCCGGGGGCTCCGGGGCGACCCCCGG
		CTTTGAGGTGAGTGGTGACTCCTGCTCCCTCCCATGTGTTGTGGGGCCTGGGAGTGGGGGTGGCAGGACCAAAGCCTCCTGGGCACCCAA
		GTCCACCATGAGGATCCAGAGGGGACGGCGGGGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGGCGGGAGTCCAG
		ATGGAGGGGATGGCGGGGTCCAGATGGAGGGGACGGCGGGGTCCAGATGGAGGGGACGGCGGGGTCCAGATGGAGGGGATGGCGGG
		GTCCAGATGGAGGGGACGGCGGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGTCGGGGCTCCAGATGGAGGGGGAC
		GGCGGGAGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGA
		GGGGACGTCGGGGCTCCAGATGGAGGGGACGGCGGGGAGTCCAGATGGAGGGGACGGCGTGGTCCAGATGGAGGGGACGGCGGGGGTCC
		AGATGGAGGGGACGTCGGGGCTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGGC

SEQ ID NO	GENE NAME	SEQUENCE
		GGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGGGACGGCGGGGGTCCAGATGGAGGG
		GACGGCGGGAGTCCAGATGGAGGGGACGGCGTGGTCCAGATGGAGGGGACGACGTCGGGGCTCCAGA
		TGGAGGGGACGGCGGGGGTCCAGATGGAGGGGATGTCGGGGGTCCAGATGGAAGGGACGGCGGGGTCCAGCAGGCAGGCTCCGGCCGTG
		CAGGGTGTGGACTGTCCCGGGGGCGCTGGGGGCTTCTGAGGGTGTCTCTGTCCGCCCTGCCCTCAGCCGCACTCTGTTCAGAAGGACCT
		TTCTGGAGGTAGGAGGGTGAGAATGTGGGTCCCCTGCTTCTGTGTGGCTCAC
252	COL6A1	GGCCGGGGAGGCGGGGAGGCTGCCCCAAGAGTAAAAGCCTTTCTGACGTGCGCAGGACGCGGCCCTGACTGGTCTAACTGACTCTTTCT
		CTTCTCCTCAGCTTGCTGTGGTGAGACCCAGGCTCTAGCTCCTGAGAGAATGGATCCCGGGGGGTCGGGGAGGCGAGGCCTGGGTCCCACA
		CATGTCACAGGACAGCACATGGCACTCTCGGTCCCCGCCCGCAGCTCCCTGCACCTGCCCGCCCCCTCTGGGGCCTGCTCCAAGCCAGCA
		GGGTTCCCGGGGTGTTGGGCTGGGCCCCGCCCTCTTTCACCCATAACTGAAATAACCAGGAGCAGGCTTGGGGGGGTCCCTGCTCCATCAT
		TCTGGCCCACAGGCCCCACCCTAGCCTGGCTGAGCAACGCCAGCCCTGACCAGCCGCCGGACAGAGCAGCCTTTACGGGGGCCATGGGAG
		GGGGTGGGCTTTTCTGGGGCTGAGACGGGGGGACCCCAACGTGTCAGGTGAGGATGTGGCAGCCAAGGAGGGGCCAGGGCGGTGGAG
		GGGAGGGGCCAGGGCACTGGAGGGGAGGGGCGTGCTCTGCTGACACCGCCCCCGCCTGCAGAATGCAAGTGCGGCCCCATCGACCTCC
		TGTTCGTGCTGGACAGCTCAGAGAGCATTGGCCTGCAGAACTTCGAGATTGCCAAGGACTTCGTCGTCAAGGTCATCGACCGGCTGAGCC
		GGGACGAGCTGGTCAAGGTGAGGCCTCGCCCCGCCCGGCTTTCTCAAGCCCAGGTGCACCCCGACCCTGCCGGCCCCCTGCCCGCG
		CCAGACCTCAGCCTCCCGAGGCCACCGCTGCATCCCTGTGACTTCCCTACTCATGACAAGGATGCCAGGCACGCGCCAGCCCGTCCAGG
		CCTCCAGCTCCACCTGGCGAGGCTGGCCCATTGTACACAGGCGCCCCAGATGAGGGAGGGTCTCCCCCTCTCCTTGAAGGGCGGTAGTC
		TGGGGTCCTGAGTGCTGGGTGTGGGCTTGTCCCTCGTGGACAGAACCCAGGAGAGGGCTTCATCCACCAAGGAAGATTGCTTTGCAGGGTAC
		CCAGGTCCCGGGGGCTGTGCCACCCTCTGGGCACCCGGAGCCAATCGCAGAGGTACCCAGGTCCCGGGGGCTGTGCCACCCTCTGTGCA
		CCCAGAGCCAATCGCAGGGGACCCAGGTCCTGAGGTCCTGGGGGCCATGCCACCCTCTGGGCACCCGCAGCCAATAGAGTCACCCTTGG
		GAAGCTTATGCGGACCTGGGGCAGCACTCGCGTCCTGACCCCGGTGCCGGTCCCACAGTTCGAGCCAGGGCAGTCGTACGCGGGTGTGG
		TGCAGTACAGCCACAGCCAGATGCAGGAGCACGTGAGCCTGCGCAGCCCCAGCATCCGGAACGTGCAGGAGCTCAAGGAAGTGAGTGCCC
		CACGCGGCCAGGACCCTCCCACCCCTCGCCCCGACCGCTGTTCCCACGGCAGGTCGGCCCTGACCCCCTGATCCCAGGTGGGCTCGGCC
		CCGCGGCAGGCCTGGCCCCAACCGGCCCTTCCTGCCCTTTGCTATGCAGAGCCATCAAGAGCCTGCAGTGGATGGCGGGCGGCACCTTC
		ACGGGGGAGGCCCTGCAGTACACGCGGGACCAGCTGCTGCCGCCCAGCCCGAACAACCGCATCGCCCTGGTCATCACTGACGGGCGCT
		CAGACACTCAGAGGGACACCACACCGCTCAACGTGCTCTGCAGCCCCGGCATCCAGGTGGGGTGGCCACCCCCCAGGCTGCACCTGCCCC
		GCCTAGGGCGCCCCCGCCAGCCAGGGTGGCCTTGTCCCCAGAAAGACGAGGGCAGAGGCAGGCTGCGCCACACCGATACTGTCTGTCCCCA
		CAGGTGGTCTCCGTGGGCATCAAAGACGTGTTTGACTTCATCCCAGGCTCAGACCAGCTCAATGTCATTTCTTGCCAAGGCCTGGCACCAT
		CCCAGGGCCGGCCCGGCCTCTCGCTGGTCAAGGAGAACTATGCAGAGCTGCTGGAGGATGCCTTCCTGAAGAATGTCACCGCCCAGATCT
		GCATAGGTGCGCATGGGGCCACCCGGGCAGTCCCAGATCTGCGTAGGTGCCGCGGGGCCGCCCGGGCAGTCCCAGATCTGCGTAGGT
		GCACGCGGGGCCGCCCCGGGCAGTCCCAGATCTGCGTAGGTGCACGCGGGGCCGCCCAGGGCCGTCCCAGATCTGTGTAGGTGCGCGCA
		GGCGCCCAGGGCTGTCCCAGAGGCCTCCTCCCAGCTCACTGTTACCTCCAGGGGCACGGCCACCCTGTAGGTGCGCACGGGGCCGCCT
		GGGGCTGTCCCACAGGCATCCTCCTCCCGGCTCGCTGTGACTTCCGGGGGGCACGGCCACCCCCTGTGCTCCGGCCGGGAGGTCCTGTGACA
		TCTCCTTGCCGGGGTTATAGGTGGAGCAGTGGGCTCACACTGCACGGCTTTTCTCTTTACAGACAAGAAGTGTCCAGATTACACCTGCCCC
		AGTGAGTACCTCGGCGGCCGGGACACGTGGGGAGGAGGGCACCGTGGTTGGGGCGAGGGCTCTGAGAGGACGGGGCTCTGGGAGGAG
		GGCCTGGCGGTCACGAGAGTAGGTGCATGGCTCACTCCGGTGGCTGAGCACCACCCGTGCCGTGCCCTCTCTGGGGAGCTTAGACGCTCT
		CTGGCCGGCCCACTGCGGCTGCATCACCAGGGCCTCATGCTAACGGCTGCCCACCCCGCCCCGCAGTCACGTTCTCCTCCCCGGCTGAC
		ATCACCATCCTGCTGGACGGCTCCGCCAGCGTGGGCAGCCACAACTTTGACACCACCAAGCGCTTCGCCAAGCGCCTGGCCGAGCGCTT
		CCTCACAGCGGGCAGGACGGACCCCGCCCACGACGTGCGGGTGGCGGTGGTGCAGTACAGCGGCACGGGCCAGCAGCGCCCAGAGCG
		GGCGTCGCTGCAGTTCCTGCAGAACTACACGGCCCTGGCCAGTGCCGTCGATGCCATGGACTTTATCAACGACGCCACCGACGTCAACGA

SEQ ID NO	GENE NAME	SEQUENCE
		TGCCCTGGGCTATGTGACCCGCTTCTACCGCGAGGCCTCGTCCGGCGCTGCCAAGAAGAGGCTGCTGCTCTTCTCAGATGGCAACTCGCA
		GGGCGCCACCGCCCGCTGCCATCGAGAAGGCCGTGCAGGAAGCCCAGCGGGGCAGGCATCGAGATCTTCGTGGTGGTCGTGGGCCGCCAG
		GTGAATGAGCCCCCACATCCGCGTCCTGGTCACCGGCAAGACGGCCGAGTACGACGTGGCCTACGGCGAGAGCCACCTGTTCCGTGTCCC
		CAGCTACCAGGCCCTGCTCCGCGGTGTCTTCCACCAGACACAGTCTCCAGGAAGGTGGCGCTGGGCTAGCCCACCCTGCACGCCGGCACCA
		AACCCTGTCCTCCCACCCCTCCCCACTCATCACTAAACAGAGTAAAATGTGATGCGAATTTTCCCGACCAACCTGATTCGCTAGATTTTTTTT
		AAGGAAAAGCTTGGAAAGCCAGGACACAACGCTGCCTGCTTTGTGCAGGGTCCTCCGGGGCTCAGCCCTGAGTTGGCATCACCTGCG
		CAGGGCCCTCTGGGGCTCAGCCCTGAGCTAGTGTCACCTGCACAGGGCCCTCAGCCCTGAGCTGGCGTCACCTGTGCAGGG
		CCCTCTGGGGCTCAGCCCTGAGCTGGCCTCACCTGGGTTCCCCACCCCGGGCTCTCCTGCCCTGCCCGCCCTCCCTCCTGC
		CTGCGCAGCTCCTTCCCTAGGCACCTCTGTGCTGCATCCCACCAGCCTGAGCAAGACGCCCTCTCGGGGCCTGTGCCGCACTAGCCTCCC
		TCTCCTCTGTCCCCATAGCTGGTTTTTCCCACCAATCCCTCACCTAACAGTTACTTTACAATTAAACTCAAAGCAAGCTCTTCCTCAGCTTG
		GGGCAGCCATTGGCCTCTGTCTCGTTTTGGGAAACCAAGGTCAGGAGGCCGTTGCAGACATAAATCTCGGCGACTCGGCCCCGTCTCCTG
		AGGGTCCTGCTGGTGACCGGCCTGGACCCTACAGCCCTGGAGGCCGCTGCTGACCAGCACTGACCCCGACCTCAGAGAGTACT
		CGCAGGGGGCCGCTGGCTGCACTCAAGACCCTCGAGATTAACGGTGCTAACCCCGTCTGCTCCTCCCTCCCGCAGAGACTGGGGCCTGGAC
		TGGACATGAGAGCCCCTTGGTGCCACAGAGGGCTGTGTCTTACTAGAAACAACGCAAACCTCTCCTTCCTCAGAATAGTGATGTTCGAC
		GTTTTATCAAAGGCCCCCCTTTCTATGTTCATGTTAGTTTTGCTCCTTCTGTGTTTTTTTCTGAACCATATCCATGTTGCTGACTTTTCCAAATAA
		AGGTTTTCACTCCTCTCCCTGTGGTTATCTTCCCCACAAAGTAAAATCCTGCCGTGTGCCCCAAAGGAGCAGTCACAGGAGGTTGGGGGGC
		GTGTGCGTGCGTGCTCACTCCCAACCCCCATCACCACCAGTCCCAGGCCAGAACCAGGGCTGCCCTTGGCTACAGCTGTCCATCCATGCC
		CCTTATCTGCGTCTGCGTCGGTGACATGGAGACCATGCTGCACCTGTGGACAGAGAGGAGGCTGAGAAGGCAACACCCTGGGCTTTGGGGT
		CGGGAGCAGATCAGGCCTCAGTGGGCTGGGCCGGCCACATCCACCACCGAGGTCAACCACAGAGGCCGGCCACAGGTTCTAGGCTTGGTAC
		TGAAATACCCCCTGGGAGCTCGGAAGGGGAGTTGAGATACTGCAGGGCCCATAGGAAGAAGTCTTGGGAGGCTCCACCTTTGGGGCAGAG
		GAAGAAGTCTTGGGAGGCTCCACCTTTGGGGCAGAAGAAGGGCCGGAGGGCAGCGAGGGCATCATCCTCAAAAGAAAGAA
		GTTAGTGGCCCCTGAATCCCAGAATCCGGGGTGCACGGCTGTTCTGGGGGCCGCTAGGGGACTAAGAGGATCGGCCGAGGGCTGGGCT
		GGAGGAGGGCAGCAGGGATGGGGCGGCGAGGGTGAGGGTGGGGCTTCCTGAAGGCCTTCACCTGCGGGGAAAATGCTGGGGACCCCTGCATTTCCG
		GTGCCACAGGCAGGGACACGCCTCGCTCGATGCGTCACACCATGTGGCCACCCAGAGCTGCGGGAAAATGCTGGGGACCCCTGCATTTCCG
		TTTCAGGTGGCGAACAAGCGCCCCTCACAGAACTGCAGGTAGAGACGGGCCCGGGGCAGACGCAGTGAGGCGGTGGGCGGGGCCCGGG
		GCAGATGCAGTGAGGCGGTGGGCGGGGCCCGGGGGCAGAGGCAGTGGGCGGGGCCCGGGGCAGACGCAGTGAGGCGGTGG
		GCGGGGGCCCGGGGGCAGAGGCAGCGGGTGGTGGCCGGGGGCCCGGGGGCAGACGCAGTGAGGCGGTGGGCGGGGCCCGGGGTAGTCGCA
		GTAGGTGGTGGGGCGGGGCCCGGGGGCAGACGCAGTGAGGTGGTGGGCGGGGCCCGGGGCAGACGCAGTGAGGCGGTGGGAGGGGCCC
		GGGGCAGACGCAGTGAGGCGGTGGGCGGGGCCCGGGTCAGAGGCAACGGGTGGTGGGCGGGGCCCGGGGCAGACGCAGTGAGGCGG
		TGGGCGGGGCCCGGGGCAGATGCAGTGAGGCGGTGGGCGGGGCCCGGGGCAGATGCAGTGAGGCGGTGGGAGGGGCCCGGGGCAGA
		CGCAGTGAGGCGGTGGGCGGGGCCCGGGGCAGACGCAGTGAGGCGGTGGGCGGGGCCCGGGGCAGACGCAGTGAGGCAGTTGCCAG
		CCTCTCTCAGCTGCCTCATGGGATTCGCACTGCAGCTGCGGCCCTGGCGCGACAAGGGCTGGACTTGGCCAGCGGGACGGTCCCTCACG
		GCGCTGAGGCCCACACTCTGCGTGGAGCCTCCCCGTGCCCAGGCTACCCTGCAAGGTCCTCGGAGAGGCTTCCTCCAGCCCCAGCCCCC
		ACACAGCTCCGGCCCAGGCCCGCTCTTCCCCATCCCAGTTGCTTTGCGCTGTATACGGCCAGGTGACCCCGAGCCGGCCCTGAGCCCTC
		GTCCCGGCTTCCTCCCCTGTAAGCTGGGTGAAGGACTCCATGGCACCCACCTGAGAGGGTTGTGGCGAGGCCCAGGCCCCTCGTGCCCA
		CACGGCCGGCGGCCCATGCCTGGCAGGGGCTGGGAGGCTGGGGCGACCAGAGGGGAGCGGCCTGTCCTGGAGGAGGCCCAGGGA
		CCCTGGTGAGAGGGTCTCTCCCAAGTGCTCTCTATGGGACCCCCTTCCTCTGCGCCCGTCCTTCACGGACCTCTCCGGGTCACCCCTGGG
		CTGCACACTGGGTTCAGGGGGGCCTTGAGGTGGGGCCCCTGTTCCCAAGTCCCGGCGGGGTTTCTCCTGAACCTCAACCCATCCTCACCT
		GCGGGCATTCCCATCCCCCCAACGCCTGGGTCACCAGGATTCCAGGCAGGAGGGGCGGTGGGGGTTACCAAGGCCCGGGTTGCCATGCA
		GAACCCCCAGCCACCACGCAGACCCCCCACGGGGCCCAGGGAAGCTCCTGGTCTCACACTGCACCTCACACTTCCTGTGGGGGCAGACTC

SEQ ID NO	GENE NAME	SEQUENCE
		CAAGGTCCCGGCCTCTCATCTTGTAGAAACTGAGGCACAGGAGGGACACACACTCCCACGGCCGGTCACCGTGGCCCCCACACCTCCCAC
		TGGACTGACACCTGGCCAGGCTCCGGACACCCGTGGCACAGCCTCAGCCCCTGCGGCCCCCTGCTCCGTGGCCCCCAGGCCCCAGCTCC
		CATGTGCACGTCCTGCCTCAGGCCTGGAGGCCCCCTCGGCCCCCAAATAATCAGACAATTCAACAGCAAAACTACTTTTTTCAGGCTGGCAGG
		ACTCTGGGCAACCCCCCTGCAACAGCCCCCCTGCCCTATCACAGCCACCCTTGCCTCCCAGGCACGGAGACCCCCACCATCAGGTCCCAGCCT
		TGGTTCATCCCCAAGCACCCTGTGTGTTGGGATGGCGATGCTGGCTGAGCCCCCTGCATCC
253	chr21:46280500-46283000	AGGGCGTTTGGGAACACCCCTCCCGGAGGGGTGAGGCGGCCCAGCCTGCGGCTGCCAGAGGACACAGGTTCTGCTGCGGAACCTGCAG
		ACATGGCCATAACAGGCCACAGTGCTCGGGCCCACACAGCCTGGACCCATGGCCCTGTGTCACCTCCTCCTCAGGGGCAGGCTTCAGGGC
		CTCGACCCTAGAGGCTGCCCCTCGGTTCTGCTCCATGGACGGCGCAGGCAGGCCCAGGCCTGTGACGAGTTCACGGAAGCTCCAGGATG
		ACCCCCGCTCTGCGCCCTCCTCCAGCATTCCAGACCACAAACCACTCTGGGCTAAAACGAGGCATCGCCAGAGCATCCCACTTCCTCGGA
		AAGCTGCGGTCTGGGGACGCGTCTTGGCCCTGAAGAGGCTCCAGATGGCTCCCATCAGGCCTCTCCGCCTACGTGCGGCCGACATGGAG
		TGACAGAGCGTCGGGGACACAGAATTCAGAGCTGGGCCTGGGGCTGCTTTGAGATACTGATGGCTGCCAGGGGGCACAGAGACCCGTCC
		TGCAGACAGGGCTGTGAGGGCCACAGGGGGCCTCGGGGAGAGGCAGTGGGAGGGAGGACAGTGGGGGCCTCCAGCTGGGTGAGCAGC
		TGGAGCGAGGGGGGCCCGGGGCTTGTGATGGTGCTGCCGACCCTAGAGGTGCCGGCCCCACGATGGAGAGCACGTAGTGCCCCCCGGG
		AGTCAGGAGGCCGGGCCTGACCTCGGGGGCTGCAGCCAGGGGAGGCCGGCACCCCAGATAACCCCCAAAGAACTGCAGGCCCTGAGGC
		GAGGCCAGAGTGGGGGCGGGGCAGGTCCCAGCCGAGGAGGTGCTCCGTGCTGCCTCAGCAGAACCCATGATGGGCTGGCCCAAGGCT
		CTGAAGGTGGAAAGGCCTCACACATTCTGCCCCGGCTGACGCCTTCCTTGGGCCAGTGCTCGGGGGTGTGTAACAAACGCCAAGACGCAT
		TGTAAAGAAGGAAGCCTGCGTTTCCATCACCGGCTTAATATCAAACAAAAGTGCAATTTTGAAAATGTAGTCCAAGGTTTTCTGTGGTGCGG
		AAATGGCCAGGCCAGACCTCCGTGGGTGGTCCTTCGTGTCCACGTCAGCGCCCTACATCCACACTGTGGGCACCATGACCTCACATGCGG
		AGCGGAGCAGGGCCGGCGCCCGGAGAGCCAGGCTGGTCACGAACGAGGCCTAGAGGGCGTCAGGCCCCAAAGCACTCACAGGCTTCTC
		CTCTGTCCTCGGGGCCTTCAGACACCTGCATGCGCCGATTCAGCCACCCGCGCGCCGCCGATTCCCCTGGCCATGGGGTTTCCAAAGTGTG
		TGCTCAGAGGACAGTTTCCTCCAGGATGACCTGTCAGTGGCTCTCTGTGCCGGGGACGTCGCGTCTGAGGCCCCCGTGCTGATGCCACGACCGAGGGAA
		AACGATTTACCCAGTTCCTTTTCTCCACTCAGGAGGCGTTTGCTGAGAGGCACAGGCTGAGCCCCCGTGCTGATGCCACGACCGAGGGAA
		CGGGTCTCCCTGTCGGCGTGAACTGACCCGGCCAGGCGTCCACTGCCACTCGGACTGTCTCCCAGGCACGTGGCGCCCACACGGGCAGA
		ACACGCCCTCCACACACGCGGCTTCGGGCAGAACACGAGGCGCCCTCCACACACGCGGCTTCGGGGCTTGTCATGAAAAAAGCTGAATG
		CTGGGGGTGCAGCTTTCACCAACAGAATCCCGTTTGGAAGGGACGCGGTGAGACATGATCCACCCTAAGTTGTGATCCTGGGTGAGCCGC
		CGTCCACACCCTGCTGAGGGTCCCTTCACCCACTTTATTCTCCAGAAAACCCTGCCCATCAGGGCTGAGTCCCACGCCTTCCCTCTCCGTC
		CAGGCCTGGCTTTGACCTCTGGGGTCGTGTGGGGCACAGGGGACACCCTATCCAGGCAGAGGCCCTACGGCTATCTGGAGGAAGTGGTG
		GGAGCTGGGCTTCTGCCTGGAGGATGCACCCAGAGGGGTCACAGTCCACACACAGAGACACACGGGTGCCTTCCAGATGGCTGAGCCAGTC
		CAGCCCAGAAGGGCCTGGGGGTTGGGGGCTGCACCTGGCCTGTCCCCACCAGCAGGGCTCAGGGCTTCCCAAGGTGTGTGGGGGACGG
		GGCAGCACCTCTCAACCAGGTCACCTGAAACCCGAACTGAAAGGCATCCTAAGTTAAGACATTAACTCCCATTGTCAAGGTGCCATCGTCA
		ATTCTGTCTCCAAATCCTTCTTTGTTATTTCATGTATTCACAGAGTGACGCTCCGTGTTTCGTTCAGCCTGCAGGCCTGCAGAAGCTGCATCT
		CGGGATGGCCAAGAGCCCGGCCAGGCCCCACGGCTGCACCCAGGACGGGATTCATGCCCCATGCCTGGCTTCTCACGACCACACAGAGTGC
		CTTTCCCGGGACTGGATGGAGGCAGAGTGAGAGAAGAGCCTGGAGCAAGTGTTTTGGACCACACAGTGATCAAACACGGAGCCCCGTGGG
254	COL6A2	AAGAAAGGCCAGACCGGGCACGGTGGCTCACGCCTGTAATCCCAACACTTGGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGT
		TCGAGACCAGCCTGGCCAACAGGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAAATTAGCCGGGCGTGGTGGCAGGCACCTGTAATC
		CCAGCTAATCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGCTGCAGTGAGCTGAGATCGCGCCACTGCACTCCC
		AGCCTGGGTGAGGGAGCGAGACTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAAGAAAGGCCCCGGTGAGATGCTTTCTCTTAAAC
		ACGGCCCTGCACGTTGAGTTGCTGCCTCCTGTGGCCTATTTCACGTTTATGCAAAGTCGGGGCGCCTGATGCGGGGCTCACCCGCCACAAG
		CAGGGGTCCTGGTGCTGCTCATGGAAGGGGCCCTACCCAGCCCGCGGGGCACTGGCTGGGACGGGGGCTGCCCAGGTCCCGCCCAGGATC

SEQ ID NO	GENE NAME	SEQUENCE
		CAAACACCCAGCCCCCGCCCAGCGGCCCTTCCTGGCCTGCAGTGGAGGCTGTAATGGGCAGGGGTGGTGGGAATCCCAGCTCACAGGGC
		GCCTGCTCTTAGAAGGGCGGCATCTGGGTCCAGAGGTCAGAAACGTCAGATGCCCATCCCAGAAGTGGCGGGGA
255	COL6A2	GGGTGAATGAGTAGATGTATGGGTGAGTAGGTGGGTAGATGGAGTGGGTGGGCCGAGTGTGTGGTTAGATGATGGATGG
		CTGAATGGATGAGTGGGGGGATGGATGGGTGAGTGGGTTAGTATGTATGGATGGGTGGGTGGATGAATGGATGGGTGCATAAA
		GGATGGATGGGATGAATGAGTTAGTGGGTTGGCAGATGGATGGAGTGGGTGAGTCAGTGGATAGATGGATGGGTGGGTGGATAGAGGATGG
		ATGGTTGGGGTAGGTGATGGGTGGATAGATGGGTATGTGAGTGAGTGGGGGGGATGGGTAGGTGGGTGGATGGATGGTTAGGT
		GAATGAGTGGATGGACAGACGGACAGTGGGTGGATGGATGAGTGAACGGATGGACCGATGGATGAATGGGTGGGTGGGTAGAGGATGGA
		CGGACAGGTGAGTGGGTGGGTGGATAGATAGATGGTAAGTGAGTGGATAGATGGAGTGGGTGGACAGAGGATGGGTGGATGAATG
		GATGGGTTAGTGGGTGGCTGGGTGGATGATGGATGGGTGACTGGGTGGATGGATGGGTTAGTGGGTGGCTGGGTGGATAGA
		TGGATGGGTGATTGGGCGAATGGGCGAATGGGTGGATGGGTGGCGTGGAGTTGGTGGGTACATGATAATGGGGGTGGAATACCCATGGA
		TTGGAATGAGCTGTTTTGGCTGCTATTTCTGGGACACCCAGCTCTGCCAGGCCCCTACCCCTCTGGTGGGCCAGGCTCTGACGGTGGCCA
		CTCATGGCCTTTCTAGCTCTGGTGCCAGCATAGGGAAGGAGGAGGCACAGCCTTGTCTTACTCCTTGCACCTGTTAGCCCCCCCCCCGC
		CAAGGGAGGACCCGTGGTTGGGGACAGCACAGGGGGCCCTGCTGTGTGCAGGGACTGTCCCTGGGGCCACTGAAGCCCACCTGTTCTTG
		TTCCTTCTCAGGCGGGATCCTGGTCCCCCTGGTGAGCCAGGCCCTCGGGGGCCAAGAGGAGTCCCAGGACCCGAGGTAGGTTGGTGGCCA
		GTCCCCATGCCCTCCCCCCAACCTGCCAGGCCAACACACACCCAAGCCTCGTGGTTCTGCCCACGGTGGACCCACGTATCAGTGGGCAGT
		GGCCTGGGAGAGACTCAGCCACCCAGCCTTGGCCCCAGAGTCTCAGCCTCATCCTTCCTTCCCCAGGGTGAGCCCGGCCCCCCTGGAGA
		CCCCGGTCTCACGGTAGGTGTCACATGGGGCAGAACCAGTGTCCTTCTCCTGCCAAAACTAGACACCAAGAGCAGCAGGGGTGGGGGAA
		GGTCAGCTGGCACGGTCAGAGAGCAAGATCAGTGGAGGAGGTCAGAGAGGCAAGGTCAGAGAGCAAGCTTGGTTGGGGAAGGTCACAGG
		GCAAGGTTGGTGGGGGAGGAGGGTGGCAGCGAGGTTGGTAGGGACAGGACCCGCCAGCCTCCCCGCATGGCTGCCTCCACACGTGGG
		CTGGAATGTCCCGGGACCCCCAGGCCAGGACCTTGCTGTGGAAACTCTTCTGGGCCCCGGGGGACTACCCTGCCTGCCGTGTGCATT
		GCAGGAGTGTGACGTCATGACCTACGTGAGGGAGACCTGCGGGTGCTGCGGTGAGGCACTGCCCACGGCAGGGTCGGGGCCCATGCAC
		CGGGTGGAGGGCGGGAGTGCAGCAGGGCTGGGTCATCGCTGGGTCCTGCATGTGCACGTGACCCTAGGGTCTCGAGGTCTCCCCGGTAC
		CCCCCGATGACCCTGCCACCCCCCCCAGACTGTGAGAAGCGCTGTGGCGCCCTGGACGTGGTCTTCGTCATCGACAGCTCCGAGAGCATT
		GGGTACACCAACTTCACACTGGAGAAGAACTTCGTCATCAACGTGGTCAACAGGCTGGGTGCCATCGCTAAGGACCCCAAGTCCGAGACA
		GGTCAGCGGGGCAGGGGCGGGTGCAGCATTGCGGGGGGCCGGGCGGGCGTGGGAGGCGATGAGATGGGAGAAGTCCAGACGCGTCC
		CTCCAACGAGGGCCTCTGCATGGCTGGGGATGCCCCCAGACCCCGAGGCCTCTGGCAACGACCTCACGCGTGCGGCTTGCAGGGACGCGT
		GTGGGCGTGGTGCAGTACAGCCACGAGGGCACCTTTGAGGCCATCCAGCTGGACGACGAACGTATCGACTCCCTGTCGAGCTTCAAGGA
		GGCTGTCAAGAACCTCGAGTGGATTGCGGGGCGGCACCTGGACACCCTCAGCCCTCAAGTTTGCCTACGACCGCCTCATCAAGGAGAGCCG
		GCGCCAGAAGACACGTGTGTTTGCGGTGGTCATCACGGACGGGGCGCCACGACCCTCGGGACGATGACCTCAACTTGCGGGCGCTGTGCG
		ACCGCGACGTCACAGTGACGGCCATCGGCATCGGGGACATGTTCCACGAGAAGCACGAGAGTGAAAACCTCTACTCCATCGCCTGCGACA
		AGCCACACAGCAGGTGCGCAACATGACGCTGTTCTCCGACCTGGTCGCTGAGAAGTTCATCGATGACATGGAGGACGTCCTCTGCCCGGGTG
		AGCGTGTGGGCGCGGGGCAGTCGGCCGAGGAGCAGCAGGCCCCAGCCGCTGTCTAGCGTGAGCCCCCAGGGACACCCCTCACCTGAGGG
		ATGAATGTGCAGCCCAGGATCTTGGGCTGTGGGTGGGAAGGGGTCGGGCCCTCTCGGGGGCTGCAGGGCAGAGGCCAGCTGCACCCTGA
		GCCTGTCTAGGCAGATCAGTGAACGGCCGCTGAGGGTTCGCTAGGGACTGACCCTGGCCTGGCCCGGCCTCTCTCCTCTTCCAGACCC
		TCAGATCGTGTGCCCAGACCTTCCCTGCCAAACAGGTAATGCAGGGCACACCCTGAGCCACCACCCCAGACTAGCAAAGCAGCCCTGGTGTC
		CTTCCTCCTCGAGGGCCGGGCTGGGGGAGGGGGCCGTGCAGGGACCCGGGGGGCGGCGGAGCCACTGCGGAGGCTGCTCCTTAGGGAG
		ATGGCCCCAGGATGGCAGCACAGGGGAGGAGGGGCTTGGGGAAGGCAGGCTCCCAGGAACGCAGGAACACGCATCACGAGGCCATGAGG
		TGGGTGCTGCTAGCCTGGCGCTGTGCTCGGCATGTGGCCACTGGTCTTGAAGGCCCACCATGGGCCTTGCAGTCTCCCTCAGCTGCCGC
		CCAGCTCCCATGGGCTGGCCGTGCATGTGCCACTCGGAGGAAGCCCTGGATTCAGTGAGTGAAACCATCCCGGGGGTGGAAGCACTGACA

SEQ ID NO	GENE NAME	SEQUENCE
		CCCCCCAGCACCAGCAGGTCTTGCTCCAACCCTGGCCTGCCTCGGAGCTGCAGCTGCGGCTCTCACATCTCTGGGAGTGGGGAGCCCA
		TGTCCCGGATGTGGCCCACGTGGGTGTGAAGCTGGAGCTGGGGGTGCCGTCCAGGCTCTGCTGGACGTGGTGCTGCCCCCATGGTGCAC
		TGCTGCACCGTACCTGGGCCCACAGGAGGTCCCCGGGGGCGTTAGGAGCTGAGTCCCCCTCAGTGAGCCGTCCCCTCCAGGAGTGTGAG
		GGTAGGGATGCCATGGAGACAGGGTGGGAGGGTCCGACCTGGAGGACCACAGGGAGGAAACCTCAGGGTCTGCGGTACGAAGTCAGCG
		CTTCCTCAGCACGCGGGTCGCGGTTCCATGGGGAGCTCCCGGTGGGTGAGCTGGGCCACTGAGCACATTCACAGG
		CCCTGAGGCTGCCCCCAGGGGAGGAGCCGTGGACTCAGAGCCGAGGTTCCCCATACGTGCTGCCGACAGAGAACCTAGGGCTTGCACCTGG
		GTCTGGCTGCCCTTCAGCAGGCGGGCAGCCTCTGGCCCCACAACAGTGGGCTGTGCTTCTGCCGCCAAGGTGCAGGCGTCCTCCCCCAG
		GGTCCACATCAGCAGCAGGGGCACCTGGACCCTGAGGGCAGGAACCAGACCTTGGCTCCTCCACCCACCCCCTCGTTCCTGATGGGGCA
		GGGAAGTCTCGGGACCCCATGATGGCGACATGGTCACTGTGGGTGCTTTGCTATCAGGTGGGGGGCCTTCCTCTCCACTCTGG
		GTCCAGTGTGAGTGGCCGCTATGGCTTCCCCTCCACTCCAGGTTCTATCGTGAGTGGGTGGGTGCTGCGTCTGTGGATGTCACGTGACCT
		TTCCTCTTTAGCCTATCATTGTAGTTGGGAGTTAGTTAGCCCGTTGAGCGTCATTGAATTTCCAGTGTTGAGCCAGCCCTGCGTGCCCGGGA
		TAAACCCACCTGGCCGTGGTGTGTGGCCCTGTTTATGCACGTGGGCCCTGATTCGCTGATGCCTGCCTGAGGGTTTGCGCTTATCGGCGA
		CATCAGCCTGCACTTTTCTTTTCTCGTGATCTCTCTGGTTCTGGCCTCAGGGTGACGTGGGCCTCGTAGGGTCCTGTGGTGGCTCCTCCCC
		AGACGGTGACATGGAGTGAGCCCATTCTCCCTCCTGGGAGTGGGTCACTCAGGCCACCAGAGCACCACAGGGAAAGCAGCCAGGGAGGA
		CACGGAGAGGCCCCTTGAAGCTCTGGCCTCTTCTGAGGCCTCCAGGACCTGACAGTGAGTGGGAGCAGCCCTGGCAGAACCCCTCCCCTCCT
		CTCGGCCGCCCTGACACCTCATCCCCGACACTCAGAGCTCATCCTCCTTCCCAGCTGTTTCCAATTTCAAAGTGAACTCGACCTTGTGGCT
		CCAGGAGATGCAGCAGGGACAGTGTTAAATCGGCTTTCACCAGCCCACACGGCCAGGCATCCTCCTCGGCCCTCCTGGGCACTGGGTGG
		ACACCACTGGCTGTGGCCTGGCCCTTCTCCAGACAGCCCTGTCCACCCCCAAAGCCCAGCCACCCTGGGCCTGCAGCAGGCCCTGT
		GGAGTTCTCAGTTGCGTGGGGACCAGAGGGTGCTGGAGAAACAAACCAGACGCAGCTGAAGGCAGTCAGGGCAGGGCGCAATCAGCGAT
		AAGAGCTGCATAGGGGCCACAGCGTAACCTGAGCTCCAGTCGGTGGAAAGAAAAGGCAGAGACGTTGCAGAGGCCAGGTCTGCTCAGGG
		GAAGACAGTTCTGGGTGTAGAGGACTCACATCCCAGAGAGGCTGAGGAAGGGTTTACCACCGCAAGCTTTCTCAGGCGGGCTCTTGAGGG
		GTGGCTGGGGTCTTCCTGGCGACGGGCCTGCGGCACTGGAAGCCCTACTGGAGTTTGGCCTGTCTCCGGCACAGGTTTGGACGGAGCTG
		TTTTGTGCTGAAAGGTTTTCTCGGGGTCCGTGGTGTCCCCCAAAGGTGCCACCGTGCGGGTCTCCTAGCTCCCTGCCAGCTTCCTGTCCCT
		GTGCTCACTGCCCCCACGCCTCCTGCCAAGGCCGAGCCACACACCCGCTCCACCTGCATTTCCTCTACCGACTCGCCAGCCCAAATGCCG
		CTCTTCACTCTGGCCTCGCTGAGCGGCTGCCCGAGGAGGAGGCTCTAGGGCCGACGCCCACCGCAGGCCTTACAGTCTTCTCTGGACGCTCC
		CTTGCAGATGCACCGTGGCCTGGCGGCGAGCCCCCGGTCACCTTCCTCCGCACGGAAGAGGGGCCGGAGCGCCACCTTCCCCAGGACCAT
		TCCCCTGATCCAACAGTTGCTAAACGCCACGGAGCTCACGCAGGACCCGGCCGCCTACTCCCAGCTGGTGGCCGTGCTGGTCTACACCGC
		CGAGCGGGGCCAAGTTCGCCACCGGGGTAGAGCGGCAGGACTGGATGGAGCTGTTCATTGACACCTTTAAGCTGGTGCACAGGGACATCG
		TGGGGGACCCCGAGACCGCGCTGGCCCTCTGCTAAAGCCCGGGCACCCGCCAGCCGGGCTGGGCCCTCCCTGCCACACTAGCTTCCC
		AGGGCTGCCCCCGACAGGCTGGCTCTCAGTGGAGGCCAGAGATCTGGAATCGGGGTCAGCGGGGCTACAGTCCTTCCAGGGGCTCTGG
		GGCAGCTCCCAGCCTCTTCCCATGCTGGTGGCCACCGTGTCCCTTGCTGCGGCTGCATCTTCCAGTCTCTCCTCCGTCTTCCTGTGGCCG
		CTCTCTTTATAAGAACCCTGGTCATTGAATTTAAGGCCCACCCCAAGTCCAGAATGACCTCGCAAGACCCTTAACTCACTCCCGTCTGCAGA
		GTCCTTCTTTGCTGCATCAGGTCACCCTCACAGGCTCCAGGGTTTGGGTGTGGAAGTCTTTGGAGGCCCTTACTTAGCGGCCCAGCTGGG
		CTGCCGTGCGTCTGGGATGGGGCTGAGGGAGGGTGCTGCCCAGGTGCTGGAGGATGTTCCAGCACCAGGTTCCAGCGGAGCCTCGGAA
		ACAGGCCCCAGAGGCTGGTGAGCCTCGCTGGGTGTGGGCACTAATCCCGTGCATGGTGACTCGTGGGCGCTCACGGCCCACCTGGTGGC
		AGGTGAAGGCTTCCGGTTGGGCAGCAGATAGTCCTGGGGGAAGCTGGCAGTCCTGGCACCATGACGTATCTGGGCTGGTGTCATGCACA
		GTAGGGCGAATGGCCACAGCTGCCTGCCAGCAGCCCTGATCCCGGGGTGTCTGCACCCTTCCAGCCCAACCTCTGGGTCTCCAAAGCAC
		AGTCGGGGGAGCATCCACCAGGCACAACCTCTGCGGTCCTCAGAGGACTGAGCAGAGAATCCCAGGGTCCACAATGTTGGGAGCGGCA
		GGGATCACCATCCAAAGGAGCGGCCCCCACGGCGAGCTGACCCCGACGTTCTGACTGCAGGAGCCCTCATCCAGGCTGGGCTCCTGCC
		GGGCACGGCTGTGACCATTTCTCAGGGCCAGGTTCTCGTCCCCACACCCACTGCACAGGGCAGGCCAGGCTGGTCTTCCCACTGTGGGG

SEQ ID NO	GENE NAME	SEQUENCE
		ATGAAGGATCCTCCACAGGAGGAGAGCAGAGTCCACACAGACATCCCAACAGCCTCAGCCTCCCTGTGCCTGGCCGGCCCCCACAGCTT
		CCCCGTCTCCTCCCAGGCCCCCACAGACACTGATGAATGGACAGAGAGACCCCCAAAACCAGCTGCCCCTTGCATGTCTGTCTCCATATGTTTGG
		TGACAGCAGTGAAAATGTTATTAGTTTTGAGGGGGTTTGGGAAGCCCAGCGGTACCTGAGGAGTTTCTGGACATTTAAGCCGGTTCCTAGG
		TGTGGCCTTAACAGGGAGGCTGCCCTTCCTTTCACTGAATGAGCTGCGTCACTCACTAAGCTCACTGAGGGAACCCCATCTGCCAGCTCGTG
		CGTGCTCAGACGGCGTCCATGTCTCAAGCGTTCTGTGAAGGCTGCGGTGCAGCGTGAGGTCACCCTGCTGTTGTTCAGAGCTTTGCTCACT
		GCCTGCGGGGGCTGGACCGTTGCACCTCCAGGGCCCCCAGAAACCGAGTTTCGGGTCAGGGTCCTCTGTGTGCATTCCTGGGGGTCCATG
		TACCAGCTGTGACGACGTCCAGGGGTTGGGCTGAGAAGCAGACACCCTTGGGGAAACTGGCTCTGTCCCTCCCCTCCCCCATCCCAGGAG
		CTGAGGTCTTGGTGAGGCCACAGGGCCAGGTCCACGCAAGGACTGTCCGTGTCCTGTCTGGTCTCTGGCCCCACGTGACACCCACAC
		GTGTGGTAGGCAGCCTGGCCTGGGTTGTGGCTATGGCCAGGCCCCCAAGCTGTCCCCGATGCCCAGGGCTGGTGACCACCCAGGCAGGT
		GGGGGCCCCACTTGGTAACAGAGTCATAGGGCAGAACCCACCTGGGCTGCCACAGAAGGTCTGGCTGCCCACTGCTCCCCA
		CCATGGCCAATCAGAAGAGTCAGGGGCTCCTGGTCTTTCCGGGAGGGACGTGGCCCAGCCAGCTCTAGGTGTTCTGAGCAGCTCTGGGA
		CCCAGCGATTGAGGGGTCAGGCTGGGGGTGTCAGAGCCAGGGTCCTCCTTAAGTACCTCCCACACTACACAGACAGTGGCCCTTTTGTGG
		GCAGCAAATTCTTGAGCCATGAAAGGATGCTTTGGGCCCCTTCCCTCCCAGGAGGGCAGCCTGTGCAGGGATGGTGCTCAGCAGGTGGAC
		AGGGCCTGGGGCCTGTGTCAGGGTCTCAGGCCTGGGAGCACCAGCAGAGGAGATGGCGGCTCCCAGCAGTGCCGCCTGAAAGTGTCTTG
		GGCTAAGGACCCACACCCAGGGCTGCCCTGCAGAAACGCCCCCCGCAGAGCCCAGTGGTCTGTGAGGTTGCAGGCAGGGTGCGAATGGAA
		GGGCACAGGTGCGGGGGCTGGCACCTGCCCGGTCCTGCCCACCTCCCCTCGCCCAGCCCGCACCTGCGTCTCCCCACAGAGCTGTCCGT
		GGCACAGTGCACGCAGCGGCCCCGTGGACATCGTCTTCCTGCTGGACGGCTCCGAGCGGCTGGGTGAGCAGAACTTCCACAAGGCCCGGC
		GCTTCGTGGAGCAGGTGGCCGCGGCTGACGCTGGCCCGGAGGGACGACGCCCTCTCAACGCACGCGTGGCGCTGCTGCAGTTTGG
		TGGCCCCCGGCGAGCAGGTGGCCTTCCCGCTGAGCCACAACCTCACGGCCATCCCACGAGGCGCTGGAGACCACACAATACCTGAACT
		CCTTCTCGCACGTGGGCGCAGGCGTGGTGCACGCCATCAATGCCATCGTGCGCAGCCCCGCGTGGCGGGGCCCGGAGGCACGCAGAGCT
		GTCCTTCGTGTTCCTCACGGACGGCGTCACGGGCAACGACAGTCTGCACGAGTCCGGCGCACTCCATGCGCAAGCAGAACGTGGTACCCA
		CCGTGCTGGCCTTGGGCAGCGACGTGGACATGGACGTGCTCACCACGCTCAGCCTGGGTGACCGCGCCGGCCGTGTTCCACGAGAAGGAC
		TATGACAGCCTGGCGCAACCCGGCTTCTTCGACCGCTTCATCCGCTGGATCTGCTAGCGCCGGCCCGGGCCCCGCAGTCGAGGGTCG
		TGAGCCCACCCCGTCCATGGTGCTAAGCGGGCCCGGGTCCCACACGGCCAGCACCGCTGCTCACTCGGACGACGCCCTCCGCAGGCTGCCTGCAC
		CTCTCCAGCTCCTCCCACGGGGTCCCCGGCCCCCGCCCAGGTCTCCCCAGGCCCTCCGCAGGCTGCCCGGCCTC
		CCTCCCCCTGCAGCCATCCCAAGGCTCCTGACCTACCTGGCCCCCTGAGCTCTGGAGCAAGCCCTGACCCAATAAAGGCTTTGAACCCATT
		GCGTGCCTGCTTGCGAGCTTCTGTGCGCAGGAGAGACCTCAAAGGTGTCTTGTGGCCAGGAGGGAAACACTGCAGCTGTCGCTCGCCCA
		CCAGGGTCAATGGCTCCCCCGGGCCCAGCCCTGACCTCCTAGGACATCAACTGCAGGTGCTGGCTGACCCCGCCTGTGCAGACCCCACA
		GCCTTGATCAGCAAACTCTCCCTCCAGCCCAGGCCCAAAGTGCTCTAAGAAGTGTCACCATGGCTGAGGGTCTTCTGTGGGTGGA
		CGCATGATTAACACTAGACGGGGAGACAGCAGGTGCTGAGCCTGTTGTGTTCTGTGTGGAGATCTCAGTGAGTTTTTGCTGTTCAGACCCC
		AGGGTCCTTCAGGCTCAGGCTCAGGAGCCCCACAGTGAACCAGAGGCTCCACAGGCAGGTGCTGACCTGACAGGAGTGGGCTTGGTGGCC
		ATCACAGGGCACCACAGACACAGCTTGAACAACTACCAGTATCGGCCACAGGCCTGGAGGCATCAGCCGGGCCATGCTTCCTCTGGAGGG
		CTAGAGGAGGACTAGAGAAGGGCCTGCCCCGGCCTCTCCCCAGCATCCCAGGGTTCCTGATCTCCTGGATAAGGATACAAGTCACCACAC
		TGGACTGGGGCTCAGCCTGCTCTAGAATACCTCACCTAAGTCACAGTGGACCAGGCTCAGCCTGCTCTAAGGTGAGCTTACCCGAGACACT
		GGACCAGAGATCAGCCTATCCTGGGATAAGCTCACCCGAGTCACACTGGACCAGGGCTCAGCCTATTCCGGGATGAGCTCACCCGAGTC
256	C21orf56	GACACTTCCATGACTGCAGCTGACCAGTCCACCTGCCAGCGGTTGACCACTCCCACTTCGCCAGCGACCGAAGGGGAGGGGAGGGGCCT
		CACCTGAGGGGCAACAGCAGAACCCACCACCTGGTCTTGCTTTACTCAGACCTGAGGGTGTGAAAGGTGCCCGTGACCTCCCGCATCAGGG
		AGCTGGCCGCCACCCTCGACTCCCGGGGAGCAGGCGTCCCGCGACCCCCCTCATCTACCAGGCCATCTGAGCTGGGCGGCGCCTCACCTC
		CGCTCCCGGGGGAGCCGGCCTCAGGGTAGGCATGCGCCCTGGGTGGGAGCAGGTCGTGGCCGCCGCCCTCCTGGCAGCTCTGGCTGAG

SEQ ID NO	GENE NAME	SEQUENCE
		CAGCCGCCGCAGCATCTGATTCTCCTTCAGGAGGCGCACCTGCTTCTTCAGGTCCGCGTTCTCGCTCAGGAGCCGGCTCATCAGCTCGCC GCCTTCAGCCATGGCGGGTGTCCCTCCCTTGTCCCTCACGGCTCCTGCAGCCCCATGGAGGTGGGAGCCCAGAGCCCGCAGGCACCAC AGAAACAGCCCAGGCACGGAGTTCCGTAGCCACCACCGCCTTCCACGCCTTGTGATGTCACTGCCCTAGTGATGAGGTGCCCAGCACCCT GCCTGCCCCCCGCGATGGCTCATGGCCCCGTTGAGGCAGTGAAGCTGGAGGCCCGTGGCGTGCACAGGCAGCCACTCCCACATTATGACC AGGGCCCGAGAATGCCAAGGACATTAGGCAGCTACGGGATGTGTAGCGACTGTACTCCAAGCCACTCCCCATTGA
257	C21orf57	AGGTGGAGGTTGCAGTGAGCCCTCCTCCCCTCCTTCCCACCTCCCATGCCCCCCTTTCTTCCTCCCACTCCCCTCCCGAG GCCCCGCTTATTCTCCCGGCCTGTGGCGGTTCGTGCACTCGCTGAGCTCAGGTTCTGGTGAAGGTGCCCGGAGCCGGGTCCCGCCTTCG GCCTGAGCTAGAGCCGCGCGGCGGCCGGCTTCCCCCCAAACCCTGTGGGAGGGGCATCCCGAGGAGGCGACCCCAGAGAGTGGGGCG CGGACACCTTCCCTGGGGAGAGGCCAG
258	C21orf57	CCTTCCAGATGTTCCAGAAGGAGAAGGCGGTGCTGGACGAGCTGGGCCGACGCACGGGGACCCGGCTGCAGCCCCTGACCCGGGGCCT CTTCGGAGGGAGCTGAGGGCCGCGTTCCTTCTGAAAGCGCGGGAGAGGGGTGGAGGCTGCGGGGAGCCGGGTCGCACACGAATA AATAACGAATGAACGTACGAGGGGAACCTCCTCTTATTTCCTTCACGTTGCATCGGGTATATTTTCGTTATTGTAAATAAAACGGTTCCGAGCC GTGGCATCGAGAGGGCGTCTGGAGTTCAGGGAACGCGTGGCCCCCCGCCCGGGAGCACCGCGCTCGCCTCTCGCCCTTCAAGGG GGTCCCCTGCCCCGGAGCCTGCGCCCCCGGAGAGGAAGGGGCTTGGGTGCCCGCAGCGCGTCCTTCCGTAGAAAAGGCTTGC GTCAGTATTTCCTGCTTTTACCTCCTGAG
259	C21orf57	CAGTATTTCCTGCTTTTACCTCCTGAGTATTGGAATATTCGAGTAAACCCTGGAGTTTCAGCGCCAGCGCCACGCTCTTCATCAGGGCAGCG CGTCGCGAGCGCGCTGGTTCCCCGGGCCTCCCGGCCACGGACACCGCTCTAGCCAGGGCCACGGCGAGGCCGCCGAGCAGCACCTCA GAGACCTGCGTGAGTTCTAAGCCTGGGGCTACTACAATTCTGCTCATCTGTTTGTCCTGTGAAATGATTCAGGGACATGAAAATGCCTTCC CACTGACTTGCGTCCTGTCTTAGCCTGGACTTGTCCCCTTGGGAACACGGGCCCAGGCCCCCTCTGTTCCTGAAGT
260	C21orf58	ATGTCTGCAGGGAAGAAGCAGGGGGACCCTGAATAAAGTTTCCGTTTTTTCCTATTTGTTAAAGTGATAGAGCATTATAGGACCAGAGAACAG GTGTGTCTGTACACTGTGCAGGTCCCCGGGGCAGGCTCTGAGTCCGTCTGCACACGGTGCGGGTCCCCGGGGCGCGCCCTGAGCCCGT CTGCACACGGTGCGGGTCCCCGGGGCGCGCCCTGAGCCCGTCTGCACACACGGTGCGGGTCCCCGGGGCGCGCCCTGAGCCCGTCTGCA CACGGTGCGGGTCCCCGGGGCGCGCCCTGAGCCCGTCTGCACACGGTGCGGGTCCCCGGGGCGCGCCCTGAGCCCGTCTGCACACGG TGCGGGTCCCCGGGGCGCGCCCTGAGCCCGTCTGTACACGGTGCGGGTCCCCGGGGCGCGCCCTGAGTCTCTACTAAAAATACAAAAAT TAGCCAGGCGTGGTGGTTCAAGCCTGTAATCCCAGCTCCTTGGGAGG
261	PRMT2	CATACATGGTTATTAGAAAAGGCATCTCATCCAAATGTGGTGGCTCGTGCTTGTAATCCCAGTGCTTCAGGAGGCCAAGGGAGGAGGATTA CTTGAGCCTAAGAGTTTGAGACCAGCCTGGGCAACACACAAGACCTTGCCTCTACAAAAAACTTAAAAACTAGCTGGGTATGATGGTGCAC ACCTGTAGTCCCAGCTACTTGGGAGGCGGGGCAGATCGCCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGATGAAACC CCGTCTCTACTAAAAATACAAAAATTAGCCGAGTGTGGTGGTGCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACT TGAACCCGGGAGGCGGAGGTTGCCATGAGCCGAGATCACGTCACTGCACTCCAGCCTGGGTGACAGAGCACAAAAGACAGGCATGACTTT GTACTTAACTGCTCAGCTTTGTAATCACTGGGGGCCCAGATGCTCACTTGGATTCTAACTTTGTTGGCATCTGGGCCTAAAAGCCGTGATGC AGGTGAGCAATGATGCAGAGGGCTCTGTGCGCCTGGCGGGCTCTGTTTGCCTGCTGGGCTCTGTGCGCCTGCTGGGCTCTGTGCGCCCG GGAAGGTGCGGCCACCCTCACGCGGAAGGCGGCCAGCGGATCCCGGTGCGCGCAGCTCCCAGCGCTGGGGTTCCAGCGCCCCGCCTCT TCCTATAGCAACCAGCGGGACCTGCCGTCCCCGGGGCACCCCGAGGGGTCTGCGCCCGCTTCTTTCCGAAACGGGAAGGCGCTGGGG GCTCGGCAGCCAGAGGGACGGGTTCAGGGAGCGTCCGGTGAGCCTAAGACGCCGCCTTTGCCGGGGTTGTCTGCCTCTCACTT AGGTATTAGGAACCGTGGCACAAATCTGTAGGTTTTCCTCTGGGGGTGGGCGGAGGCTCCAAACCGGACGGTTTTCTCCTGGAGGACTGT

SEQ ID NO	GENE NAME	SEQUENCE
		GTTCAGACAGATACTGGTTTCCTTATCCGCAGGTGTGCGCGGCGCTCGCAAGTGGTCAGCATAACGCCGGGCGAATTCGGAAAGCCCGTG CGTCCGTGGACGACCCACTTGGAAGGAGTTGGGAGAAGTCCTTGTTCCCACGCGCGGACGCTTCCCTCCGTGTGTCCTTCGAGCCACAAA AAGCCCAGACCCTAACCCGCTCCTTTCTCCCGCCGCGTCCATGCAGAACTCCGCCGTTCCTGGGAGGGGAAGCCCGCGAGGCGTCGGGA GAGGCACGTCCTCCGTGAGCAAAGAGCTCCTCCGAGCGCGCGGCGGGACGCTGGGCCGACACAGGGGACCGCGGGGGCAGGGCGGAGA GGACCCGCCCTCGAGTCGGCCCAGCCCTAACACTCAGGACCGCCTCCAGCCGGAGGTCTGCGCCCTTCTGAGGACCCTGCCTGGGGGA GCTTATTGCGGTTCTTTTGCAAATACCCGCTGCGCTTGGACGGAGGAAGCGCCCACGCGTCGACCCCGGAAACGAAGGCCTCCCTGATGG GAACGCATGCGTCCAGGAGCCTTTATTTACTCTTAATTCTGCCCGATGCTTGTACGTGTGTGAAATGCTTCAGATGCTTTTGGGAGCGAGGT GTTACATAAATCATGGAAATGCCTCCTGGTCTCACCACACCCAGGGTGACAGCTGAGATGCGGCTTCTCCAGGGTGGAGCCTCCTCGTTTT CCAGAGCTGCTTGTTGAAGTCTTCCCAGGGCCCCTGACTTGCACTGGAAACTGCTCACCTTGGCATCGGGATGTGGAGCAAGAAATGCTTT TGTTTTCATTCATCCTAGTGTTCATAAAATGGAAAACAAATAAGGACATACAAAAACATTAATAAAATAAATTAATGGAACTAGATTTTTCAGAA AGCACAACAAACACAAAATCCAAGTATTGCCATGTCAGCAACACATTCCTACTTTAAGTTTTATGAAGTTAATTGGAGTAGTGGAGAACAAAA GTGGATGTGGGGCAG

Example 14: Fetal DNA quantification using massively parallel shotgun sequencing

In this example, fetal-specific DNA methylation markers were utilized to quantify the fraction of circulating cell-free fetal DNA in maternal plasma, using a massively parallel shotgun sequencing (MPSS) platform. For this Example, four types of DNA markers were assayed: 1) fetal-specific methylation markers which allowed selective enrichment and subsequent quantification of fetal DNA (e.g., SOX14, TBX), 2) Y-chromosome markers which confirmed fetal DNA quantification (for samples with a male fetus; e.g., SRY1, SRY2, UTY), 3) total markers avoid of restriction sites which were used to quantify total cell-free DNA, including fetal and maternal DNA (e.g., ALB, APOE, RNAseP, and 4) digestion control markers which monitored the completeness of restriction digestion and hence the accuracy of methylation marker-based fetal quantification (e.g., LDHA, POP5).

Methylation-specific restriction digestion

Fetal methylation DNA markers were enriched by selective digestion of unmethylated maternal DNA, using methylation-sensitive restriction enzymes. Digestion was performed according to the parameters specified in Table 7 below.

TABLE 7

TABLE 7: Methylation-specific restriction digestion

Reagent	**Concentration in reaction**	**Reagent Volume (µL) for n=1**
H2O	N/A	16.7
10x PCR Buffer (20mM MgCl2, Roche)	1	3.5
25 mM MgCl2 (Roche)	2	2.8
ExoI [U/µl] (NEB)	0.2857	0.5
HhaI [U/µl] (NEB)	0.2857	0.5
HpaII [U/µl] (NEB)	1.4285	1
DNA [µl]		**10**
Final Vol:		**35**

TABLE 7: Methylation-specific restriction digestion		
	Reaction conditions:	
Digestion	41ºC 60'	
Inactivation	98ºC 10'	

Competitive PCR

The digested samples were amplified by PCR together with known copy numbers of competitor oligonucleotides. The competitors were synthetic oligonucleotides having the same nucleotide sequences as the target DNA, except for one base difference at the synthetic target site, which differentiated the target DNA from the competitor. Competitive PCR using target-specific primers allowed for independent quantification of each marker. Competitive PCR was performed according to the parameters specified in Table 8 below.

TABLE 8

TABLE 8: PCR amplification		
Reagent	**Concentration in reaction**	**Reagent Volume (µL) for n=1**
Water, HPLC grade	N/A	6.64
10x PCR Buffer (20mM MgCL2, Roche)	1x (2 mM MgCl2)	1.5
25 mM MgCl2 (Roche)	2 mM	1.2
dNTPs (25 mM, Roche)	500 µM	1
PCR primer (1 uM each)	0.1 µM	5
FASTSTART PCR Enzyme (5U/µl, Roche)	0.1 U/µl	1
Competitor MIX (8000/800c/ul)(1:0.1c/ul)		0.38
DNA (from restriction digestion)		35
Total		50

TABLE 8: PCR amplification		
	PCR Cycling conditions:	
	95°C, 5 min	
	95°C, 45 sec	35 cycles
	60°C, 30 sec	
	72°C, 45 sec	
	72°C, 3 min	
	4°C hold	

Adaptor oligonucleotide ligation

Illumina adaptor oligonucleotides (TRUSEQ adaptors) were ligated to the amplicons generated in the competitive PCR described above. The adaptor-ligated amplicons were subsequently sequenced using the Illumina HISEQ 2000 platform (Illumina, San Diego CA). Two different ligation-based approaches were used to flank the amplicons with the adaptors. The ligation procedure was optimized to maximize the amount of double ligation products (i.e., adaptor oligonucleotides ligated to both ends of the amplicon), and minimize single ligation and/or empty ligation (i.e., two adaptor oligonucleotides ligate to each other without amplicon insertion).

Direct ligation of adaptors

To render the PCR amplicons compatible for MPSS, the amplicons (which had 3' adenine (A) overhangs generated by Taq polymerase during the PCR reaction) were ligated to adaptor oligonucleotides having 3' thymine (T) overhangs (see FIG. 158). Prior to the ligation reaction, AMPURE XP beads at 2-fold volume of PCR reaction volume were used to remove single-stranded primers and amplicons generated by asymmetric PCR. Cleaned amplicons were quantified by Agilent Bioanalyzer and mixed with Illumina TRUSEQ library adaptors at an 8:1 ratio. 2 µL of T4 DNA ligase (Enzymatics) and 17.5 µL of 2x ligase buffer (Enzymatics) were added, and the ligation reaction was carried out at room temperature for 15 minutes.

Unidirectional adaptor ligation

In some cases, a modified protocol to improve ligation efficiency and to ensure unidirectional ligation was used. Single base overhang ligation can be less efficient compared to ligation of longer cohesive ends. Additionally, using single base overhang ligation, PCR amplicons can ligate with Illumina TRUSEQ adaptors in either orientation such that, when the ligated product were sequenced, only about half of the sequence reads covered the target sites for copy number calculation. Modifications of the ligation procedure were thus developed to overcome such limitations. First, tag sequences that were 5 nucleotides long were designed to replace the original tag sequence (10 nucleotides long) in the PCR primers (for the competitive PCR above; provided in Table 9 below). The tags were of different sequences for reverse or forward PCR primers and each had a deoxyuridine at the junction between tag sequence and target-specific sequence. The modified primers were used at equal molar ratio in the competitive PCR reaction above.

After PCR amplification, the tags were cleaved from the amplicons by uracil N-glycosylase (UNG; UDG) and EndoVIII digestion, creating a 5 base overhang that selectively ligated the PCR amplicon to universal or indexed adaptors (provided in Table 9 below) with high efficiency (see FIG. 159). Specifically, 1 µL UDG (5 U/µL, NEB) and 5 µL EndoVIII (10 U/µL, NEB) were added to each reaction and incubated at 37°C for 30 minutes. The reaction was stopped by heating at 95°C for 10 minutes to inactivate UDG, after which it was gradually cooled to 25°C. The amplicons were cleaned by AMPURE XP beads prior to the ligation reaction.

Pre-annealed index adaptor and index-linker was prepared by mixing at equal molar ratio, heating to 95°C for 5 minutes, and gradually cooled to 25°C. Universal adaptor and pre-annealed index adaptor at equal molar ratio were mixed with the UDG/EndoVIII-digested PCR amplicons (having 5 nucleotide overhangs). The ratio of adaptor to amplicon varied from 8:1 to 2:1. 2 µL of T4 DNA ligase (Enzymatics) and 17.5 µL of 2x ligase buffer (Enzymatics) were added, and the ligation reaction was carried out at room temperature for 15 minutes.

For both ligation approaches, the ligated product (5 µL) was amplified using Illumina TRUSEQ PCR mixture and primers as specified in Table 10 below. Amplified libraries were purified using AMPURE XP beads to remove free primers/adaptors and DNA fragments of smaller size.

TABLE 9

Table 9: Primer and adaptor sequences		
Target	**Forward_Primer (SEQ ID NOS 338-350, respectively, in order of appearance)**	**Reverse_Primer (SEQ ID NOS 351-363, respectively, in order of appearance)**
ALB	TAGCUGCGTAGCAACCTGTTACATATT	GATCUATACTGAGCAAAGGCAATCAAC
APOE	TAGCUCAGTTTCTCCTTCCCCAGAC	GATCUGAATGTGACCAGCAACGCAG
RNAseP	TAGCUGGTCAGCTCTTCCCTTCATC	GATCUCCTCCCACATGTAATGTGTTG
CDC42EP1	TAGCUAGCTGGTGCGGAGGGTGGG	GATCUATGGGGGAGATGGCCGGTGGA
LDHA	TAGCUGGCCTTTGCAACAAGGATCAC	GATCUCGCAATACTAGAAACCAGGGC
MGC15523	TAGCUTCTGGTGACCCCCGCGCTTC	GATCUCATCTCTGGGTGCGCCTTG
POP5	TAGCUCCCTCCACATCCCGCCATC	GATCUCAGCCGCCTGCTCCATCG
SOX14	TAGCUACGGAATCCCGGCTCTGTG	GATCUCCTTCCTAGTGTGAGAACCG
SPN	TAGCUGGCCCTGCTGGCGGTCATA	GATCUTGCTCAGCACGAGGGCCCCA
SRY1	TAGCUAGCAACGGGACCGCTACAG	GATCUTCTAGGTAGGTCTTTGTAGCC
SRY2	TAGCUTAAGTTTCGAACTCTGGCACC	GATCUGAAGCATATGATTGCATTGTCAA
TBX3	TAGCUCTCCTCTTTGTCTCTGCGTG	GATCUTTAATCACCCAGCGCATGGC
UTY	TAGCUTGATGCCCGATGCCGCCCTT	GATCUGTCTGTGCTGGGTGTTTTTGC
Adaptors (SEQ ID NOS 364-366, respectively, in order of appearance)		
Universal_adaptor	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	
Index_linker	GCTCTTCCGATCTATAGCT	
Index_adaptor	5'phos/ GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTATGCCGTCTTCTGCTTG	

TABLE 10

TABLE 10: PCR amplification of ligation products	
Reagent	**Reagent Volume (µL) for n=1**
Water, HPLC grade	11
TRUSEQ PCR master mix	20
TRUSEQ PCR primers	4

Ligation product	5
Total	40
PCR Cycling conditions	
98°C, 5 min	
98°C, 10 sec	10 cycles
65°C, 30 sec	
72°C, 30 sec	
72°C, 3 min	
4°C hold	

Amplified libraries were retained on an Illumina flow cell and bridge amplified to generate clusters for subsequent sequencing on Illumina's HISEQ 2000. Use of indexed adaptors allowed for sequencing of multiple samples in a single lane on the flow cell.

Nucleotide sequence read analysis and fetal DNA quantification

Nucleotide sequence reads were analyzed and used to calculate copy number of individual markers and fetal percentage. 50 base pair (bp) nucleotide sequence reads were uniquely aligned to expected chromosome positions, allowing up to 5 mismatches outside the target sites/synthetic target sites. Reads having quality score greater than 13 at the target site with expected target DNA or competitor alleles were used to calculate the copy number of each marker. Specifically, the following formula was used:

$$Copy(DNA) = Copy(comp) \times \frac{Read\ Counts(expected\ DNA\ allele)}{Read\ Counts(expected\ comp\ allele)}$$

Fetal DNA, Y-chromosome DNA and total DNA copy numbers were represented by the mean value of methylation markers, Y-markers and total DNA markers, respectively. Fetal percentage was calculated according to the following formulas:

$$Fetal\ Fraction(methyl) = \frac{mean\ copy\ number(methylation\ markers)}{mean\ copy\ number(total\ markers)}$$

and

$$Fetal\ Fraction(Y) = 2 \times \frac{mean\ copy\ number(Y\ markers)}{mean\ copy\ number(total\ markers)}$$

Digestion efficiency was calculated by

$$digestion\ efficiency = 1 - \frac{mean\ copy\ number(digestion\ markers)}{mean\ copy\ number(total\ markers)}$$

Results

The fetal DNA quantification method using MPSS described in this Example was applied to ccfDNA extracted from 48 plasma samples from pregnant women. The results were compared to those obtained from another method that used mass spectrometry (e.g., MASSARRAY) as a detection method instead of MPSS. The results from both methods were highly correlated (see FIGS. 160 and 161). With exception of digestion markers (LDHA and POP5, which were detected at higher levels by the MPSS method), the R^2 values were in the range of 0.965-0.998. The fetal fractions derived from methylation markers also were highly correlated between MPSS and mass spectrometry methods (see FIG. 162).

Example 15: SNP allele frequency based method for fetal fraction quantification

In this example, single nucleotide polymorphism (SNP) markers were utilized to detect and quantify circulating cell-free (CCF) fetal DNA in maternal plasma (i.e. fetal fraction). In some cases, fetal fraction was determined by measuring single nucleotide polymorphism alleles using a single tube multiplex PCR for amplicon sequencing via massively parallel shotgun sequencing (MPSS). Advantages of this methodology include, for example: 1) the ability to detect CCF fraction of DNA from both male and female fetuses without prior knowledge of maternal or paternal SNP genotypes; 2) a simplified workflow that generates MPSS ready products without the need for traditional library generation and 3) an ability to perform MPSS fetal fraction quantification on samples multiplexed with genomic libraries on the same flow cell lane.

Materials and methods

CCF DNA was extracted from 4 mL plasma from 46 pregnant women using QIAAMP Circulating Nucleic Acid kit in an elution volume of 55µl. DNA also was extracted from maternal buffy coat

samples for confirmation of maternal genotypes. Gestational age at collection ranged from 10-17 weeks. Maternal age ranged from 18-42 years. Ethnic background of samples included African American, Asian, Caucasian and Hispanic ethnicities. 15µl of CCF DNA underwent PCR for each SNP panel using a single tube multiplex of forward and reverse PCR primers that included adapter sequences to allow secondary amplification with universal PCR primers designed to incorporate index tags. Amplicon libraries with index tags were clustered on the cBOT and sequenced on the HiSeq 2000 for 36 cycles or 27 cycles to generate amplicon sequence reads and 7 cycles to determine the index tag sequence. Reads were aligned to the human genome (hg19) and matched read counts for expected SNP alleles were used to calculate the allele ratio of each SNP within each CCF DNA. 15µl of CCF DNA also was used for quantification of fetal fraction by fetal specific methylation patterns for comparison with SNP based quantification.

Detection of paternally inherited alleles

CCF fetal DNA in maternal plasma contains both maternally and paternally inherited DNA (e.g., SNP alleles). Detection of paternal SNP alleles not present in the maternal genome can allow confirmation of the presence of fetal DNA. Additionally, quantification of paternal:maternal SNP allele ratios can provide for a determination of fetal DNA fraction in maternal plasma. The likelihood of detecting a paternally inherited allele at a single locus is dependent upon allele frequency and individual inheritance patterns. FIG. 163, for example, provides a summary of expected genotypes and the associated population frequency of each genotype based a SNP having a minor allele population frequency of 0.4. A SNP with a high minor allele frequency may increase the chance that paternal and maternal alleles will differ at a given SNP locus. Provided enough SNPs are interrogated, a high probability can be established that the fetus will contain some paternal alleles that differ from the maternal alleles. Thus, use of multiple SNP alleles increases the likelihood of informative fetal and maternal genotype combinations. Often, no prior knowledge of the paternal genotypes is required because paternal alleles can be inferred by the presence of non-maternal alleles in the maternal/fetal cell free DNA mixture. FIGS. 164 and 165 show how fetal fraction can be calculated using SNP allele frequency.

SNP panels

High minor allele frequency SNPs that contain only 2 known alleles were identified. Two panels of SNPs were generated: a 67 SNP panel (SNP panel 1) and an 86 SNP panel (SNP panel 2).

Individual SNP identifiers for each panel are provided in Table 11A and Table 12A below. Tables 11B and 12B include chromosome identity for each SNP.

TABLE 11A

TABLE 11A: SNP Panel 1				
rs10413687	rs2001778	rs4453265	rs539344	rs7176924
rs10949838	rs2323659	rs447247	rs551372	rs7525374
rs1115649	rs2427099	rs4745577	rs567681	rs870429
rs11207002	rs243992	rs484312	rs585487	rs949312
rs11632601	rs251344	rs499946	rs600933	rs9563831
rs11971741	rs254264	rs500090	rs619208	rs970022
rs12660563	rs2827530	rs500399	rs622994	rs985462
rs13155942	rs290387	rs505349	rs639298	
rs1444647	rs321949	rs505662	rs642449	
rs1572801	rs348971	rs516084	rs6700732	
rs17773922	rs390316	rs517316	rs677866	
rs1797700	rs3944117	rs517914	rs683922	
rs1921681	rs425002	rs522810	rs686851	
rs1958312	rs432586	rs531423	rs6941942	
rs196008	rs444016	rs537330	rs7045684	

5

TABLE 11B

TABLE 11B: SNP Panel 1					
SNP_ID	**Chromosome**	**SNP_ID**	**Chromosome**	**SNP_ID**	**Chromosome**
rs10413687	chr19	rs290387	chr20	rs537330	chr8
rs10949838	chr7	rs321949	chr19	rs539344	chr19
rs1115649	chr21	rs348971	chr2	rs551372	chr11
rs11207002	chr1	rs390316	chr14	rs567681	chr11
rs11632601	chr15	rs3944117	chr7	rs585487	chr19
rs11971741	chr7	rs425002	chr4	rs600933	chr1
rs12660563	chr6	rs432586	chr12	rs619208	chr11
rs13155942	chr5	rs444016	chr5	rs622994	chr13
rs1444647	chr12	rs4453265	chr11	rs639298	chr1

TABLE 11B: SNP Panel 1					
SNP_ID	**Chromosome**	**SNP_ID**	**Chromosome**	**SNP_ID**	**Chromosome**
rs1572801	chr6	rs447247	chr6	rs642449	chr1
rs17773922	chr19	rs4745577	chr9	rs6700732	chr1
rs1797700	chr12	rs484312	chr13	rs677866	chr13
rs1921681	chr4	rs499946	chr7	rs683922	chr15
rs1958312	chr14	rs500090	chr11	rs686851	chr6
rs196008	chr16	rs500399	chr10	rs6941942	chr6
rs2001778	chr11	rs505349	chr11	rs7045684	chr9
rs2323659	chr17	rs505662	chr6	rs7176924	chr15
rs2427099	chr20	rs516084	chr1	rs7525374	chr1
rs243992	chr4	rs517316	chr1	rs870429	chr3
rs251344	chr5	rs517914	chr4	rs949312	chr18
rs254264	chr19	rs522810	chr13	rs9563831	chr13
rs2827530	chr21	rs531423	chr1	rs970022	chr4
				rs985462	chr10

TABLE 12A

TABLE 12A: SNP Panel 2				
rs1005241	rs1432515	rs2906237	rs654065	rs849084
rs1006101	rs1452396	rs2929724	rs6576533	rs873870
rs10745725	rs1518040	rs3742257	rs6661105	rs9386151
rs10776856	rs16853186	rs3764584	rs669161	rs9504197
rs10790342	rs1712497	rs3814332	rs6703320	rs9690525
rs11076499	rs1792205	rs4131376	rs675828	rs9909561
rs11103233	rs1863452	rs4363444	rs6814242	
rs11133637	rs1991899	rs4461567	rs6989344	
rs11974817	rs2022958	rs4467511	rs7120590	
rs12102203	rs2099875	rs4559013	rs7131676	
rs12261	rs2108825	rs4714802	rs7214164	
rs12460763	rs2132237	rs4775899	rs747583	
rs12543040	rs2195979	rs4817609	rs768255	

TABLE 12A: SNP Panel 2				
rs12695642	rs2248173	rs488446	rs768708	
rs13137088	rs2250246	rs4950877	rs7828904	
rs13139573	rs2268697	rs530913	rs7899772	
rs1327501	rs2270893	rs6020434	rs7900911	
rs13438255	rs244887	rs6442703	rs7925270	
rs1360258	rs2736966	rs6487229	rs7975781	
rs1421062	rs2851428	rs6537064	rs8111589	

TABLE 12B

TABLE 12B: SNP Panel 2					
SNP_ID	**Chromosome**	**SNP_ID**	**Chromosome**	**SNP_ID**	**Chromosome**
rs1518040	chr1	rs11974817	chr7	rs10745725	chr12
rs16853186	chr1	rs13438255	chr7	rs2250246	chr12
rs2268697	chr1	rs2736966	chr7	rs2270893	chr12
rs3814332	chr1	rs2906237	chr7	rs6487229	chr12
rs4363444	chr1	rs4131376	chr7	rs7975781	chr12
rs4950877	chr1	rs849084	chr7	rs12261	chr13
rs6661105	chr1	rs9690525	chr7	rs3742257	chr13
rs6703320	chr1	rs12543040	chr8	rs675828	chr13
rs1432515	chr2	rs1863452	chr8	rs12102203	chr15
rs12695642	chr3	rs2022958	chr8	rs4775899	chr15
rs2132237	chr3	rs6989344	chr8	rs6576533	chr15
rs6442703	chr3	rs7828904	chr8	rs11076499	chr16
rs13137088	chr4	rs10776856	chr9	rs244887	chr16
rs13139573	chr4	rs11103233	chr9	rs654065	chr16
rs1452396	chr4	rs1327501	chr9	rs7214164	chr17
rs1712497	chr4	rs1360258	chr9	rs9909561	chr17
rs4461567	chr4	rs1421062	chr10	rs12460763	chr19
rs4467511	chr4	rs2248173	chr10	rs2108825	chr19

TABLE 12B: SNP Panel 2					
SNP_ID	**Chromosome**	**SNP_ID**	**Chromosome**	**SNP_ID**	**Chromosome**
rs6537064	chr4	rs768255	chr10	rs2195979	chr19
rs6814242	chr4	rs7899772	chr10	rs3764584	chr19
rs747583	chr4	rs7900911	chr10	rs8111589	chr19
rs1006101	chr5	rs10790342	chr11	rs873870	chr19
rs11133637	chr5	rs1792205	chr11	rs530913	chr20
rs2929724	chr5	rs1991899	chr11	rs6020434	chr20
rs4559013	chr5	rs2099875	chr11	rs4817609	chr21
rs4714802	chr6	rs2851428	chr11	rs1005241	chr22
rs669161	chr6	rs488446	chr11		
rs9386151	chr6	rs7120590	chr11		
rs9504197	chr6	rs7131676	chr11		
		rs768708	chr11		
		rs7925270	chr11		

Generation of Illumina Sequencer ready amplicons

For SNP panel 1, PCR primers were designed to amplify the 67 targeted SNPs plus a flanking region of 35 base pairs (bp) surrounding the SNP site. The 67 targeted regions were amplified in a single multiplex reaction. For SNP panel 2, PCR primers were designed to amplify the 86 targeted SNPs plus a flanking region of 26 base pairs (bp) surrounding the SNP site. The 86 targeted regions were amplified in a single multiplex reaction.

PCR primers were modified such that Illumina sequencing adapters could be added via universal tag sequences incorporated onto the 5' end of the SNP-specific PCR primers. Illumina tags were added using two separate PCR reactions (see FIG. 166 and Table 13 below): 1) a loci-specific PCR which incorporated a section of the Illumina sequencing adapters followed by 2) a universal PCR whose primers annealed to the tags in the loci-specific PCR to complete the addition of the adapters whilst allowing the addition of a sample specific index sequence via the reverse primer in the universal PCR. A 3rd single cycle PCR was performed to remove heteroduplex secondary

structure that can arise in the amplicons during the universal PCR stage due to cross-annealing of shared adapter sequences between different amplicons in the same multiplex. Loci-specific PCR and universal PCR were performed under standard conditions using primers synthesized from Integrated DNA Technologies (IDT; Coralville, IA) with no special modifications.

TABLE 13

TABLE 13: Sequencing adaptors, loci specific PCR primer tags and universal PCR primer tags	
Name	Sequence (SEQ ID NOS 364 and 367-373, respectively, in order of appearance)
TRUSEQ P5 Adapter	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT- 3'
TRUSEQ Read 1 sequencing primer	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
TRUSEQ P7 adapter, Index 13	5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTCAAATCTCGTATGCCGTCTTCTGCTTG-3'
TRUSEQ index read primer	5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'
Loci PCR forward tag	5'-TCTTTCCCTACACGACGCTCTTCCGATCT-3'
Loci PCR reverse tag	5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
UNIV PCR forward primer	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC-3'
UNIV PCR reverse index 13 primer	5'-CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTG-3'

Amplicon sequencing by Illumina NGS

Universal PCR products were quantified using standard DNA fragment analysis methods such as Caliper LabChip GX or Agilent Bioanalyzer. The sequencer-ready amplicons from up to 12 samples were pooled and sequenced on an Illumina HISEQ apparatus. For SNP panel 1, 36 cycles were used to sequence the target SNP plus the 35 bp flanking region. For SNP panel 2, 27 cycles were used to sequence the target SNP plus the 26 bp flanking region. Samples were de-mulitplexed using a 6 bp index identifier incorporated at the universal PCR stage.

Assignment of informative alleles and fetal fraction determination

Reads were aligned to the human genome (hg19) with up to 3 mismatches in each read to allow for sequencing error and variant alleles at target SNP position. The frequency of each SNP allele was determined by counting the number of reads having the allele of interest and dividing it by the total number of reads for each SNP locus (i.e., (# reads allele 1)/(# reads allele 1 + # reads allele 2)). Based on the frequency value generated from this data, the sequenced genotypes were assigned as Type 0 non-informative genotypes, Type 1 informative genotypes or Type 2 informative genotypes. A Type 0 non-informative genotype is a fetal genotype that cannot be distinguished from the maternal genotype because the fetus has the same genotype as the mother (e.g., mother is “Aa” and fetus is “Aa”). A Type I informative genotype is the situation where the mother is homozygous (AA) and the fetus is heterozygous (Aa). This genotype is informative because allele “a” is from the father. The frequency of a Type 1 informative allele can be indicative of the percentage fetal DNA in the mixture. A Type 2 informative genotype is the situation where the mother is heterozygous (Aa) and the fetus is homozygous (AA). The genotype is informative because the frequency of the maternal allele “a” will deviate from the expected Mendelian frequency of 0.5 when there is fetal DNA contributing additional “A” alleles. This deviation in value from 0.5 can be used to compute the fetal fraction.

Allele frequencies for each of the SNPs was calculated for each sample based on the number of reads containing each allele, as described above. Variation of expected allele frequency could be due to the presence of fetal DNA with a different paternal allele or could be due to mis-incorporated sequences by the Illumina Sequencer (e.g., background noise). In some cases, the amount of background noise associated with each particular SNP amplicon was determined to establish a dynamic cutoff value for each SNP. Maternal DNA (i.e. buffy coat) samples were sequenced and the deviations from the expected Mendelian ratios of 1 for homozygotes and 0.5 for heterozygotes were observed. From these values a median-adjusted deviation (MAD score) was identified for each SNP assay. In some cases, a genotype was identified as being a Type I informative genotype when the paternal allele frequency measured was greater than 3x MAD score. In some cases, multiple Type 1 informative genotypes were identified and an average allele frequency was determined. Fetal fraction was calculated by multiplying the average Type 1 informative allele frequency by 2. For example, an average informative allele frequency of 4.15% indicated a fetal fraction of 8.3%. Fetal Fraction also can be calculated from Type 2 informative genotypes by

determining maternal allele "a" frequencies deviating from 0.5 by greater than 3x MAD, for example. Fetal fraction can be identified by multiplying this deviation by 2.

In some cases, informative genotypes were assigned without prior knowledge of maternal or paternal genotypes. Allele frequencies for each SNP (of SNP panel 1) were plotted as shown in FIG. 167 and FIG. 168 for two of the 46 samples tested. Homozygous allele frequencies in maternal buffy coat were close to 0 or 1. Type 1 informative SNPs were identified by allele frequencies that deviated from the expected allele frequency of 0 or 1 due to the presence of a paternal allele from the fetus. The size of the deviation was dependent on the size of the fetal fraction of CCF DNA. A maximum background allele frequency of 0.007 was observed for maternal buffy coat DNA. For this approach, fixed cutoff frequency value of 0.01 was used to distinguish non-informative homozygotes from informative genotypes in plasma samples (see FIGS. 169 and 170, showing the assignment of certain Type 1 informative genotypes). A fixed cutoff value of 0.25 was used to distinguish non-informative heterozygotes from other genotypes. Fetal fractions were calculated for 46 plasma samples by taking the mean of the informative genotype allele frequencies and multiplying this value by 2. Informative genotypes assigned per sample ranged from 1 to 26. Fetal fractions ranged from 2.5% to 14% (see FIG. 171).

To assess performance of the above method, fetal fractions also were determined for the 46 plasma samples using a differential methylation-based fetal quantifier assay. SNP-based fetal fraction estimates showed a linear association with the methylation-based estimates ($r^2 = 0.72$). FIG. 172 shows linear regression of fetal fraction estimate methods as a diagonal line.

Amplicon sequence coverage

Various amounts of SNP amplicon libraries were combined (i.e. diluted) with TRUSEQ libraries to demonstrate that allele frequency determinations can be made at varying levels of amplicon sequence coverage. SNP amplicon libraries from 6 plasma samples and 6 buffy coat samples were combined with 11 TRUSEQ libraries and co-sequenced on a HISEQ 2000 apparatus in the same flowcell lane. Percent (%) of SNP amplicon library combined with TRUSEQ libraries ranged from 50% to 0.8%. After alignment coverage per SNP for each amplicon library ranged from 71619x per SNP (50% amplicon library) to 1413x per SNP (0.8% amplicon library). Fetal fraction estimates were not significantly different even at lowest coverage level (see FIG. 173). These findings indicate that less than 1% of the flowcell clusters on a HISEQ 2000 apparatus can be used

to co-sequence amplicon libraries and that high levels of sample multiplexing (e.g., greater than 96) can be achieved.

Example 16: Using Fetal Fraction Derived from Sequence Reads to Determine Aneuploidy

MPSS samples were classified as euploid or T13/T18/T21 based on fitted ploidy values derived from fetal fraction estimates based on ChrY representation derived from whole-genome sequencing results.

Examples 9 and 10 have demonstrated the accuracy and precision of sequencing-based fetal fractions determinations in male fetuses. In this example, those measurements were used to resolve difficult trisomy cases with low fetal fraction values. As a proof of principle, the method was applied to male pregnancy subsets of the LDTv2CE data set. The same approach can be applied to female pregnancies as well using the methods in Examples 11-15 or similar high-accuracy fetal fraction measurements. FIGS. 175-177 show fitted ploidy values obtained by combining PERUN profiles of chromosomes 13, 18, and 21 with fetal fractions derived from PERUN chromosome Y representations. Similar results were obtained with chromosome X-based fetal fractions.

FIGS. 175-177 illustrate that fitted ploidy extracted from sequencing-based fetal fraction measurements perfectly and accurately distinguished T13, T18, and T21 samples from euploids. Ideally, trisomy and euploid samples should have ploidy values of 1.5 and 1, respectively. The LDTv2CE male trisomy cases have fitted ploidy values that always exceed 1.4, while euploids were always below 1.3 and in a great majority of the cases below 1.2.

This method is insensitive to fetal fraction in that it correctly classified T21, T18, and T13 cases even at extremely low fetal fractions. These samples tend to have borderline Z-scores. Fitted fetal fraction does not have any borderline cases irrespective of the fetal fraction.

Example 17: Detecting Chromosome Abnormalities Using PERUN Technology

The methods, or parts or combinations thereof, described herein (e.g., Examples 1 to 8 above) where used to determine the presence or absence of chromosome abnormalities (e.g., chromosome aneuploidies, microduplications, microdeletions, trisomies, the like and/or combinations thereof). The methods described herein were used to detect a trisomy of each of

chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, X and/or Y in a fetus. In addition, the methods described herein were used to detect a microdeletion and/or microduplication in each of chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22 and/or X in a fetus. The methods described herein were also used to detect complex genetic variations in a fetus that included one or more trisomies, one or more microduplications, one or more microdeletions, and combinations thereof. Representative examples are presented herein. Samples comprising circulating cell-free nucleic acid were obtained from a blood product obtained from one or more pregnant female subjects.

As described herein, PERUN technology was applied to detect fetal microdeletions, microduplications, autosomal trisomies and multiple trisomies. Some of these fetal aberrations were known (either as being present in a general population with non-negligible frequency or as being confirmed in a given sample by conventional karyotyping techniques). In some examples, PERUN technology facilitated identification of these aberrations. In other examples, PERUN enables discovery of previously unknown aberrations or aberrations that have been missed by conventional karyotyping.

PERUN technology was utilized to detect a fetal trisomy 16 (FIG. 190A-F). Six cases of fetal trisomy 16 were determined. All T16 calls (based on Z-scores) were confirmed by conventional karyotyping. The elevation of chromosome 16 sometimes varied with the fetal fraction. Visual inspection was sufficient to classify all six samples as T16. Due to its high GC content, chromosome 16 is prone to GC bias and is typically difficult to quantify. This example demonstrated PERUN's successful removal of systematic biases from an otherwise highly biased chromosome.

PERUN technology was utilized to detect a fetal trisomy 20 (FIG. 179A-C) in three separate samples and a trisomy 22 (FIG. 180) in another sample. Trisomy 20 is another aberration known to occasionally occur in the general population. Not unlike chromosome 16, the quantification of these chromosomes are also complicated by their high GC content and vulnerability to GC bias. The use of PERUN greatly facilitated the detection of a chromosome aneuploidy of both chromosomes.

The use of PERUN technology was extended from the detection of whole single chromosome aberrations to trisomies involving multiple chromosomes. A trisomy of chromosome 21 was

detected in all samples (FIG. 181 and FIG. 182). Figures 181A to 181C refer to the same patient. In addition, a trisomy of chromosome 4 (FIG. 181) and of chromosome 9 (FIG. 182) were detected. Only one aberration (T21) in both cases, was discovered using conventional karyotyping. However, the use of conventional karyotyping did not detect the second chromosome aneuploidy.

PERUN was also used to detect the presence of a whole chromosome aneuploidy of chromosome 16 (FIG. 183A-B). Conventional karyotyping failed to detect this chromosome aneuploidy.

PERUN was used to detect microdeletions in chromosome X (FIG. 184A-C) and microdeletions in chromosome 7 (FIG. 185). Tables 2 and 3 list some of the genes conditions associated with the observed microdeletions. The fetal fraction for the chromosome X determination was measured twice using FQA, giving the values of 35% and 21%, in line with the observed depression of ~11%. The microdeletion in chromosome X ranged from chrX:35,500,001 to chrX:48,500,000 (or from chrX:35,950,000 to chrX:48,350,000, depending on which PERUN bin selection was used or which set of bins were included in the normalized profiles).

PERUN was used to detect microdeletions in chromosome 18 (FIG. 186A-B), chromosome 3 (FIG. 187) and chromosome 4 (FIG. 188). The duplication in chromosome 18 (p arm, p11.32-p11.21) started at chr18:300,001 and ended at chr18:18,700,000, spanning 18.4 Mbp. The chromosome 3 microduplication encompassed positions chr3:1-40,500,000. The microduplication in chromosome 4 extended from chr4:167,150,001 to chr4:180,050,000. These regions contained numerous genes. Some of the genes and disorders associated with these microduplications are presented in Tables 4-6. According to a truth table, the pregnancy represented in FIG. 187 was terminated due to multiple fetal defects involving heart and cleft. The mother was shown to have a balanced translocation 46,XX, t(1;3)(p36.3;p23). Regarding the chromosome 4 duplication shown in FIG. 188, a truth table only indicated 46,XY. Although chromosome Y appears to be elevated, visual inspection of chromosome X does not suggest a depression. In some embodiments, elevated Y with a lack of depression in X suggests that, according to sequencing data, the actual karyotype is 47,XXY (Klinefelter).

PERUN technology was also used to detect complex fetal abnormalities involving duplications in chromosome 5 (the first 15 Mbp) and chromosome 8 (starting at chr8:119,750,000), as well as a trisomy-X. The microduplication in chromosome 8 (Fig. 189A-B) may be an artifact, as it is shared

with many other samples. The elevations of chromosomes 11 and 14 demonstrated unusual patterns involving extensive deletions and duplications.

The known unbalanced versions of translocation t(11;14)(q13;q32) involved in mantle cell lymphoma (Van den Berghe, et al., A new characteristic karyotypic anomaly in lymphoproliferative disorders. Cancer (1979); **44:** 188-95) differed from the aberration discovered in this sample. Van den Berghe, et al., reported that band 11q13 started at chr11:63,400,000 and ended at chr11:77,100,000, close to the middle portion of the chromosome 11. In contrast, the observed depression of chromosome 11 in this sample started at around chr11:93,000,000 and extended all the way to the end of the q arm. In addition, while 14q32 occupies the end of the q arm of chromosome 14, starting at chr4:89,800,000, the observed deletion of chromosome 14 started at chr14:58,300,000. To our knowledge, Fig. 189 showed the first detection and quantification of this particular aberration. The conventional karyotyping call failed to observe any of these multiple and sizeable abnormalities. Consequently, a truth table for this sample incorrectly indicated euploidy (46,XX). The discovery of this complex genomic rearrangement was enabled by PERUN.

The findings detailed herein only include a representative subset of a much larger set of microdeletions and microduplications detected in the WI and LDTv2CE studies. The whole chromosome abnormalities shown in Figures 178-182 (T4, T9, T16, T20, T22) may be confined to placenta (e.g., fetal placenta) without necessarily affecting the fetus, in which case the pregnancy can be carried to term without any complications due to the observed trisomies. In addition, without additional samples from the mother, the possibility that some of the detected aberrations may originate from a minority population of maternal cells cannot be excluded.

Example 18: Examples of embodiments

Listed hereafter are non-limiting examples of certain embodiments of the technology.

A1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) from the counts in (a), generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

A1.1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fraction of the fetal nucleic acid in the blood of the pregnant female by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) produce from the counts in (a) an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

A2. The method of embodiment A1 or A1.1, wherein the X chromosome representation in (c)(ii) is an experimental X chromosome representation.

A3. The method of embodiment A1, A1.1 or A2, wherein the X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof.

A4. The method of any one of embodiments A1 to A3, wherein the X chromosome representation in (c)(ii) is determined from a set of pregnant females bearing a male fetus.

A5. The method of any one of embodiments A1 to A4, wherein the X chromosome representation in (c)(ii) is determined for pregnant females bearing a male fetus.

A6. The method of any one of embodiments A1 to A5, wherein the fitted relationship is linear.

A7. The method of embodiment A6, wherein the fraction of the fetal nucleic acid is determined according to the slope and intercept of the fitted relationship, the experimental Y chromosome representation generated in (b) and a median X chromosome representation for a set of pregnant females bearing a female fetus.

A8. The method of embodiment A7, wherein the fraction of the fetal nucleic acid is determined according to equation (62):

$$f = 2\frac{I+S\langle x\rangle-y}{S\langle x\rangle} \qquad (62)$$

wherein *I* is the intercept, *S* is the slope, (*x*) is the median X chromosome representation for a set of pregnant females bearing a female fetus and *y* is the experimental Y chromosome representation generated in (b).

A9. The method of any one of embodiments A1 to A8, wherein the median chromosome X representation is a median experimental X chromosome representation.

A10. The method of any one of embodiments A1 to A9, where the median chromosome Y representation is a median experimental Y chromosome representation.

A11. The method of any one of embodiments A1 to A10, wherein the fitted relationship in (c) is determined prior to (a).

A12. The method of any one of embodiments A1 to A11, wherein the fitted relationship in (c) is determined prior to (b).

A13. The method of any one of embodiments A1 to A12, wherein the counts in (a) are obtained from a pregnant female bearing a male fetus having a chromosome aneuploidy.

A14. The method of embodiment A13, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

A15. The method of embodiment A13, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

A16. The method of any one of embodiments A1 to A15, wherein the set of pregnant females in (c)(i) is a set of about 500 females or more.

A17. The method of any one of embodiments A1 to A16, wherein the set of pregnant females in (c)(ii) is a set of about 500 females or more.

A18. The method of any one of embodiments A1 to A17, wherein the set of pregnant females in (c)(i) and (c)(ii) are the same set.

A19. The method of any one of embodiments A1 to A18, wherein the median chromosome X representation and a median chromosome Y representation is determined for a set of about 500 pregnant females or more.

A20. The method of any one of embodiments A1 to A19, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is provided by a fetal fraction module.

A21. The method of any one of embodiments A1 to A20, comprising normalizing the counts in (a), thereby providing normalized counts mapped to the genomic sections of the reference genome; and which counts in (b) are normalized counts.

A22. The method of embodiment A21, wherein the counts in (b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

A23. The method of embodiment A19 or A22, wherein the normalized counts are provided by a normalization module.

A24. The method of any one of embodiments A1 to A23, wherein the experimental Y chromosome representation in (b) is provided by an experimental representation module.

A25. The method of any one of embodiments A1 to A24, wherein the fetal fraction in (c) is provided by a fetal fraction module.

A26. The method of any one of embodiments A1 to A25, wherein the fitted relationship is provided by a relationship module.

A27. The method of any one of embodiments A24 to A26, wherein the normalized counts are transferred to the experimental representation module from the normalization module.

A28. The method of any one of embodiments A25 to A27, wherein the experimental Y chromosome representation is transferred to the fetal fraction module from the experimental representation module.

A29. The method of any one of embodiments A1 to A28, which comprises obtaining nucleic acid sequence reads.

A30. The method of embodiment A29, wherein the nucleic acid sequence reads are generated by a sequencing module.

A31. The method of any one of embodiments A1 to A30, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

A32. The method of any one of embodiments A1 to A31, which comprises mapping the nucleic acid sequence reads to genomic sections in a segment of the reference genome or to an entire reference genome.

A33. The method of embodiment A32, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

A34. The method of any one of embodiments A1 to A33, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

A35. The method of embodiment A33 or A34, wherein the sequence reads are transferred to the mapping module from the sequencing module.

A36. The method of embodiment A34 or A35, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

A37. The method of any one of embodiments A34 to A36, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

A38. The method of any one of embodiments A34 to A37, wherein an apparatus comprises one or more of the sequencing module, a sequence receiving module, the mapping module, the counting module, the normalization module, the experimental representation module, the relationship module, the fetal fraction module, a comparison module, a range setting module, a categorization

module, an adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

A39. The method of embodiment A38, wherein a first apparatus comprises one or more of the normalization module, the experimental representation module, the relationship module and the fetal fraction module.

A40. The method of embodiment A38 or A39, wherein a second apparatus comprises the mapping module and the counting module.

A41. The method of any one of embodiments A38 to A40, wherein a third apparatus comprises the sequencing module.

A41.1. The method of any one of embodiments A39 to A40, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

A42. The method of any one of embodiments A21 to A41.1, wherein the counts that are normalized are raw counts.

A43. The method of any one of embodiments A21 to A42, wherein the counts that are normalized are filtered.

A44. The method of any one of embodiments A21 to A42, wherein the counts that are normalized are not filtered.

A45. The method of any one of embodiments A1 to A44, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

A46. The method of any one of embodiments A1 to A45, wherein the genomic sections of the reference genome are one or more bins.

A47. The method of embodiment A46, wherein each bin is of about an equal number of contiguous nucleotides.

A48. The method of embodiment A46 or A47, wherein each bin is about 50 kb.

A49. The method of any one of embodiments A1 to A48, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

A50. The method of any one of embodiments A1 to A49, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

A51. The method of any one of embodiments A1 to A50, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

A52. The method of embodiment A51, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

A53. The method of embodiment A51 or A52, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

A54. The method of any one of embodiments A51 to A53, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

A55. The method of any one of embodiments A51 to A53, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

A56. The method of any one of embodiments A1 to A55, which comprises determining the gender of the fetus.

A57. The method of embodiment A56, wherein the gender of the fetus is determined before performing (b).

A58. The method of embodiment A56 or A57, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

A59. The method of embodiment A56, wherein the gender of the fetus is determined before performing (a).

A60. The method of any one of embodiments A56 to A59, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

A61. The method of any one of embodiments A56 to A59, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

A62. The method of any one of embodiments A1 to A61, which comprises determining the presence or absence of a Y chromosome in the fetus.

A63. The method of any one of embodiments A1 to A62, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

A64. The method of embodiment A63, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

A65. The method of any one of embodiments A1 to A63, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

A66. The method of any one of embodiments A1 to A65, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

A67. The method of embodiment A66, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

A68. The method of any one of embodiments A1 to A67, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

A69. The method of any one of embodiments A1 to A68, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

A70. The method of embodiment A69, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

A71. The method of any one of embodiments A1 to A68, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

B1. A system comprising one or more microprocessors and memory,
which memory comprises instructions executable by the one or more microprocessors, and which instructions executable by the one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus

(b) from the counts in (a), generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof,

wherein the sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

(c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

B1.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to genomic sections of a reference genome, and generate counts of the sequence reads mapped to the genomic sections;

(b) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

B2. The system of embodiment B1 or B2, wherein the X chromosome representation in (c)(ii) is an experimental X chromosome representation.

B3. The system of embodiment B1, B1.1 or B2, wherein the X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof.

B4. The system of any one of embodiments B1 to B3, wherein the X chromosome representation in (c)(ii) is determined from a set of pregnant females bearing a male fetus.

B5. The system of any one of embodiments B1 to B4, wherein the X chromosome representation in (c)(ii) is determined for pregnant females bearing a male fetus.

B6. The system of any one of embodiments B1 to B5, wherein the fitted relationship is linear.

B7. The system of embodiment B6, wherein the fraction of the fetal nucleic acid is determined according to the slope and intercept of the fitted relationship, the experimental Y chromosome representation generated in (b) and a median X chromosome representation for a set of pregnant females bearing a female fetus.

B8. The system of embodiment B7, wherein the fraction of the fetal nucleic acid is determined according to equation (62):

$$f = 2\frac{I+S\langle x\rangle - y}{S\langle x\rangle} \qquad (62)$$

wherein *I* is the intercept, *S* is the slope, (x) is the median X chromosome representation for a set of pregnant females bearing a female fetus and *y* is the experimental Y chromosome representation generated in (b).

B9. The system of any one of embodiments B1 to B8, wherein the median chromosome X representation is a median experimental X chromosome representation.

B10. The system of any one of embodiments B1 to B9, where the median chromosome Y representation is a median experimental Y chromosome representation.

B11. The system of any one of embodiments B1 to B10, wherein the fitted relationship in (c) is determined prior to (a).

B12. The system of any one of embodiments B1 to B11, wherein the fitted relationship in (c) is determined prior to (b).

B13. The system of any one of embodiments B1 to B12, wherein the counts in (a) are obtained from a pregnant female bearing a male fetus having a chromosome aneuploidy.

B14. The system of embodiment B13, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

B15. The system of embodiment B13, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

B16. The system of any one of embodiments B1 to B15, wherein the set of pregnant females in (c)(i) is a set of about 500 females or more.

B17. The system of any one of embodiments B1 to B16, wherein the set of pregnant females in (c)(ii) is a set of about 500 females or more.

B18. The system of any one of embodiments B1 to B17, wherein the set of pregnant females in (c)(i) and (c)(ii) are the same set.

B19. The system of any one of embodiments B1 to B18, wherein the median chromosome X representation and a median chromosome Y representation is determined for a set of about 500 pregnant females or more.

B20. The system of any one of embodiments B1 to B19, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is provided by a fetal fraction module.

B21. The system of any one of embodiments B1 to B20, which instructions are configured to normalize the counts in (a), thereby providing normalized counts mapped to the genomic sections of the reference genome; and which counts in (b) are normalized counts.

B22. The system of embodiment B21, wherein the counts in (b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

B23. The system of embodiment B19 or B22, wherein the normalized counts are provided by a normalization module.

B24. The system of any one of embodiments B1 to B23, wherein the experimental Y chromosome representation in (b) is provided by an experimental representation module.

B25. The system of any one of embodiments B1 to B24, wherein the fetal fraction in (c) is provided by a fetal fraction module.

B26. The system of any one of embodiments B1 to B25, wherein the fitted relationship is provided by a relationship module.

B27. The system of any one of embodiments B24 to B26, wherein the normalized counts are transferred to the experimental representation module from the normalization module.

B28. The system of any one of embodiments B25 to B27, wherein the experimental Y chromosome representation is transferred to the fetal fraction module from the experimental representation module.

B29. The system of any one of embodiments B1 to B28, which instructions are configured to obtain nucleic acid sequence reads.

B30. The system of embodiment B29, wherein the nucleic acid sequence reads are generated by a sequencing module.

B31. The system of any one of embodiments B1 to B30, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

B32. The system of any one of embodiments B1 to B31, which instructions are configured to map the nucleic acid sequence reads to genomic sections in a segment of the reference genome or to an entire reference genome.

B33. The system of embodiment B32, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

B34. The system of any one of embodiments B1 to B33, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

B35. The system of embodiment B33 or B34, wherein the sequence reads are transferred to the mapping module from the sequencing module.

B36. The system of embodiment B34 or B35, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

B37. The system of any one of embodiments B34 to B36, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

B38. The system of any one of embodiments B34 to B37, wherein an apparatus comprises one or more of the sequencing module, a sequence receiving module, the mapping module, the counting module, the normalization module, the experimental representation module, the relationship module, the fetal fraction module, a comparison module, a range setting module, a categorization module, an adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

B39. The system of embodiment B38, wherein a first apparatus comprises one or more of the normalization module, the experimental representation module, the relationship module and the fetal fraction module.

B40. The system of embodiment B38 or B39, wherein a second apparatus comprises the mapping module and the counting module.

B41. The system of any one of embodiments B38 to B40, wherein a third apparatus comprises the sequencing module.

B41.1. The method of any one of embodiments A39 to A40, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

B42. The system of any one of embodiments B21 to B41.1, wherein the counts that are normalized are raw counts.

B43. The system of any one of embodiments B21 to B42, wherein the counts that are normalized are filtered.

B44. The system of any one of embodiments B21 to B42, wherein the counts that are normalized are not filtered.

B45. The system of any one of embodiments B1 to B44, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

B46. The system of any one of embodiments B1 to B45, wherein the genomic sections of the reference genome are one or more bins.

B47. The system of embodiment B46, wherein each bin is of about an equal number of contiguous nucleotides.

B48. The system of embodiment B46 or B47, wherein each bin is about 50 kb.

B49. The system of any one of embodiments B1 to B48, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

B50. The system of any one of embodiments B1 to B49, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

B51. The system of any one of embodiments B1 to B50, which instructions are configured to determine the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

B52. The system of embodiment B51, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

B53. The system of embodiment B51 or B52, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

B54. The system of any one of embodiments B51 to B53, which instructions are configured to determine the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

B55. The system of any one of embodiments B51 to B53, which instructions are configured to determine the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

B56. The system of any one of embodiments B1 to B55, which instructions are configured to determine the gender of the fetus.

B57. The system of embodiment B56, wherein the gender of the fetus is determined before performing (b).

B58. The system of embodiment B56 or B57, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

B59. The system of embodiment B56, wherein the gender of the fetus is determined before performing (a).

B60. The system of any one of embodiments B56 to B59, which instructions are configured to determine whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

B61. The system of any one of embodiments B56 to B59, which instructions are configured to determine whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

B62. The system of any one of embodiments B1 to B61, which instructions are configured to determine the presence or absence of a Y chromosome in the fetus.

B63. The system of any one of embodiments B1 to B62, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

B64. The system of embodiment B63, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

B65. The system of any one of embodiments B1 to B63, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

B66. The system of any one of embodiments B1 to B65, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

B67. The system of embodiment B66, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

B68. The system of any one of embodiments B1 to B67, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

B69. The system of any one of embodiments B1 to B68, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

B70. The system of embodiment B69, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

B71. The system of any one of embodiments B1 to B68, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

C1. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) from the counts, generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

D1. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) from the counts in (a), generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof;

(c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

E1. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental X chromosome representation and an expected X chromosome representation, which expected X chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the X chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

E1.1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fraction of the fetal nucleic acid in the blood of the pregnant female by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) generate, from the counts in (a), an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determine the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental X chromosome representation and an expected X chromosome representation, which expected X chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the X chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

E2. The method of embodiment E1 or E1.1, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined in (c) according to a ratio of the experimental X chromosome representation and the expected X chromosome representation.

E2.1. The method of embodiment E1, E1.1 or E2, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined according to equation AC.

E3. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

E3.1. The method of embodiment E3, wherein the fraction of fetal nucleic acid in the blood of the pregnant female bearing a fetus having a chromosome aneuploidy is determined according to equation AB.

E3.2. The method of embodiment E3 or E3.1, wherein the fraction of fetal nucleic acid determined in (c)(i) and the experimental X chromosome representation in (c)(ii) are derived from greater than about 500 subjects.

E3.3. The method of any one of embodiments E1 to E3.2, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is provided by a fetal fraction module.

E4. The method of any one of embodiments E3 to E3.2, wherein the relationship is a linear relationship.

E4.1. The method of any one of embodiments E3 to E4, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

E4.2. The method of any one of embodiments E3 to E4.1, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

E4.3. The method of any one of embodiments E3 to E4.2, wherein the relationship is:

$$F = k - r(MCRx),$$

wherein F is the fraction, MCRx is the experimental X chromosome representation, k is an intercept from the linear relationship and r is a slope from the linear relationship.

E4.3.1. The method of embodiment E4.3, wherein the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female is determined by the relationship in E4.3 and the experimental X chromosome representation.

E4.4. The method of any one of embodiments E3 to E4.3.1, wherein the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that comprises use of sequence reads mapped to genomic sections of a reference genome.

E4.5. The method of any one of embodiments E3 to E4.3.1, wherein the fraction of fetal nucleic acid determined for nucleic acid from nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that does not utilize sequence reads mapped to genomic sections of a reference genome.

E4.6. The method of embodiment E4.5, wherein the process comprises mass spectrometry.

E4.7. The method of any one of embodiments E3 to E4.6, wherein the pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is different from the pregnant female in (a).

E4.7.1 The method of embodiment E4.7, wherein the pregnant female bearing a fetus having a chromosome aneuploidy is bearing a male fetus.

E4.8. The method of any one of embodiments E3 to E4.7.1, wherein the nucleic acid from the blood of the pregnant female in (c)(i) is circulating cell-free nucleic acid.

E5. The method of any one of embodiments E1 to E4.8, comprising normalizing the counts mapped to the genomic sections of the reference genome, thereby providing normalized counts mapped to the genomic sections of the reference genome; and which counts in E1(b) and E3(b) are normalized counts.

E5.1. The method of any one of embodiments E1 to E5, wherein the counts in E1(b) and E3(b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

E6. The method of embodiment E5, wherein the normalized counts are provided by a normalization module.

E7. The method of any one of embodiments E1 to E6, wherein the experimental X chromosome representation in E1(b) and E3(b) is provided by a representation module.

E8. The method of any one of embodiments E1 to E7, wherein expected X chromosome representation in E1(c) and E3(c) is determined by an expected representation module.

E9. The method of any one of embodiments E1 to E8, wherein the fetal fraction in E1(c) and E3(c) is provided by a fetal fraction module.

E10. The method of any one of embodiments E3 to E9, wherein the relationship is provided by a relationship module.

E11. The method of any one of embodiments E7 to E10, wherein the normalized mapped counts are transferred to the experimental representation module from the normalization module.

E12. The method of any one of embodiments E8 to E11, wherein the normalized mapped counts are transferred to the expected representation module from the normalization module.

E13. The method of any one of embodiments E9 to E12, wherein the experimental X chromosome representation is transferred to the fetal fraction module from the experimental representation module.

E14. The method of any one of embodiments E9 to E13, wherein the expected X chromosome representation is transferred to the fetal fraction module from the expected representation module.

E15. The method of any one of embodiments E1 to E14, which comprises obtaining nucleic acid sequence reads.

E16. The method of embodiment E15, wherein the nucleic acid sequence reads are generated by a sequencing module.

E16.1 The method of embodiment E15 or E16, wherein the nucleic acid sequencing reads are generated by massively parallel sequencing (MPS).

E17. The method of embodiment E15 or E16, which comprises mapping the nucleic acid sequence reads to the genomic sections of the reference genome or to an entire reference genome.

E18. The method of embodiment E17, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

E19. The method of any one of embodiments E1 to E18, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

E20. The method of any one of embodiments E18 to E19, wherein the sequence reads are transferred to the mapping module from the sequencing module.

E21. The method of any one of embodiments E19 to E20, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

E22. The method of any one of embodiments E19 to E21, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

E23. The method of any one of embodiments E1 to E22, wherein an apparatus comprises one or more of a sequencing module, sequence receiving module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module or logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

E24. The method of embodiment E23, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

E24.1. The method of any one of embodiments E19 to E24, wherein a second apparatus comprises the mapping module and the counting module.

E25. The method of any one of embodiments E16 to E24.1, wherein a third apparatus comprises the sequencing module.

E25.1. The method of any one of embodiments E24 to E25, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

E26. The method of any one of embodiments E5 to E25.1, wherein the counts that are normalized are raw counts.

E27. The method of any one of embodiments E5 to E26, wherein the counts that are normalized are filtered.

E27.1 The method of any one of embodiments E5 to E26, wherein the counts that are normalized are not filtered.

E28. The method of any one of embodiments E1 to E27, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

E29. The method of any one of embodiments E1 to E28, wherein the genomic sections of the reference genome are one or more bins.

E30. The method of embodiment E29, wherein each bin is of about an equal number of contiguous nucleotides.

E31. The method of embodiment E29 or E30, wherein each bin is about 50 kb.

E32. The method of any one of embodiments E1 to E31, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

E33. The method of any one of embodiments E1 to E32, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

E34. The method of any one of embodiments E1 to E33, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

E35. The method of embodiment E34, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

E36. The method of embodiment E34 or E35, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

E37. The method of any one of embodiments E34 to E36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

E38. The method of any one of embodiments E34 to E36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

E39. The method of any one of embodiments E1 to E38, which comprises determining the gender of the fetus.

E40. The method of embodiment E39, wherein the gender of the fetus is determined before performing (b).

E41. The method of embodiment E39 or E40, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

E42. The method of embodiment E39, wherein the gender of the fetus is determined before performing (a).

E43. The method of any one of embodiments E39 to E42, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

E44. The method of any one of embodiments E39 to E42, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

E45. The method of any one of embodiments E1 to E44, which comprises determining the presence or absence of a Y chromosome in the fetus.

E46. The method of any one of embodiments E1 to E45, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

E47. The method of embodiment E46, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

E48. The method of any one of embodiments E1 to E46, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

E49. The method of any one of embodiments E1 to E48, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

E50. The method of embodiment E49, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

E51. The method of any one of embodiments E1 to E50, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

E52. The method of any one of embodiments E1 to E51, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

E53. The method of embodiment E52, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

E54. The method of any one of embodiments E1 to E51, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

F1. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental Y chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation and an expected Y chromosome representation, which expected Y chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the Y chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

F1.1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fraction of the fetal nucleic acid in the blood of the pregnant female by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) generate, from the counts in (a), an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence

reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental Y chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation and an expected Y chromosome representation, which expected Y chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the Y chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

F2. The method of embodiment F1 or F1.1, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined in (c) according to a ratio of the experimental Y chromosome representation and the expected Y chromosome representation.

F2.1. The method of embodiment F1, F1.1 or F2, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined according to equation AC.

F3. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental Y chromosome representation, determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

F3.1. The method of embodiment F3, wherein the fraction of fetal nucleic acid in the blood of the pregnant female bearing a fetus having a chromosome aneuploidy is determined according to equation AB.

F3.2. The method of embodiment F3 or F3.1 wherein the fraction of fetal nucleic acid determined in (c)(i) and the experimental Y chromosome representation in (c)(ii) are derived from greater than about 500 subjects.

F4. The method of embodiment F3, wherein the relationship is a linear relationship.

F4.1. The method of embodiment F3 or F4, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

F4.2. The method of embodiment F3 or F4.1, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

F4.3. The method of any one of embodiments F4 to F4.2, wherein the relationship is:

$$F = k - r(MCR_Y),$$

wherein F is the fraction, MCR_Y is the experimental Y chromosome representation, k is an intercept from the linear relationship and r is a slope from the linear relationship.

F4.3.1. The method of embodiment F4.3 wherein the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female is determined by the relationship in F4.3 and the experimental Y chromosome representation.

F4.4. The method of any one of embodiments F3 to F4.3.1, wherein the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that comprises use of sequence reads mapped to genomic sections of a reference genome.

F4.5. The method of any one of embodiments F3 to F4.3.1, wherein the fraction of fetal nucleic acid determined for nucleic acid from nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that does not utilize sequence reads mapped to genomic sections of a reference genome.

F4.6. The method of embodiment F4.5, wherein the process comprises mass spectrometry.

F4.7. The method of any one of embodiments F3 to F4.6, wherein the pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is different from the pregnant female in (a).

F4.7.1 The method of F.4.7 wherein the pregnant female bearing a fetus having a chromosome aneuploidy is bearing a male fetus.

F4.8. The method of any one of embodiments F3 to F4.7.1, wherein the nucleic acid from the blood of the pregnant female in (c)(i) is circulating cell-free nucleic acid.

F5. The method of any one of embodiments F1 to F4.8, comprising normalizing the counts mapped to the genomic sections of the reference genome, thereby providing normalized counts of the genomic sections of the reference genome; and which counts in F1(b) and F3(b) are normalized counts.

F5.1. The method of any one of embodiments F1 to F5, wherein the counts in F1(b) and F3(b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

F6. The method of embodiment F5, wherein the normalized counts are provided by a normalization module.

F7. The method of any one of embodiments F1 to F6, wherein the experimental Y chromosome representation in F1(b) and F3(b) is provided by a representation module.

F8. The method of any one of embodiments F1 to F7, wherein expected Y chromosome representation in F1(c) and F3(c) is determined by an expected representation module.

F9. The method of any one of embodiments F1 to F8, wherein the fetal fraction in F1(c) and F3(c) is provided by a fetal fraction module.

F10. The method of any one of embodiments F3 to F9, wherein the relationship is provided by a relationship module.

F11. The method of any one of embodiments F7 to F10, wherein the normalized mapped counts are transferred to the Y experimental module from the normalization module.

F12. The method of any one of embodiments F8 to F11, wherein the normalized mapped counts are transferred to the Y expected module from the normalization module.

F13. The method of any one of embodiments F9 to F12, wherein the experimental Y chromosome representation is transferred to the fetal fraction module from the Y experimental module.

F14. The method of any one of embodiments F9 to F13, wherein the expected Y chromosome representation is transferred to the fetal fraction module from the Y expected module.

F15. The method of any one of embodiments F1 to F14, which comprises obtaining nucleic acid sequence reads.

F16. The method of embodiment F15, wherein the nucleic acid sequence reads are generated by a sequencing module.

F16.1 The method of embodiment F15 or F16, wherein the nucleic acid sequencing reads are generated by massively parallel sequencing (MPS).

F17. The method of embodiment F15 or F16, which comprises mapping the nucleic acid sequence reads to the genomic sections of the reference genome.

F18. The method of embodiment F17, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

F19. The method of any one of embodiments F1 to F18, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

F20. The method of any one of embodiments F18 to F19, wherein the sequence reads are transferred to the mapping module from the sequencing module.

F21. The method of any one of embodiments F19 to F20, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

F22. The method of any one of embodiments F19 to F21, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

F23. The method of any one of embodiments F1 to F22, wherein an apparatus comprises one or more of a sequencing module, sequence receiving module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module, logic processing module, representation module, relationship module or fetal fraction module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

F24. The method of embodiment F23, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

F24.1. The method of any one of embodiments F19 to F24, wherein a second apparatus comprises the mapping module and the counting module.

F25. The method of any one of embodiments F16 to F24.1, wherein a third apparatus comprises the sequencing module.

F25.1. The method of any one of embodiments F24 to F25, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

F26. The method of any one of embodiments F5 to F25.1, wherein the counts that are normalized are raw counts.

F27. The method of any one of embodiments F5 to F26, wherein the counts that are normalized are filtered.

F27.1 The method of any one of embodiments F5 to F26, wherein the counts that are normalized are not filtered.

F28. The method of any one of embodiments F1 to F27, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

F29. The method of any one of embodiments F1 to F28, wherein the genomic sections of the reference genome are one or more bins.

F30. The method of embodiment F29, wherein each bin is of about an equal number of contiguous nucleotides.

F31. The method of embodiment F29 or F30, wherein each bin is about 50 kb.

F32. The method of any one of embodiments F1 to F31, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

F33. The method of any one of embodiments F1 to F32, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

F34. The method of any one of embodiments F1 to F33, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

F35. The method of embodiment F34, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

F36. The method of embodiment F34 or F35, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the

counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

F37. The method of any one of embodiments F34 to F36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

F38. The method of any one of embodiments F34 to F36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

F39. The method of any one of embodiments F1 to F38, which comprises determining the gender of the fetus.

F40. The method of embodiment F39, wherein the gender of the fetus is determined before performing (b).

F41. The method of embodiment F39 or F40, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

F42. The method of embodiment F39, wherein the gender of the fetus is determined before performing (a).

F43. The method of any one of embodiments F39 to F42, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

F44. The method of any one of embodiments F39 to F42, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

F45. The method of any one of embodiments F1 to F44, which comprises determining the presence or absence of a Y chromosome in the fetus.

F46. The method of any one of embodiments F1 to F45, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

F47. The method of embodiment F46, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

F48. The method of any one of embodiments F1 to F46, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

F49. The method of any one of embodiments F1 to F48, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

F50. The method of embodiment F49, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

F51. The method of any one of embodiments F1 to F50, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

F52. The method of any one of embodiments F1 to F51, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

F53. The method of embodiment F52, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

F54. The method of any one of embodiments F1 to F51, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

G1. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and

generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation and experimental Y chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to:

the experimental X chromosome representation and an expected X chromosome representation, which expected X chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the X chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof, and

the experimental Y chromosome representation and an expected Y chromosome representation, which expected Y chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the Y chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

G1.1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or loading the sequencing apparatus with a modified variant

of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fraction of the fetal nucleic acid in the blood of the pregnant female by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) generate, from the counts in (a) an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and

generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation and experimental Y chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to:

the experimental X chromosome representation and an expected X chromosome representation, which expected X chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the X chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof, and

the experimental Y chromosome representation and an expected Y chromosome representation, which expected Y chromosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the Y chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

G2. The method of embodiment G1 or G1.1, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined in (c) according to (i) a ratio of the experimental X

chromosome representation and the expected X chromosome representation, and (ii) a ratio of the experimental Y chromosome representation and the expected Y chromosome representation.

G2.1. The method of embodiment G1, G1.1 or G2, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined according to equation AC.

G3. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof;

generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation and experimental Y chromosome representation, determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

G3.1. The method of embodiment G3, wherein the fraction of fetal nucleic acid in the blood of the pregnant female bearing a fetus having a chromosome aneuploidy is determined according to equation AB.

G3.2. The method of embodiment G3 or G3.1 wherein the fraction of fetal nucleic acid determined in (c)(i) and the experimental X and the experimental Y chromosome representation in (c)(ii) are derived from greater than about 500 subjects.

G4. The method of embodiment G3, wherein the relationship is a linear relationship.

G4.1. The method of embodiment G3 or G4, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

G4.2. The method of embodiment G3 or G4, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

G4.3. The method of any one of embodiments G4 to G4.2, wherein the relationship is:

$$F = k - r(MCR_x) + t(MCR_y)$$

wherein F is the fraction, MCR_Y is the experimental Y chromosome representation, MCR_X is the experimental X chromosome representation, k is an intercept from the linear relationship and r is a slope from the linear relationship.

G4.3.1. The method of embodiment G4.3, wherein the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female is determined by the relationship in G4.3, the experimental X chromosome representation and the experimental Y chromosome representation.

G4.4. The method of any one of embodiments G3 to G4.3.1, wherein the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that comprises use of sequence reads mapped to genomic sections of a reference genome.

G4.5. The method of any one of embodiments G3 to G4.3.1, wherein the fraction of fetal nucleic acid determined for nucleic acid from nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that does not utilize sequence reads mapped to genomic sections of a reference genome.

G4.6. The method of embodiment G4.5, wherein the process comprises mass spectrometry.

G4.7. The method of any one of embodiments G3 to G4.6, wherein the pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is different from the pregnant female in (a).

G4.7.1 The method of embodiment G4.7, wherein the pregnant female bearing a fetus having a chromosome aneuploidy is bearing a male fetus.

G4.8. The method of any one of embodiments G3 to G4.7.1, wherein the nucleic acid from the blood of the pregnant female in (c)(i) is circulating cell-free nucleic acid.

G5. The method of any one of embodiments G1 to G4.8, comprising normalizing the counts mapped to the genomic sections of the reference genome, thereby providing normalized counts of the genomic sections of the reference genome in a genomic section; and which counts in G1(b) and G3(b) are normalized counts.

G5.1 The method of any one of embodiments G1 to G5, wherein the counts in G1(b) and G3(b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

G6. The method of embodiment G5 or G5.1, wherein the normalized counts are provided by a normalization module.

G7. The method of any one of embodiments G1 to G6, wherein the experimental X chromosome representation in G1(b) and G3(b) is provided by a representation module and the expected X chromosome representation in G1(c) and G3(c) is determined by an expected representation module.

G8. The method of any one of embodiments G1 to G7, wherein the experimental Y chromosome representation in G1(b) and G3(b) is provided by a representation module and the expected Y chromosome representation in G1(c) and G3(c) is determined by an expected representation module.

G9. The method of any one of embodiments G1 to G8, wherein the fetal fraction in G1(c) and G3(c) is provided by a fetal fraction module.

G10. The method of any one of embodiments G3 to G9, wherein the relationship is provided by a relationship module.

G11. The method of any one of embodiments G7 to G10, wherein the normalized mapped counts are transferred to the experimental representation module and expected representation module from the normalization module.

G12. The method of any one of embodiments G8 to G11, wherein the normalized mapped counts are transferred to the Y experimental module and Y expected module from the normalization module.

G13. The method of any one of embodiments G9 to G12, wherein the experimental X chromosome representation is transferred to the fetal fraction module from the experimental representation module and the expected X chromosome representation is transferred to the fetal fraction module from the expected representation module.

G14. The method of any one of embodiments G9 to G12, wherein the experimental Y chromosome representation is transferred to the fetal fraction module from the Y experimental module and the expected Y chromosome representation is transferred to the fetal fraction module from the Y expected module.

G15. The method of any one of embodiments G1 to G14, which comprises obtaining nucleic acid sequence reads.

G16. The method of embodiment G15, wherein the nucleic acid sequence reads are generated by a sequencing module.

G16.1 The method of embodiment G15 or G16, wherein the nucleic acid sequencing reads are generated by massively parallel sequencing (MPS).

G17. The method of embodiment G15 or G16, which comprises mapping the nucleic acid sequence reads to the genomic sections of the reference genome.

G18. The method of embodiment G17, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

G19. The method of any one of embodiments G1 to G18, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

G20. The method of any one of embodiments G18 to G19, wherein the sequence reads are transferred to the mapping module from the sequencing module.

G21. The method of any one of embodiments G19 to G20, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

G22. The method of any one of embodiments G19 to G21, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

G23. The method of any one of embodiments G1 to G22, wherein an apparatus comprises one or more of a sequencing module, sequence receiving module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module or logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

G24. The method of embodiment G23, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

G24.1. The method of any one of embodiments G19 to G24, wherein a second apparatus comprises the mapping module and the counting module.

G25. The method of any one of embodiments G16 to G24.1, wherein a third apparatus comprises the sequencing module.

G25.1. The method of any one of embodiments G24 to G25, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

G26. The method of any one of embodiments G5 to G25.1, wherein the counts that are normalized are raw counts.

G27. The method of any one of embodiments G5 to G26, wherein the counts that are normalized are filtered.

G27.1 The method of any one of embodiments G5 to G26, wherein the counts that are normalized are not filtered.

G28. The method of any one of embodiments G1 to G27, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

G29. The method of any one of embodiments G1 to G28, wherein the genomic sections of the reference genome are one or more bins.

G30. The method of embodiment G29, wherein each bin is of about an equal number of contiguous nucleotides.

G31. The method of embodiment G29 or G30, wherein each bin is about 50 kb.

G32. The method of any one of embodiments G1 to G31, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

G33. The method of any one of embodiments G1 to G32, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

G34. The method of any one of embodiments G1 to G33, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

G35. The method of embodiment G34, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

G36. The method of embodiment G34 or G35, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

G37. The method of any one of embodiments G34 to G36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

G38. The method of any one of embodiments G34 to G36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

G39. The method of any one of embodiments G1 to G38, which comprises determining the gender of the fetus.

G40. The method of embodiment G39, wherein the gender of the fetus is determined before performing (b).

G41. The method of embodiment G39 or G40, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

G42. The method of embodiment G39, wherein the gender of the fetus is determined before performing (a).

G43. The method of any one of embodiments G39 to G42, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

G44. The method of any one of embodiments G39 to G42, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

G45. The method of any one of embodiments G1 to G44, which comprises determining the presence or absence of a Y chromosome in the fetus.

G46. The method of any one of embodiments G1 to G45, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

G47. The method of embodiment G46, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

G48. The method of any one of embodiments G1 to G46, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

G49. The method of any one of embodiments G1 to G48, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

G50. The method of embodiment G49, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

G51. The method of any one of embodiments G1 to G50, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

G52. The method of any one of embodiments G1 to G51, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

G53. The method of embodiment G52, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

G54. The method of any one of embodiments G1 to G51, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

H1. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome;

(b) generating an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental affected chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental affected autosome representation and an expected affected autosome representation, which expected affected autosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the affected autosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

H1.1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected

autosome, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fraction of the fetal nucleic acid in the blood of the pregnant female by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) generate, from the counts in (a) an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental affected chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental affected autosome representation and an expected affected autosome representation, which expected affected autosome representation is a ratio of (i) the number of the genomic sections of the reference genome in the affected autosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

H2. The method of embodiment H1 or H1.1, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined in (c) according to a ratio of the experimental affected autosome representation and the expected a e representation.

H2.1. The method of embodiment H1, H1.1 or H2, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined according to equation AB.

H3. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the

blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome;

(b) generating an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental affected chromosome representation, determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

H3.1. The method of embodiment H3, wherein the fraction of fetal nucleic acid in the blood of the pregnant female bearing a fetus having a chromosome aneuploidy is determined according to equation AB.

H3.2. The method of embodiment H3 or H3.1 wherein the fraction of fetal nucleic acid determined in (c)(i) and the experimental affected autosome representation in (c)(ii) are derived from greater than about 500 subjects.

H4. The method of embodiment H3, wherein the relationship is a linear relationship.

H4.1. The method of embodiment H3 or H4, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

H4.2. The method of embodiment H3 or H4, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

H4.3. The method of any one of embodiments H4 to H4.2, wherein the relationship is:

$$F = k - r(MCR_n),$$

wherein F is the fraction, MCR*n* is the experimental affected autosome representation, k is an intercept from the linear relationship and r is a slope from the linear relationship.

H4.3.1. The method of embodiment H4.3, wherein the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female is determined by the relationship in H4.3 and the experimental affected autosome representation.

H4.4. The method of any one of embodiments H3 to H4.3, wherein the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that comprises use of sequence reads mapped to genomic sections of a reference genome.

H4.5. The method of any one of embodiments H3 to H4.3, wherein the fraction of fetal nucleic acid determined for nucleic acid from nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that does not utilize sequence reads mapped to genomic sections of a reference genome.

H4.6. The method of embodiment H4.5, wherein the process comprises mass spectrometry.

H4.7. The method of any one of embodiments H3 to H4.6, wherein the pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is different from the pregnant female in (a).

H4.8. The method of any one of embodiments H3 to H4.7, wherein the nucleic acid from the blood of the pregnant female in (c)(i) is circulating cell-free nucleic acid.

H5. The method of any one of embodiments H1 to H4, comprising normalizing the counts mapped to the genomic sections of the reference genome, thereby providing normalized counts of the genomic sections of the reference genome in a genomic section; and which counts in H1(b) and H3(b) are normalized counts.

H5.1 The method of any one of embodiments H1 to H5, wherein the counts in H1(b) and H3(b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

H6. The method of embodiment H5, wherein the normalized counts are provided by a normalization module.

H7. The method of any one of embodiments H1 to H6, wherein the experimental affected autosome representation in H1(b) and H3(b) is provided by a representation module.

H8. The method of any one of embodiments H1 to H7, wherein the expected affected autosome representation in H1(c) and H3(c) is determined by an expected representation module.

H9. The method of any one of embodiments H2 to H8, wherein the fetal fraction is provided by a fetal fraction module.

H10. The method of any one of embodiments H3 to H9, wherein the relationship is provided by a relationship module.

H11. The method of any one of embodiments H7 to H10, wherein the normalized mapped counts are transferred to the autosome experimental module from the normalization module.

H12. The method of any one of embodiments H8 to H11, wherein the normalized mapped counts are transferred to the autosome expected module from the normalization module.

H13. The method of any one of embodiments H9 to H12, wherein the experimental affected autosome representation is transferred to the fetal fraction module from the autosome experimental module.

H14. The method of any one of embodiments H9 to H12, wherein the expected affected autosome representation is transferred to the fetal fraction module from the autosome expected module.

H15. The method of any one of embodiments H1 to H14, which comprises obtaining nucleic acid sequence reads.

H16. The method of embodiment H15, wherein the nucleic acid sequence reads are generated by a sequencing module.

H16.1 The method of embodiment H15 or H16, wherein the nucleic acid sequencing reads are generated by massively parallel sequencing (MPS).

H17. The method of embodiment H15 or H16, which comprises mapping the nucleic acid sequence reads to the genomic sections of the reference genome.

H18. The method of embodiment H17, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

H19. The method of any one of embodiments H1 to H18, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

H20. The method of any one of embodiments H18 to H19, wherein the sequence reads are transferred to the mapping module from the sequencing module.

H21. The method of any one of embodiments H19 to H20, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

H22. The method of any one of embodiments H19 to H21, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

H23. The method of any one of embodiments H1 to H22, wherein an apparatus comprises one or more of a sequencing module, sequence receiving module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module or logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

H24. The method of embodiment H23, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

H24.1. The method of any one of embodiments H20 to H24.1, wherein a second apparatus comprises the mapping module and the counting module.

H25. The method of any one of embodiments H16 to H24.1, wherein a third apparatus comprises the sequencing module.

H25.1. The method of any one of embodiments H24 to H25, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

H26. The method of any one of embodiments H5 to H25.1, wherein the counts that are normalized are raw counts.

H27. The method of any one of embodiments H5 to H26, wherein the counts that are normalized are filtered.

H27.1 The method of any one of embodiments H5 to H26, wherein the counts that are normalized are not filtered.

H28. The method of any one of embodiments H1 to H27, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

H29. The method of any one of embodiments H1 to H28, wherein the genomic sections of the reference genome are one or more bins.

H30. The method of embodiment H29, wherein each bin is of about an equal number of contiguous nucleotides.

H31. The method of embodiment H29 or H30, wherein each bin is about 50 kb.

H32. The method of any one of embodiments H1 to H31, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

H33. The method of any one of embodiments H1 to H32, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

H33.1. The method of any one of embodiments H1 to H33, comprising determining the presence or absence of a fetal aneuploidy.

H34. The method of any one of embodiments H1 to H33.1, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

H35. The method of embodiment H34, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

H36. The method of embodiment H34 or H35, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

H37. The method of any one of embodiments H34 to H36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

H38. The method of any one of embodiments H34 to H36, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

H39. The method of any one of embodiments H1 to H38, which comprises determining the gender of the fetus.

H40. The method of embodiment H39, wherein the gender of the fetus is determined before performing (b).

H41. The method of embodiment H39 or H40, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

H42. The method of embodiment H39, wherein the gender of the fetus is determined before performing (a).

H43. The method of any one of embodiments H39 to H42, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

H44. The method of any one of embodiments H39 to H42, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

H45. The method of any one of embodiments H1 to H44, which comprises determining the presence or absence of a Y chromosome in the fetus.

H46. The method of any one of embodiments H1 to H45, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

H47. The method of embodiment H46, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

H48. The method of any one of embodiments H1 to H46, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

H49. The method of any one of embodiments H1 to H48, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

H50. The method of embodiment H49, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

H51. The method of any one of embodiments H1 to H50, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

H52. The method of any one of embodiments H1 to H51, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

H53. The method of embodiment H52, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

H54. The method of any one of embodiments H1 to H51, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

I1. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental X chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

I1.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to genomic sections of a reference genome, and generate counts of the sequence reads mapped to the genomic sections;

(b) from the counts in (a), generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

I2. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental X chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the

fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

I3. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

J1. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

J1.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to genomic sections of a reference genome, and generate counts of the sequence reads mapped to the genomic sections;

(b) from the counts in (a), generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

J2. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

J3. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental Y chromosome representation.

K1. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and

generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic

sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental X and the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

K1.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to genomic sections of a reference genome, and generate counts of the sequence reads mapped to the genomic sections;

(b) from the counts in (a), generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and

generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X and the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

K2. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and

generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental X and the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

K3. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generate an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof, and

generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X and the experimental Y chromosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X and the experimental Y chromosome representation.

L1. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental affected autosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

L1.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to genomic sections of a reference genome, and generate counts of the sequence reads mapped to the genomic sections;

(b) from the counts in (a), generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental affected autosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

L2. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome; and

which instructions executable by the one or more processors are configured to:

(a) generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) from the experimental affected autosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

L3. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) access counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus bearing a trisomy of an autosome, which autosome is an affected autosome;

(b) generate an experimental affected autosome representation, which experimental affected autosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the affected autosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental affected autosome representation, determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) a fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental affected autosome representation.

M1. A method for determining fetal ploidy according to nucleic acid sequence reads for a test sample obtained from a pregnant female, comprising:

(a) obtaining counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from sample nucleic acid;

(b) calculating a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels;

(c) determining a fraction of fetal nucleic acid in the test sample according to calculated genomic section levels for a first subset of portions of the reference genome;

(d) determining fetal ploidy according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (c).

M1.1. A method for determining fetal ploidy according to nucleic acid sequence reads for a test sample obtained from a pregnant female, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the

one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fetal ploidy by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to portions of a reference genome and generate counts of the sequence reads mapped to the portions;

(b) calculate, from the counts in (a), a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels;

(c) determine a fraction of fetal nucleic acid in the test sample according to calculated genomic section levels for a first subset of portions of the reference genome; and

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (c).

M2. The method of embodiment M1 or M1.1, wherein the calculating in (b) comprises:

(1) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and

(2) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

M3. The method of embodiment M1, M1.1 or M2, wherein the fraction of fetal nucleic acid determined in (c) is determined by any one of embodiments A1 to A71, E1 to E54, F1 to F54, G1 to G54, or H1 to H54.

M4. The method of any one of embodiments M1 to M3, wherein the first subset of portions of the reference genome are portions of a sex chromosome or a segment thereof.

M5. The method of any one of embodiments M1 to M4, wherein the first subset of portions of the reference genome are portions of a Y chromosome or a segment thereof.

M6. The method of any one of embodiments M1 to M5, wherein the second subset of portions of the reference genome are portions of one or more autosomes or a segment thereof.

M7. The method of embodiments M6, wherein the one or more autosomes are selected from chromosomes 18, 13 and 21.

M8. The method of any one of embodiments M1 to M7, wherein the second subset of portions of the reference genome are portions of an X chromosome or a segment thereof.

M9. The method of any one of embodiments M1 to M8, wherein the second subset of portions of the reference genome are portions of all chromosomes or a segment thereof, from which counts were obtained.

M10. The method of any one of embodiments M1 to M9, wherein the fetal ploidy in (d) is determined according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (c) and maternal ploidy.

M11. The method of any one of embodiments M1 to M10, wherein the fetal ploidy in (d) is determined according to y_i where y_i represents calculated genomic section levels of portions i of a reference genome.

M12. The method of any one of embodiments M1 to M11, wherein the fraction of fetal nucleic acid determined in (c) is fixed at its determined value and fetal ploidy in (d) is determined from the equation below or a derivation thereof:

$$y_i = (1 - F)M_i f_i + FXf_i \quad ,$$

where F represents the fetal fraction, X represents the fetal ploidy, and M_i represents maternal ploidy assigned to each portion i.

M13. The method of any one of embodiments M1 to M12, wherein the fetal fraction determined in (c) is fixed at its determined value and fetal ploidy in (d) is varied to optimize the sum of squared residuals.

M14. The method of any one of embodiments M1 to M13, wherein the fetal ploidy in (d) is determined according to equation 20 below:

$$X = \frac{\sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^{N} \frac{M f_i^2}{\sigma_i^2}}{F \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}}.$$

M15. The method of any one of embodiments M1 to M14, wherein the fetal ploidy in (d) is determined according to equation 21 below:

$$X = \frac{\Sigma_{fy} - (1-F)\Sigma_{ff}}{F\Sigma_{ff}} = \frac{\Sigma_{fy}}{F\Sigma_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Sigma_{fy}}{\Sigma_{ff}} - 1\right).$$

M16. The method of any one of embodiments M1 to M15, comprising determining the presence or absence of a fetal chromosome aneuploidy according to the fetal ploidy determined in (d).

M17. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample obtained from the blood of a pregnant female bearing a fetus; and

which instructions executable by the one or more processors are configured to:

(a) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels;

(b) determine a fraction of fetal nucleic acid in the test sample according to calculated genomic section levels for a first subset of portions of the reference genome;

(c) determine fetal ploidy according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (b).

M18. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) access counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample;

(b) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels;

(c) determine a fraction of fetal nucleic acid in the test sample according to calculated genomic section levels for a first subset of portions of the reference genome;

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (c).

M19. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from test sample obtained from a pregnant female bearing a fetus; and

which instructions executable by the one or more processors are configured to: (a) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels;

(b) determine a fraction of fetal nucleic acid in the test sample according to calculated genomic section levels for a first subset of portions of the reference genome;

(c) determine fetal ploidy according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (b).

M19.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to portions of a reference genome, and generate counts of the sequence reads mapped to the portions;

(b) from the counts in (a), calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels;

(c) determine a fraction of fetal nucleic acid in the test sample according to calculated genomic section levels for a first subset of portions of the reference genome;

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a second subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (c).

N1. A method for determining fetal ploidy according to nucleic acid sequence reads, comprising:

(a) determining a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus;

(b) obtaining counts of sequence reads mapped to portions of a reference genome, which sequence reads are from the nucleic acid in the sample;

(c) calculating a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(d) determining fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

N1.1. A method for determining fetal ploidy according to nucleic acid sequence reads for a test sample obtained from a pregnant female, comprising:

(a) determining a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus;

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the fetal ploidy by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(b) align the sequence reads to portions of a reference genome and generate counts of the sequence reads mapped to the portions;

(c) calculate, from the counts in (b), a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(e) determining fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

N2. The method of embodiment N1 or N1.1, wherein the fetal fraction is determined from a first part of the test sample and the genomic section levels are determined from a second part of the test sample.

N3. The method of embodiment N1, N1.1 or N2, wherein calculating the genomic section level for each of the portions of the reference genome comprises normalizing counts of reads mapped to the reference genome according to guanine and cytosine (GC) content for each of the portions.

N4. The method of embodiment N3, comprising:

(1) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and

(2) calculating the genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

N5. The method of any one of embodiments N1 to N3, wherein the subset of portions of the reference genome in (d)(i) is portions of one or more autosomes or segment thereof.

N6. The method of embodiment N5, wherein the subset is portions of one autosome or segment thereof.

N7. The method of embodiment N5 or N6, wherein the autosomes are chosen from chromosome 13, chromosome 18 and chromosome 21.

N8. The method of any one of embodiments N1 to N7, wherein the subset of portions of the reference genome are portions of all autosomes or segment thereof.

N9. The method of any one of embodiments N1 to N7, wherein the subset of portions of the reference genome are portions of all chromosomes or segment thereof.

N10. The method of embodiment N8 or N9, wherein the autosomes or segment thereof, or chromosomes or segment thereof, are those from which counts were obtained.

N11. The method of any one of embodiments N1 to N10, wherein the subset of portions of the reference genome are portions of an X chromosome or a segment thereof.

N12. The method of any one of embodiments N1 to N11, comprising determining a reference count.

N12.1. The method of embodiment N12, wherein the reference count is determined according to calculated genomic section levels for a subset of portions of the reference genome for one or more pregnant female bearing a fetus wherein the subset of portions of the reference genome are known to be euploid.

N12.2. The method of embodiment N12, wherein the reference count is not determined from the sample.

N12.3. The method of embodiment N12, wherein the reference count is determined from the same subset of portions of the reference genome as in (d).

N12.4. The method of any one of embodiments N12 to N12.3, wherein the reference count is normalized by bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof.

N12.5. The method of any one of embodiments N12 to N12.4, wherein the subset of portions of the reference genome are known to be euploid for the fetus.

N12.6. The method of any one of embodiments N12 to N12.5, wherein the subset of portions of the reference genome are known to be euploid for the mother and the fetus.

N12.7. The method of any one of embodiments N12 to N12.6, wherein the fetal ploidy in (d) is determined according to the reference count.

N12.8. The method of any one of embodiments N1 to N12.7, comprising determining a maternal ploidy.

N12.9. The method of embodiment N12.8, wherein fetal ploidy in (d) is determined according the maternal ploidy.

N12.10. The method of embodiment N12.8 or N12.9, wherein the maternal ploidy is about 1.

N13. The method of any one of embodiments N1 to N12.10, wherein the fetal ploidy is determined in (d) according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a) and (iii) a maternal ploidy.

N13.1. The method of any one of embodiments N1 to N13, wherein the fetal ploidy is determined according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a), (iii) a maternal ploidy, (iv) the reference count and (v) an uncertainty value σ for the reference count.

N13.2. The method of embodiment N13.1, wherein the fraction of fetal nucleic acid determined in (a) is fixed at its determined value and fetal ploidy *X* is determined according to Equation 8 below, or a derivation thereof:

$$y_i = (1 - F)M_i f_i + FXf_i \qquad (8)$$

where y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, *X* represents the fetal ploidy, and M_i represents the maternal ploidy of portion *i*.

N13.3. The method of N13.2, comprising determining the sum of squared residuals according to equation (8) and for multiple bins *i* for a subset of portions of the reference genome.

N14. The method of embodiment N13.2 or N13.3, wherein the fetal fraction is fixed at a value determined in (a) and the fetal ploidy is varied to optimize the sum of squared residuals according to equation (8) or a variation thereof.

N15. The method of embodiment N14, comprising determining a linear regression according to the sum of square residuals.

N16. The method of any one of embodiments N1 to N15, wherein the fetal ploidy is determined according to the reference count and an uncertainty value σ for the reference count.

N17. The method of embodiment N16, wherein the fetal ploidy is determined according to Equation 20 below:

$$X = \frac{\sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^{N} \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}} \quad . \qquad (20)$$

wherein y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, σ represents the uncertainty value for f_i, *X* represents the fetal ploidy, and M_i represents the maternal ploidy of portion *i*.

N18. The method of any one of embodiments N1 to N17, wherein the fetal ploidy is determined according to Equation 20, wherein the maternal ploidy is 1.

N19. The method of embodiment N18, wherein the fetal ploidy is determined according to Equation 21 below:

$$X = \frac{\Xi_{fy} - (1-F)\Xi_{ff}}{F\Xi_{ff}} = \frac{\Xi_{fy}}{F\Xi_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Xi_{fy}}{\Xi_{ff}} - 1\right) \quad (21)$$

wherein $\Xi_{ff} = \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}$, $\Xi_{fy} = \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2}$, y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, σ represents the uncertainty value for f_i, and *X* represents the fetal ploidy.

N20. The method of any one of embodiments N1 to N19, comprising determining the presence or absence of a fetal chromosome aneuploidy according to the fetal ploidy determined in (d).

N21. The method of embodiment N20, wherein a fetal ploidy of about 1.4 or greater indicates the presence of a fetal chromosome aneuploidy.

N21.1. The method of embodiment N20, wherein a fetal ploidy of about 1.4 to about 1.8 indicates the presence of a fetal chromosome aneuploidy.

N21.2. The method of embodiment N20, wherein a fetal ploidy of about 1.3 or less indicates the absence of a fetal chromosome aneuploidy.

N21.3. The method of embodiment N20, wherein a fetal ploidy of 1.2 or less indicates the absence of a fetal chromosome aneuploidy.

N21.4. The method of embodiment N20, wherein a fetal ploidy of about 1.3 to about 0.8 indicates the absence of a fetal chromosome aneuploidy.

N21.5. The method of embodiment N20, wherein a fetal ploidy of about 1.2 to about 0.8 indicates the absence of a fetal chromosome aneuploidy.

N21.6. The method of any one of embodiments N20 to N21.4, wherein the fetal chromosome aneuploidy is a trisomy.

N21.7. The method of embodiment N21.6, wherein the trisomy is selected from a trisomy of chromosome 13, 18 and 21.

N22. The method of any one of embodiments N1 to N21.7, wherein determining the fraction of fetal nucleic acid comprises analyzing all or a subset of the sequence reads.

N23. The method of embodiment N22, wherein determining the fraction of fetal nucleic acid comprises analyzing the sequence reads mapped to all or a subset of the portions of the reference genome.

N24. The method of embodiment N22 or N23, wherein determining the fraction of fetal nucleic acid comprises analyzing the calculated genomic sections levels for all or a subset of portions of the reference genome.

N25. The method of embodiment N24, wherein determining the fraction of fetal nucleic acid comprises analyzing the calculated genomic sections levels for a subset of portions of the reference genome, which subset is a first subset and the subset in (d) is a second subset.

N26. The method of embodiment N25, wherein the first subset is different than the second subset.

N27. The method of embodiment N25 or N26, wherein the first subset of portions of the reference genome is portions of a sex chromosome or a segment thereof.

N28. The method of embodiment N25 or N26, wherein the first subset of portions of the reference genome is portions of a Y chromosome or a segment thereof.

N29. The method of embodiment N27 or N28, wherein the fetal fraction is determined according to a method of any one of embodiments A1 to A71, E1 to E54, F1 to F54, G1 to G54, H1 to H54, AA1 to AA39, AB1 to AB35, AC1 to AC35 or AE1 to AE41.

N30. The method of any one of embodiments N1 to N29, wherein the sequence reads are obtained by a massively parallel sequencing (MPS) process.

N31. The method of embodiment N30, which MPS process comprises use of a flowcell.

N32. The method of embodiment N31, wherein the sequence reads used for determining the fetal fraction and determining fetal ploidy are obtained in part using the same flowcell.

N33. The method of any one of embodiments N1 to N32, wherein determining the fraction of fetal nucleic acid comprises analyzing one or more loci in sample nucleic acid, wherein at least one of the one or more loci vary between fetal nucleic acid and maternal nucleic acid.

N34. The method of embodiment N33, wherein an amount of the one or more loci is determined.

N35. The method of embodiment N33 or N34, wherein a nucleotide sequence is determined for the one or more loci.

N36. The method of embodiment N35, wherein the sequence is obtained by a massively parallel sequencing (MPS) process.

N37. The method of embodiment N36, which MPS process comprises use of a flowcell.

N38. The method of any one of embodiments N33 to N37, wherein the one or more loci comprise one or more polymorphic sites.

N39. The method of embodiment N38, comprising:

(1) enriching nucleic acid in a first part of the test sample for a plurality of polymorphic sites;

(2) obtaining nucleotide sequences for some or all of the polymorphic sites by a sequencing process;

(3) analyzing the nucleotide sequences of (2); and

(4) determining the fraction of fetal nucleic acid based on the analysis of (3), wherein the polymorphic sites and number thereof result in at least five polymorphic sites being informative for determining the fetal fraction for at least 90% of samples.

N40. The method of embodiment N38 or N39, wherein one or more polymorphic sites comprise one or more single nucleotide polymorphisms (SNPs).

N41. The method of any one of embodiments N33 to N40, wherein the one or more loci comprise one or more methylation regions.

N42. The method of embodiment N41, comprising:

(1) contacting the test sample with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid; and

(2) determining the fraction of fetal nucleic acid in the sample based on the differentially modified nucleic acid.

N43. The method of embodiment N42, comprising:

contacting under amplification conditions the differentially modified nucleic acid with:

(i) a first set of amplification primers that specifically amplify a first region in sample nucleic acid comprising one or more loci that are differentially methylated between the fetal nucleic acid and maternal nucleic acid, and

(ii) a second set of amplification primers that amplify a second region in the sample nucleic acid allowing for a determination of total nucleic acid in the sample, wherein the first region and the second region are different, thereby generating fetal nucleic acid amplification products and total nucleic acid amplification products.

N44. The method of embodiment N42, comprising:

contacting under amplification conditions the differentially modified nucleic acid with:

(i) a first set of amplification primers that specifically amplify a first region in sample nucleic acid comprising one or more loci that are differentially methylated between the fetal nucleic acid and maternal nucleic acid; and

(ii) a predetermined copy number of one or more first competitor oligonucleotides that compete with the first region for hybridization of primers of the first amplification primer set, thereby generating fetal nucleic acid amplification products and competitor amplification products.

N45. The method of embodiment N43 or N44 comprising:

(i) incorporating adaptor oligonucleotides into the amplification products thereby generating adaptor-modified amplification products;

(ii) obtaining nucleotide sequences of the adaptor-modified amplification products by a sequencing process, thereby generating sequence reads; and

(iii) quantifying the sequence reads of the adaptor-modified amplification products.

N46. The method of embodiment N45, wherein determining the fraction of fetal nucleic acid is based on a quantification of the sequence reads in (iii) of N45.

N47. The method of embodiment N44 comprising:

(i) incorporating adaptor oligonucleotides into the amplification products thereby generating adaptor-modified amplification products;

(ii) obtaining nucleotide sequences of the adaptor-modified amplification products by a sequencing process, thereby generating sequence reads;

(iii) quantifying the sequence reads of the adaptor-modified amplification products; and

(iv) determining the copy number of fetal nucleic acid in the sample based on a quantification of the sequence reads of the adaptor-modified amplification products and the amount of competitor oligonucleotide used.

N48. The method of embodiment N47, wherein the fraction of fetal nucleic acid is determined according to the copy number of fetal nucleic acid.

N49. The method of any one of embodiments N42 to N48, wherein the one or more agents are methylation sensitive restriction enzymes.

O1. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample obtained from a pregnant female bearing a fetus; and

which instructions executable by the one or more processors are configured to:

(a) determine a fraction of fetal nucleic acid in the test sample;

(b) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(c) determine fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

O2. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) determine a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus;

(b) access counts of sequence reads mapped to portions of a reference genome, which sequence reads are from the nucleic acid in the sample;

(c) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

O3. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a test sample obtained from a pregnant female bearing a fetus; and

which instructions executable by the one or more processors are configured to: (a) determine a fraction of fetal nucleic acid in the test sample;

(b) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(c) determine fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

AA1. A method for determining the amount of fetal nucleic acid in a sample comprising:

(a) contacting a sample nucleic acid with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid;

(b) contacting under amplification conditions the differentially modified sample nucleic acid with:

(i) a first set of amplification primers that specifically amplify a first region in sample nucleic acid comprising one or more loci that are differentially methylated between the fetal nucleic acid and maternal nucleic acid, and

(ii) a second set of amplification primers that amplify a second region in the sample nucleic acid allowing for a determination of total nucleic acid in the sample, wherein the first region and the second region are different, thereby generating fetal nucleic acid amplification products and total nucleic acid amplification products;

(c) incorporating adaptor oligonucleotides into the amplification products in (b); thereby generating adaptor-modified amplification products;

(d) obtaining nucleotide sequences of the adaptor-modified amplification products in (c) by a sequencing process, thereby generating sequence reads;

(e) quantifying the sequence reads; and

(f) determining the amount of fetal nucleic acid in the sample based on a quantification of the sequence reads in (e).

AA2. The method of embodiment AA1, wherein the first region comprises one or more loci which each contain a restriction site for a methylation-sensitive restriction enzyme.

AA3. The method of embodiment AA2, wherein the one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid comprise one or more methylation sensitive restriction enzymes.

AA4. The method of embodiment AA2 or AA3, wherein the second region comprises one or more loci which do not contain a restriction site for a methylation-sensitive restriction enzyme.

AA5. The method of embodiment AA1, wherein the one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid comprise bisulfite.

AA6. The method of any one of embodiments AA1 to AA5, wherein the adaptor oligonucleotides are incorporated into the amplification products by ligation.

AA7. The method of embodiment AA6, wherein the ligation is unidirectional ligation.

AA8. The method of any one of embodiments AA1 to AA5, wherein the adaptor oligonucleotides are incorporated into the amplification products using amplification primers comprising the adaptor oligonucleotide sequences.

AA9. The method of any one of embodiments AA1 to AA8, wherein the adaptor oligonucleotides comprise one or more index sequences.

AA10. The method of embodiment AA9, wherein the one or more index sequences comprise a sample-specific index.

AA11. The method of embodiment AA9, wherein the one or more index sequences comprise an aliquot-specific index.

AA12. The method of any one of embodiments AA1 to AA11, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-261, or a fragment thereof.

AA13. The method of embodiment AA12, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-89, or a fragment thereof.

AA14. The method of embodiment AA12, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:90-261, or a fragment thereof.

AA15. The method of embodiment AA12, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-59 and SEQ ID NOs:86-89, or a fragment thereof.

AA16. The method of embodiment AA12, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-59, or a fragment thereof.

AA17. The method of embodiment AA12, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NO:42, SEQ ID NO:52, SEQ ID NO:154, SEQ ID NO:158 and SEQ ID NO:163.

AA18. The method of any one of embodiments AA1 to AA17, wherein the sequencing process is a sequencing by synthesis method.

AA19. The method of any one of embodiments AA1 to AA18, wherein the sequencing process is a reversible terminator-based sequencing method.

AA20. The method of any one of embodiments AA1 to AA19, wherein the amount of fetal nucleic acid determined is the fraction of fetal nucleic acid in the sample based on the amount of each of the fetal nucleic acid amplification products and total nucleic acid amplification products.

AA21. The method of embodiment AA20, wherein the fraction of fetal nucleic acid is a ratio of fetal nucleic acid amplification product amount to total nucleic acid amplification product amount.

AA22. The method of any one of embodiments AA1 to AA21, further comprising contacting under amplification conditions the nucleic acid sample with a third set of amplification primers that amplify a third region in the sample nucleic acid allowing for a determination of the presence or absence of fetal specific nucleic acid.

AA23. The method of embodiment AA22, wherein the fetal specific nucleic acid is Y chromosome nucleic acid.

AA24. The method of embodiment AA23, wherein the third region comprises one or more loci within chromosome Y.

AA25. The method of any one of embodiments AA3 to AA24, further comprising contacting under amplification conditions the nucleic acid sample with a fourth set of amplification primers that amplify a fourth region in the sample nucleic acid allowing for a determination of the amount of digested or undigested nucleic acid, as an indicator of digestion efficiency.

AA26. The method of embodiment AA25, wherein the fourth region comprises one or more loci present in both fetal nucleic acid and maternal nucleic acid and unmethylated in both fetal nucleic acid and maternal nucleic acid.

AA27. The method of any one of embodiments AA1 to AA26, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more first competitor oligonucleotides that compete with the first region for hybridization of primers of the first amplification primer set.

AA28. The method of any one of embodiments AA1 to AA27, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more second competitor oligonucleotides that compete with the second region for hybridization of primers of the second amplification primer set.

AA29. The method of any one of embodiments AA22 to AA28, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more third competitor oligonucleotides that compete with the third region for hybridization of primers of the third amplification primer set.

AA30. The method of any one of embodiments AA25 to AA29, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more fourth competitor oligonucleotides that compete with the fourth region for hybridization of primers of the fourth amplification primer set.

AA31. The method of any one of embodiments AA27 to AA30, wherein the amount of fetal nucleic acid determined is the copy number of fetal nucleic acid based on the amount of competitor oligonucleotide used.

AA32. The method of any one of embodiments AA1 to AA26, wherein the amount of fetal nucleic acid determined is the copy number of fetal nucleic acid based on a quantification of sequence reads.

AA33. The method of any one of embodiments AA1 to AA32, wherein the sample nucleic acid is extracellular nucleic acid.

AA34. The method of any one of embodiments AA1 to AA33, wherein the nucleic acid sample is obtained from a pregnant female subject.

AA35. The method of embodiment AA34, wherein the subject is human.

AA36. The method of any one of embodiments AA1 to AA35, wherein the sample nucleic acid is from plasma or serum.

AA37. The method of any one of embodiments AA1 to AA36, wherein two or more independent loci in the first region are assayed.

AA38. The method of any one of embodiments AA1 to AA37, wherein the amount of fetal nucleic acid is substantially equal to the amount of fetal nucleic acid determined using a mass spectrometry method.

AA39. The method of any one of embodiments AA1 to AA38, wherein the amount of fetal nucleic acid is determined with an R^2 value of 0.97 or greater when compared to an amount of fetal nucleic acid determined using a mass spectrometry method.

AAB1. A method for determining the amount of fetal nucleic acid in a sample comprising:

(a) contacting a sample nucleic acid with one or more methylation sensitive restriction enzymes, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially digested sample nucleic acid;

(b) contacting under amplification conditions the digested sample nucleic acid with:

(i) a first set of amplification primers that specifically amplify a first region in sample nucleic acid comprising one or more loci that are differentially methylated between the fetal nucleic acid and maternal nucleic acid, and

(ii) a second set of amplification primers that amplify a second region in the sample nucleic acid allowing for a determination of total nucleic acid in the sample, wherein the first region and the second region are different, thereby generating fetal nucleic acid amplification products and total nucleic acid amplification products;

(c) incorporating adaptor oligonucleotides into the amplification products in (b); thereby generating adaptor-modified amplification products;

(d) obtaining nucleotide sequences of the adaptor-modified amplification products in (c) by a sequencing process, thereby generating sequence reads;

(e) quantifying the sequence reads; and

(f) determining the amount of fetal nucleic acid in the sample based on a quantification of the sequence reads in (e).

AAB2. The method of embodiment AB1, wherein the adaptor oligonucleotides are incorporated into the amplification products by ligation.

AB3. The method of embodiment AB2, wherein the ligation is unidirectional ligation.

AB4. The method of any one of embodiments AB1 to AB3, wherein the adaptor oligonucleotides are incorporated into the amplification products using amplification primers comprising the adaptor oligonucleotide sequences.

AB5. The method of any one of embodiments AB1 to AB4, wherein the adaptor oligonucleotides comprise one or more index sequences.

AB6. The method of embodiment AB5, wherein the one or more index sequences comprise a sample-specific index.

AB7. The method of embodiment AB5, wherein the one or more index sequences comprise an aliquot-specific index.

AB8. The method of any one of embodiments AB1 to AB7, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-261, or a fragment thereof.

AB9. The method of embodiment AB8, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-89, or a fragment thereof.

AB10. The method of embodiment AB8, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:90-261, or a fragment thereof.

AB11. The method of embodiment AB8, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-59 and SEQ ID NOs:86-89, or a fragment thereof.

AB12. The method of embodiment AB8, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-59, or a fragment thereof.

AB13. The method of embodiment AB8, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NO:42, SEQ ID NO:52, SEQ ID NO:154, SEQ ID NO:158 and SEQ ID NO:163.

AB14. The method of any one of embodiments AB1 to AB13, wherein the sequencing process is a sequencing by synthesis method.

AB15. The method of any one of embodiments AB1 to AB13, wherein the sequencing process is a reversible terminator-based sequencing method.

AB16. The method of any one of embodiments AB1 to AB15, wherein the amount of fetal nucleic acid determined is the fraction of fetal nucleic acid in the sample based on the amount of each of the fetal nucleic acid amplification products and total nucleic acid amplification products.

AB17. The method of embodiment AB16, wherein the fraction of fetal nucleic acid is a ratio of fetal nucleic acid amplification product amount to total nucleic acid amplification product amount.

AB18. The method of any one of embodiments AB1 to AB17, further comprising contacting under amplification conditions the nucleic acid sample with a third set of amplification primers that amplify a third region in the sample nucleic acid allowing for a determination of the presence or absence of fetal specific nucleic acid.

AB19. The method of embodiment AB18, wherein the fetal specific nucleic acid is Y chromosome nucleic acid.

AB20. The method of embodiment AB19, wherein the third region comprises one or more loci within chromosome Y.

AB21. The method of any one of embodiments AB1 to AB20, further comprising contacting under amplification conditions the nucleic acid sample with a fourth set of amplification primers that amplify a fourth region in the sample nucleic acid allowing for a determination of the amount of digested or undigested nucleic acid, as an indicator of digestion efficiency.

AB22. The method of embodiment AB21, wherein the fourth region comprises one or more loci present in both fetal nucleic acid and maternal nucleic acid and unmethylated in both fetal nucleic acid and maternal nucleic acid.

AB23. The method of any one of embodiments AB1 to AB22, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more first competitor oligonucleotides that compete with the first region for hybridization of primers of the first amplification primer set.

AB24. The method of any one of embodiments AB1 to AB23, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more second competitor oligonucleotides that compete with the second region for hybridization of primers of the second amplification primer set.

AB25. The method of any one of embodiments AB18 to AB24, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more third competitor oligonucleotides that compete with the third region for hybridization of primers of the third amplification primer set.

AB26. The method of any one of embodiments AB21 to AB25, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more fourth competitor oligonucleotides that compete with the fourth region for hybridization of primers of the fourth amplification primer set.

AB27. The method of any one of embodiments AB23 to AB26, wherein the amount of fetal nucleic acid determined is the copy number of fetal nucleic acid based on the amount of competitor oligonucleotide used.

AB28. The method of any one of embodiments AB1 to AB27, wherein the amount of fetal nucleic acid determined is the copy number of fetal nucleic acid based on a quantification of sequence reads.

AB29. The method of any one of embodiments AB1 to AB28, wherein the sample nucleic acid is extracellular nucleic acid.

AB30. The method of any one of embodiments AB1 to AB29, wherein the nucleic acid sample is obtained from a pregnant female subject.

AB31. The method of embodiment AB30, wherein the subject is human.

AB32. The method of any one of embodiments AB1 to AB31, wherein the sample nucleic acid is from plasma or serum.

AB33. The method of any one of embodiments AB1 to AB32, wherein two or more independent loci in the first region are assayed.

AB34. The method of any one of embodiments AB1 to AB33, wherein the amount of fetal nucleic acid is substantially equal to the amount of fetal nucleic acid determined using a mass spectrometry method.

AB35. The method of any one of embodiments AB1 to AB34, wherein the amount of fetal nucleic acid is determined with an R^2 value of 0.97 or greater when compared to an amount of fetal nucleic acid determined using a mass spectrometry method.

AC1. A method for determining the copy number of fetal nucleic acid in a sample comprising:

(a) contacting a sample nucleic acid with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises

differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid;

(b) contacting under amplification conditions the differentially modified sample nucleic acid with:

(i) a first set of amplification primers that specifically amplify a first region in sample nucleic acid comprising one or more loci that are differentially methylated between the fetal nucleic acid and maternal nucleic acid, and

(ii) a predetermined copy number of one or more first competitor oligonucleotides that compete with the first region for hybridization of primers of the first amplification primer set, thereby generating fetal nucleic acid amplification products and competitor amplification products;

(c) incorporating adaptor oligonucleotides into the amplification products in (b); thereby generating adaptor-modified amplification products;

(d) obtaining nucleotide sequences of the adaptor-modified amplification products in (c) by a sequencing process, thereby generating sequence reads;

(e) quantifying the sequence reads; and

(f) determining the copy number of fetal nucleic acid in the sample based on a quantification of the sequence reads in (e) and the amount of competitor oligonucleotide used.

AC2. The method of embodiment AC1, wherein the first region comprises one or more loci which each contain a restriction site for a methylation-sensitive restriction enzyme.

AC3. The method of embodiment AC2, wherein the one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid comprise one or more methylation sensitive restriction enzymes.

AC4. The method of embodiment AC1, wherein the one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid comprise bisulfite.

AC5. The method of any one of embodiments AC1 to AC4, wherein the adaptor oligonucleotides are incorporated into the amplification products by ligation.

AC6. The method of embodiment AC5, wherein the ligation is unidirectional ligation.

AC7. The method of any one of embodiments AC1 to AC4, wherein the adaptor oligonucleotides are incorporated into the amplification products using amplification primers comprising the adaptor oligonucleotide sequences.

AC8. The method of any one of embodiments AC1 to AC7, wherein the adaptor oligonucleotides comprise one or more index sequences.

AC9. The method of embodiment AC8, wherein the one or more index sequences comprise a sample-specific index.

AC10. The method of embodiment AC8, wherein the one or more index sequences comprise an aliquot-specific index.

AC11. The method of any one of embodiments AC1 to AC10, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-261, or a fragment thereof.

AC12. The method of embodiment AC11, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-89, or a fragment thereof.

AC13. The method of embodiment AC11, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:90-261, or a fragment thereof.

AC14. The method of embodiment AC11, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-59 and SEQ ID NOs:86-89, or a fragment thereof.

AC15. The method of embodiment AC11, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NOs:1-59, or a fragment thereof.

AC16. The method of embodiment AC11, wherein at least one of the one or more loci in the first region comprises a nucleotide sequence selected from among SEQ ID NO:42, SEQ ID NO:52, SEQ ID NO:154, SEQ ID NO:158 and SEQ ID NO:163.

AC17. The method of any one of embodiments AC1 to AC16, wherein the sequencing process is a sequencing by synthesis method.

AC18. The method of any one of embodiments AC1 to AC16, wherein the sequencing process is a reversible terminator-based sequencing method.

AC19 The method of any one of embodiments AC1 to AC18, further comprising contacting under amplification conditions the nucleic acid sample with a second set of amplification primers that amplify a second region in the sample nucleic acid allowing for a determination of total nucleic acid in the sample, wherein the first region and the second region are different.

AC20. The method of embodiment AC19, wherein the second region comprises one or more loci which do not contain a restriction site for a methylation-sensitive restriction enzyme.

AC21. The method of any one of embodiments AC1 to AC20, further comprising contacting under amplification conditions the nucleic acid sample with a third set of amplification primers that amplify a third region in the sample nucleic acid allowing for a determination of the presence or absence of fetal specific nucleic acid.

AC22. The method of embodiment AC21, wherein the fetal specific nucleic acid is Y chromosome nucleic acid.

AC23. The method of embodiment AC22, wherein the third region comprises one or more loci within chromosome Y.

AC24. The method of any one of embodiments AC3 to AC23, further comprising contacting under amplification conditions the nucleic acid sample with a fourth set of amplification primers that amplify a fourth region in the sample nucleic acid allowing for a determination of the amount of digested or undigested nucleic acid, as an indicator of digestion efficiency.

AC25. The method of embodiment AC24, wherein the fourth region comprises one or more loci present in both fetal nucleic acid and maternal nucleic acid and unmethylated in both fetal nucleic acid and maternal nucleic acid.

AC26. The method of any one of embodiments AC19 to AC25, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more second competitor oligonucleotides that compete with the second region for hybridization of primers of the second amplification primer set.

AC27. The method of any one of embodiments AC21 to AC26, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more third competitor oligonucleotides that compete with the third region for hybridization of primers of the third amplification primer set.

AC28. The method of any one of embodiments AC24 to AC27, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more fourth competitor oligonucleotides that compete with the fourth region for hybridization of primers of the fourth amplification primer set.

AC29. The method of any one of embodiments AC1 to AC28, wherein the sample nucleic acid is extracellular nucleic acid.

AC30. The method of any one of embodiments AC1 to AC29, wherein the nucleic acid sample is obtained from a pregnant female subject.

AC31. The method of embodiment AC30, wherein the subject is human.

AC32. The method of any one of embodiments AC1 to AC31, wherein the sample nucleic acid is from plasma or serum.

AC33. The method of any one of embodiments AC1 to AC32, wherein two or more independent loci in the first region are assayed.

AC34. The method of any one of embodiments AC1 to AC33, wherein the copy number of fetal nucleic acid is substantially equal to the copy number of fetal nucleic acid determined using a mass spectrometry method.

AC35. The method of any one of embodiments AC1 to AC34, wherein the copy number of fetal nucleic acid is determined with an R^2 value of 0.97 or greater when compared to a copy number of fetal nucleic acid determined using a mass spectrometry method.

AD1. A method for detecting the presence or absence of a fetal aneuploidy in a sample comprising:

(a) contacting a sample nucleic acid with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid;

(b) contacting under amplification conditions the differentially modified sample nucleic acid with:

(i) a first set of amplification primers that specifically amplify one or more loci in a target chromosome that are differentially methylated between the fetal nucleic acid and maternal nucleic acid, and

(ii) a second set of amplification primers that specifically amplify one or more loci in a reference chromosome that are differentially methylated between the fetal nucleic acid and maternal nucleic acid, thereby generating target chromosome amplification products and reference chromosome amplification products;

(c) incorporating adaptor oligonucleotides into the amplification products in (b); thereby generating adaptor-modified amplification products;

(d) obtaining nucleotide sequences of the adaptor-modified amplification products in (c) by a sequencing process, thereby generating sequence reads;

(e) quantifying the sequence reads; and

(f) detecting the presence or absence of a fetal aneuploidy in the sample based on a quantification of the sequence reads in (e).

AD2. The method of embodiment AD1, wherein the target chromosome comprises one or more loci which each contain a restriction site for a methylation-sensitive restriction enzyme.

AD3. The method of embodiment AD1 or AD2, wherein the reference chromosome comprises one or more loci which each contain a restriction site for a methylation-sensitive restriction enzyme.

AD4. The method of embodiment AD2 or AD3, wherein the one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid comprise one or more methylation sensitive restriction enzymes.

AD5. The method of embodiment AD1, wherein the one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid comprise bisulfite.

AD6. The method of any one of embodiments AD1 to AD5, wherein the adaptor oligonucleotides are incorporated into the amplification products by ligation.

AD7. The method of embodiment AD6, wherein the ligation is unidirectional ligation.

AD8. The method of any one of embodiments AD1 to AD5, wherein the adaptor oligonucleotides are incorporated into the amplification products using amplification primers comprising the adaptor oligonucleotide sequences.

AD9. The method of any one of embodiments AD1 to AD8, wherein the adaptor oligonucleotides comprise one or more index sequences.

AD10. The method of embodiment AD9, wherein the one or more index sequences comprise a sample-specific index.

AD11. The method of embodiment AD9, wherein the one or more index sequences comprise an aliquot-specific index.

AD12. The method of any one of embodiments AD1 to AD11, wherein at least one of the one or more loci in the target chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-261, or a fragment thereof.

AD13. The method of embodiment AD12, wherein at least one of the one or more loci in the target chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-89, or a fragment thereof.

AD14. The method of embodiment AD12, wherein at least one of the one or more loci in the target chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:90-261, or a fragment thereof.

AD15. The method of embodiment AD12, wherein at least one of the one or more loci in target chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-59 and SEQ ID NOs:86-89, or a fragment thereof.

AD16. The method of embodiment AD12, wherein at least one of the one or more loci in the target chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-59, or a fragment thereof.

AD17. The method of embodiment AD12, wherein at least one of the one or more loci in the target chromosome comprises a nucleotide sequence selected from among SEQ ID NO:42, SEQ ID NO:52, SEQ ID NO:154, SEQ ID NO:158 and SEQ ID NO:163.

AD18. The method of any one of embodiments AD1 to AD17, wherein at least one of the one or more loci in the reference chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-261, or a fragment thereof.

AD19. The method of embodiment AD18, wherein at least one of the one or more loci in the reference chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-89, or a fragment thereof.

AD20. The method of embodiment AD18, wherein at least one of the one or more loci in the reference chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:90-261, or a fragment thereof.

AD21. The method of embodiment AD18, wherein at least one of the one or more loci in reference chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-59 and SEQ ID NOs:86-89, or a fragment thereof.

AD22. The method of embodiment AD18, wherein at least one of the one or more loci in the reference chromosome comprises a nucleotide sequence selected from among SEQ ID NOs:1-59, or a fragment thereof.

AD23. The method of embodiment AD18, wherein at least one of the one or more loci in the reference chromosome comprises a nucleotide sequence selected from among SEQ ID NO:42, SEQ ID NO:52, SEQ ID NO:154, SEQ ID NO:158 and SEQ ID NO:163.

AD24. The method of any one of embodiments AD1 to AD23, wherein the sequencing process is a sequencing by synthesis method.

AD25. The method of any one of embodiments AD1 to AD23, wherein the sequencing process is a reversible terminator-based sequencing method.

AD26. The method of any one of embodiments AD1 to AD25, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more first competitor oligonucleotides that compete with the target chromosome for hybridization of primers of the first amplification primer set.

AD27. The method of any one of embodiments AD1 to AD26, further comprising contacting under amplification conditions the nucleic acid sample with a predetermined copy number of one or more second competitor oligonucleotides that compete with the reference chromosome for hybridization of primers of the second amplification primer set.

AD28. The method of any one of embodiments AD1 to AD27, wherein the sample nucleic acid is extracellular nucleic acid.

AD29. The method of any one of embodiments AD1 to AD28, wherein the nucleic acid sample is obtained from a pregnant female subject.

AD30. The method of embodiment AD29, wherein the subject is human.

AD31. The method of any one of embodiments AD1 to AD30, wherein the sample nucleic acid is from plasma or serum.

AD32. The method of any one of embodiments AD1 to AD31, wherein two or more independent loci in the target chromosome are assayed.

AD33. The method of any one of embodiments AD1 to AD32, wherein two or more independent loci in the reference chromosome are assayed.

AD34. The method of any one of embodiments AD1 to AD33, wherein the target chromosome is chromosome 13.

AD35. The method of any one of embodiments AD1 to AD33, wherein the target chromosome is chromosome 18.

AD36. The method of any one of embodiments AD1 to AD33, wherein the target chromosome is chromosome 21.

AE1. A method for determining fetal fraction in a sample comprising:

(a) enriching a sample nucleic acid for a plurality of polymorphic nucleic acid targets, which sample nucleic acid comprises fetal nucleic acid and maternal nucleic acid;

(b) obtaining nucleotide sequences for some or all of the nucleic acid targets by a sequencing process;

(c) analyzing the nucleotide sequences of (b); and

(d) determining fetal fraction based on the analysis of (c), wherein the polymorphic nucleic acid targets and number thereof result in at least five polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 90% of samples.

AE2. The method of embodiment AE1, wherein the enriching comprises amplifying the plurality of polymorphic nucleic acid targets.

AE3. The method of embodiment AE1 or AE2, wherein the enriching comprises generating amplification products in an amplification reaction.

AE4. The method of embodiment AE3, wherein the amplification reaction is performed in a single vessel.

AE5. The method of any one of embodiments AE1 to AE4, wherein the maternal genotype and the paternal genotype at each of the polymorphic nucleic acid targets are not known prior to (a).

AE5.1 The method of any one of embodiments AE1 to AE5, wherein polymorphic nucleic acid targets having a minor allele population frequency of about 40% or more are selected.

AE6. The method of any one of embodiments AE1 to AE5.1, comprising determining an allele frequency in the sample for each of the polymorphic nucleic acid targets.

AE7. The method of embodiment AE6, wherein determining which polymorphic nucleic acid targets are informative comprises identifying informative genotypes by comparing each allele frequency to one or more fixed cutoff frequencies.

AE7.1 The method of embodiment AE7, wherein the fixed cutoff for identifying informative genotypes from non-informative homozygotes is about a 1% or greater shift in allele frequency.

AE7.2 The method of embodiment AE7, wherein the fixed cutoff for identifying informative genotypes from non-informative homozygotes is about a 2% or greater shift in allele frequency.

AE7.3 The method of embodiment AE7, wherein the fixed cutoff for identifying informative genotypes from non-informative heterozygotes is about a 25% or greater shift in allele frequency.

AE7.4 The method of embodiment AE7, wherein the fixed cutoff for identifying informative genotypes from non-informative heterozygotes is about a 50% or greater shift in allele frequency.

AE8. The method of embodiment AE6, wherein determining which polymorphic nucleic acid targets are informative comprises identifying informative genotypes by comparing each allele frequency to one or more target-specific cutoff frequencies.

AE9. The method of embodiment AE8, wherein the one or more target-specific cutoff frequencies are determined for each polymorphic nucleic acid target.

AE10. The method of embodiment AE8 or AE9, wherein each target-specific cutoff frequency is determined based on the allele frequency variance for the corresponding polymorphic nucleic acid target.

AE11. The method of any one of embodiments AE6 to AE10, further comprising determining an allele frequency mean.

AE12. The method of embodiment AE11, wherein fetal fraction is determined based, in part, on the allele frequency mean.

AE13. The method of any one of embodiments AE1 to AE12, wherein the fetal genotype at one or more informative polymorphic nucleic acid targets is heterozygous.

AE14. The method of any one of embodiments AE1 to AE13, wherein the fetal genotype at one or more informative polymorphic nucleic acid targets is homozygous.

AE15. The method of any one of embodiments AE1 to AE14, wherein fetal fraction is determined with a coefficient of variance (CV) of 0.20 or less.

AE16. The method of embodiment AE15, wherein fetal fraction is determined with a coefficient of variance (CV) of 0.10 or less.

AE17. The method of embodiment AE16, wherein fetal fraction is determined with a coefficient of variance (CV) of 0.05 or less.

AE18. The method of any one of embodiments AE1 to AE17, wherein the polymorphic nucleic acid targets each comprise at least one single nucleotide polymorphism (SNP).

AE19. The method of embodiment AE18, wherein the SNPs are selected from:

rs10413687, rs10949838, rs1115649, rs11207002, rs11632601, rs11971741, rs12660563, rs13155942, rs1444647, rs1572801, rs17773922, rs1797700, rs1921681, rs1958312, rs196008, rs2001778, rs2323659, rs2427099, rs243992, rs251344, rs254264, rs2827530, rs290387, rs321949, rs348971, rs390316, rs3944117, rs425002, rs432586, rs444016, rs4453265, rs447247, rs4745577, rs484312, rs499946, rs500090, rs500399, rs505349, rs505662, rs516084, rs517316, rs517914, rs522810, rs531423, rs537330, rs539344, rs551372, rs567681,rs585487, rs600933, rs619208, rs622994, rs639298, rs642449, rs6700732, rs677866, rs683922, rs686851, rs6941942, rs7045684, rs7176924, rs7525374, rs870429, rs949312, rs9563831, rs970022, rs985462, rs1005241, rs1006101, rs10745725, rs10776856, rs10790342, rs11076499, rs11103233, rs11133637, rs11974817, rs12102203, rs12261, rs12460763, rs12543040, rs12695642, rs13137088, rs13139573, rs1327501, rs13438255, rs1360258, rs1421062, rs1432515, rs1452396, rs1518040, rs16853186, rs1712497, rs1792205, rs1863452, rs1991899, rs2022958, rs2099875, rs2108825, rs2132237, rs2195979, rs2248173, rs2250246, rs2268697, rs2270893, rs244887, rs2736966, rs2851428, rs2906237, rs2929724, rs3742257, rs3764584, rs3814332, rs4131376, rs4363444, rs4461567, rs4467511, rs4559013, rs4714802, rs4775899, rs4817609, rs488446, rs4950877, rs530913, rs6020434, rs6442703, rs6487229, rs6537064, rs654065, rs6576533, rs6661105, rs669161, rs6703320, rs675828, rs6814242, rs6989344, rs7120590, rs7131676, rs7214164, rs747583, rs768255, rs768708, rs7828904, rs7899772, rs7900911, rs7925270, rs7975781, rs8111589, rs849084, rs873870, rs9386151, rs9504197, rs9690525, and rs9909561.

AE20. The method of embodiment AE19, wherein the SNPs are selected from: rs10413687, rs10949838, rs1115649, rs11207002, rs11632601, rs11971741, rs12660563, rs13155942, rs1444647, rs1572801, rs17773922, rs1797700, rs1921681, rs1958312, rs196008, rs2001778, rs2323659, rs2427099, rs243992, rs251344, rs254264, rs2827530, rs290387, rs321949, rs348971, rs390316, rs3944117, rs425002, rs432586, rs444016, rs4453265, rs447247, rs4745577, rs484312, rs499946, rs500090, rs500399, rs505349, rs505662, rs516084, rs517316, rs517914, rs522810, rs531423, rs537330, rs539344, rs551372, rs567681,rs585487, rs600933, rs619208, rs622994, rs639298, rs642449, rs6700732, rs677866, rs683922, rs686851, rs6941942, rs7045684, rs7176924, rs7525374, rs870429, rs949312, rs9563831, rs970022, and rs985462.

AE21. The method of embodiment AE19, wherein the SNPs are selected from: rs1005241, rs1006101, rs10745725, rs10776856, rs10790342, rs11076499, rs11103233, rs11133637, rs11974817, rs12102203, rs12261, rs12460763, rs12543040, rs12695642, rs13137088, rs13139573, rs1327501, rs13438255, rs1360258, rs1421062, rs1432515, rs1452396,

rs1518040, rs16853186, rs1712497, rs1792205, rs1863452, rs1991899, rs2022958, rs2099875, rs2108825, rs2132237, rs2195979, rs2248173, rs2250246, rs2268697, rs2270893, rs244887, rs2736966, rs2851428, rs2906237, rs2929724, rs3742257, rs3764584, rs3814332, rs4131376, rs4363444, rs4461567, rs4467511, rs4559013, rs4714802, rs4775899, rs4817609, rs488446, rs4950877, rs530913, rs6020434, rs6442703, rs6487229, rs6537064, rs654065, rs6576533, rs6661105, rs669161, rs6703320, rs675828, rs6814242, rs6989344, rs7120590, rs7131676, rs7214164, rs747583, rs768255, rs768708, rs7828904, rs7899772, rs7900911, rs7925270, rs7975781, rs8111589, rs849084, rs873870, rs9386151, rs9504197, rs9690525, and rs9909561.

AE22. The method of any one of embodiments AE1 to AE21, wherein the polymorphic nucleic acid targets and number thereof result in at least five polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 95% of samples.

AE23. The method of embodiment AE22, wherein the polymorphic nucleic acid targets and number thereof result in at least five polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 99% of samples.

AE24. The method of any one of embodiments AE1 to AE21, wherein the polymorphic nucleic acid targets and number thereof result in at least ten polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 90% of samples.

AE25. The method of embodiment AE24, wherein the polymorphic nucleic acid targets and number thereof result in at least ten polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 95% of samples.

AE26. The method of embodiment AE25, wherein the polymorphic nucleic acid targets and number thereof result in at least ten polymorphic nucleic acid targets being informative for determining the fetal fraction for at least 99% of samples.

AE27. The method of any one of embodiments AE1 to AE26, wherein 10 or more polymorphic nucleic acid targets are enriched.

AE27.1. The method of embodiment AE27, wherein about 40 to about 100 polymorphic nucleic acid targets are enriched.

AE28. The method of embodiment AE27, wherein 50 or more polymorphic nucleic acid targets are enriched.

AE29. The method of embodiment AE28, wherein 100 or more polymorphic nucleic acid targets are enriched.

AE30. The method of embodiment AE29, wherein 500 or more polymorphic nucleic acid targets are enriched.

AE31. The method of any one of embodiments AE1 to AE30, wherein the sequencing process comprises a sequencing by synthesis method.

AE31.1. The method of embodiment AE31, wherein the sequencing by synthesis method comprises a plurality of synthesis cycles.

AE31.2. The method of embodiment AE31.1, wherein the sequencing by synthesis method comprises about 36 cycles.

AE31.3. The method of embodiment AE31.1, wherein the sequencing by synthesis method comprises about 27 cycles.

AE32. The method of any one of embodiments AE1 to AE30, wherein the sequencing process comprises a sequencing by ligation method.

AE33. The method of any one of embodiments AE1 to AE30, wherein the sequencing process comprises a single molecule sequencing method.

AE34. The method of any one of embodiments AE1 to AE33, wherein the sequencing process comprises sequencing a plurality of samples in a single compartment.

AE35. The method of embodiment AE34, wherein the fetal fraction is determined for 10 or more samples.

AE36. The method of embodiment AE35, wherein the fetal fraction is determined for 100 or more samples.

AE37. The method of embodiment AE36, wherein the fetal fraction is determined for 1000 or more samples.

AE38. The method of any one of embodiments AE1 to AE37, wherein the sample nucleic acid is cell-free DNA.

AE39. The method of any one of embodiments AE1 to AE38, wherein the sample nucleic acid is obtained from a pregnant female subject.

AE40. The method of embodiment AE39, wherein the subject is human.

AE41. The method of any one of embodiments AE1 to AE40, wherein the sample nucleic acid is from plasma or serum.

P1. A method for determining the presence or absence of a chromosome aneuploidy, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determining the presence or absence of a chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, wherein the determination is generated from the nucleic acid sequence reads, and the chromosome aneuploidy is not an aneuploidy of chromosome 13, 18, or 21.

P1.1. A method for determining the presence or absence of a chromosome aneuploidy, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the presence or absence of a chromosome aneuploidy by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) produce from the counts in (a) a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, wherein the determination is generated from the nucleic acid sequence reads, and the chromosome aneuploidy is not an aneuploidy of chromosome 13, 18, or 21.

P2. The method of embodiment P1 or P1.1, wherein the aneuploidy is a trisomy.

P3. The method of embodiment P1, P1.1 or P2, wherein the chromosome aneuploidy is a double chromosome aneuploidy.

P4. The method of any one of embodiments P1 to P3, wherein the chromosome aneuploidy is selected from an aneuploidy of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 20 22, X and Y or a combination thereof.

P5. The method of any one of embodiments P1 to P4, wherein the chromosome aneuploidy is selected from an aneuploidy of chromosome 4, 9, 16, 20, and 22 or a combination thereof.

P5.1. The method of any one of embodiments P1 to P5, wherein the chromosome aneuploidy is a fetal aneuploidy.

P6. The method of any one of embodiments P1 to P5.1, wherein the portions of the reference genome are in a chromosome.

P7. The method of any one of embodiments P1 to P6, wherein the portions of the reference genome are in a segment of a chromosome.

P8. The method of any one of embodiments P1 to P7, which comprises prior to (b) calculating a measure of error for the counts of sequence reads mapped to some or all of the portions of the reference genome and removing or weighting the counts of sequence reads for certain portions of the reference genome according to a threshold of the measure of error.

P9. The method of embodiment P8, wherein the threshold is selected according to a standard deviation gap between a first genomic section level and a second genomic section level of 3.5 or greater.

P10. The method of embodiment P8 or P9, wherein the measure of error is an R factor.

P11. The method of embodiment P10, wherein the counts of sequence reads for a portion of the reference genome having an R factor of about 7% to about 10% are removed prior to (b).

P12. The method of any one of embodiments P1 to P11, wherein the fitted relation in (b) is a fitted linear relation.

P13. The method of embodiment P12, wherein a slope of the fitted relation in (b) is determined by linear regression.

P14. The method of embodiment P12 or P13, wherein each GC bias is a GC bias coefficient, which GC bias coefficient is the slope of the linear relationship between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) the GC content for each of the portions.

P15. The method of any one of embodiments P1 to P11, wherein the fitted relation in (b) is a fitted non-linear relation.

P16. The method of embodiment P15, wherein each GC bias comprises a GC curvature estimation.

P17. The method of any one of embodiments P1 to P16, wherein the fitted relation in (c) is linear.

P18. The method of embodiment P17, wherein a slope of the fitted relation in (c) is determined by a linear regression.

P19. The method of any one of embodiments P1 to P14, P17 or P18, wherein the fitted relation in (b) is linear, the fitted relation in (c) is linear and the genomic section level L_i is determined for each of the portions of the reference genome according to Equation α:

$$L_i = (m_i - G_i S)\, I^{-1} \qquad \text{Equation } \alpha$$

wherein G_i is the GC bias, I is the intercept of the fitted relation in (c), S is the slope of the relation in (c), m_i is measured counts mapped to each portion of the reference genome and i is a sample.

P20. The method of any one of embodiments P1 to P19, wherein the number of portions of the reference genome is about 40,000 or more portions.

P21. The method of any one of embodiments P1 to P20, wherein each portion of the reference genome comprises a nucleotide sequence of a predetermined length.

P22. The method of embodiment P21, wherein the predetermined length is about 50 kilobases.

P23. The method of any one of embodiments P1 to P22, wherein the GC bias in (b) is determined by a GC bias module.

P24. The method of any one of embodiments P1 to P22, wherein the genomic section level in (c) is calculated by a level module.

P25. The method of embodiment P24, wherein the genomic section level in (c) is provided by the level module according to equation α.

P26. The method of embodiment P24 or P25, wherein the level module receives a determination of GC bias from the GC bias module.

P27. The method of any one of embodiments P1 to P25 comprising:

(i) identifying a first elevation of the calculated genomic section levels significantly different than a second elevation of the calculated genomic section levels, which first elevation is for a first set of portions, and which second elevation is for a second set of portions;

(ii) determining an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome;

(iii) adjusting the first elevation by a predetermined value when the first elevation is within the expected elevation range, thereby providing an adjustment of the first elevation, and thereby providing adjusted genomic section levels; and

(iv) determining the presence or absence of a chromosome aneuploidy in the sample from the pregnant female in (a) according to the calculated genomic section levels comprising the adjustment of (iii), wherein the determination is made with a reduced likelihood of being a false positive or false negative.

P28. The method of any one of embodiments P1 to P27, which comprises obtaining nucleic acid sequence reads.

P29. The method of embodiment P28, wherein the nucleic acid sequence reads are generated by a sequencing module.

P30. The method of embodiment P28 or P29, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

P31. The method of any one of embodiments P29 to P30, which comprises mapping the nucleic acid sequence reads to a segment of the reference genome or to an entire reference genome.

P32. The method of embodiment P31, wherein the nucleic acid sequence reads are mapped by a mapping module.

P33. The method of any one of embodiments P1 to P32, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are counted by a counting module.

P34. The method of embodiment P32 or P33, wherein the sequence reads are transferred to the mapping module from the sequencing module.

P35. The method of embodiment P33 or P34, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the counting module from the mapping module.

P36. The method of any one of embodiments P1 to P35, wherein the normalized counts in (b) are provided by a normalization module.

P37. The method of any one of embodiments P34 to P36, wherein the counts of the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the normalization module from the counting module.

P38. The method of any one of embodiments P34 to P37, wherein an apparatus comprises one or more of the sequencing module, the mapping module, the counting module, the normalization module, the comparison module, the range setting module, a categorization module, the adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

P39. The method of embodiment P38, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

P40. The method of any one of embodiments P33 to P39, wherein a second apparatus comprises the mapping module and the counting module.

P41. The method of any one of embodiments P29 to P40, wherein a third apparatus comprises the sequencing module.

P41.1. The method of any one of embodiments P39 to P41, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

P42. The method of any one of embodiments P1 to P41.1, wherein determining the presence or absence of a chromosome aneuploidy in the sample from the pregnant female in (a) comprises determining the presence or absence of a chromosome aneuploidy in a fetus

P43. The method of any one of embodiments P1 to P42, wherein determining the presence or absence of a chromosome aneuploidy comprises determining an outcome for a fetus.

Q1. A method for determining the presence or absence of a double chromosome aneuploidy, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels,

whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determining the presence or absence of a double chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

Q1.1 A method for determining the presence or absence of a double chromosome aneuploidy, comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the presence or absence of a double chromosome aneuploidy for the sample by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) determine from the counts in (a) a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a double chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

Q2. The method of embodiment Q1 or Q1.1, comprising identifying a first calculated genomic section level and a second calculated genomic section level different than a third calculated genomic section level, which first calculated genomic section level is for a first set of portions, which second calculated genomic section level is for a second set of portions and which third calculated genomic section level is for a third set of portions.

Q3. The method of embodiment Q1, Q1.1 or Q2, wherein one or both of the aneuploidies in the double chromosome aneuploidy is a trisomy.

Q4. The method of any one of embodiments Q1 to Q3, wherein one of the chromosome aneuploidies is selected from an aneuploidy of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X and Y or a combination thereof.

Q5. The method of any one of embodiments Q1 to Q4, wherein one of the chromosome aneuploidies is selected from an aneuploidy of chromosome 13, 18 and 21.

Q6. The method of any one of embodiments Q1 to Q5, wherein the chromosome aneuploidy is a fetal aneuploidy.

Q7. The method of any one of embodiments Q1 to Q6, wherein the portions of the reference genome are in a chromosome.

Q8. The method of any one of embodiments Q1 to Q7, wherein the portions of the reference genome are in a segment of a chromosome.

Q9. The method of any one of embodiments Q1 to Q8, which comprises prior to (b) calculating a measure of error for the counts of sequence reads mapped to some or all of the portions of the reference genome and removing or weighting the counts of sequence reads for certain portions of the reference genome according to a threshold of the measure of error.

Q10. The method of embodiment Q9, wherein the threshold is selected according to a standard deviation gap between a first genomic section level and a second genomic section level of 3.5 or greater.

Q11. The method of embodiment Q9 or Q10, wherein the measure of error is an R factor.

Q12. The method of embodiment Q11, wherein the counts of sequence reads for a portion of the reference genome having an R factor of about 7% to about 10% are removed prior to (b).

Q13. The method of any one of embodiments Q1 to Q12, wherein the fitted relation in (b) is a fitted linear relation.

Q14. The method of embodiment Q13, wherein a slope of the fitted relation in (b) is determined by a linear regression.

Q15. The method of embodiment Q13 or Q14, wherein each GC bias is a GC bias coefficient, which GC bias coefficient is the slope of the linear relationship between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) the GC content for each of the portions.

Q16. The method of any one of embodiments Q1 to Q12, wherein the fitted relation in (b) is a fitted non-linear relation.

Q17. The method of embodiment Q16, wherein each GC bias comprises a GC curvature estimation.

Q18. The method of any one of embodiments Q1 to Q17, wherein the fitted relation in (c) is linear.

Q19. The method of embodiment Q18, wherein a slope of the fitted relation in (c) is determined by a linear regression.

Q20. The method of any one of embodiments Q1 to Q15, Q18 or Q19, wherein the fitted relation in (b) is linear, the fitted relation in (c) is linear and the genomic section level L_i is determined for each of the portions of the reference genome according to Equation α:

$$L_i = (m_i - G_i S)\, I^{-1} \qquad \text{Equation } \alpha$$

wherein G_i is the GC bias, I is the intercept of the fitted relation in (c), S is the slope of the relation in (c), m_i is measured counts mapped to each portion of the reference genome and i is a sample.

Q21. The method of any one of embodiments Q1 to Q20, wherein the number of portions of the reference genome is about 40,000 or more portions.

Q22. The method of any one of embodiments Q1 to Q21, wherein each portion of the reference genome comprises a nucleotide sequence of a predetermined length.

Q23. The method of embodiment Q22, wherein the predetermined length is about 50 kilobases.

Q24. The method of any one of embodiments Q1 to Q23, wherein the GC bias in (b) is determined by a GC bias module.

Q25. The method of any one of embodiments Q1 to Q24, wherein the genomic section level in (c) is calculated by a level module.

Q26. The method of embodiment Q25, wherein the genomic section level in (c) is provided by the level module according to equation α.

Q27. The method of embodiment Q25 or Q26, wherein the level module receives a determination of GC bias from the GC bias module.

Q28. The method of any one of embodiments Q1 to Q27 comprising:

(i) identifying a first elevation of the calculated genomic section levels significantly different than a second elevation of the calculated genomic section levels, which first elevation is for a first set of portions, and which second elevation is for a second set of portions;

(ii) determining an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome;

(iii) adjusting the first elevation by a predetermined value when the first elevation is within the expected elevation range, thereby providing an adjustment of the first elevation, and thereby providing adjusted genomic section levels; and

(iv) determining the presence or absence of a chromosome aneuploidy in the sample from the pregnant female in (a) according to the calculated genomic section levels comprising the

adjustment of (iii), wherein the determination is made with a reduced likelihood of being a false positive or false negative.

Q29. The method of any one of embodiments Q1 to Q28, which comprises obtaining nucleic acid sequence reads.

Q30. The method of embodiment Q29, wherein the nucleic acid sequence reads are generated by a sequencing module.

Q31. The method of embodiment Q29 or Q30, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

Q32. The method of any one of embodiments Q30 to Q31, which comprises mapping the nucleic acid sequence reads to a segment of the reference genome or to an entire reference genome.

Q33. The method of embodiment Q32, wherein the nucleic acid sequence reads are mapped by a mapping module.

Q34. The method of any one of embodiments Q1 to Q33, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are counted by a counting module.

Q35. The method of embodiment Q33 or Q34, wherein the sequence reads are transferred to the mapping module from the sequencing module.

Q36. The method of embodiment Q34 or Q35, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the counting module from the mapping module.

Q37. The method of any one of embodiments Q1 to Q36, wherein the normalized counts in (b) are provided by a normalization module.

Q38. The method of any one of embodiments Q35 to Q37, wherein the counts of the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the normalization module from the counting module.

Q39. The method of any one of embodiments Q35 to Q38, wherein an apparatus comprises one or more of the sequencing module, the mapping module, the counting module, the normalization module, the comparison module, the range setting module, a categorization module, the adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

Q40. The method of embodiment Q39, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

Q41. The method of any one of embodiments Q34 to Q40, wherein a second apparatus comprises the mapping module and the counting module.

Q42. The method of any one of embodiments Q30 to Q41, wherein a third apparatus comprises the sequencing module.

Q42.1. The method of any one of embodiments Q40 to Q42, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

Q43. The method of any one of embodiments Q1 to Q42.1, wherein determining the presence or absence of a double chromosome aneuploidy in the sample from the pregnant female in (a) comprises determining the presence or absence of a double chromosome aneuploidy in a fetus

Q44. The method of any one of embodiments Q1 to Q43, wherein determining the presence or absence of a double chromosome aneuploidy comprises determining an outcome for a fetus.

R1. A method for determining the presence or absence of a microdeletion or microduplication comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determining the presence or absence of a microdeletion or microduplication for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

R1.1. A method for determining the presence or absence of a microdeletion or microduplication comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the one or more computing apparatus in the system comprise memory and one or more processors, and

determining the presence or absence of a microdeletion or microduplication for the sample by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) determine from the counts in (a) a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to

each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determining the presence or absence of a microdeletion or microduplication for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

R2. The method of embodiment R1 or R1.1, comprising determining the presence or absence of a microdeletion wherein the microdeletion is a deletion of a segment of chromosome selected from chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X and Y.

R2.1. The method of embodiment R1, R1.1 or R2, comprising determining the presence or absence of a microdeletion wherein the microdeletion is a deletion of a segment of chromosome X or chromosome 7.

R3. The method of embodiment R1, comprising determining the presence or absence of a microduplication wherein the microduplication is a duplication of a segment of a chromosome selected from chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, X and Y.

R3.1. The method of embodiment R1 or R3, comprising determining the presence or absence of a microduplication wherein the microduplication is a duplication of a segment of a chromosome selected from chromosome 3, 4, 5, 7, 11, 14, 18 and X.

R4. The method of any one of embodiments R1 to R3.1, wherein the microdeletion or microduplication is a fetal microdeletion or a fetal microduplication.

R5. The method of any one of embodiments R1 to R4, wherein the portions of the reference genome are in a chromosome.

R6. The method of any one of embodiments R1 to R5, wherein the portions of the reference genome are in a segment of a chromosome.

R7. The method of any one of embodiments R1 to R6, which comprises prior to (b) calculating a measure of error for the counts of sequence reads mapped to some or all of the portions of the reference genome and removing or weighting the counts of sequence reads for certain portions of the reference genome according to a threshold of the measure of error.

R8. The method of embodiment R7, wherein the threshold is selected according to a standard deviation gap between a first genomic section level and a second genomic section level of 3.5 or greater.

R9. The method of embodiment R7 or R8, wherein the measure of error is an R factor.

R10. The method of embodiment R9, wherein the counts of sequence reads for a portion of the reference genome having an R factor of about 7% to about 10% are removed prior to (b).

R11. The method of any one of embodiments R1 to R10, wherein the fitted relation in (b) is a fitted linear relation.

R12. The method of embodiment R11, wherein a slope of the fitted relation in (b) is determined by a linear regression.

R13. The method of embodiment R11 or R12, wherein each GC bias is a GC bias coefficient, which GC bias coefficient is the slope of the linear relationship between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) the GC content for each of the portions.

R14. The method of any one of embodiments R1 to R110, wherein the fitted relation in (b) is a fitted non-linear relation.

R15. The method of embodiment R14, wherein each GC bias comprises a GC curvature estimation.

R16. The method of any one of embodiments R1 to R15, wherein the fitted relation in (c) is linear.

R17. The method of embodiment R16, wherein a slope of the fitted relation in (c) is determined by a linear regression.

R18. The method of any one of embodiments R1 to R17, R16 or R17, wherein the fitted relation in (b) is linear, the fitted relation in (c) is linear and the genomic section level L_i is determined for each of the portions of the reference genome according to Equation α:

$$L_i = (m_i - G_i S)\, I^{-1} \qquad \text{Equation } \alpha$$

wherein G_i is the GC bias, I is the intercept of the fitted relation in (c), S is the slope of the relation in (c), m_i is measured counts mapped to each portion of the reference genome and i is a sample.

R19. The method of any one of embodiments R1 to R18, wherein the number of portions of the reference genome is about 40,000 or more portions.

R20. The method of any one of embodiments R1 to R19, wherein each portion of the reference genome comprises a nucleotide sequence of a predetermined length.

R21. The method of embodiment R20, wherein the predetermined length is about 50 kilobases.

R22. The method of any one of embodiments R1 to R21, wherein the GC bias in (b) is determined by a GC bias module.

R23. The method of any one of embodiments R1 to R22, wherein the genomic section level in (c) is calculated by a level module.

R24. The method of embodiment R23, wherein the genomic section level in (c) is provided by the level module according to equation α.

R25. The method of embodiment R23 or R24, wherein the level module receives a determination of GC bias from the GC bias module.

R26. The method of any one of embodiments R1 to R25 comprising:

(i) identifying a first elevation of the calculated genomic section levels significantly different than a second elevation of the calculated genomic section levels, which first elevation is for a first set of portions, and which second elevation is for a second set of portions;

(ii) determining an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome;

(iii) adjusting the first elevation by a predetermined value when the first elevation is within the expected elevation range, thereby providing an adjustment of the first elevation, and thereby providing adjusted genomic section levels; and

(iv) determining the presence or absence of a chromosome aneuploidy in the sample from the pregnant female in (a) according to the calculated genomic section levels comprising the adjustment of (iii), wherein the determination is made with a reduced likelihood of being a false positive or false negative.

R27. The method of any one of embodiments R1 to R26, which comprises obtaining nucleic acid sequence reads.

R28. The method of embodiment R27, wherein the nucleic acid sequence reads are generated by a sequencing module.

R29. The method of embodiment R27 or R28, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

R30. The method of any one of embodiments R28 to R29, which comprises mapping the nucleic acid sequence reads to a segment of the reference genome or to an entire reference genome.

R31. The method of embodiment R30, wherein the nucleic acid sequence reads are mapped by a mapping module.

R32. The method of any one of embodiments R1 to R31, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are counted by a counting module.

R33. The method of embodiment R31 or R32, wherein the sequence reads are transferred to the mapping module from the sequencing module.

R34. The method of embodiment R32 or R33, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the counting module from the mapping module.

R35. The method of any one of embodiments R1 to R34, wherein the normalized counts in (b) are provided by a normalization module.

R36. The method of embodiment R35, wherein the counts of the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the normalization module from the counting module.

R37. The method of embodiment R35 or R36, wherein an apparatus comprises one or more of the sequencing module, the mapping module, the counting module, the normalization module, the comparison module, the range setting module, a categorization module, the adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

R38. The method of embodiment R37, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

R39. The method of any one of embodiments R32 to R38, wherein a second apparatus comprises the mapping module and the counting module.

R40. The method of any one of embodiments R28 to R39, wherein a third apparatus comprises the sequencing module.

R40.1. The method of any one of embodiments R38 to R40, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

R41. The method of any one of embodiments R1 to R40.1, wherein determining the presence or absence of a microdeletion or microduplication in the sample from the pregnant female in (a) comprises determining the presence or absence of a microdeletion or microduplication in a fetus.

R42. The method of any one of embodiments R1 to R41, wherein determining the presence or absence of a microdeletion or microduplication comprises determining an outcome for a fetus.

T1. A method for determining the presence or absence of a complex genetic variation comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determining a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determining the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more genetic variations chosen from a whole chromosome aneuploidy, a microduplication and a microdeletion, and wherein the determination is generated from the nucleic acid sequence reads.

T1.1. A method for determining the presence or absence of a complex genetic variation comprising:

loading a sequencing apparatus with circulating cell-free nucleic acid from the blood of a pregnant female, or loading the sequencing apparatus with a modified variant of the nucleic acid, which sequencing apparatus produces signals corresponding to nucleotide bases of the nucleic acid;

generating sequence reads from the signals of the nucleic acid by, after optionally transferring the signals to, a system comprising one or more computing apparatus, wherein the

one or more computing apparatus in the system comprise memory and one or more processors, and

determining the presence or absence of a complex genetic variation for the sample by the system, wherein one computing apparatus, or combination of computing apparatus, in the system is configured to:

(a) align the sequence reads to genomic sections of a reference genome and generate counts of the sequence reads mapped to the genomic sections;

(b) determine from the counts in (a) a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculating a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determining the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more genetic variations chosen from a whole chromosome aneuploidy, a microduplication and a microdeletion, and wherein the determination is generated from the nucleic acid sequence reads.

T2. The method of embodiment T1 or T1.1, comprising identifying at least a first calculated genomic section level and at least a second calculated genomic section level significantly different than a third calculated genomic section level, which first calculated genomic section level is for a first set of portions, which second calculated genomic section level is for a second set of portions and which third calculated genomic section level is for a third set of portions.

T2.1. The method of embodiment T1 or T2, wherein the first calculated genomic section level and the second calculated genomic section level comprise a genetic variation.

T3. The method of any one of embodiments T1 to T2.1, wherein the complex genetic variation comprises a whole chromosome aneuploidy and one or more microduplications.

T4. The method of any one of embodiments T1 to T2.1, wherein the complex genetic variation comprises a whole chromosome aneuploidy and one or more microdeletions.

T5. The method of any one of embodiments T1 to T2.1, wherein the complex genetic variation comprises a whole chromosome aneuploidy and one or more of a microdeletion and/or a microduplication.

T6. The method of any one of embodiments T1 to T5, wherein the whole chromosome aneuploidy is a trisomy.

T7. The method of embodiment T6, wherein the trisomy is a trisomy of chromosome X, 13, 18 or 21.

T8. The method of any one of embodiments T1 to T7, wherein the whole chromosome aneuploidy is a chromosome duplication.

T9. The method of any one of embodiments T1 to T7, wherein the whole chromosome aneuploidy is a chromosome deletion.

T10. The method of any one of embodiments T1 to T9, wherein the complex genetic variation comprises a trisomy and one or more of a microdeletion and/or a microduplication.

T11. The method of any one of embodiments T2 to T10, wherein the at least first calculated genomic section level, the at least second calculated genomic section level and/or the at least third calculated genomic section level is a genomic section level of a chromosome, or a segment thereof.

T12. The method of any one of embodiments T1 to T11, wherein the complex genetic variation is a complex genetic variation in a fetus.

T13. The method of any one of embodiments T1 to T12, wherein the portions of the reference genome are in a chromosome.

T14. The method of any one of embodiments T1 to T13, wherein the portions of the reference genome are in a segment of a chromosome.

T15. The method of any one of embodiments T1 to T14, which comprises prior to (b) calculating a measure of error for the counts of sequence reads mapped to some or all of the portions of the reference genome and removing or weighting the counts of sequence reads for certain portions of the reference genome according to a threshold of the measure of error.

T16. The method of embodiment T15, wherein the threshold is selected according to a standard deviation gap between a first genomic section level and a second genomic section level of 3.5 or greater.

T17. The method of embodiment T15 or T16, wherein the measure of error is an R factor.

T18. The method of embodiment T17, wherein the counts of sequence reads for a portion of the reference genome having an R factor of about 7% to about 10% are removed prior to (b).

T19. The method of any one of embodiments T1 to T18, wherein the fitted relation in (b) is a fitted linear relation.

T20. The method of embodiment T19, wherein a slope of the fitted relation in (b) is determined by a linear regression.

T21. The method of embodiment T19 or T20, wherein each GC bias is a GC bias coefficient, which GC bias coefficient is the slope of the linear relationship between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) the GC content for each of the portions.

T22. The method of any one of embodiments T1 to T18, wherein the fitted relation in (b) is a fitted non-linear relation.

T23. The method of embodiment T22, wherein each GC bias comprises a GC curvature estimation.

T24. The method of any one of embodiments T1 to T23, wherein the fitted relation in (c) is linear.

T25. The method of embodiment T24, wherein a slope of the fitted relation in (c) is determined by a linear regression.

T26. The method of any one of embodiments T1 to T25, T24 or T25, wherein the fitted relation in (b) is linear, the fitted relation in (c) is linear and the genomic section level L_i is determined for each of the portions of the reference genome according to Equation α:

$$L_i = (m_i - G_i S)\, I^{-1} \qquad \text{Equation } \alpha$$

wherein G_i is the GC bias, I is the intercept of the fitted relation in (c), S is the slope of the relation in (c), m_i is measured counts mapped to each portion of the reference genome and i is a sample.

T27. The method of any one of embodiments T1 to T26, wherein the number of portions of the reference genome is about 40,000 or more portions.

T28. The method of any one of embodiments T1 to T27, wherein each portion of the reference genome comprises a nucleotide sequence of a predetermined length.

T29. The method of embodiment T28, wherein the predetermined length is about 50 kilobases.

T30. The method of any one of embodiments T1 to T29, wherein the GC bias in (b) is determined by a GC bias module.

T31. The method of any one of embodiments T1 to T30, wherein the genomic section level in (c) is calculated by a level module.

T32. The method of embodiment T31, wherein the genomic section level in (c) is provided by the level module according to equation α.

T33. The method of embodiment T31 or T32, wherein the level module receives a determination of GC bias from the GC bias module.

T34. The method of any one of embodiments T1 to T33, which comprises obtaining nucleic acid sequence reads.

T35. The method of embodiment T34, wherein the nucleic acid sequence reads are generated by a sequencing module.

T36. The method of embodiment T34 or T35, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

T37. The method of any one of embodiments T35 to T36, which comprises mapping the nucleic acid sequence reads to a segment of the reference genome or to an entire reference genome.

T38. The method of embodiment T37, wherein the nucleic acid sequence reads are mapped by a mapping module.

T39. The method of any one of embodiments T1 to T38, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are counted by a counting module.

T40. The method of embodiment T38 or T39, wherein the sequence reads are transferred to the mapping module from the sequencing module.

T41. The method of embodiment T39 or T40, wherein the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the counting module from the mapping module.

T42. The method of any one of embodiments T1 to T41, wherein the normalized counts in (b) are provided by a normalization module.

T43. The method of embodiment T42, wherein the counts of the nucleic acid sequence reads mapped to the portions of the reference genome are transferred to the normalization module from the counting module.

T44. The method of embodiment T42 or T43, wherein an apparatus comprises one or more of the sequencing module, the mapping module, the counting module, the normalization module, the

comparison module, the range setting module, a categorization module, the adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

T45. The method of embodiment T44, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

T46. The method of any one of embodiments T39 to T45, wherein a second apparatus comprises the mapping module and the counting module.

T47. The method of any one of embodiments T35 to T46, wherein a third apparatus comprises the sequencing module.

T47.1. The method of any one of embodiments T45 to T47, wherein the one or more computing apparatus comprise the first apparatus, the second apparatus, the third apparatus or a combination thereof.

T48. The method of any one of embodiments T1 to T47.1, wherein determining the presence or absence of a complex genetic variation in the sample from the pregnant female in (a) comprises determining the presence or absence of a complex genetic variation in a fetus.

T49. The method of any one of embodiments T1 to T48, wherein determining the presence or absence of a complex genetic variation comprises determining an outcome for a fetus.

U1. A system comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and
which instructions executable by the one or more processors are configured to:

(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the

sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, wherein the determination is generated from the nucleic acid sequence reads, and the chromosome aneuploidy is not an aneuploidy of chromosome 13, 18, or 21.

U2. A system comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and
which instructions executable by the one or more processors are configured to:

(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a double chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

U3. A system comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which

sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and

which instructions executable by the one or more processors are configured to:

(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a microdeletion or microduplication for the sample from the pregnant female in (a) according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

U4. A system comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and

which instructions executable by the one or more processors are configured to:

(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more of a whole chromosome aneuploidy, a microduplication, a

microdeletion, or a combination thereof, whereby the determination is generated from the nucleic acid sequence reads.

V1. An apparatus comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and
which instructions executable by the one or more processors are configured to:
(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;
(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and
(c) determine the presence or absence of a chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, wherein the determination is generated from the nucleic acid sequence reads, and the chromosome aneuploidy is not an aneuploidy of chromosome 13, 18, or 21.

V2. An apparatus comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and
which instructions executable by the one or more processors are configured to:
(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a double chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

V3. An apparatus comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and
which instructions executable by the one or more processors are configured to:
(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a microdeletion or microduplication for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

V4. An apparatus comprising one or more processors and memory,
which memory comprises instructions executable by the one or more processors and which memory comprises counts of sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female; and
which instructions executable by the one or more processors are configured to:

(a) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(b) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(c) determine the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more of a whole chromosome aneuploidy, a microduplication, a microdeletion, or a combination thereof, and whereby the determination is generated from the nucleic acid sequence reads.

W1. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, wherein the determination is generated from the nucleic acid sequence reads, and the chromosome aneuploidy is not an aneuploidy of chromosome 13, 18, or 21.

W2. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a double chromosome aneuploidy for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

W3. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a microdeletion or microduplication for the sample from the pregnant female according to the calculated genomic section levels, whereby the determination is generated from the nucleic acid sequence reads.

W4. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:
(a) obtain counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;
(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;
(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and
(d) determine the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more of a whole chromosome aneuploidy, a microduplication, a microdeletion, or a combination thereof, whereby the determination is generated from the nucleic acid sequence reads.

X1. A system comprising one or more microprocessors and memory,
which memory comprises instructions executable by the one or more microprocessors, and
which instructions executable by the one or more processors are configured to:
(a) determine a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus;
(b) obtain counts of sequence reads mapped to portions of a reference genome, which sequence reads are from the nucleic acid in the sample;
(c) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

X1.1 A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) determine a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus;

(b) produce sequence reads from the signals, align the sequence reads to portions of a reference genome, and generate counts of the sequence reads mapped to the portions;

(c) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

X2. The system of embodiment X1 or X1.1, wherein the fetal fraction is determined from a first part of the test sample and the genomic section levels are determined from a second part of the test sample.

X3. The system of embodiment X1, X1.1 or X2, wherein calculating the genomic section level for each of the portions of the reference genome comprises normalizing counts of reads mapped to the reference genome according to guanine and cytosine (GC) content for each of the portions.

X4. The system of any one of embodiments X1 to X3, comprising instructions configured to:

(1) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and

(2) calculate the genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

X5. The system of any one of embodiments X1 to X4, wherein the subset of portions of the reference genome in (d)(i) is portions of a chromosome or segment thereof.

X6. The system of embodiment X5, wherein the chromosome is chosen from chromosome 13, chromosome 18 and chromosome 21.

X7. The system of any one of embodiments X1 to X6, comprising instructions configured to determine a reference count and an uncertainty value according to the reference count:

wherein the reference count is determined according to calculated genomic section levels for a subset of portions of the reference genome for one or more pregnant females bearing a fetus;

wherein the subset of portions of the reference genome for one or more pregnant females are known to be euploid for the fetus and/or the mother;

wherein the reference count is not determined from the sample; and

wherein the reference count is determined from the same subset of portions of the reference genome as in (d).

X8. The system of embodiment X7, wherein the reference count is normalized by bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof.

X9. The system of any one of embodiments X1 to X8, comprising instructions configured to determine a maternal ploidy.

X10. The system of embodiment X9, wherein the fetal ploidy is determined in (d) according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a) and (iii) a maternal ploidy.

X11. The system of embodiment X10, wherein the fetal ploidy is determined according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a), (iii) a maternal ploidy, (iv) the reference count and (v) an uncertainty value σ for the reference count.

X12. The system of embodiment X11, wherein the fraction of fetal nucleic acid determined in (a) is fixed at its determined value and fetal ploidy *X* is determined according to Equation 8 below, or a derivation thereof:

$$y_i = (1 - F)M_i f_i + FXf_i \qquad (8)$$

where y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, *X* represents the fetal ploidy, and M_i represents the maternal ploidy of portion *i.*

X13. The system of embodiment X12, comprising instructions configured to determine the sum of squared residuals according to equation (8) and for multiple bins *i* for a subset of portions of the reference genome.

X14. The system of embodiment X13, wherein the fetal fraction is fixed at a value determined in (a) and the fetal ploidy is varied to optimize the sum of squared residuals according to equation (8) or a variation thereof.

X15. The system of embodiment X14, comprising instructions configured to determine a linear regression according to the sum of square residuals.

X16. The system of embodiment X11, wherein the fetal ploidy is determined according to Equation 20 below:

$$X = \frac{\sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^{N} \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}} . \qquad (20)$$

wherein y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, σ represents the uncertainty value for f_i, *X* represents the fetal ploidy, and M_i represents the maternal ploidy of portion *i*.

X17. The system of embodiment X11, wherein the fetal ploidy is determined according to Equation 21 below:

$$X = \frac{\Xi_{fy} - (1-F)\Xi_{ff}}{F\Xi_{ff}} = \frac{\Xi_{fy}}{F\Xi_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Xi_{fy}}{\Xi_{ff}} - 1\right) \quad (21)$$

wherein $\Xi_{ff} = \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}$, $\Xi_{fy} = \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2}$, y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, σ represents the uncertainty value for f_i, and *X* represents the fetal ploidy.

X18. The system of any one of embodiments X1 to X17, comprising instructions configured to determine the presence or absence of a fetal chromosome aneuploidy according to the fetal ploidy determined in (d).

X19. The system of embodiment X18, wherein the fetal ploidy determined in (d) is greater than about 1.25 and the presence of a fetal chromosome aneuploidy is determined.

X20. The system of embodiment X19, wherein the fetal ploidy determined in (d) is less than about 1.25 and the absence of a fetal chromosome aneuploidy is determined.

X21. The system of embodiment X19, wherein the fetal chromosome aneuploidy is a trisomy.

X22. The system of embodiment X21, wherein the trisomy is selected from a trisomy of chromosome 13, 18 and 21.

X23. The system of any one of embodiments X1 to X22, wherein determining the fraction of fetal nucleic acid comprises analyzing the calculated genomic sections levels for a subset of portions of

the reference genome, which subset is a first subset, the subset in (d) is a second subset, and the first subset of portions of the reference genome is portions of a Y chromosome or a segment thereof.

X24. The system of any one of embodiments X1 to X22, wherein determining the fraction of fetal nucleic acid comprises:

(1) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(2) from the counts in (1), generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof;

(3) determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (2) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

X25. The system of any one of embodiments X1 to X22, wherein determining the fraction of fetal nucleic acid comprises analyzing one or more loci in sample nucleic acid, wherein at least one of the one or more loci vary between fetal nucleic acid and maternal nucleic acid.

X26. The system of embodiment X25, wherein the one or more loci comprise one or more polymorphic sites, comprising:

(1) enriching nucleic acid in a first part of the test sample for a plurality of polymorphic sites;

(2) obtaining nucleotide sequences for some or all of the polymorphic sites by a sequencing process;

(3) analyzing the nucleotide sequences of (2); and

(4) determining the fraction of fetal nucleic acid based on the analysis of (3), wherein the polymorphic sites and number thereof result in at least five polymorphic sites being informative for determining the fetal fraction for at least 90% of samples.

X27. The system of embodiment X25, wherein the one or more loci comprise one or more methylation regions, comprising:

(1) contacting the test sample with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid; and

(2) determining the fraction of fetal nucleic acid in the sample based on the differentially modified nucleic acid.

X28. The system of embodiment X27, wherein the one or more agents are methylation sensitive restriction enzymes.

Y1. A system comprising one or more microprocessors and memory,
which memory comprises instructions executable by the one or more microprocessors, and
which instructions executable by the one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex

genetic variation comprises two or more genetic variations chosen from a whole chromosome aneuploidy, a microduplication and a microdeletion, and wherein the determination is generated from the nucleic acid sequence reads.

Y1.1. A system comprising a sequencing apparatus and one or more computing apparatus,

which sequencing apparatus is configured to produce signals corresponding to nucleotide bases of a nucleic acid loaded in the sequencing apparatus, which nucleic acid is circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus, or which nucleic acid loaded in the sequencing apparatus is a modified variant of the circulating cell-free nucleic acid; and

which one or more computing apparatus comprise memory and one or more processors, which memory comprises instructions executable by the one or more processors and which instructions executable by the one or more processors are configured to:

(a) produce sequence reads from the signals, align the sequence reads to portions of a reference genome, and generate counts of the sequence reads mapped to the portions;

(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more genetic variations chosen from a whole chromosome aneuploidy, a microduplication and a microdeletion, and wherein the determination is generated from the nucleic acid sequence reads.

Y2. The system of claim Y1 or Y1.1, wherein the whole chromosome aneuploidy is a trisomy of a chromosome selected from chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 22, X and Y.

Y3. The system of claim Y1, Y1.1 or Y2, wherein the complex genetic variation comprises a double chromosome aneuploidy.

Y4. The system of any one of claims Y1 to Y3, wherein the complex genetic variation comprises a fetal genetic variation.

Y5. The system of any one of claims Y1 to Y4, wherein the microdeletion and/or microduplication is located in a chromosome selected from chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, and X.

Y6. The system of any one of claims Y1 to Y5, comprising instructions configured to, prior to (b), calculate a measure of error for the counts of sequence reads mapped to some or all of the portions of the reference genome and remove or weight the counts of sequence reads for certain portions of the reference genome according to a threshold of the measure of error.

Y7. The system of claim Y6, wherein the threshold is selected according to a standard deviation gap between a first genomic section level and a second genomic section level of 3.5 or greater.

Y8. The system of claim Y6, wherein the measure of error is an R factor.

Y9. The system of any one of claims Y1 to Y8, wherein the counts of sequence reads for a portion of the reference genome having an R factor of about 7% to about 10% are removed prior to (b).

Y10. The system of any one of claims Y1 to Y9, wherein the fitted relation in (b) is a fitted linear relation.

Y11. The system of any one of claims Y1 to Y10, wherein a slope of the fitted relation in (b) is determined by linear regression.

Y12. The system of any one of claims Y1 to Y11, wherein each GC bias is a GC bias coefficient, which GC bias coefficient is the slope of the linear relationship between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) the GC content for each of the portions.

Y13. The system of claim of any one of claims Y1 to Y9, wherein the fitted relation in (b) is a fitted non-linear relation.

Y14. The system of claim Y13, wherein each GC bias comprises a GC curvature estimation.

Y15. The system of any one of claims Y1 to Y14, wherein the fitted relation in (c) is linear.

Y16. The system of any one of claims Y1 to Y15, wherein a slope of the fitted relation in (c) is determined by a linear regression.

Y17. The system of any one of claims Y1 to Y16, wherein the fitted relation in (b) is linear, the fitted relation in (c) is linear and the genomic section level L_i is determined for each of the portions of the reference genome according to Equation α:

$$L_i = (m_i - G_i S)\, I^{-1} \qquad \text{Equation } \alpha$$

wherein G_i is the GC bias, I is the intercept of the fitted relation in (c), S is the slope of the relation in (c), m_i is measured counts mapped to each portion of the reference genome and i is a sample.

Y18. The system of any one of claims Y1 to Y17, wherein the number of portions of the reference genome is about 40,000 or more portions.

Y19. The system of any one of claims Y1 to Y18, wherein each portion of the reference genome comprises a nucleotide sequence of a predetermined length.

Y20. The system of any one of claims Y1 to Y19, comprising instructions configured to:

(i) identify a first elevation of the calculated genomic section levels significantly different than a second elevation of the calculated genomic section levels, which first elevation is for a first set of portions, and which second elevation is for a second set of portions;

(ii) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome;

(iii) adjust the first elevation by a predetermined value when the first elevation is within the expected elevation range, thereby providing an adjustment of the first elevation, and thereby providing adjusted genomic section levels; and

(iv) determine the presence or absence of a chromosome aneuploidy in the sample from the pregnant female in (a) according to the calculated genomic section levels comprising the adjustment of (iii), wherein the determination is made with a reduced likelihood of being a false positive or false negative.

* * *

The entirety of each patent, patent application, publication and document referenced herein hereby is incorporated by reference. Citation of the above patents, patent applications, publications and documents is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

Modifications can be made to the foregoing without departing from the basic aspects of the technology. Although the technology has been described in substantial detail with reference to one or more specific embodiments, those of ordinary skill in the art will recognize that changes can be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the technology.

The technology illustratively described herein suitably can be practiced in the absence of any element(s) not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of," and "consisting of" can be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and use of such terms and expressions do not exclude any equivalents of the features shown and described or segments thereof, and various modifications are possible within the scope of the technology claimed. The term "a" or "an" can refer to one of or a plurality of the elements it modifies (e.g., "a reagent" can mean one or more reagents) unless it is contextually clear either one of the elements or more than one of the elements is described. The term "about" as used herein refers to a value within 10% of the underlying parameter (i.e., plus or minus 10%), and use of the term "about" at the beginning of a string of values modifies each of the values (i.e., "about 1, 2 and 3" refers to about 1, about 2 and about 3). For example, a weight of "about 100 grams" can include weights between 90 grams and 110 grams. Further, when a listing of values is described herein (e.g., about 50%, 60%, 70%, 80%, 85% or 86%) the listing includes all intermediate and fractional values thereof (e.g., 54%, 85.4%). Thus, it should be understood that although the present technology has been specifically disclosed by representative embodiments and optional features, modification and variation of the concepts herein disclosed

can be resorted to by those skilled in the art, and such modifications and variations are considered within the scope of this technology.

Certain embodiments of the technology are set forth in the claim(s) that follow(s).

What is claimed is:

1. A method for determining a fraction of fetal nucleic acid in circulating cell-free nucleic acid from blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) from the counts in (a), generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof;

(c) determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

2. The method of claim 1, wherein the X chromosome representation in (c)(ii) is an experimental X chromosome representation.

3. The method of claim 1 or 2, wherein the X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof.

4. The method of any one of claims 1 to 3, wherein the X chromosome representation in (c)(ii) is determined from a set of pregnant females bearing a male fetus.

5. The method of any one of claims 1 to 4, wherein the X chromosome representation in (c)(ii) is determined for pregnant females bearing a male fetus.

6. The method of any one of claims 1 to 5, wherein the fitted relationship is linear.

7. The method of claim 6, wherein the fraction of the fetal nucleic acid is determined according to the slope and intercept of the fitted relationship, the experimental Y chromosome representation generated in (b) and a median X chromosome representation for a set of pregnant females bearing a female fetus.

8. The method of claim 7, wherein the fraction of the fetal nucleic acid is determined according to equation (62):

$$f = 2\frac{I+S\langle x\rangle - y}{S\langle x\rangle} \qquad (62)$$

wherein *I* is the intercept, *S* is the slope, $\langle x\rangle$ is the median X chromosome representation for a set of pregnant females bearing a female fetus and *y* is the experimental Y chromosome representation generated in (b).

9. The method of any one of claims 1 to 8, wherein the median chromosome X representation is a median experimental X chromosome representation.

10. The method of any one of claims 1 to 9, where the median chromosome Y representation is a median experimental Y chromosome representation.

11. The method of any one of claims 1 to 10, wherein the fitted relationship in (c) is determined prior to (a).

12. The method of any one of claims 1 to 11, wherein the fitted relationship in (c) is determined prior to (b).

13. The method of any one of claims 1 to 12, wherein the counts in (a) are obtained from a pregnant female bearing a male fetus having a chromosome aneuploidy.

14. The method of claim 13, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

15. The method of claim 13, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

16. The method of any one of claims 1 to 15, wherein the set of pregnant females in (c)(i) is a set of about 500 females or more.

17. The method of any one of claims 1 to 16, wherein the set of pregnant females in (c)(ii) is a set of about 500 females or more.

18. The method of any one of claims 1 to 17, wherein the set of pregnant females in (c)(i) and (c)(ii) are the same set.

19. The method of any one of claims 1 to 18, wherein the median chromosome X representation and a median chromosome Y representation is determined for a set of about 500 pregnant females or more.

20. The method of any one of claims 1 to 19, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is provided by a fetal fraction module.

21. The method of any one of claims 1 to 20, comprising normalizing the counts in (a), thereby providing normalized counts mapped to the genomic sections of the reference genome; and which counts in (b) are normalized counts.

22. The method of claim 21, wherein the counts in (b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

23. The method of claim 19 or 22, wherein the normalized counts are provided by a normalization module.

24. The method of any one of claims 1 to 23, wherein the experimental Y chromosome representation in (b) is provided by an experimental representation module.

25. The method of any one of claims 1 to 24, wherein the fetal fraction in (c) is provided by a fetal fraction module.

26. The method of any one of claims 1 to 25, wherein the fitted relationship is provided by a relationship module.

27. The method of any one of claims 24 to 26, wherein the normalized counts are transferred to the experimental representation module from the normalization module.

28. The method of any one of claims 25 to 27, wherein the experimental Y chromosome representation is transferred to the fetal fraction module from the experimental representation module.

29. The method of any one of claims 1 to 28, which comprises obtaining nucleic acid sequence reads.

30. The method of claim 29, wherein the nucleic acid sequence reads are generated by a sequencing module.

31. The method of any one of claims 1 to 30, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

32. The method of any one of claims 1 to 31, which comprises mapping the nucleic acid sequence reads to genomic sections in a segment of the reference genome or to an entire reference genome.

33. The method of claim 32, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

34. The method of any one of claims 1 to 33, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

35. The method of claim 33 or 34, wherein the sequence reads are transferred to the mapping module from the sequencing module.

36. The method of claim 34 or 35, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

37. The method of any one of claims 34 to 36, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

38. The method of any one of claims 34 to 37, wherein an apparatus comprises one or more of the sequencing module, a sequence receiving module, the mapping module, the counting module, the normalization module, the experimental representation module, the relationship module, the fetal fraction module, a comparison module, a range setting module, a categorization module, an adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

39. The method of claim 38, wherein a first apparatus comprises one or more of the normalization module, the experimental representation module, the relationship module and the fetal fraction module.

40. The method of claim 38 or 39, wherein a second apparatus comprises the mapping module and the counting module.

41. The method of any one of claims 38 to 40, wherein a third apparatus comprises the sequencing module.

42. The method of any one of claims 21 to 41, wherein the counts that are normalized are raw counts.

43. The method of any one of claims 21 to 42, wherein the counts that are normalized are filtered.

44. The method of any one of claims 21 to 42, wherein the counts that are normalized are not filtered.

45. The method of any one of claims 1 to 44, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

46. The method of any one of claims 1 to 45, wherein the genomic sections of the reference genome are one or more bins.

47. The method of claim 46, wherein each bin is of about an equal number of contiguous nucleotides.

48. The method of claim 46 or 47, wherein each bin is about 50 kb.

49. The method of any one of claims 1 to 48, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

50. The method of any one of claims 1 to 49, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

51. The method of any one of claims 1 to 50, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

52. The method of claim 51, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

53. The method of claim 51 or 52, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined

from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

54. The method of any one of claims 51 to 53, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

55. The method of any one of claims 51 to 53, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

56. The method of any one of claims 1 to 55, which comprises determining the gender of the fetus.

57. The method of claim 56, wherein the gender of the fetus is determined before performing (b).

58. The method of claim 56 or 57, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

59. The method of claim 56, wherein the gender of the fetus is determined before performing (a).

60. The method of any one of claims 56 to 59, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

61. The method of any one of claims 56 to 59, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

62. The method of any one of claims 1 to 61, which comprises determining the presence or absence of a Y chromosome in the fetus.

63. The method of any one of claims 1 to 62, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

64. The method of claim 63, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

65. The method of any one of claims 1 to 64, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

66. The method of any one of claims 1 to 65, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

67. The method of claim 66, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

68. The method of any one of claims 1 to 67, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

69. The method of any one of claims 1 to 68, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

70. The method of claim 69, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

71. The method of any one of claims 1 to 68, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

72. A system comprising one or more microprocessors and memory,

which memory comprises instructions executable by the one or more microprocessors, and which instructions executable by the one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus

(b) from the counts in (a), generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof,

wherein the sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

(b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

73. The system of claim 72, wherein the X chromosome representation in (c)(ii) is an experimental X chromosome representation.

74. The system of claim 72 or 73, wherein the X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference

genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof.

75. The system of any one of claims 72 to 74, wherein the X chromosome representation in (c)(ii) is determined from a set of pregnant females bearing a male fetus.

76. The system of any one of claims 72 to 75, wherein the X chromosome representation in (c)(ii) is determined for pregnant females bearing a male fetus.

77. The system of any one of claims 72 to 76, wherein the fitted relationship is linear.

78. The system of claim 77, wherein the fraction of the fetal nucleic acid is determined according to the slope and intercept of the fitted relationship, the experimental Y chromosome representation generated in (b) and a median X chromosome representation for a set of pregnant females bearing a female fetus.

79. The system of claim 78, wherein the fraction of the fetal nucleic acid is determined according to equation (62):

$$f = 2\frac{I+S\langle x\rangle - y}{S\langle x\rangle} \qquad (62)$$

wherein *I* is the intercept, *S* is the slope, (x) is the median X chromosome representation for a set of pregnant females bearing a female fetus and *y* is the experimental Y chromosome representation generated in (b).

80. The system of any one of claims 72 to 79, wherein the median chromosome X representation is a median experimental X chromosome representation.

81. The system of any one of claims 72 to 80, where the median chromosome Y representation is a median experimental Y chromosome representation.

82. The system of any one of claims 72 to 81, wherein the fitted relationship in (c) is determined prior to (a).

83. The system of any one of claims 72 to 82, wherein the fitted relationship in (c) is determined prior to (b).

84. The system of any one of claims 72 to 83, wherein the counts in (a) are obtained from a pregnant female bearing a male fetus having a chromosome aneuploidy.

85. The system of claim 84, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

86. The system of claim 84, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

87. The system of any one of claims 72 to 86, wherein the set of pregnant females in (c)(i) is a set of about 500 females or more.

88. The system of any one of claims 72 to 87, wherein the set of pregnant females in (c)(ii) is a set of about 500 females or more.

89. The system of any one of claims 72 to 88, wherein the set of pregnant females in (c)(i) and (c)(ii) are the same set.

90. The system of any one of claims 72 to 18, wherein the median chromosome X representation and a median chromosome Y representation is determined for a set of about 500 pregnant females or more.

91. The system of any one of claims 72 to 90, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is provided by a fetal fraction module.

92. The system of any one of claims 72 to 91, which instructions are configured to normalize the counts in (a), thereby providing normalized counts mapped to the genomic sections of the reference genome; and which counts in (b) are normalized counts.

93. The system of claim 92, wherein the counts in (b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

94. The system of claim 90 or 93, wherein the normalized counts are provided by a normalization module.

95. The system of any one of claims 72 to 94, wherein the experimental Y chromosome representation in (b) is provided by an experimental representation module.

96. The system of any one of claims 72 to 95, wherein the fetal fraction in (c) is provided by a fetal fraction module.

97. The system of any one of claims 72 to 96, wherein the fitted relationship is provided by a relationship module.

98. The system of any one of claims 95 to 97, wherein the normalized counts are transferred to the experimental representation module from the normalization module.

99. The system of any one of claims 96 to 98, wherein the experimental Y chromosome representation is transferred to the fetal fraction module from the experimental representation module.

100. The system of any one of claims 72 to 99, which instructions are configured to obtain nucleic acid sequence reads.

101. The system of claim 100, wherein the nucleic acid sequence reads are generated by a sequencing module.

102. The system of any one of claims 72 to 101, wherein the nucleic acid sequence reads are generated by massively parallel sequencing (MPS).

103. The system of any one of claims 72 to 102, which instructions are configured to map the nucleic acid sequence reads to genomic sections in a segment of the reference genome or to an entire reference genome.

104. The system of claim 103, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

105. The system of any one of claims 72 to 104, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

106. The system of claim 104 or 105, wherein the sequence reads are transferred to the mapping module from the sequencing module.

107. The system of claim 105 or 106, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

108. The system of any one of claims 105 to 107, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

109. The system of any one of claims 105 to 108, wherein an apparatus comprises one or more of the sequencing module, a sequence receiving module, the mapping module, the counting module, the normalization module, the experimental representation module, the relationship module, the fetal fraction module, a comparison module, a range setting module, a categorization module, an adjustment module, a plotting module, an outcome module, a data display organization module or a logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

110. The system of claim 109, wherein a first apparatus comprises one or more of the normalization module, the experimental representation module, the relationship module and the fetal fraction module.

111. The system of claim 109 or 110, wherein a second apparatus comprises the mapping module and the counting module.

112. The system of any one of claims 109 to 111, wherein a third apparatus comprises the sequencing module.

113. The system of any one of claims 92 to 112, wherein the counts that are normalized are raw counts.

114. The system of any one of claims 92 to 113, wherein the counts that are normalized are filtered.

115. The system of any one of claims 92 to 113, wherein the counts that are normalized are not filtered.

116. The system of any one of claims 72 to 115, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

117. The system of any one of claims 72 to 116, wherein the genomic sections of the reference genome are one or more bins.

118. The system of claim 117, wherein each bin is of about an equal number of contiguous nucleotides.

119. The system of claim 117 or 118, wherein each bin is about 50 kb.

120. The system of any one of claims 72 to 119, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

121. The system of any one of claims 72 to 120, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

122. The system of any one of claims 72 to 121, which instructions are configured to determine the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

123. The system of claim 122, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

124. The system of claim 122 or 123, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

125. The system of any one of claims 122 to 124, which instructions are configured to determine the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

126. The system of any one of claims 122 to 124, which instructions are configured to determine the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

127. The system of any one of claims 72 to 126, which instructions are configured to determine the gender of the fetus.

128. The system of claim 127, wherein the gender of the fetus is determined before performing (b).

129. The system of claim 127 or 128, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

130. The system of claim 127, wherein the gender of the fetus is determined before performing (a).

131. The system of any one of claims 127 to 130, which instructions are configured to determine whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

132. The system of any one of claims 127 to 130, which instructions are configured to determine whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

133. The system of any one of claims 72 to 132, which instructions are configured to determine the presence or absence of a Y chromosome in the fetus.

134. The system of any one of claims 72 to 133, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

135. The system of claim 134, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

136. The system of any one of claims 72 to 134, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

137. The system of any one of claims 72 to 136, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

138. The system of claim 137, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

139. The system of any one of claims 72 to 138, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

140. The system of any one of claims 72 to 139, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

141. The system of claim 140, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

142. The system of any one of claims 72 to 139, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

143. An apparatus comprising one or more processors and memory,

which memory comprises instructions executable by the one or more processors and which memory comprises counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus; and

which instructions executable by the one or more processors are configured to:

(a) from the counts, generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(b) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and

the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

144. A computer program product tangibly embodied on a computer-readable medium, comprising instructions that when executed by one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) from the counts in (a), generate an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof;

(c) determine the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (b) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and
the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

145. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determining the fraction of fetal nucleic acid in the blood of the pregnant female according to the experimental X chromosome representation and an expected X chromosome representation, which expected X chromosome representation is a ratio of (i) the number of the genomic

sections of the reference genome in the X chromosome, and (ii) the number of the genomic sections of the reference genome in the genome or segment thereof.

146. The method of claim 145, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined in (c) according to a ratio of the experimental X chromosome representation and the expected X chromosome representation.

147. The method of claim 145 or 146, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is determined according to equation AC.

148. A method for determining the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female, comprising:

(a) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(b) generating an experimental X chromosome representation, which experimental X chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the X chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof; and

(c) from the experimental X chromosome representation, determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to a relationship determined from (i) the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy, and (ii) the experimental X chromosome representation.

149. The method of claim 148, wherein the fraction of fetal nucleic acid in the blood of the pregnant female bearing a fetus having a chromosome aneuploidy is determined according to equation AB.

150. The method of claim 148 or 149, wherein the fraction of fetal nucleic acid determined in (c)(i) and the experimental X chromosome representation in (c)(ii) are derived from greater than about 500 subjects.

151. The method of any one of claims 145 to 150, wherein the fraction of fetal nucleic acid in the blood of the pregnant female is provided by a fetal fraction module.

152. The method of any one of claims 148 to 151, wherein the relationship is a linear relationship.

153. The method of any one of claims 148 to 152, wherein the chromosome aneuploidy is a trisomy 21, trisomy 18 and/or trisomy 13.

154. The method of any one of claims 148 to 153, wherein the chromosome aneuploidy is a sex chromosome aneuploidy.

155. The method of any one of claims 148 to 154, wherein the relationship is:

$$F = k - r(MCRx),$$

wherein F is the fraction, MCRx is the experimental X chromosome representation, k is an intercept from the linear relationship and r is a slope from the linear relationship.

156. The method of claim 155, wherein the fraction of fetal nucleic acid in circulating cell-free nucleic acid from the blood of a pregnant female is determined by the relationship in 4.3 and the experimental X chromosome representation.

157. The method of any one of claims 148 to 156, wherein the fraction of fetal nucleic acid determined for nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that comprises use of sequence reads mapped to genomic sections of a reference genome.

158. The method of any one of claims 148 to 156, wherein the fraction of fetal nucleic acid determined for nucleic acid from nucleic acid from the blood of a pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is determined by a process that does not utilize sequence reads mapped to genomic sections of a reference genome.

159. The method of claim 158, wherein the process comprises mass spectrometry.

160. The method of any one of claims 148 to 159, wherein the pregnant female bearing a fetus having a chromosome aneuploidy in (c)(i) is different from the pregnant female in (a).

161. The method of claim 160, wherein the pregnant female bearing a fetus having a chromosome aneuploidy is bearing a male fetus.

162. The method of any one of claims 148 to 161, wherein the nucleic acid from the blood of the pregnant female in (c)(i) is circulating cell-free nucleic acid.

163. The method of any one of claims 145 to 162, comprising normalizing the counts mapped to the genomic sections of the reference genome, thereby providing normalized counts mapped to the genomic sections of the reference genome; and which counts in 1(b) and 3(b) are normalized counts.

164. The method of any one of claims 145 to 163, wherein the counts in 1(b) and 3(b) are normalized by GC content, bin-wise normalization, GC LOESS, PERUN, GCRM, or combinations thereof.

165. The method of claim 163, wherein the normalized counts are provided by a normalization module.

166. The method of any one of claims 145 to 165, wherein the experimental X chromosome representation in 1(b) and 3(b) is provided by a representation module.

167. The method of any one of claims 145 to 166, wherein expected X chromosome representation in 1(c) and 3(c) is determined by an expected representation module.

168. The method of any one of claims 145 to 167, wherein the fetal fraction in 1(c) and 3(c) is provided by a fetal fraction module.

169. The method of any one of claims 148 to 168, wherein the relationship is provided by a relationship module.

170. The method of any one of claims 166 to 169, wherein the normalized mapped counts are transferred to the experimental representation module from the normalization module.

171. The method of any one of claims 167 to 170, wherein the normalized mapped counts are transferred to the expected representation module from the normalization module.

172. The method of any one of claims 168 to 171, wherein the experimental X chromosome representation is transferred to the fetal fraction module from the experimental representation module.

173. The method of any one of claims 168 to 172, wherein the expected X chromosome representation is transferred to the fetal fraction module from the expected representation module.

174. The method of any one of claims 145 to 173, which comprises obtaining nucleic acid sequence reads.

175. The method of claim 174, wherein the nucleic acid sequence reads are generated by a sequencing module.

176 The method of claim 174 or 175, wherein the nucleic acid sequencing reads are generated by massively parallel sequencing (MPS).

177. The method of claim 174 or 175, which comprises mapping the nucleic acid sequence reads to the genomic sections of the reference genome or to an entire reference genome.

178. The method of claim 177, wherein the nucleic acid sequence reads are mapped to the genomic sections of the reference genome by a mapping module.

179. The method of any one of claims 145 to 178, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are counted by a counting module.

180. The method of any one of claims 178 to 179, wherein the sequence reads are transferred to the mapping module from the sequencing module.

181. The method of any one of claims 179 to 180, wherein the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the counting module from the mapping module.

182. The method of any one of claims 179 to 181, wherein the counts of the nucleic acid sequence reads mapped to the genomic sections of the reference genome are transferred to the normalization module from the counting module.

183. The method of any one of claims 145 to 182, wherein an apparatus comprises one or more of a sequencing module, sequence receiving module, mapping module, counting module, normalization module, comparison module, range setting module, categorization module, adjustment module, plotting module, outcome module, data display organization module or logic processing module, which apparatus comprises, or is in communication with, a processor that is capable of implementing instructions from one or more of the modules.

184. The method of claim 183, wherein a first apparatus comprises one or more of the normalization module, the comparison module, the range setting module, the adjustment module, and the outcome module.

185. The method of any one of claims 179 to 184, wherein a second apparatus comprises the mapping module and the counting module.

186. The method of any one of claims 176 to 185, wherein a third apparatus comprises the sequencing module.

187. The method of any one of claims 163 to 186, wherein the counts that are normalized are raw counts.

188. The method of any one of claims 163 to 187, wherein the counts that are normalized are filtered.

189 The method of any one of claims 163 to 187, wherein the counts that are normalized are not filtered.

190. The method of any one of claims 145 to 189, wherein the genomic sections of the reference genome are chromosomes or genomic sections thereof.

191. The method of any one of claims 145 to 190, wherein the genomic sections of the reference genome are one or more bins.

192. The method of claim 191, wherein each bin is of about an equal number of contiguous nucleotides.

193. The method of claim 191 or 192, wherein each bin is about 50 kb.

194. The method of any one of claims 145 to 193, wherein the fetal fraction is provided with an accuracy of equal to or greater than 90% and/or a precision equal to or greater than 90%.

195. The method of any one of claims 145 to 194, wherein the sequence reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus are from a test sample obtained from the pregnant female bearing a male fetus.

196. The method of any one of claims 145 to 195, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome in the reference genome.

197. The method of claim 196, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome is performed before performing (b).

198. The method of claim 196 or 35, wherein determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

199. The method of any one of claims 196 to 198, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (b), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (b) and (c).

200. The method of any one of claims 196 to 198, which comprises determining the presence or absence of nucleic acid sequence reads mapped to the Y chromosome before performing (a), and if nucleic acid sequence reads mapped to the Y chromosome are present, then performing (a), (b) and (c).

201. The method of any one of claims 145 to 198, which comprises determining the gender of the fetus.

202. The method of claim 201, wherein the gender of the fetus is determined before performing (b).

203. The method of claim 201 or 202, wherein the gender of the fetus is determined from the counts of nucleic acid sequence reads mapped to genomic sections of the reference genome obtained in (a).

204. The method of claim 201, wherein the gender of the fetus is determined before performing (a).

205. The method of any one of claims 201 to 204, which comprises determining whether the gender of the fetus is male or female before performing (b), and if the gender of the fetus is determined as being male, then performing (b) and (c).

206. The method of any one of claims 201 to 204, which comprises determining whether the gender of the fetus is male or female before performing (a), and if the gender of the fetus is determined as being male, then performing (a), (b) and (c).

207. The method of any one of claims 145 to 206, which comprises determining the presence or absence of a Y chromosome in the fetus.

208. The method of any one of claims 145 to 207, wherein the genomic sections of the Y chromosome in (b)(ii) are a subset of genomic sections of the Y chromosome.

209. The method of claim 208, wherein the subset of genome sections of the Y chromosome comprises one or more polynucleotides located within the first 28 Mb from the 5' end of the Y chromosome.

210. The method of any one of claims 145 to 208, wherein counts of sequence reads that map to both chromosome Y and chromosome X are excluded before performing (b).

211. The method of any one of claims 145 to 210, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome are substantially unique to the Y chromosome.

212. The method of claim 211, wherein greater than 80% or more of the genomic sections in the Y chromosome are substantially unique to the Y chromosome.

213. The method of any one of claims 145 to 212, wherein the counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome do not map to genomic sections of the reference genome in the X chromosome.

214. The method of any one of claims 145 to 213, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) are counts of sequence reads mapped to autosomes.

215. The method of claim 214, wherein the counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof in (b)(ii) do not include sequence reads mapped to sex chromosomes.

216. The method of any one of claims 145 to 213, wherein the counts of sequence reads mapped to the reference genome in the genome or a segment thereof in (b)(ii) are counts of sequence reads mapped to all chromosomes from which reads are obtained.

217. A system comprising one or more microprocessors and memory,

which memory comprises instructions executable by the one or more microprocessors, and which instructions executable by the one or more processors are configured to:

(a) determine a fraction of fetal nucleic acid in a sample, which sample comprises circulating cell-free nucleic acid from the blood of a pregnant female bearing a fetus;

(b) obtain counts of sequence reads mapped to portions of a reference genome, which sequence reads are from the nucleic acid in the sample;

(c) calculate a genomic section level for each of the portions of the reference genome, thereby providing calculated genomic section levels; and

(d) determine fetal ploidy according to (i) the calculated genomic section levels for a subset of portions of the reference genome and (ii) the fraction of fetal nucleic acid determined in (a).

218. The system of claim 217, wherein the fetal fraction is determined from a first part of the test sample and the genomic section levels are determined from a second part of the test sample.

219. The system of claim 217, wherein calculating the genomic section level for each of the portions of the reference genome comprises normalizing counts of reads mapped to the reference genome according to guanine and cytosine (GC) content for each of the portions.

220. The system of claim 219, comprising instructions configured to:

(1) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i)

the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions; and

(2) calculate the genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels.

221. The system of claim 217, wherein the subset of portions of the reference genome in (d)(i) is portions of a chromosome or segment thereof.

222. The system of claim 221, wherein the chromosome is chosen from chromosome 13, chromosome 18 and chromosome 21.

223. The system of claim 217, comprising instructions configured to determine a reference count and an uncertainty value according to the reference count:

wherein the reference count is determined according to calculated genomic section levels for a subset of portions of the reference genome for one or more pregnant females bearing a fetus;

wherein the subset of portions of the reference genome for one or more pregnant females are known to be euploid for the fetus and/or the mother;

wherein the reference count is not determined from the sample; and

wherein the reference count is determined from the same subset of portions of the reference genome as in (d).

224. The system of claim 223, wherein the reference count is normalized by bin-wise normalization, normalization by GC content, linear and nonlinear least squares regression, LOESS, GC LOESS, LOWESS, PERUN, RM, GCRM and combinations thereof.

225. The system of claim 217, comprising instructions configured to determine a maternal ploidy.

226. The system of claim 225, wherein the fetal ploidy is determined in (d) according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a) and (iii) a maternal ploidy.

227. The system of claim 226, wherein the fetal ploidy is determined according to (i) the calculated genomic section levels for a subset of portions of the reference genome, (ii) the fraction of fetal nucleic acid determined in (a), (iii) a maternal ploidy, (iv) the reference count and (v) an uncertainty value σ for the reference count.

228. The system of claim 227, wherein the fraction of fetal nucleic acid determined in (a) is fixed at its determined value and fetal ploidy *X* is determined according to Equation 8 below, or a derivation thereof:

$$y_i = (1 - F)M_i f_i + FXf_i \qquad (8)$$

where y_i represents the calculated genomic section level for portion *i* of a reference genome, *F* represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for *i*, *X* represents the fetal ploidy, and M_i represents the maternal ploidy of portion *i.*

229. The system of claim 228, comprising instructions configured to determine the sum of squared residuals according to equation (8) and for multiple bins *i* for a subset of portions of the reference genome.

230. The system of claim 229, wherein the fetal fraction is fixed at a value determined in (a) and the fetal ploidy is varied to optimize the sum of squared residuals according to equation (8) or a variation thereof.

231. The system of claim 230, comprising instructions configured to determine a linear regression according to the sum of square residuals.

232. The system of claim 227, wherein the fetal ploidy is determined according to Equation 20 below:

$$X = \frac{\sum_{i=1}^{N} \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^{N} \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}} \quad . \qquad (20)$$

wherein y_i represents the calculated genomic section level for portion i of a reference genome, F represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for i, σ represents the uncertainty value for f_i, X represents the fetal ploidy, and M_i represents the maternal ploidy of portion i.

233. The system of claim 227, wherein the fetal ploidy is determined according to Equation 21 below:

$$X = \frac{\Xi_{fy} - (1-F)\Xi_{ff}}{F\Xi_{ff}} = \frac{\Xi_{fy}}{F\Xi_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Xi_{fy}}{\Xi_{ff}} - 1\right) \qquad (21)$$

wherein $\Xi_{ff} = \sum_{i=1}^{N} \frac{f_i^2}{\sigma_i^2}$, $\Xi_{fy} = \sum_{i=1}^{N} \frac{y_i f_i}{\sigma_i^2}$, y_i represents the calculated genomic section level for portion i of a reference genome, F represents the fraction of fetal nucleic acid determined in (a), f_i represents a reference count for i, σ represents the uncertainty value for f_i, and X represents the fetal ploidy.

234. The system of claim 217, comprising instructions configured to determine the presence or absence of a fetal chromosome aneuploidy according to the fetal ploidy determined in (d).

235. The system of claim 234, wherein the fetal ploidy determined in (d) is greater than about 1.25 and the presence of a fetal chromosome aneuploidy is determined.

236. The system of claim 235, wherein the fetal ploidy determined in (d) is less than about 1.25 and the absence of a fetal chromosome aneuploidy is determined.

237. The system of claim 235, wherein the fetal chromosome aneuploidy is a trisomy.

238. The system of claim 237, wherein the trisomy is selected from a trisomy of chromosome 13, 18 and 21.

239. The system of claim 217, wherein determining the fraction of fetal nucleic acid comprises analyzing the calculated genomic sections levels for a subset of portions of the reference genome, which subset is a first subset, the subset in (d) is a second subset, and the first subset of portions of the reference genome is portions of a Y chromosome or a segment thereof.

240. The system of claim 217, wherein determining the fraction of fetal nucleic acid comprises:

(1) obtaining counts of nucleic acid sequence reads mapped to genomic sections of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from the blood of a pregnant female bearing a male fetus;

(2) from the counts in (1), generating an experimental Y chromosome representation, which experimental Y chromosome representation is a ratio of (i) counts of sequence reads mapped to the genomic sections of the reference genome in the Y chromosome, and (ii) counts of sequence reads mapped to genomic sections of the reference genome in the genome or segment thereof;

(3) determining the fraction of the fetal nucleic acid in the blood of the pregnant female according to the experimental Y chromosome representation generated in (2) and a fitted relationship, wherein:

the fitted relationship is between (i) an experimental Y chromosome representation determined from a set of pregnant females bearing a male fetus and (ii) an X chromosome representation determined from a set of pregnant females; and the fitted relationship is fitted to a median chromosome X representation and a median chromosome Y representation for a set of pregnant females bearing a female fetus.

241. The system of claim 217, wherein determining the fraction of fetal nucleic acid comprises analyzing one or more loci in sample nucleic acid, wherein at least one of the one or more loci vary between fetal nucleic acid and maternal nucleic acid.

242. The system of claim 241, wherein the one or more loci comprise one or more polymorphic sites, comprising:

(1) enriching nucleic acid in a first part of the test sample for a plurality of polymorphic sites;

(2) obtaining nucleotide sequences for some or all of the polymorphic sites by a sequencing process;

(3) analyzing the nucleotide sequences of (2); and

(4) determining the fraction of fetal nucleic acid based on the analysis of (3), wherein the polymorphic sites and number thereof result in at least five polymorphic sites being informative for determining the fetal fraction for at least 90% of samples.

243. The system of claim 241, wherein the one or more loci comprise one or more methylation regions, comprising:

(1) contacting the test sample with one or more agents that differentially modify methylated nucleic acid and unmethylated nucleic acid, which sample nucleic acid comprises differentially methylated fetal nucleic acid and maternal nucleic acid, the combination of the fetal nucleic acid and the maternal nucleic acid comprising total nucleic acid in the sample, thereby generating differentially modified sample nucleic acid; and

(2) determining the fraction of fetal nucleic acid in the sample based on the differentially modified nucleic acid.

244. The system of claim 243, wherein the one or more agents are methylation sensitive restriction enzymes.

245. A system comprising one or more microprocessors and memory,

which memory comprises instructions executable by the one or more microprocessors, and which instructions executable by the one or more processors are configured to:

(a) obtain counts of nucleic acid sequence reads mapped to portions of a reference genome, which sequence reads are reads of circulating cell-free nucleic acid from a sample from a pregnant female;

(b) determine a guanine and cytosine (GC) bias for each of the portions of the reference genome for multiple samples from a fitted relation for each sample between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) GC content for each of the portions;

(c) calculate a genomic section level for each of the portions of the reference genome from a fitted relation between (i) the GC bias and (ii) the counts of the sequence reads mapped to each of the portions of the reference genome, thereby providing calculated genomic section levels, whereby bias in the counts of the sequence reads mapped to each of the portions of the reference genome is reduced in the calculated genomic section levels; and

(d) determine the presence or absence of a complex genetic variation for the sample from the pregnant female according to the calculated genomic section levels, wherein the complex genetic variation comprises two or more genetic variations chosen from a whole chromosome aneuploidy, a microduplication and a microdeletion, and wherein the determination is generated from the nucleic acid sequence reads.

246. The system of claim 245, wherein the whole chromosome aneuploidy is a trisomy of a chromosome selected from chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 22, X and Y.

247. The system of claim 245 or 246, wherein the complex genetic variation comprises a double chromosome aneuploidy.

248. The system of any one of claims 245 to 247, wherein the complex genetic variation comprises a fetal genetic variation.

249. The system of any one of claims 245 to 248, wherein the microdeletion and/or microduplication is located in a chromosome selected from chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, and X.

250. The system of any one of claims 245 to 249, comprising instructions configured to, prior to (b), calculate a measure of error for the counts of sequence reads mapped to some or all of the portions of the reference genome and remove or weight the counts of sequence reads for certain portions of the reference genome according to a threshold of the measure of error.

251. The system of claim 250, wherein the threshold is selected according to a standard deviation gap between a first genomic section level and a second genomic section level of 3.5 or greater.

252. The system of claim 250, wherein the measure of error is an R factor.

253. The system of any one of claims 245 to 252, wherein the counts of sequence reads for a portion of the reference genome having an R factor of about 7% to about 10% are removed prior to (b).

254. The system of any one of claims 245 to 253, wherein the fitted relation in (b) is a fitted linear relation.

255. The system of any one of claims 245 to 254, wherein a slope of the fitted relation in (b) is determined by linear regression.

256. The system of any one of claims 245 to 255, wherein each GC bias is a GC bias coefficient, which GC bias coefficient is the slope of the linear relationship between (i) the counts of the sequence reads mapped to each of the portions of the reference genome, and (ii) the GC content for each of the portions.

257. The system of claim of any one of claims 245 to 253, wherein the fitted relation in (b) is a fitted non-linear relation.

258. The system of claim 257, wherein each GC bias comprises a GC curvature estimation.

259. The system of any one of claims 245 to 258, wherein the fitted relation in (c) is linear.

260. The system of any one of claims 245 to 259, wherein a slope of the fitted relation in (c) is determined by a linear regression.

261. The system of any one of claims 245 to 260, wherein the fitted relation in (b) is linear, the fitted relation in (c) is linear and the genomic section level L_i is determined for each of the portions of the reference genome according to Equation α:

$$L_i = (m_i - G_iS)\, I^{-1} \qquad \text{Equation } \alpha$$

wherein G_i is the GC bias, I is the intercept of the fitted relation in (c), S is the slope of the relation in (c), m_i is measured counts mapped to each portion of the reference genome and i is a sample.

262. The system of any one of claims 245 to 261, wherein the number of portions of the reference genome is about 40,000 or more portions.

263. The system of any one of claims 245 to 262, wherein each portion of the reference genome comprises a nucleotide sequence of a predetermined length.

264. The system of any one of claims 245 to 263, comprising instructions configured to:

(i) identify a first elevation of the calculated genomic section levels significantly different than a second elevation of the calculated genomic section levels, which first elevation is for a first set of portions, and which second elevation is for a second set of portions;

(ii) determine an expected elevation range for a homozygous and heterozygous copy number variation according to an uncertainty value for a segment of the genome;

(iii) adjust the first elevation by a predetermined value when the first elevation is within the expected elevation range, thereby providing an adjustment of the first elevation, and thereby providing adjusted genomic section levels; and

(iv) determine the presence or absence of a chromosome aneuploidy in the sample from the pregnant female in (a) according to the calculated genomic section levels comprising the adjustment of (iii), wherein the determination is made with a reduced likelihood of being a false positive or false negative.

265. The system of any one of claims 72 to 142 or 217 to 264, comprising a machine configured to generate sequence reads from circulating cell-free nucleic acid obtained from a pregnant female.

266. The system of claim 265, where the machine configured to generate sequence reads transfers the sequence reads to a machine comprising the one or more microprosessors and memory, which memory comprises the instructions configured to carry out any one of claims 72 to 142 or 217 to 264.

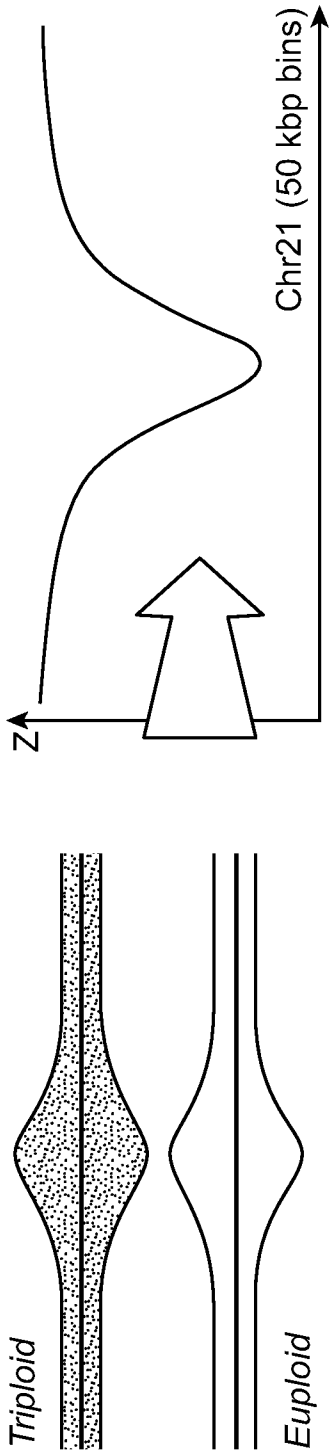
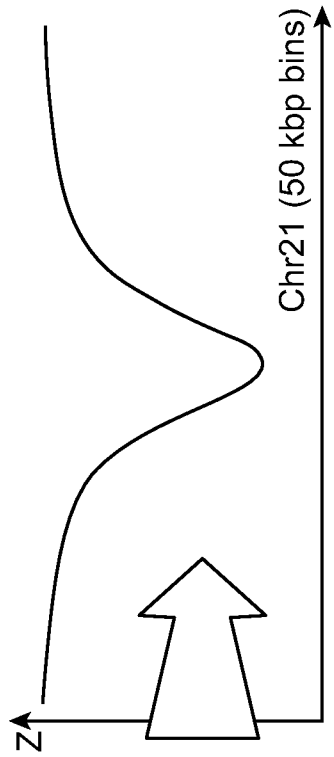


FIG. 1



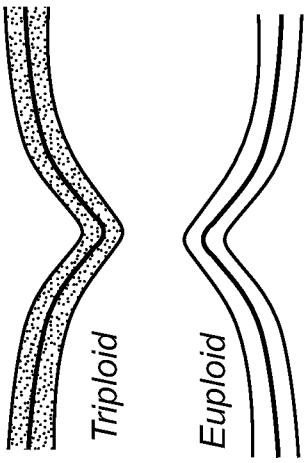


FIG. 2

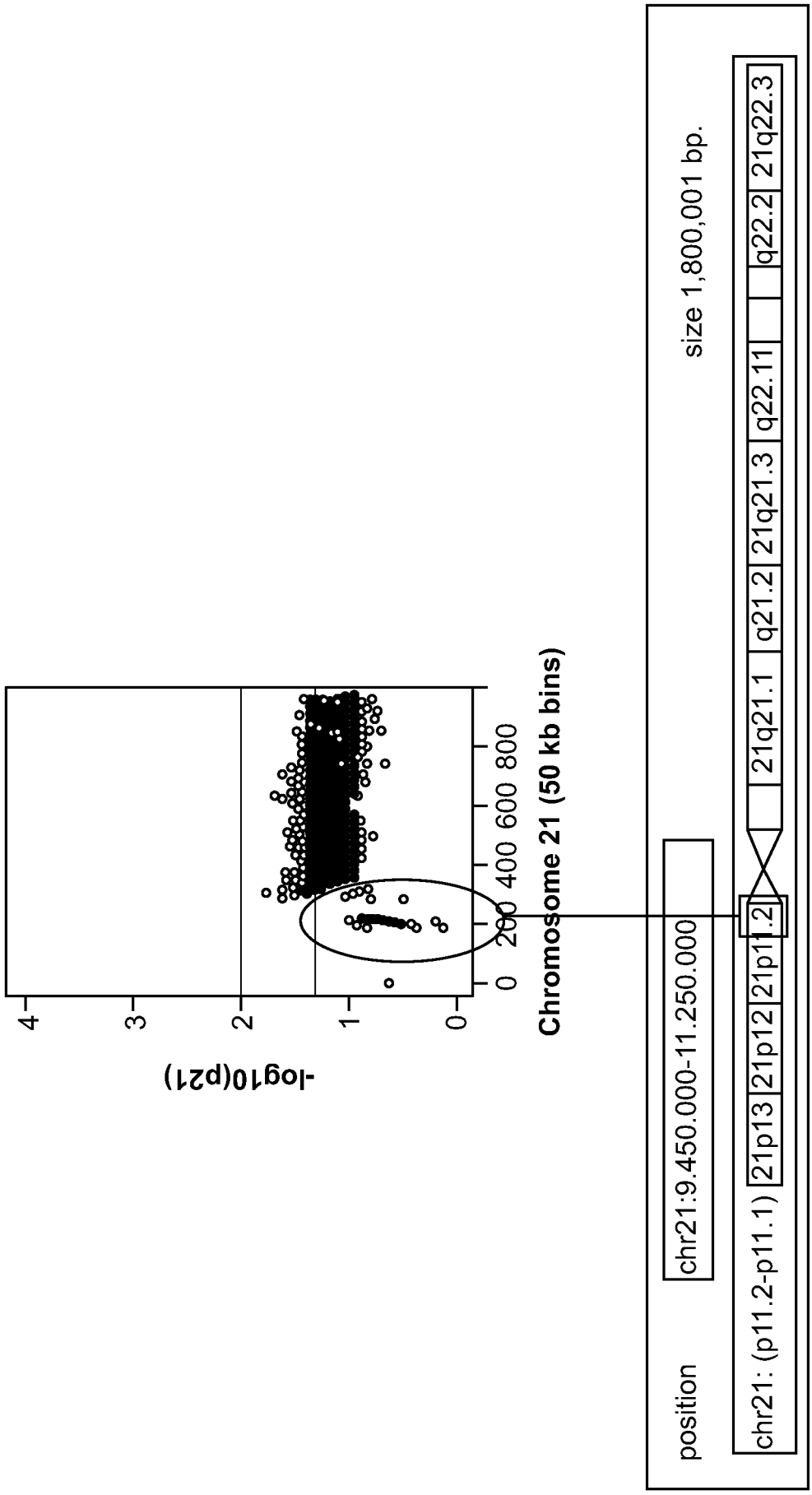

-log10(p21)
0
1
2
3
4
0
200
400
600
800
Chromosome 21 (50 kb bins)
position
chr21:9.450.000-11.250.000
size 1,800,001 bp.
chr21: (p11.2-p11.1)
21p13
21p12
21p11.2
21q11.1
21q11.2
21q21.3
21q22.11
21q22.2
21q22.3

FIG. 3

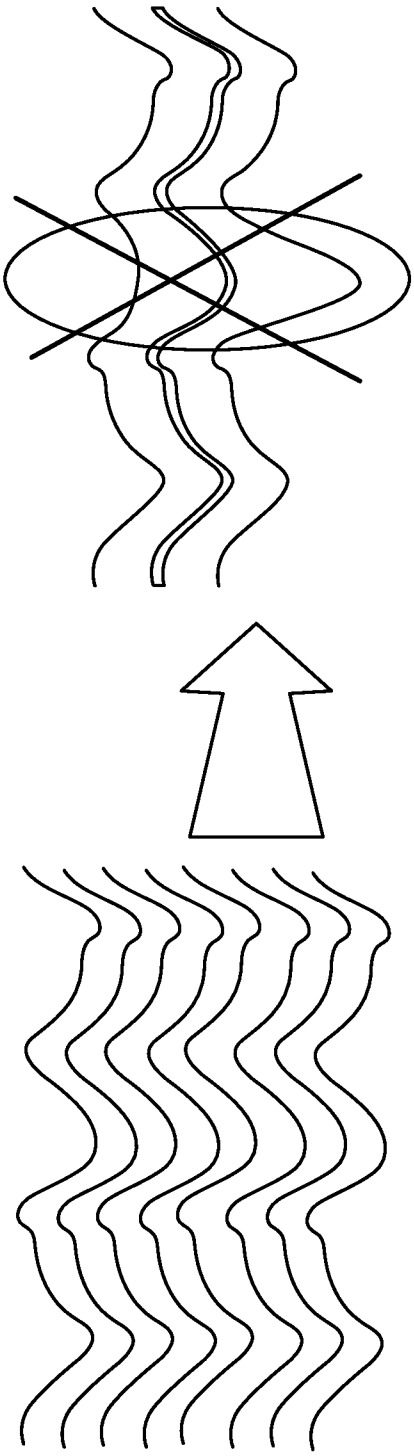


FIG. 4

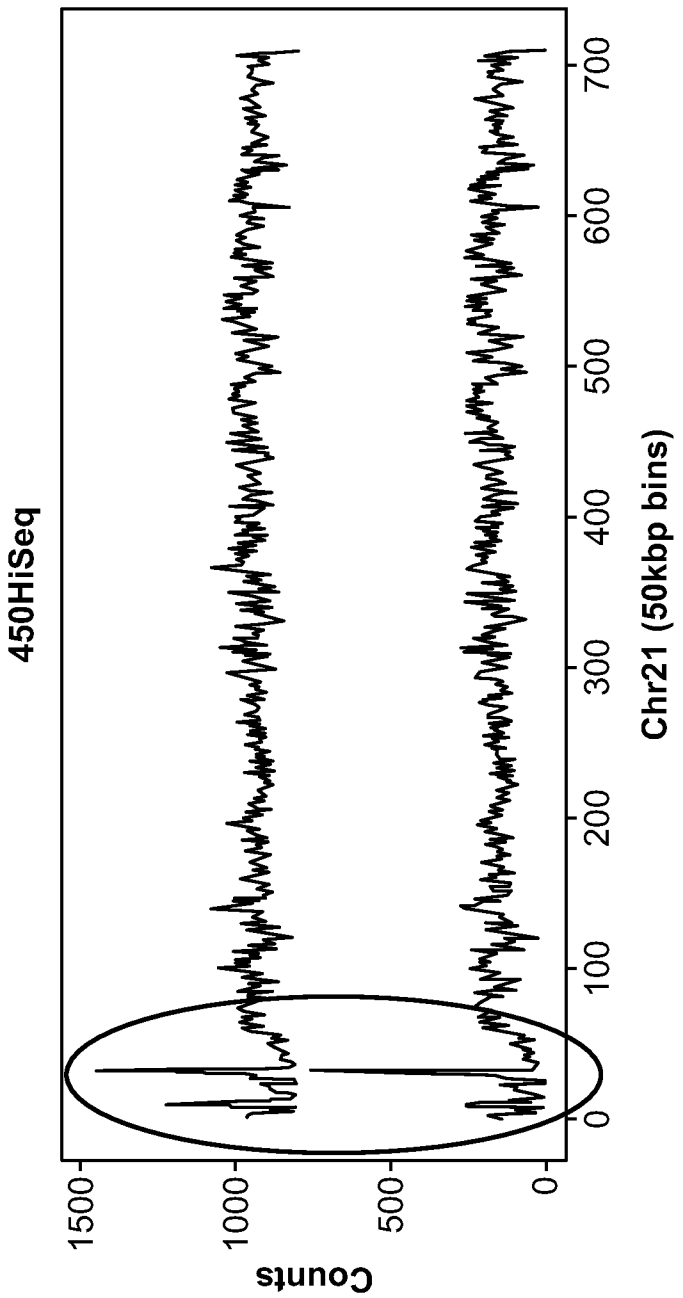

450HiSeq
Counts
0
500
1000
1500
0
100
200
300
400
500
600
700
Chr21 (50kbp bins)

FIG. 5

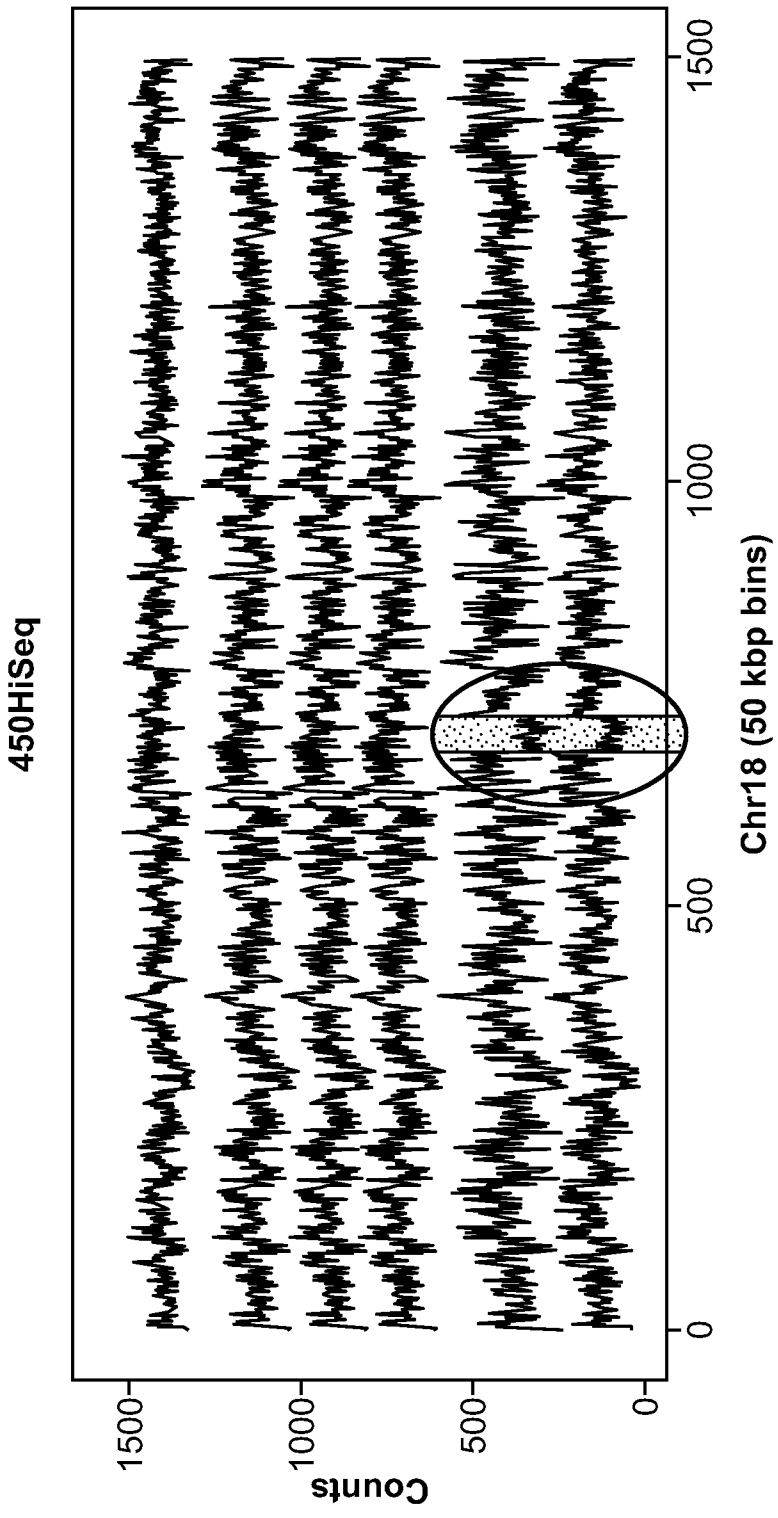

450HiSeq
Counts
0
500
1000
1500
Chr18 (50 kbp bins)
0
500
1000
1500

FIG. 6

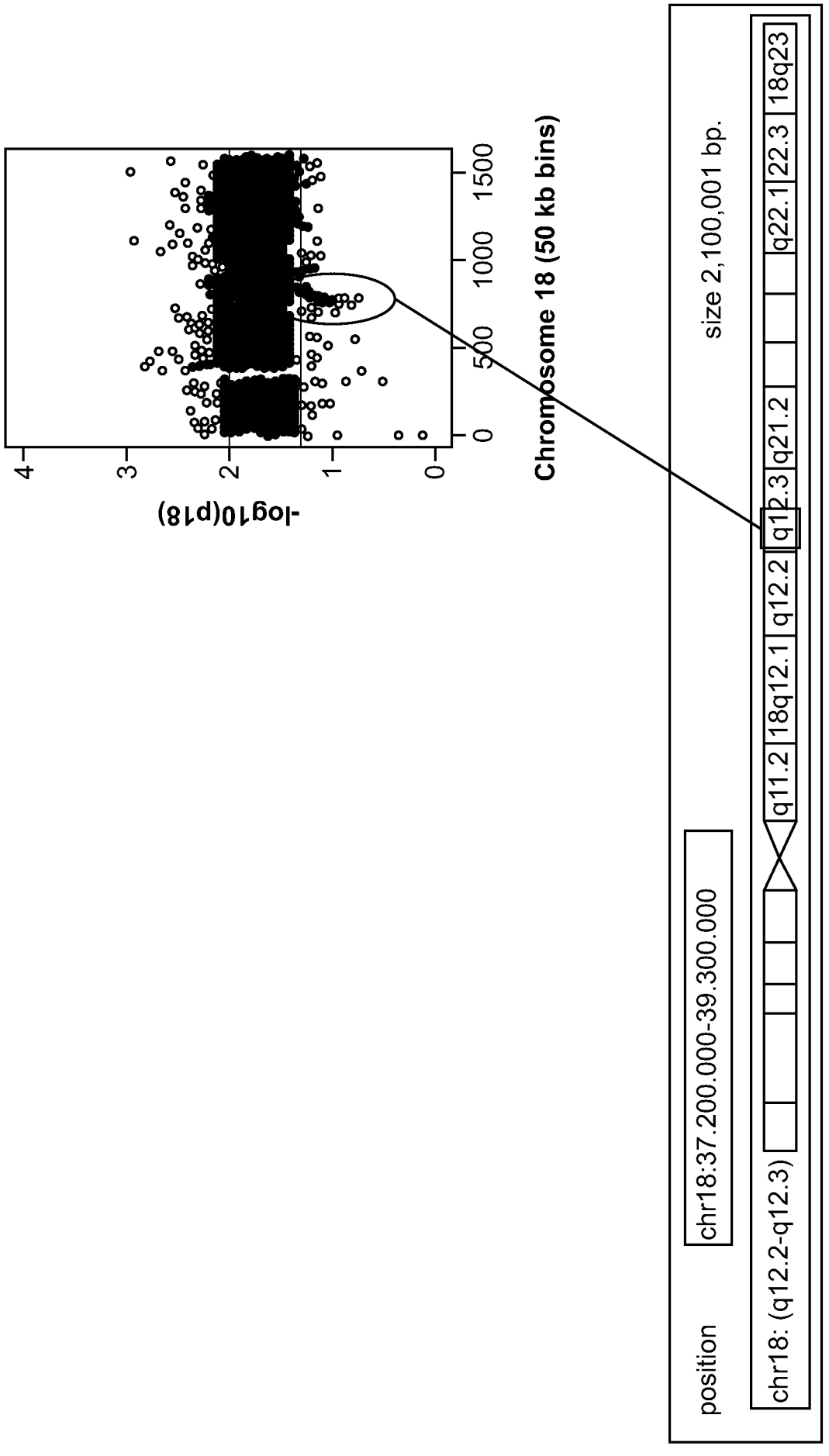

-log10(p18)
0
1
2
3
4
0
500
1000
1500
Chromosome 18 (50 kb bins)
position
chr18:37.200.000-39.300.000
size 2,100,001 bp.
chr18: (q12.2-q12.3)
q11.2
18q12.1
q12.2
q12.3
q21.2
q22.1
22.3
18q23

FIG. 7

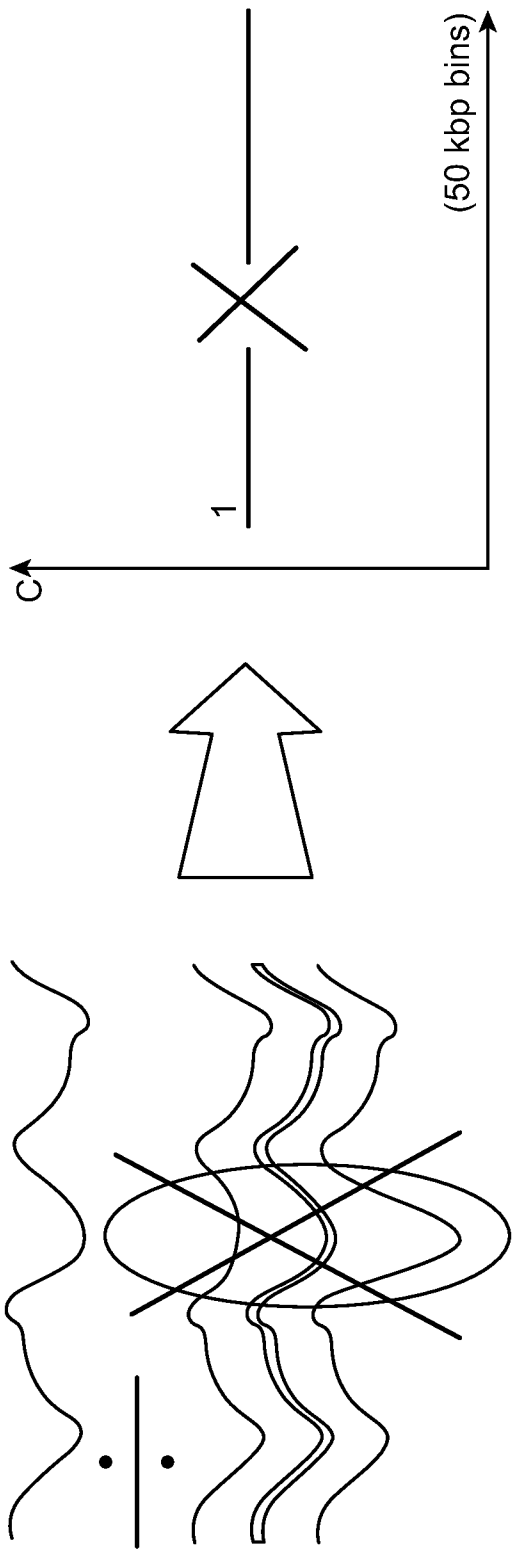


FIG. 8

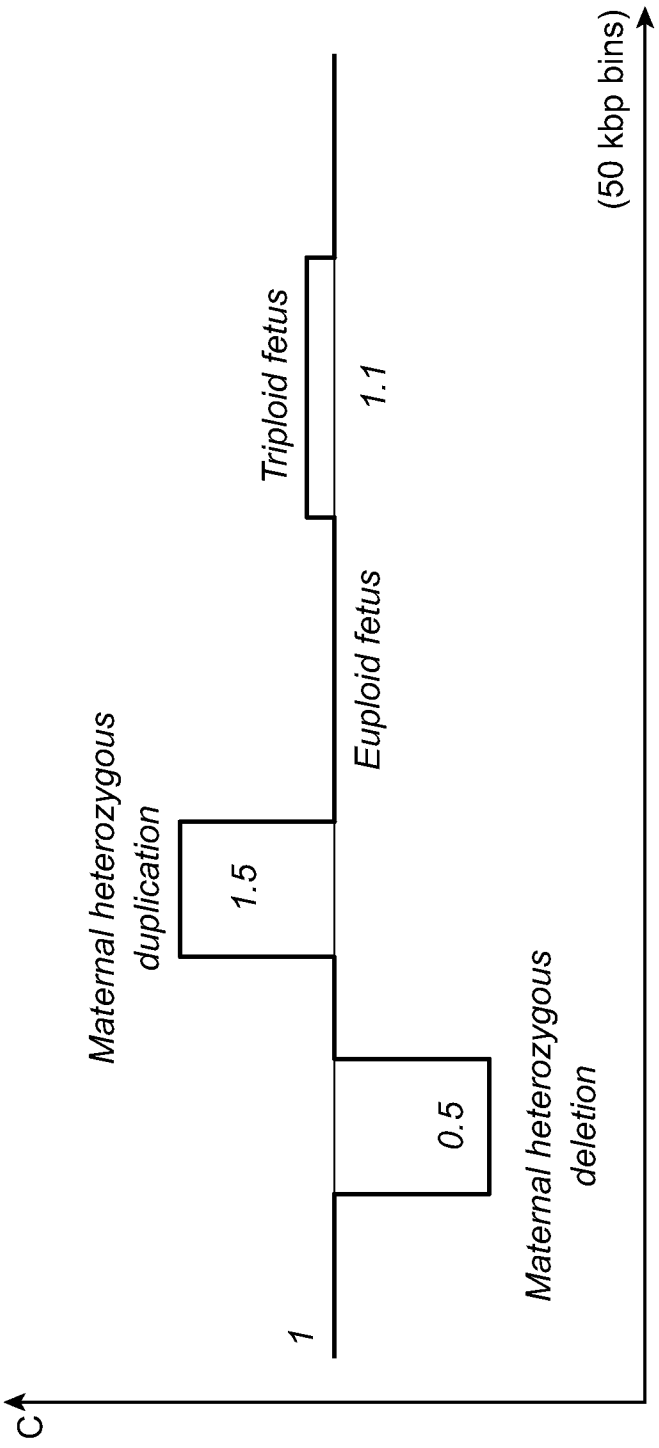

C
(50 kbp bins)
1
0.5
1.5
1.1
Maternal heterozygous deletion
Euploid fetus
Maternal heterozygous duplication
Triploid fetus

FIG. 9

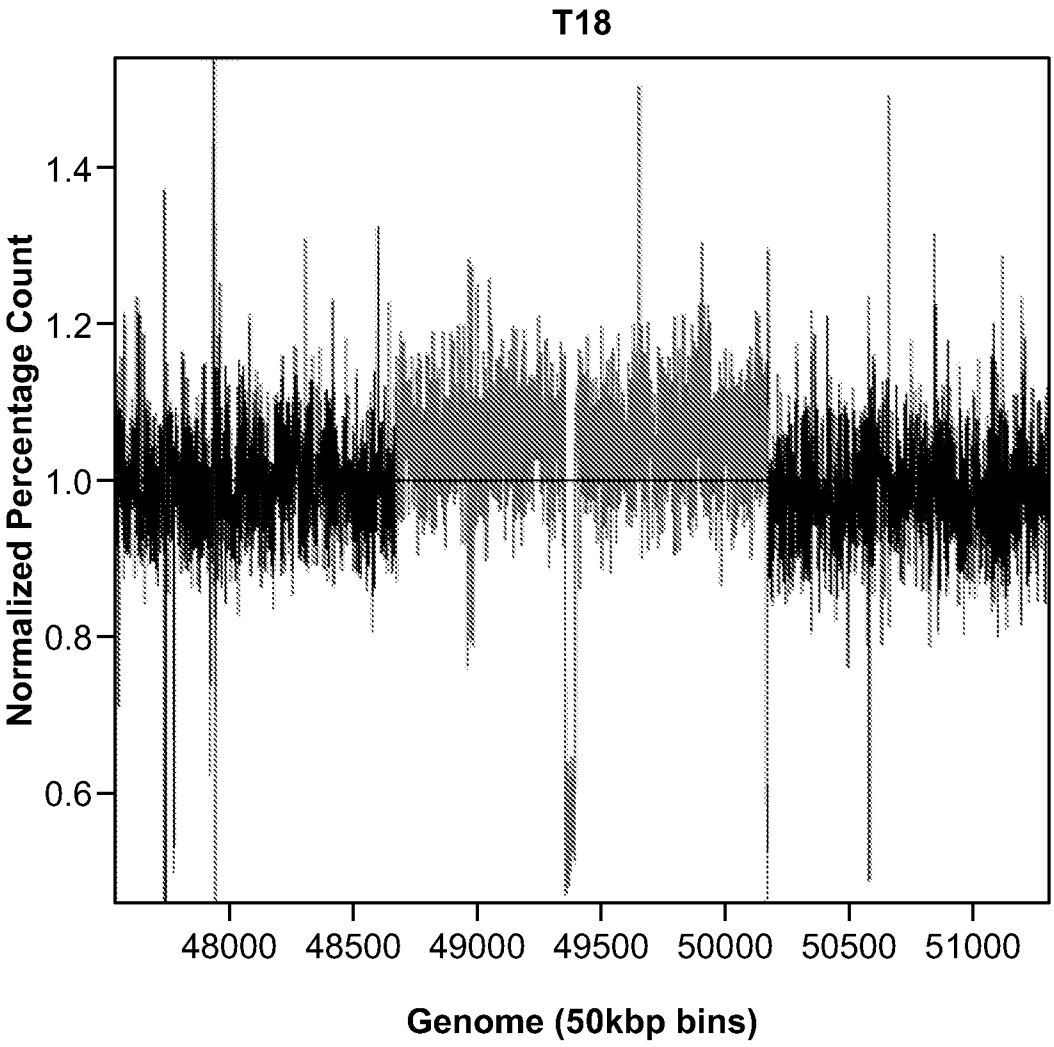

T18
Normalized Percentage Count
1.4
1.2
1.0
0.8
0.6
48000
48500
49000
49500
50000
50500
51000
Genome (50kbp bins)

FIG. 10

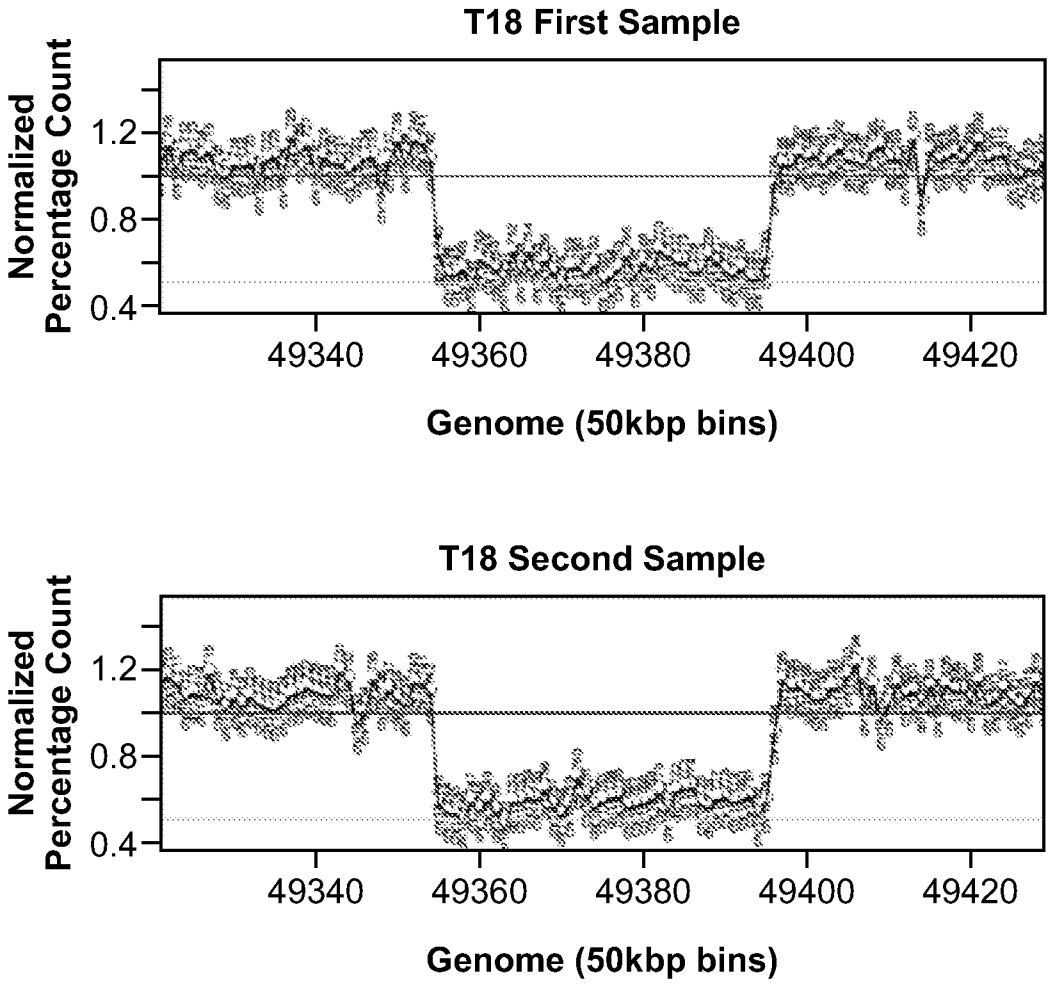

T18 First Sample
Normalized Percentage Count
1.2
0.8
0.4
49340
49360
49380
49400
49420
Genome (50kbp bins)
T18 Second Sample
Normalized Percentage Count
1.2
0.8
0.4
49340
49360
49380
49400
49420
Genome (50kbp bins)

FIG. 11

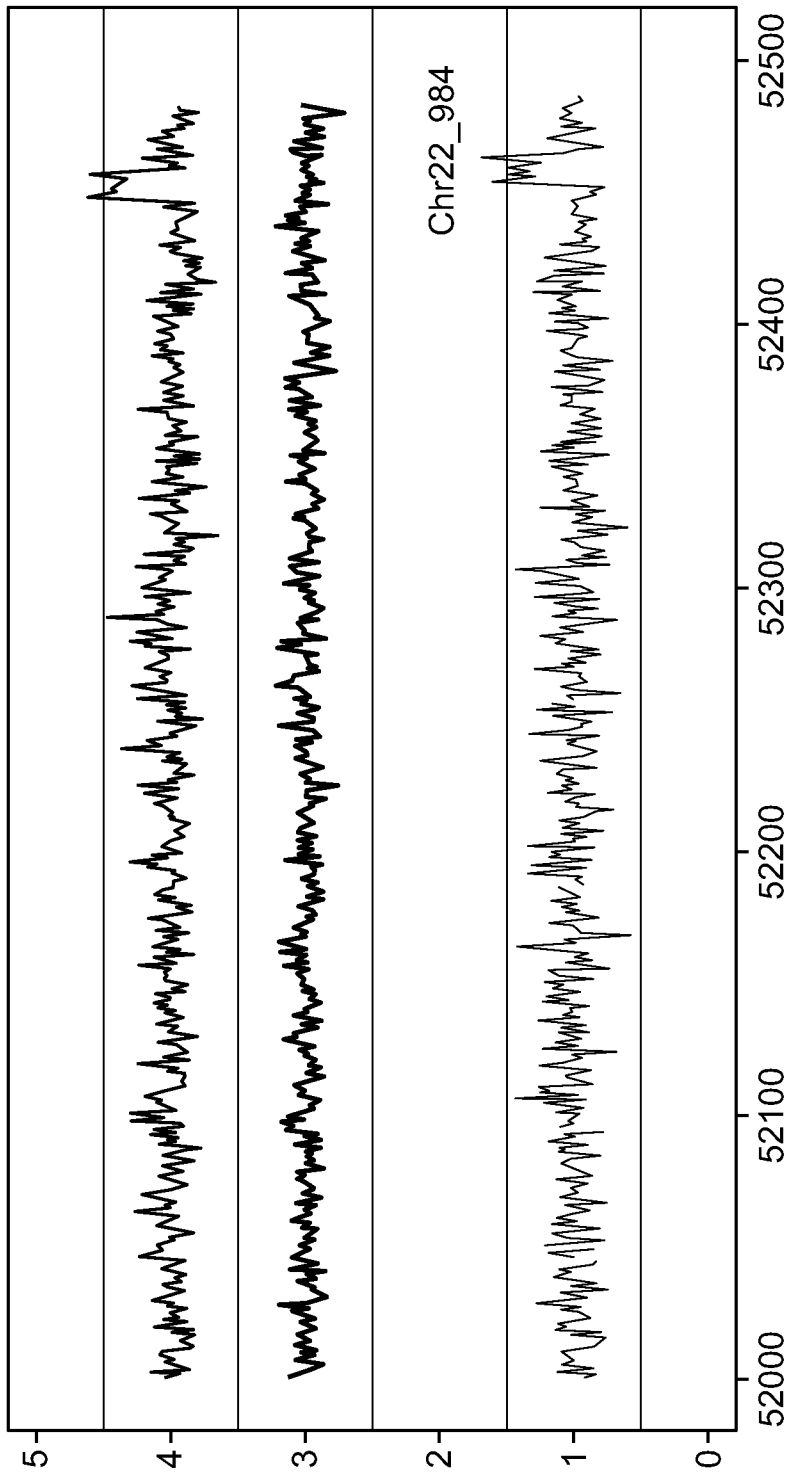

Chr22_984
5
4
3
2
1
0
52000
52100
52200
52300
52400
52500
50 kbp bins

FIG. 12

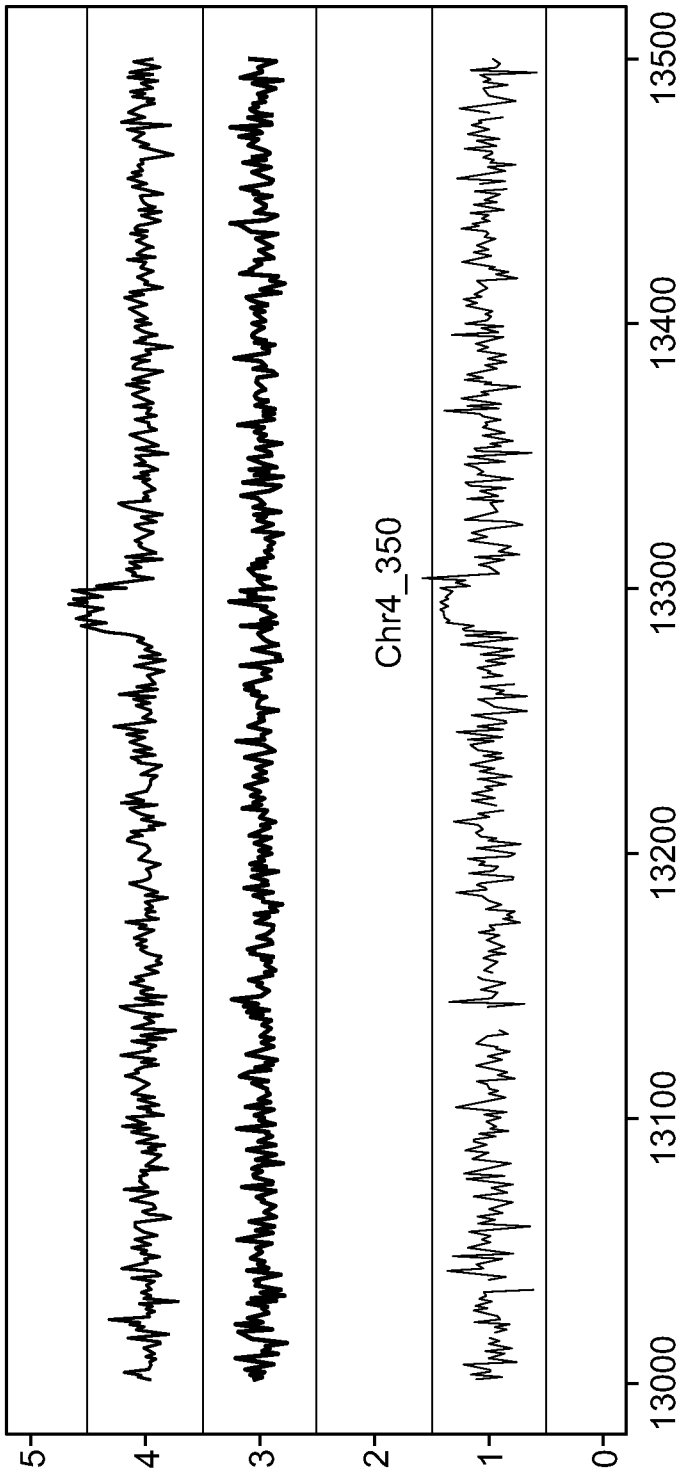

5
4
3
2
1
0
Chr4_350
13000
13100
13200
13300
13400
13500
50 kbp bins

FIG. 13

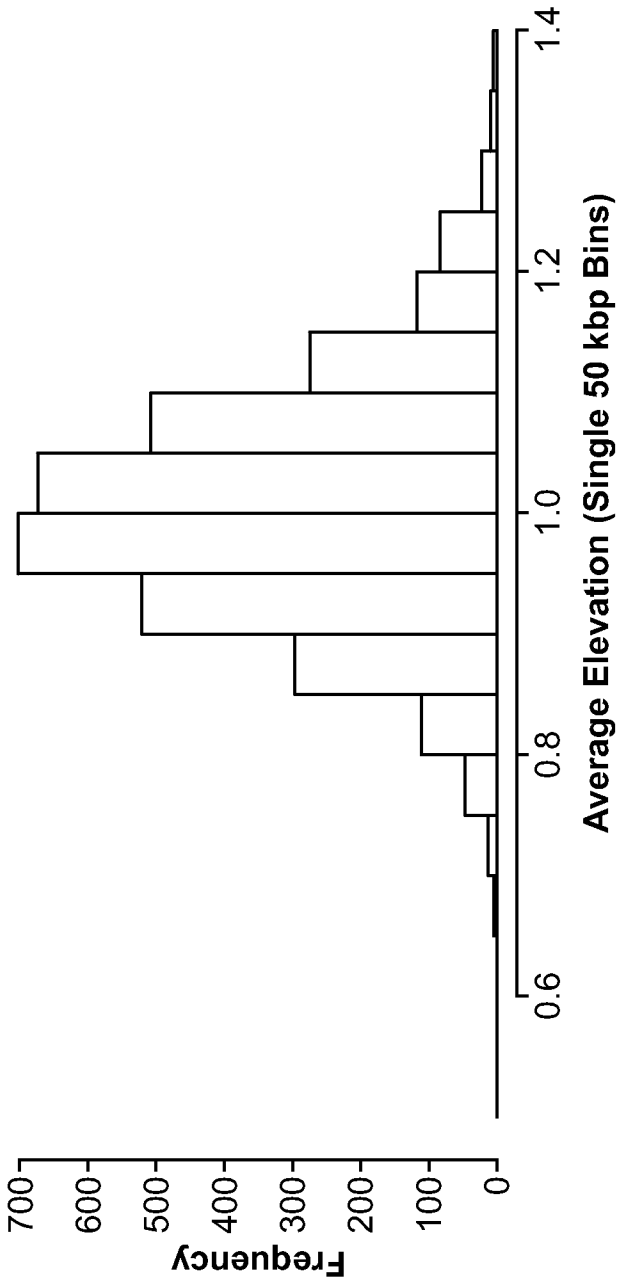

Frequency
0
100
200
300
400
500
600
700
0.6
0.8
1.0
1.2
1.4
Average Elevation (Single 50 kbp Bins)

FIG. 14

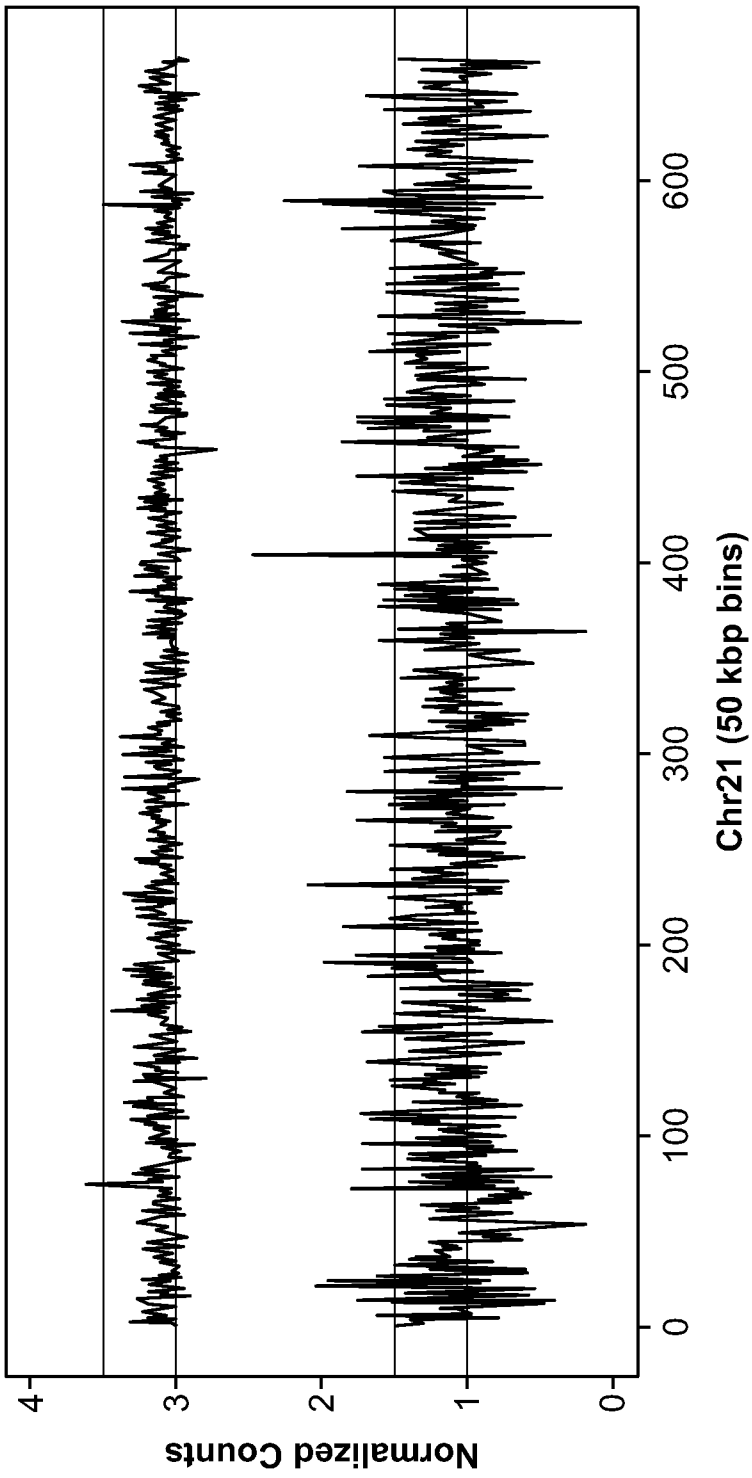

Normalized Counts
0
1
2
3
4
0
100
200
300
400
500
600
Chr21 (50 kbp bins)

FIG. 15

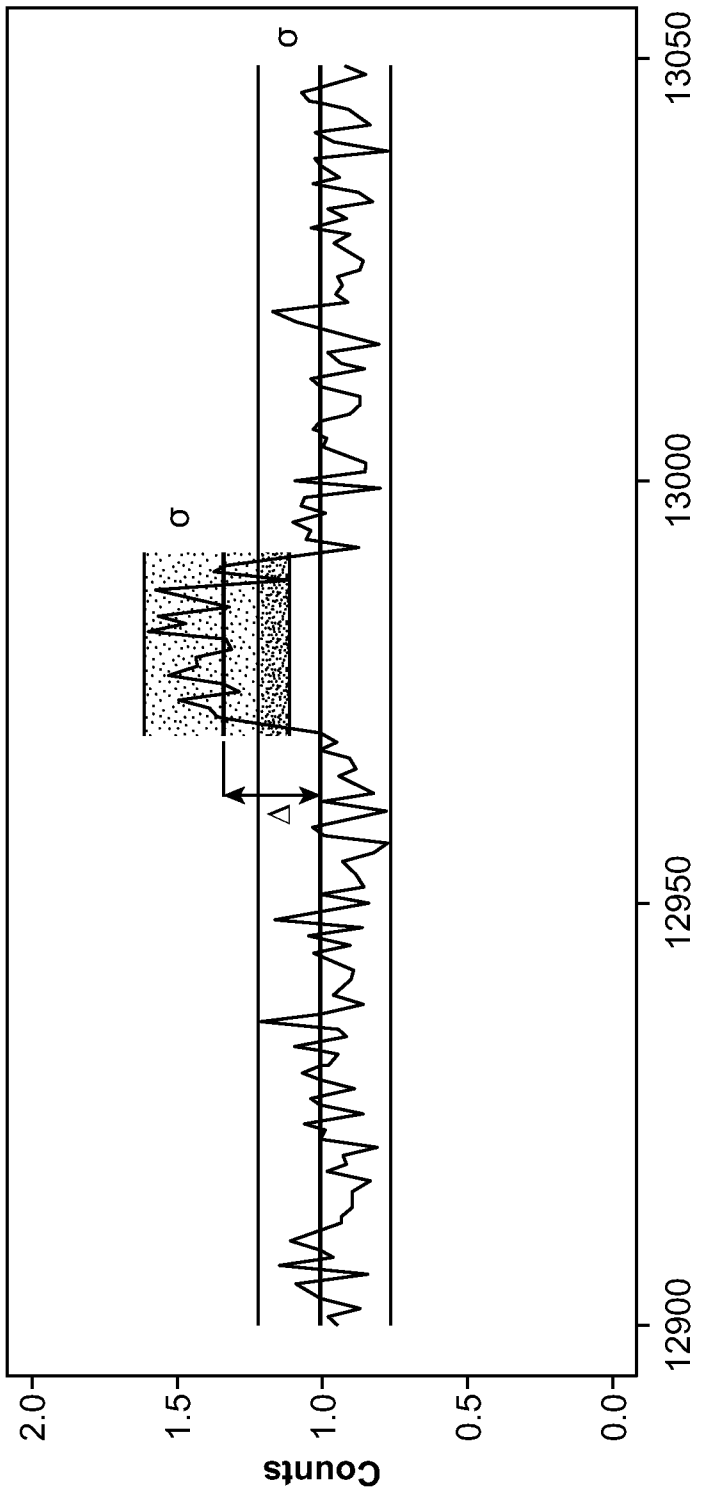

Counts
2.0
1.5
1.0
0.5
0.0
12900
12950
13000
13050
Chr4 (50 kbp bins)
σ
σ
Δ

FIG. 16

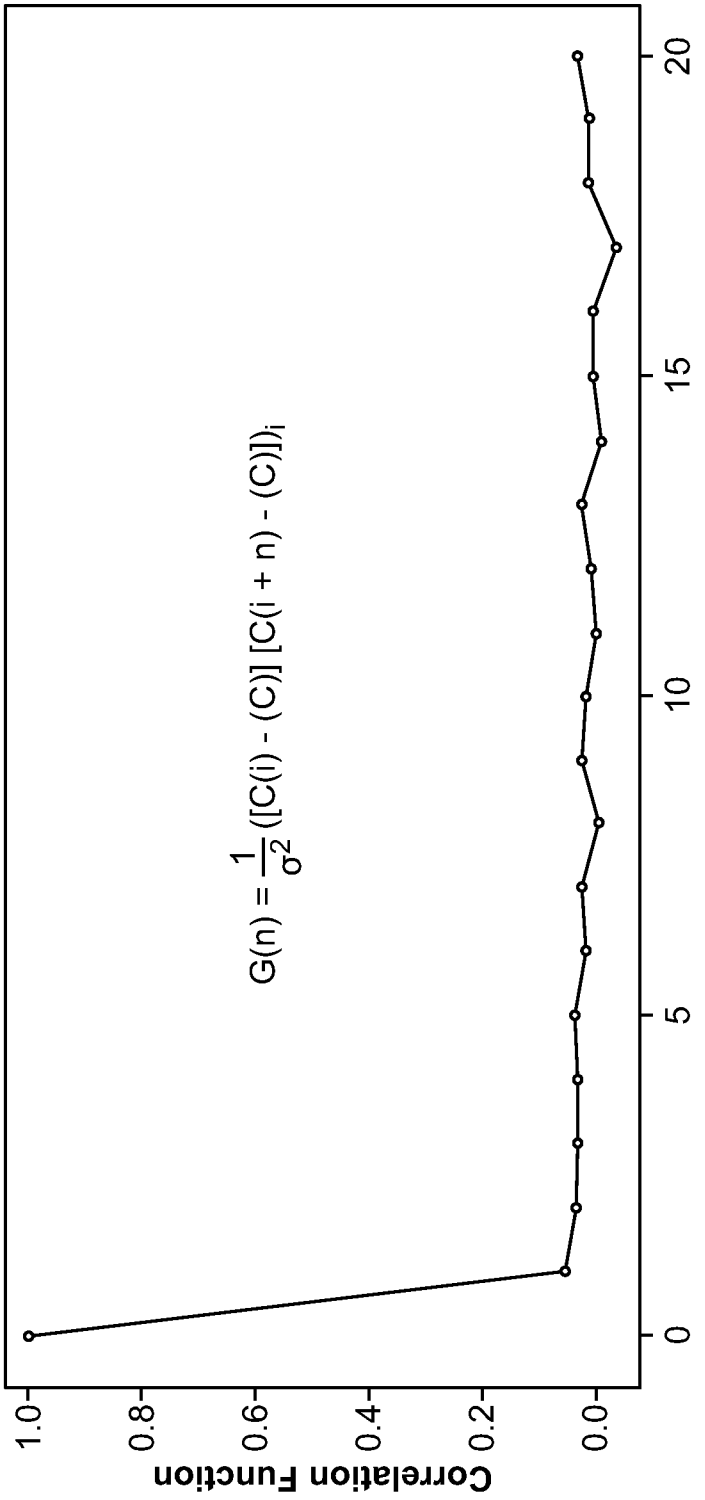

$G(n) = \frac{1}{\sigma^2}([C(i) - (C)][C(i+n) - (C)])_i$
Correlation Function
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
n (Chr5, 50kbp bins)

FIG. 17

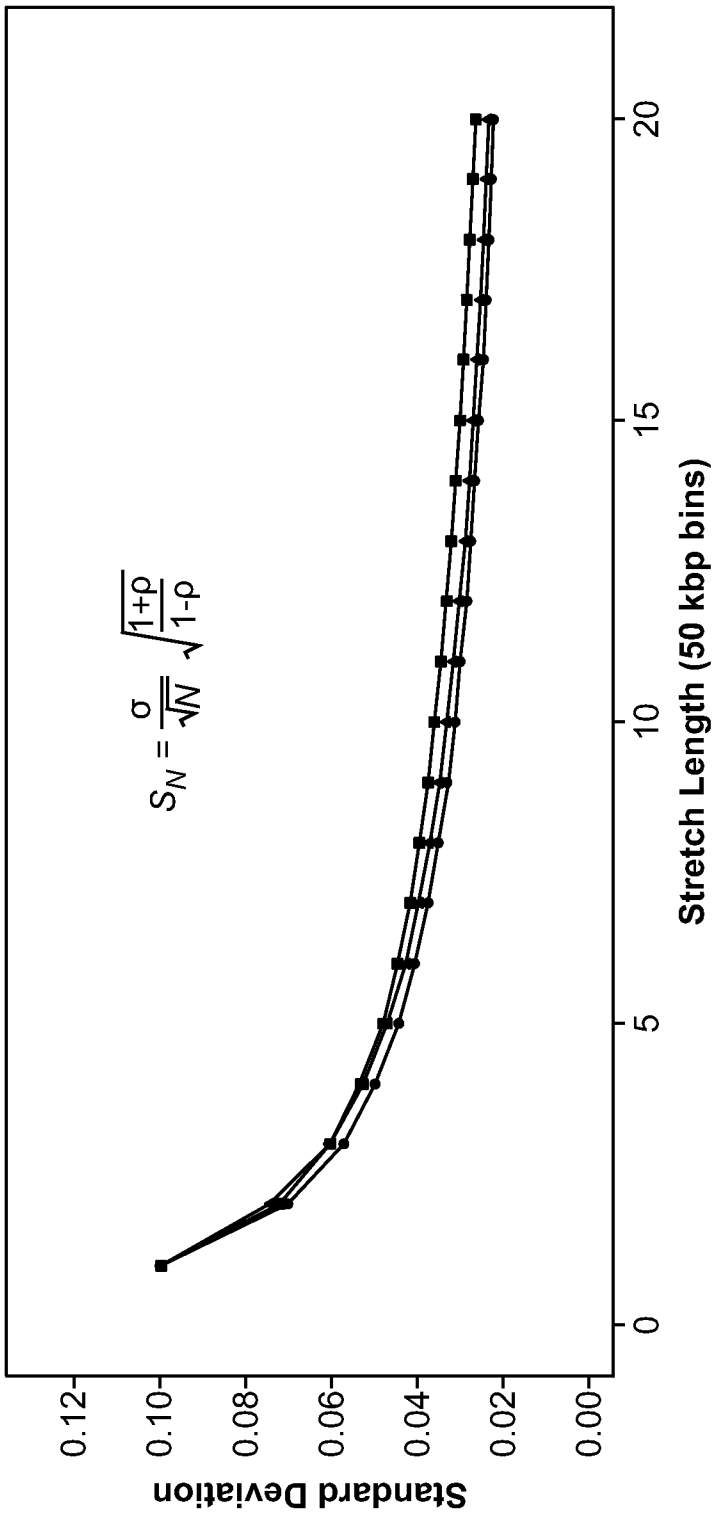

Standard Deviation
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
Stretch Length (50 kbp bins)
$S_N = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{1+\rho}{1-\rho}}$

FIG. 18

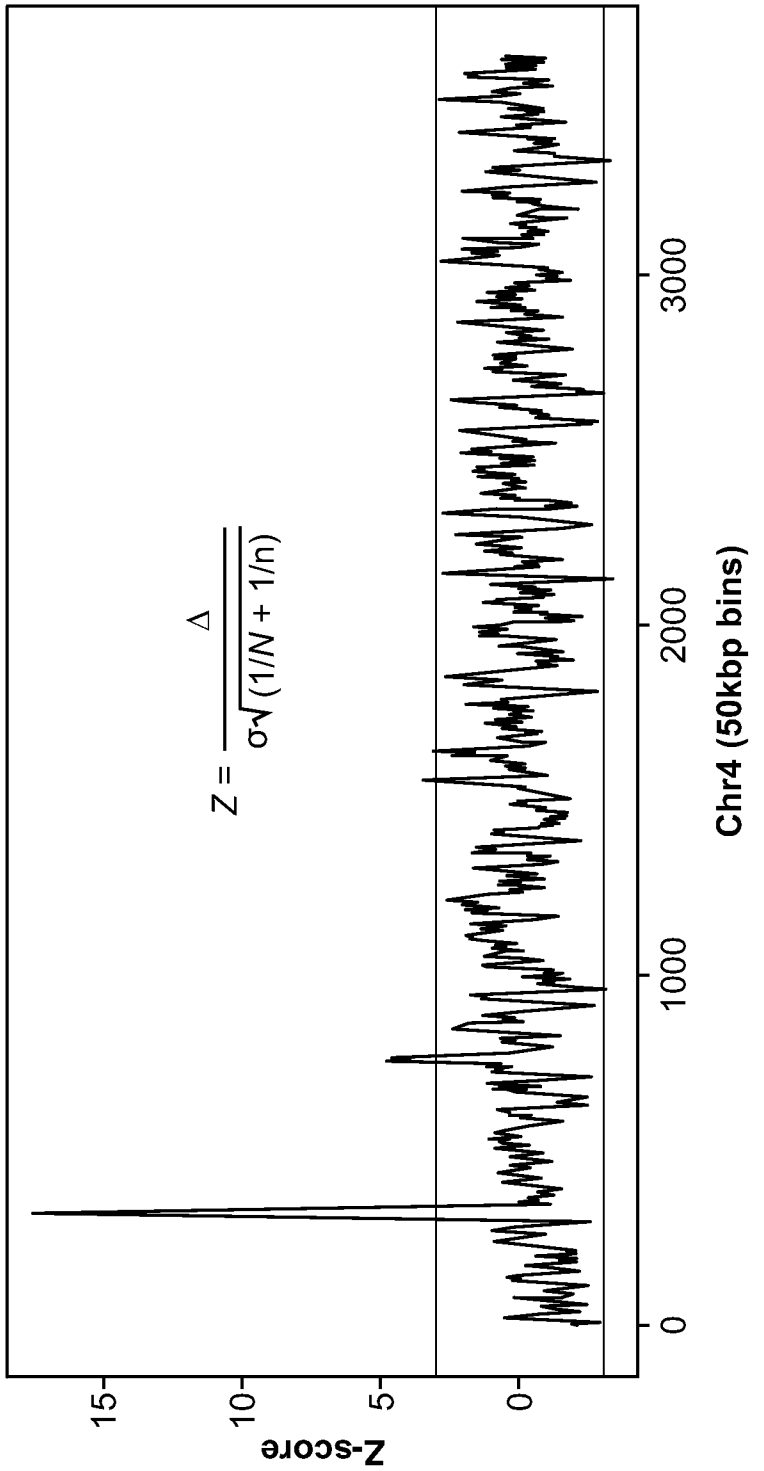


FIG. 19

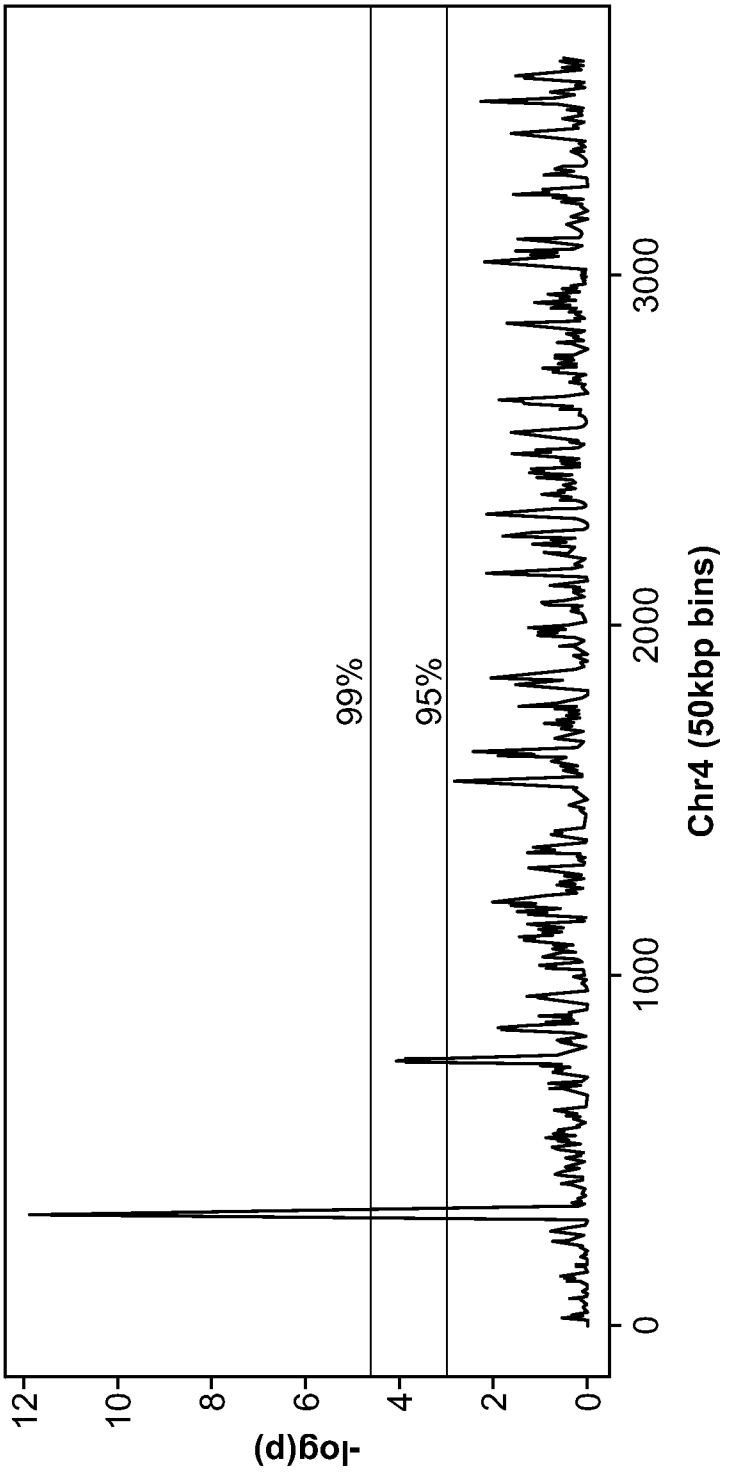


FIG. 20

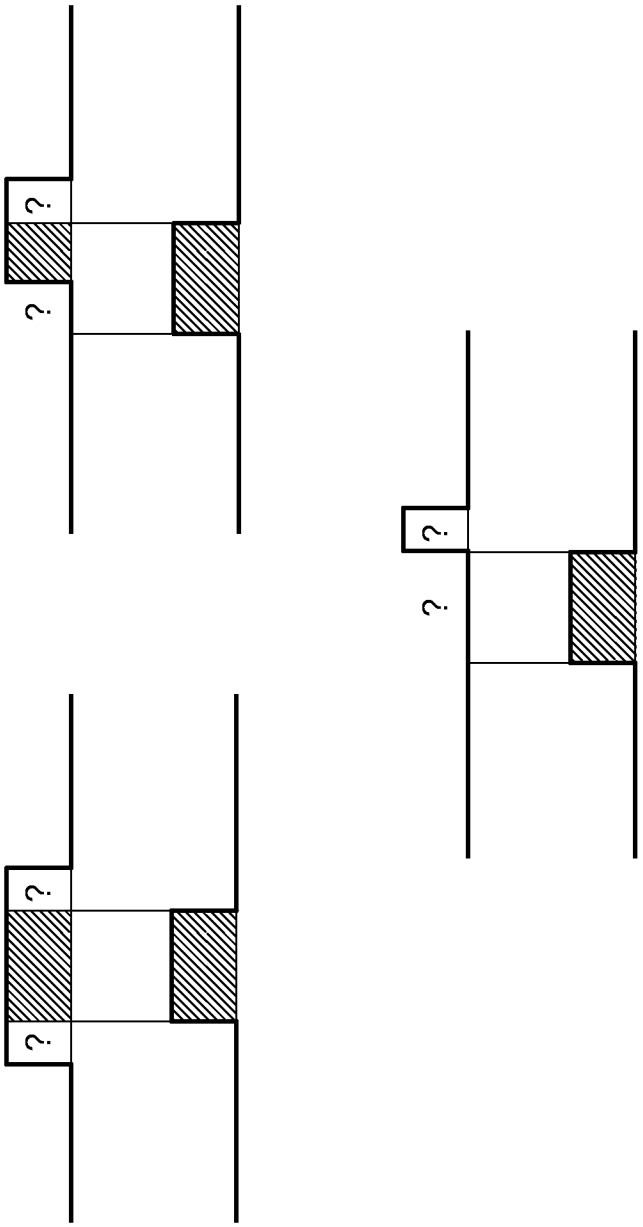


FIG. 21

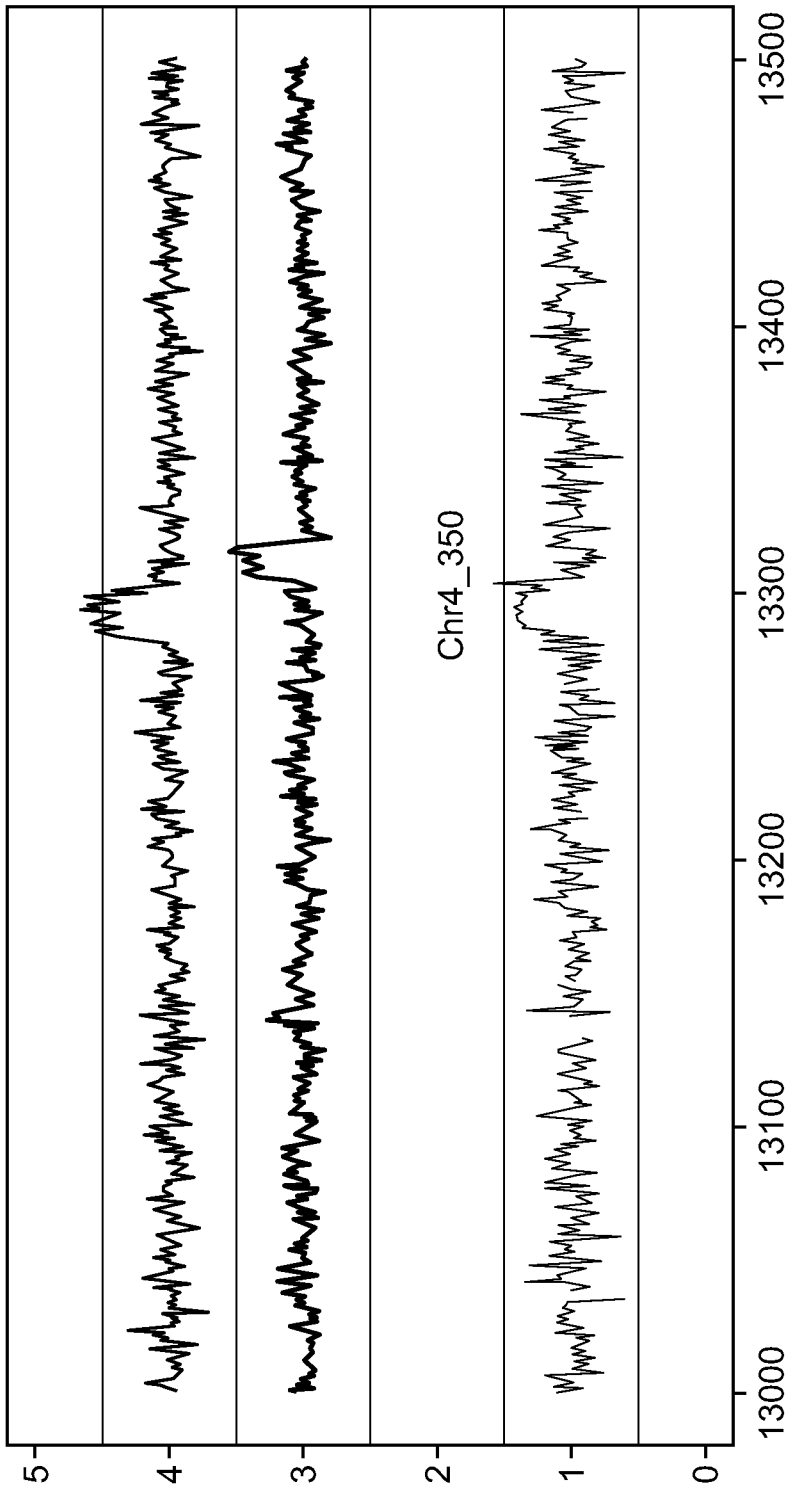

5
4
3
2
1
0
13000
13100
13200
13300
13400
13500
Chr4_350
Chr4 (50kbp bins)

FIG. 22

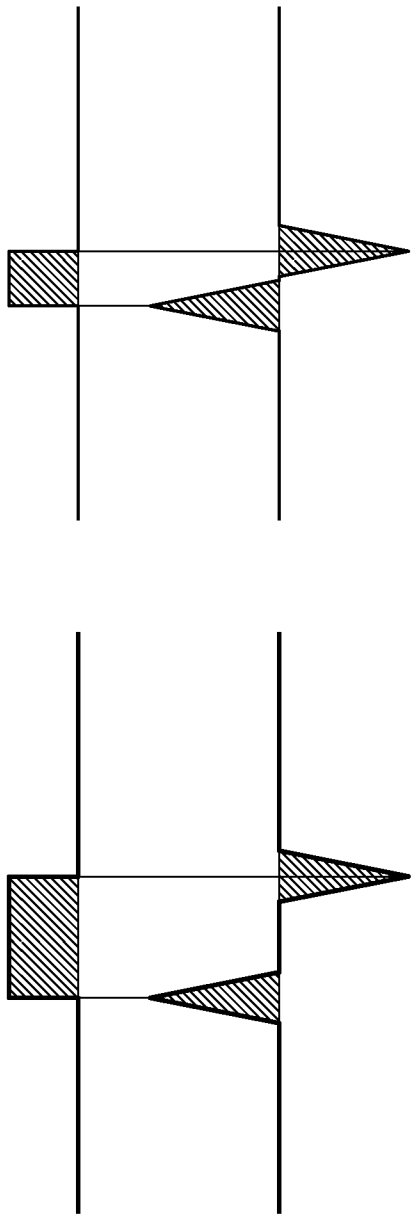


FIG. 23

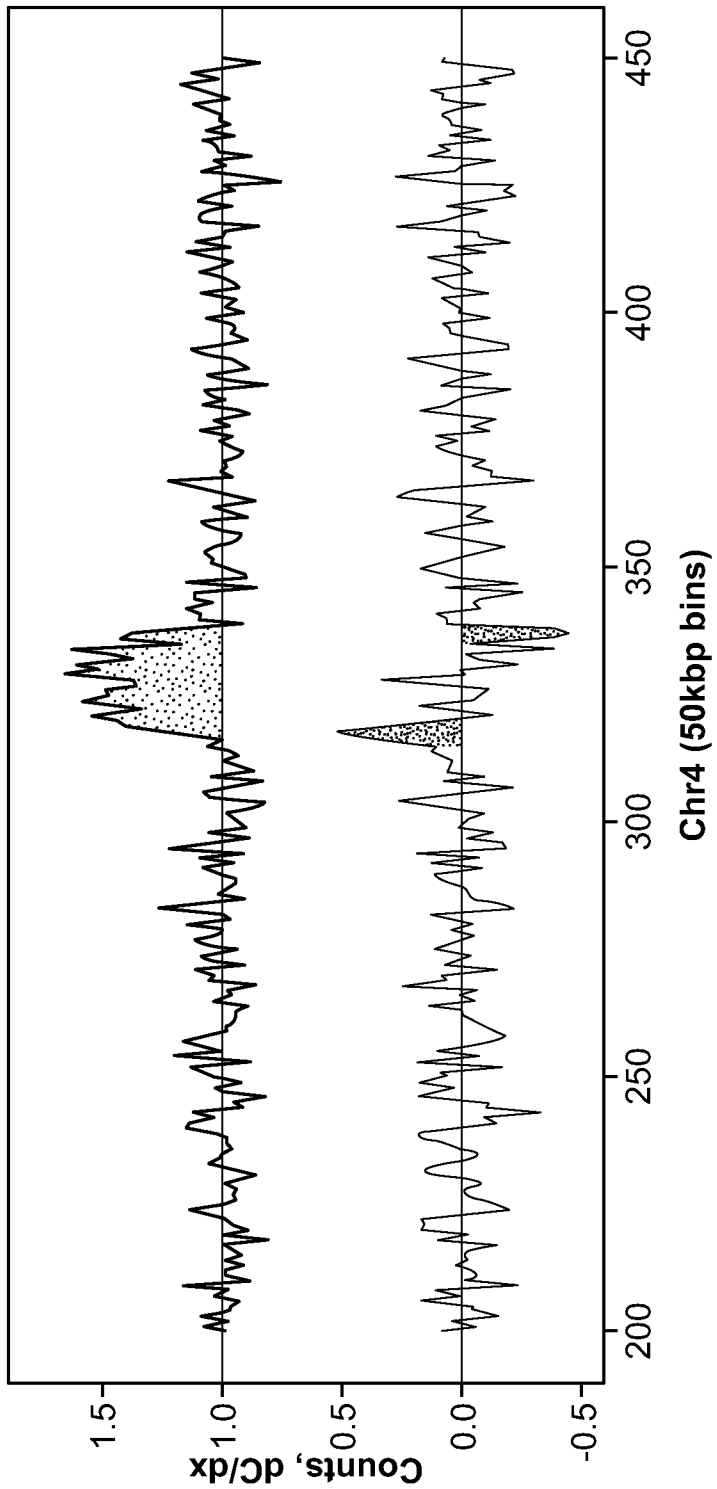

Counts, dC/dx
1.5
1.0
0.5
0.0
-0.5
200
250
300
350
400
450
Chr4 (50kbp bins)

FIG. 24

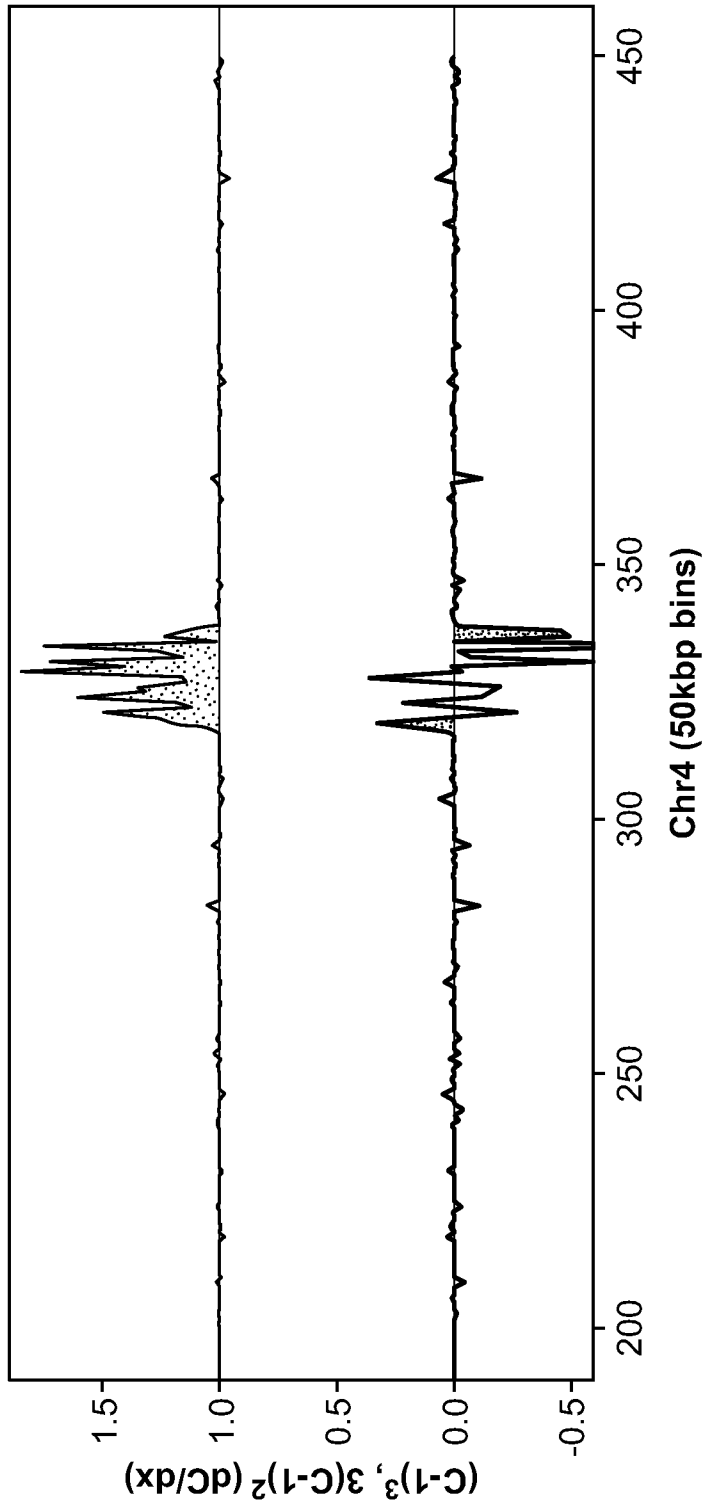

(C-1)3, 3(C-1)2 (dC/dx)
1.5
1.0
0.5
0.0
-0.5
200
250
300
350
400
450
Chr4 (50kbp bins)

FIG. 25

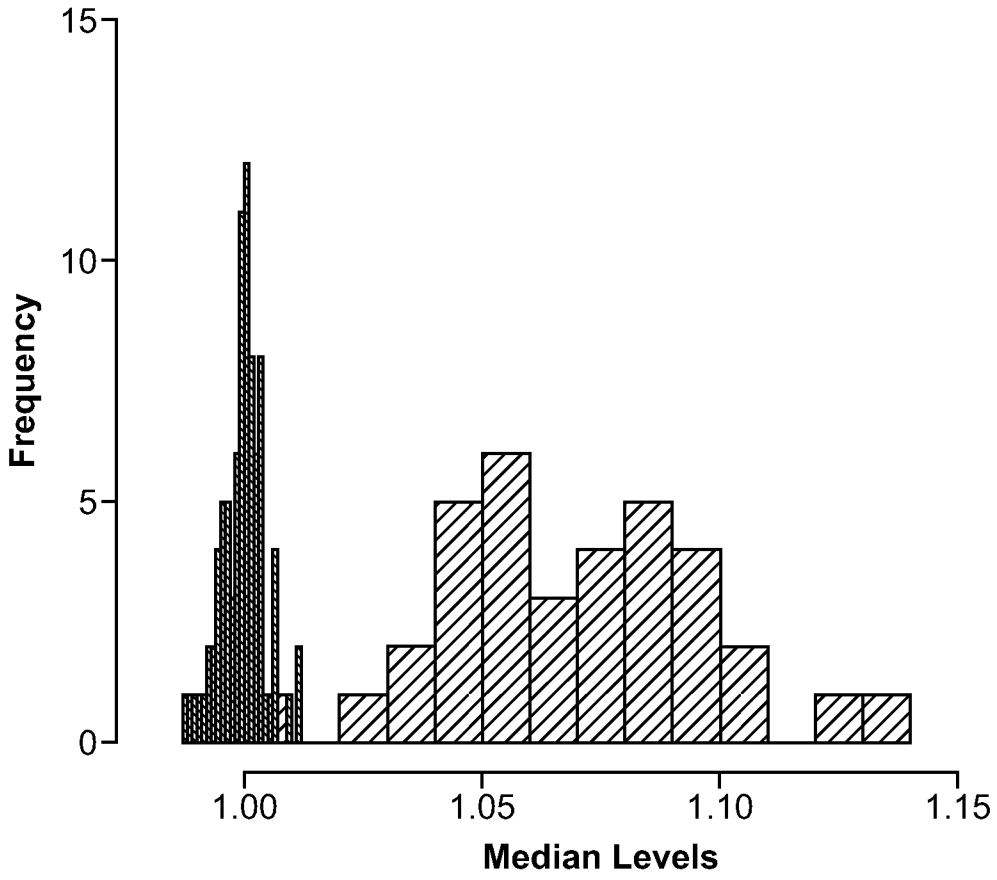


FIG. 26

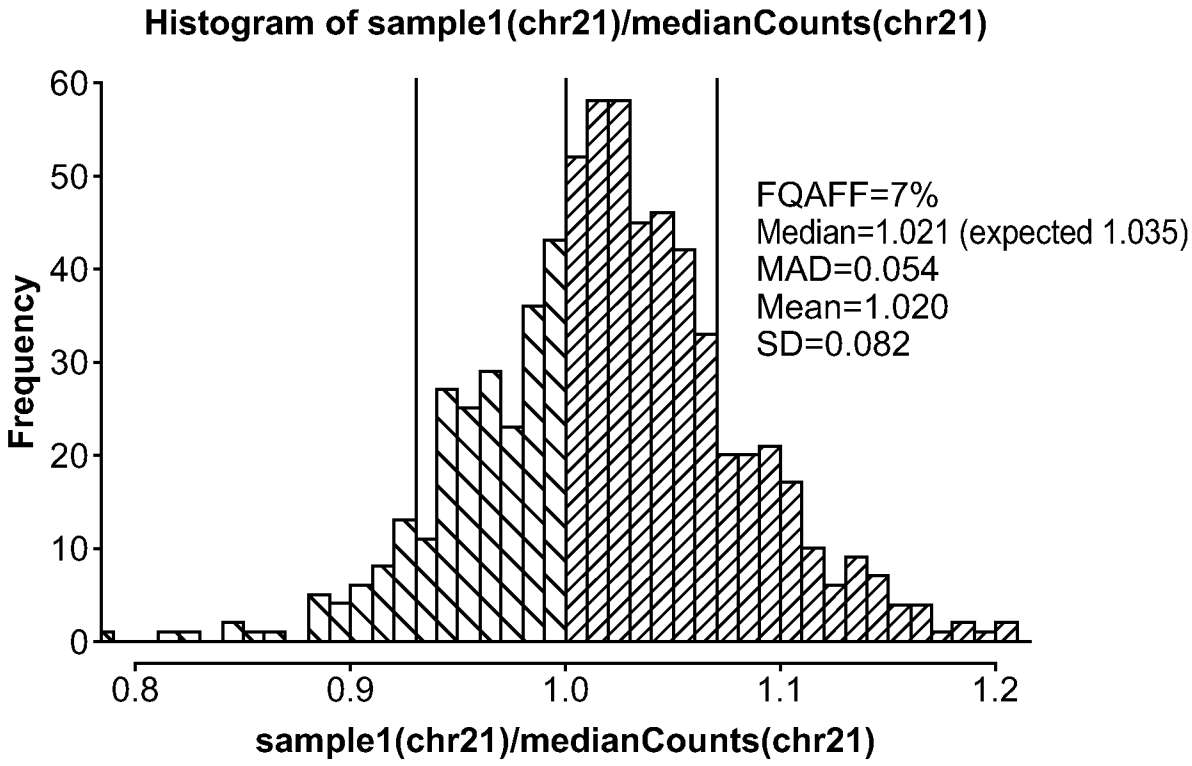

Histogram of sample1(chr21)/medianCounts(chr21)
Frequency
60
50
40
30
20
10
0
0.8
0.9
1.0
1.1
1.2
sample1(chr21)/medianCounts(chr21)
FQAFF=7%
Median=1.021 (expected 1.035)
MAD=0.054
Mean=1.020
SD=0.082

FIG. 27

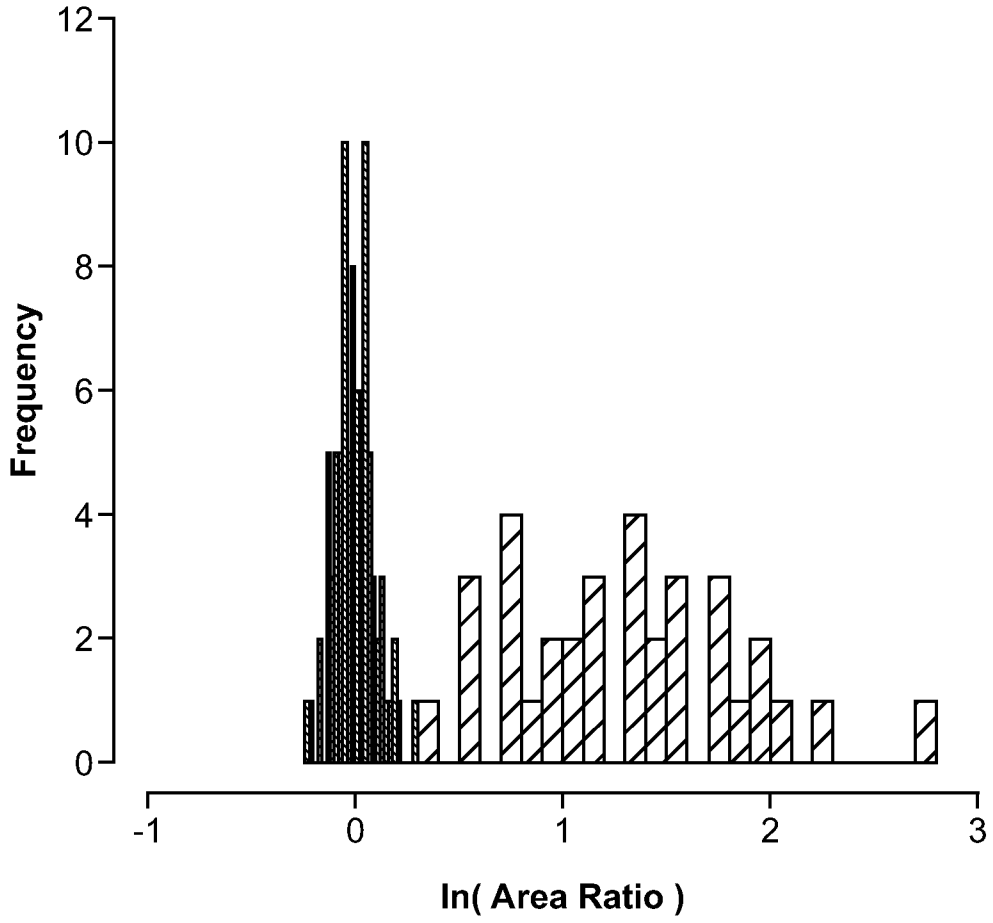


FIG. 28

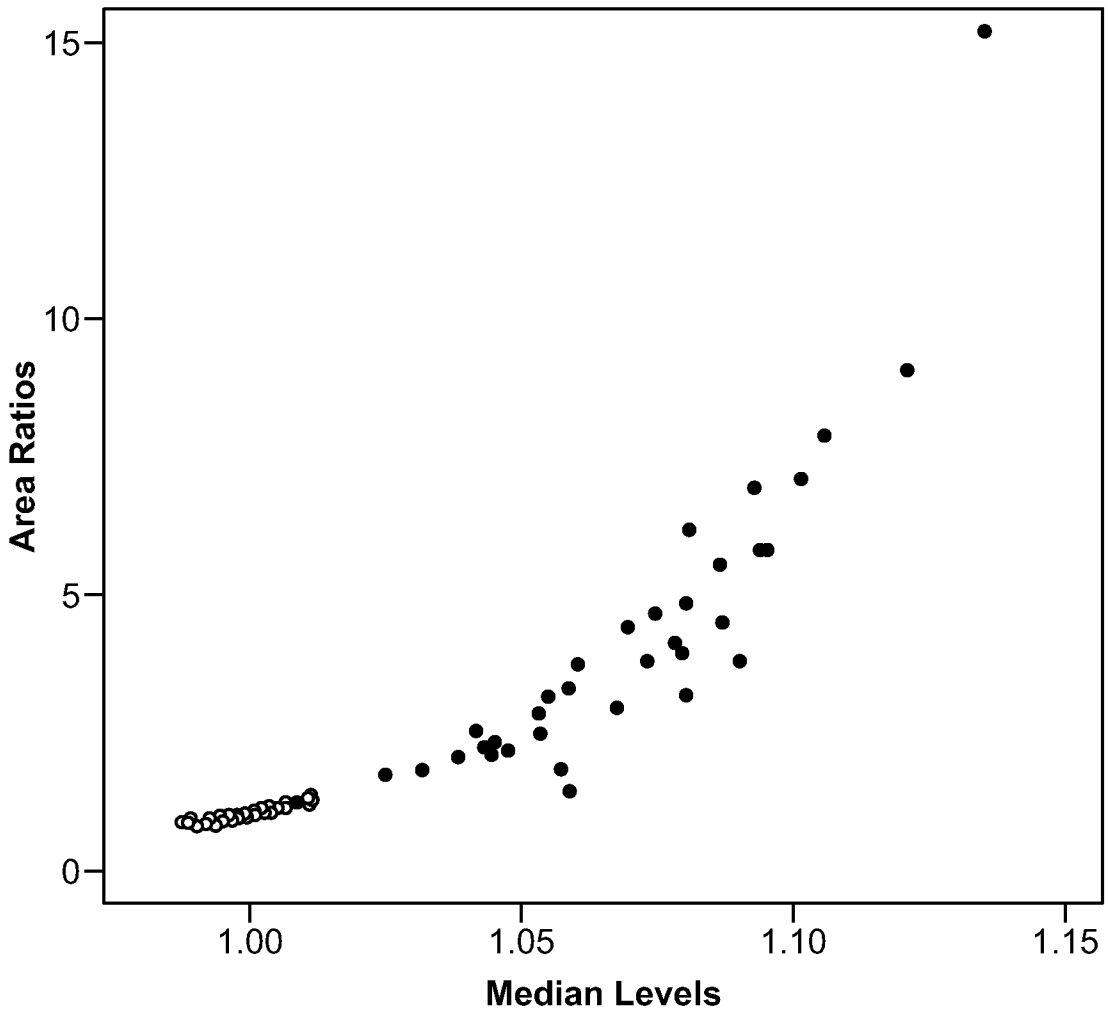


FIG. 29

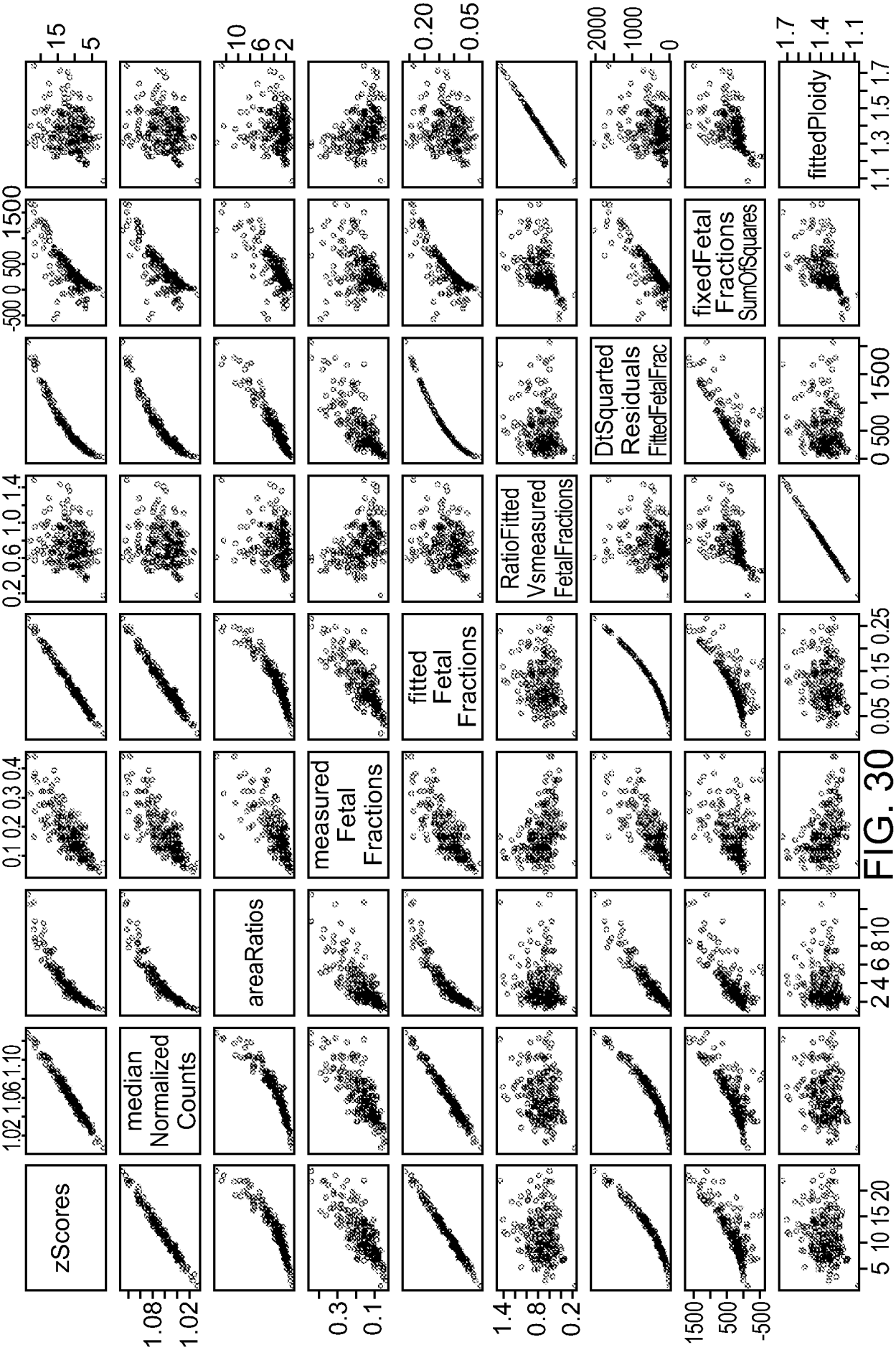

zScores
median Normalized Counts
areaRatios
measured Fetal Fractions
fitted Fetal Fractions
RatioFitted Vsmeasured FetalFractions
DtSquarted Residuals FittedFetalFrac
fixedFetal Fractions SumOfSquares
fittedPloidy

FIG. 30

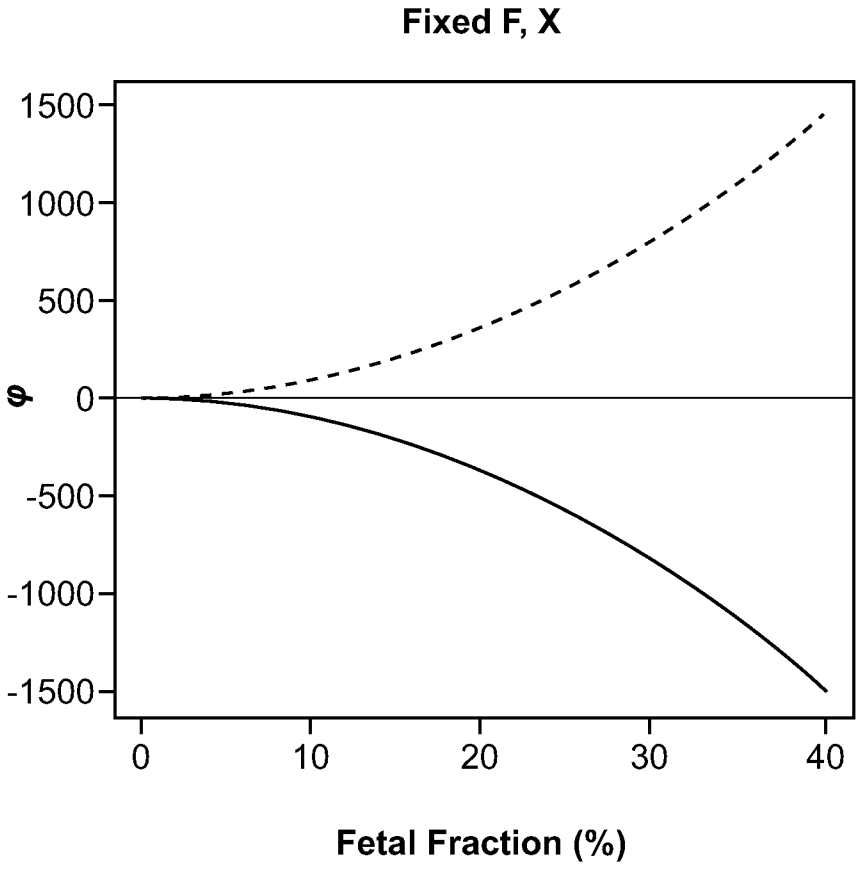

Fixed F, X
1500
1000
500
0
-500
-1000
-1500
0
10
20
30
40
Fetal Fraction (%)

FIG. 31

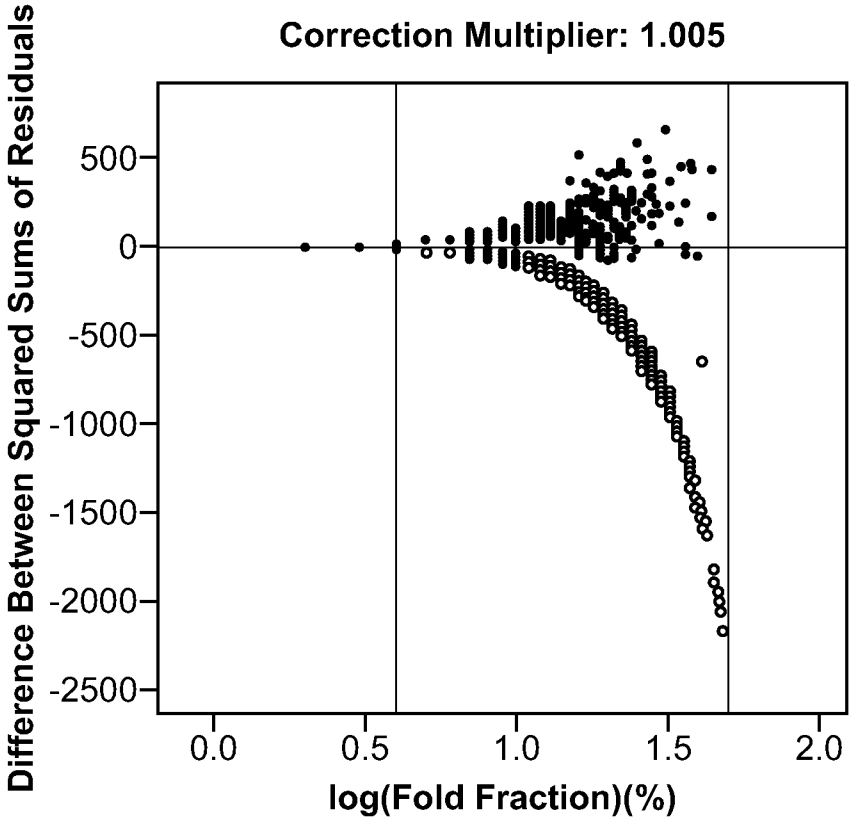

Correction Multiplier: 1.005
Difference Between Squared Sums of Residuals
500
0
-500
-1000
-1500
-2000
-2500
0.0
0.5
1.0
1.5
2.0
log(Fold Fraction)(%)

FIG. 32

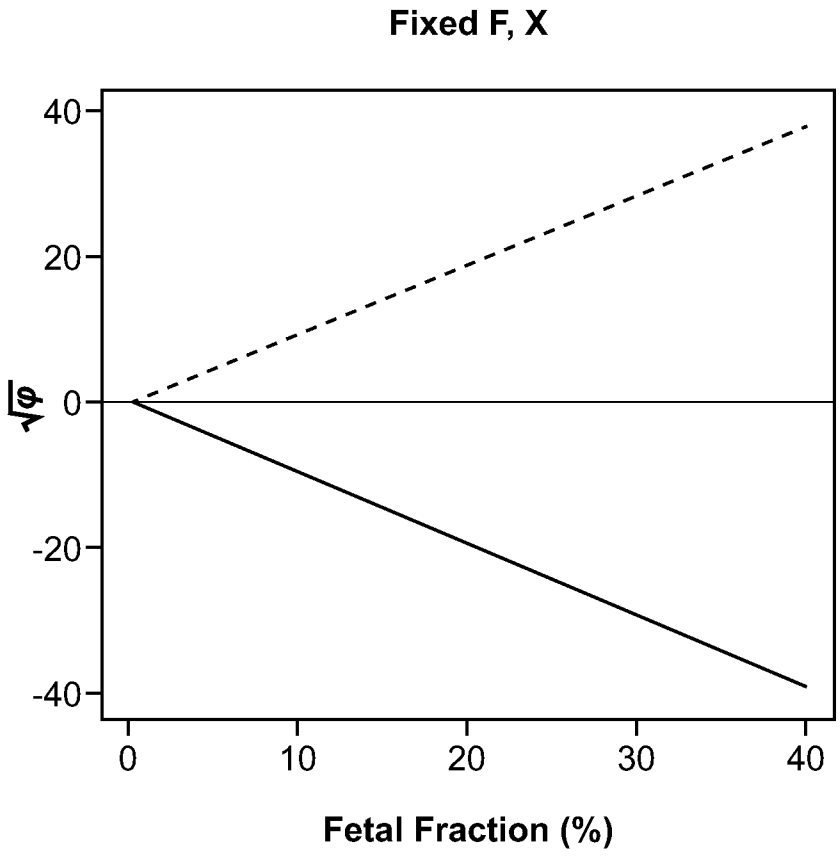


FIG. 33

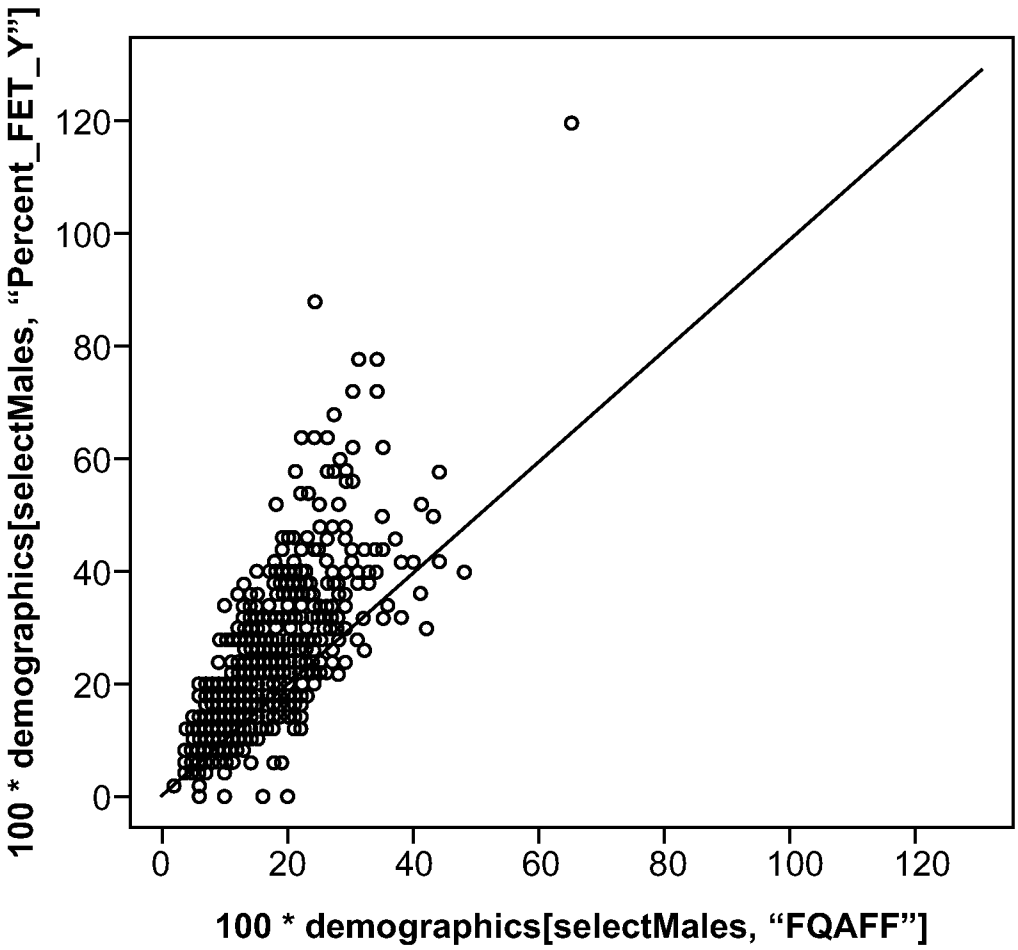

100 * demographics[selectMales, "Percent_FET_Y"]
120
100
80
60
40
20
0
0
20
40
60
80
100
120
100 * demographics[selectMales, "FQAFF"]

FIG. 34

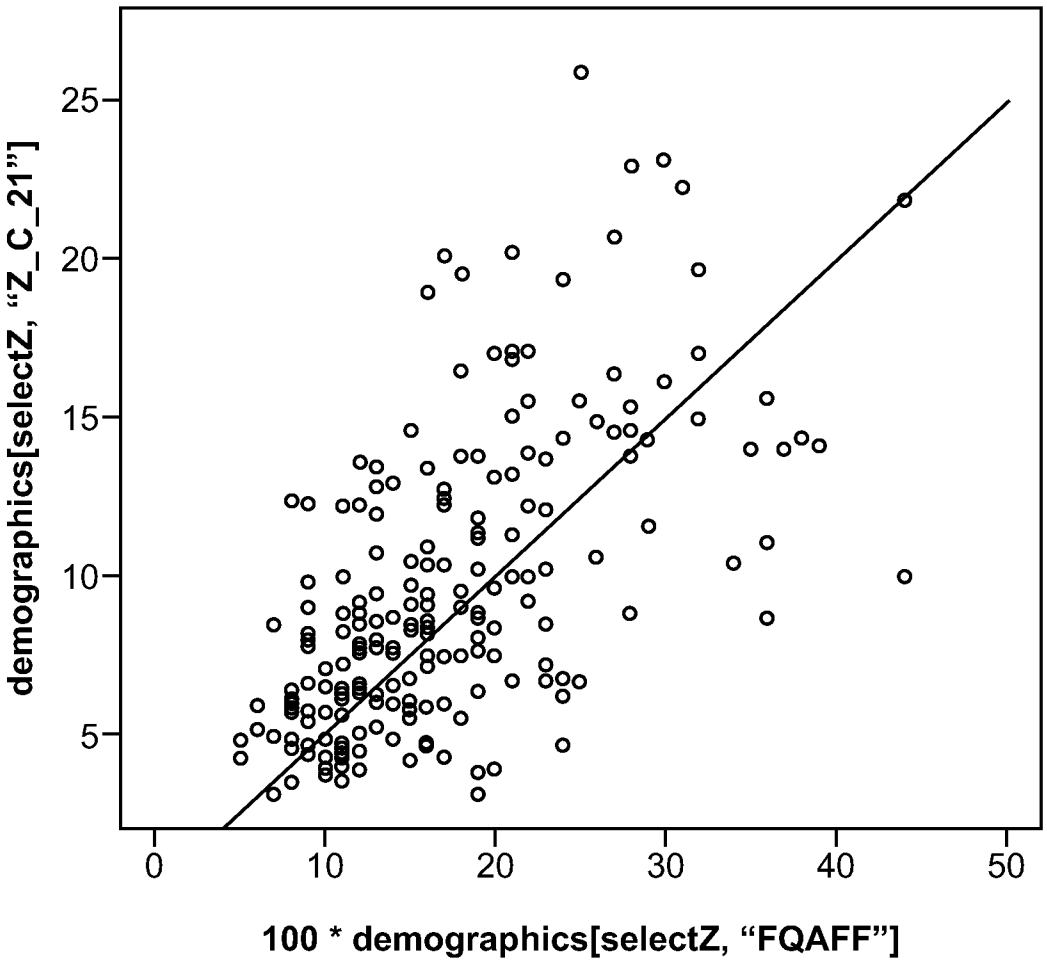

demographics[selectZ, "Z_C_21"]
25
20
15
10
5
0
10
20
30
40
50
100 * demographics[selectZ, "FQAFF"]

FIG. 35

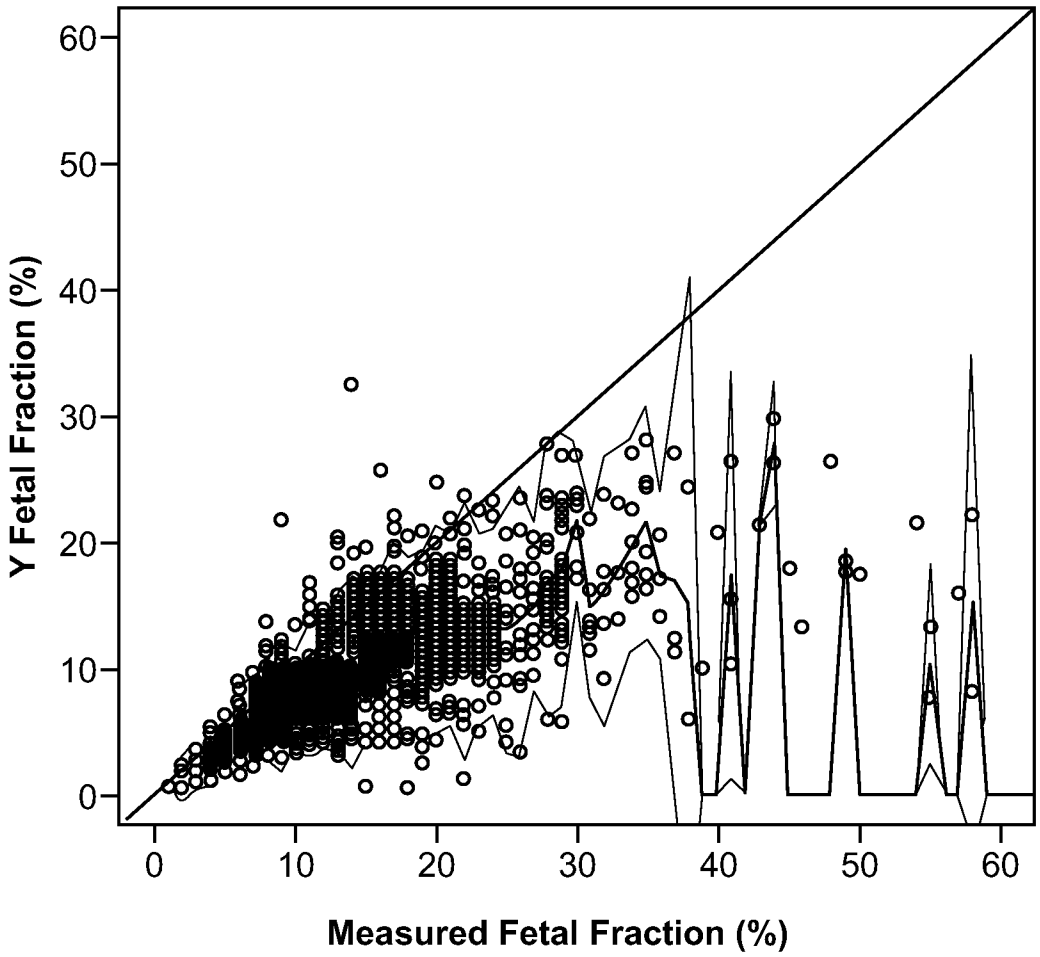
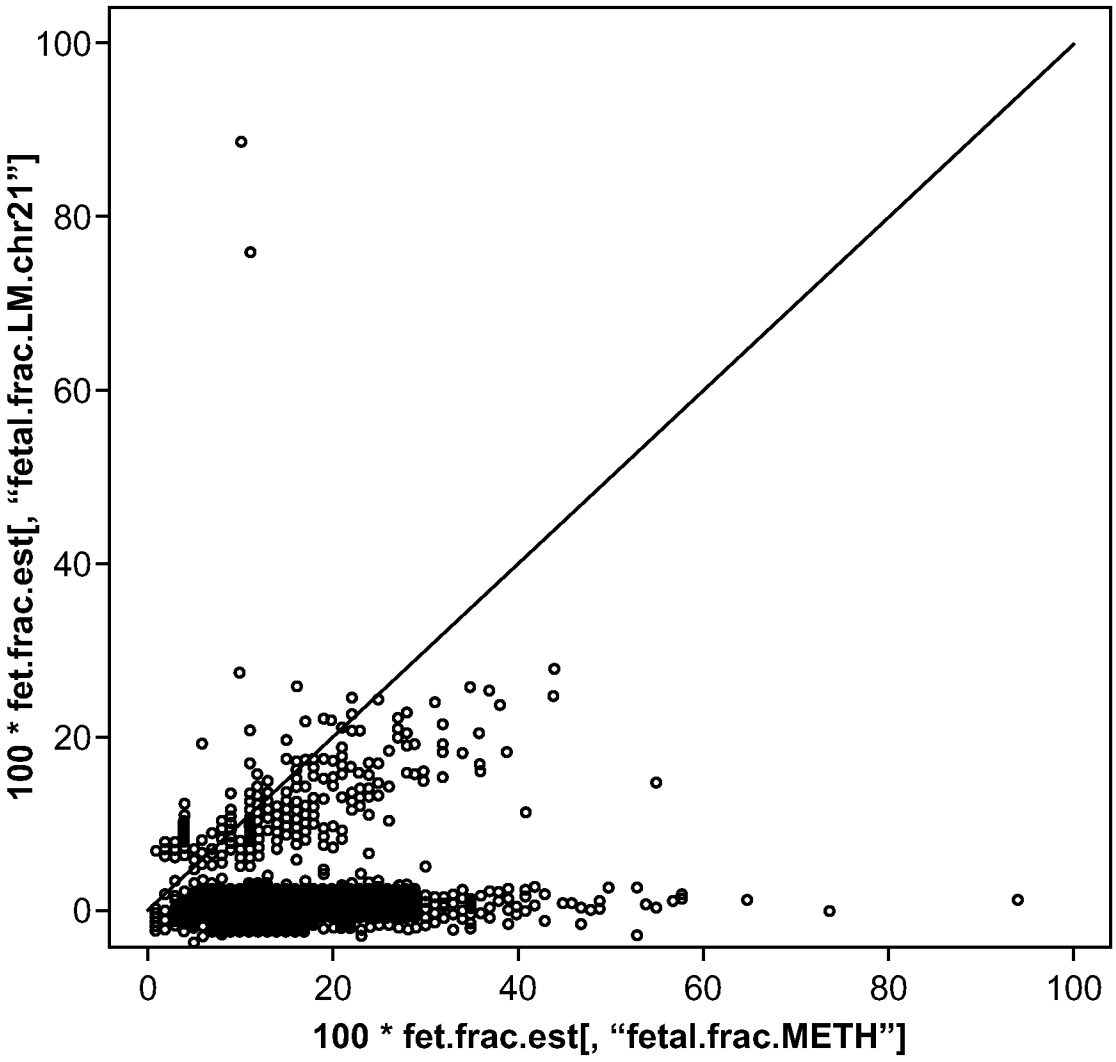


FIG. 36



Fetal fraction estimates based on
Chr 21 vs. measured fetal fractions.

FIG. 37

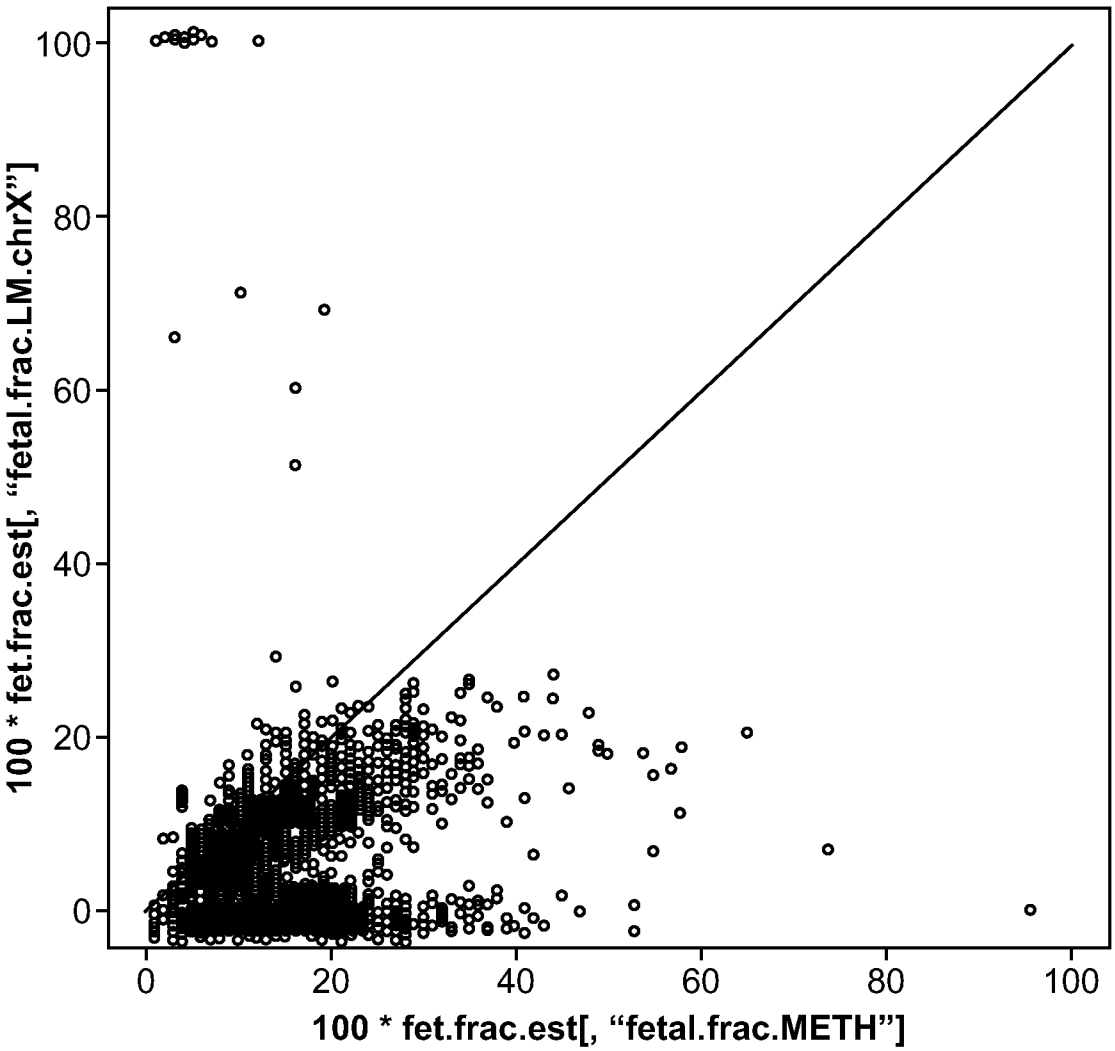

100
80
60
40
20
0
0
20
40
60
80
100
100 * fet.frac.est[, "fetal.frac.LM.chrX"]
100 * fet.frac.est[, "fetal.frac.METH"]

FIG. 38

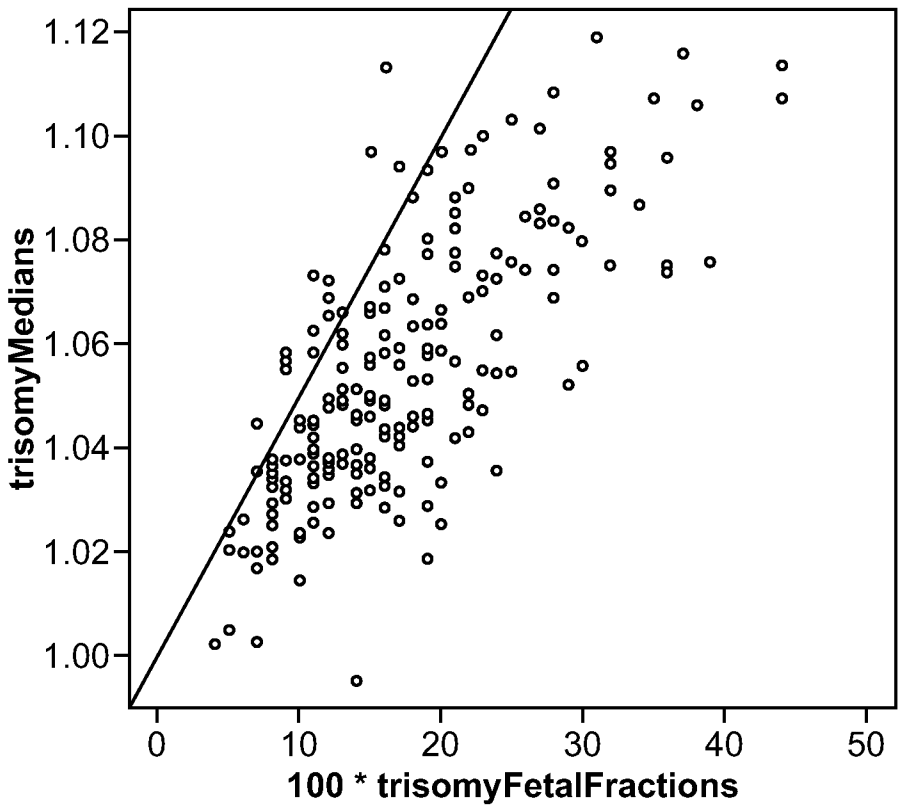

1.12
1.10
1.08
1.06
1.04
1.02
1.00
trisomyMedians
0
10
20
30
40
50
100 * trisomyFetalFractions

FIG. 39

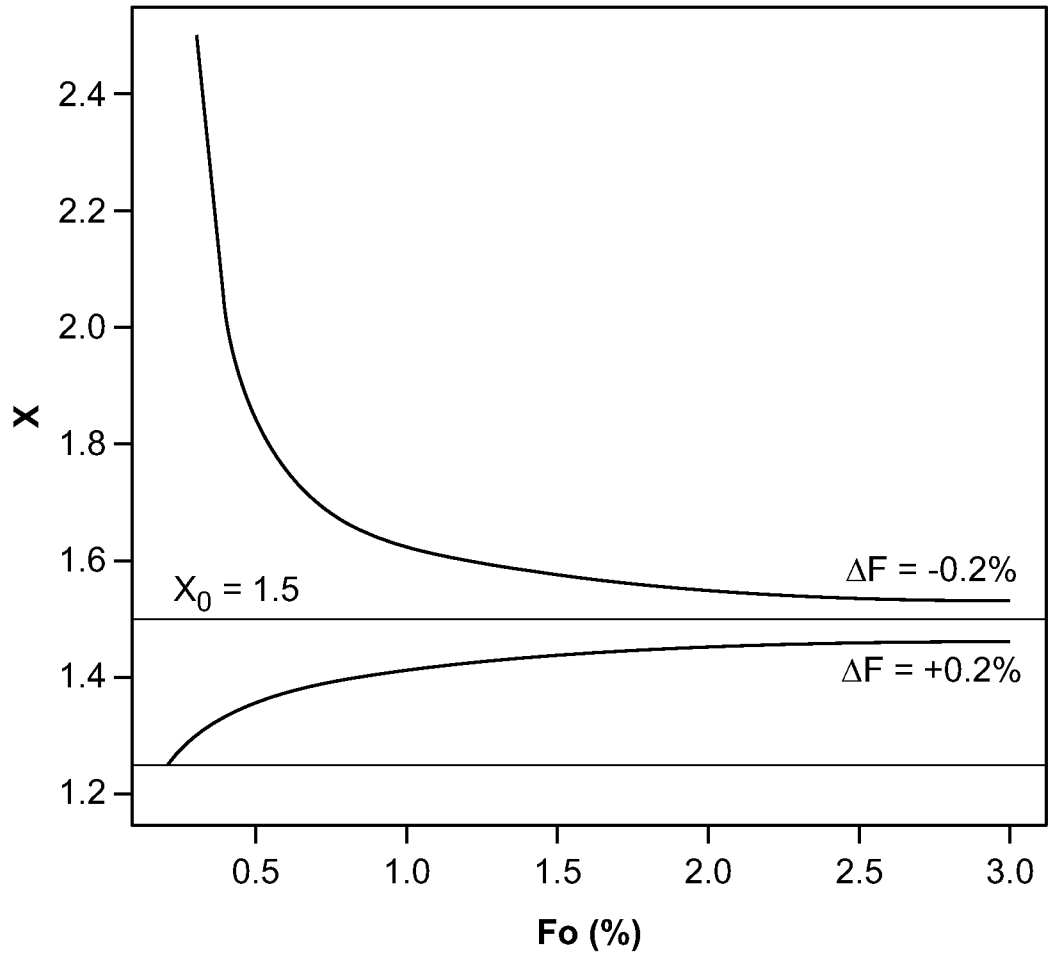


FIG. 40

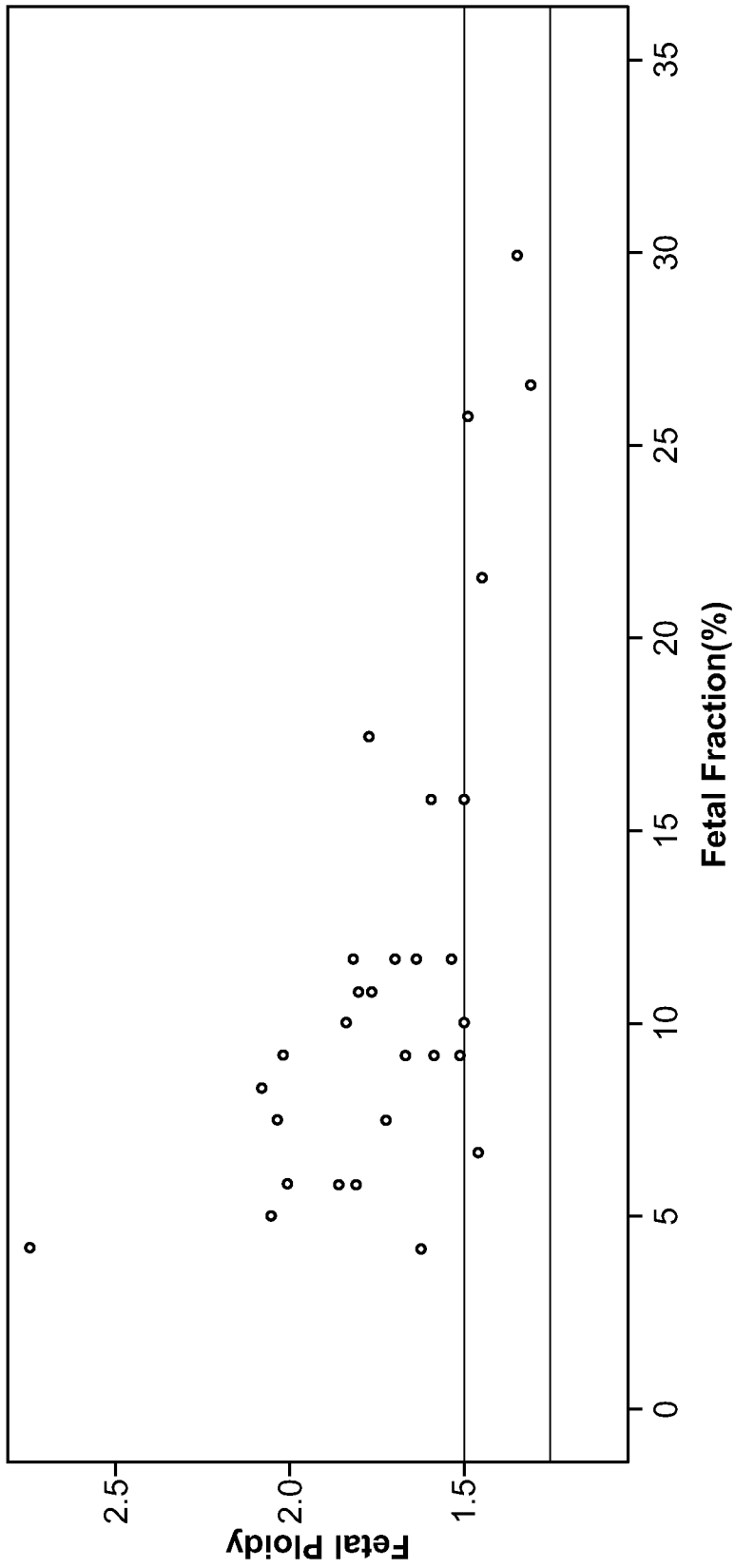


FIG. 41

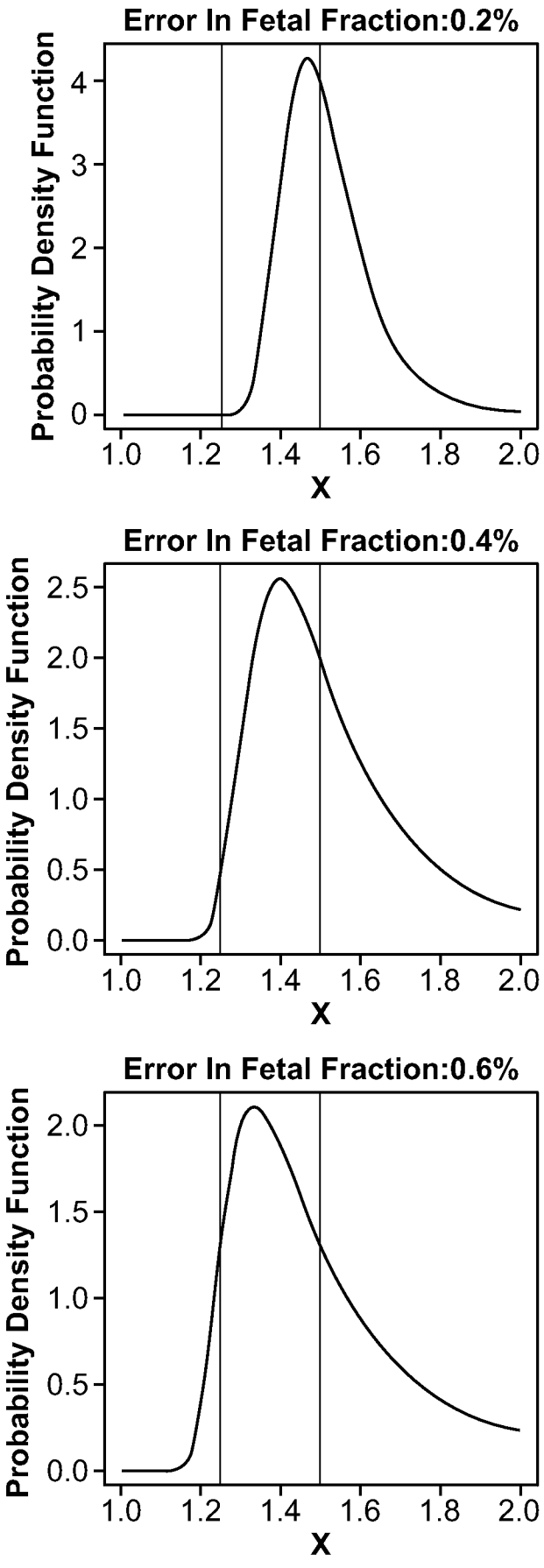

Error In Fetal Fraction:0.2%
Probability Density Function
0
1
2
3
4
1.0
1.2
1.4
1.6
1.8
2.0
X
Error In Fetal Fraction:0.4%
Probability Density Function
0.0
0.5
1.0
1.5
2.0
2.5
1.0
1.2
1.4
1.6
1.8
2.0
X
Error In Fetal Fraction:0.6%
Probability Density Function
0.0
0.5
1.0
1.5
2.0
1.0
1.2
1.4
1.6
1.8
2.0
X

FIG. 42

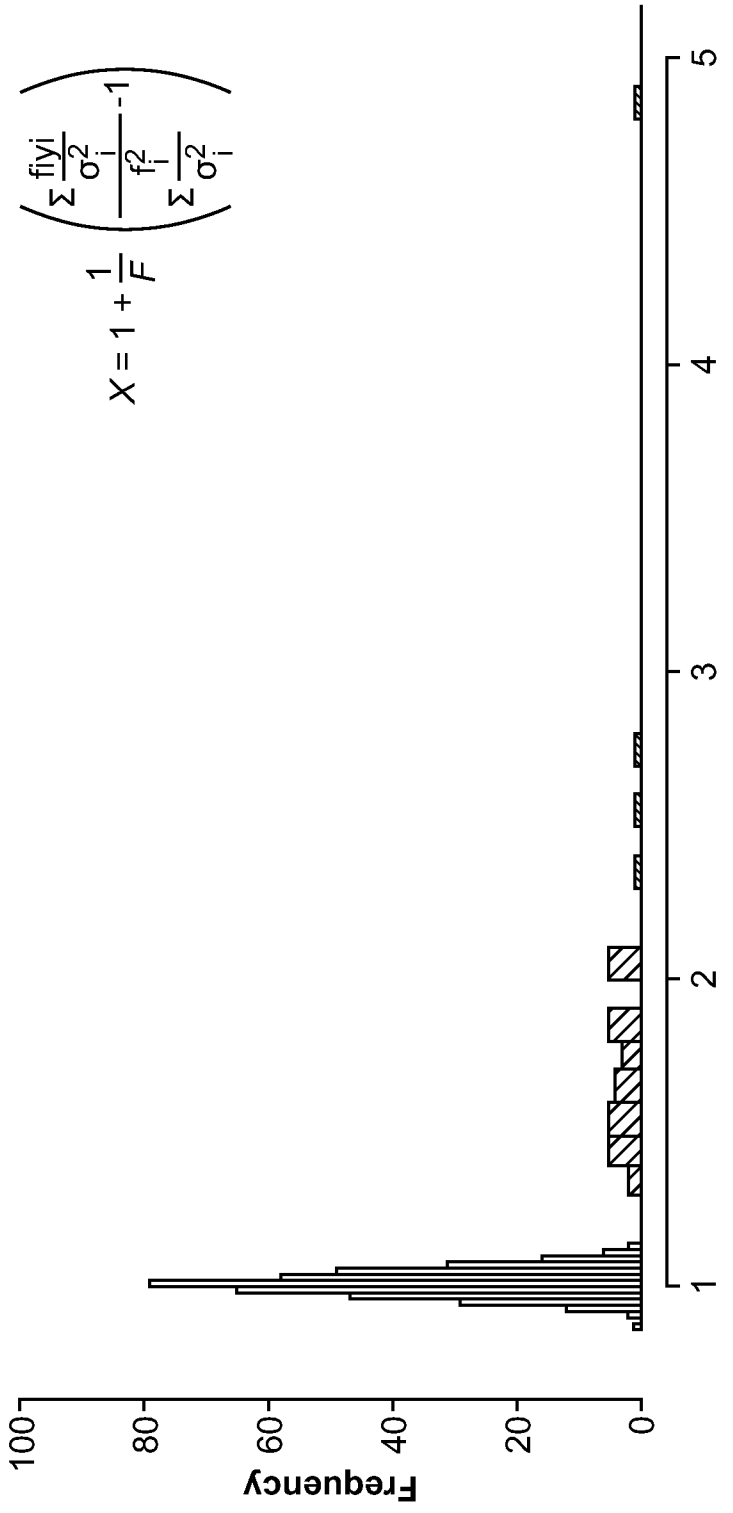

480v2
$X = 1 + \frac{1}{F}\left(\frac{\sum \frac{f_i y_i}{\sigma_i^2}}{\frac{f_i^2}{\sum \frac{1}{\sigma_i^2}}} - 1\right)$
Frequency
0
20
40
60
80
100
1
2
3
4
5
Ploidy

FIG. 43

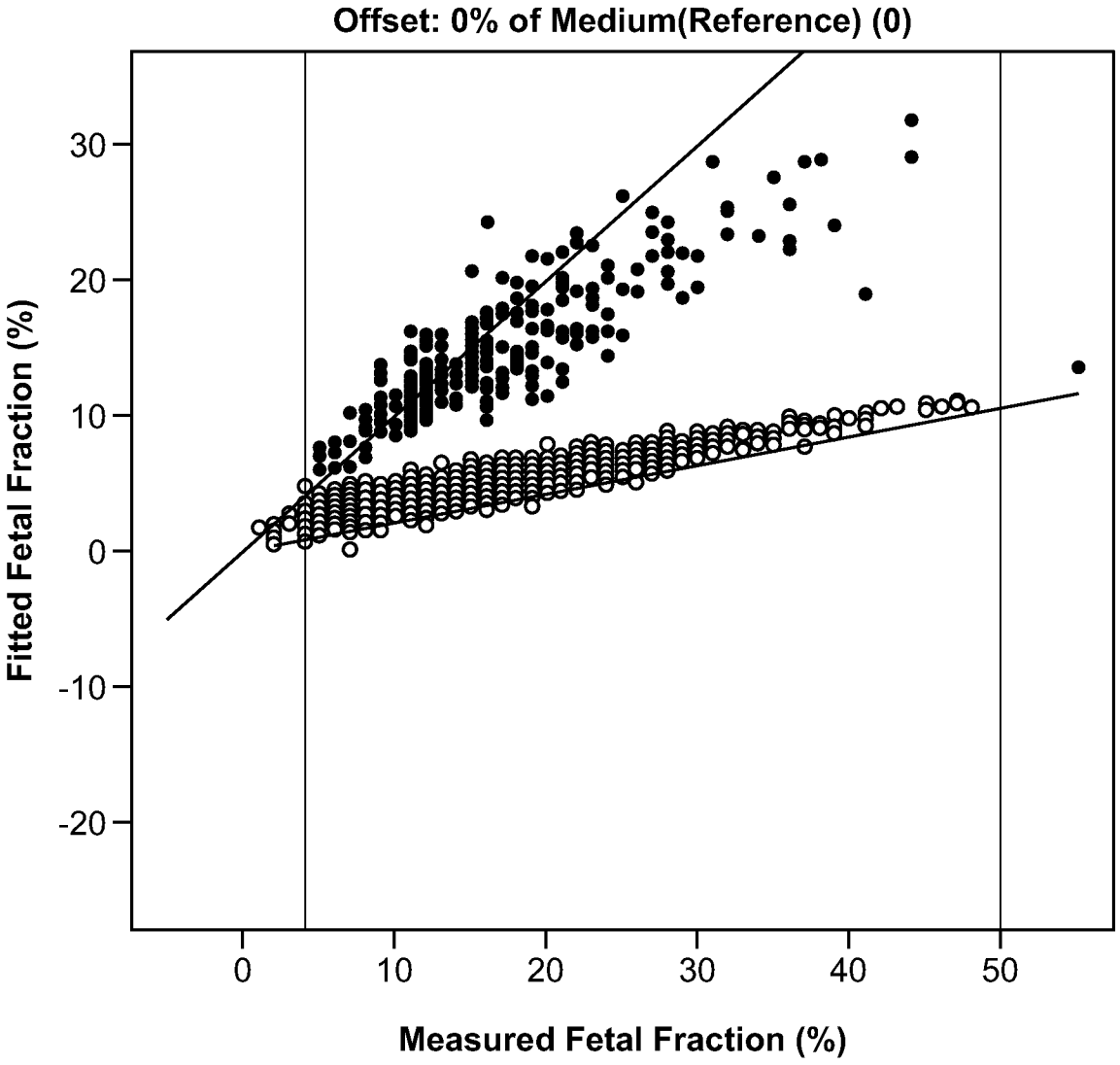

Offset: 0% of Medium(Reference) (0)
Fitted Fetal Fraction (%)
30
20
10
0
-10
-20
0
10
20
30
40
50
Measured Fetal Fraction (%)

FIG. 44

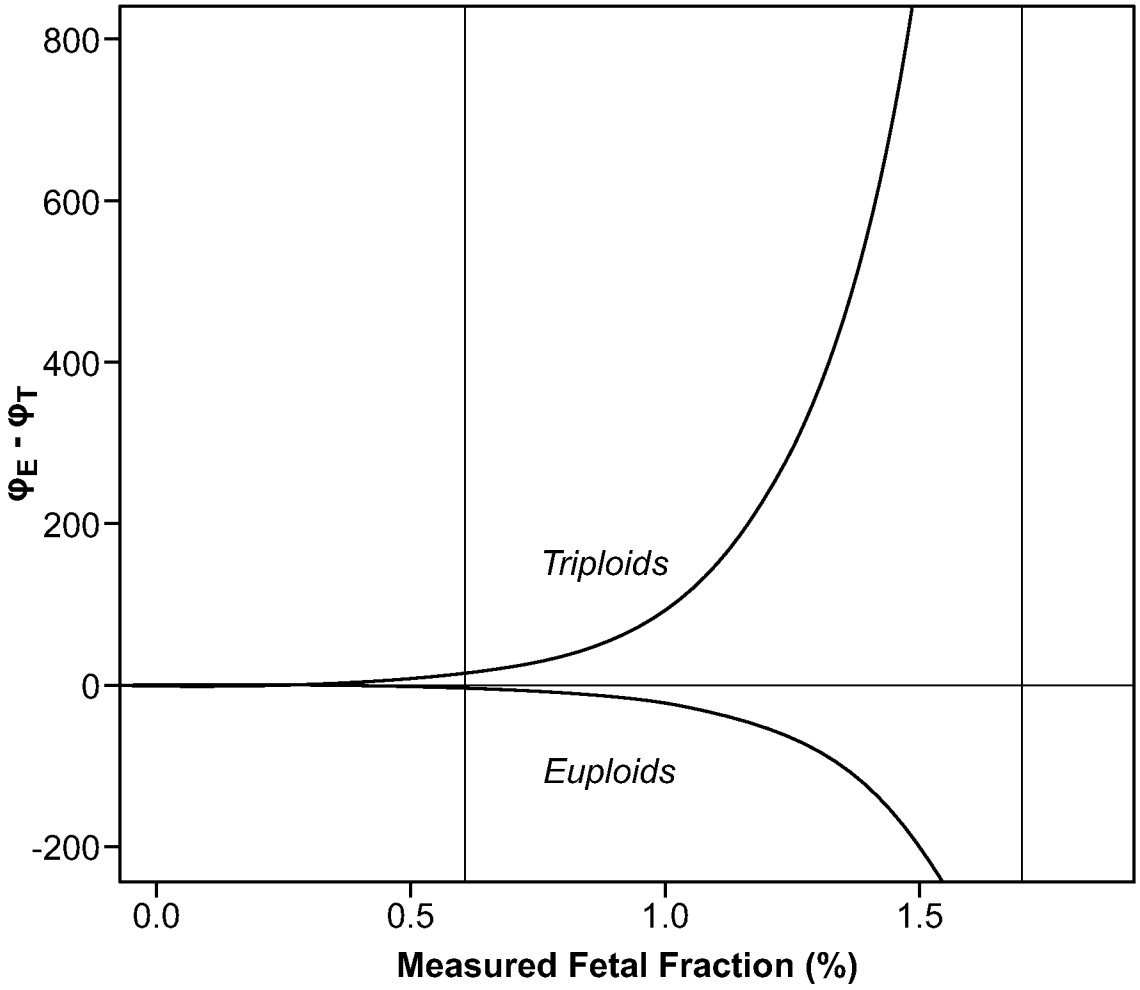

800
600
400
200
0
-200
$\varphi_E - \varphi_T$
Triploids
Euploids
0.0
0.5
1.0
1.5
Measured Fetal Fraction (%)

FIG. 45

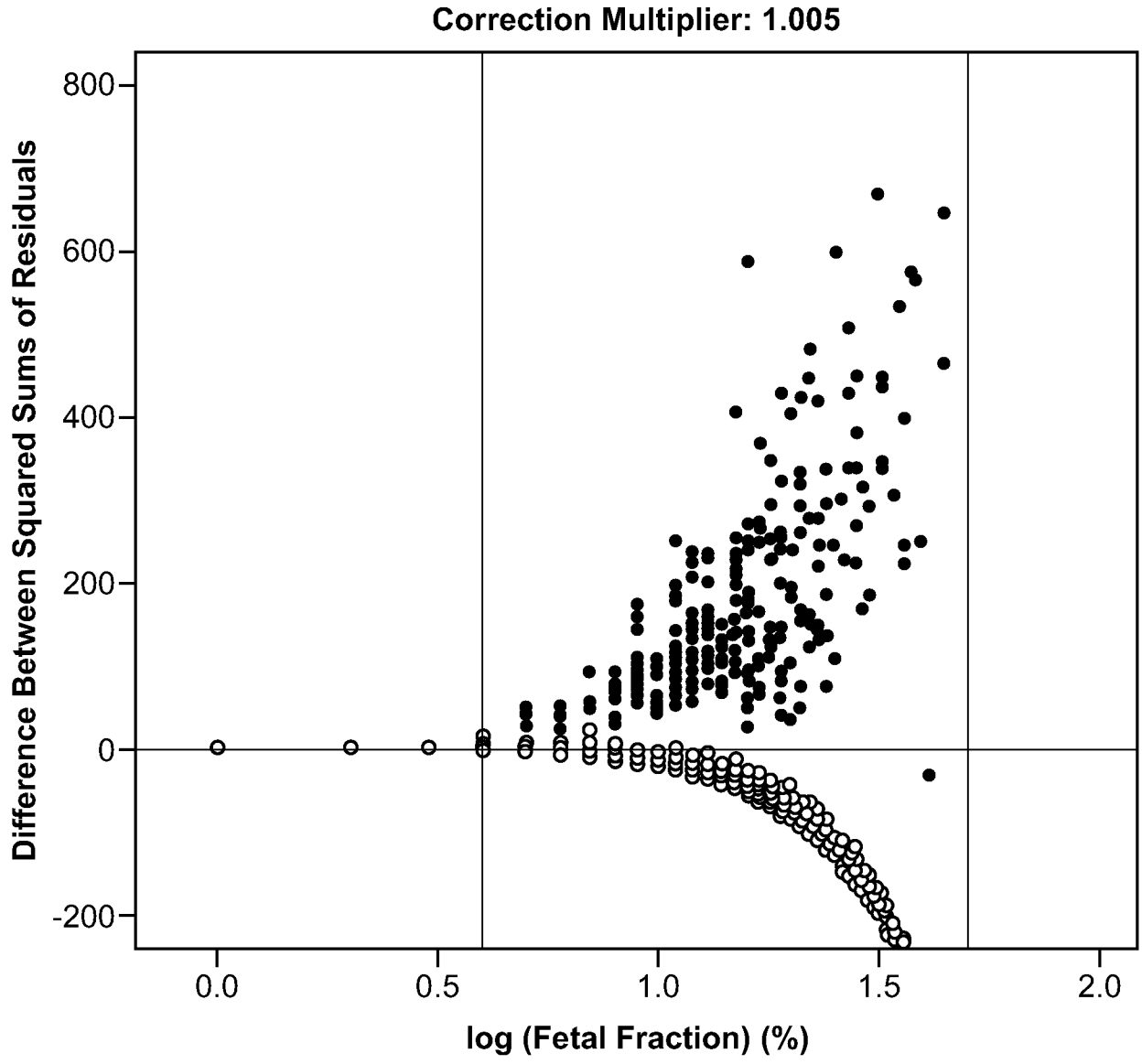

Correction Multiplier: 1.005
Difference Between Squared Sums of Residuals
800
600
400
200
0
-200
0.0
0.5
1.0
1.5
2.0
log (Fetal Fraction) (%)

FIG. 46

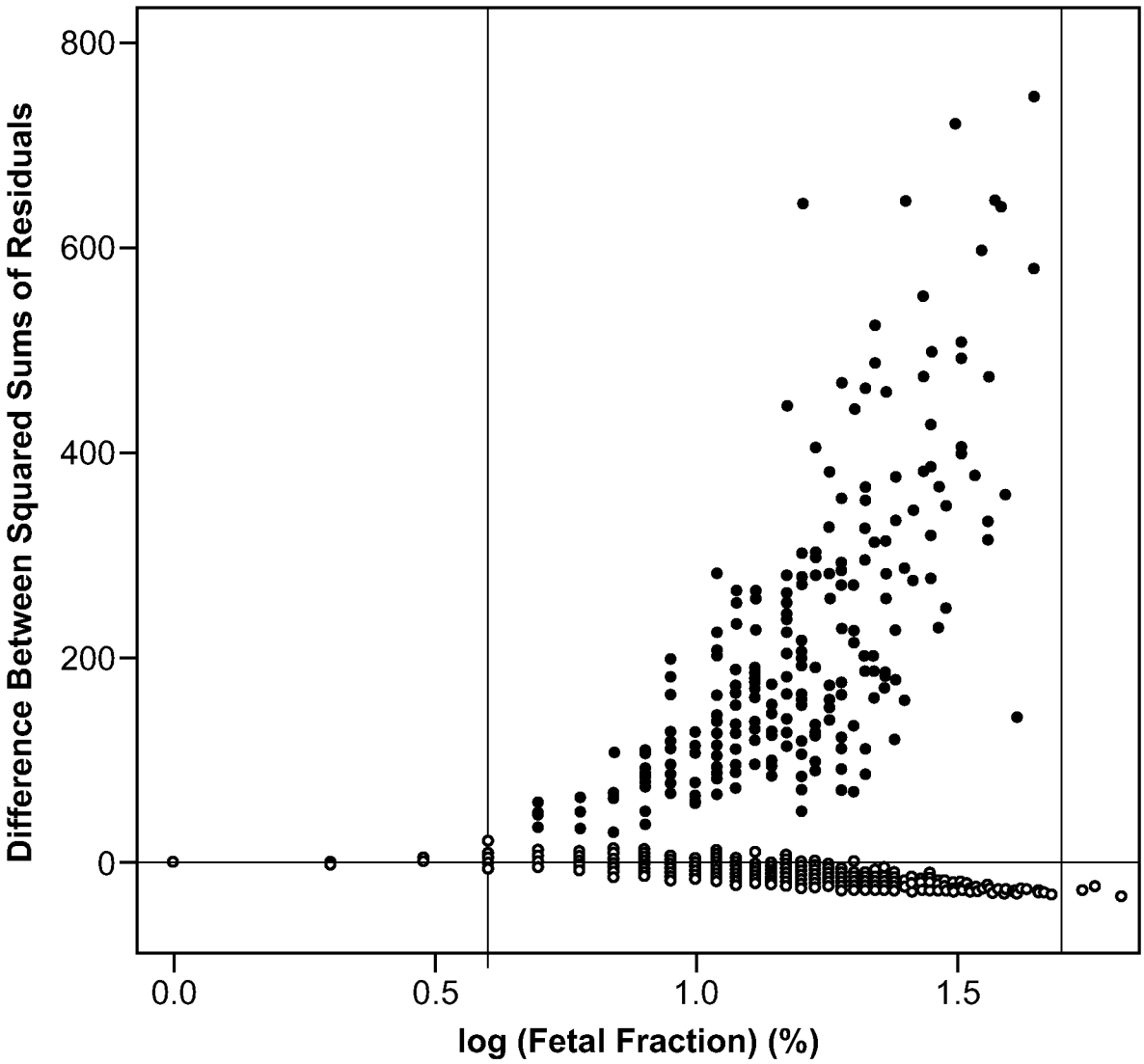

800
600
400
200
0
Difference Between Squared Sums of Residuals
0.0
0.5
1.0
1.5
log (Fetal Fraction) (%)

FIG. 47

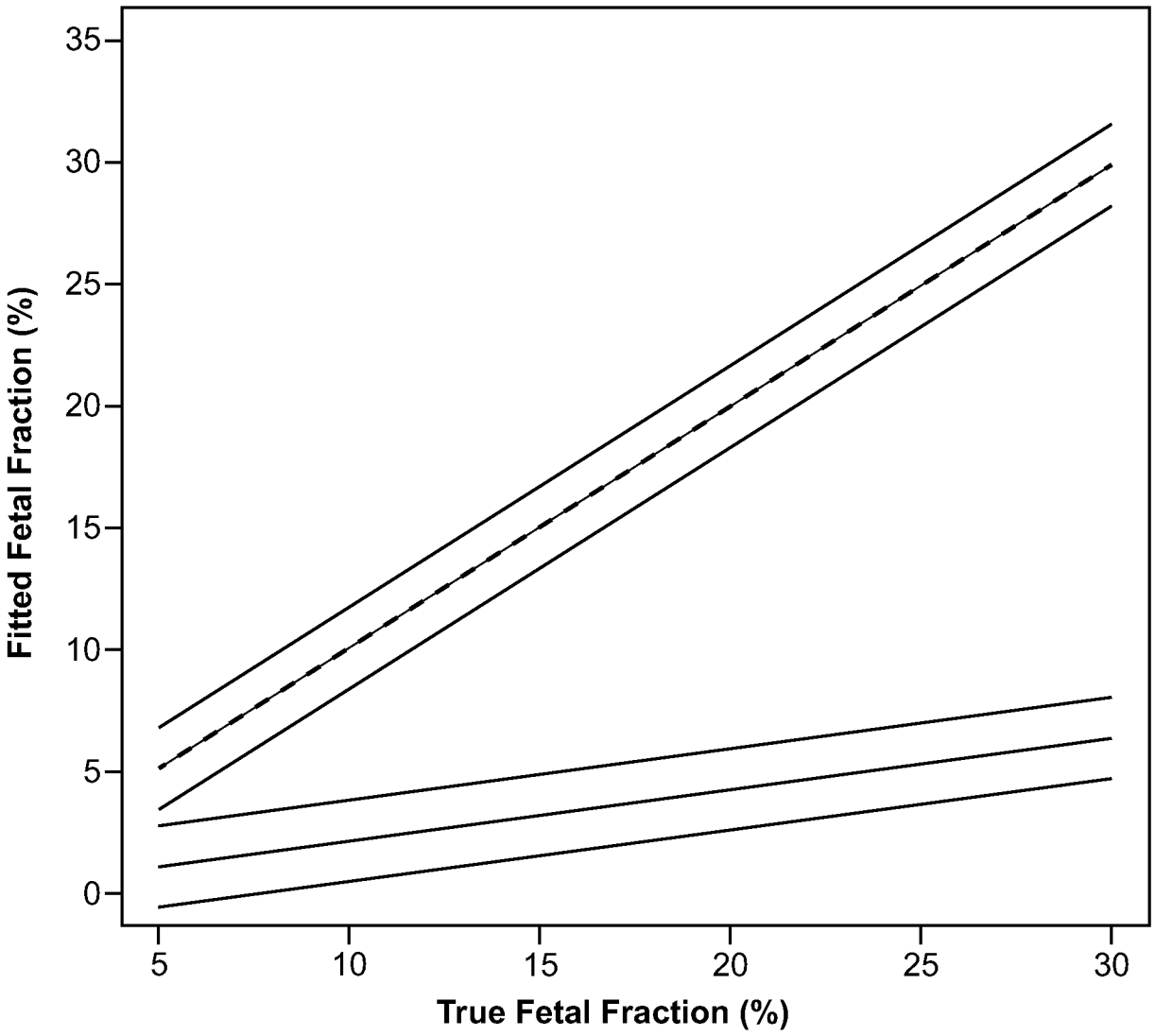

Fitted Fetal Fraction (%)
35
30
25
20
15
10
5
0
5
10
15
20
25
30
True Fetal Fraction (%)

FIG. 48

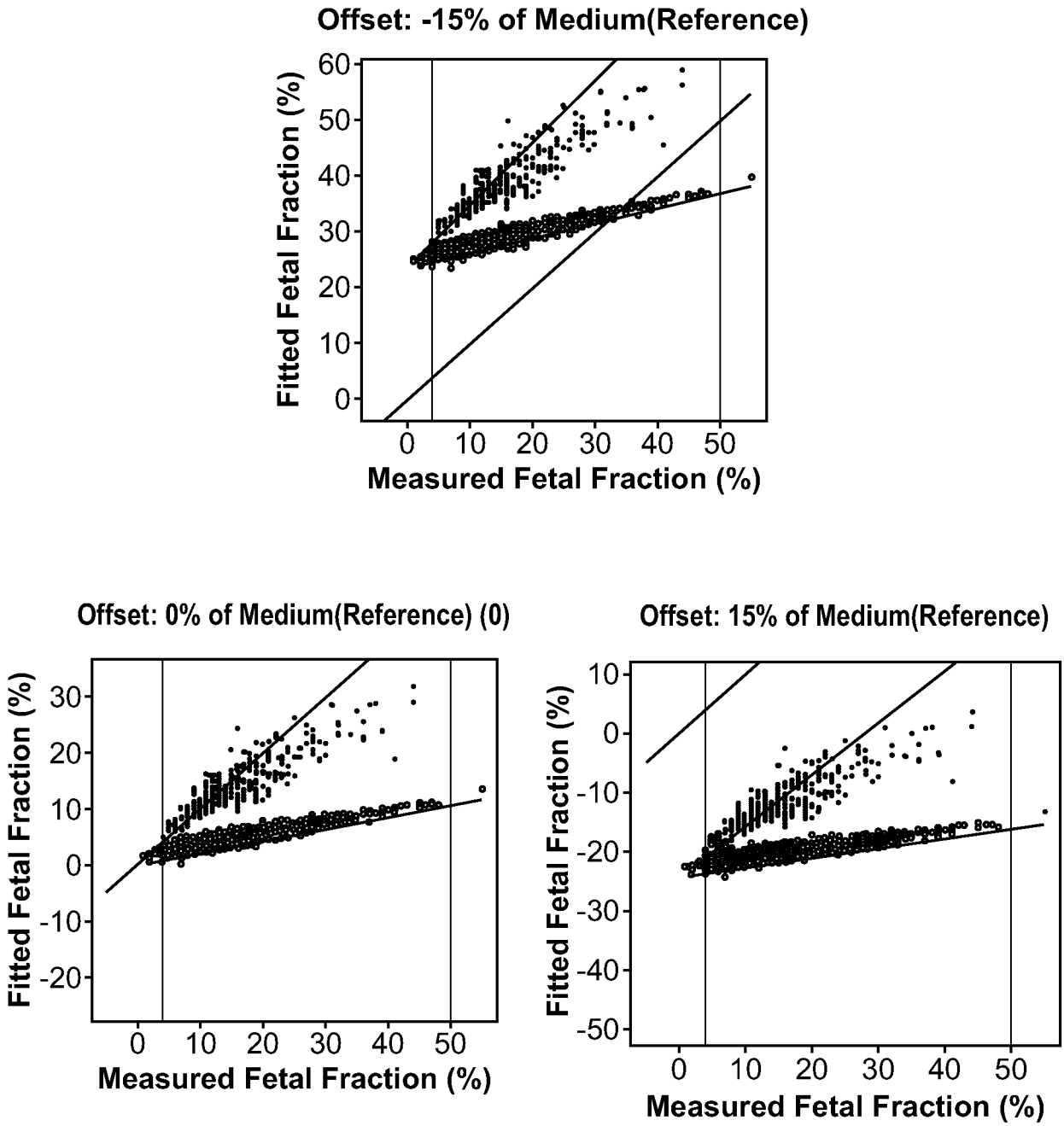

Offset: -15% of Medium(Reference)
Fitted Fetal Fraction (%)
Measured Fetal Fraction (%)
Offset: 0% of Medium(Reference) (0)
Fitted Fetal Fraction (%)
Measured Fetal Fraction (%)
Offset: 15% of Medium(Reference)
Fitted Fetal Fraction (%)
Measured Fetal Fraction (%)

FIG. 49

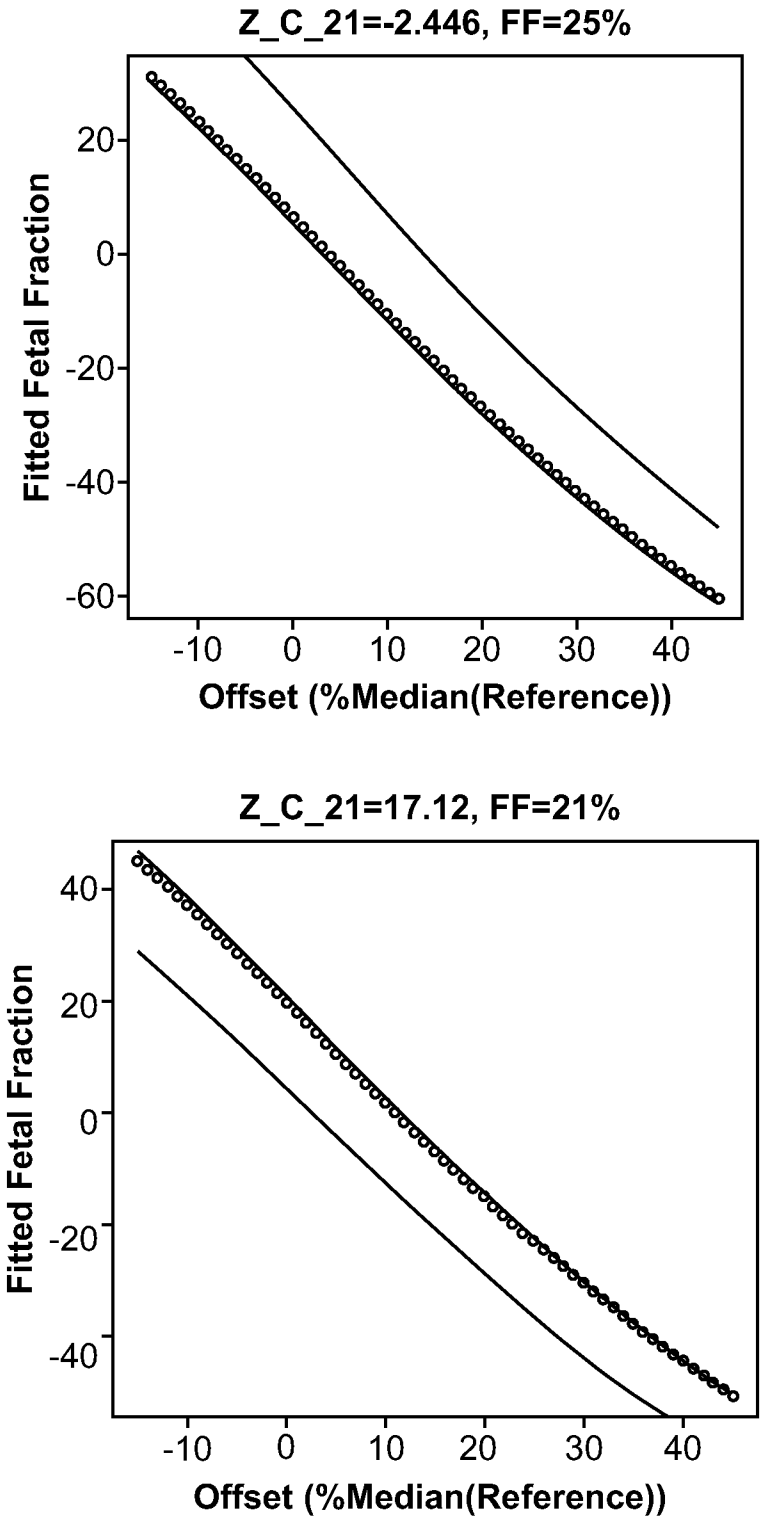

Z_C_21=-2.446, FF=25%
Fitted Fetal Fraction
20
0
-20
-40
-60
-10
0
10
20
30
40
Offset (%Median(Reference))
Z_C_21=17.12, FF=21%
Fitted Fetal Fraction
40
20
0
-20
-40
-10
0
10
20
30
40
Offset (%Median(Reference))

FIG. 50

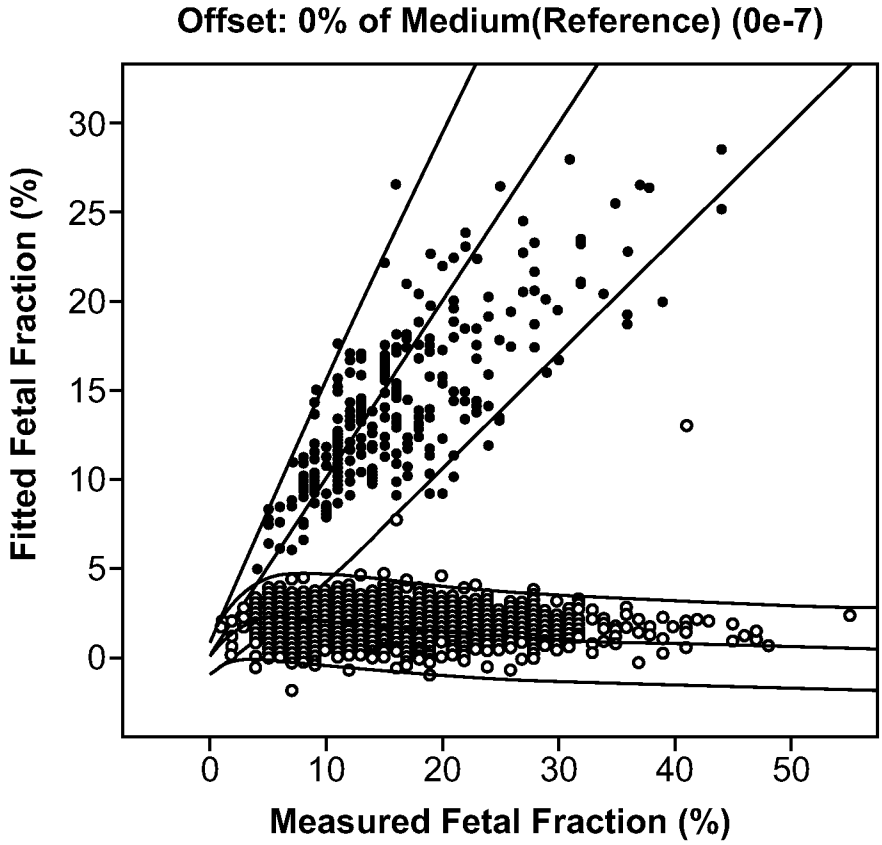

Offset: 0% of Medium(Reference) (0e-7)
Fitted Fetal Fraction (%)
30
25
20
15
10
5
0
0
10
20
30
40
50
Measured Fetal Fraction (%)

FIG. 51

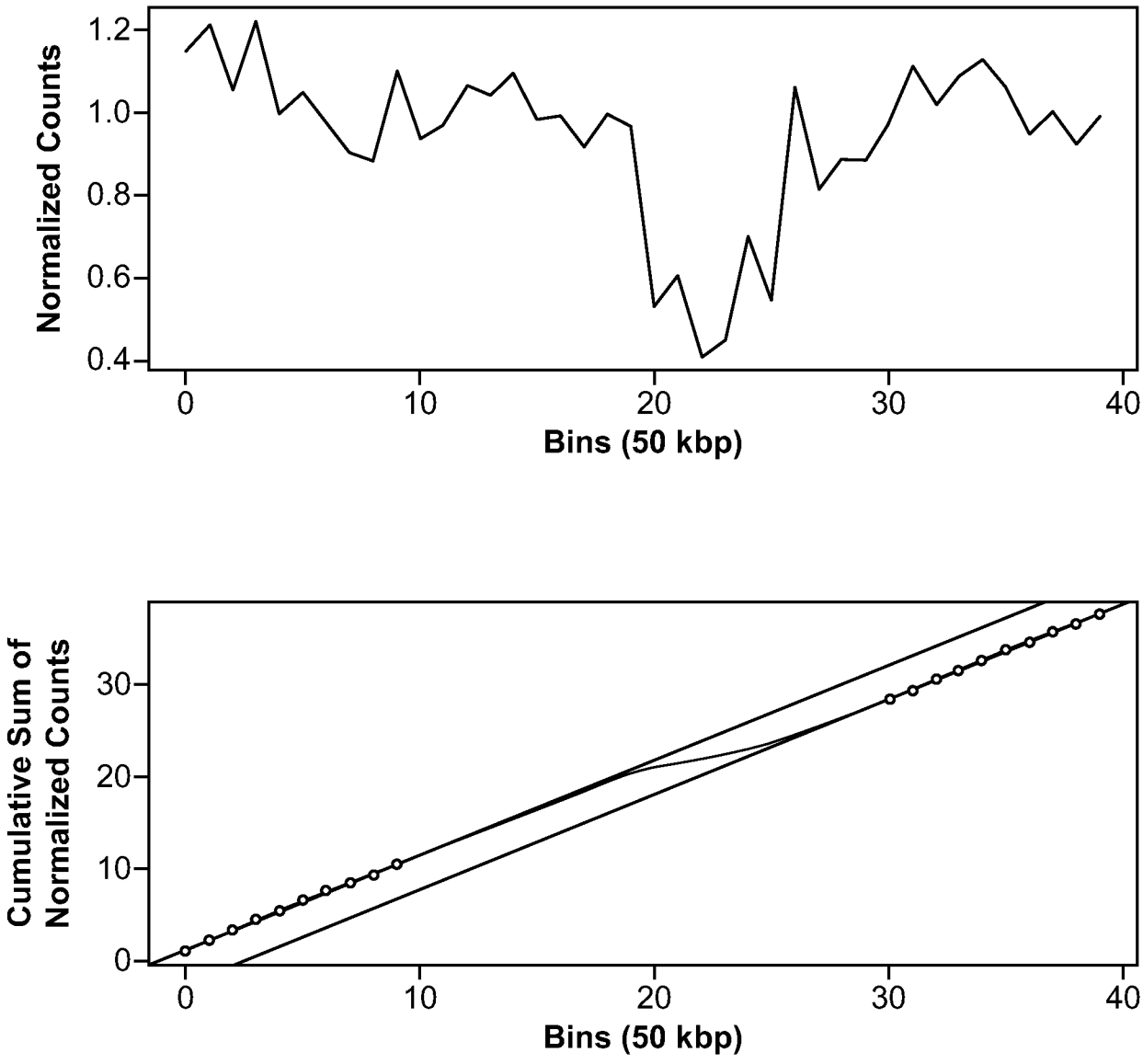

Normalized Counts
1.2
1.0
0.8
0.6
0.4
0
10
20
30
40
Bins (50 kbp)
Cumulative Sum of Normalized Counts
30
20
10
0
0
10
20
30
40
Bins (50 kbp)

FIG. 52

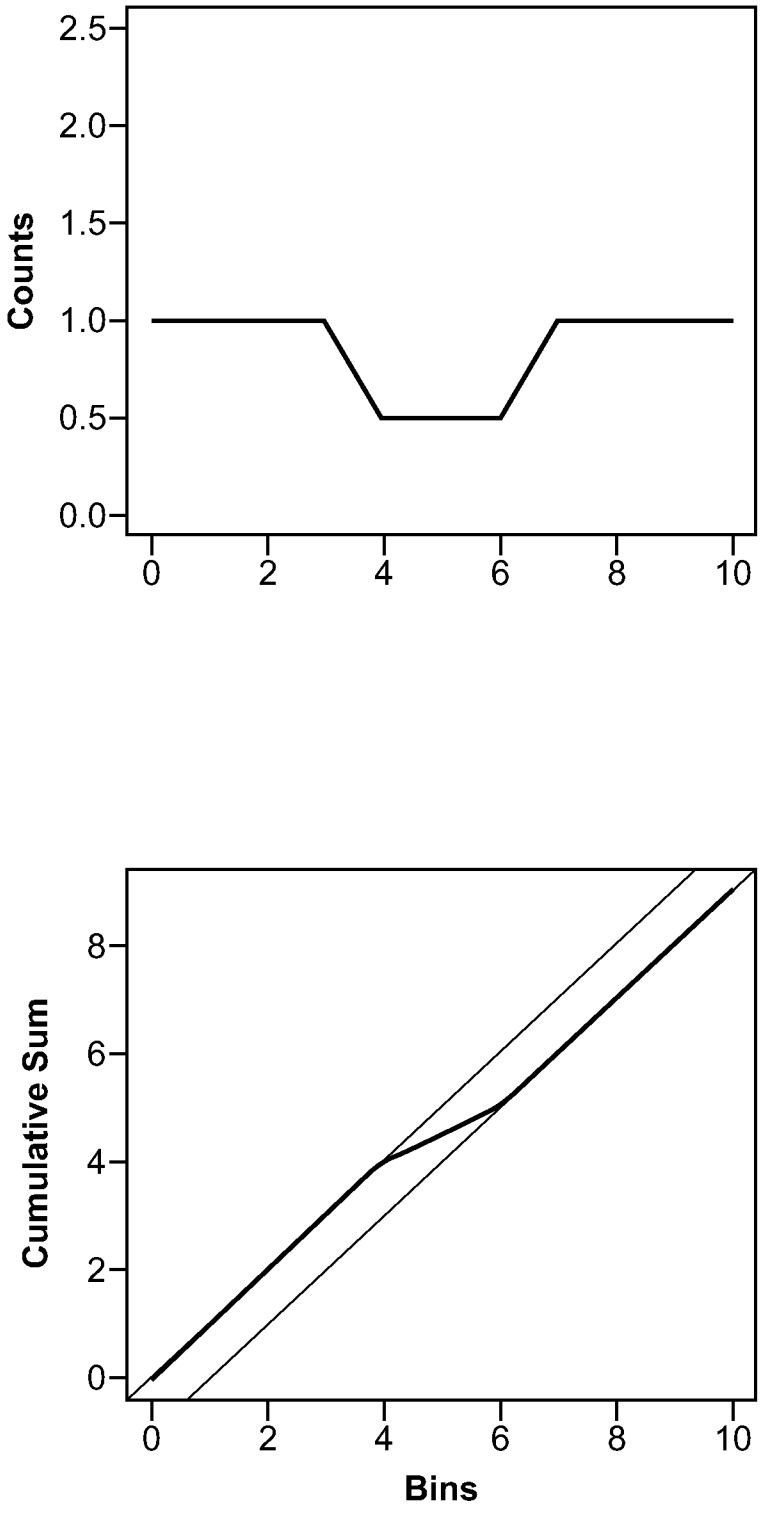

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
8
6
4
2
0
0
2
4
6
8
10
Bins

FIG. 53

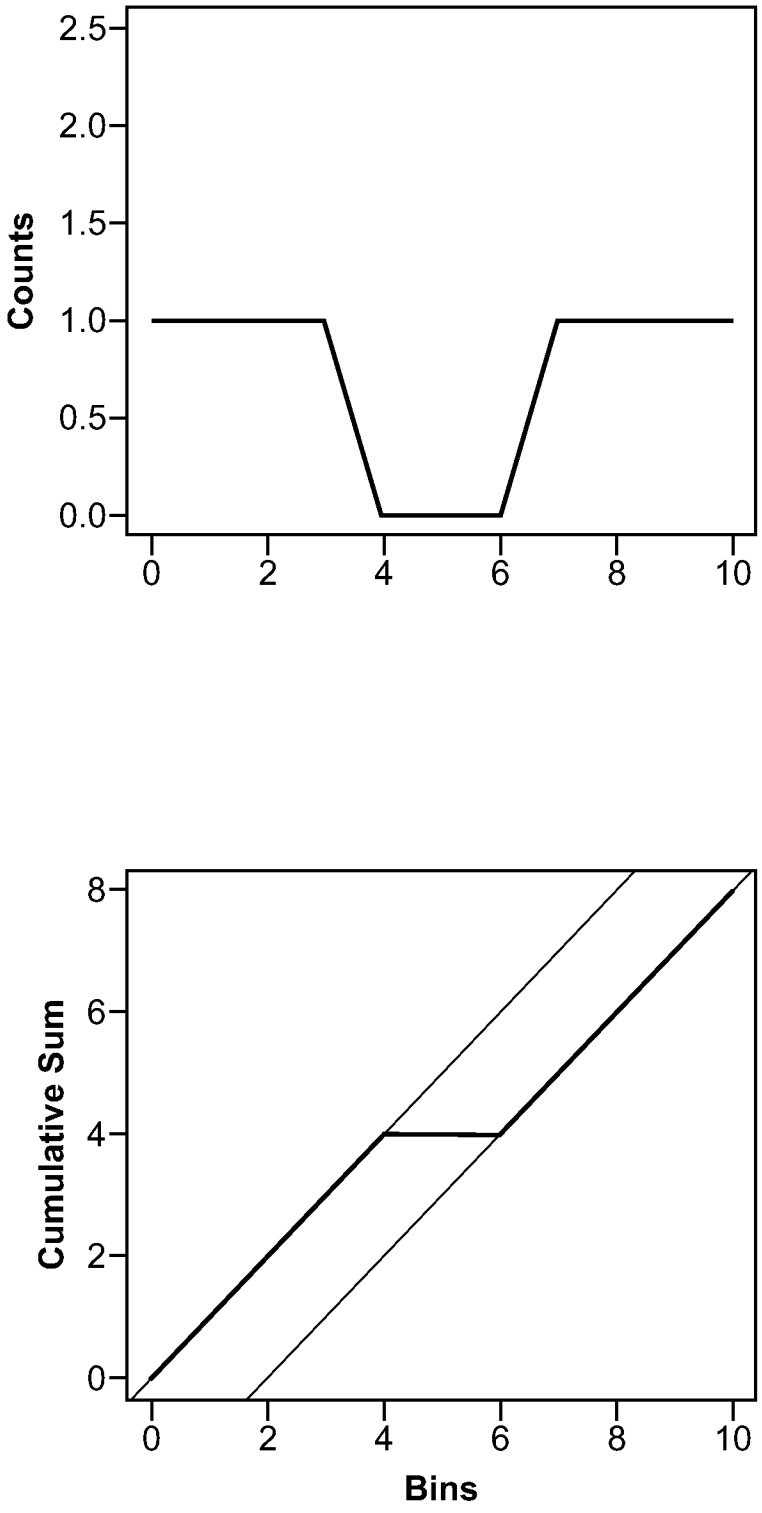

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
8
6
4
2
0
0
2
4
6
8
10
Bins

FIG. 54

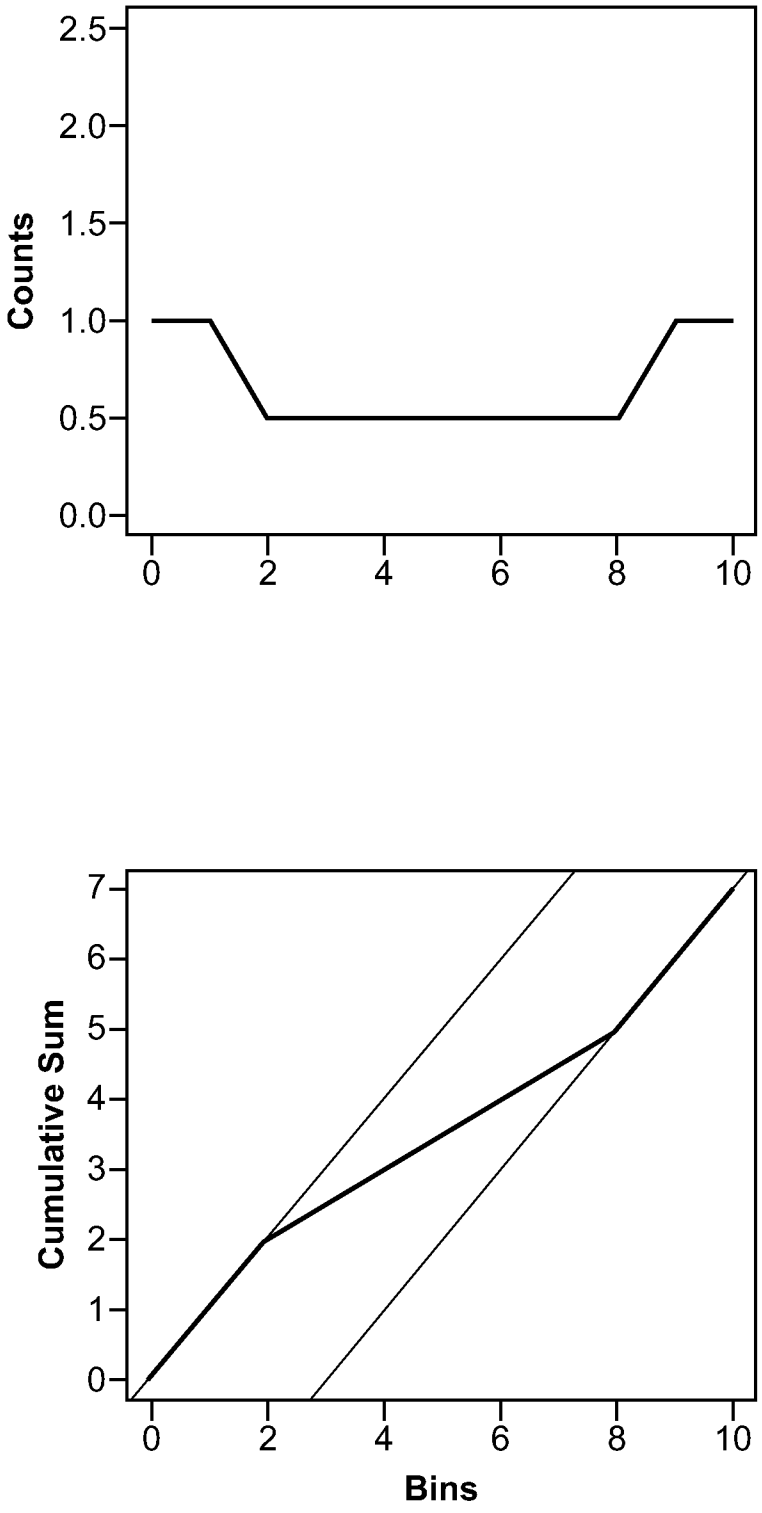

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
7
6
5
4
3
2
1
0
0
2
4
6
8
10
Bins

FIG. 55

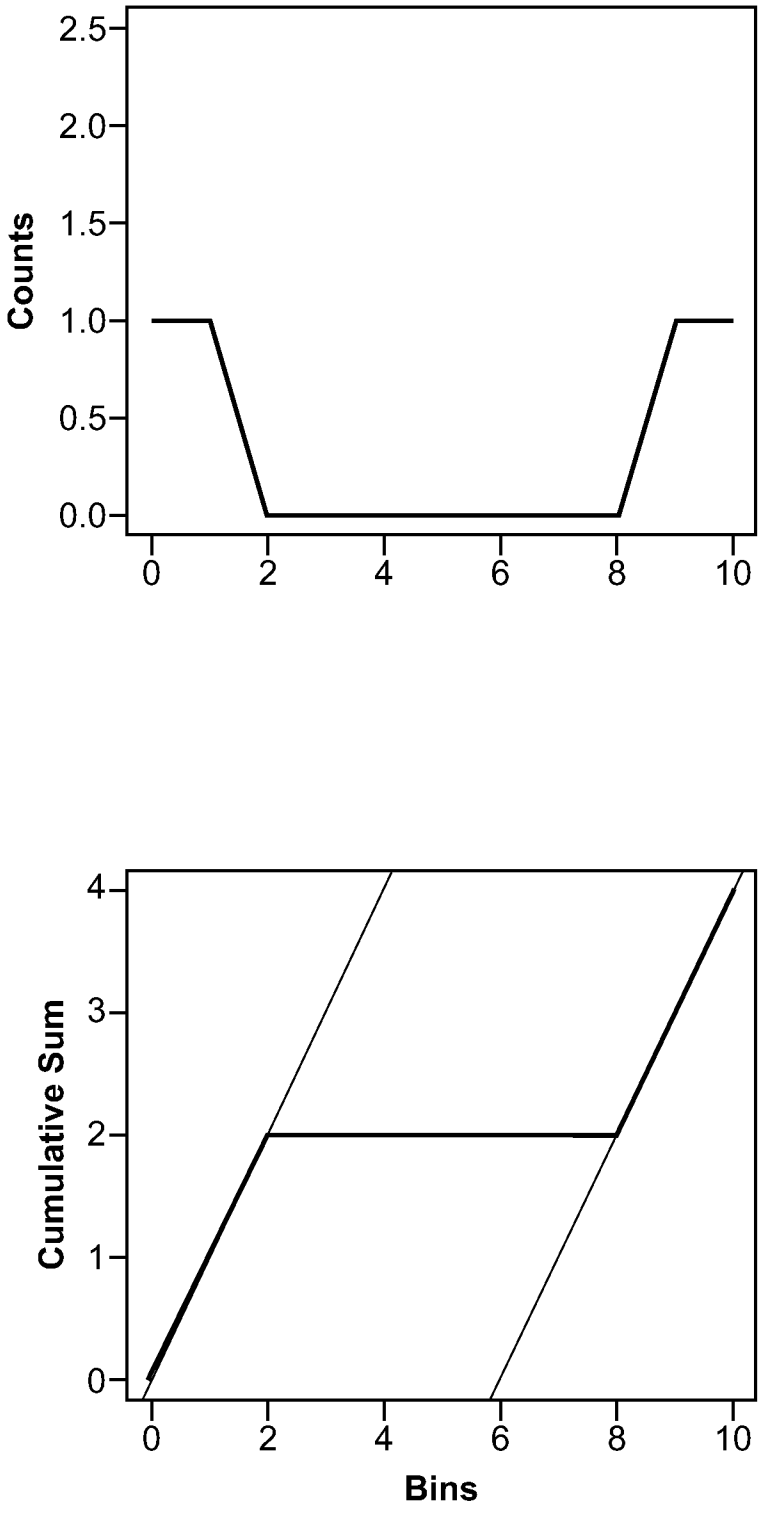

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
4
3
2
1
0
0
2
4
6
8
10
Bins

FIG. 56

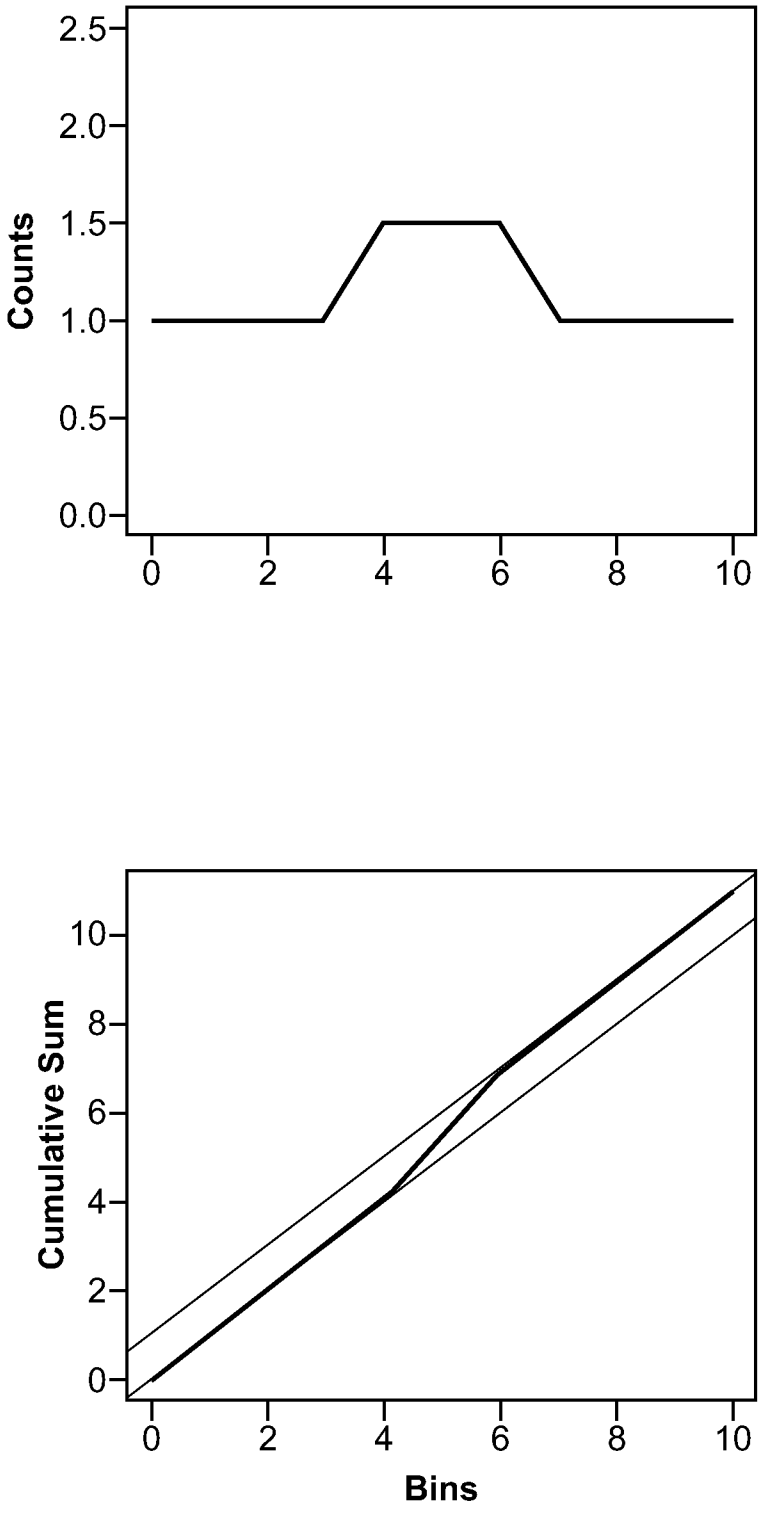

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
10
8
6
4
2
0
0
2
4
6
8
10
Bins

FIG. 57

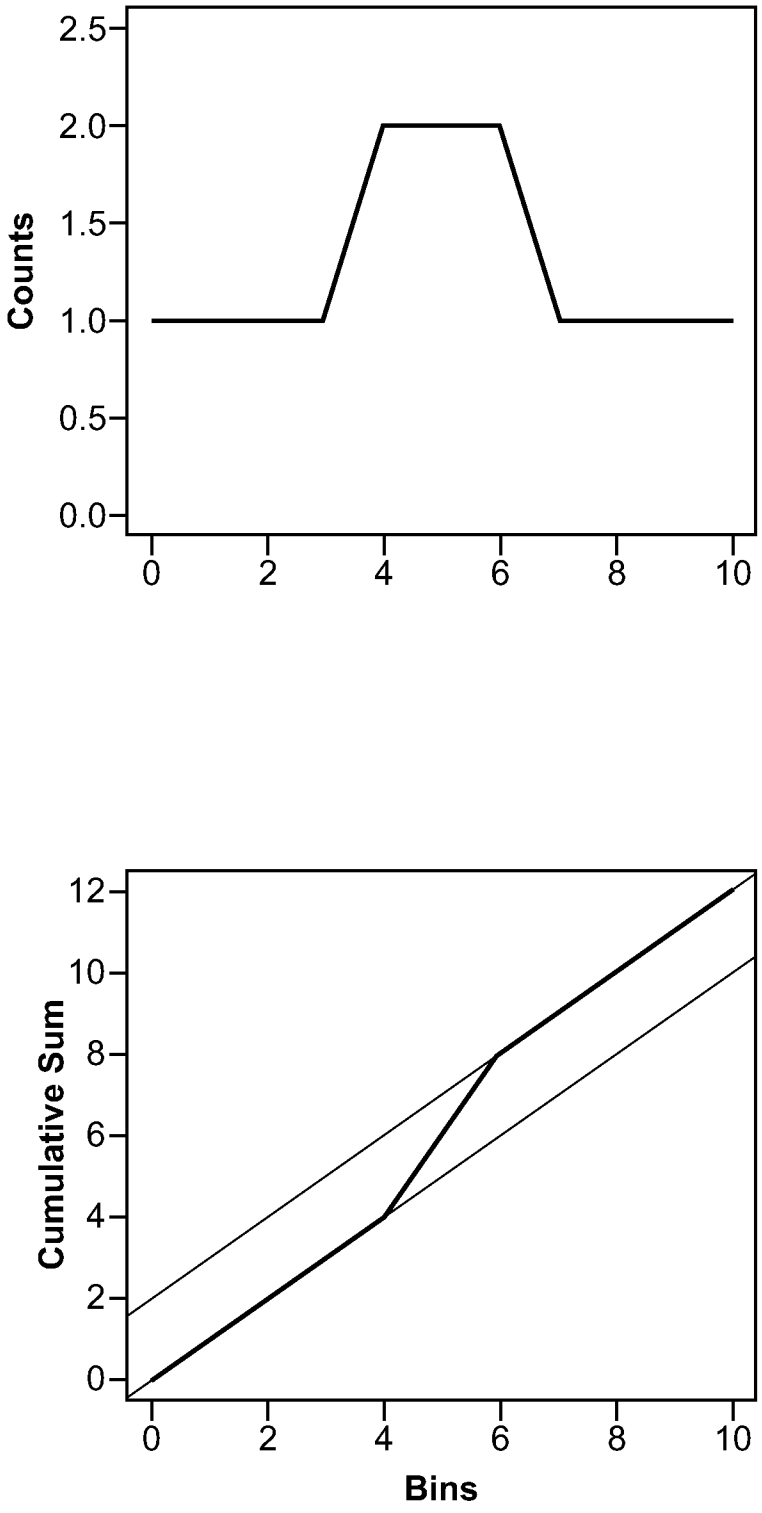

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
12
10
8
6
4
2
0
0
2
4
6
8
10
Bins

FIG. 58

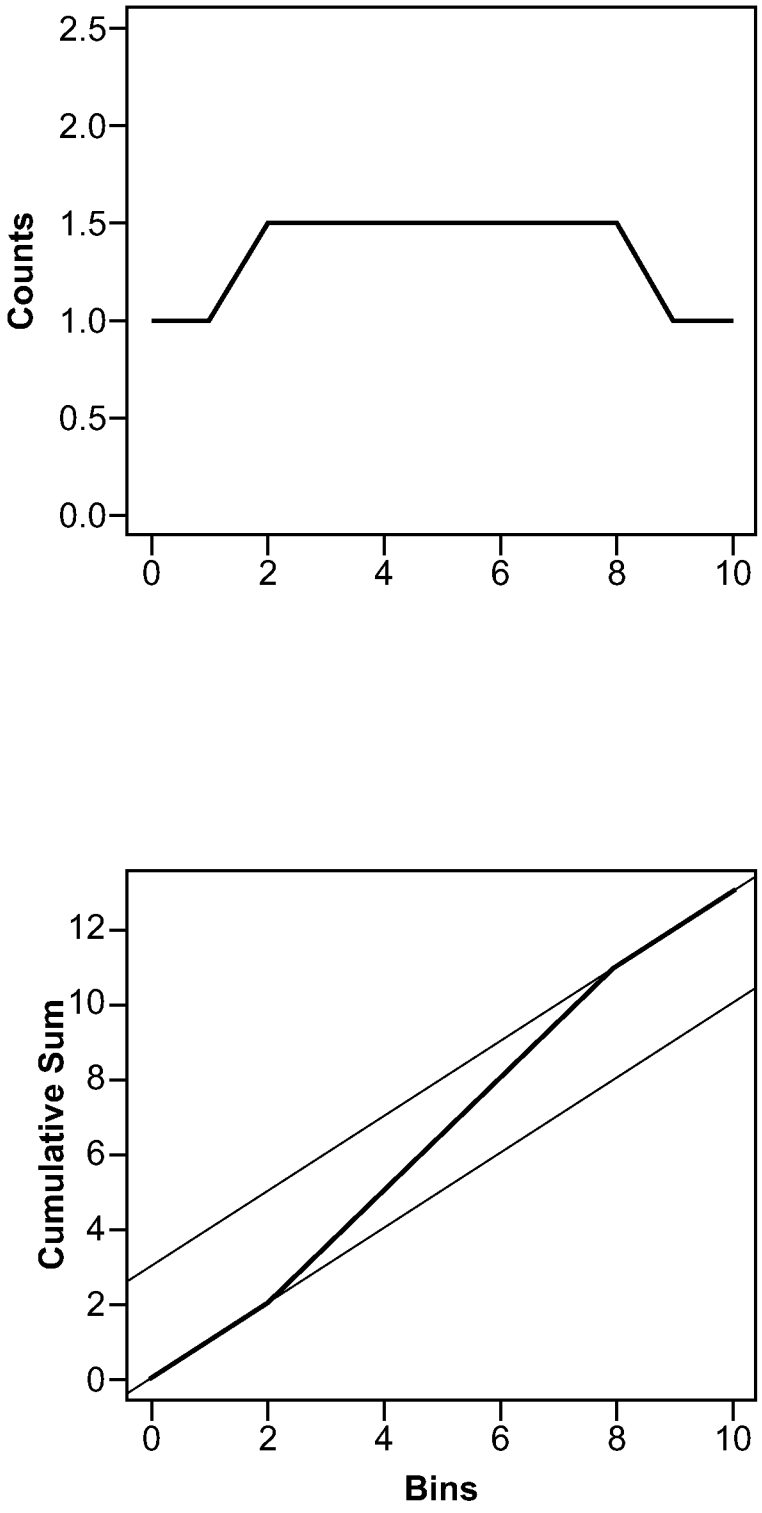

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
12
10
8
6
4
2
0
0
2
4
6
8
10
Bins

FIG. 59

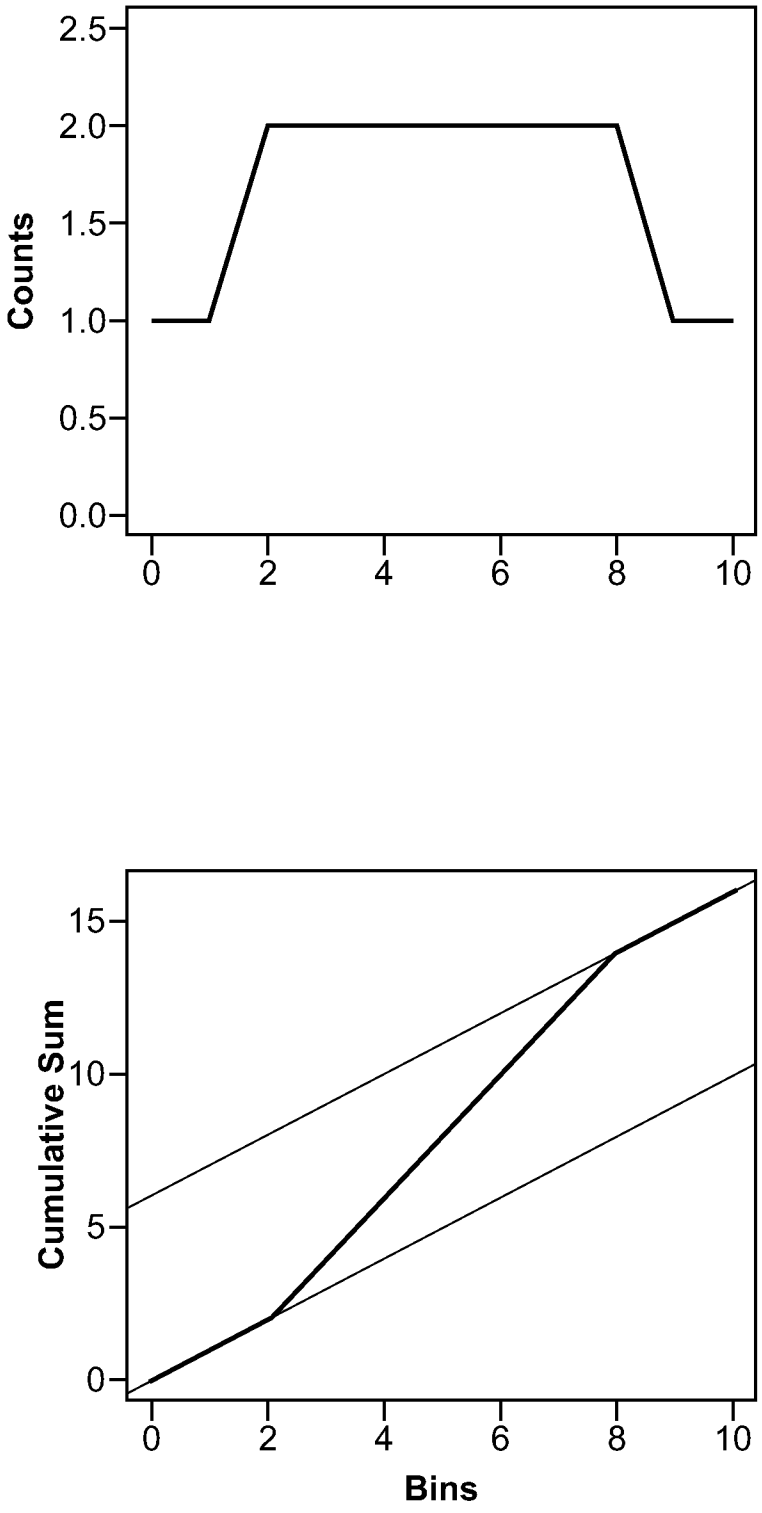

Counts
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Cumulative Sum
15
10
5
0
0
2
4
6
8
10
Bins

FIG. 60

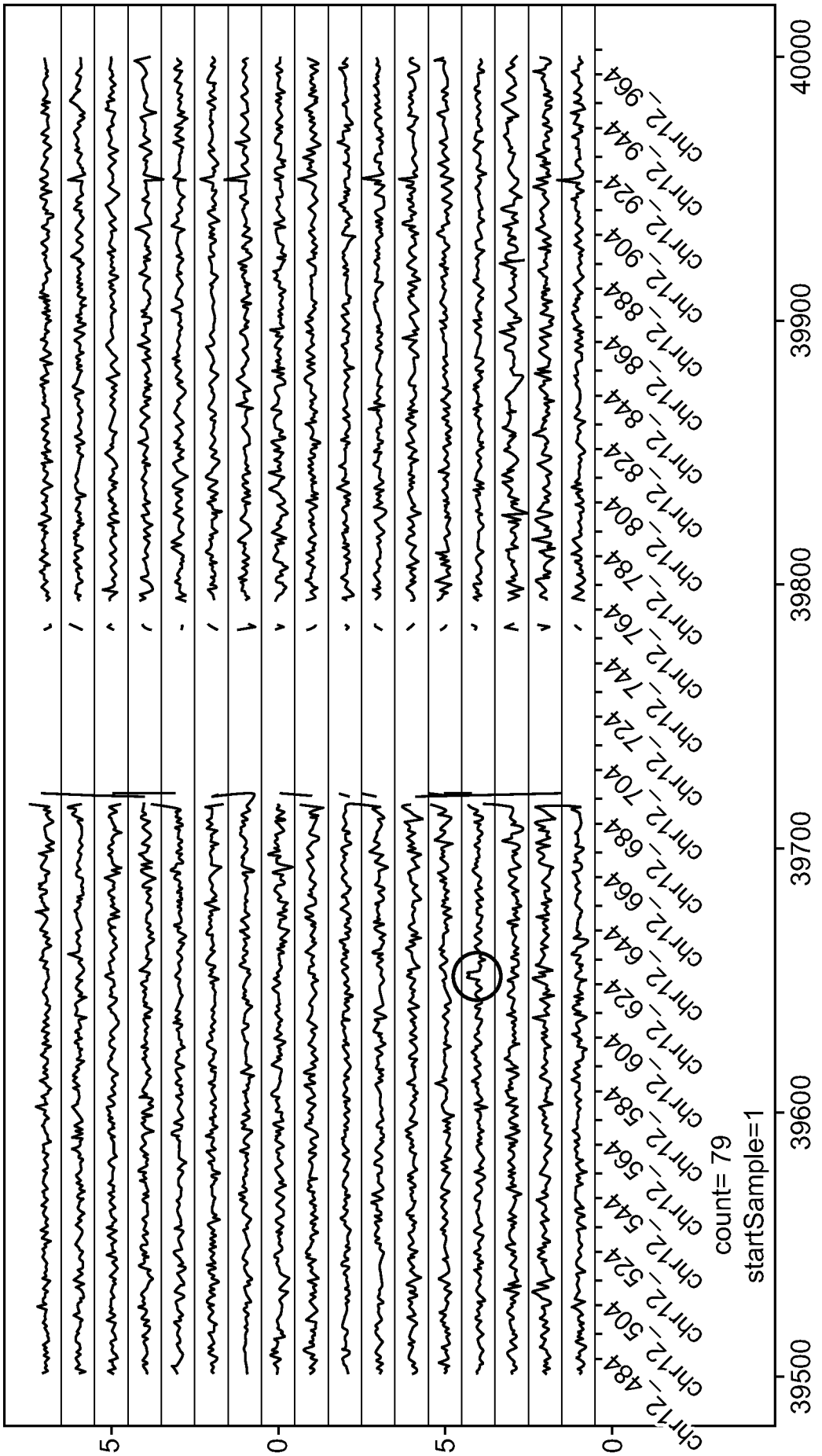

15
10
5
0
39500
39600
39700
39800
39900
40000
startSample=1
count= 79
chr12_48a
chr12_50a
chr12_52a
chr12_54a
chr12_56a
chr12_58a
chr12_60a
chr12_62a
chr12_64a
chr12_66a
chr12_68a
chr12_70a
chr12_72a
chr12_74a
chr12_76a
chr12_78a
chr12_80a
chr12_82a
chr12_84a
chr12_86a
chr12_88a
chr12_90a
chr12_92a
chr12_94a
chr12_96a

FIG. 61A

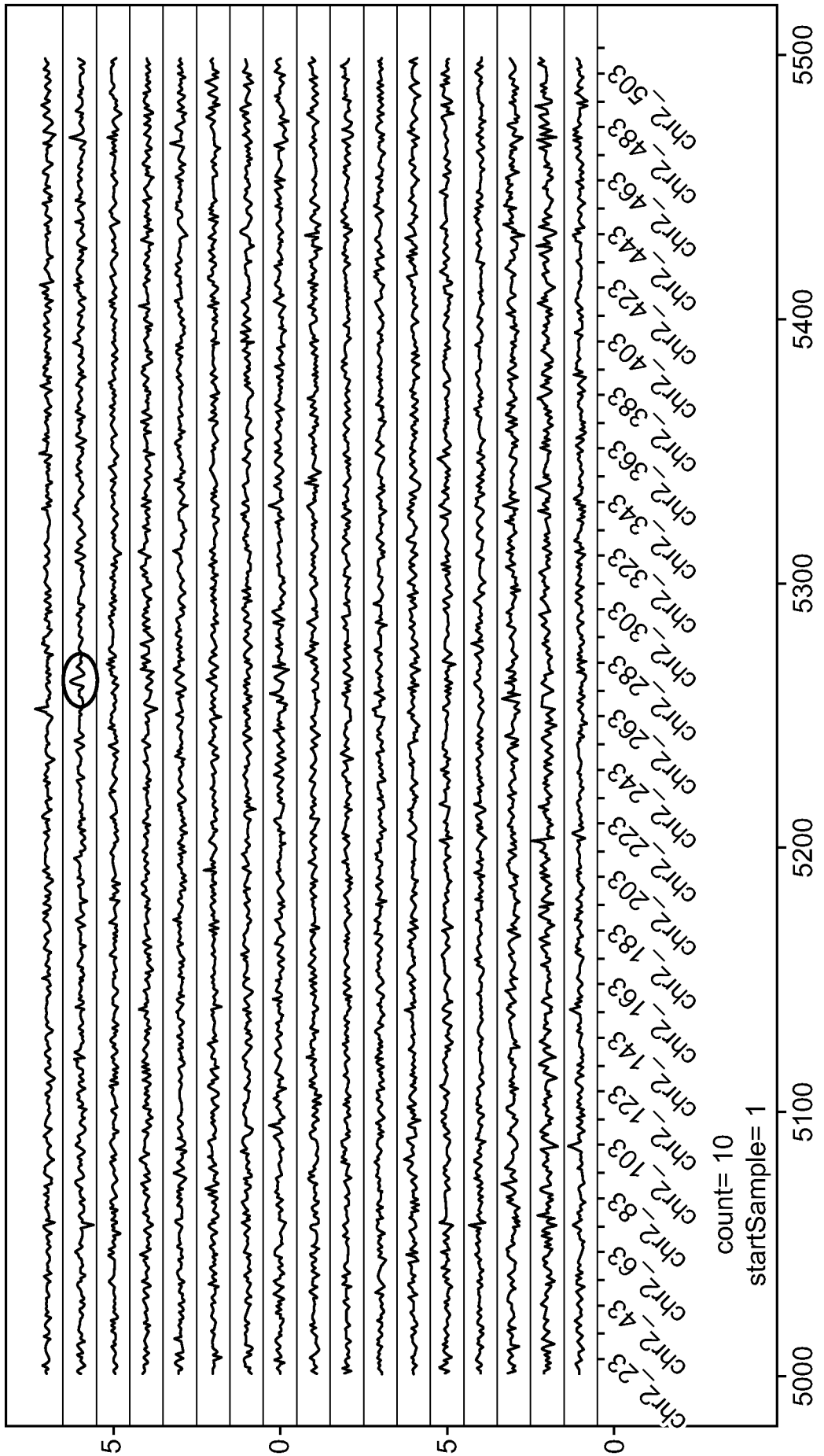

0
5
10
15
5000
5100
5200
5300
5400
5500
startSample= 1
count= 10
chr2_23
chr2_43
chr2_63
chr2_83
chr2_103
chr2_123
chr2_143
chr2_163
chr2_183
chr2_203
chr2_223
chr2_243
chr2_263
chr2_283
chr2_303
chr2_323
chr2_343
chr2_363
chr2_383
chr2_403
chr2_423
chr2_443
chr2_463
chr2_483
chr2_503

FIG. 61B

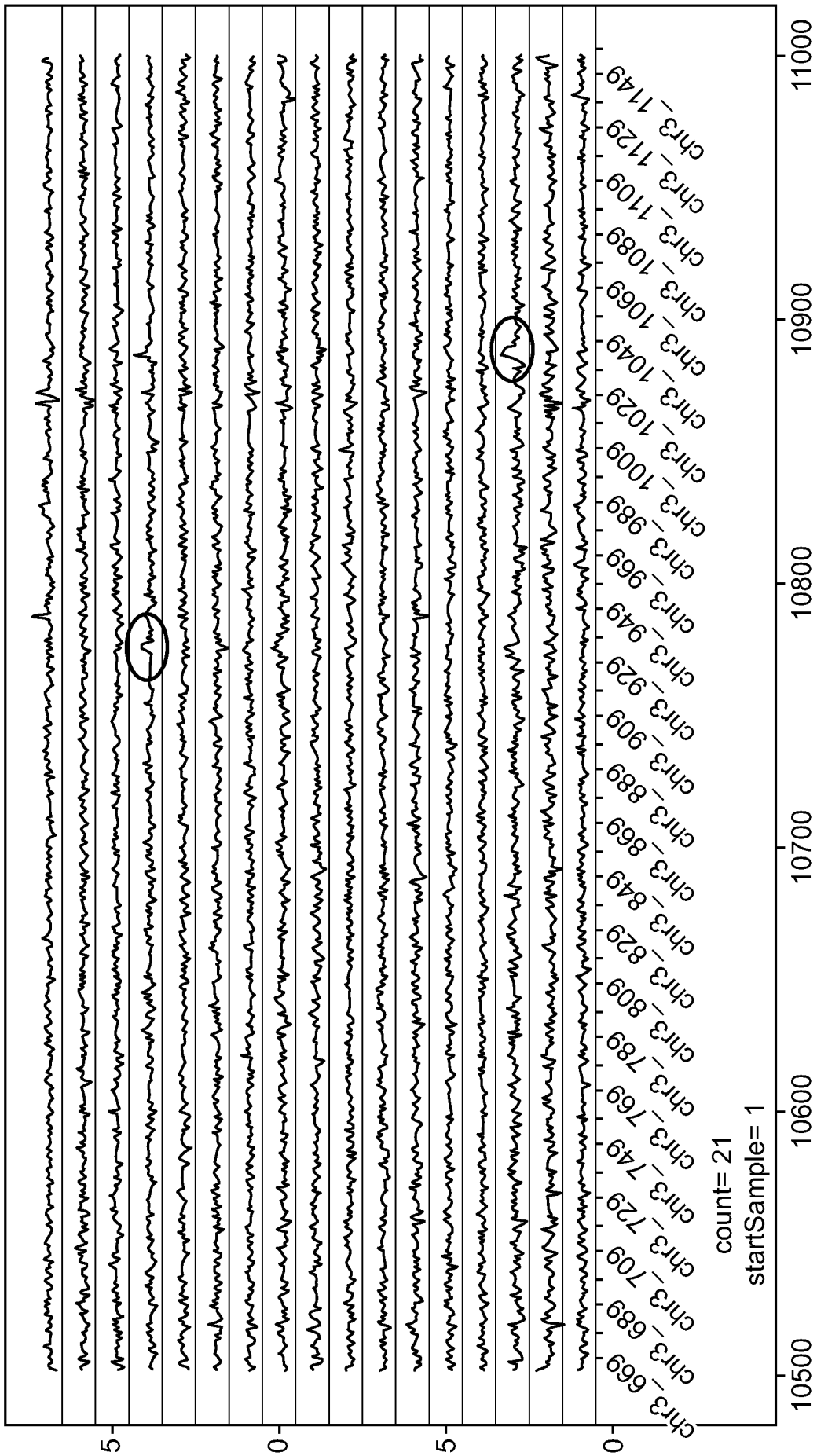

15
10
5
0
10500
10600
10700
10800
10900
11000
count= 21
startSample= 1
chr3_669
chr3_689
chr3_709
chr3_729
chr3_749
chr3_769
chr3_789
chr3_809
chr3_829
chr3_849
chr3_869
chr3_889
chr3_909
chr3_929
chr3_949
chr3_969
chr3_989
chr3_1009
chr3_1029
chr3_1049
chr3_1069
chr3_1089
chr3_1109
chr3_1129
chr3_1149

FIG. 61C

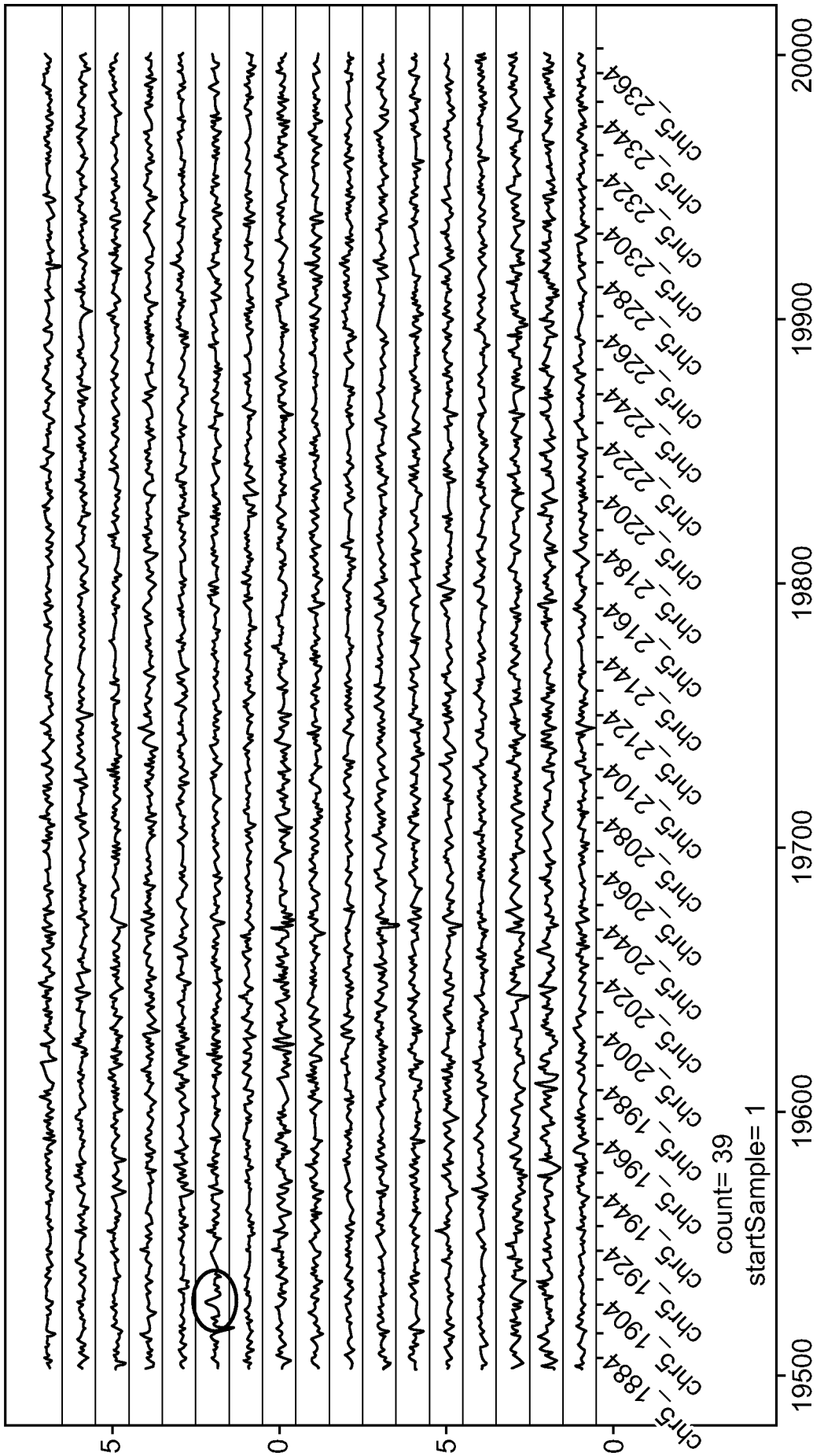

15
10
5
0
19500
19600
19700
19800
19900
20000
startSample= 1
count= 39
chr5_188A
chr5_190A
chr5_192A
chr5_194A
chr5_196A
chr5_198A
chr5_200A
chr5_202A
chr5_204A
chr5_206A
chr5_208A
chr5_210A
chr5_212A
chr5_214A
chr5_216A
chr5_218A
chr5_220A
chr5_222A
chr5_224A
chr5_226A
chr5_228A
chr5_230A
chr5_232A
chr5_234A
chr5_236A

FIG. 61D

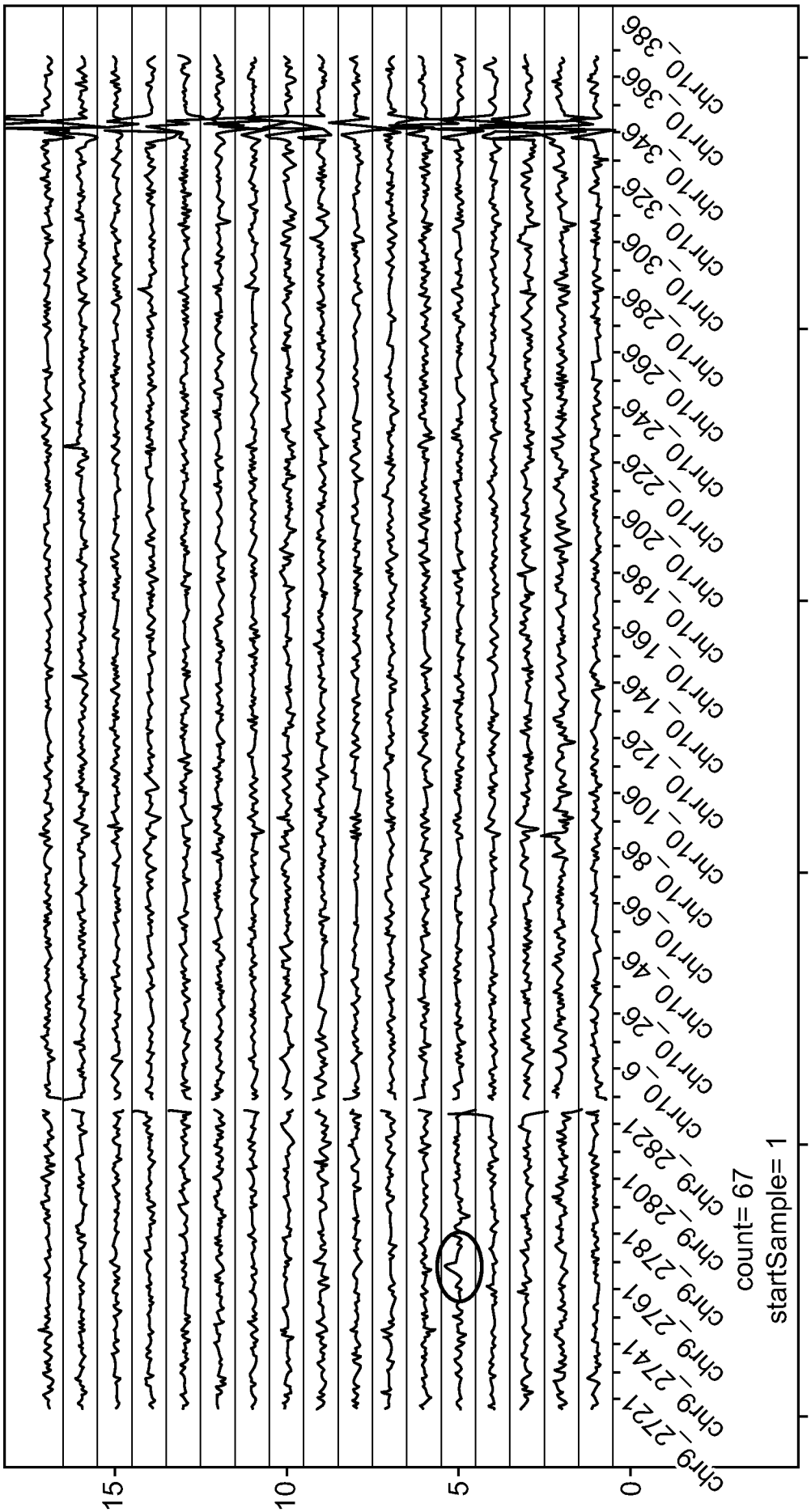

15
10
5
0
33500
33600
33700
33800
33900
34000
chr9_2721
chr9_2741
chr9_2761
chr9_2781
chr9_2801
chr9_2821
chr10_6
chr10_26
chr10_46
chr10_66
chr10_86
chr10_106
chr10_126
chr10_146
chr10_166
chr10_186
chr10_206
chr10_226
chr10_246
chr10_266
chr10_286
chr10_306
chr10_326
chr10_346
chr10_366
chr10_386
count= 67
startSample= 1

FIG. 61E

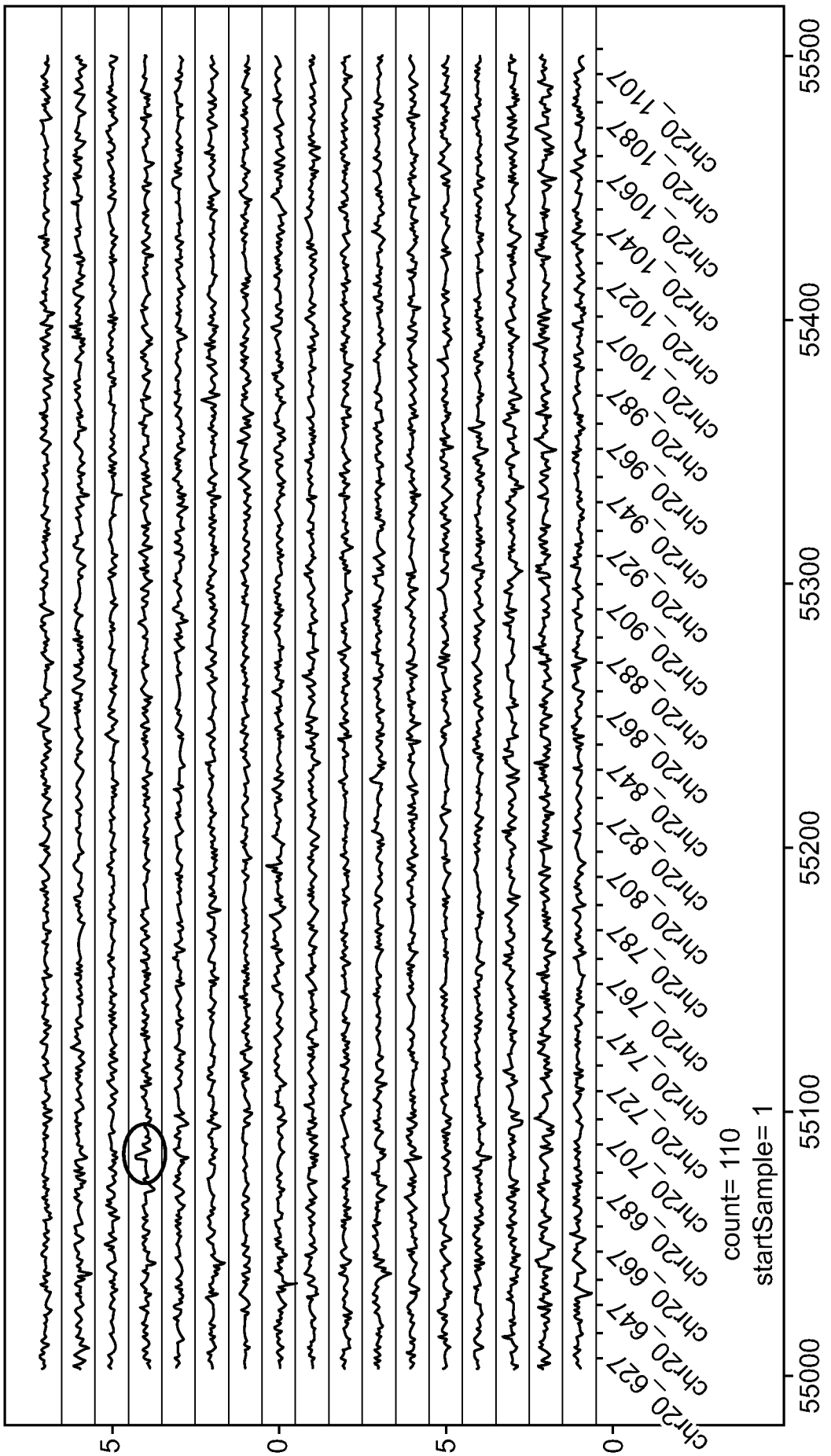

15
10
5
0
55000
55100
55200
55300
55400
55500
count= 110
startSample= 1
chr20_627
chr20_647
chr20_667
chr20_687
chr20_707
chr20_727
chr20_747
chr20_767
chr20_787
chr20_807
chr20_827
chr20_847
chr20_867
chr20_887
chr20_907
chr20_927
chr20_947
chr20_967
chr20_987
chr20_1007
chr20_1027
chr20_1047
chr20_1067
chr20_1087
chr20_1107

FIG. 61F

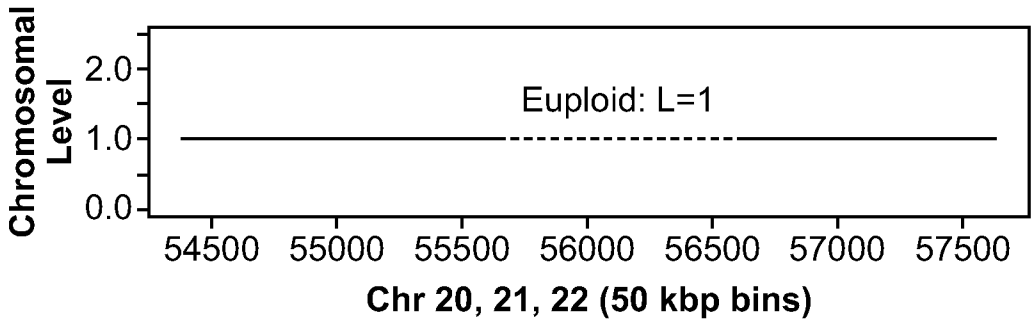

Chromosomal Level
2.0
1.0
0.0
Euploid: L=1
54500
55000
55500
56000
56500
57000
57500
Chr 20, 21, 22 (50 kbp bins)

FIG. 62

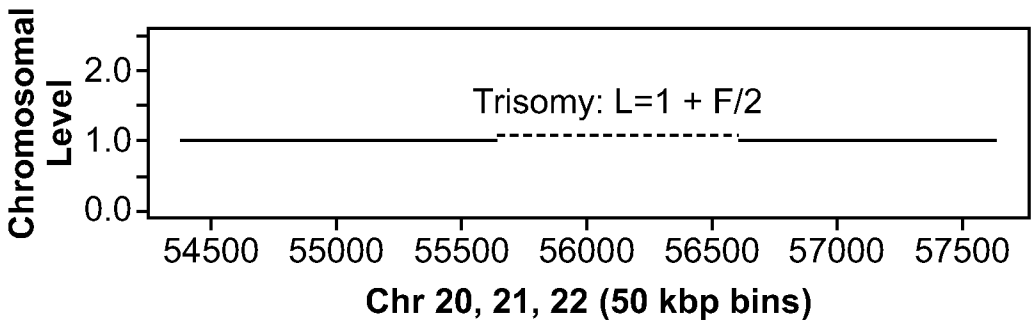

Chromosomal Level
2.0
1.0
0.0
Trisomy: L=1 + F/2
54500
55000
55500
56000
56500
57000
57500
Chr 20, 21, 22 (50 kbp bins)

FIG. 63

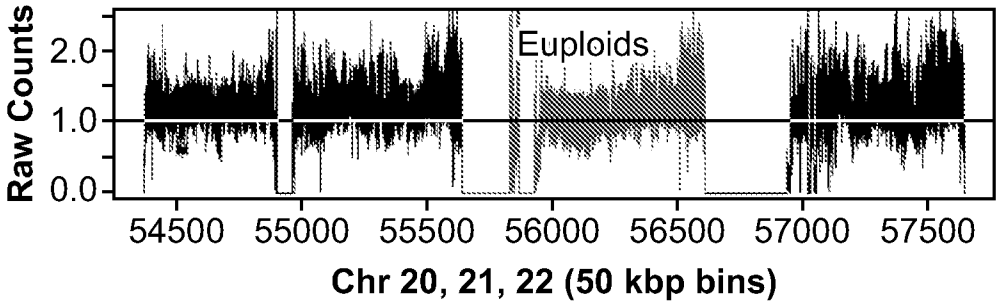

Raw Counts
2.0
1.0
0.0
Euploids
54500
55000
55500
56000
56500
57000
57500
Chr 20, 21, 22 (50 kbp bins)

FIG. 64

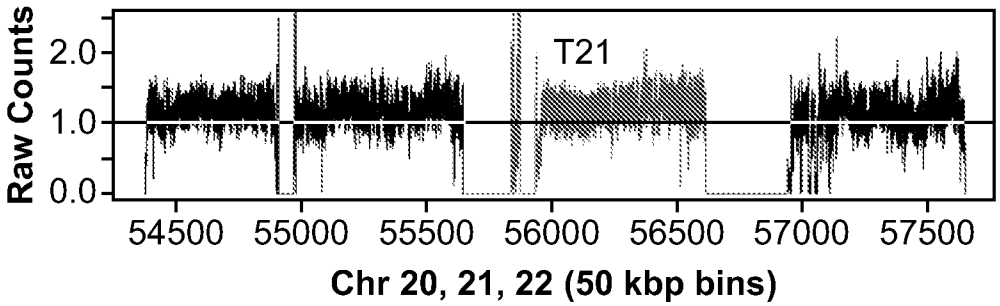

Raw Counts
2.0
1.0
0.0
T21
54500
55000
55500
56000
56500
57000
57500
Chr 20, 21, 22 (50 kbp bins)

FIG. 65

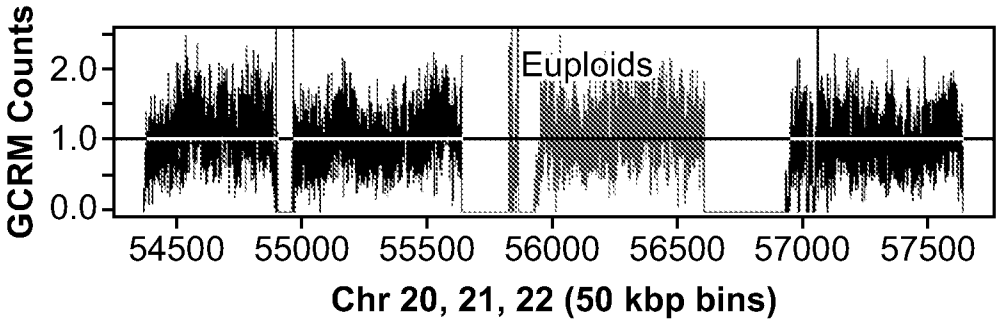

GCRM Counts
2.0
1.0
0.0
Euploids
54500
55000
55500
56000
56500
57000
57500
Chr 20, 21, 22 (50 kbp bins)

FIG. 66

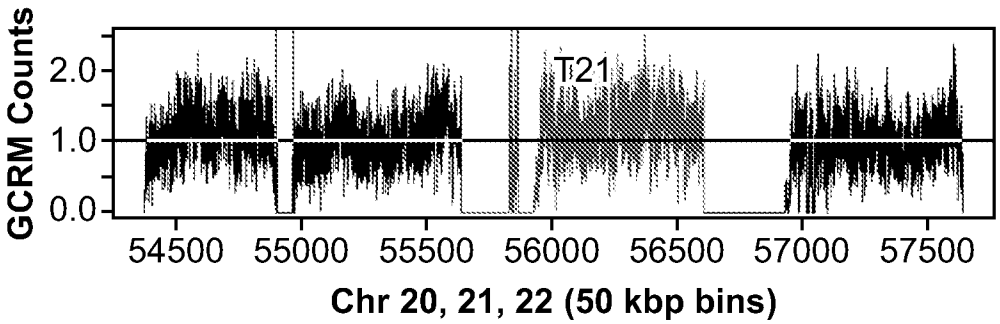

GCRM Counts
2.0
1.0
0.0
T21
54500
55000
55500
56000
56500
57000
57500
Chr 20, 21, 22 (50 kbp bins)

FIG. 67

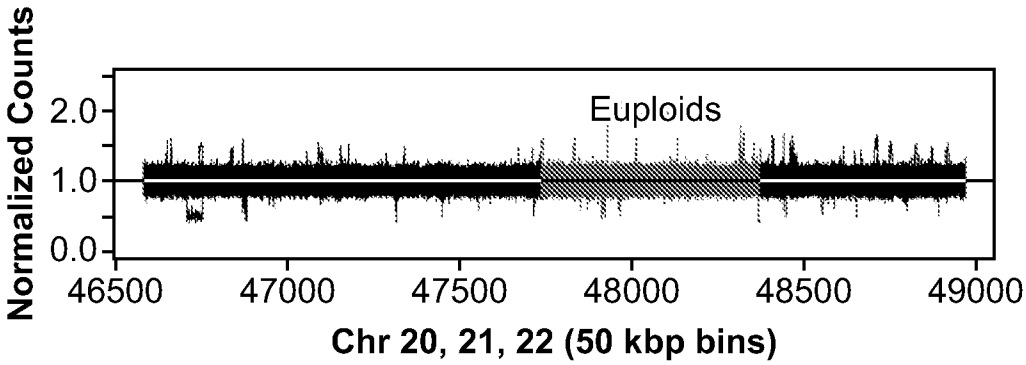

Normalized Counts
2.0
1.0
0.0
Euploids
46500
47000
47500
48000
48500
49000
Chr 20, 21, 22 (50 kbp bins)

FIG. 68

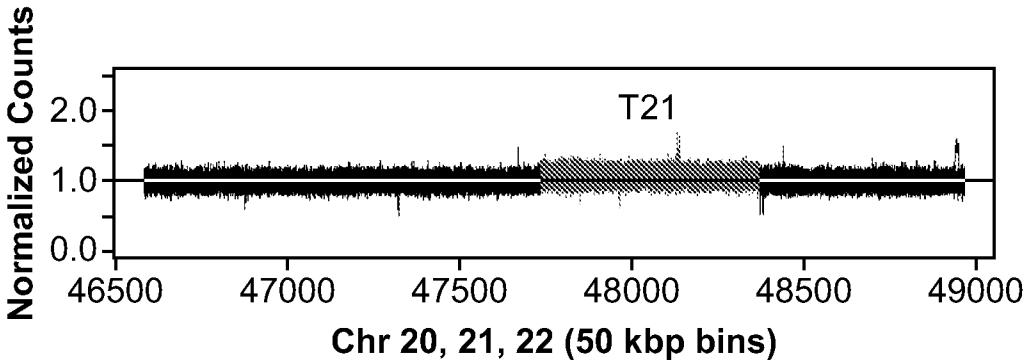
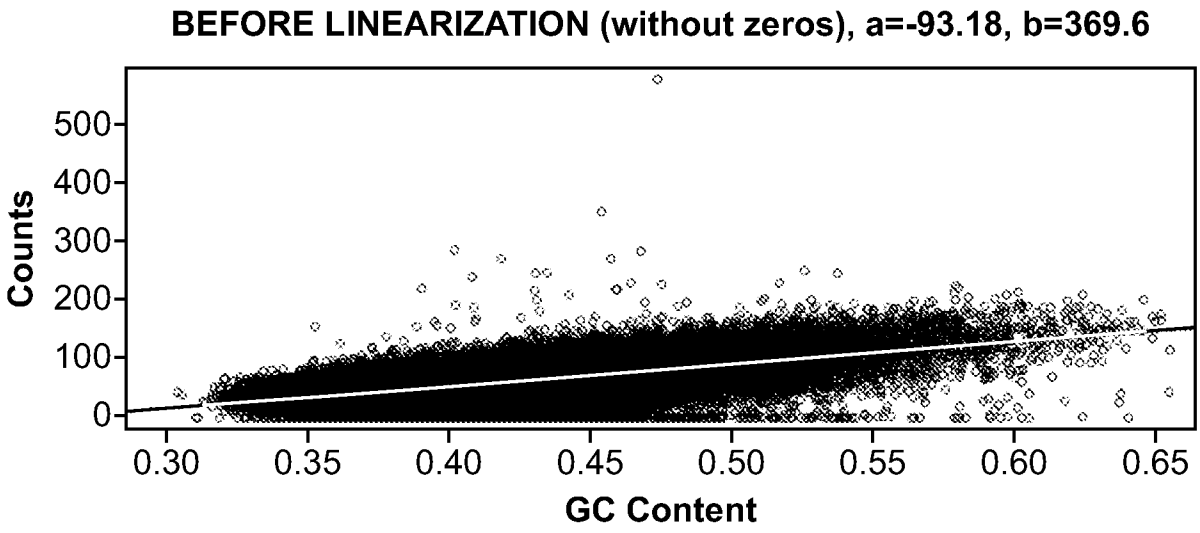

Normalized Counts
2.0
1.0
0.0
T21
46500
47000
47500
48000
48500
49000
Chr 20, 21, 22 (50 kbp bins)

FIG. 69


BEFORE LINEARIZATION (without zeros), a=-93.18, b=369.6
Counts
500
400
300
200
100
0
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
GC Content

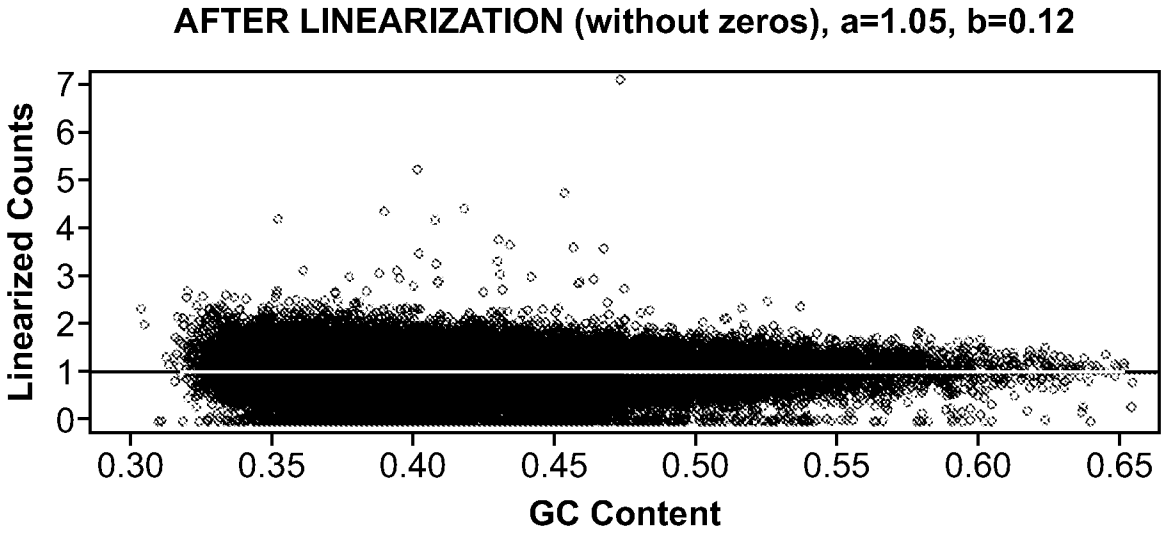

AFTER LINEARIZATION (without zeros), a=1.05, b=0.12
Linearized Counts
7
6
5
4
3
2
1
0
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
GC Content

FIG. 70

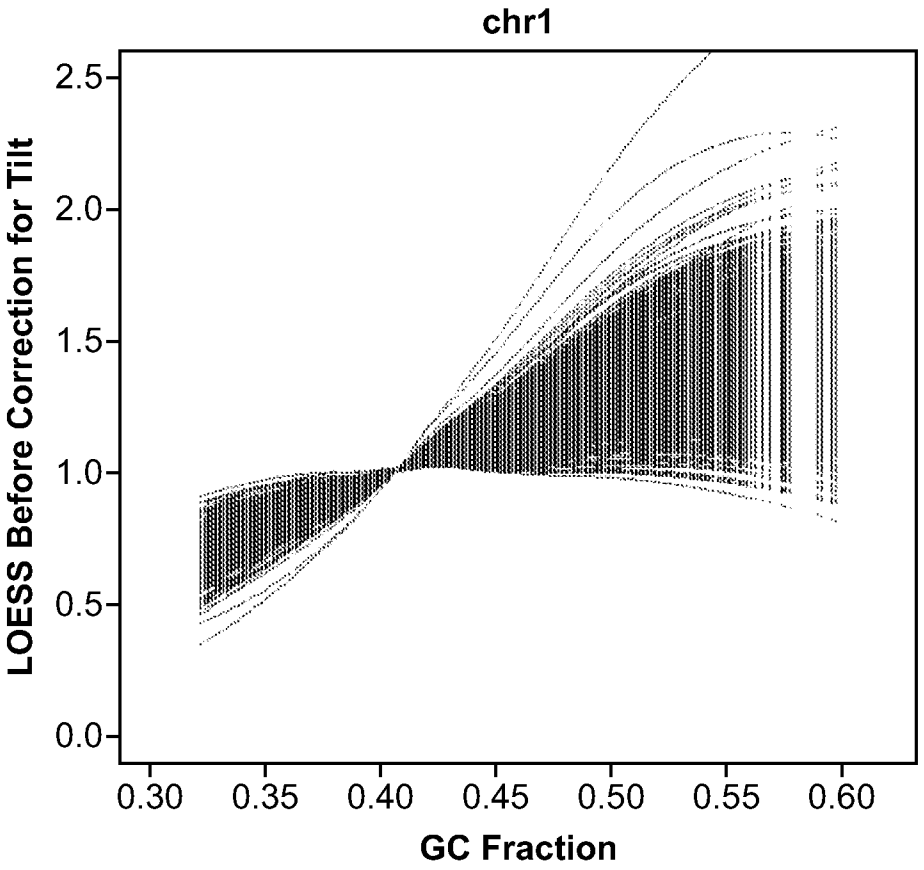

chr1
LOESS Before Correction for Tilt
2.5
2.0
1.5
1.0
0.5
0.0
0.30
0.35
0.40
0.45
0.50
0.55
0.60
GC Fraction

FIG. 71

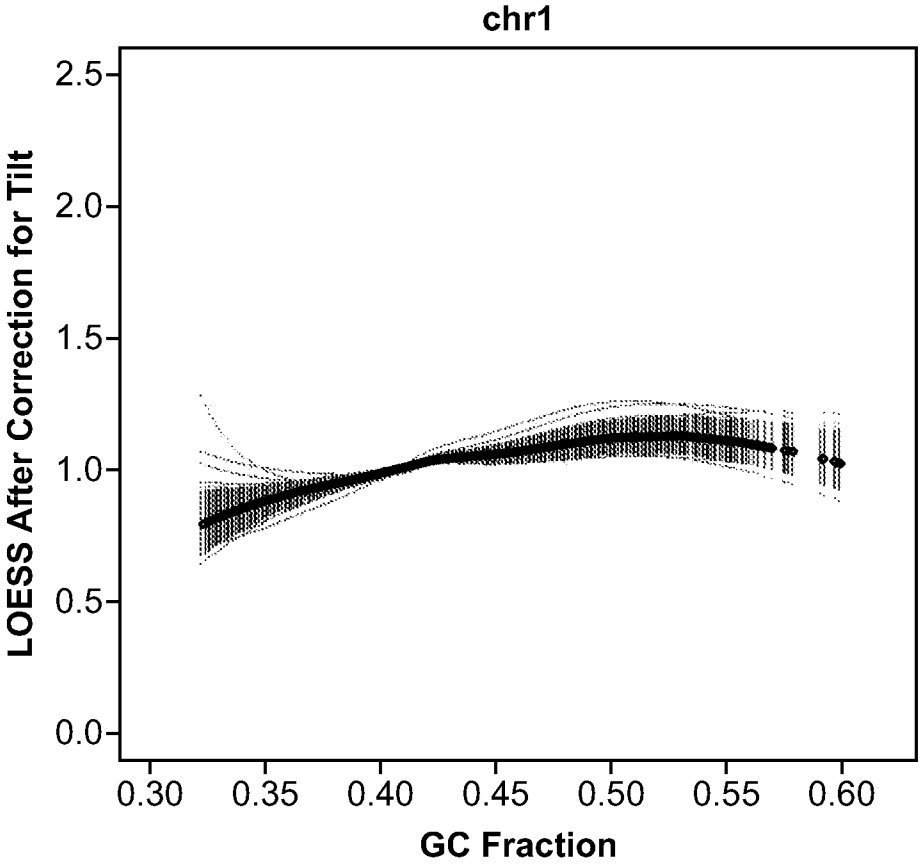

chr1
LOESS After Correction for Tilt
2.5
2.0
1.5
1.0
0.5
0.0
0.30
0.35
0.40
0.45
0.50
0.55
0.60
GC Fraction

FIG. 72

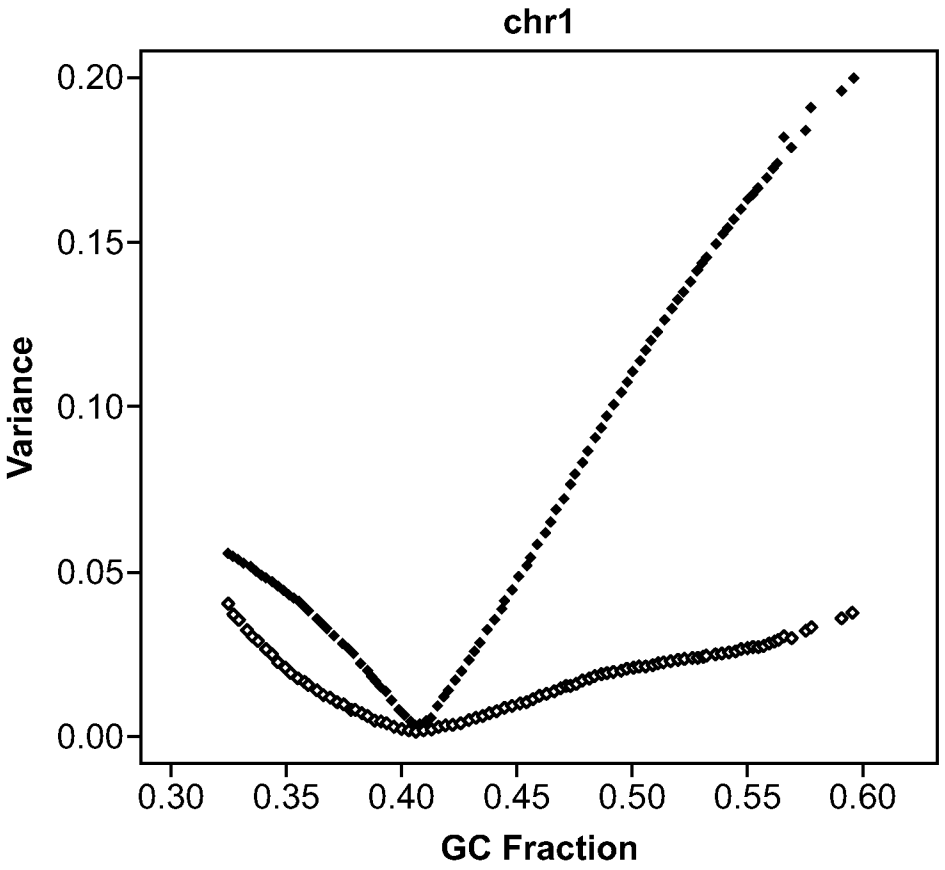

chr1
0.20
0.15
0.10
0.05
0.00
Variance
0.30
0.35
0.40
0.45
0.50
0.55
0.60
GC Fraction

FIG. 73

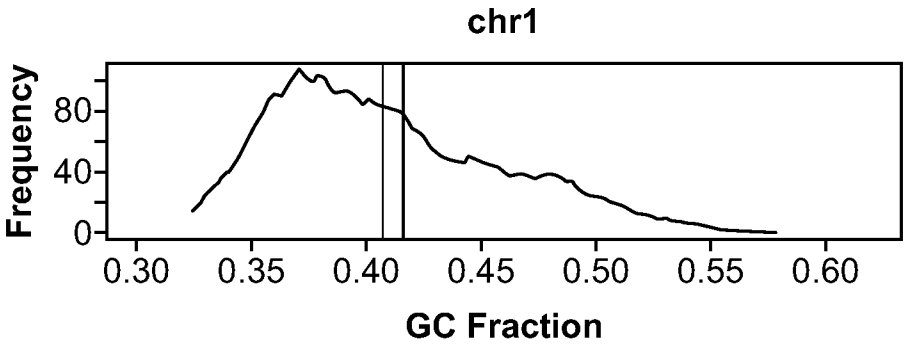


FIG. 74

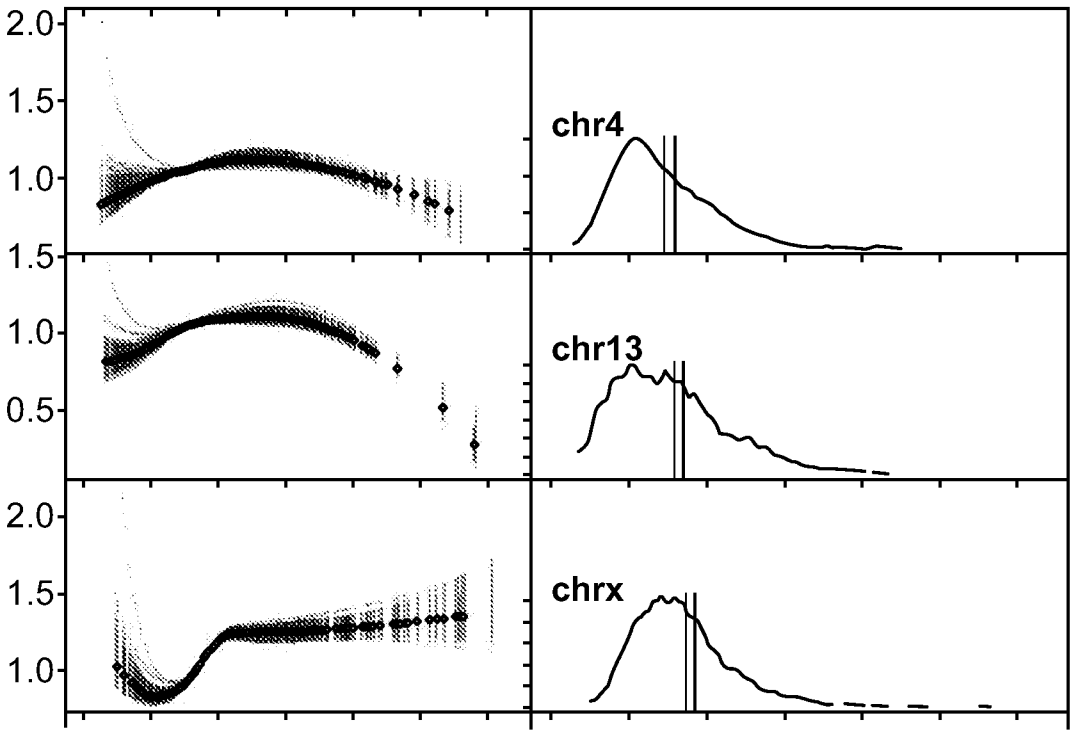


FIG. 75A

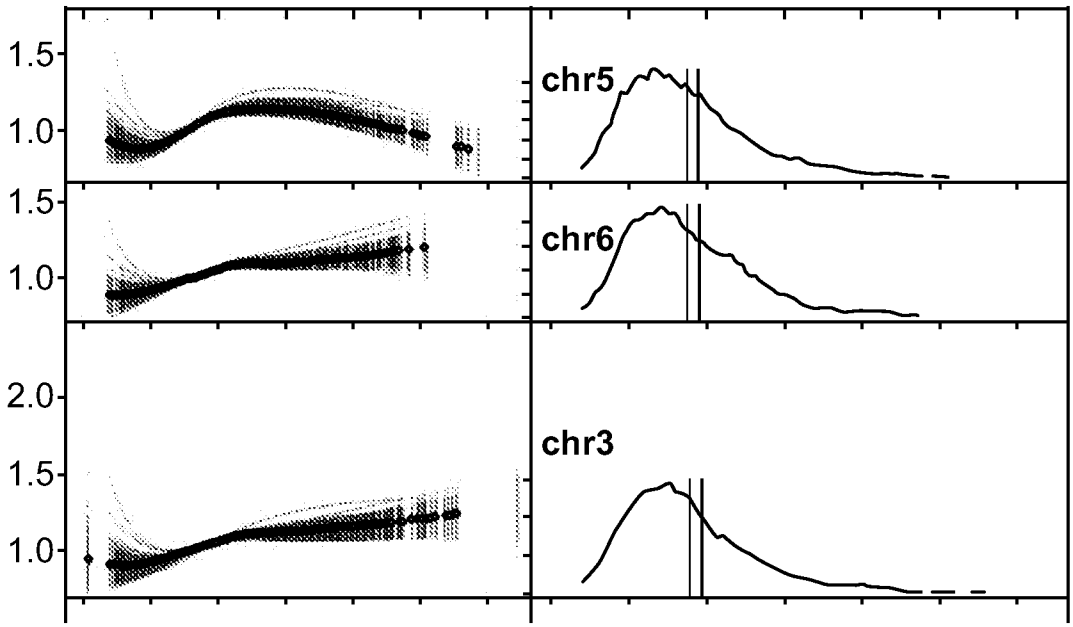


FIG. 75B

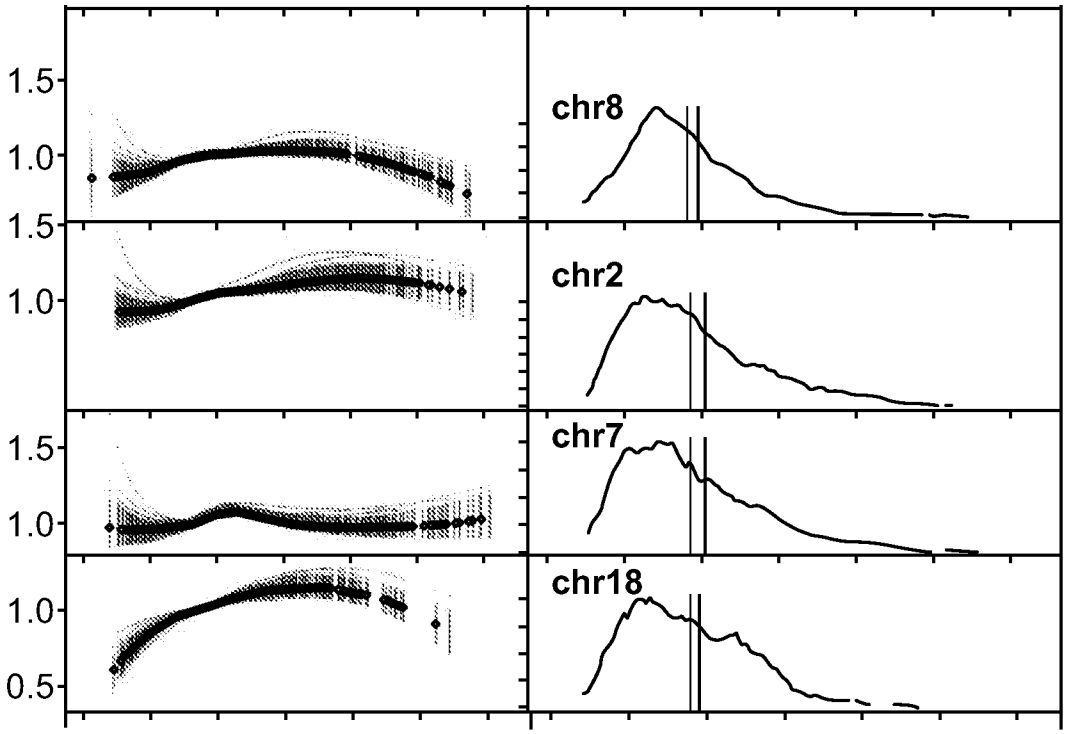


FIG. 75C

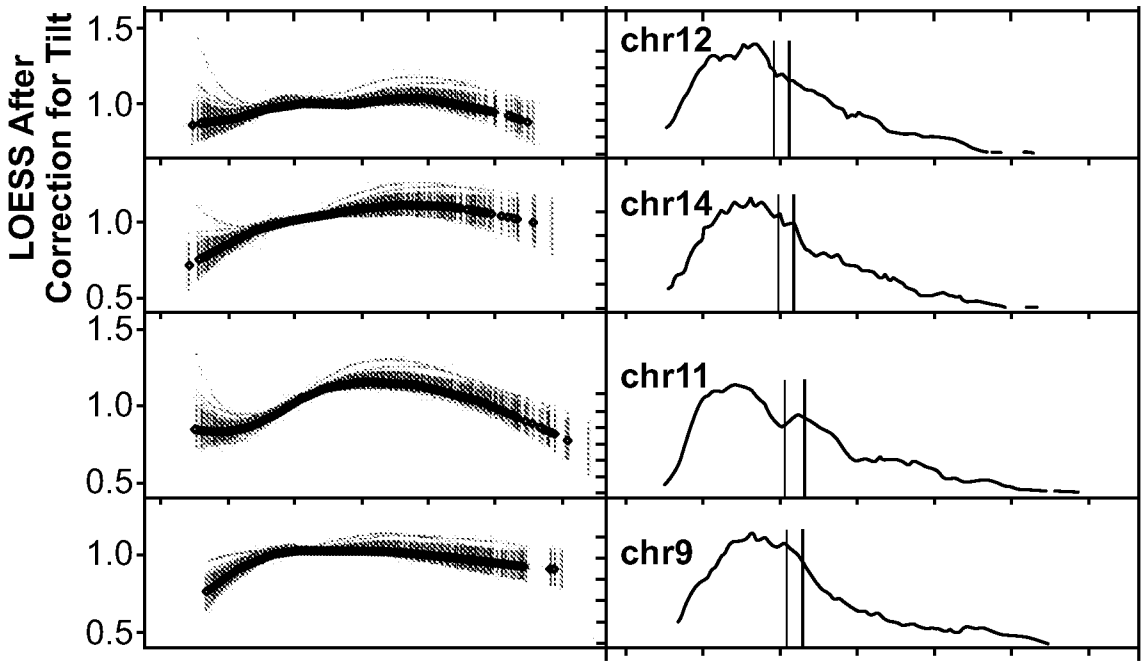


FIG. 75D

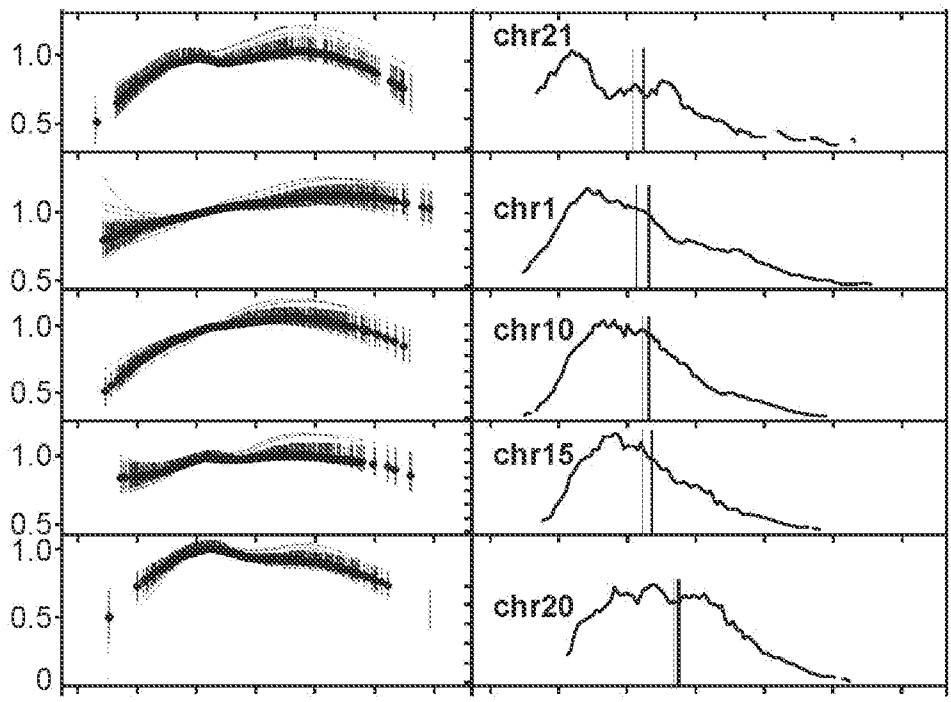


FIG. 75E

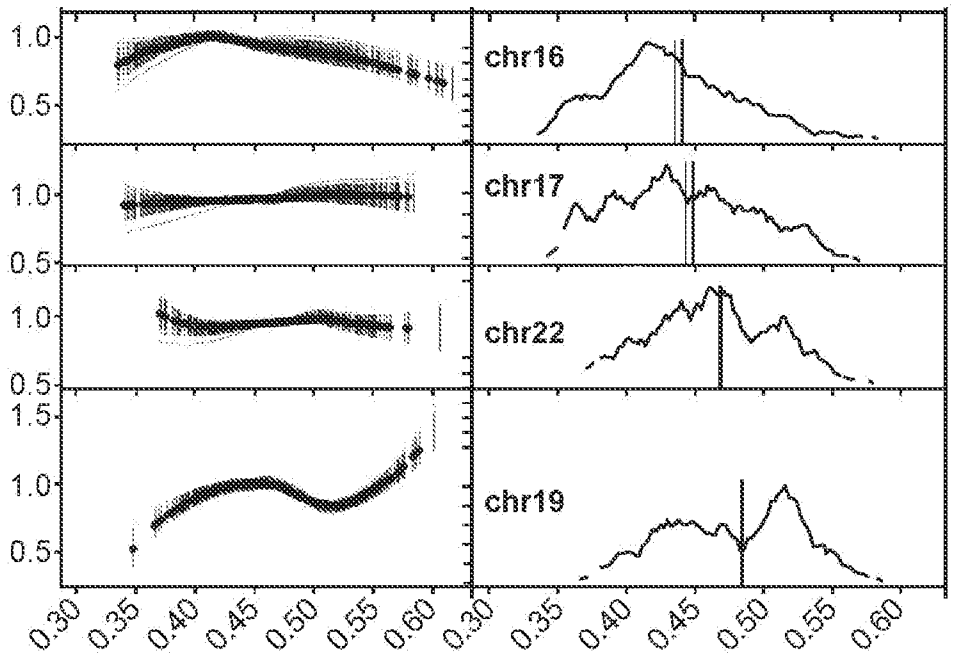


FIG. 75F

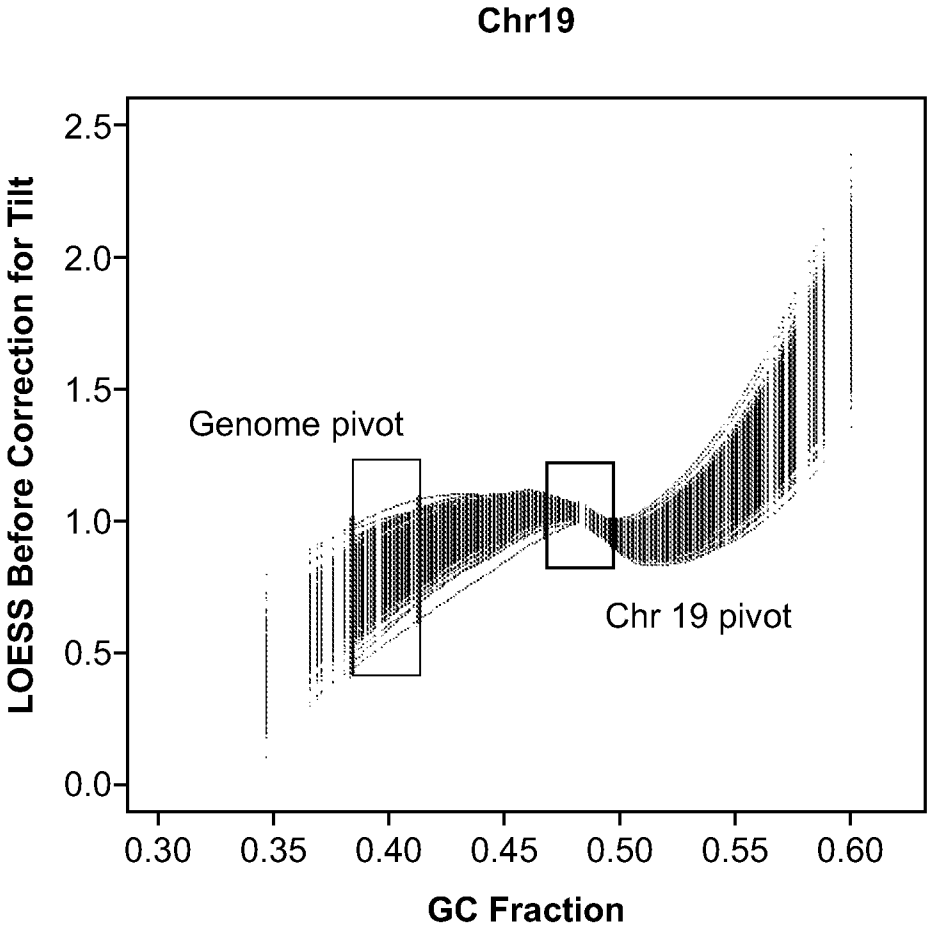

Chr19
LOESS Before Correction for Tilt
2.5
2.0
1.5
1.0
0.5
0.0
Genome pivot
Chr 19 pivot
0.30
0.35
0.40
0.45
0.50
0.55
0.60
GC Fraction

FIG. 76

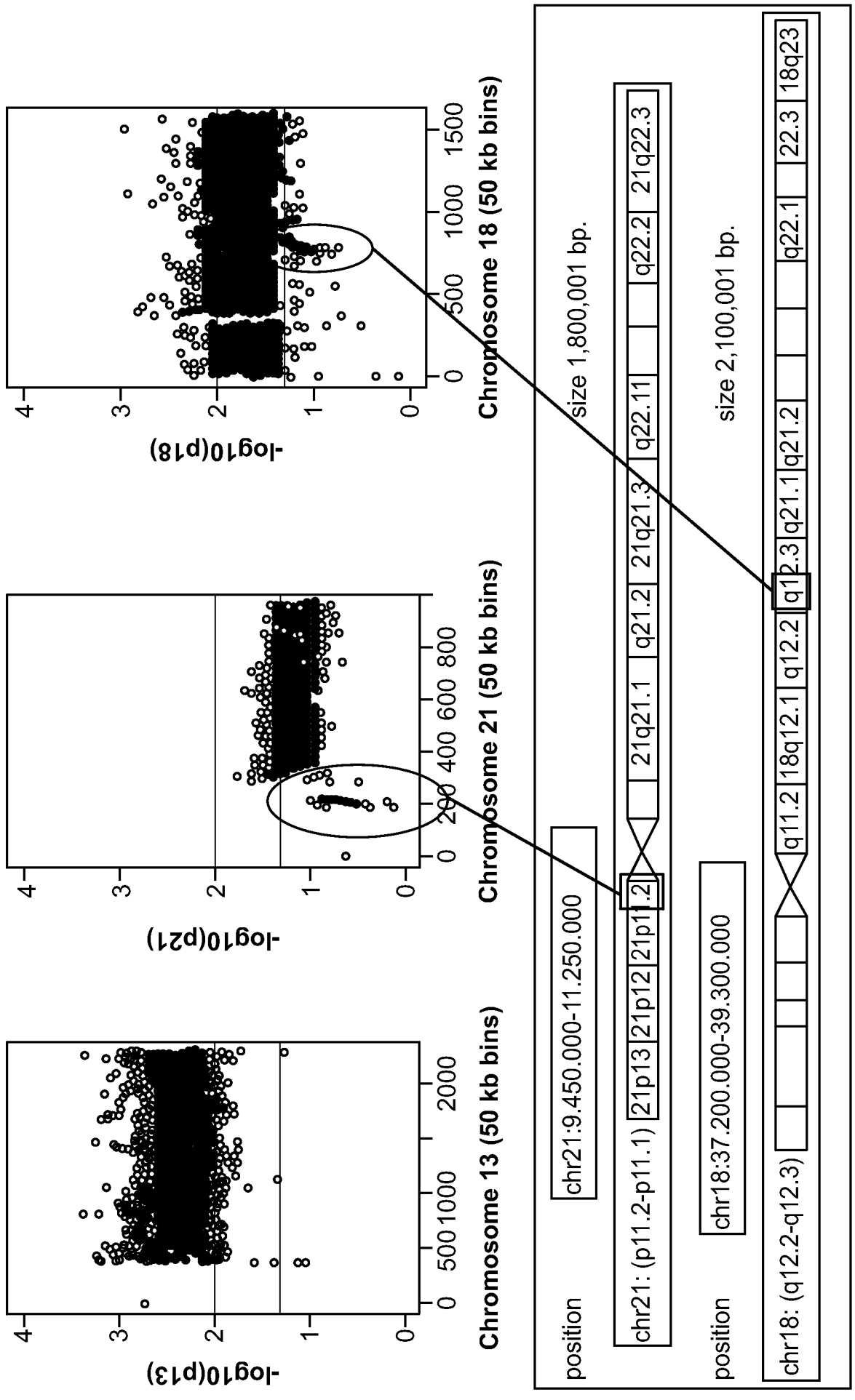

-log10(p13)
Chromosome 13 (50 kb bins)
-log10(p21)
Chromosome 21 (50 kb bins)
-log10(p18)
Chromosome 18 (50 kb bins)
position
chr21:9.450.000-11.250.000
size 1,800,001 bp.
chr21: (p11.2-p11.1)
21p13
21p12
21p11.2
21q21.1
21q21.2
21q21.3
q22.11
q22.2
21q22.3
position
chr18:37.200.000-39.300.000
size 2,100,001 bp.
chr18: (q12.2-q12.3)
q11.2
18q12.1
q12.2
q12.3
q21.1
q21.2
q22.1
22.3
18q23

FIG. 77

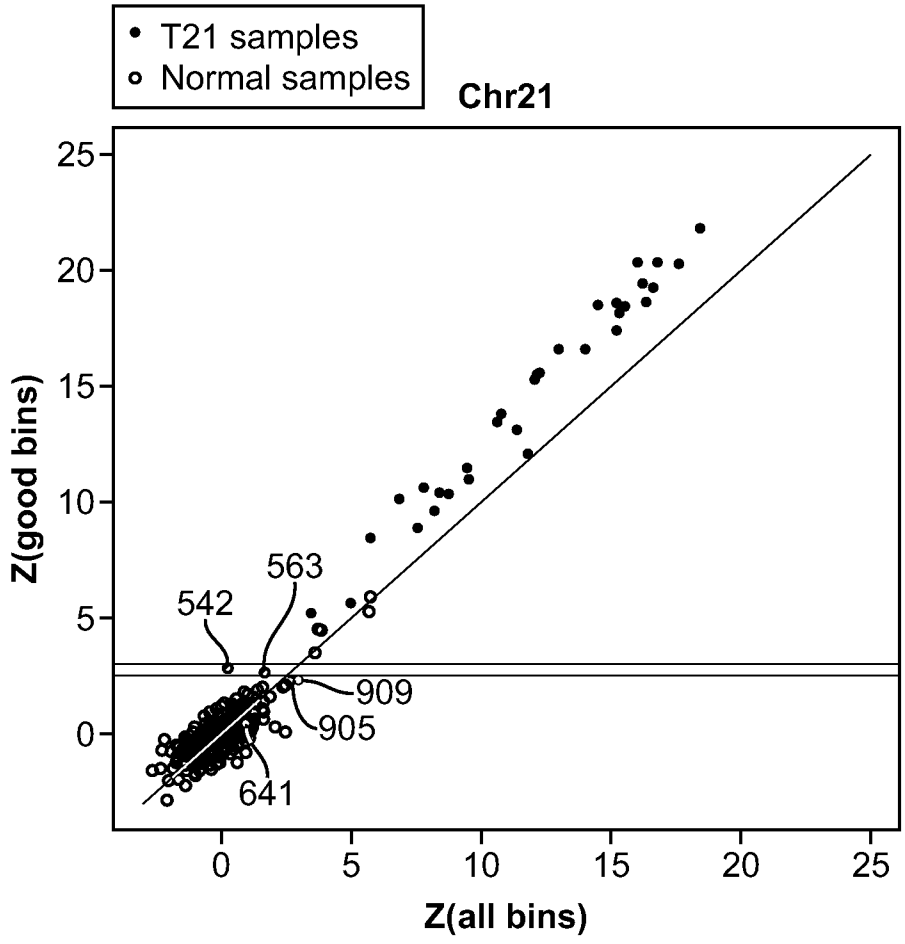

T21 samples
Normal samples
Chr21
Z(good bins)
Z(all bins)
25
20
15
10
5
0
0
5
10
15
20
25
563
542
909
905
641

FIG. 78

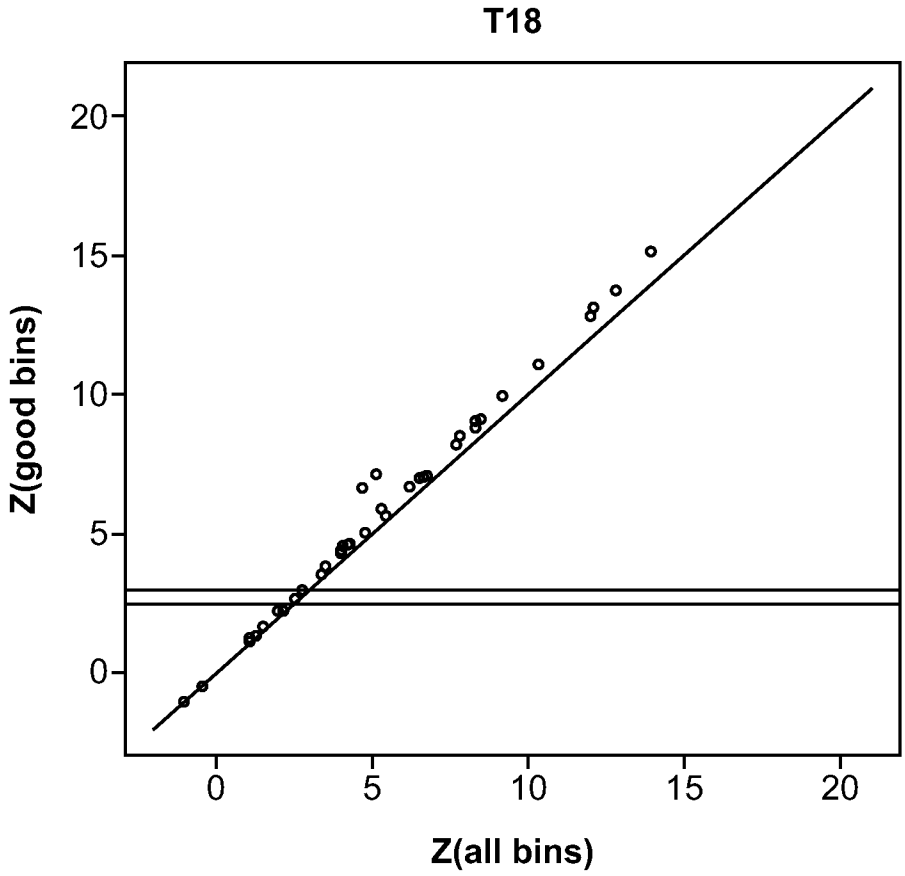


FIG. 79

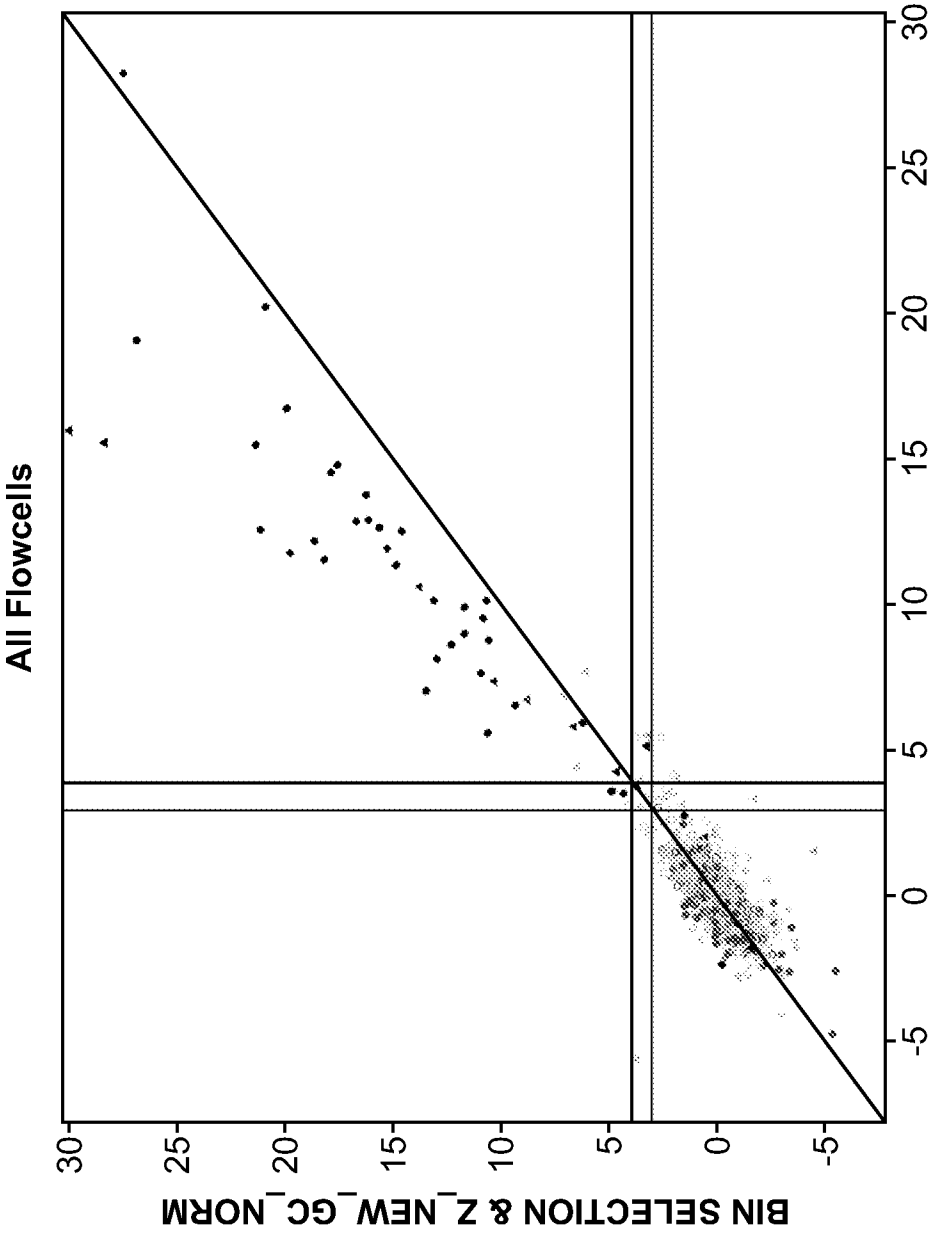

All Flowcells
BIN SELECTION & Z_NEW_GC_NORM
Z_C_18
30
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25
30

FIG. 80

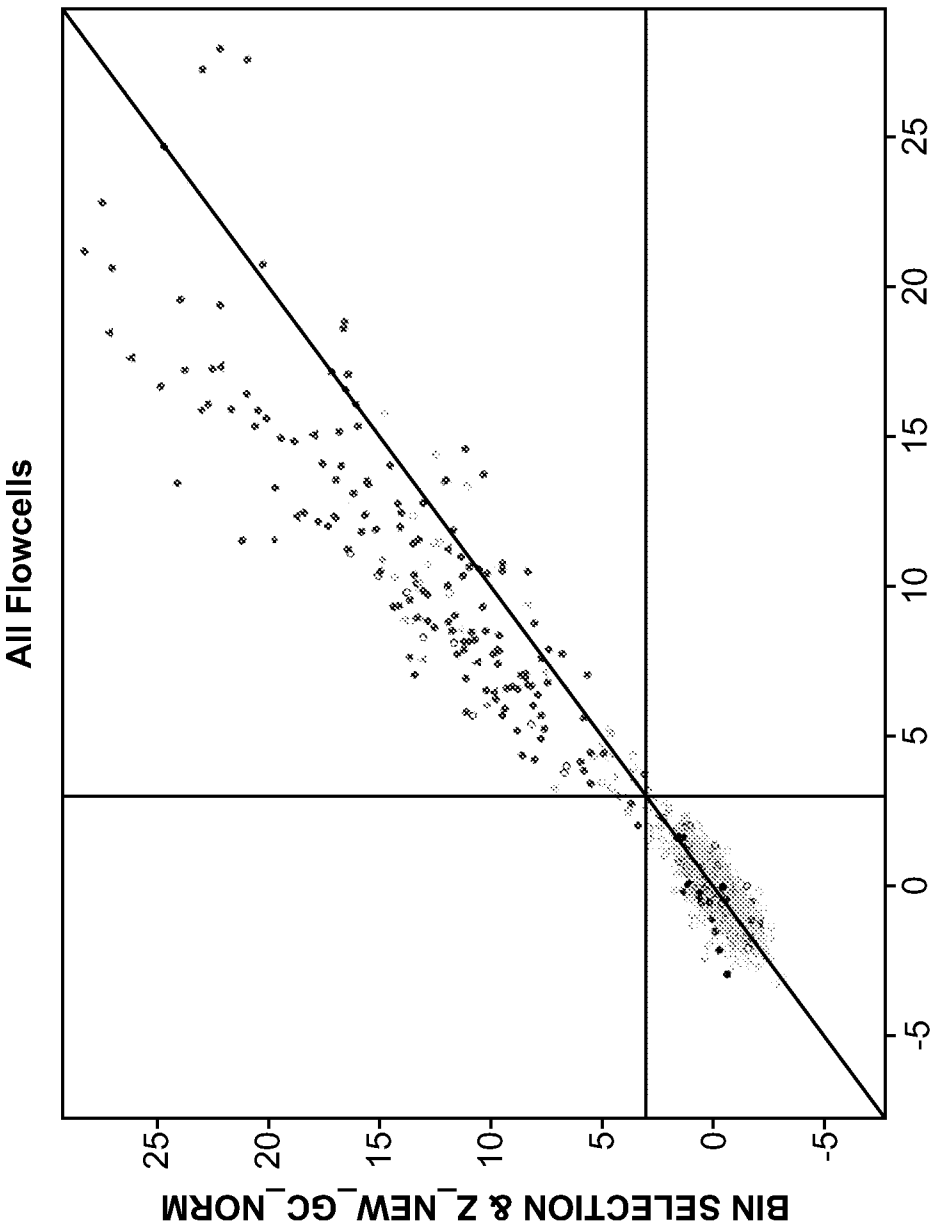

All Flowcells
BIN SELECTION & Z_NEW_GC_NORM
Z_C_21
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25

FIG. 81

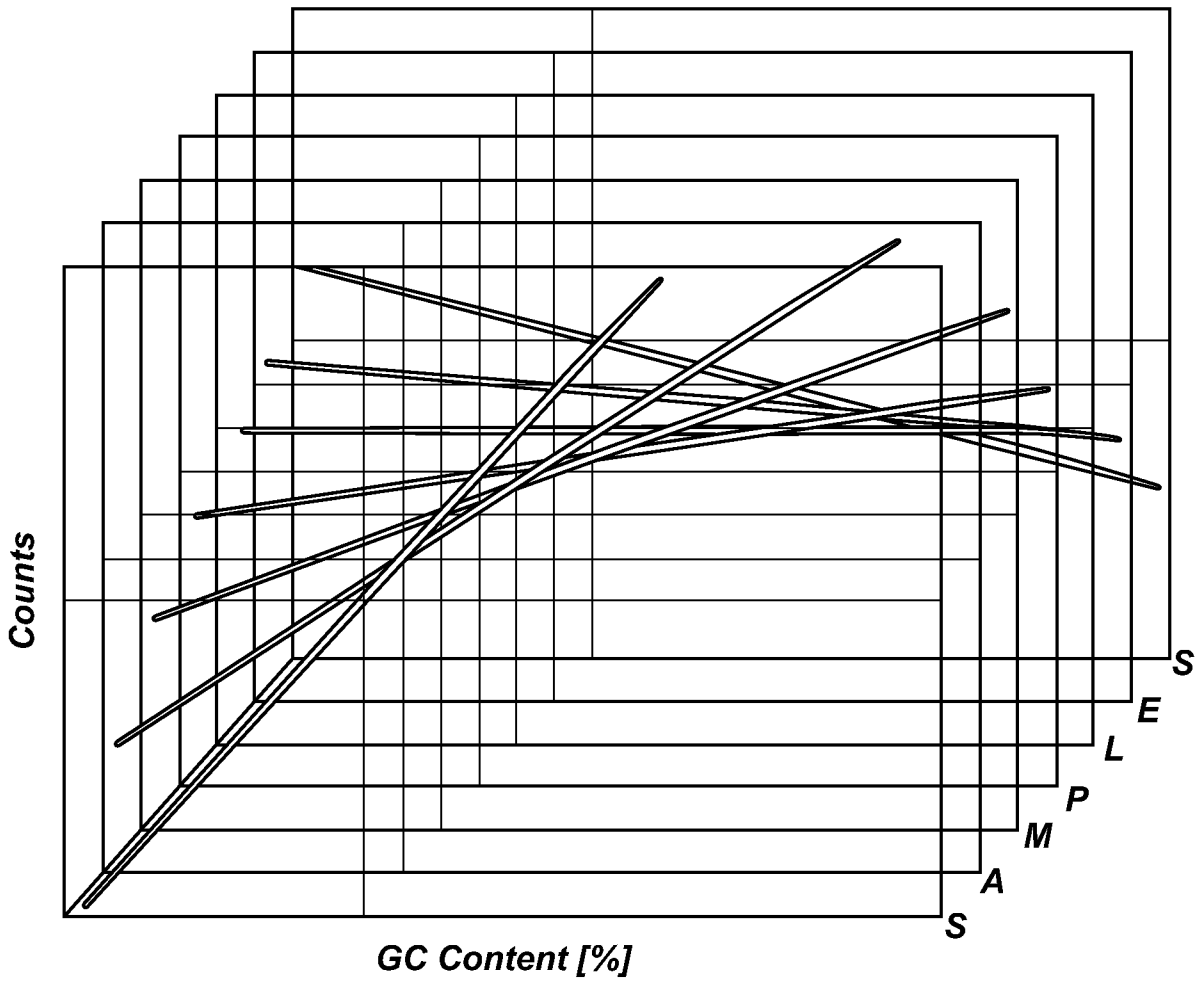


FIG. 82

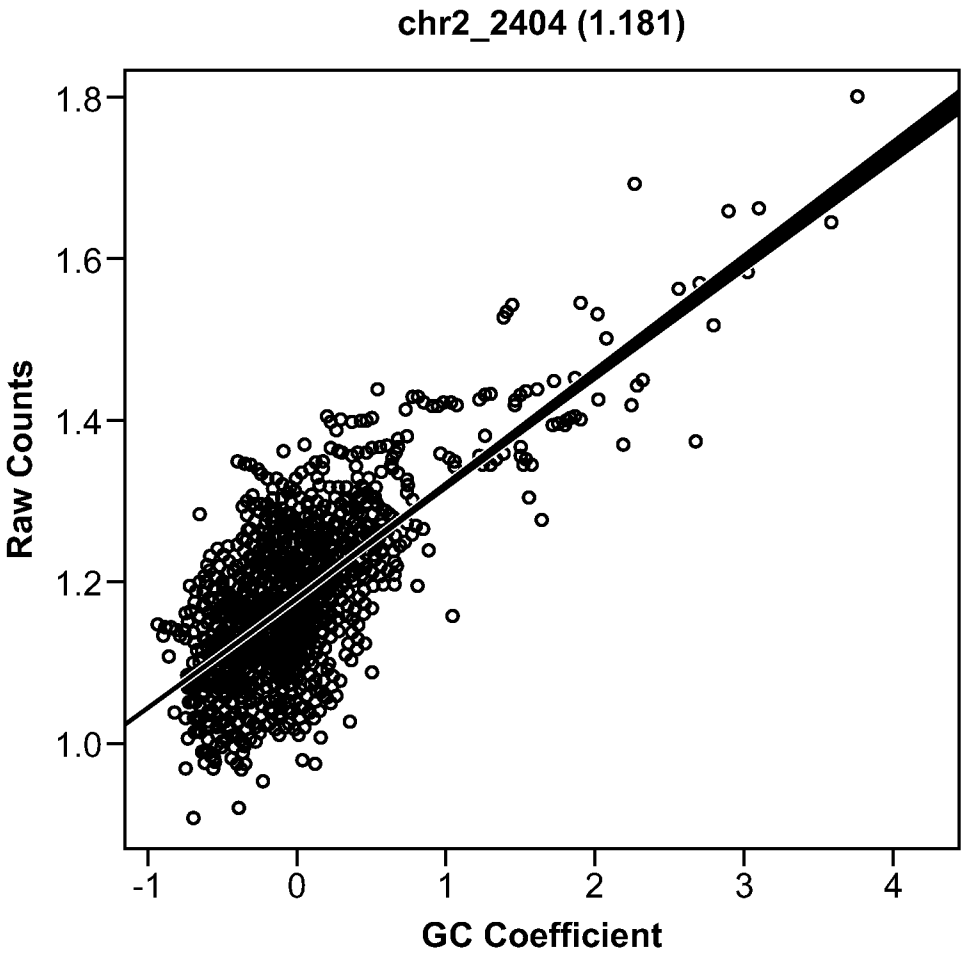

chr2_2404 (1.181)
Raw Counts
1.8
1.6
1.4
1.2
1.0
-1
0
1
2
3
4
GC Coefficient

FIG. 83

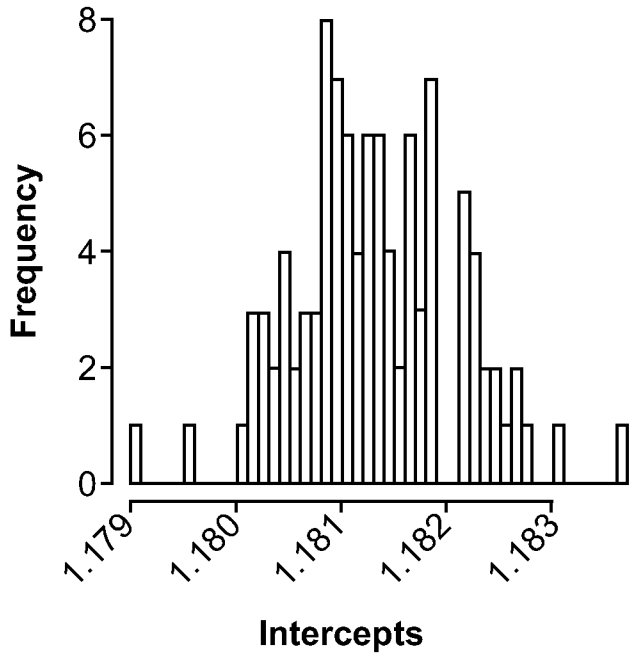

Frequency
0
2
4
6
8
1.179
1.180
1.181
1.182
1.183
Intercepts

FIG. 84

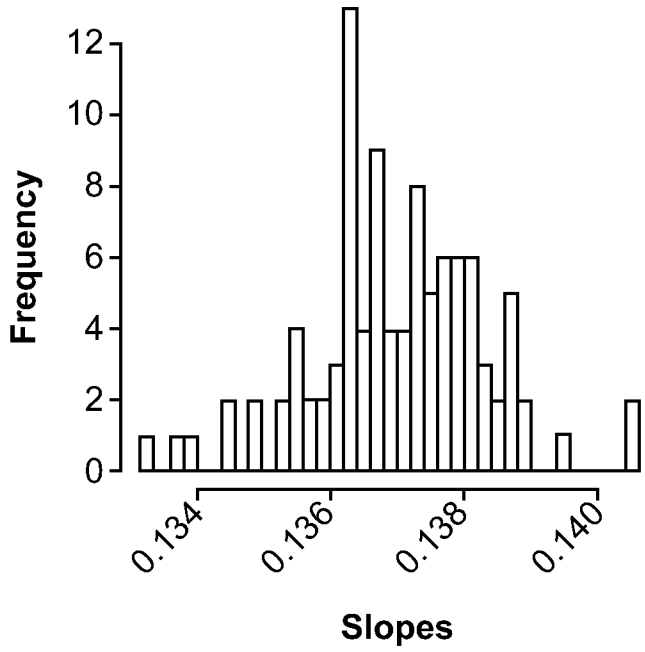

Frequency
0
2
4
6
8
10
12
0.134
0.136
0.138
0.140
Slopes

FIG. 85

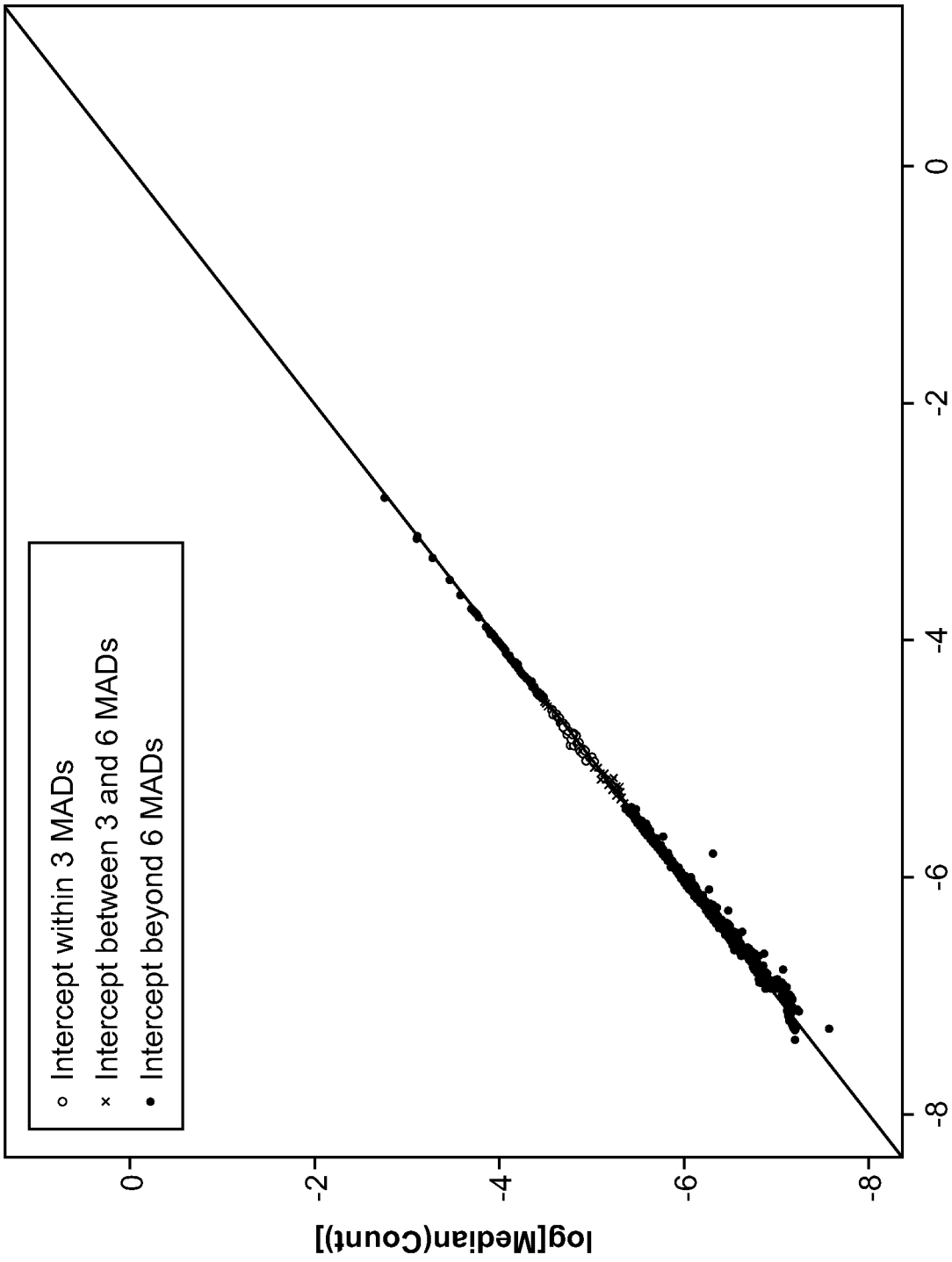

log[Median(Count)]
0
-2
-4
-6
-8
log(intercept)
0
-2
-4
-6
-8
Intercept within 3 MADs
Intercept between 3 and 6 MADs
Intercept beyond 6 MADs

FIG. 86

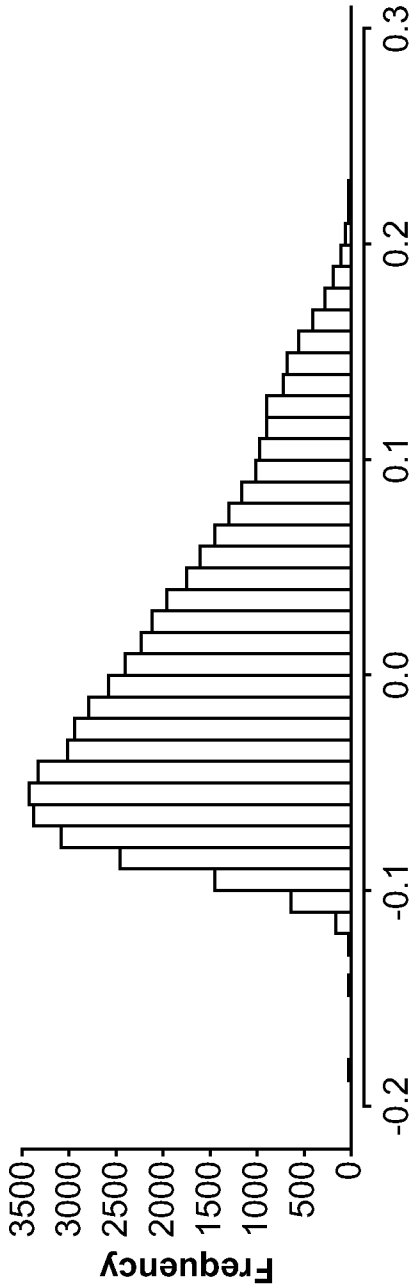

Frequency
3500
3000
2500
2000
1500
1000
500
0
-0.2
-0.1
0.0
0.1
0.2
0.3
Slope

FIG. 87

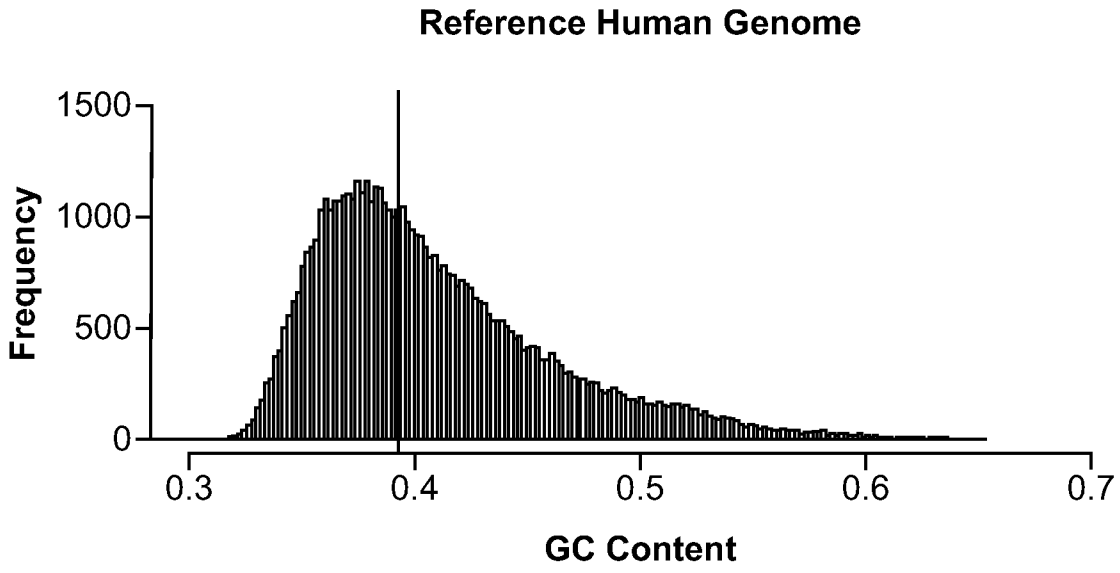

Reference Human Genome
Frequency
1500
1000
500
0
0.3
0.4
0.5
0.6
0.7
GC Content

FIG. 88

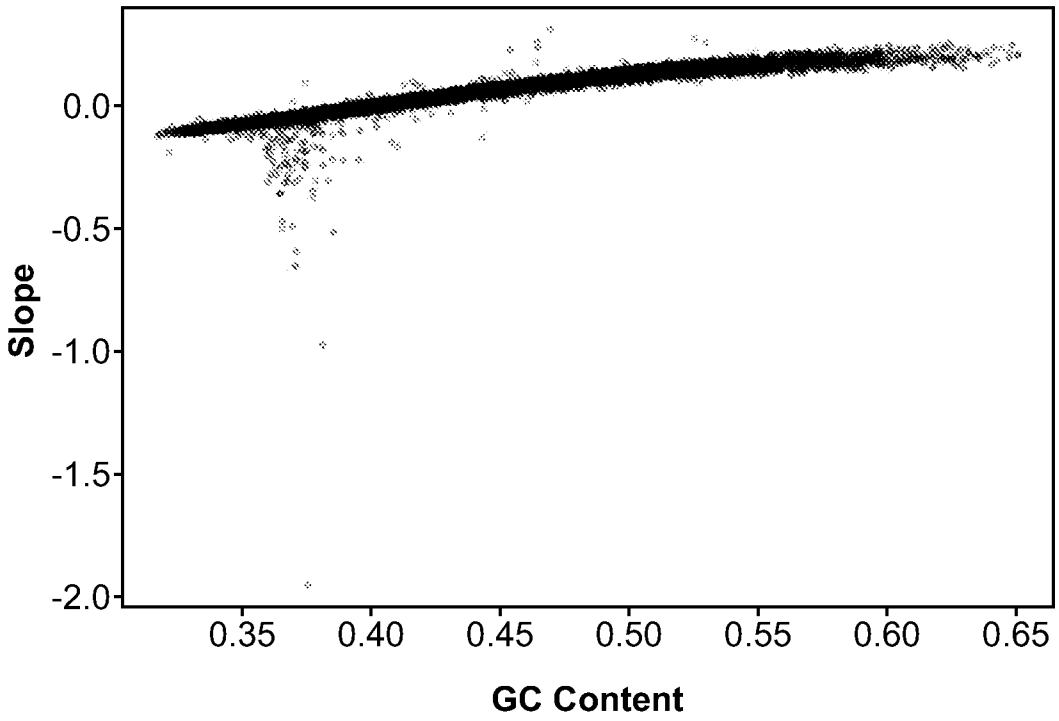


FIG. 89

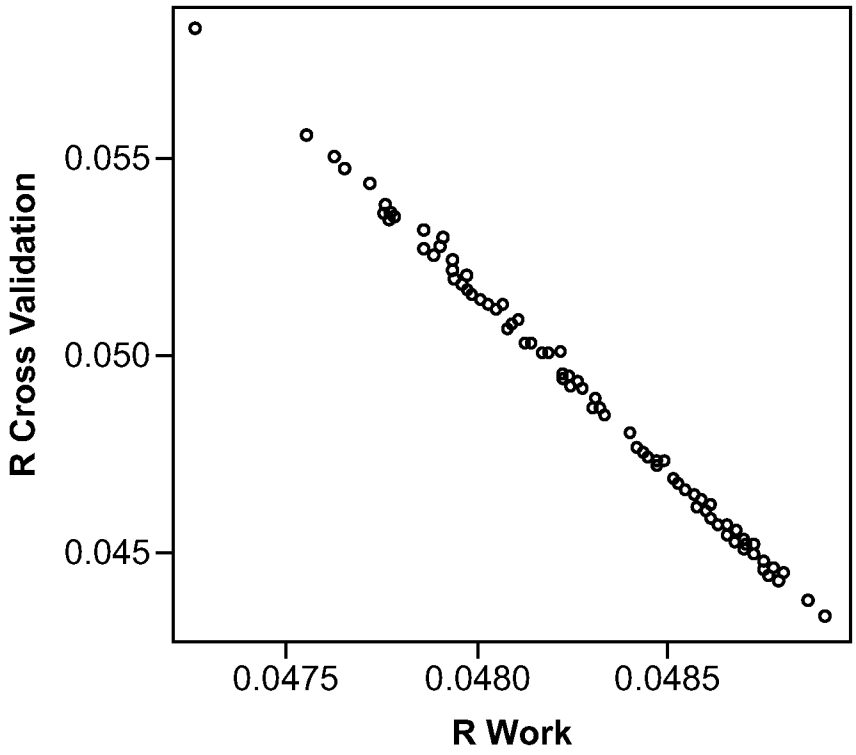


FIG. 90

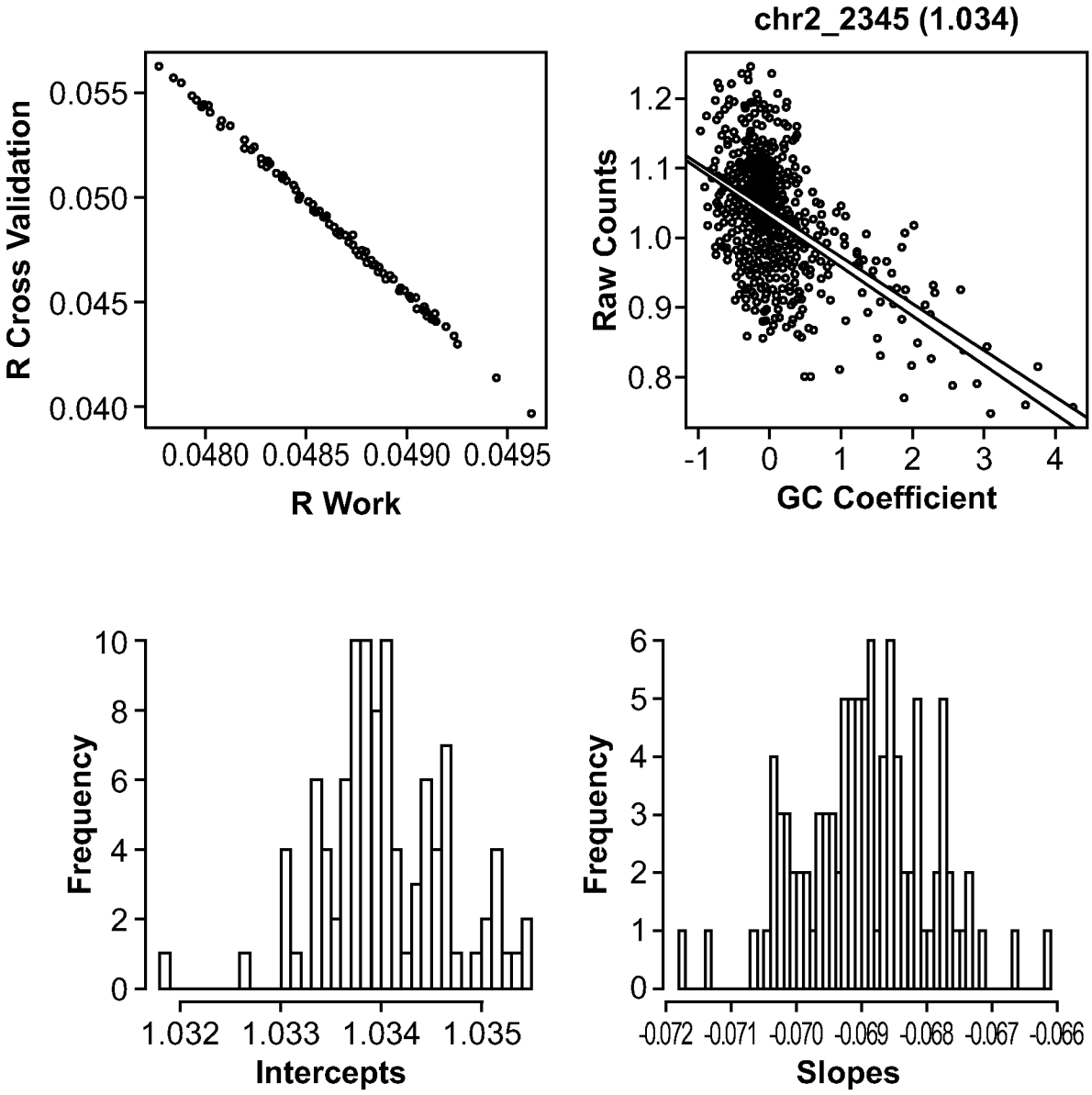

R Cross Validation
0.055
0.050
0.045
0.040
0.0480 0.0485 0.0490 0.0495
R Work
chr2_2345 (1.034)
Raw Counts
1.2
1.1
1.0
0.9
0.8
-1 0 1 2 3 4
GC Coefficient
Frequency
10
8
6
4
2
0
1.032 1.033 1.034 1.035
Intercepts
Frequency
6
5
4
3
2
1
0
-0.072 -0.071 -0.070 -0.069 -0.068 -0.067 -0.066
Slopes

FIG. 91

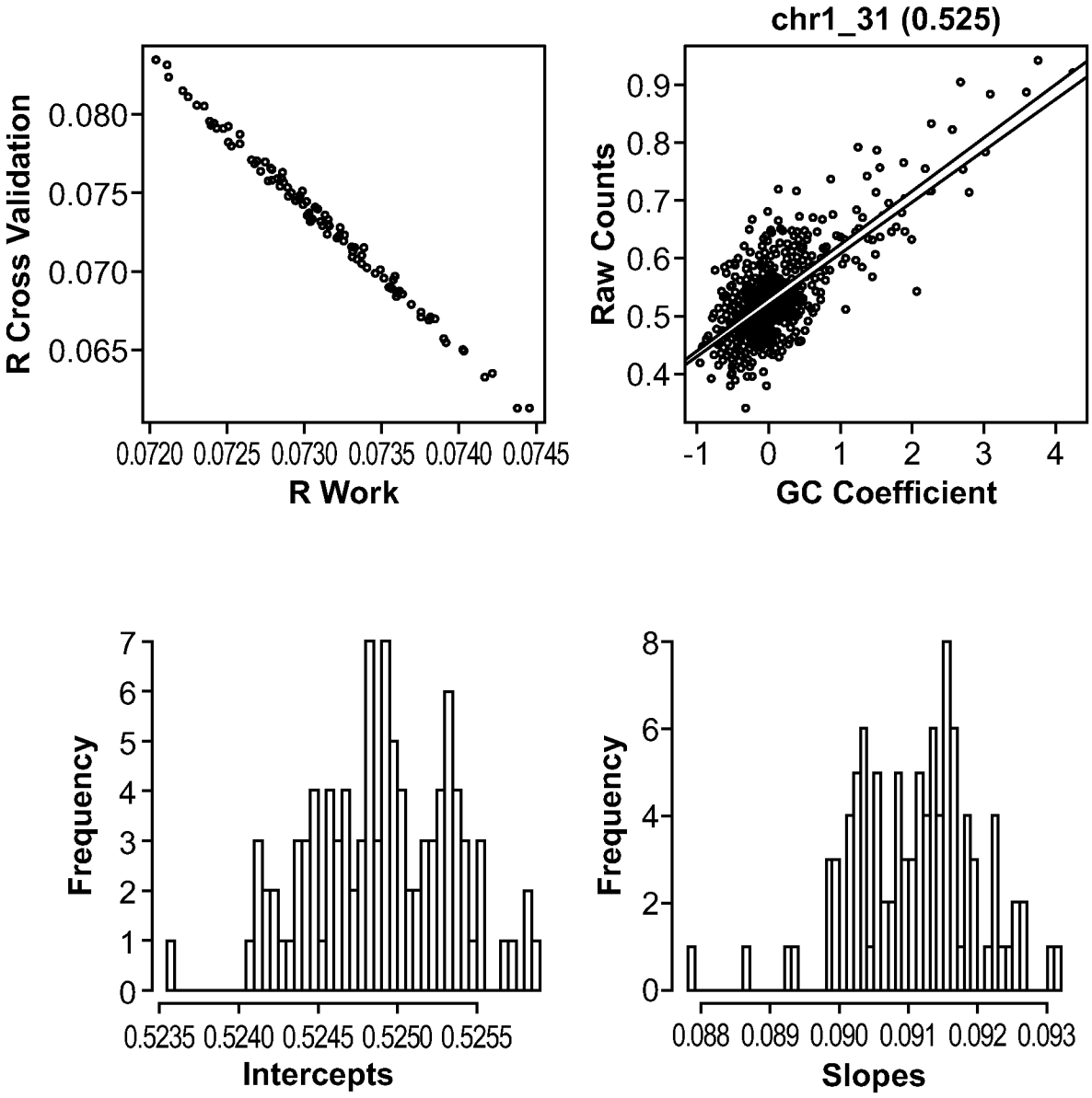

R Cross Validation
0.080
0.075
0.070
0.065
0.0720 0.0725 0.0730 0.0735 0.0740 0.0745
R Work
chr1_31 (0.525)
Raw Counts
0.9
0.8
0.7
0.6
0.5
0.4
-1 0 1 2 3 4
GC Coefficient
Frequency
7
6
5
4
3
2
1
0
0.5235 0.5240 0.5245 0.5250 0.5255
Intercepts
Frequency
8
6
4
2
0
0.088 0.089 0.090 0.091 0.092 0.093
Slopes

FIG. 92

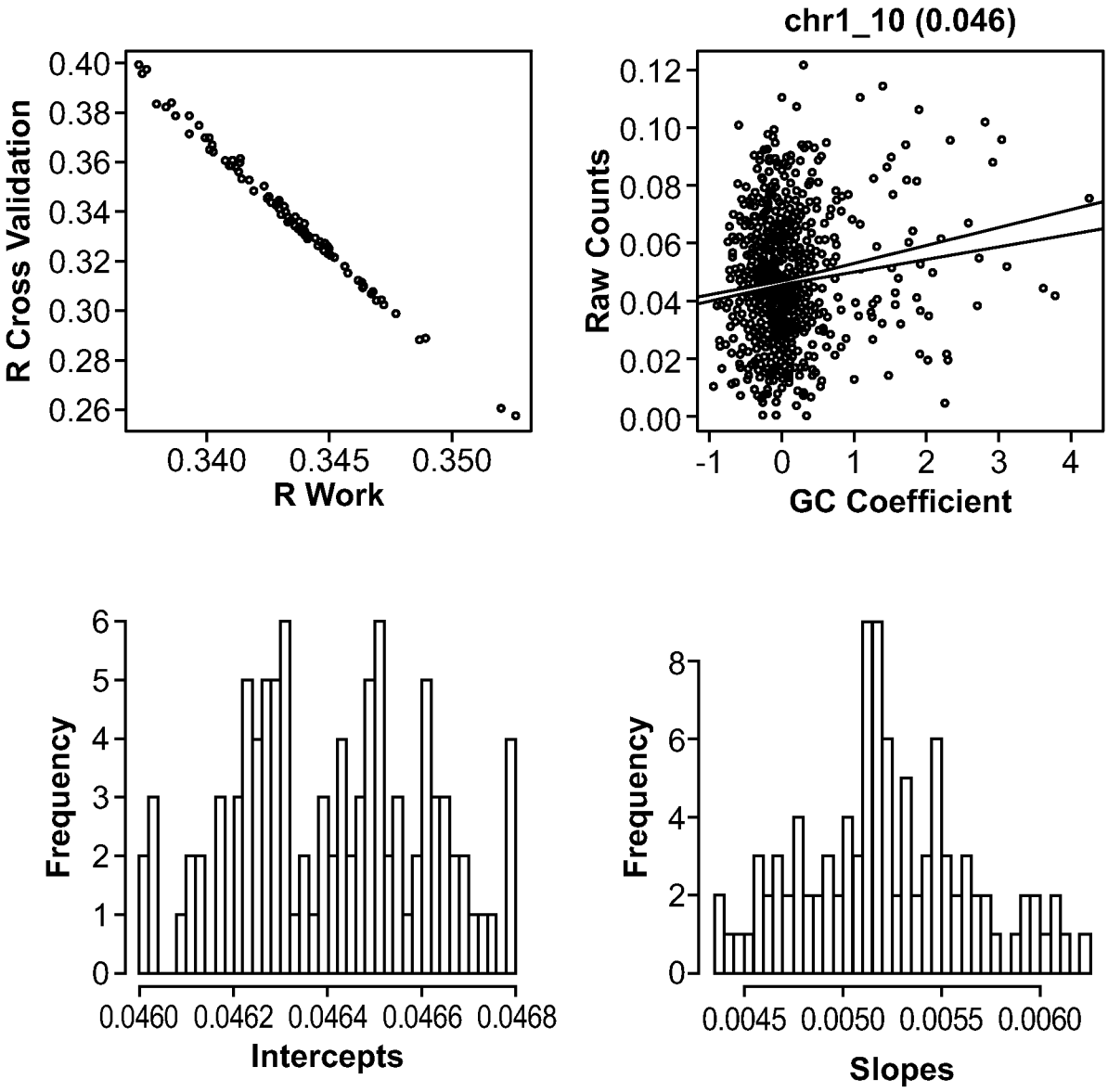

R Cross Validation
0.40
0.38
0.36
0.34
0.32
0.30
0.28
0.26
0.340
0.345
0.350
R Work
chr1_10 (0.046)
Raw Counts
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-1
0
1
2
3
4
GC Coefficient
Frequency
6
5
4
3
2
1
0
0.0460
0.0462
0.0464
0.0466
0.0468
Intercepts
Frequency
8
6
4
2
0
0.0045
0.0050
0.0055
0.0060
Slopes

FIG. 93

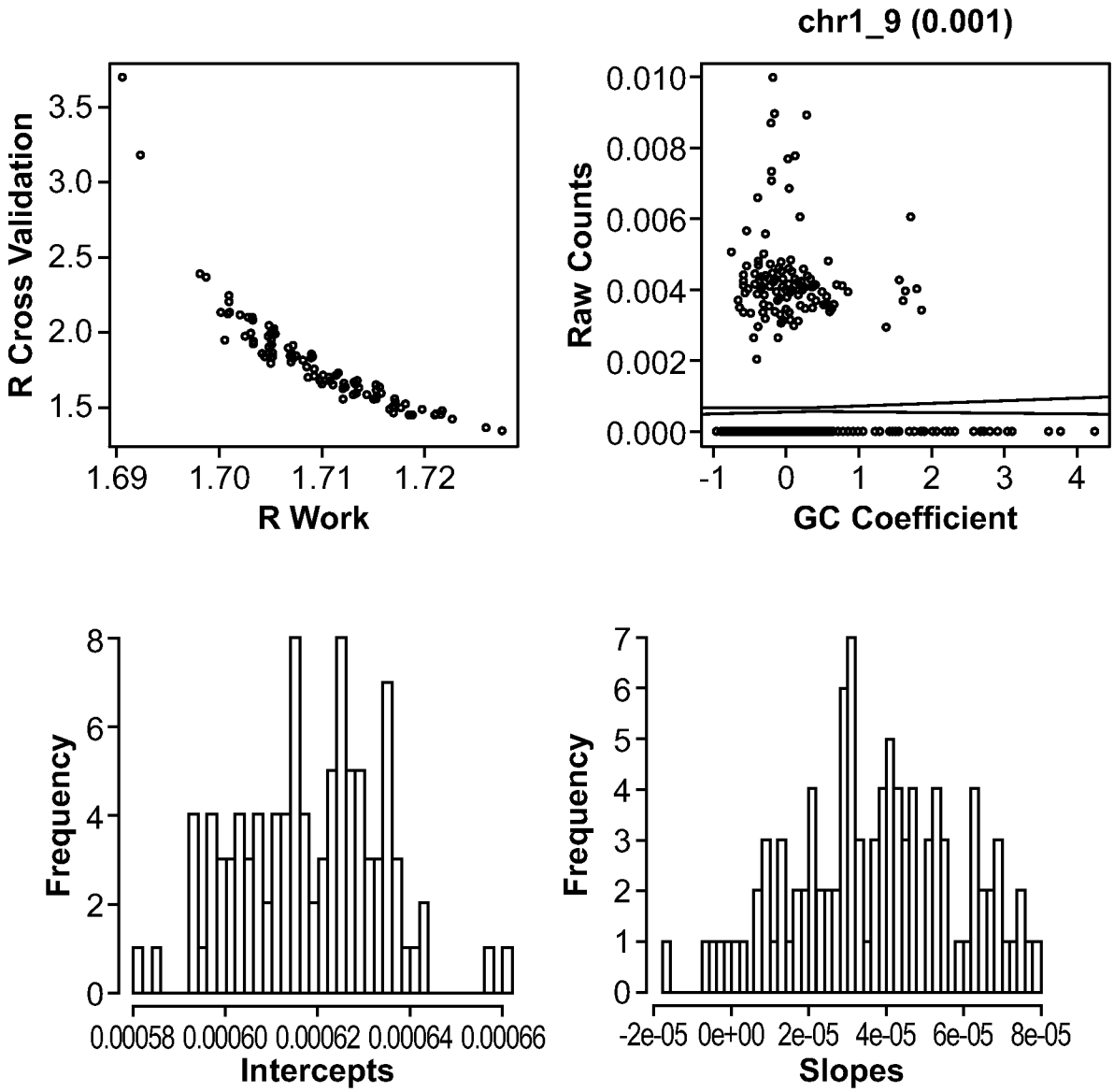

chr1_9 (0.001)
R Cross Validation
3.5
3.0
2.5
2.0
1.5
1.69
1.70
1.71
1.72
R Work
Raw Counts
0.010
0.008
0.006
0.004
0.002
0.000
-1
0
1
2
3
4
GC Coefficient
Frequency
8
6
4
2
0
0.00058
0.00060
0.00062
0.00064
0.00066
Intercepts
Frequency
7
6
5
4
3
2
1
0
-2e-05
0e+00
2e-05
4e-05
6e-05
8e-05
Slopes

FIG. 94

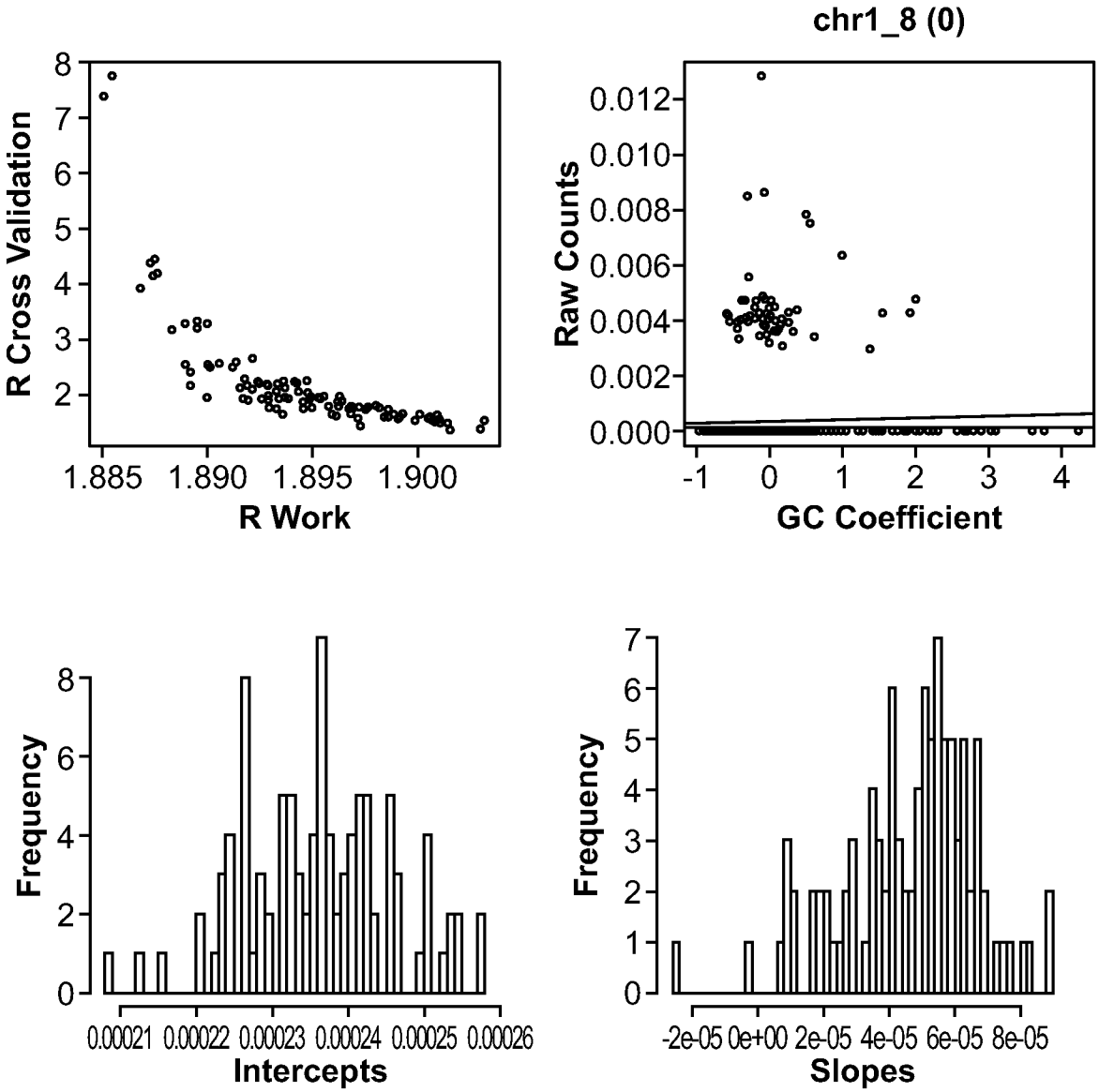

R Cross Validation
R Work
chr1_8 (0)
Raw Counts
GC Coefficient
Frequency
Intercepts
Frequency
Slopes

FIG. 95

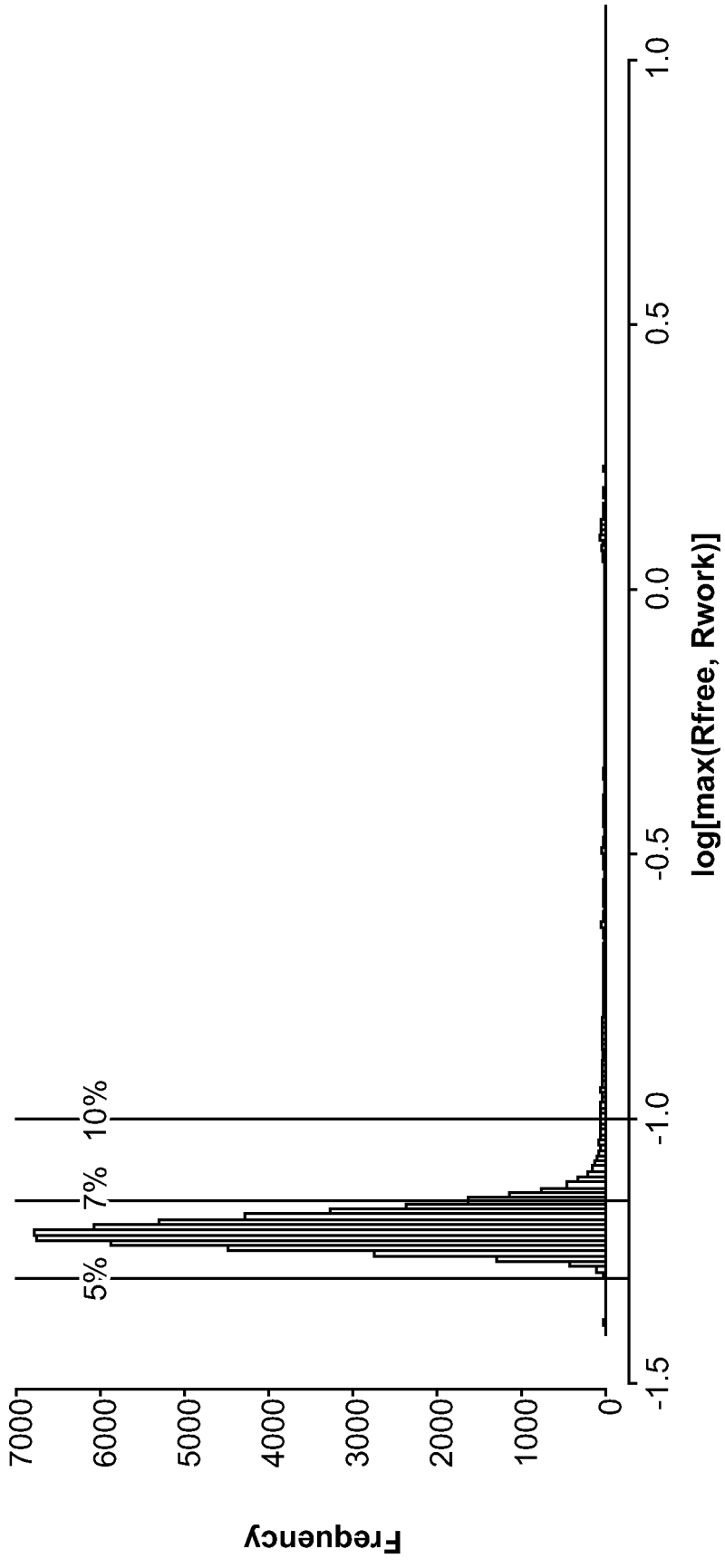

Cross-Validation Error in Bin Parameters
Frequency
7000
6000
5000
4000
3000
2000
1000
0
5%
7%
10%
-1.5
-1.0
-0.5
0.0
0.5
1.0
log[max(Rfree, Rwork)]

FIG. 96

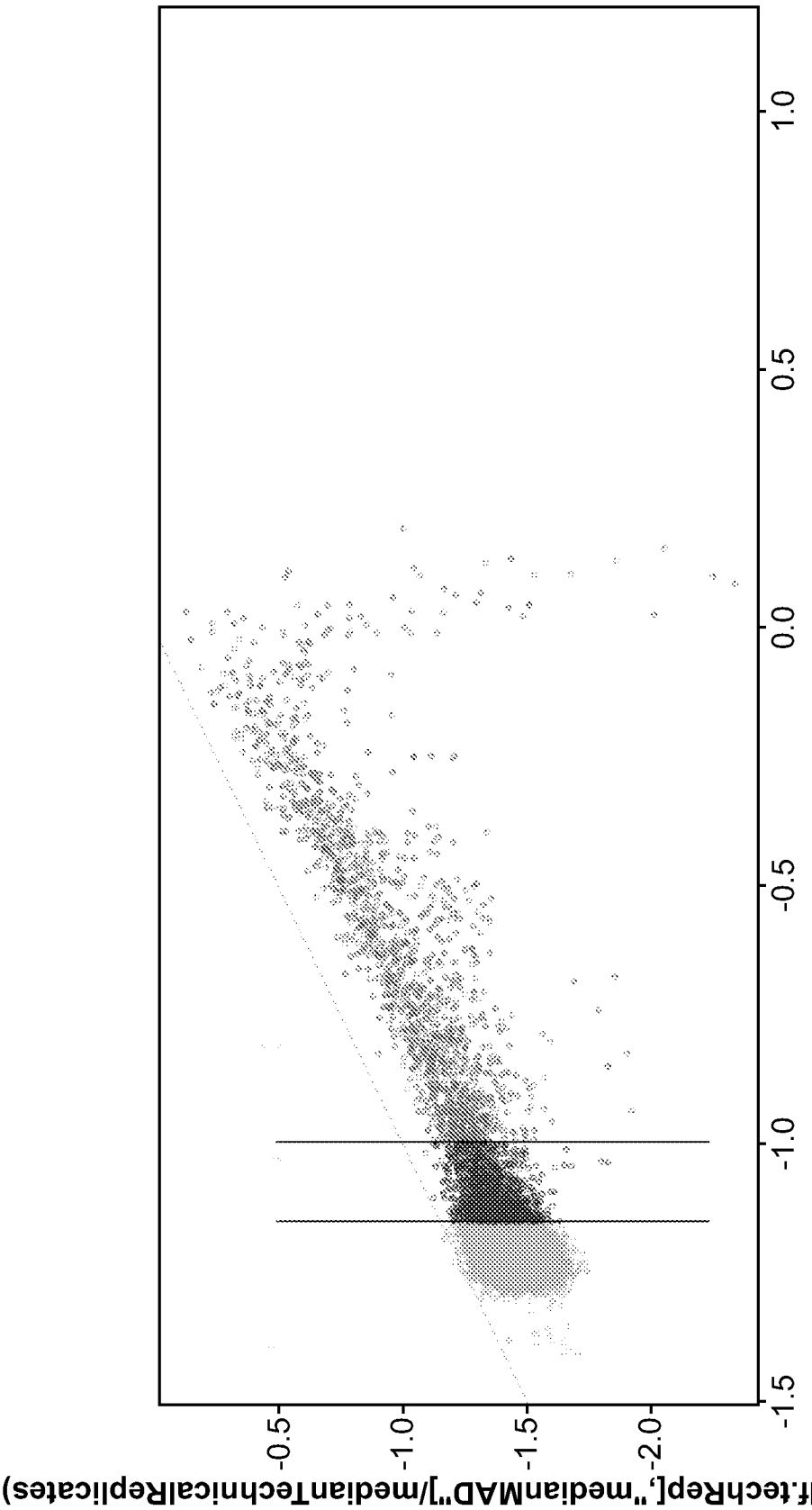

log10(df.techRep[,"medianMAD"]/medianTechnicalReplicates)
-0.5
-1.0
-1.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
log10(max Cross Validation Errors)

FIG. 97

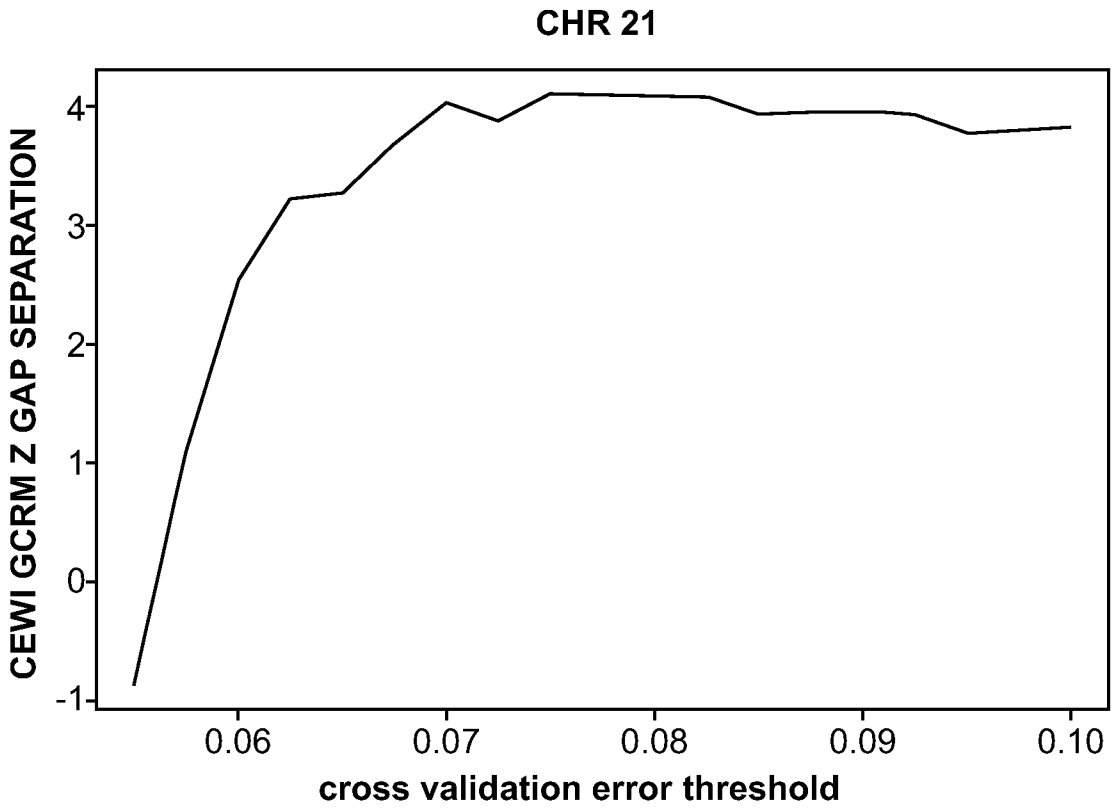

CHR 21
CEWI GCRM Z GAP SEPARATION
4
3
2
1
0
-1
0.06
0.07
0.08
0.09
0.10
cross validation error threshold

FIG. 98

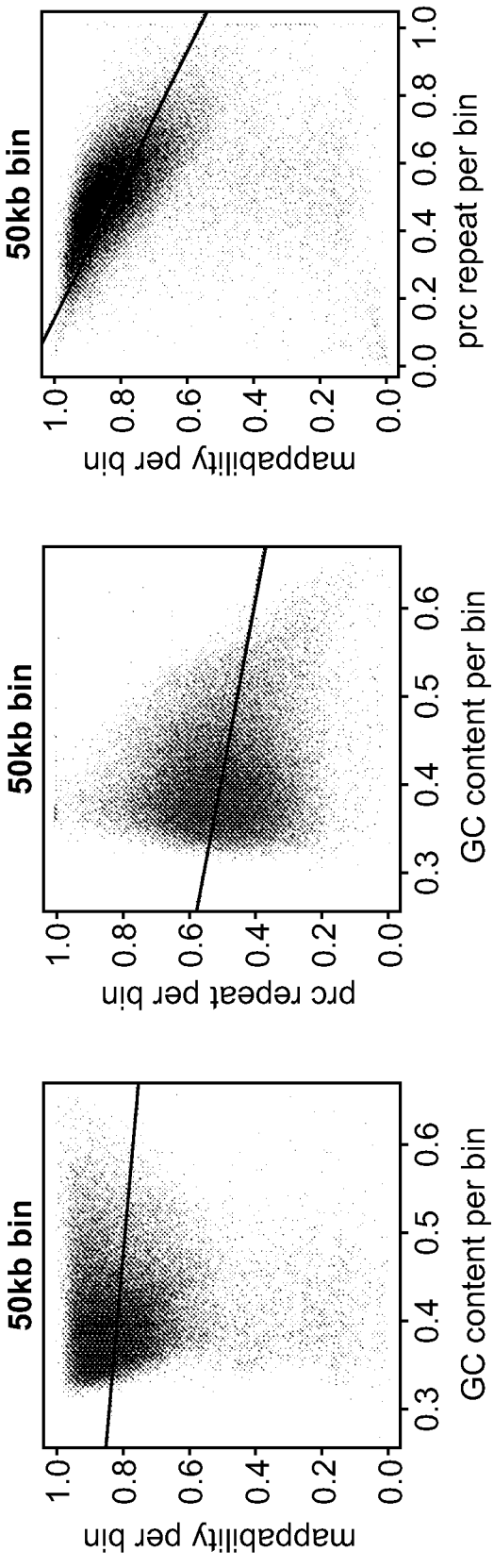


FIG. 99A

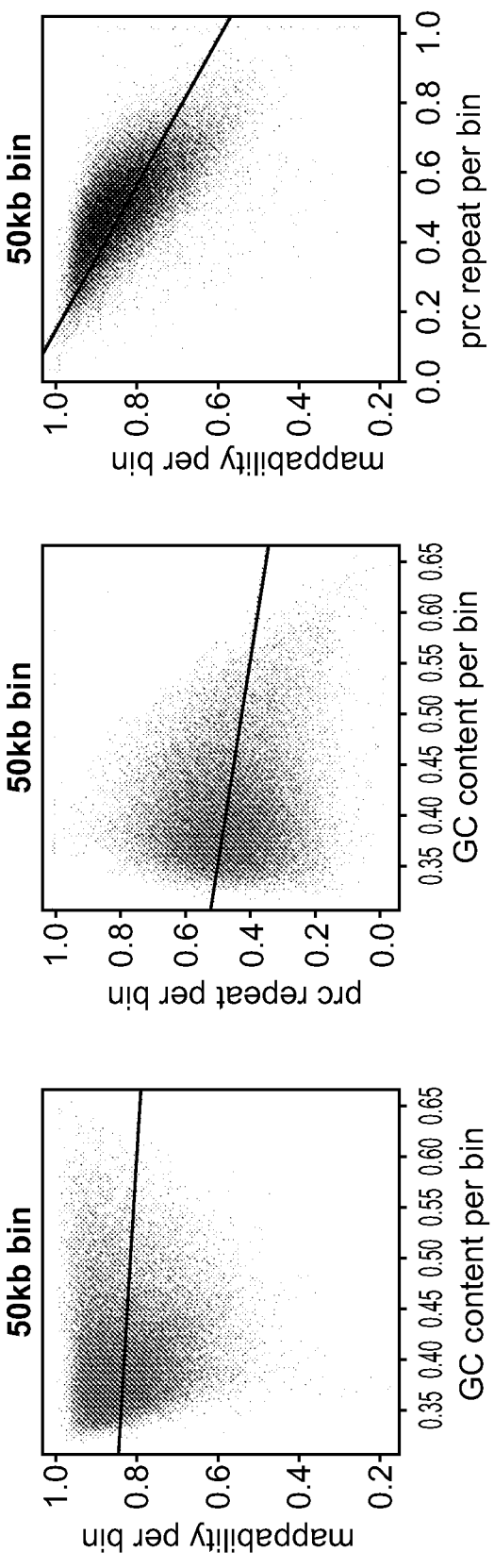


FIG. 99B

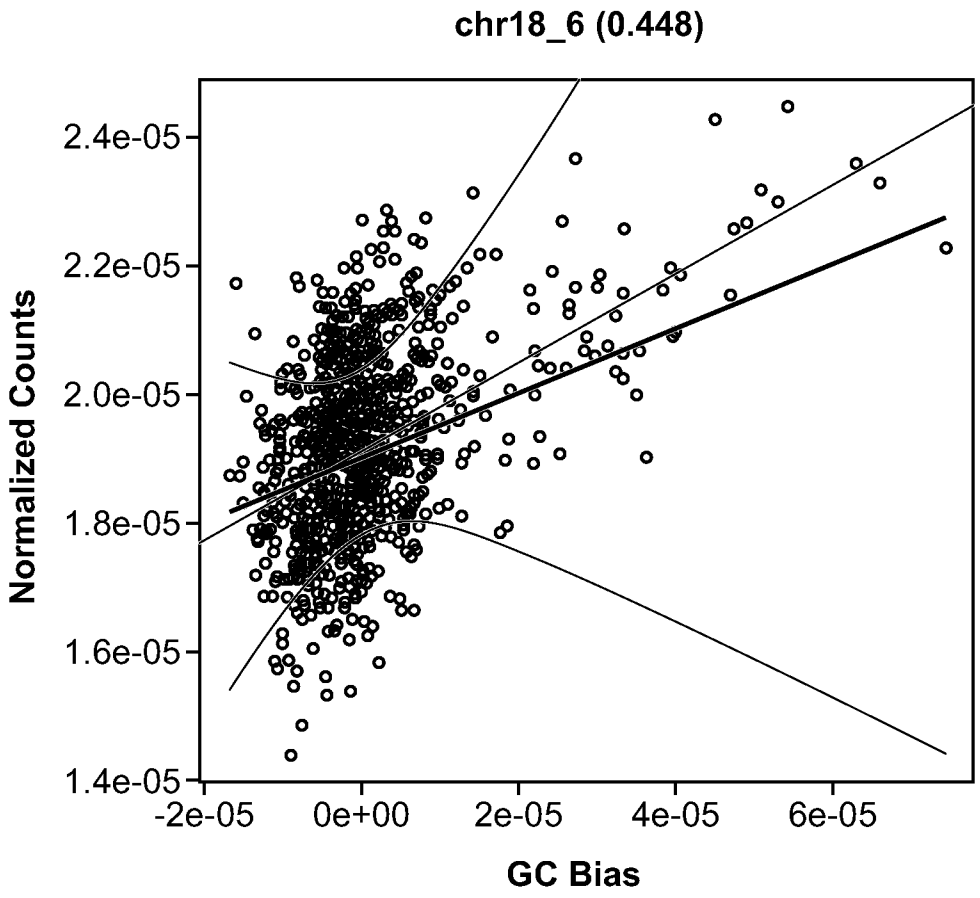

chr18_6 (0.448)
Normalized Counts
2.4e-05
2.2e-05
2.0e-05
1.8e-05
1.6e-05
1.4e-05
-2e-05
0e+00
2e-05
4e-05
6e-05
GC Bias

FIG. 100

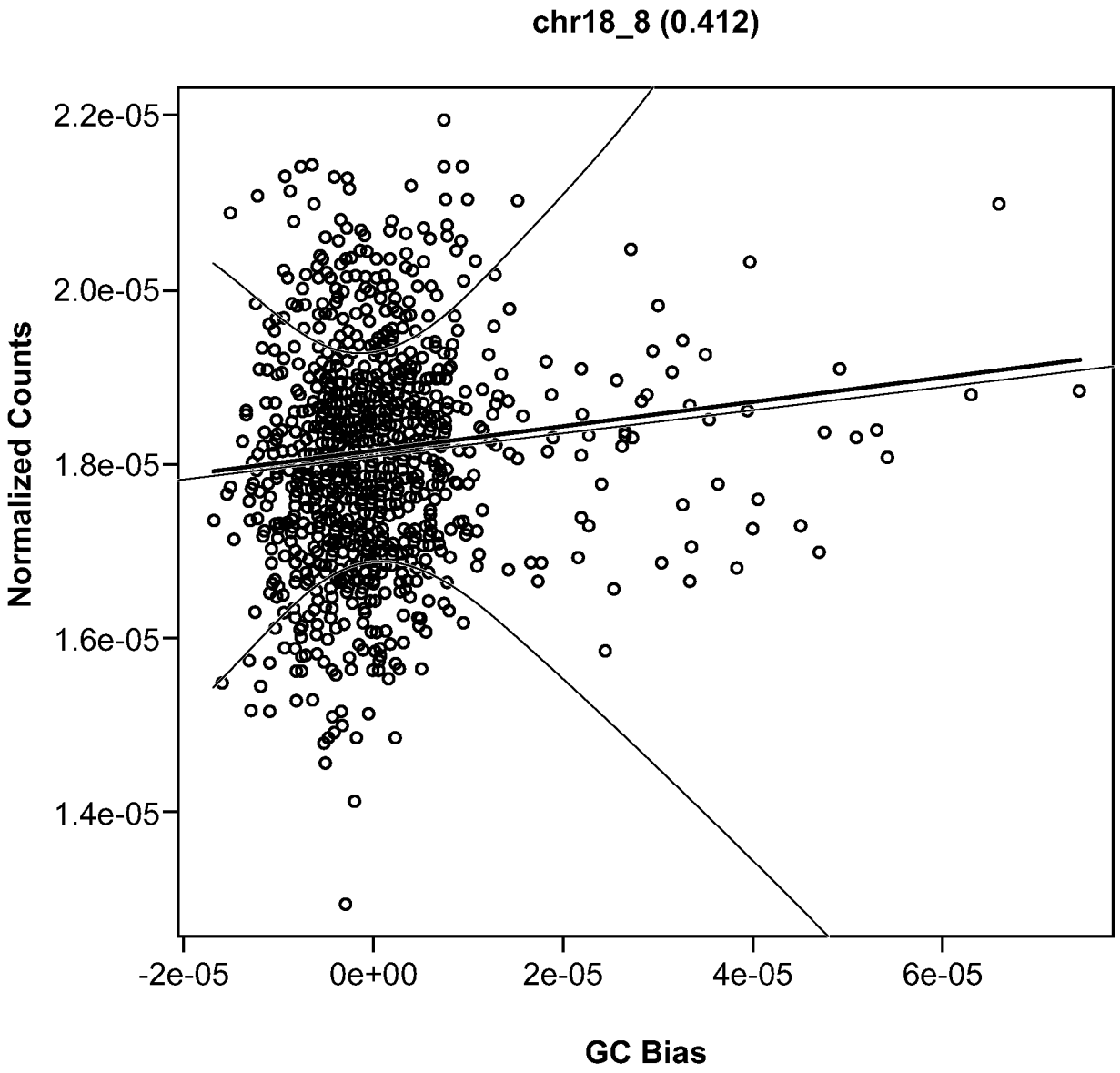

chr18_8 (0.412)
Normalized Counts
2.2e-05
2.0e-05
1.8e-05
1.6e-05
1.4e-05
-2e-05
0e+00
2e-05
4e-05
6e-05
GC Bias

FIG. 101

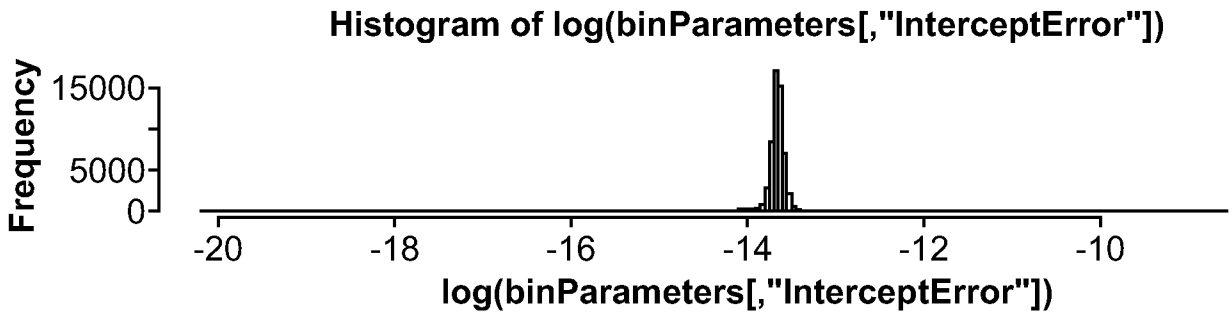

Histogram of log(binParameters[,"InterceptError"])
Frequency
15000
5000
0
-20
-18
-16
-14
-12
-10
log(binParameters[,"InterceptError"])

FIG. 102

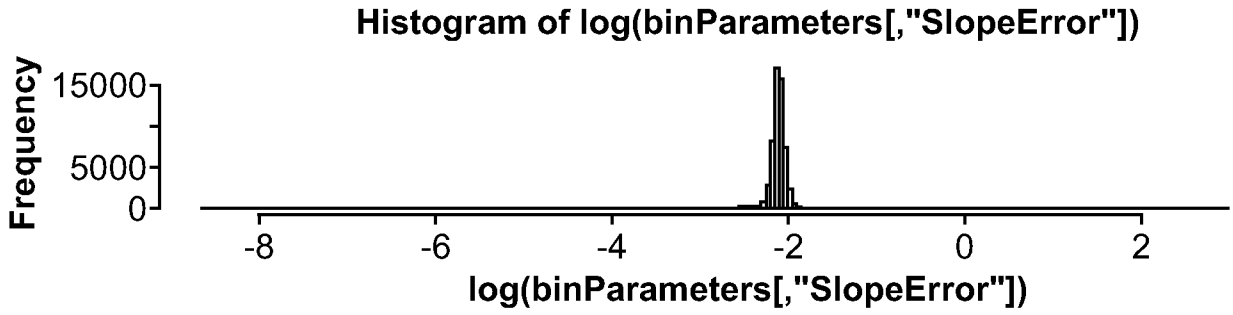

Histogram of log(binParameters[,"SlopeError"])
Frequency
15000
5000
0
-8
-6
-4
-2
0
2
log(binParameters[,"SlopeError"])

FIG. 103

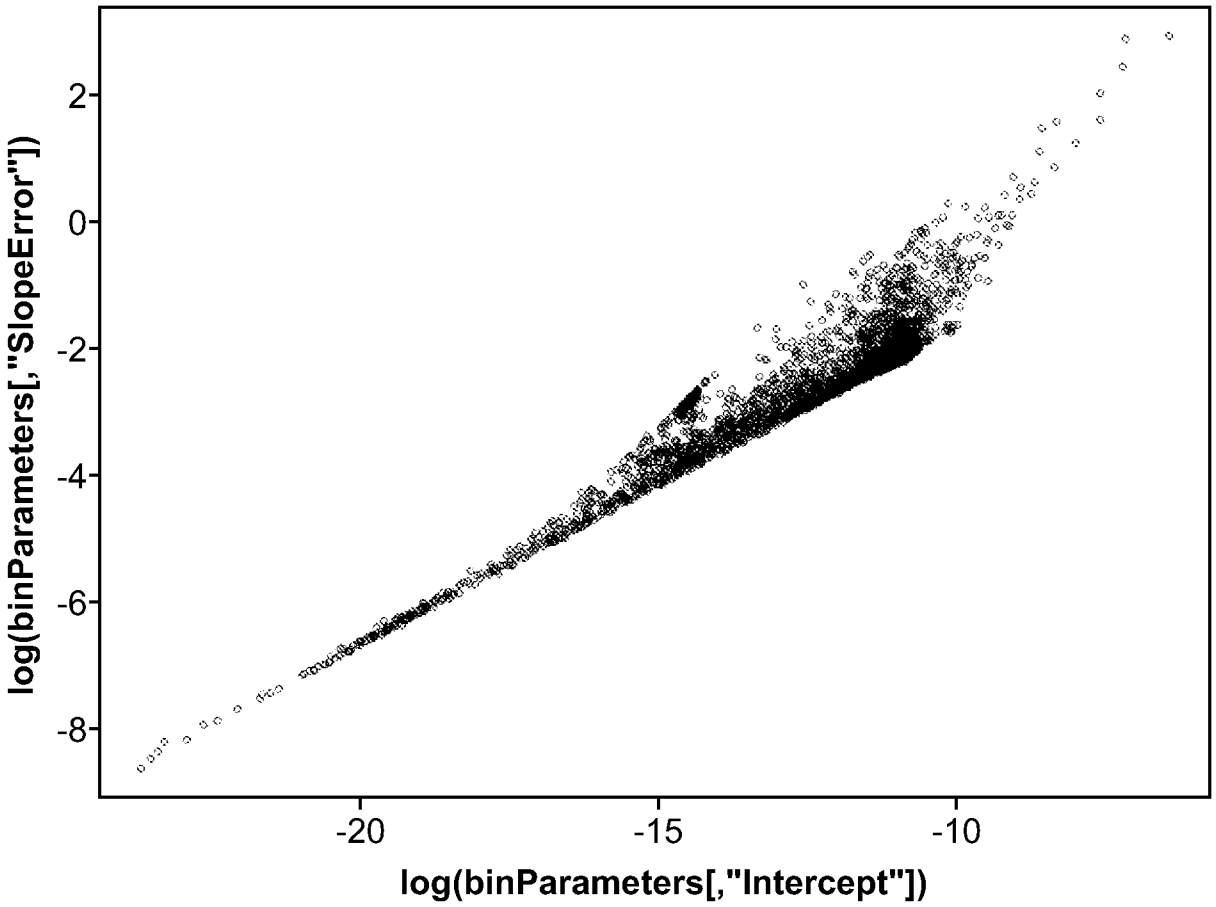

log(binParameters[,"SlopeError"])
2
0
-2
-4
-6
-8
-20
-15
-10
log(binParameters[,"Intercept"])

FIG. 104

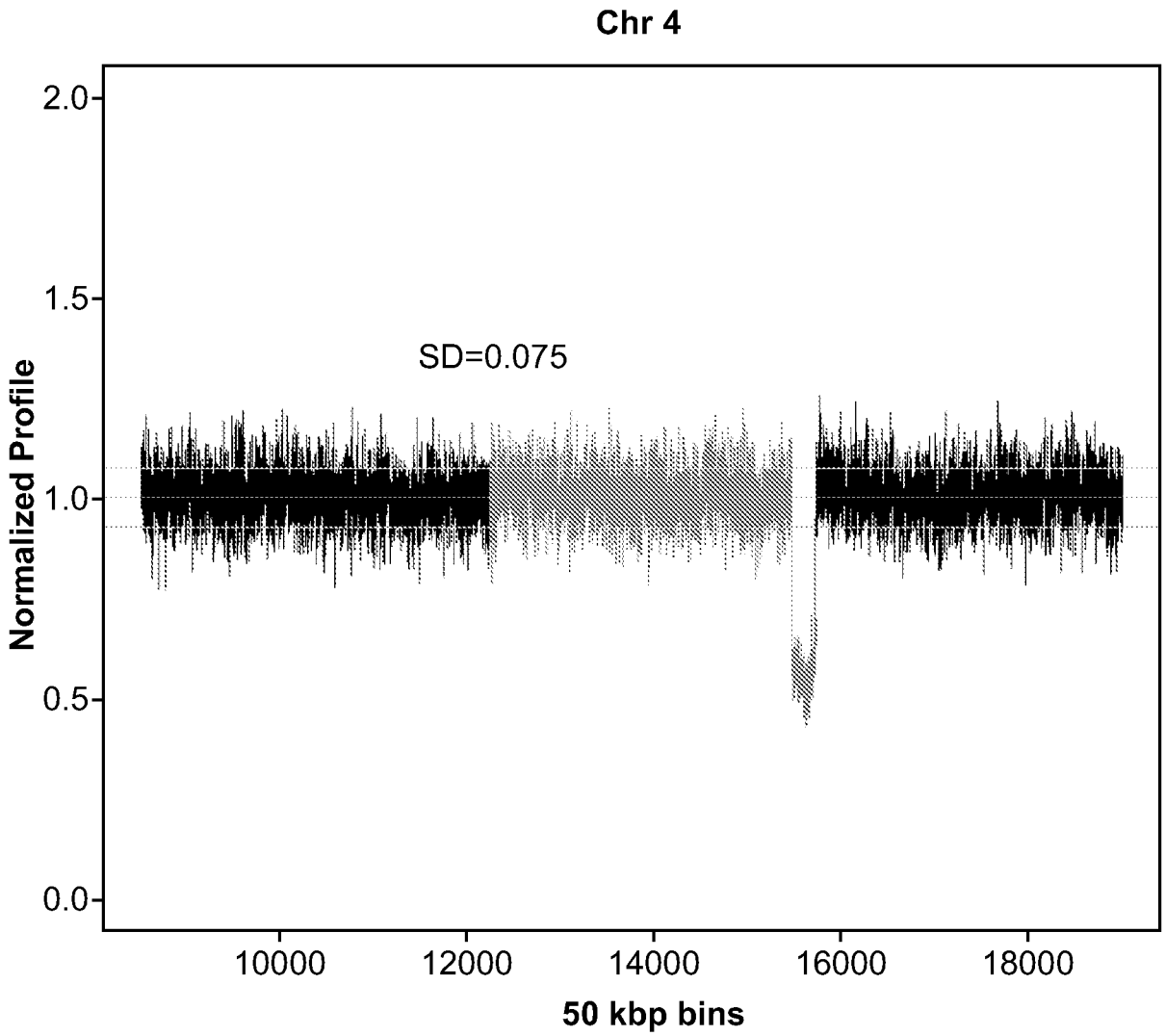

Chr 4
2.0
1.5
1.0
0.5
0.0
SD=0.075
Normalized Profile
10000
12000
14000
16000
18000
50 kbp bins

FIG. 105

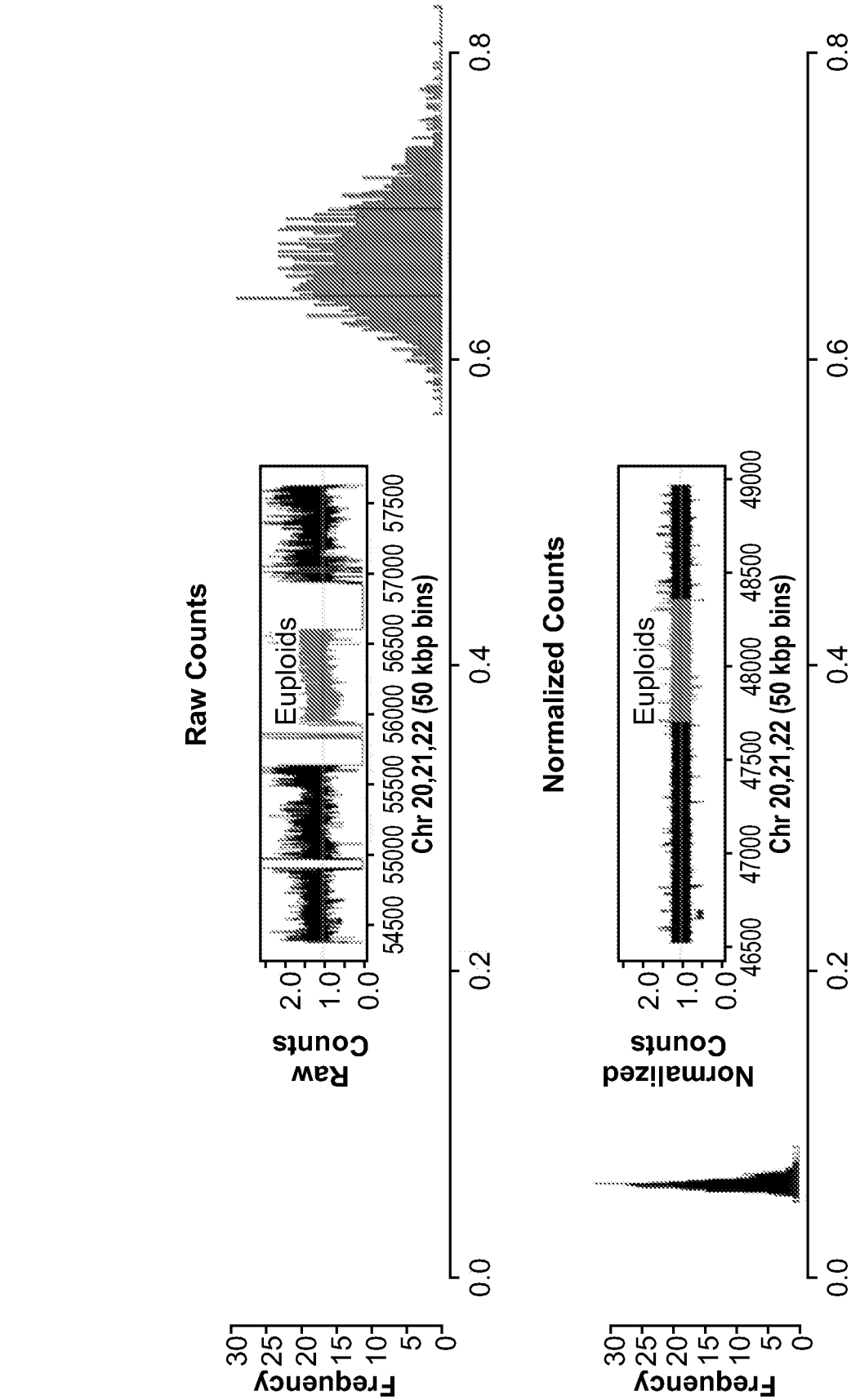

Raw Counts
Frequency
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
Raw Counts
2.0
1.0
0.0
Euploids
54500 55000 55500 56000 56500 57000 57500
Chr 20,21,22 (50 kbp bins)
Normalized Counts
Frequency
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
SD
Normalized Counts
2.0
1.0
0.0
Euploids
46500 47000 47500 48000 48500 49000
Chr 20,21,22 (50 kbp bins)

FIG. 106

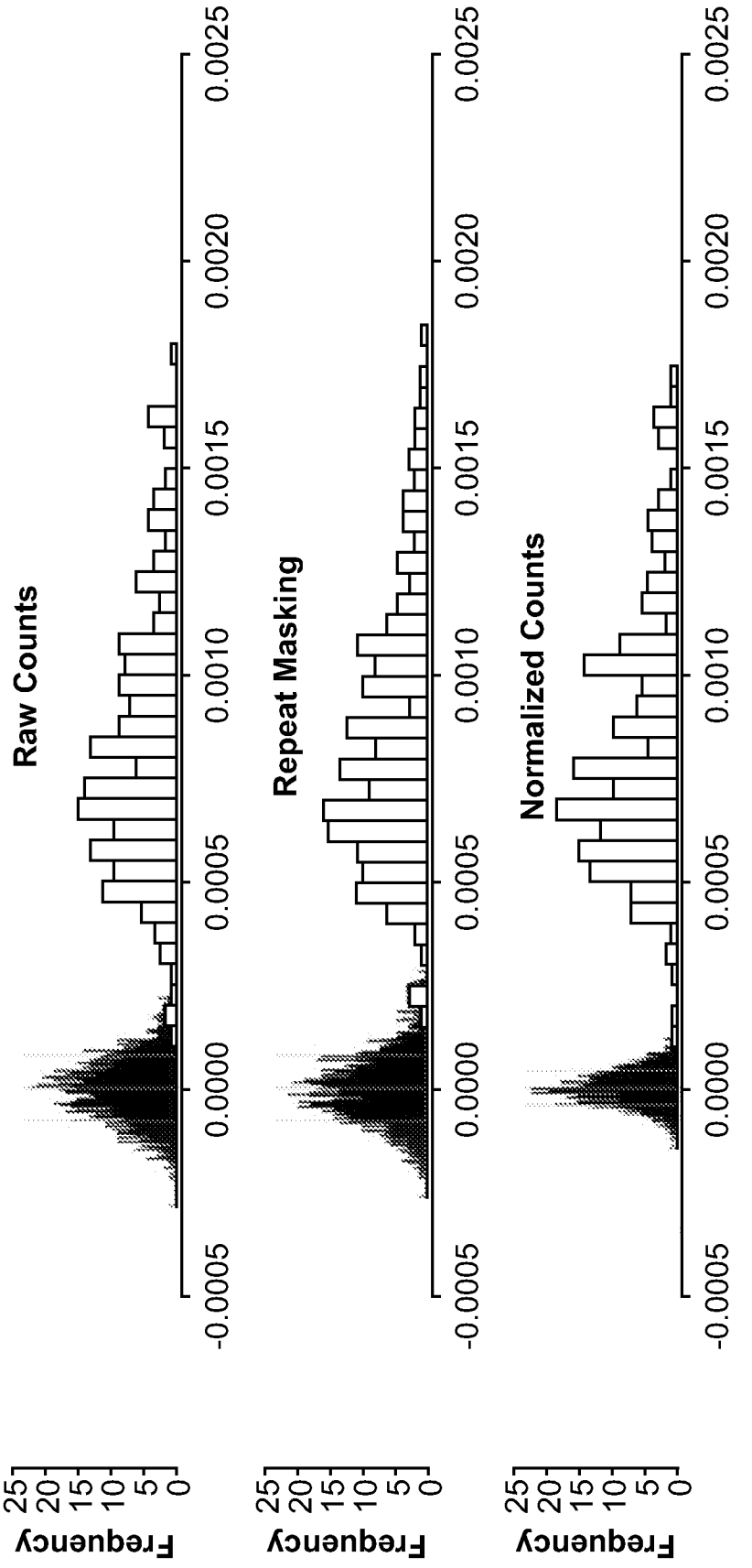

Raw Counts
Repeat Masking
Normalized Counts
Frequency
25
20
15
10
5
0
-0.0005
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
WI Chr 21 Fraction

FIG. 107

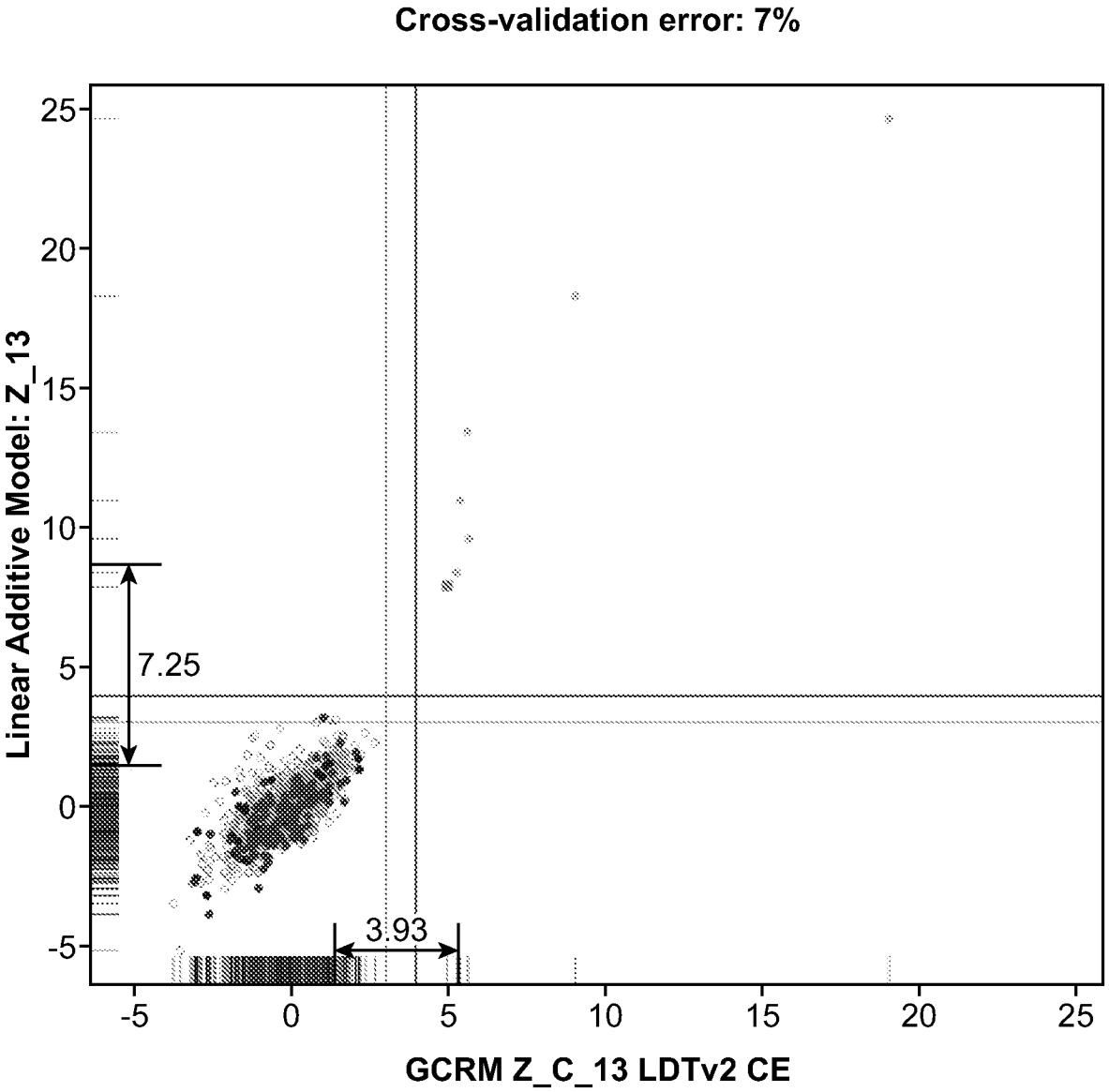

Cross-validation error: 7%
Linear Additive Model: Z_13
GCRM Z_C_13 LDTv2 CE
7.25
3.93
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25

FIG. 108

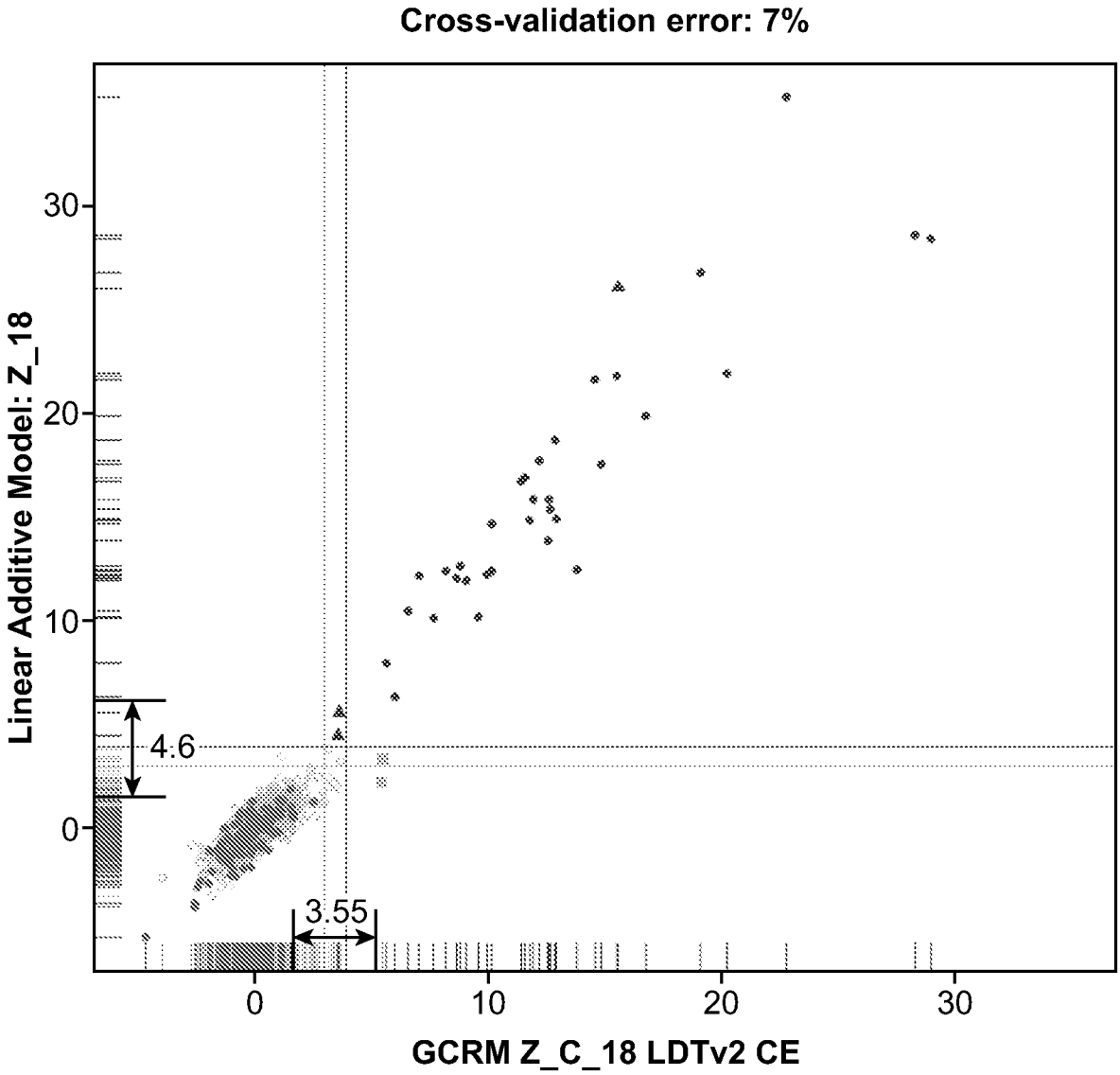

Cross-validation error: 7%
Linear Additive Model: Z_18
GCRM Z_C_18 LDTv2 CE
30
20
10
0
4.6
3.55
0
10
20
30

FIG. 109

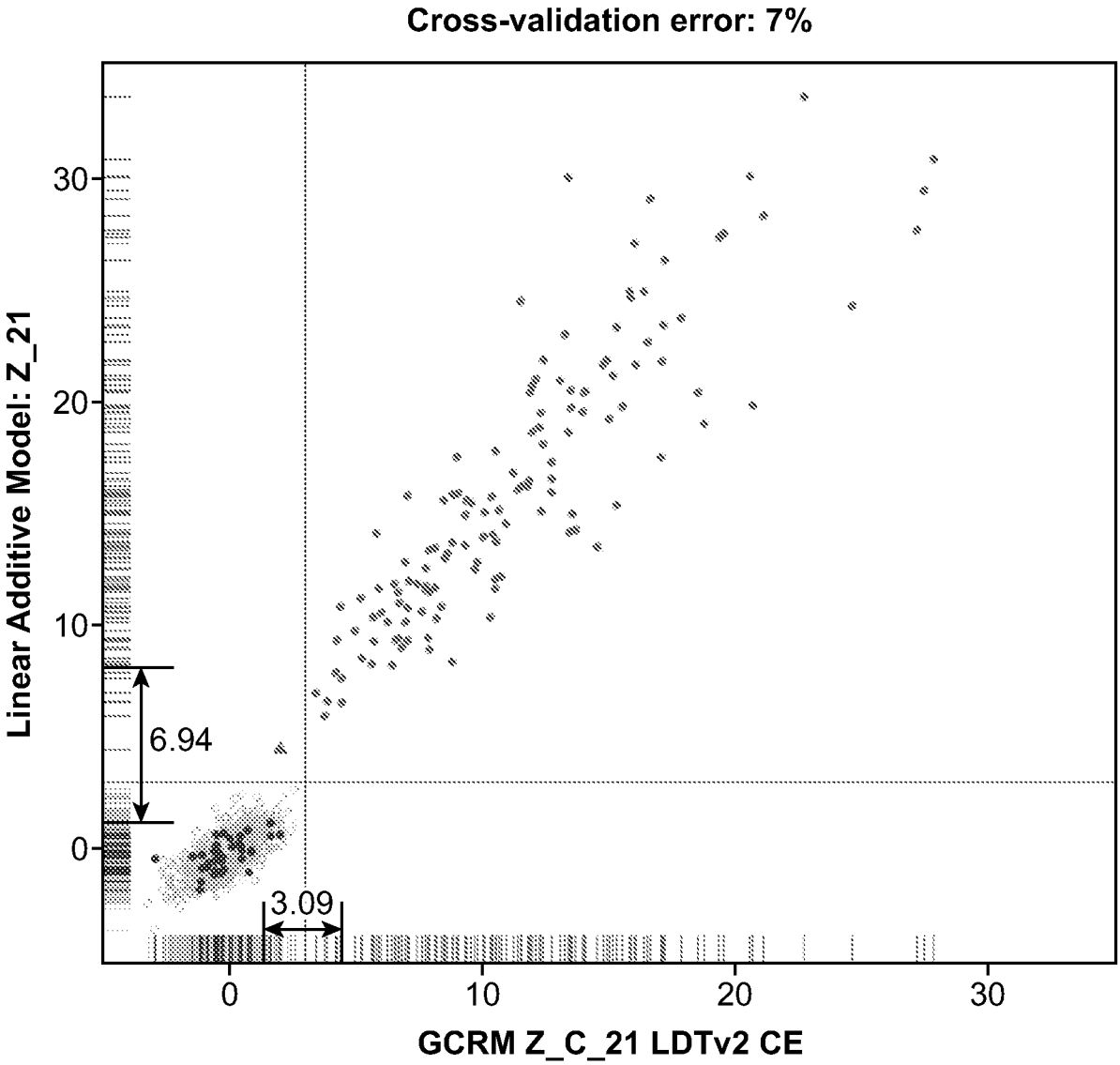

Cross-validation error: 7%
Linear Additive Model: Z_21
GCRM Z_C_21 LDTv2 CE
6.94
3.09
0
10
20
30

FIG. 110

high negative control >3, low fetal percentage with boderline Z-score

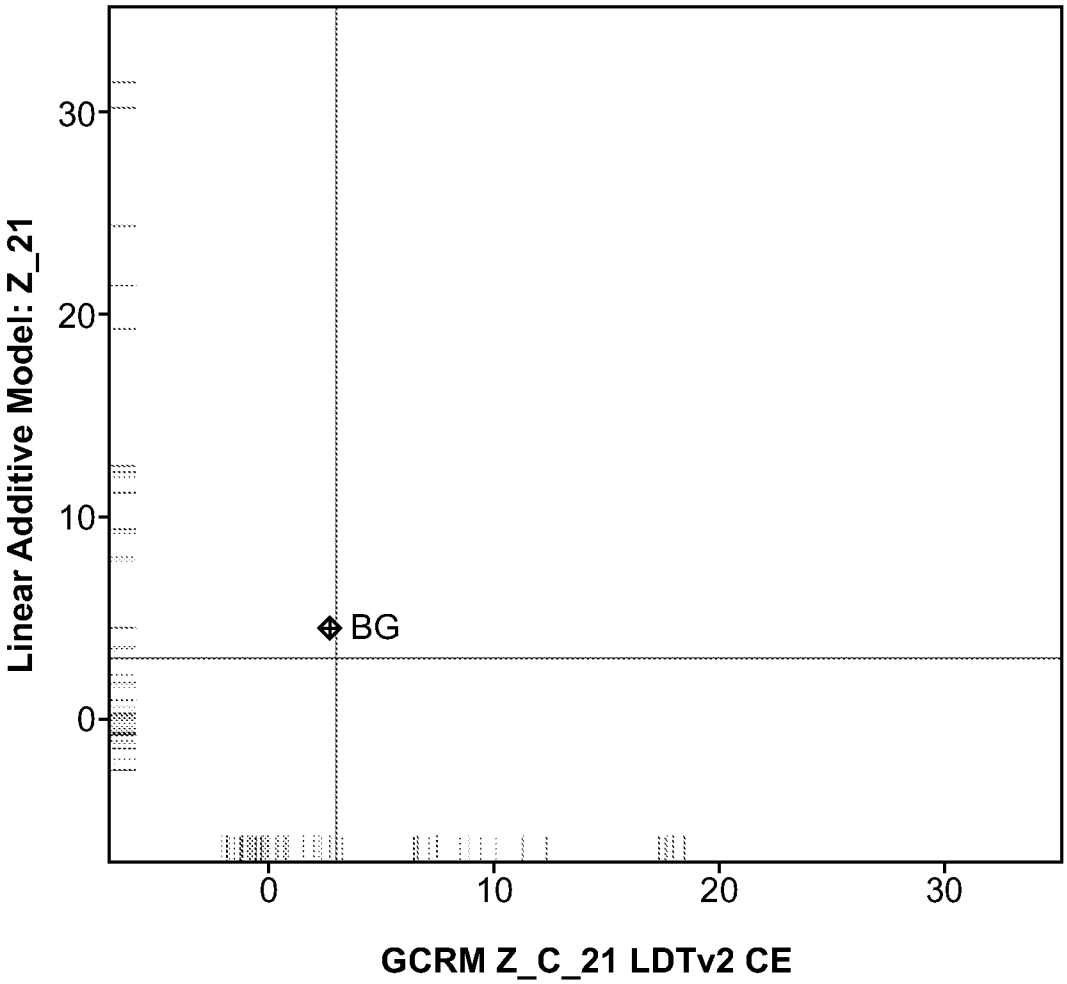


FIG. 111

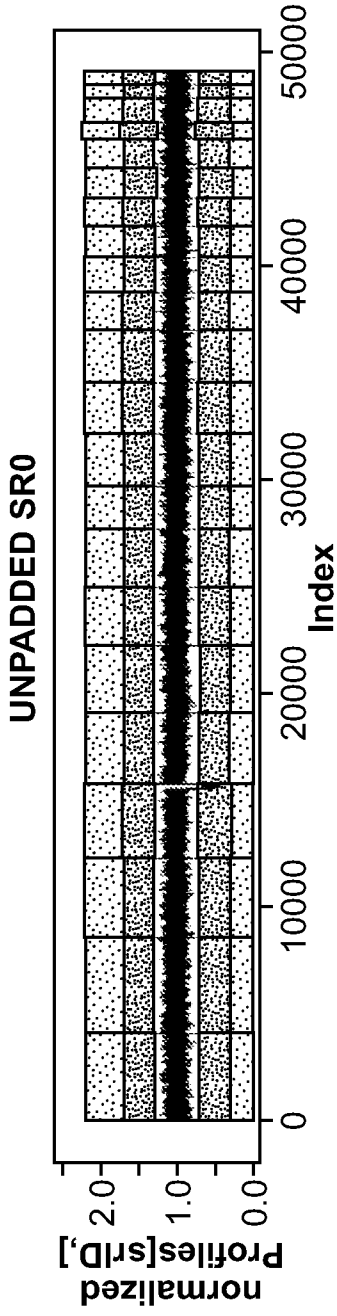


FIG. 112A

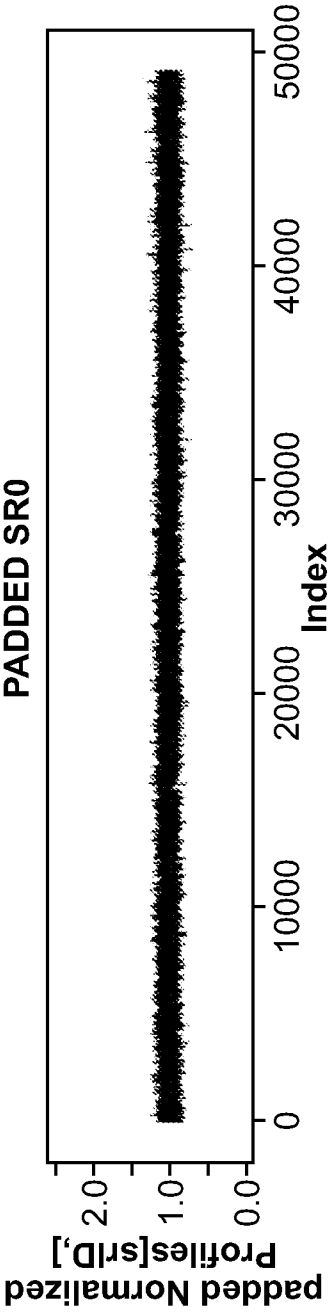


FIG. 112B

FIG. 112C

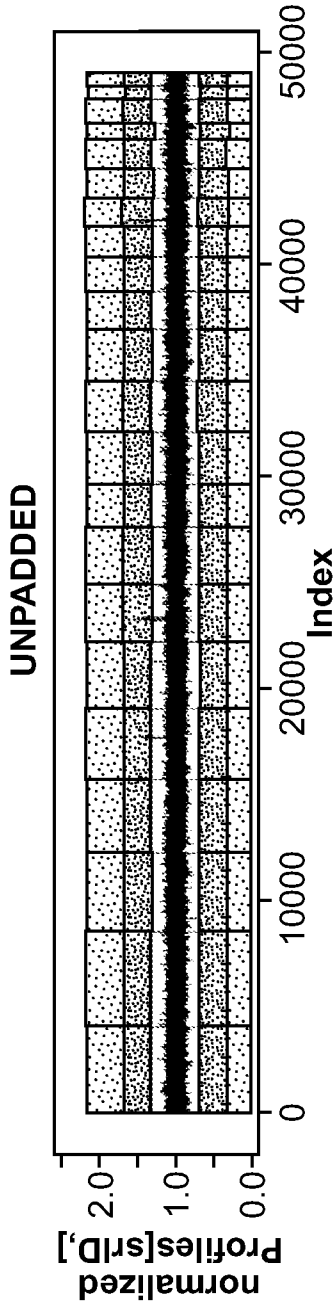


FIG. 113A

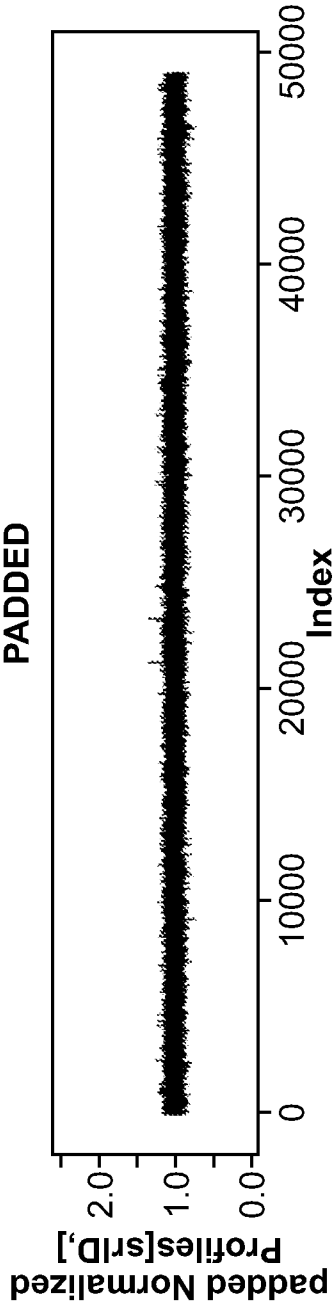


FIG. 113B

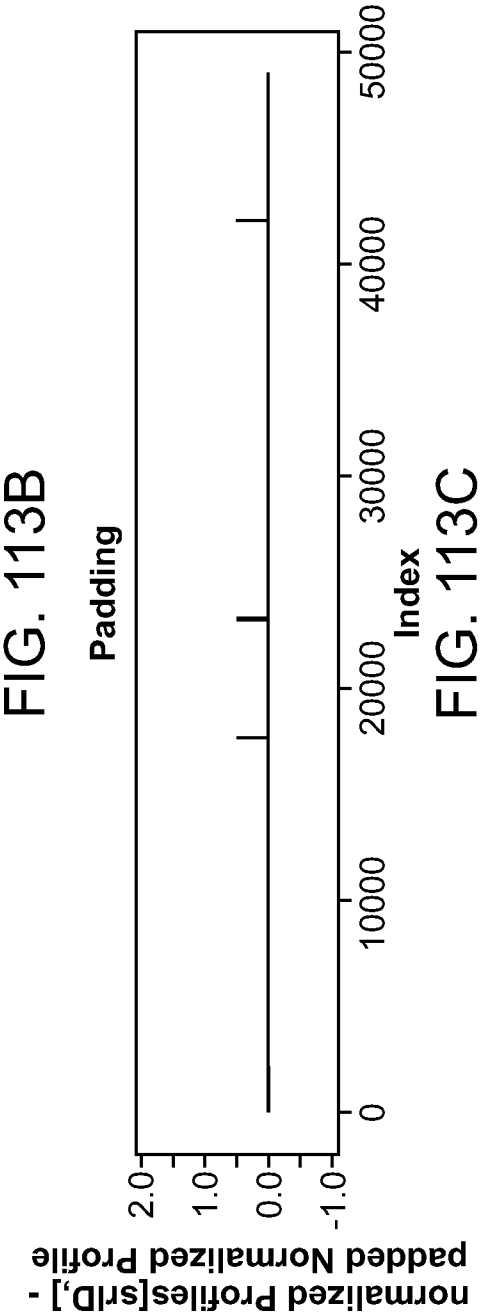


FIG. 113C

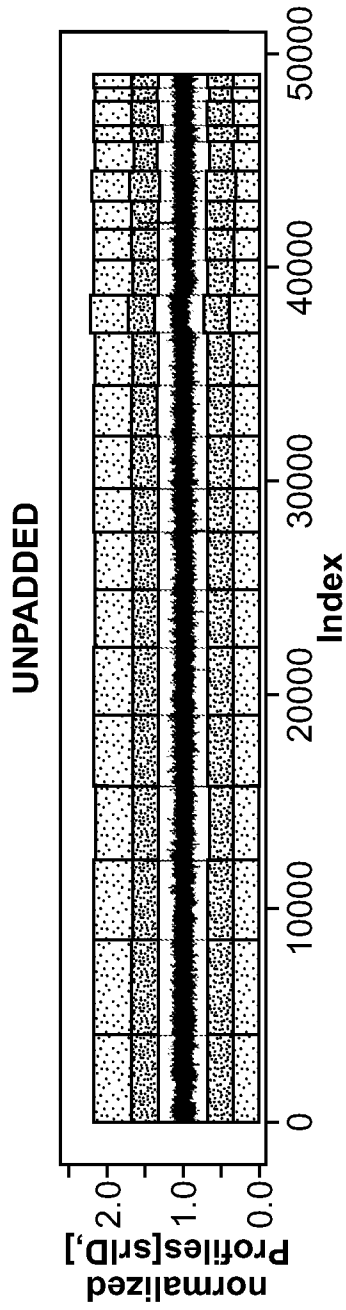


FIG. 114A

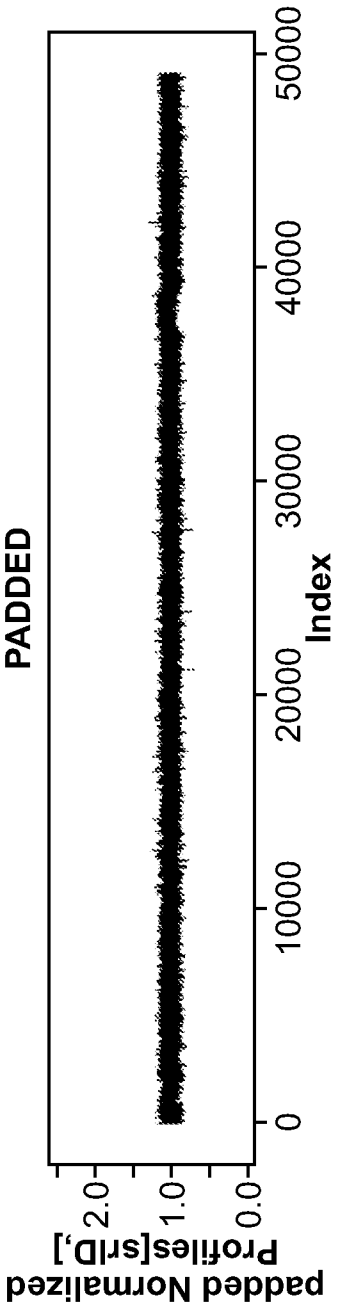


FIG. 114B

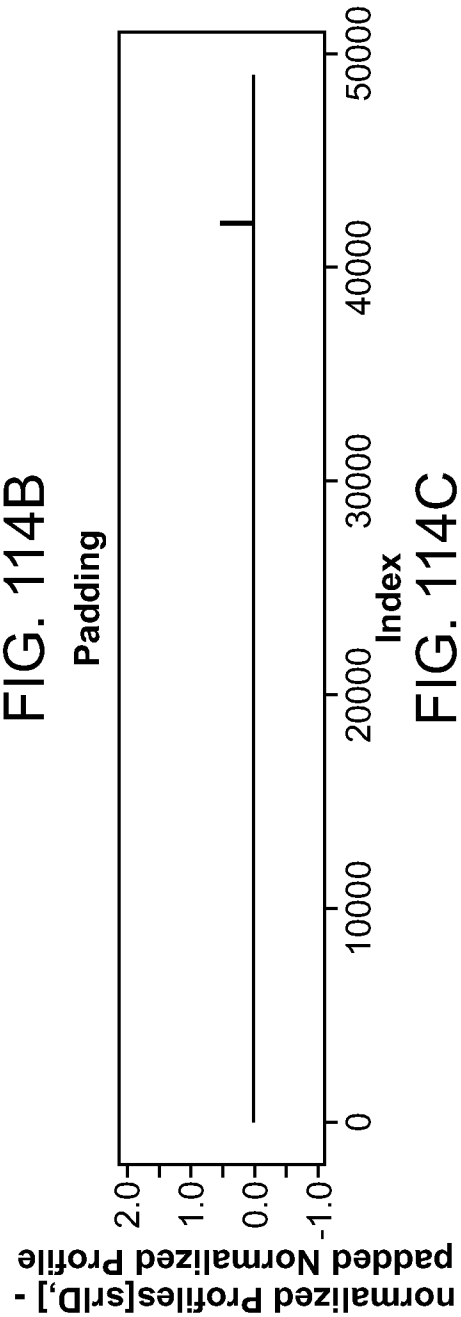


FIG. 114C

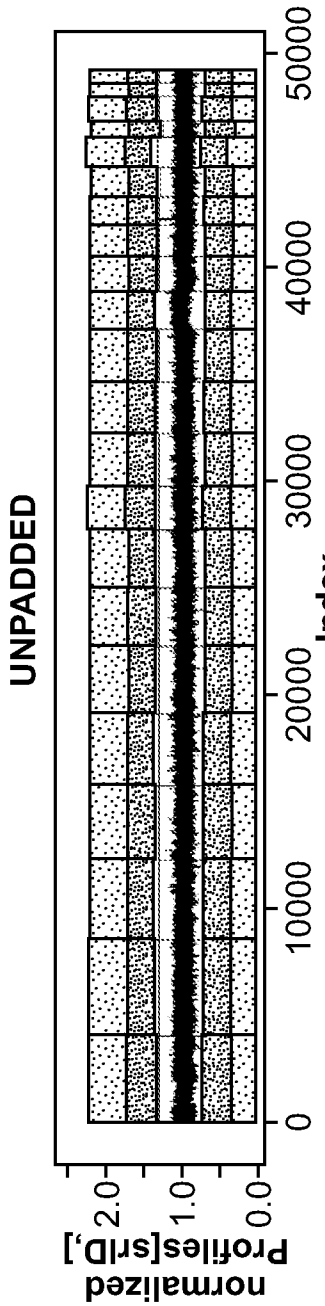


FIG. 115A

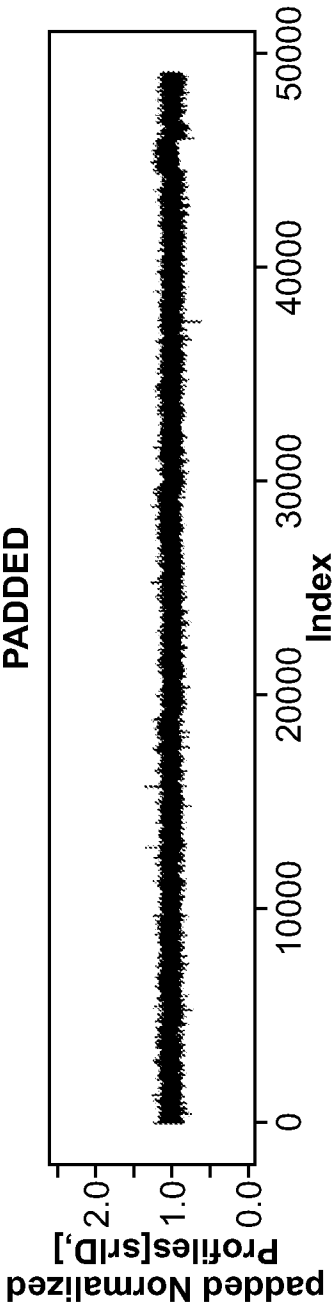


FIG. 115B

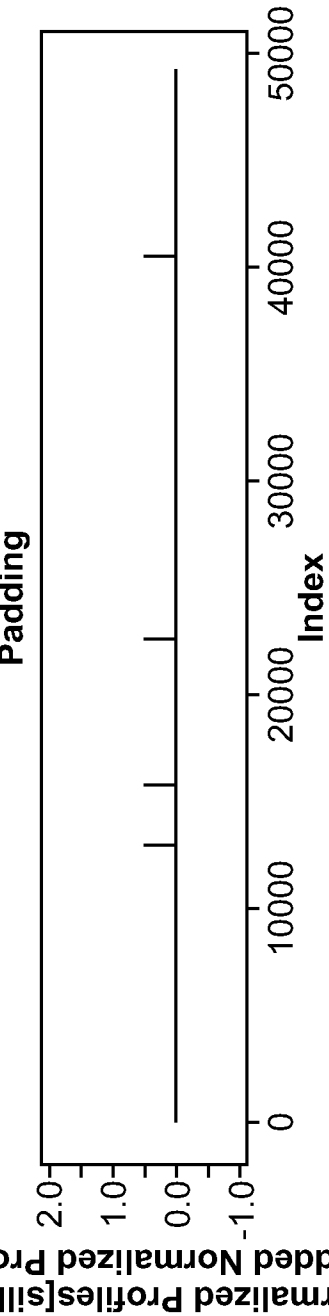


FIG. 115C

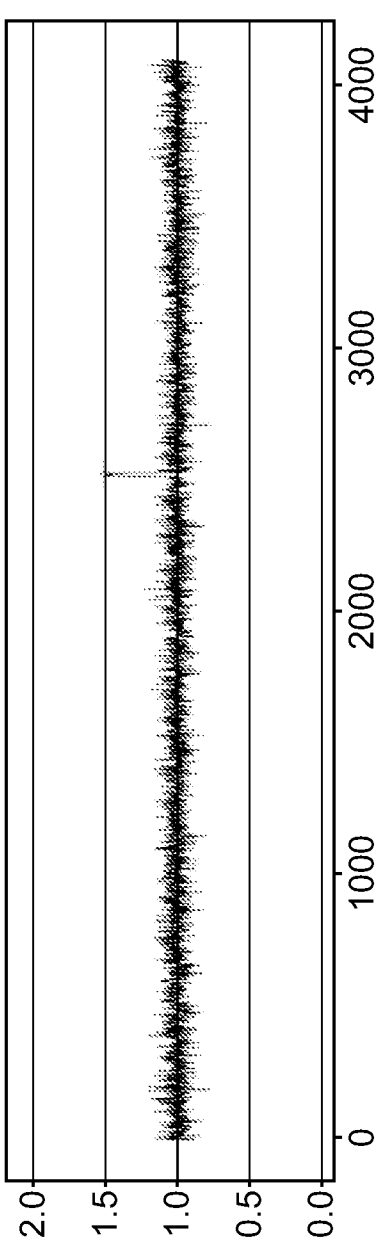


FIG. 116

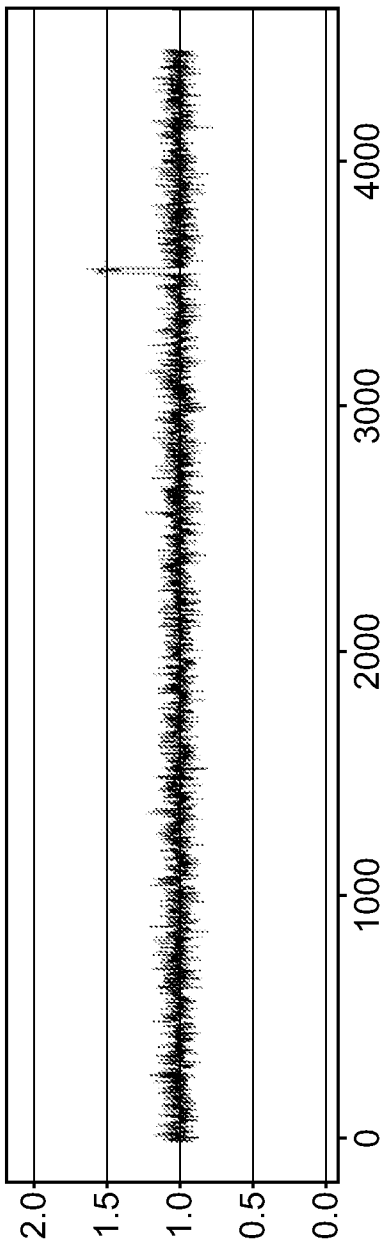


FIG. 117

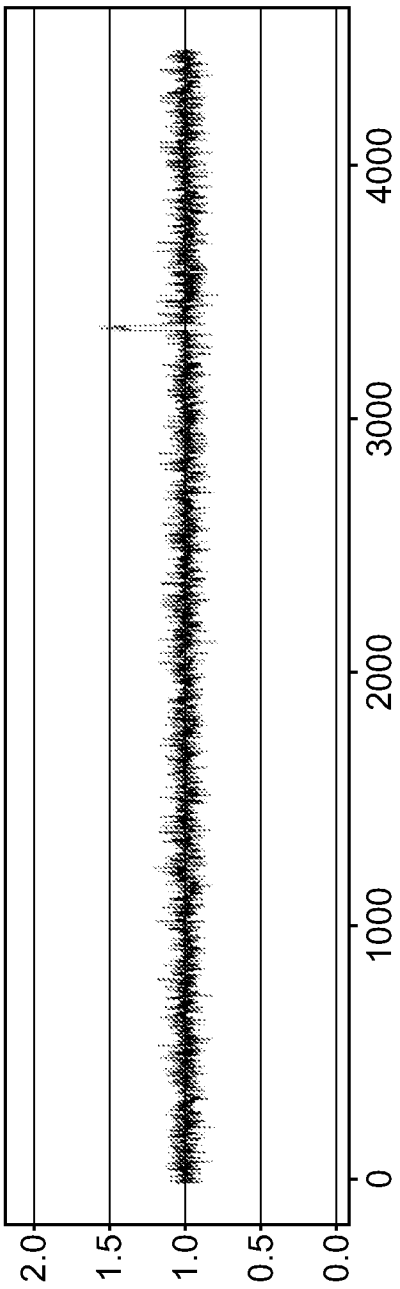


FIG. 118

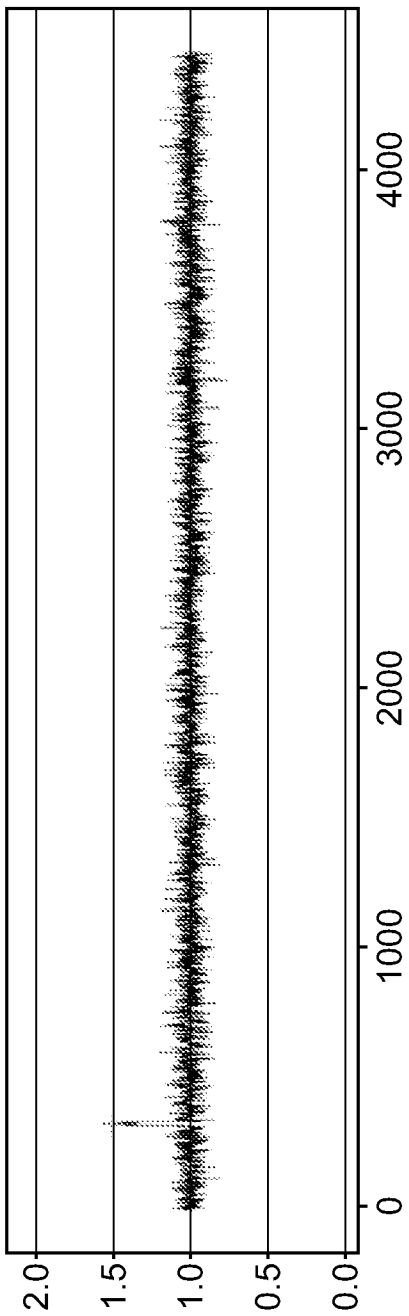


FIG. 119

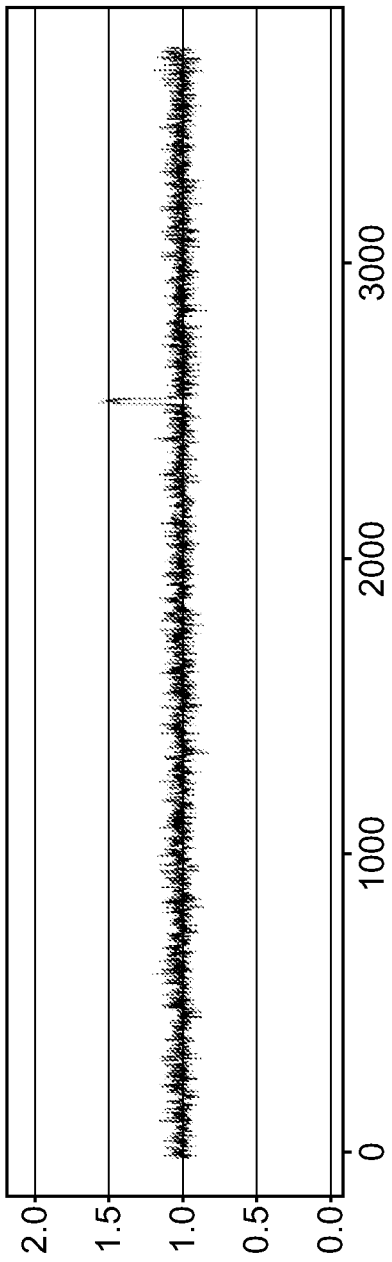


FIG. 120

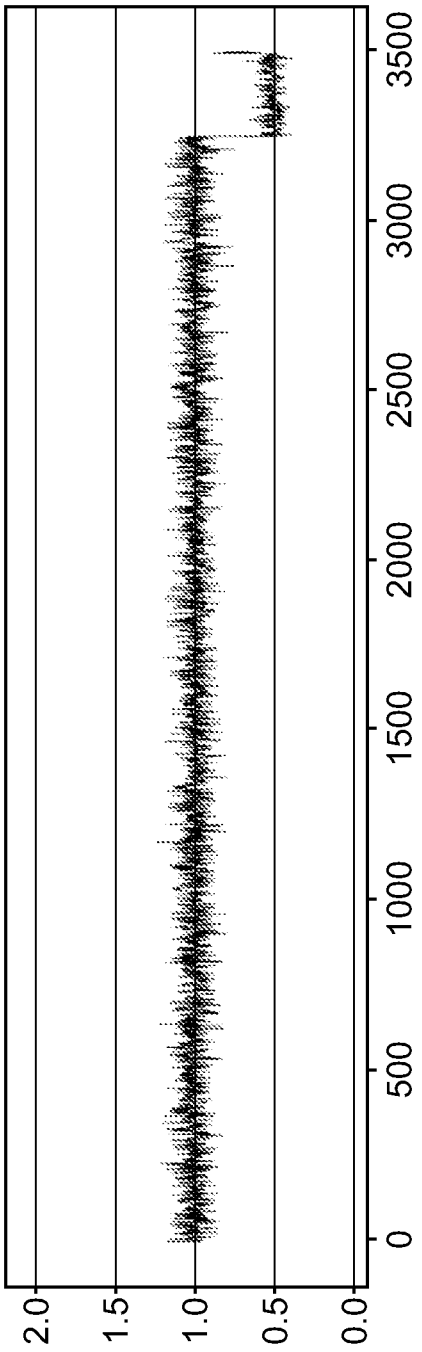


FIG. 121

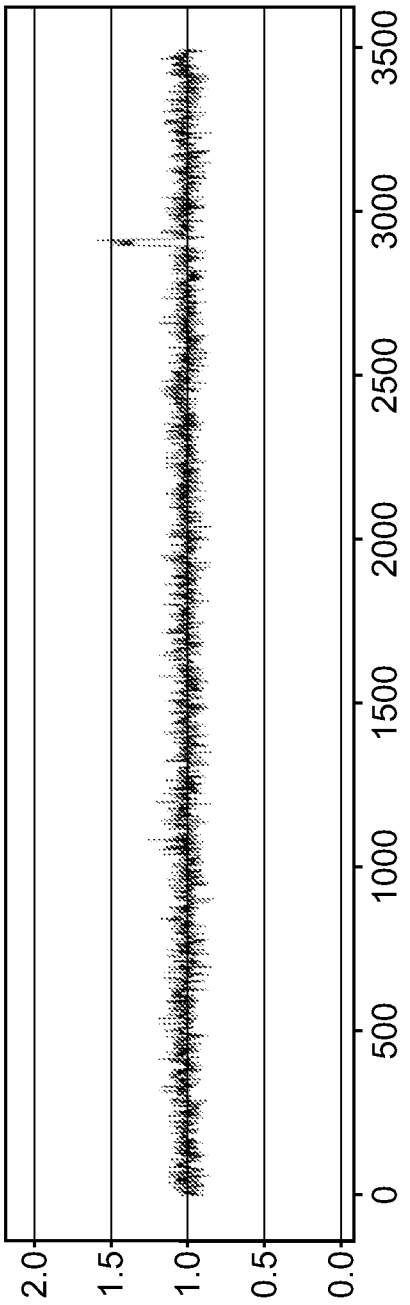


FIG. 122

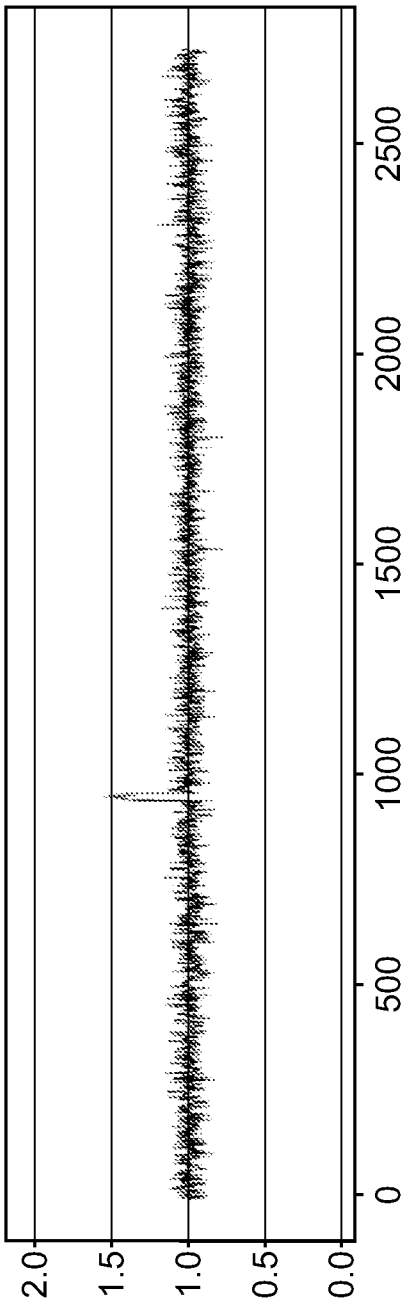


FIG. 123

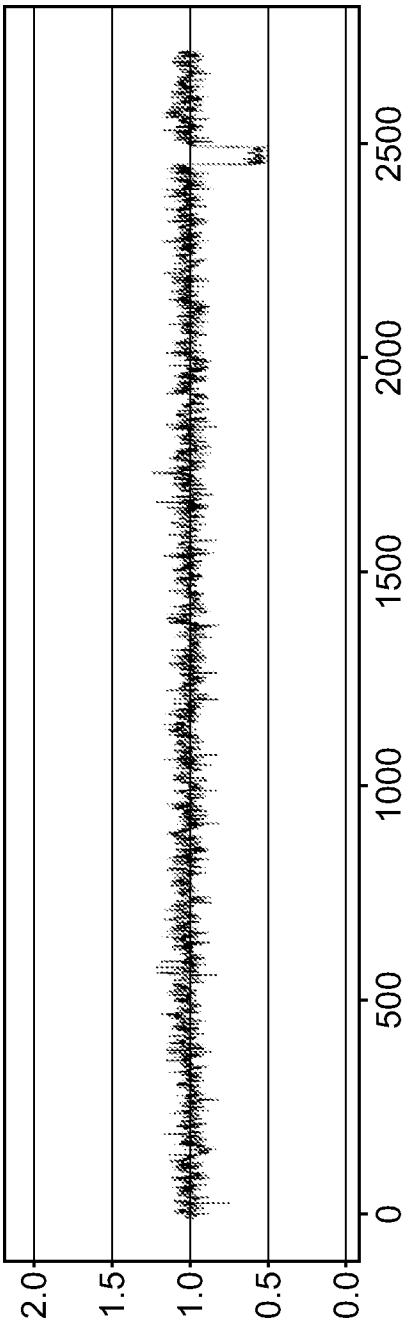


FIG. 124

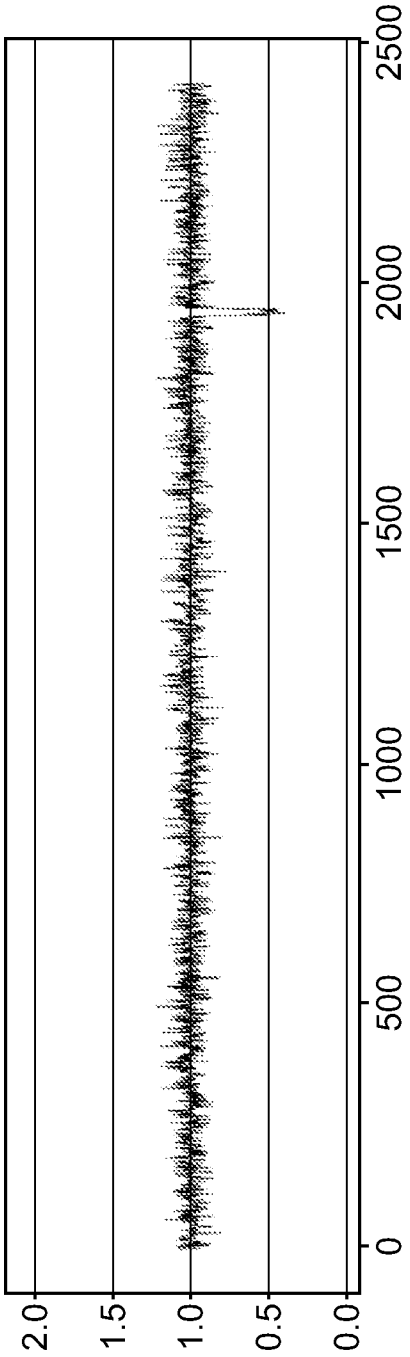


FIG. 125

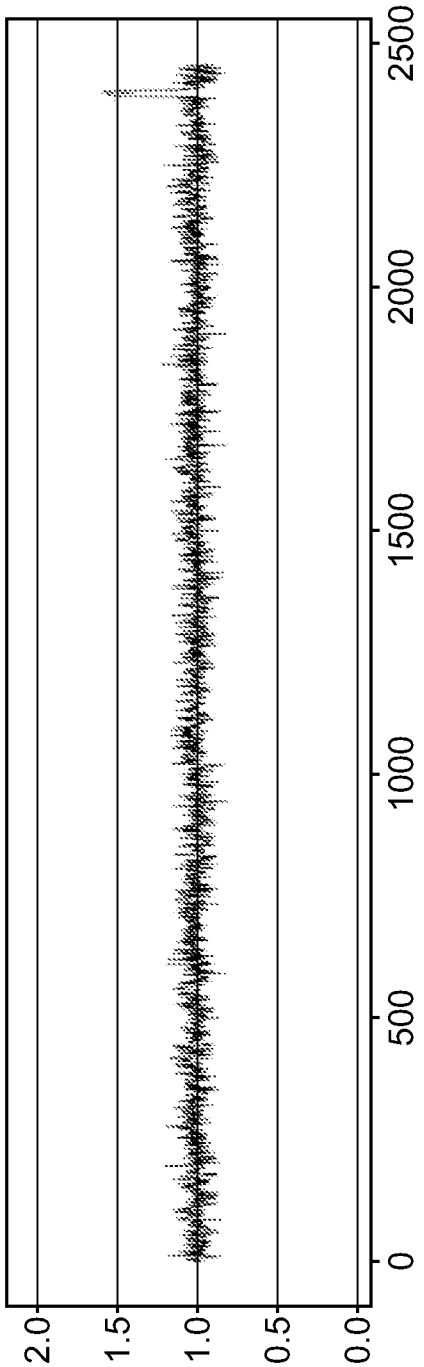


FIG. 126

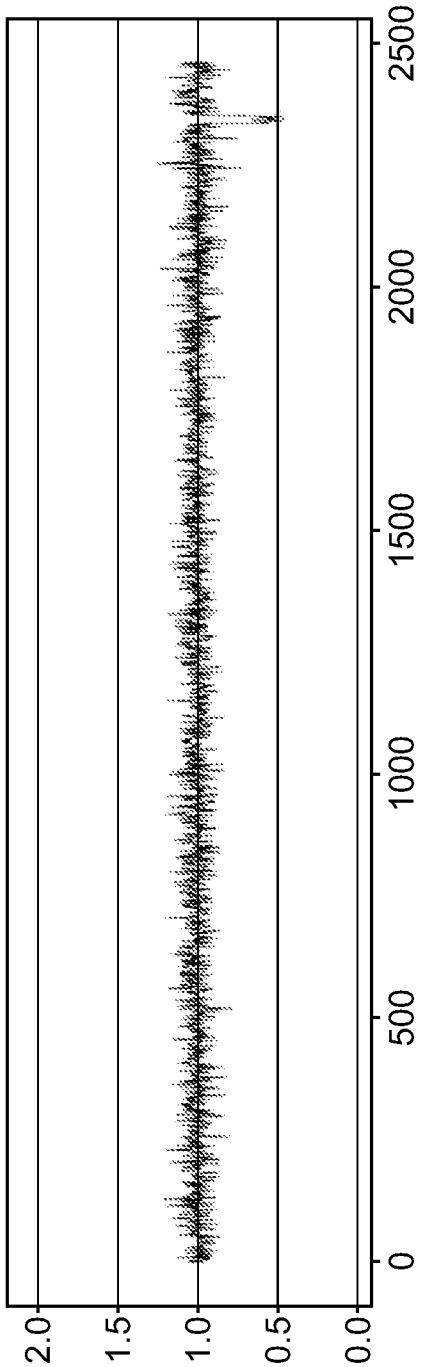


FIG. 127

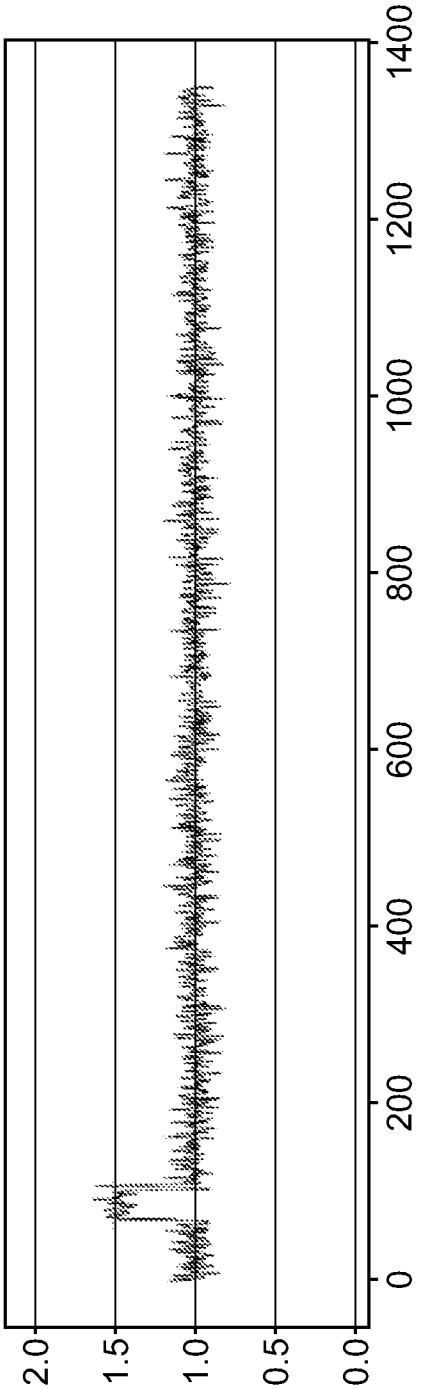


FIG. 128

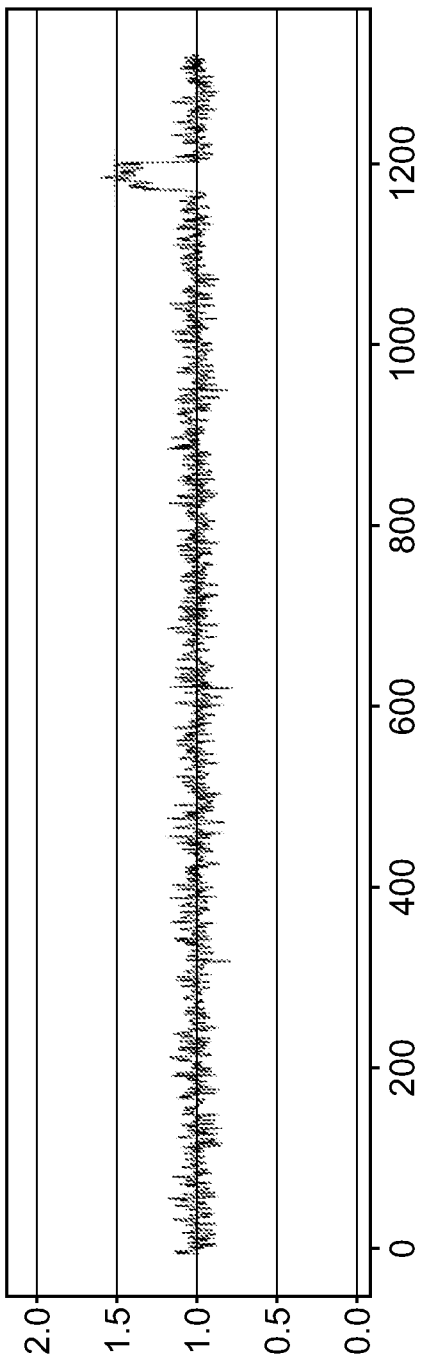


FIG. 129

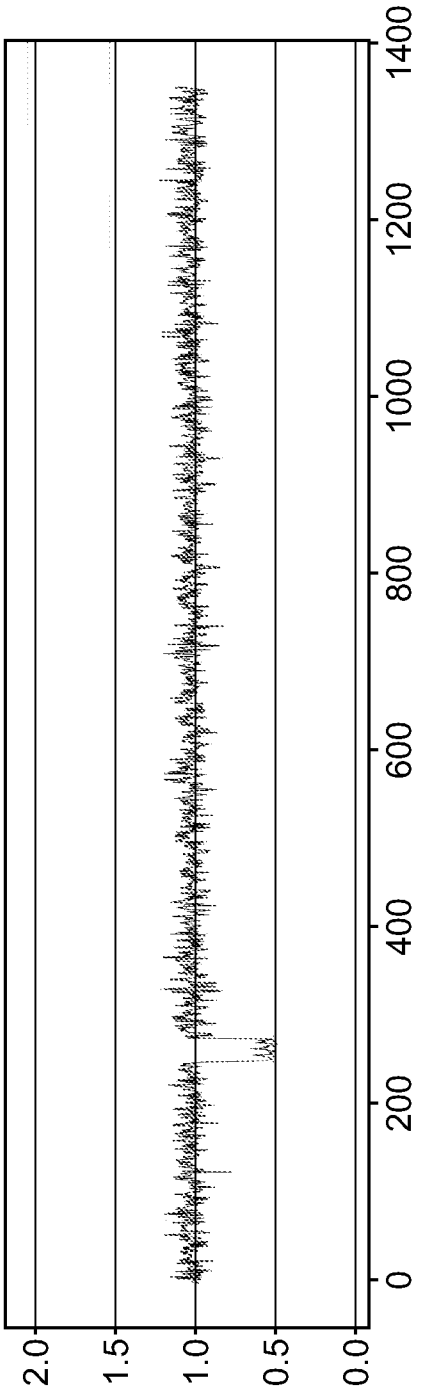


FIG. 130

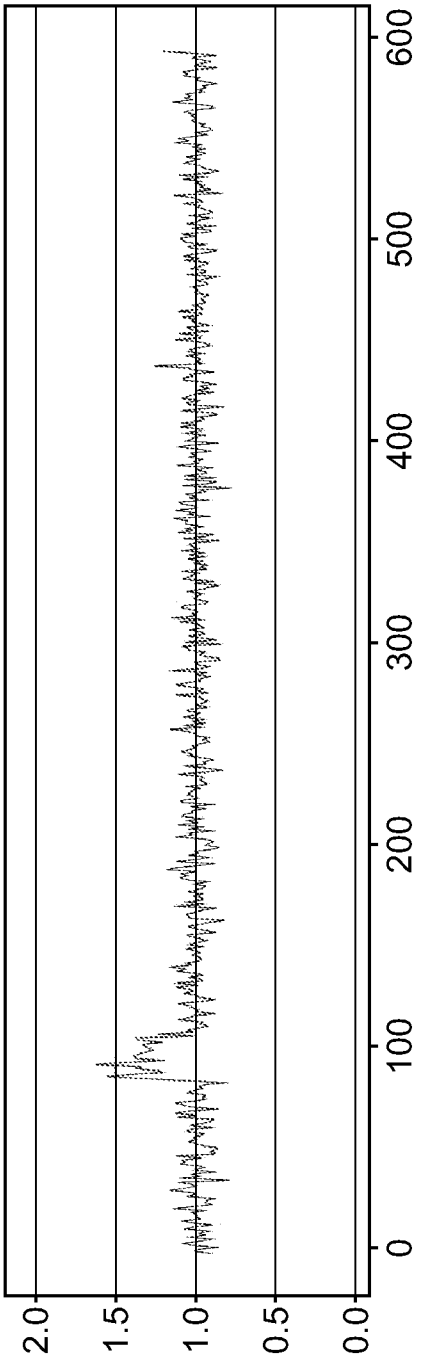


FIG. 131

FIG. 132

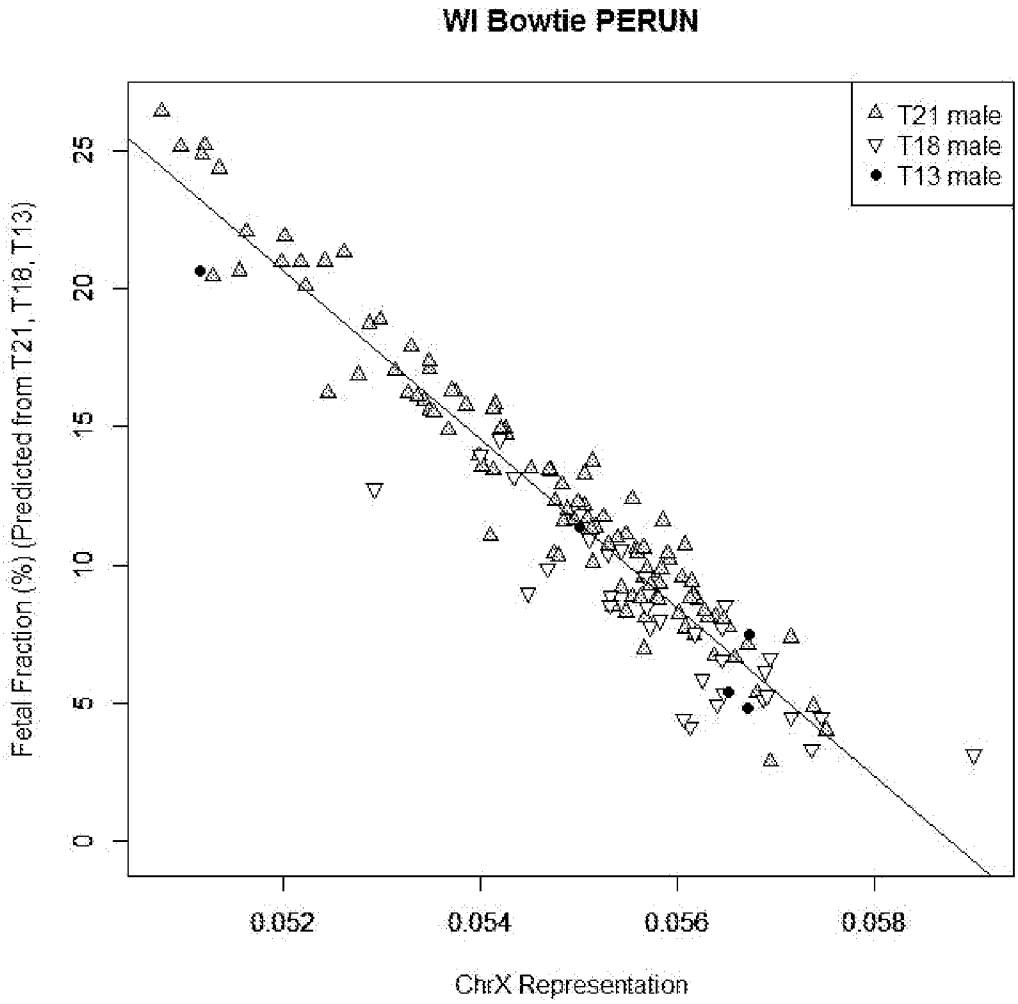

WI Bowtie PERUN
T21 male
T18 male
T13 male
Fetal Fraction (%) (Predicted from T21, T18, T13)
0
5
10
15
20
25
0.052
0.054
0.056
0.058
ChrX Representation

FIG. 133

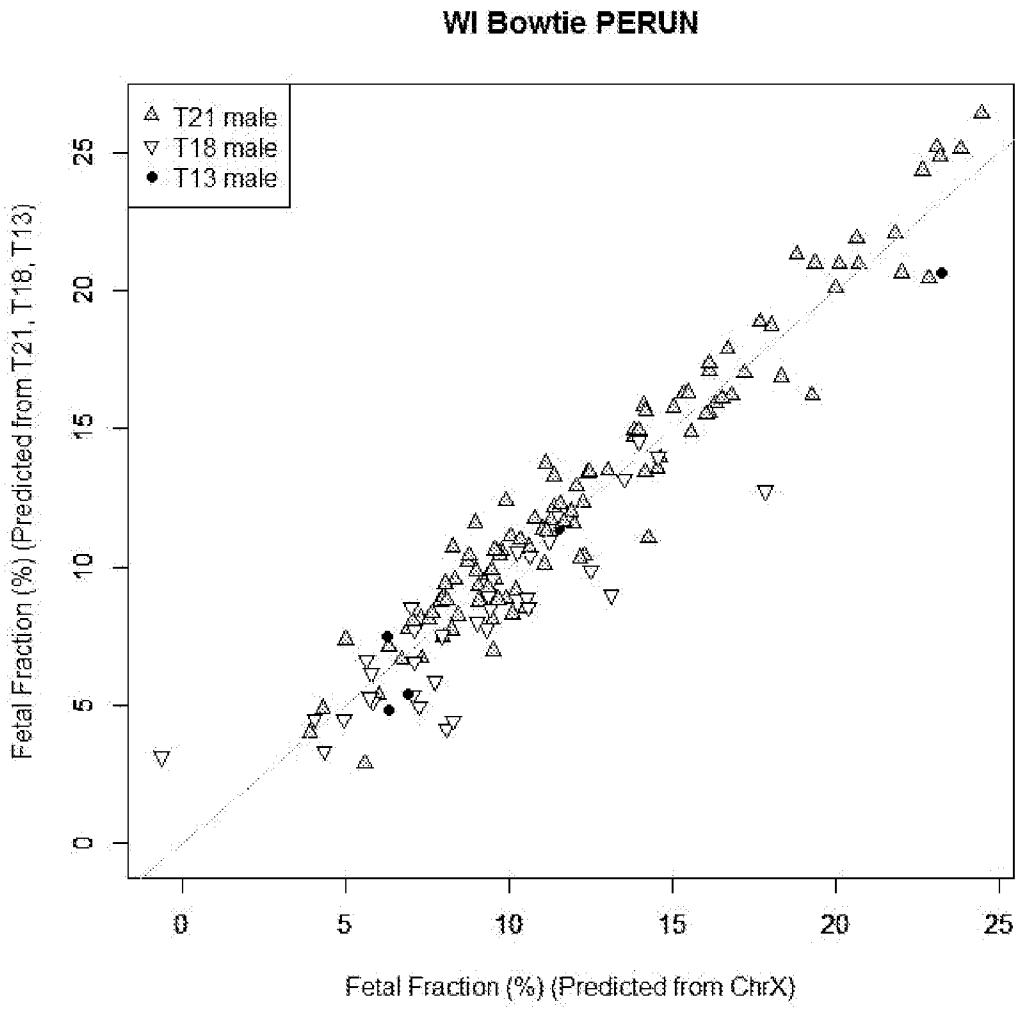

WI Bowtie PERUN
T21 male
T18 male
T13 male
Fetal Fraction (%) (Predicted from T21, T18, T13)
Fetal Fraction (%) (Predicted from ChrX)
0
5
10
15
20
25

FIG. 134

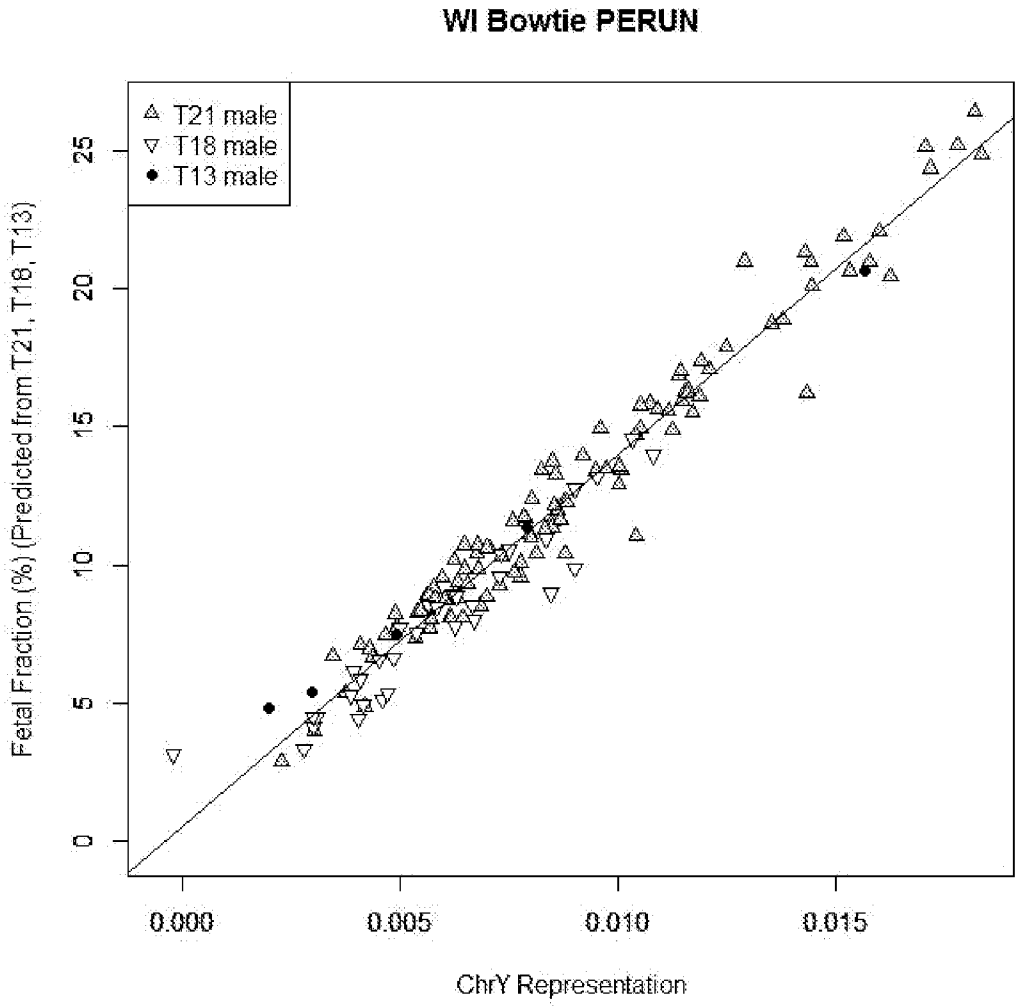

WI Bowtie PERUN
T21 male
T18 male
T13 male
Fetal Fraction (%) (Predicted from T21, T18, T13)
0
5
10
15
20
25
0.000
0.005
0.010
0.015
ChrY Representation

FIG. 135

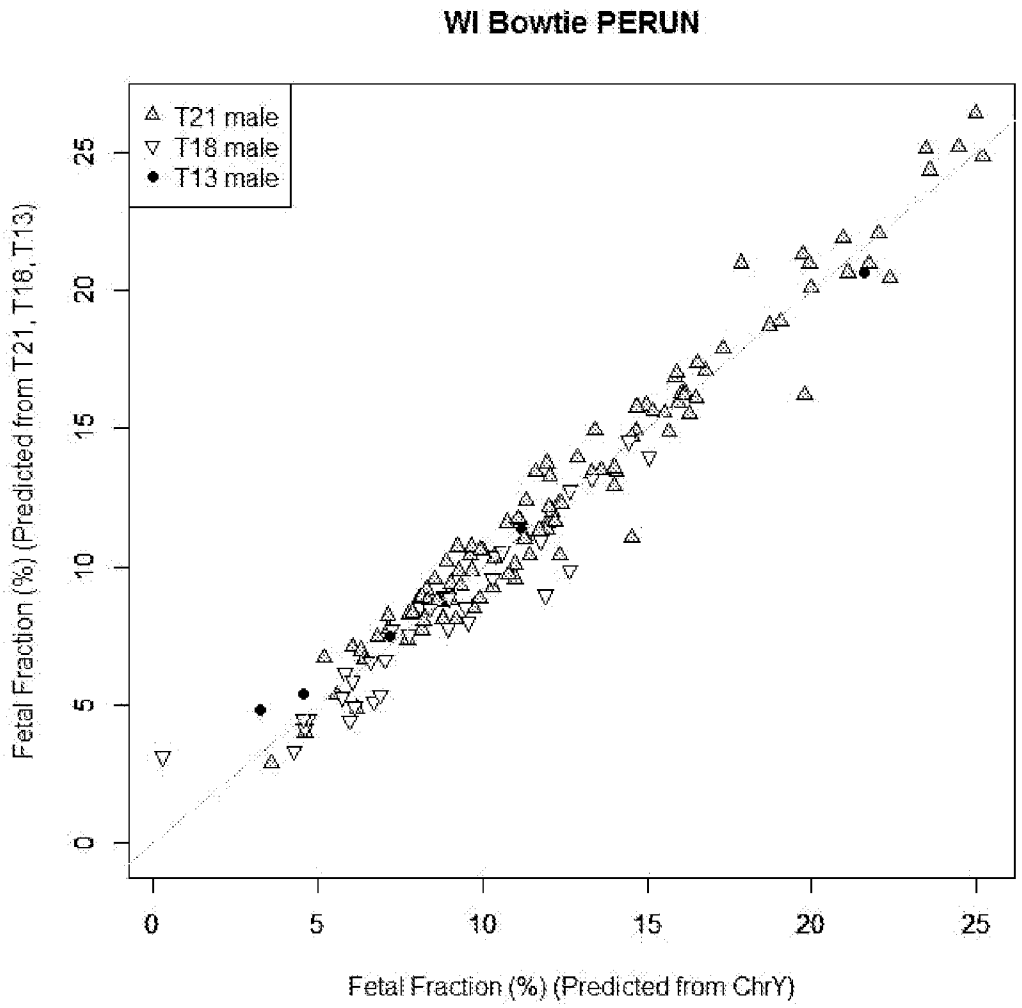

WI Bowtie PERUN
T21 male
T18 male
T13 male
Fetal Fraction (%) (Predicted from T21, T18, T13)
Fetal Fraction (%) (Predicted from ChrY)
0
5
10
15
20
25

FIG. 136

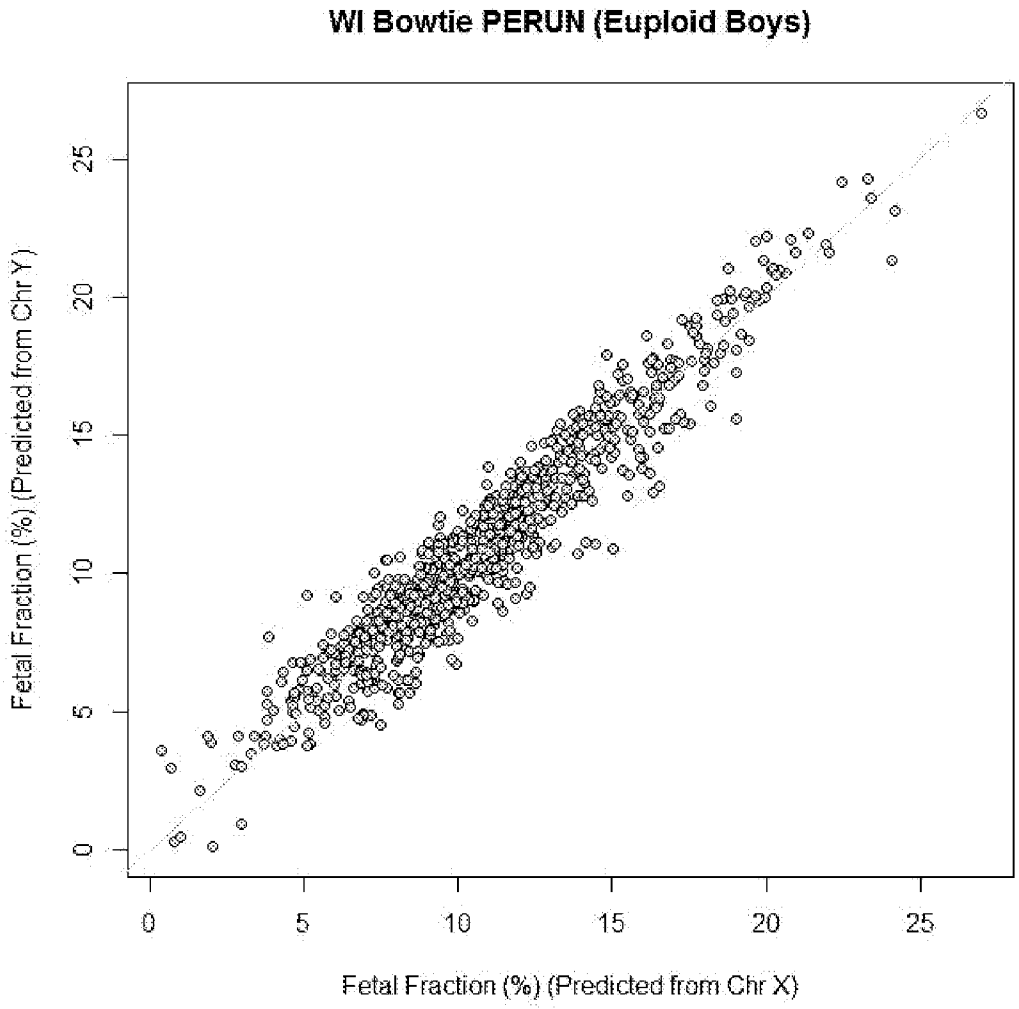

WI Bowtie PERUN (Euploid Boys)
Fetal Fraction (%) (Predicted from Chr Y)
Fetal Fraction (%) (Predicted from Chr X)
0
5
10
15
20
25

FIG. 137

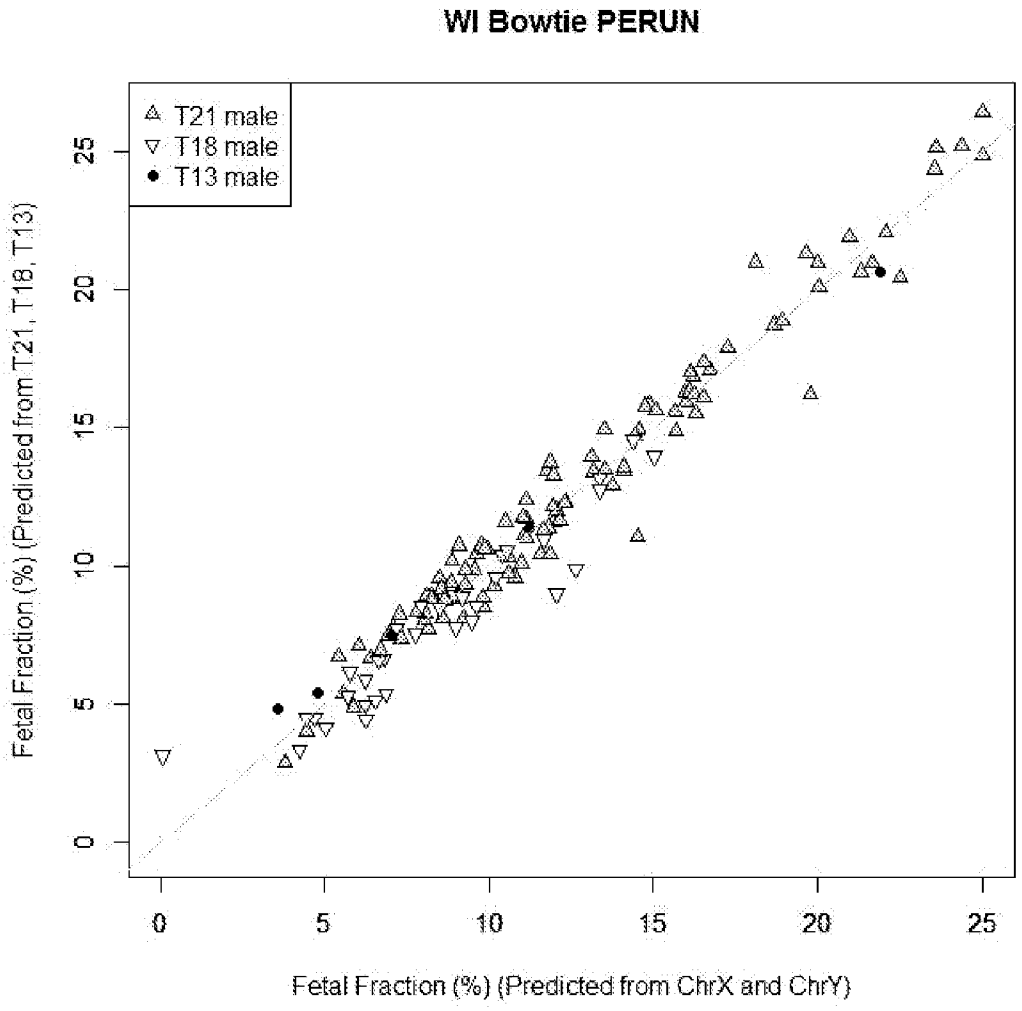

WI Bowtie PERUN
T21 male
T18 male
T13 male
Fetal Fraction (%) (Predicted from T21, T18, T13)
Fetal Fraction (%) (Predicted from ChrX and ChrY)
0
5
10
15
20
25

FIG. 138

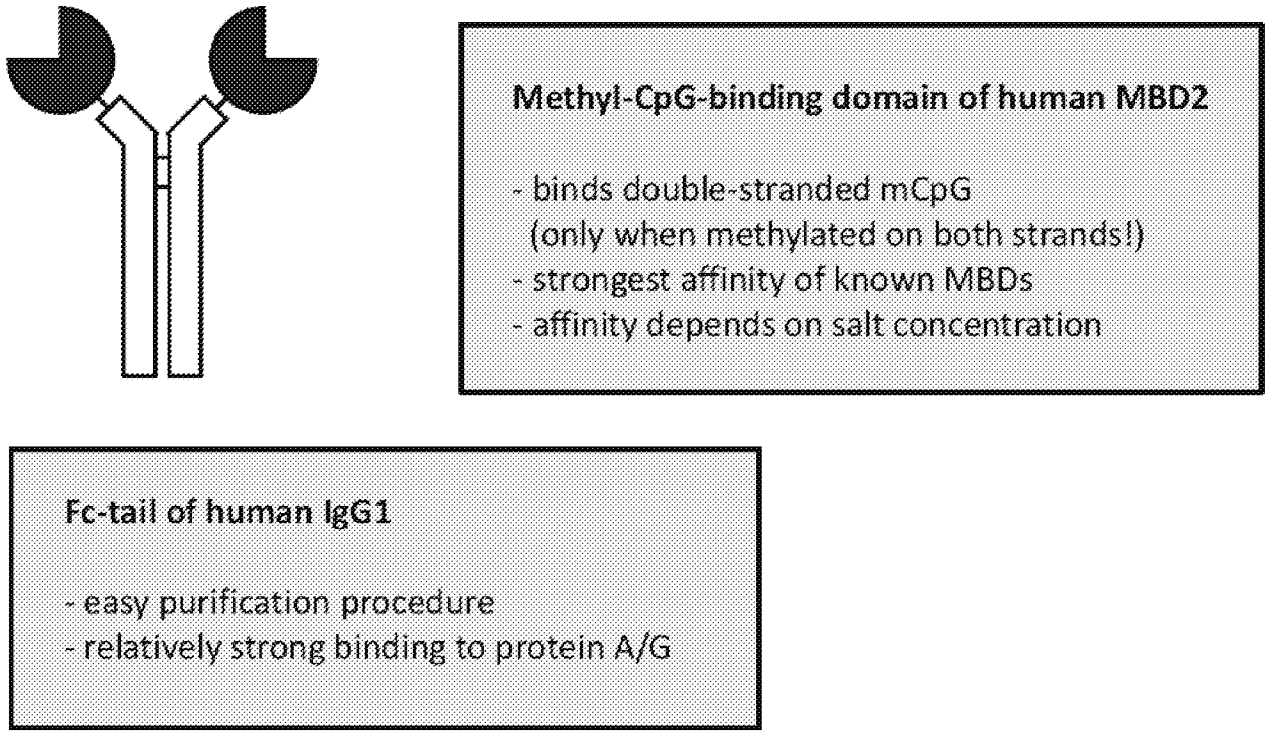

Design of the recombinant MBD-Fc protein
Methyl-CpG-binding domain of human MBD2
- binds double-stranded mCpG
(only when methylated on both strands!)
- strongest affinity of known MBDs
- affinity depends on salt concentration
Fc-tail of human IgG1
- easy purification procedure
- relatively strong binding to protein A/G

FIG. 139

MBD Protein

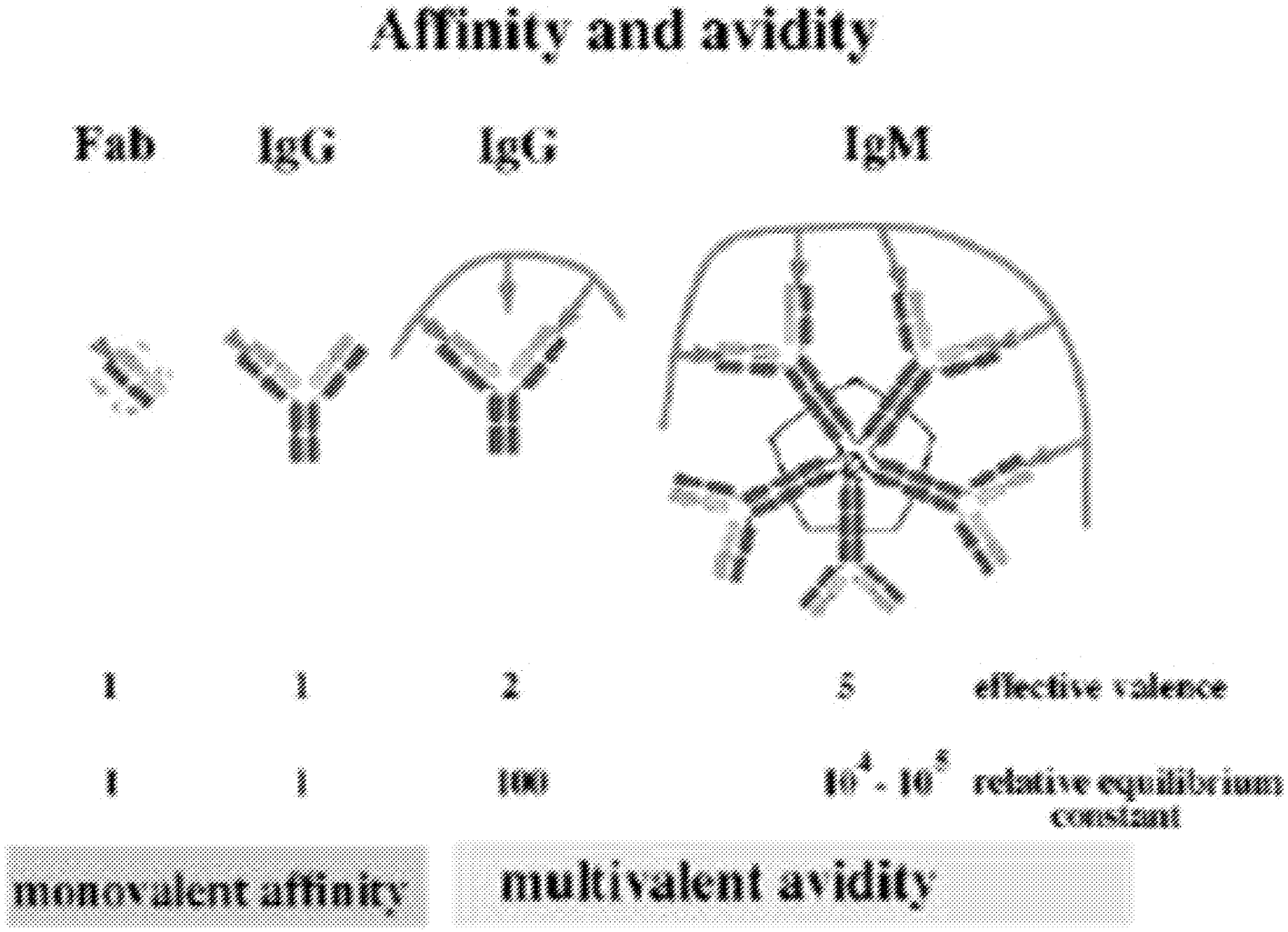
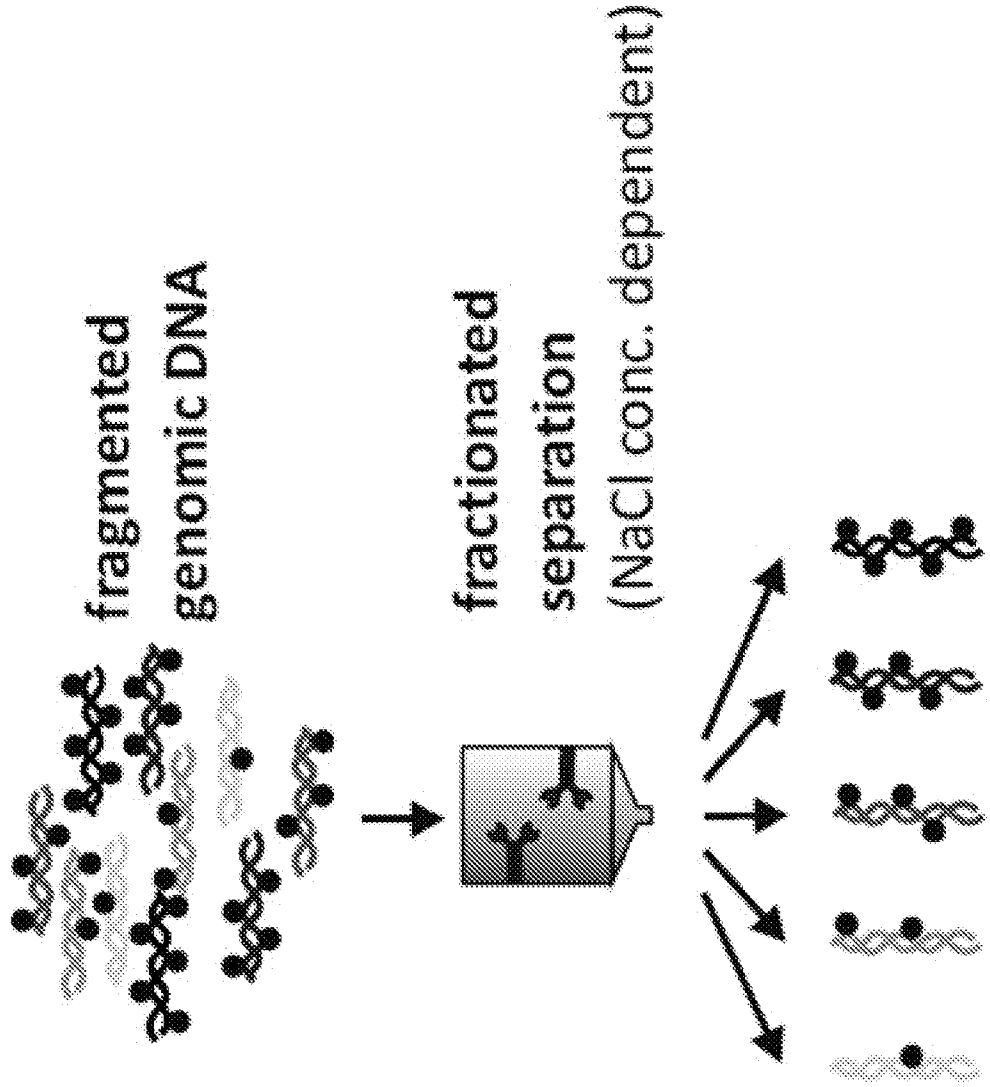


FIG. 140

Fractionating DNA
Based on Methylation



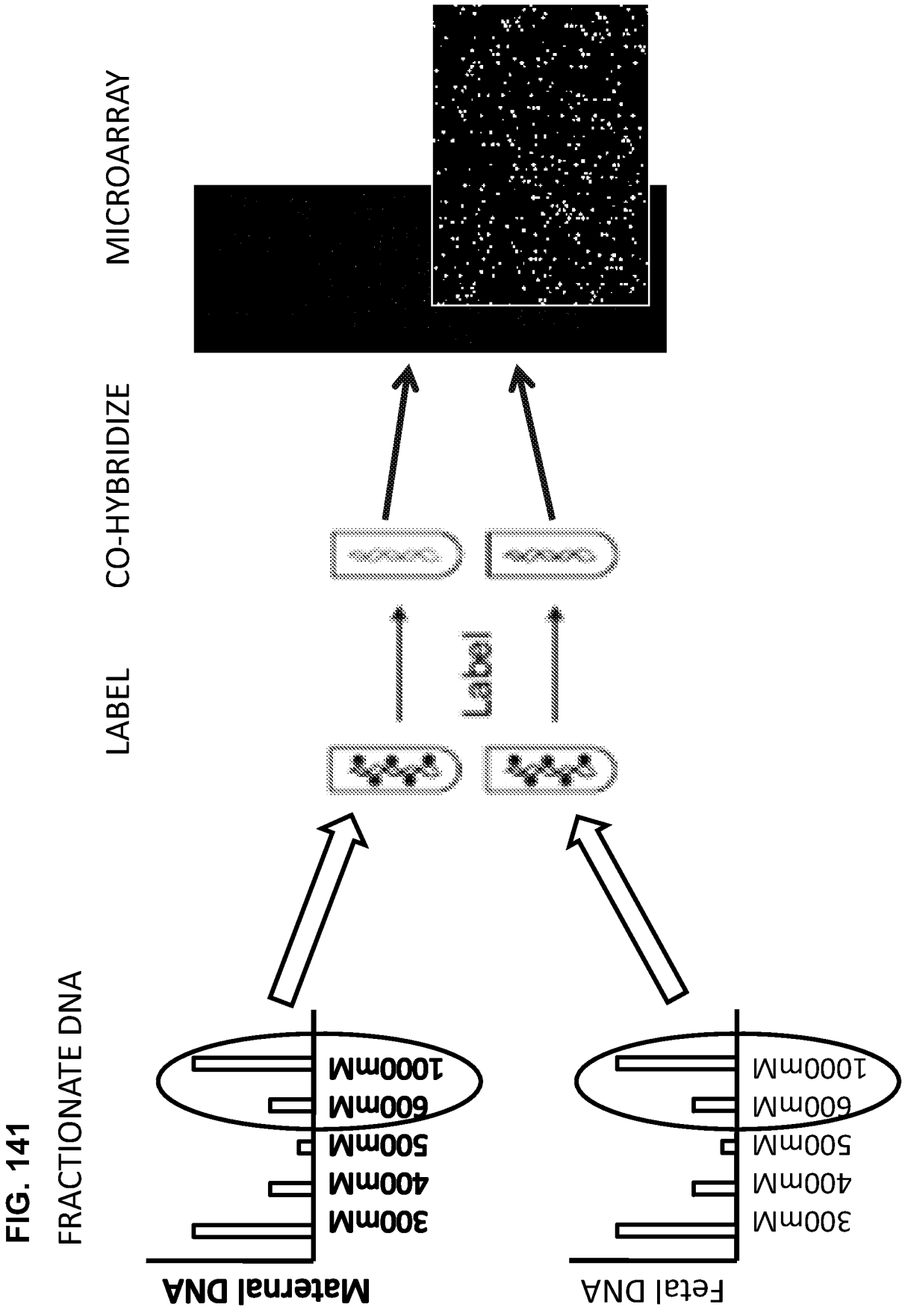

FRACTIONATE DNA
LABEL
CO-HYBRIDIZE
MICROARRAY
Label
Maternal DNA
300mM
400mM
500mM
600mM
1000mM
Fetal DNA
300mM
400mM
500mM
600mM
1000mM

FIG. 141

FIG. 142

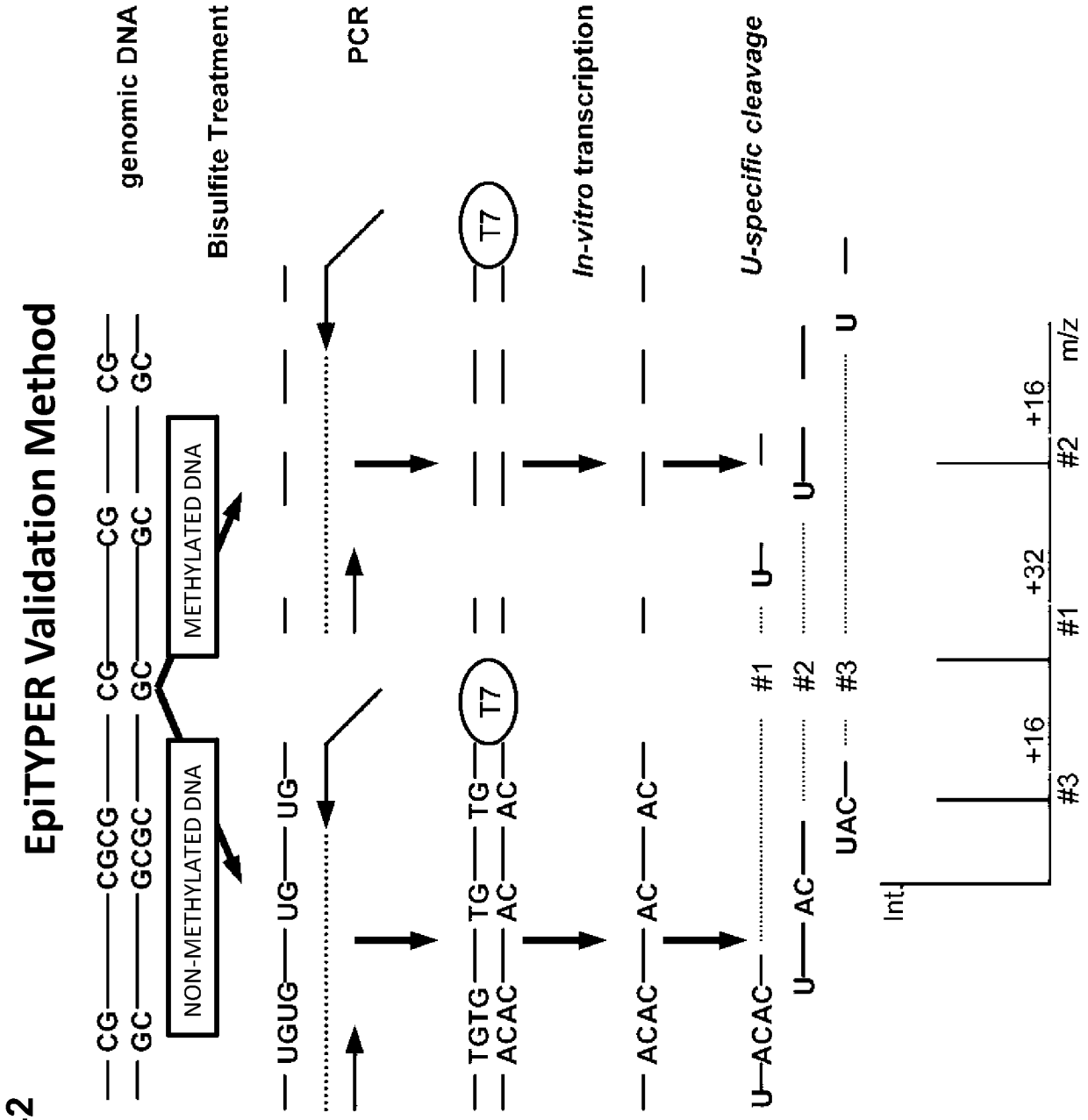
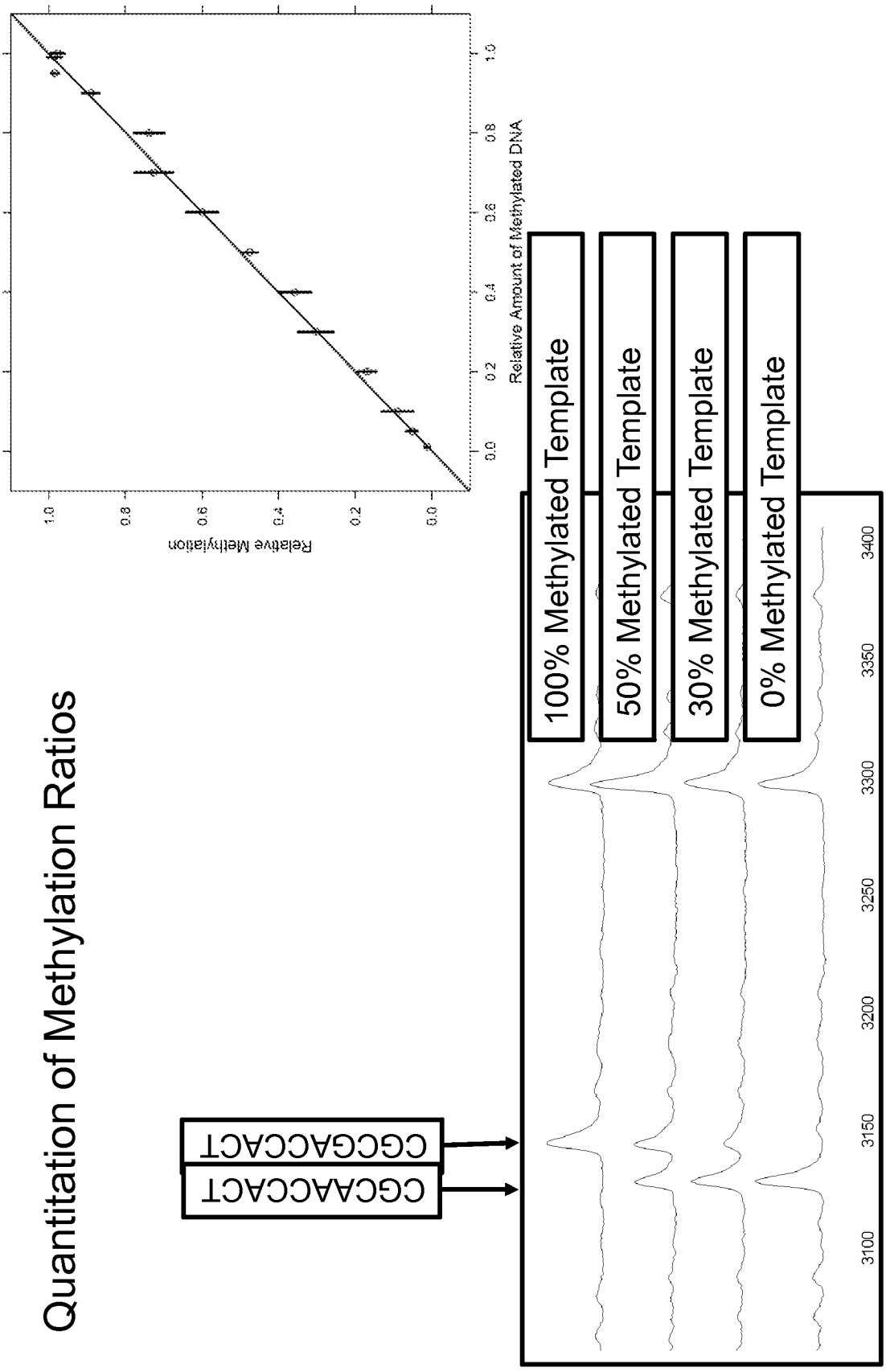

EpiTYPER Validation Method
genomic DNA
Bisulfite Treatment
NON-METHYLATED DNA
METHYLATED DNA
PCR
T7
In-vitro transcription
U-specific cleavage
Int.
+16
+32
+16
#1
#2
#3
m/z

FIG. 143

EpiTYPER Results

Quantitation of Methylation Ratios



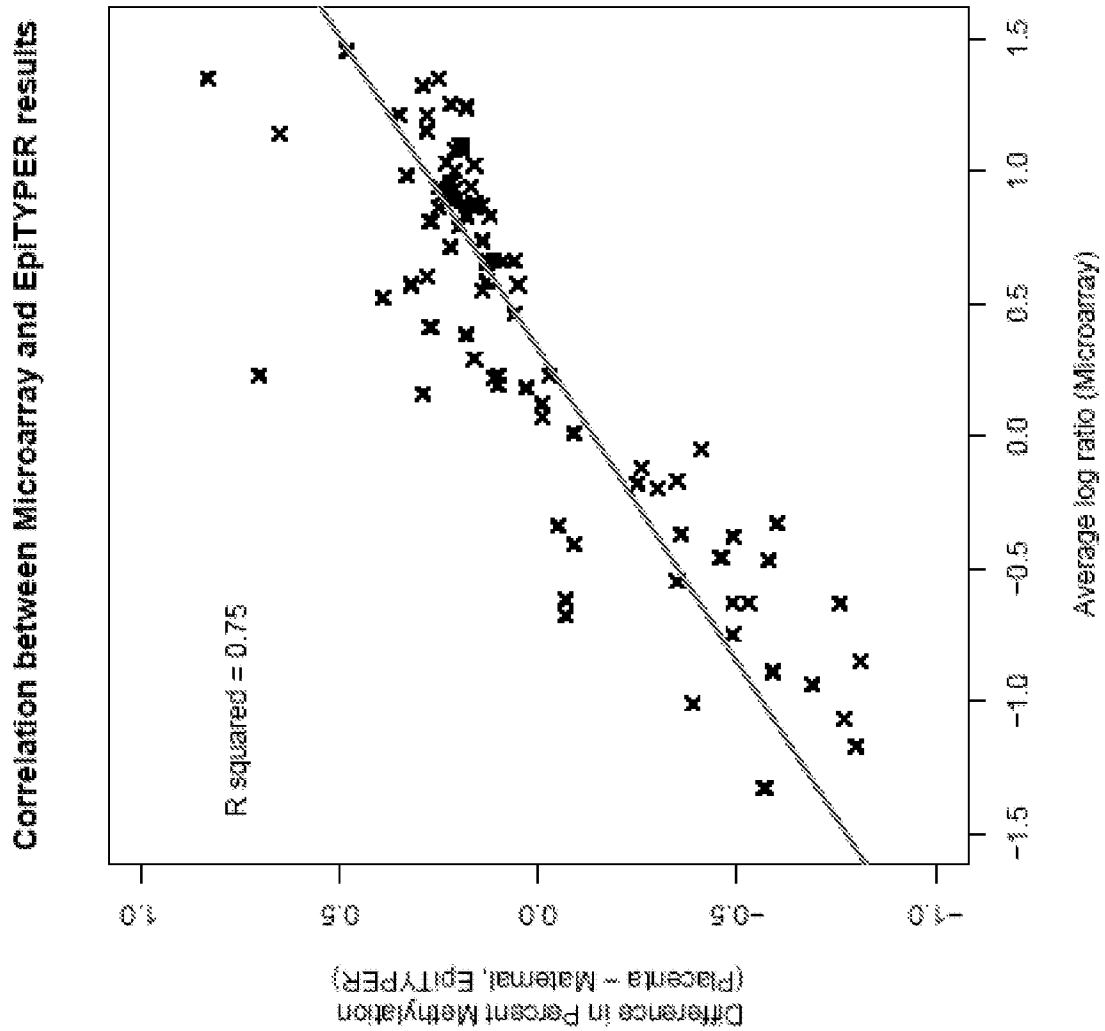

Correlation between Microarray and EpiTYPER results
R squared = 0.75
Difference in Percent Methylation
(Placenta ~ Maternal, EpiTYPER)
Average log ratio (Microarray)
1.0
0.5
0.0
−0.5
−1.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

FIG. 144

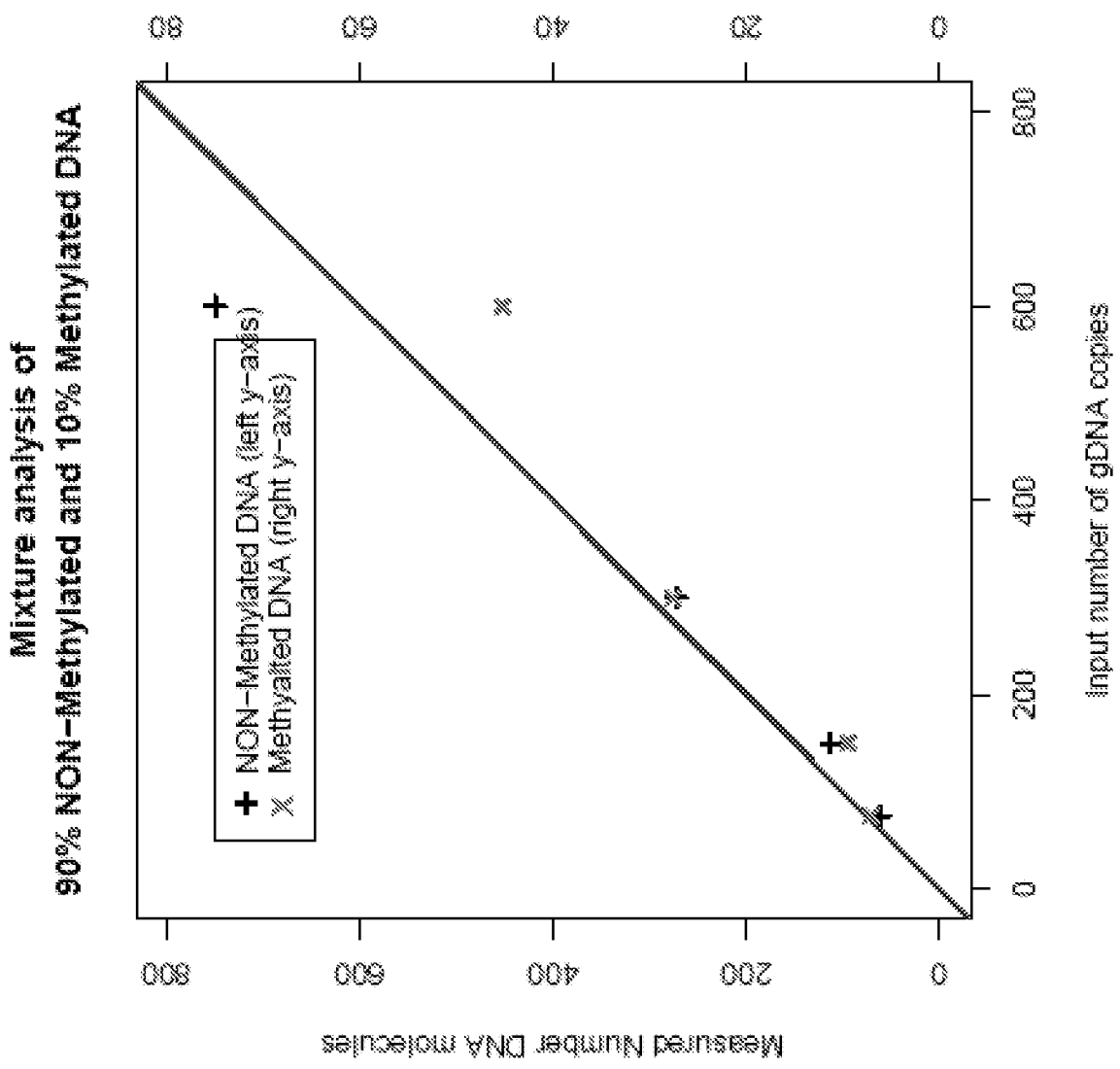

Mixture analysis of
90% NON-Methylated and 10% Methylated DNA
NON-Methylated DNA (left y-axis)
Methyated DNA (right y-axis)
Measured Number DNA molecules
0
200
400
600
800
0
20
40
60
80
input number of gDNA copies
0
200
400
600
800

FIG. 145

FIG. 146A

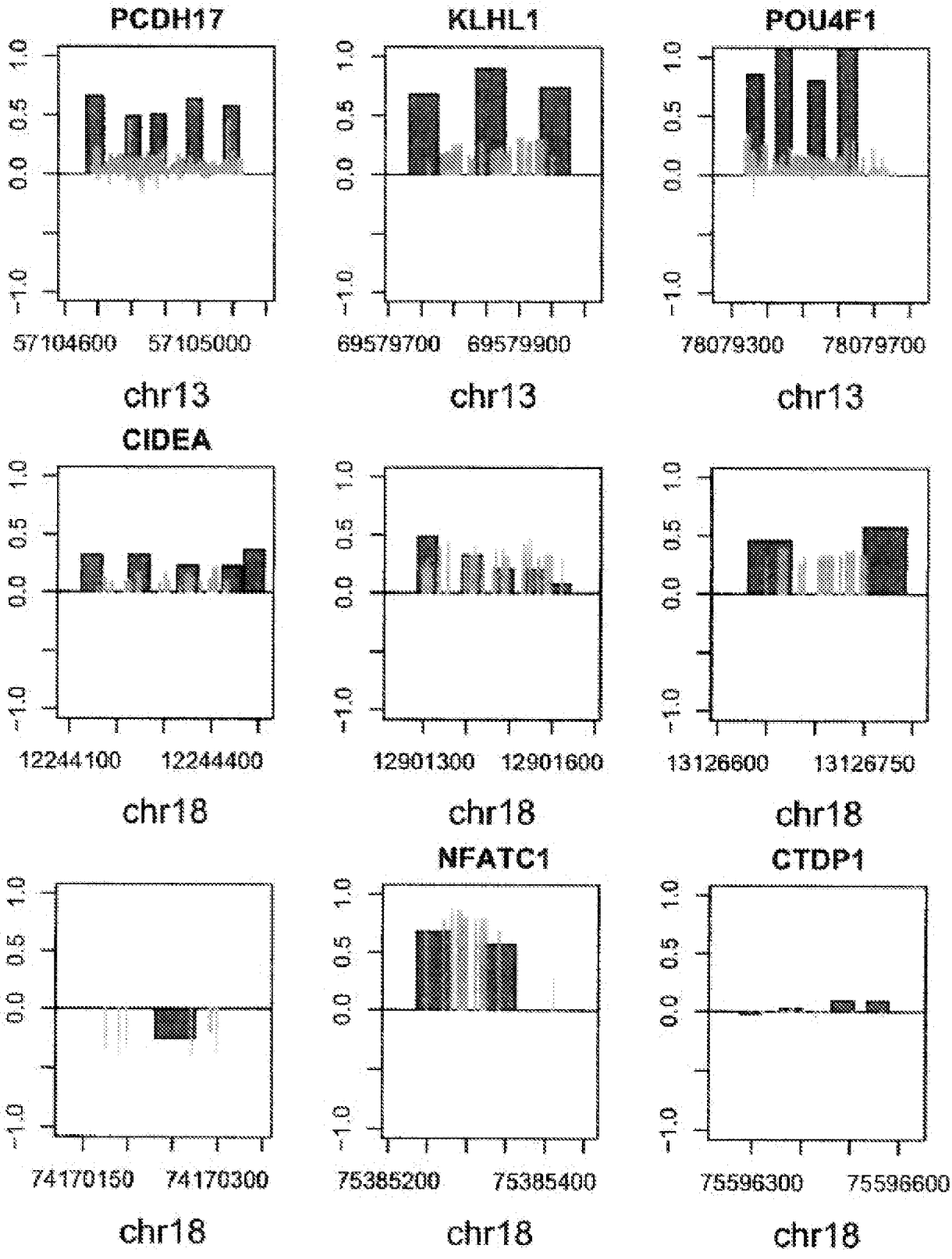

PCDH17
1.0
0.5
0.0
−1.0
57104600
57105000
chr13
KLHL1
69579700
69579900
chr13
POU4F1
78079300
78079700
chr13
CIDEA
12244100
12244400
chr18
12901300
12901600
chr18
13126600
13126750
chr18
74170150
74170300
chr18
NFATC1
75385200
75385400
chr18
CTDP1
75596300
75596600
chr18

FIG. 146B

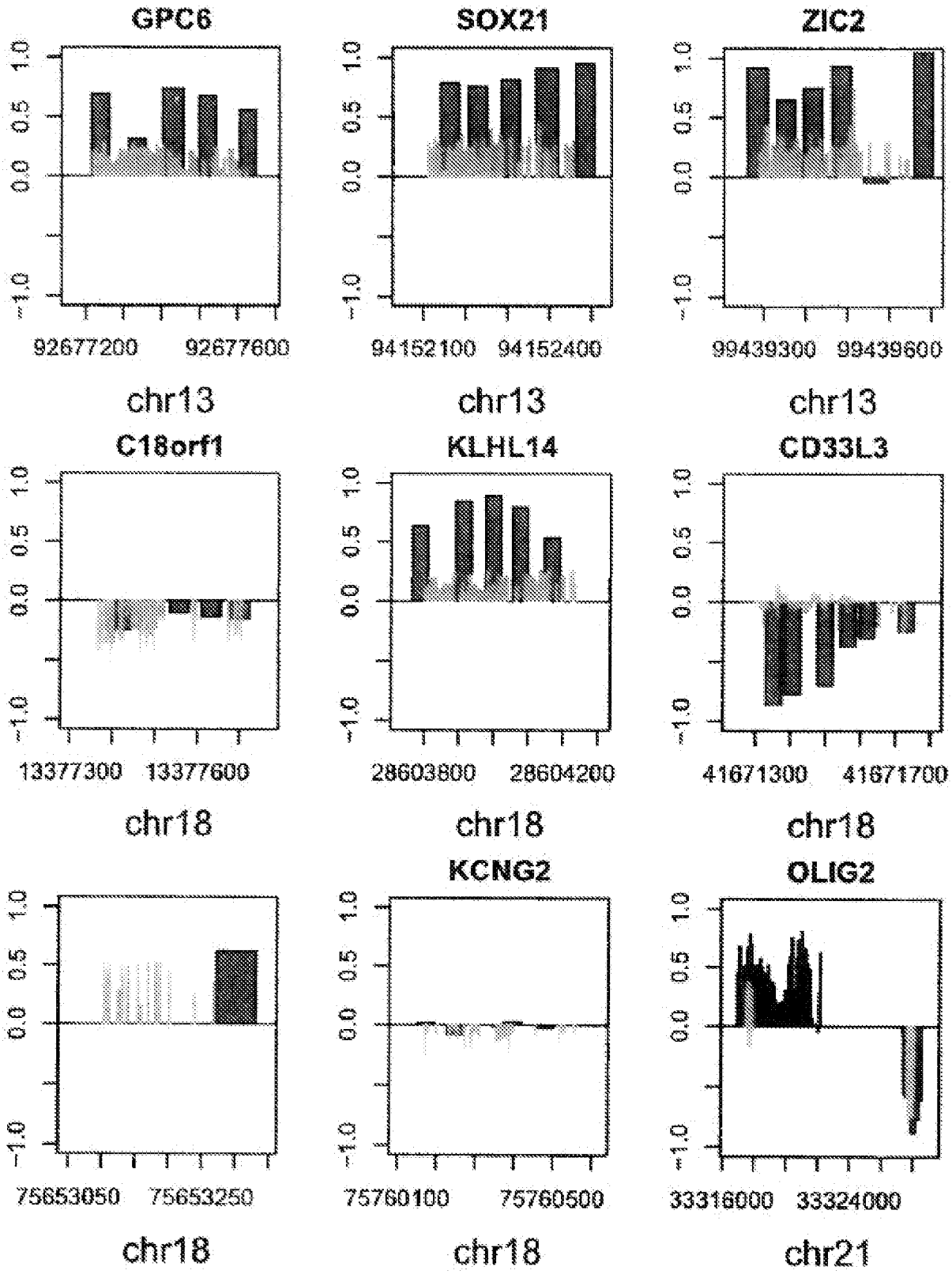

GPC6
1.0
0.5
0.0
-1.0
92677200
92677600
chr13
SOX21
94152100
94152400
chr13
ZIC2
99439300
99439600
chr13
C18orf1
13377300
13377600
chr18
KLHL14
28603800
28604200
chr18
CD33L3
41671300
41671700
chr18
KCNG2
75653050
75653250
chr18
KCNG2
75760100
75760500
chr18
OLIG2
33316000
33324000
chr21

FIG. 146C

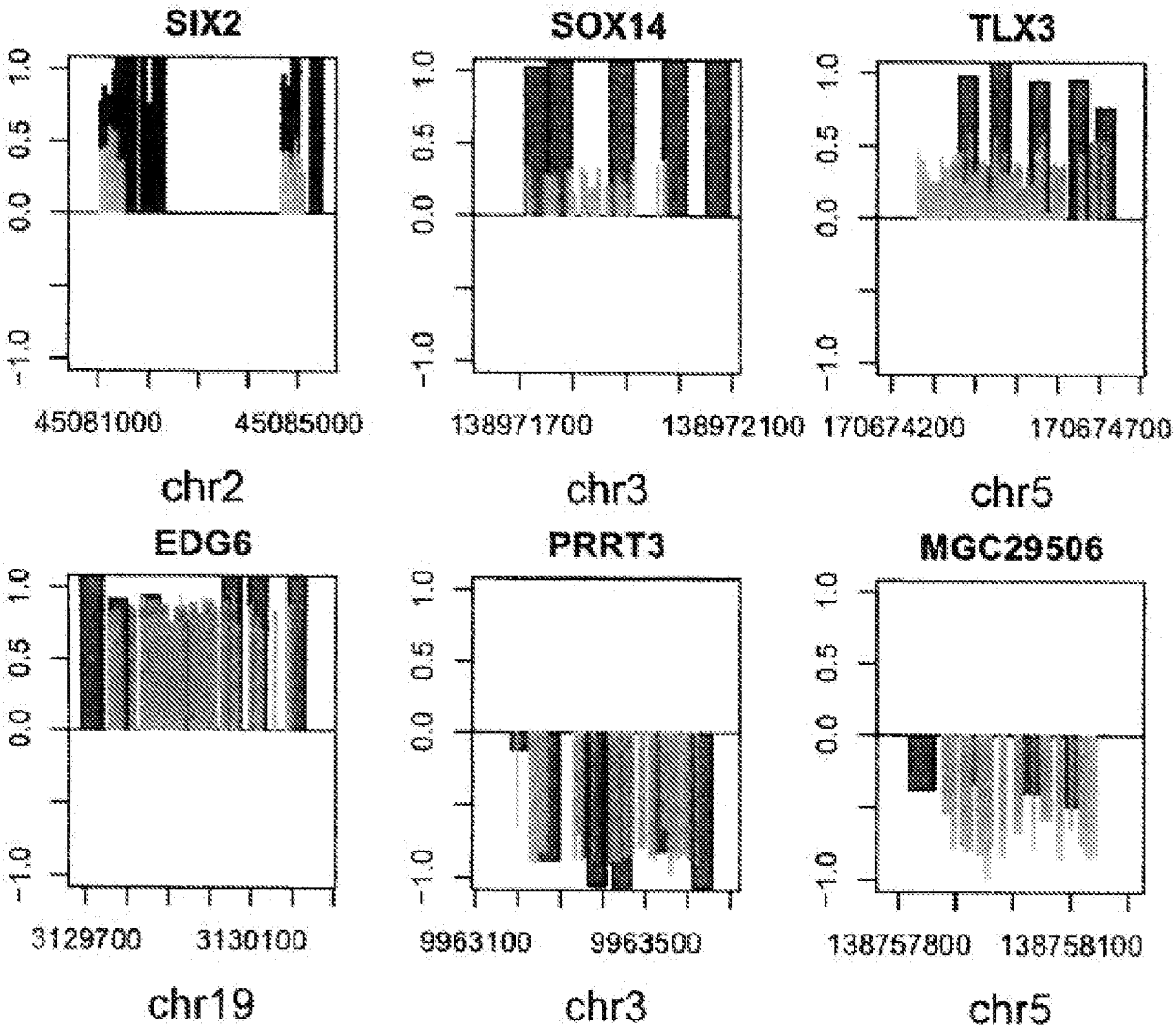

SIX2
1.0
0.5
0.0
-1.0
45081000
45085000
chr2
SOX14
1.0
0.5
0.0
-1.0
138971700
138972100
chr3
TLX3
1.0
0.5
0.0
-1.0
170674200
170674700
chr5
EDG6
1.0
0.5
0.0
-1.0
3129700
3130100
chr19
PRRT3
1.0
0.5
0.0
-1.0
9963100
9963500
chr3
MGC29506
1.0
0.5
0.0
-1.0
138757800
138758100
chr5

FIG. 146D

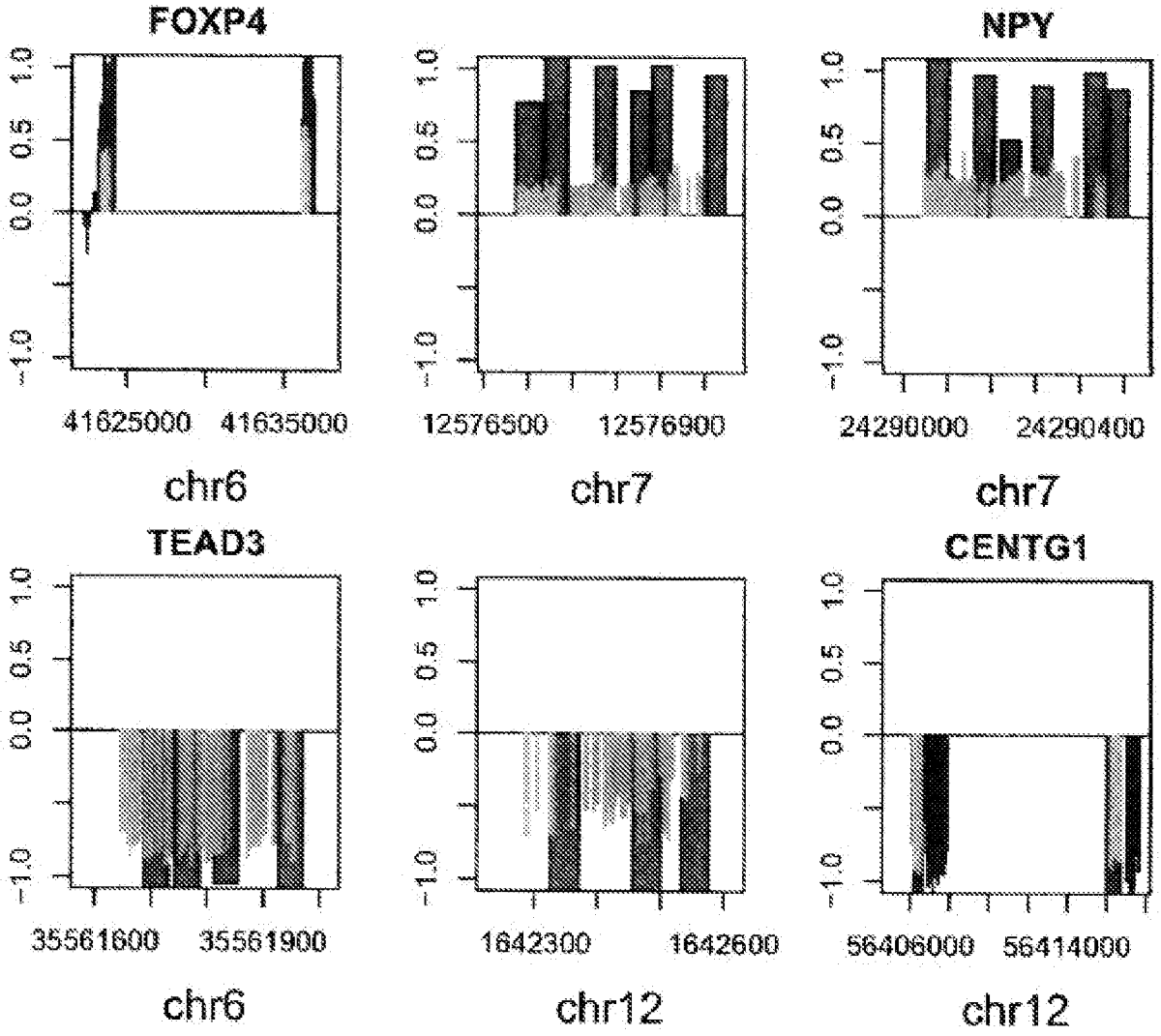

FOXP4
1.0
0.5
0.0
-1.0
41625000
41635000
chr6
12576500
12576900
chr7
NPY
24290000
24290400
chr7
TEAD3
35561600
35561900
chr6
1642300
1642600
chr12
CENTG1
56406000
56414000
chr12

FIG. 146E

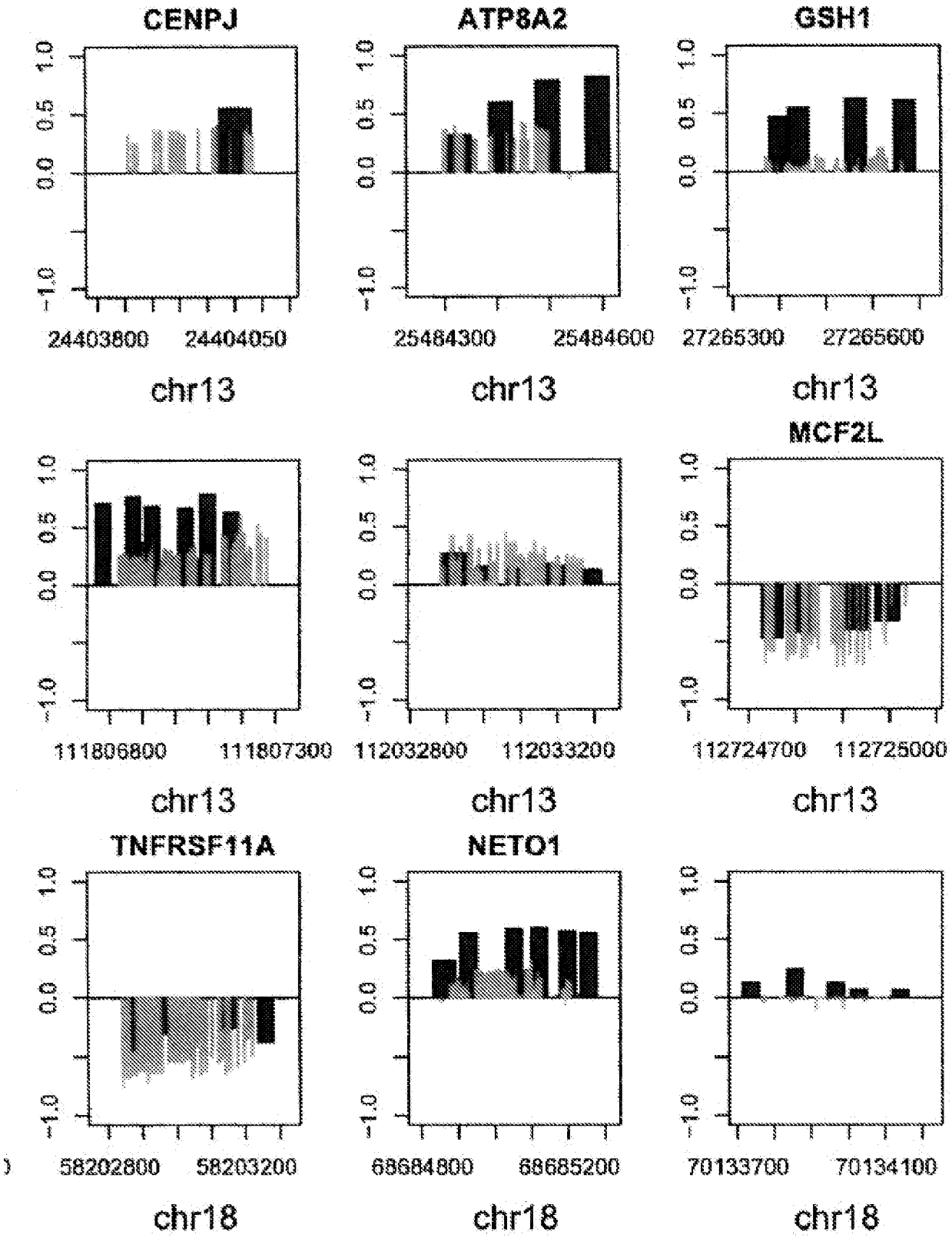

CENPJ
ATP8A2
GSH1
1.0
0.5
0.0
-1.0
24403800
24404050
25484300
25484600
27265300
27265600
chr13
chr13
chr13
MCF2L
111806800
111807300
112032800
112033200
112724700
112725000
chr13
chr13
chr13
TNFRSF11A
NETO1
58202800
58203200
68684800
68685200
70133700
70134100
chr18
chr18
chr18

FIG. 146F

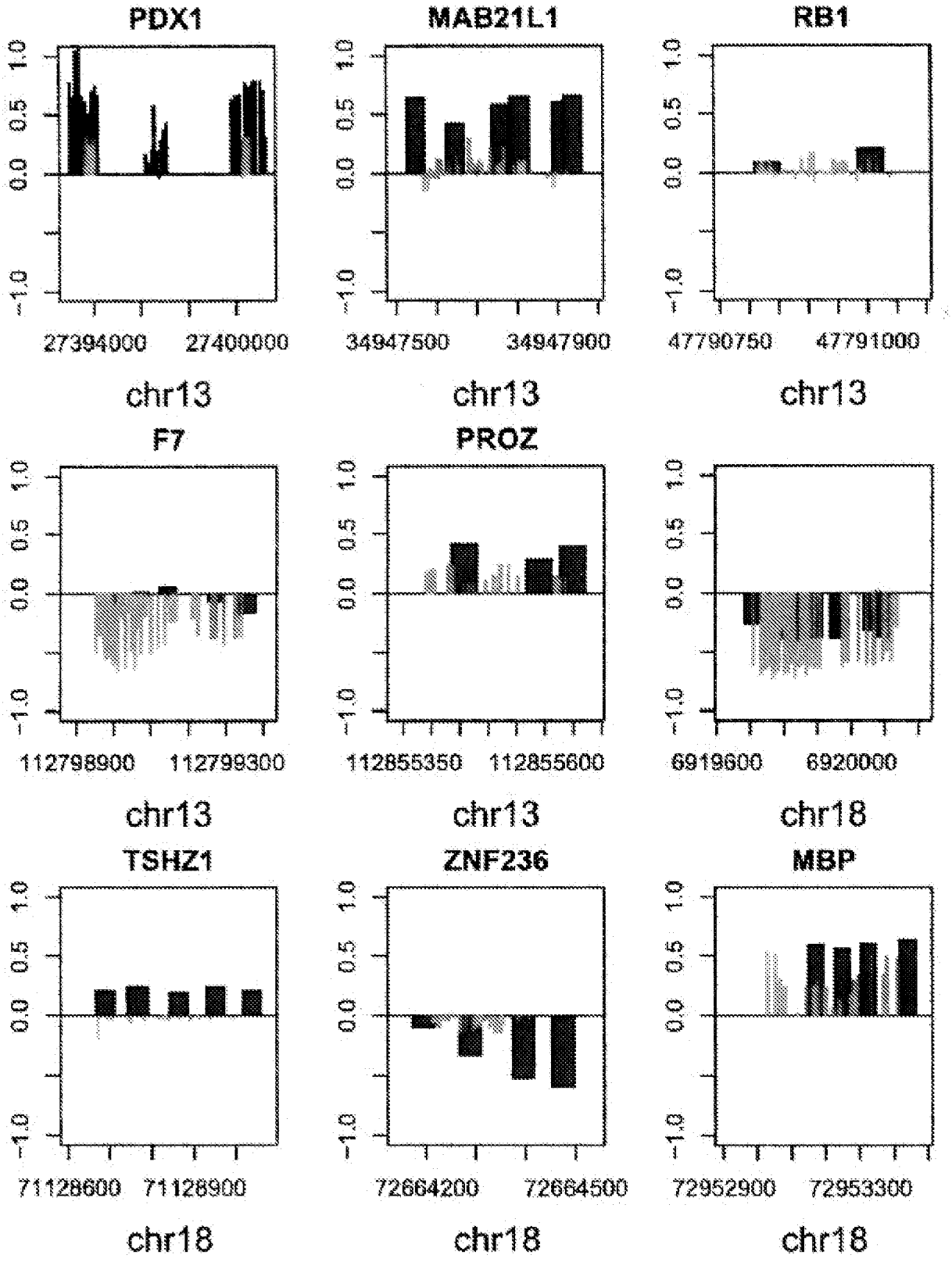

PDX1
1.0
0.5
0.0
-1.0
27394000
27400000
chr13
MAB21L1
34947500
34947900
chr13
RB1
47790750
47791000
chr13
F7
112798900
112799300
chr13
PROZ
112855350
112855600
chr13
6919600
6920000
chr18
TSHZ1
71128600
71128900
chr18
ZNF236
72664200
72664500
chr18
MBP
72952900
72953300
chr18

FIG. 146G

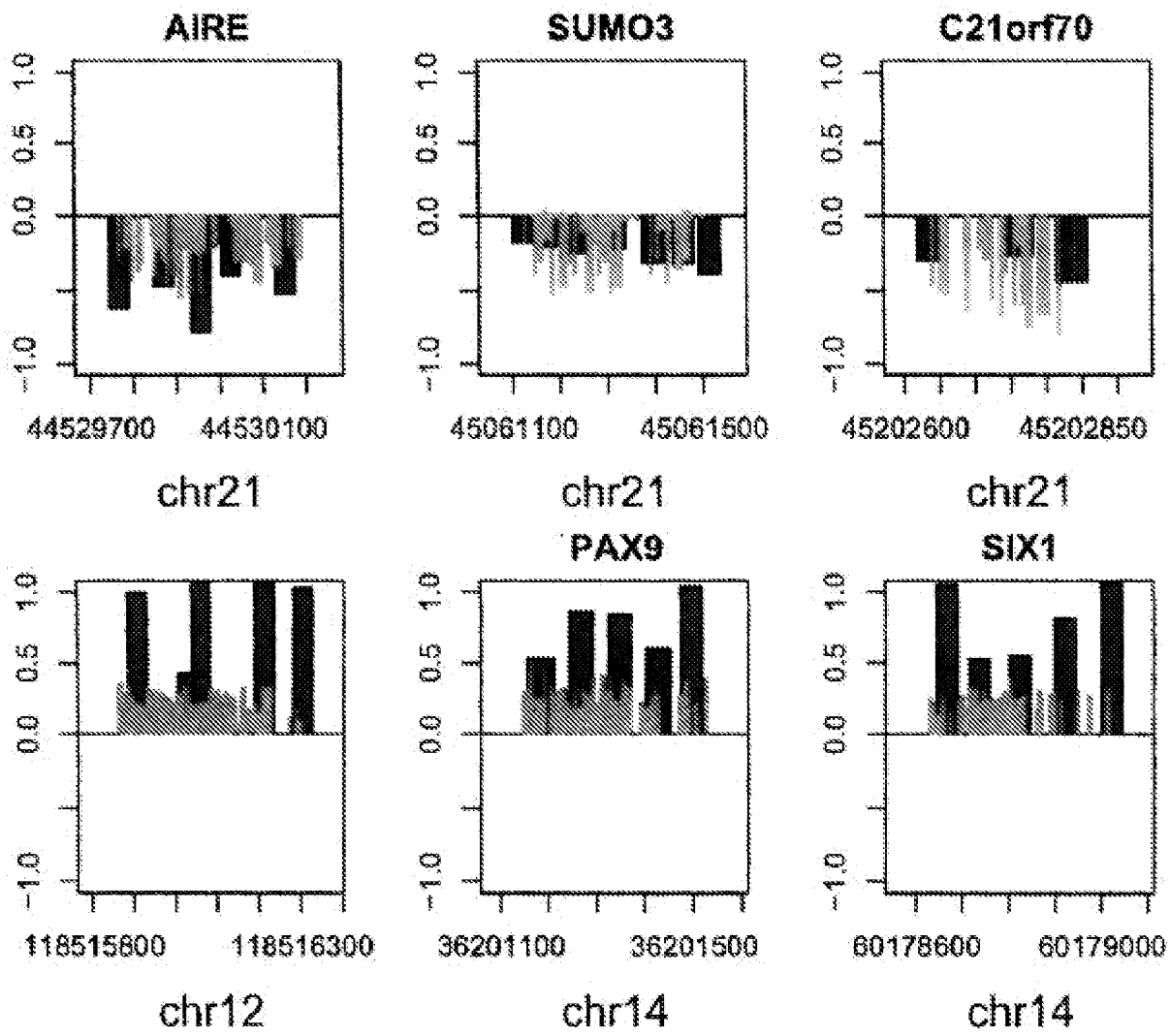

AIRE
1.0
0.5
0.0
-1.0
44529700
44530100
chr21
SUMO3
45061100
45061500
chr21
C21orf70
45202600
45202850
chr21
118515800
118516300
chr12
PAX9
36201100
36201500
chr14
SIX1
60178600
60179000
chr14

FIG. 146H

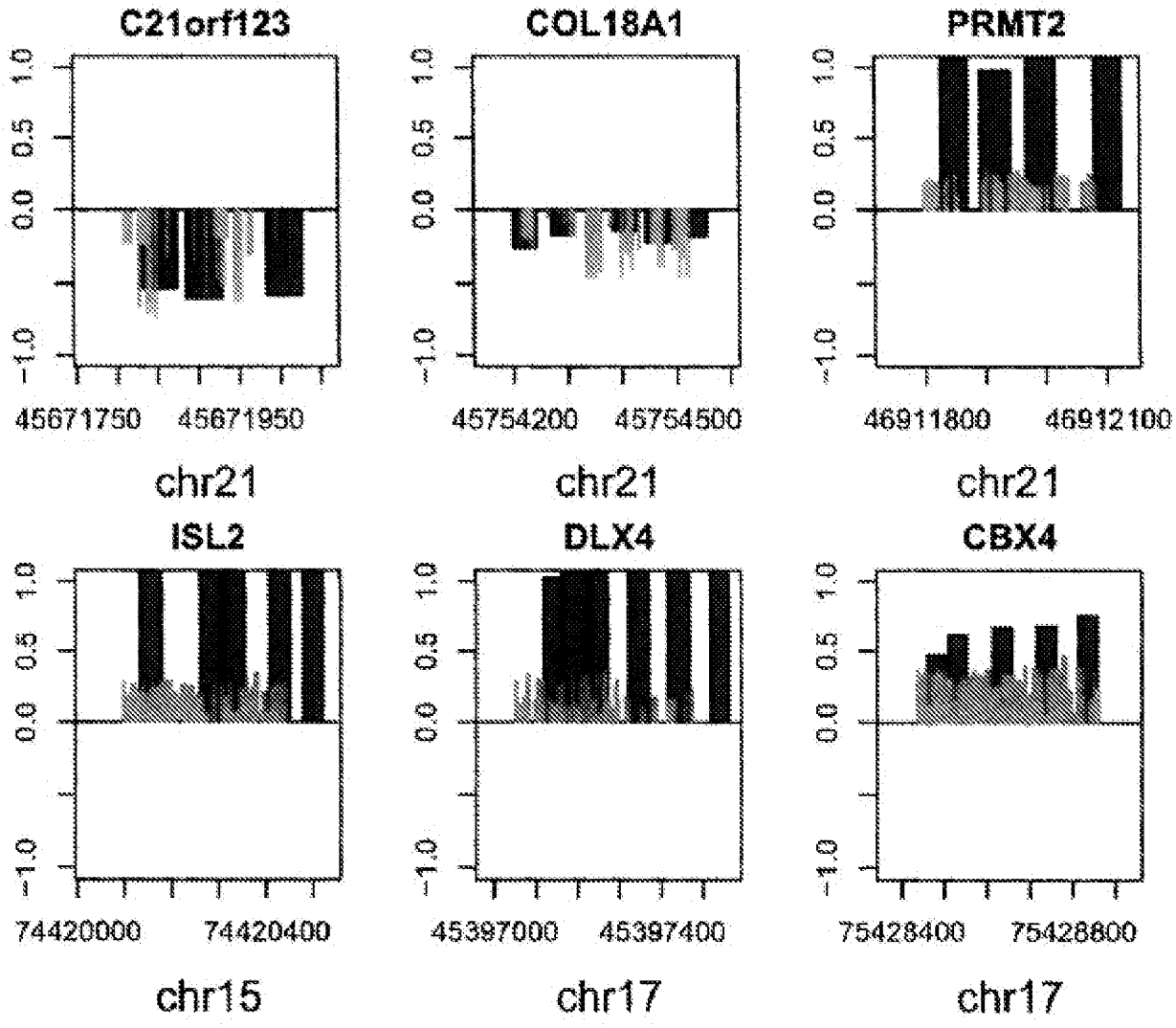

C21orf123
1.0
0.5
0.0
-1.0
45671750
45671950
chr21
COL18A1
45754200
45754500
chr21
PRMT2
46911800
46912100
chr21
ISL2
74420000
74420400
chr15
DLX4
45397000
45397400
chr17
CBX4
75428400
75428800
chr17

FIG. 146I

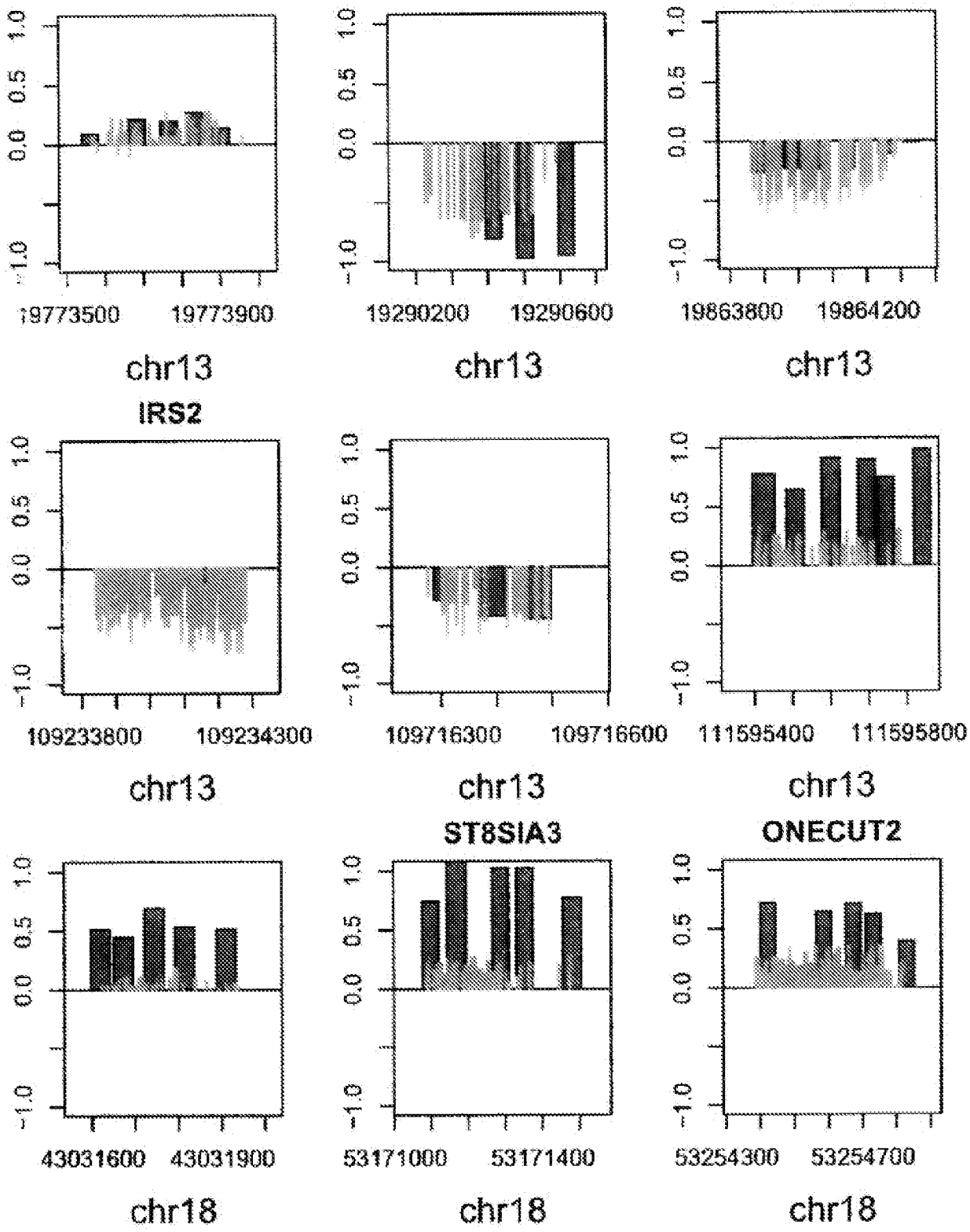

1.0
0.5
0.0
-1.0
19773500
19773900
chr13
19290200
19290600
chr13
19863800
19864200
chr13
IRS2
109233800
109234300
chr13
109716300
109716600
chr13
111595400
111595800
chr13
ST8SIA3
43031600
43031900
chr18
53171000
53171400
chr18
ONECUT2
53254300
53254700
chr18

FIG. 146J

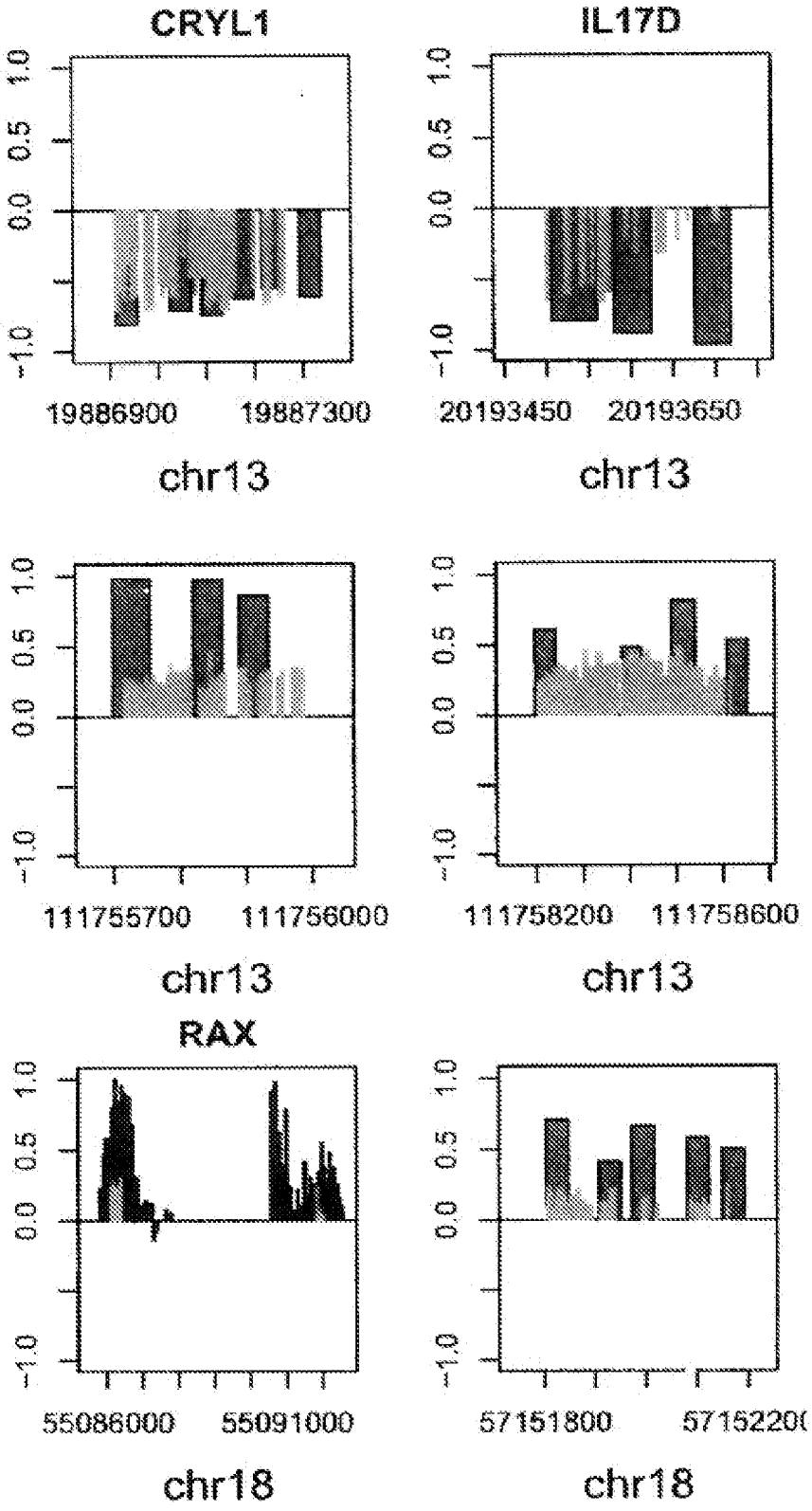

CRYL1
1.0
0.5
0.0
-1.0
19886900
19887300
chr13
IL17D
1.0
0.5
0.0
-1.0
20193450
20193650
chr13
1.0
0.5
0.0
-1.0
111755700
111756000
chr13
1.0
0.5
0.0
-1.0
111758200
111758600
chr13
RAX
1.0
0.5
0.0
-1.0
55086000
55091000
chr18
1.0
0.5
0.0
-1.0
57151800
chr18

FIG. 146K

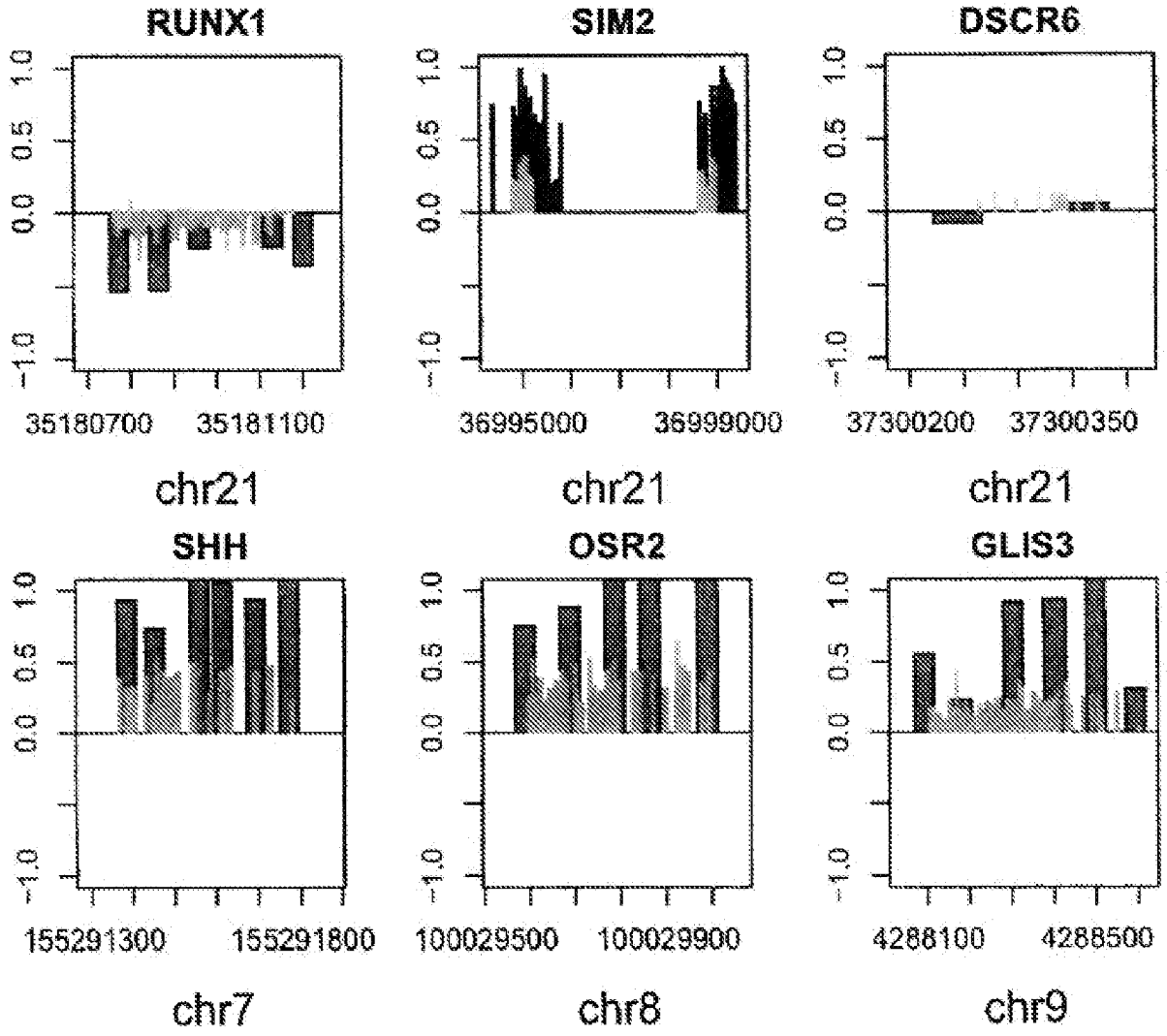

RUNX1
1.0
0.5
0.0
−1.0
35180700
35181100
chr21
SIM2
1.0
0.5
0.0
−1.0
36995000
36999000
chr21
DSCR6
1.0
0.5
0.0
−1.0
37300200
37300350
chr21
SHH
1.0
0.5
0.0
−1.0
155291300
155291800
chr7
OSR2
1.0
0.5
0.0
−1.0
100029500
100029900
chr8
GLIS3
1.0
0.5
0.0
−1.0
4288100
4288500
chr9

FIG. 146L

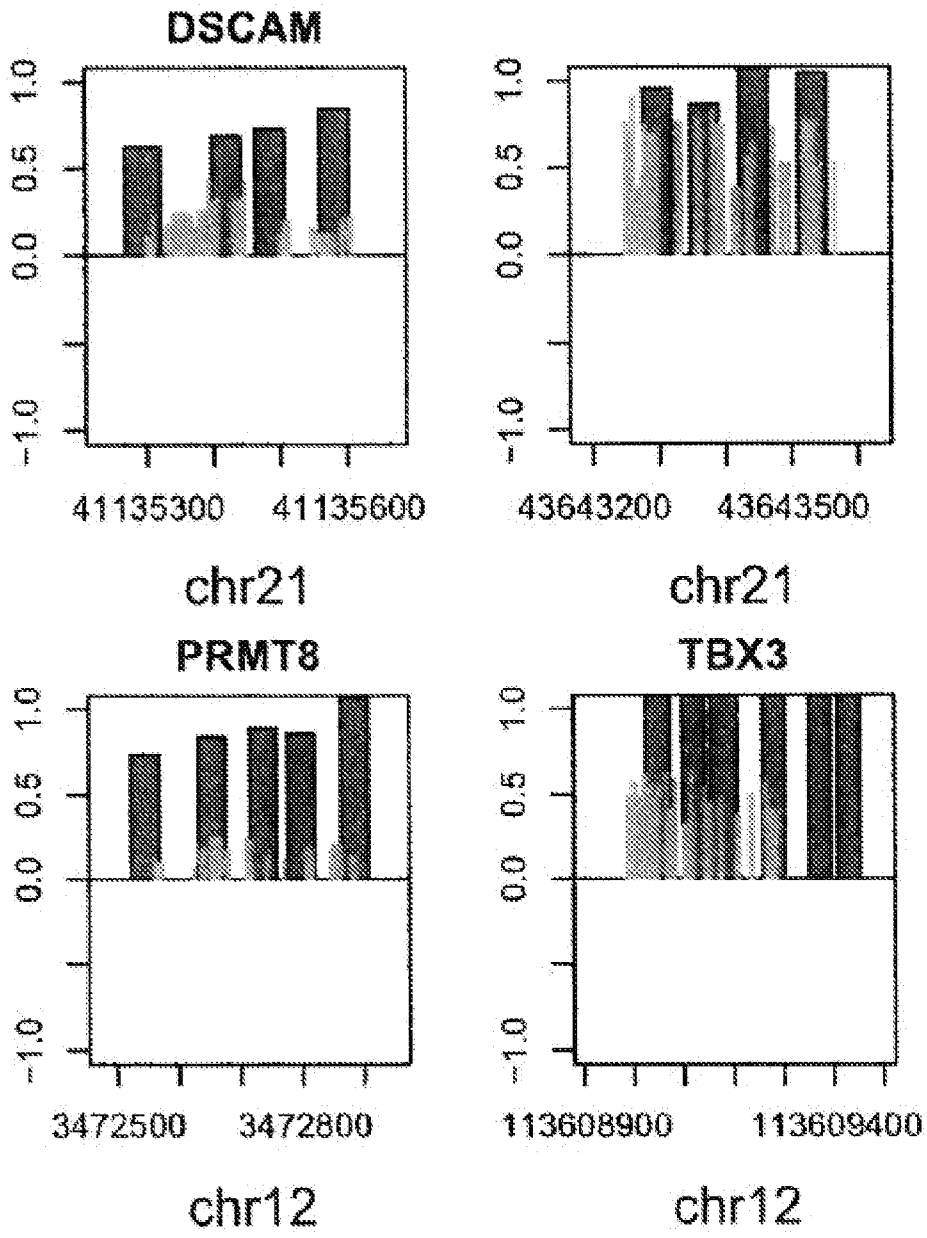

DSCAM
1.0
0.5
0.0
−1.0
41135300
41135600
chr21
43643200
43643500
chr21
PRMT8
3472500
3472800
chr12
TBX3
113608900
113609400
chr12

FIG. 147

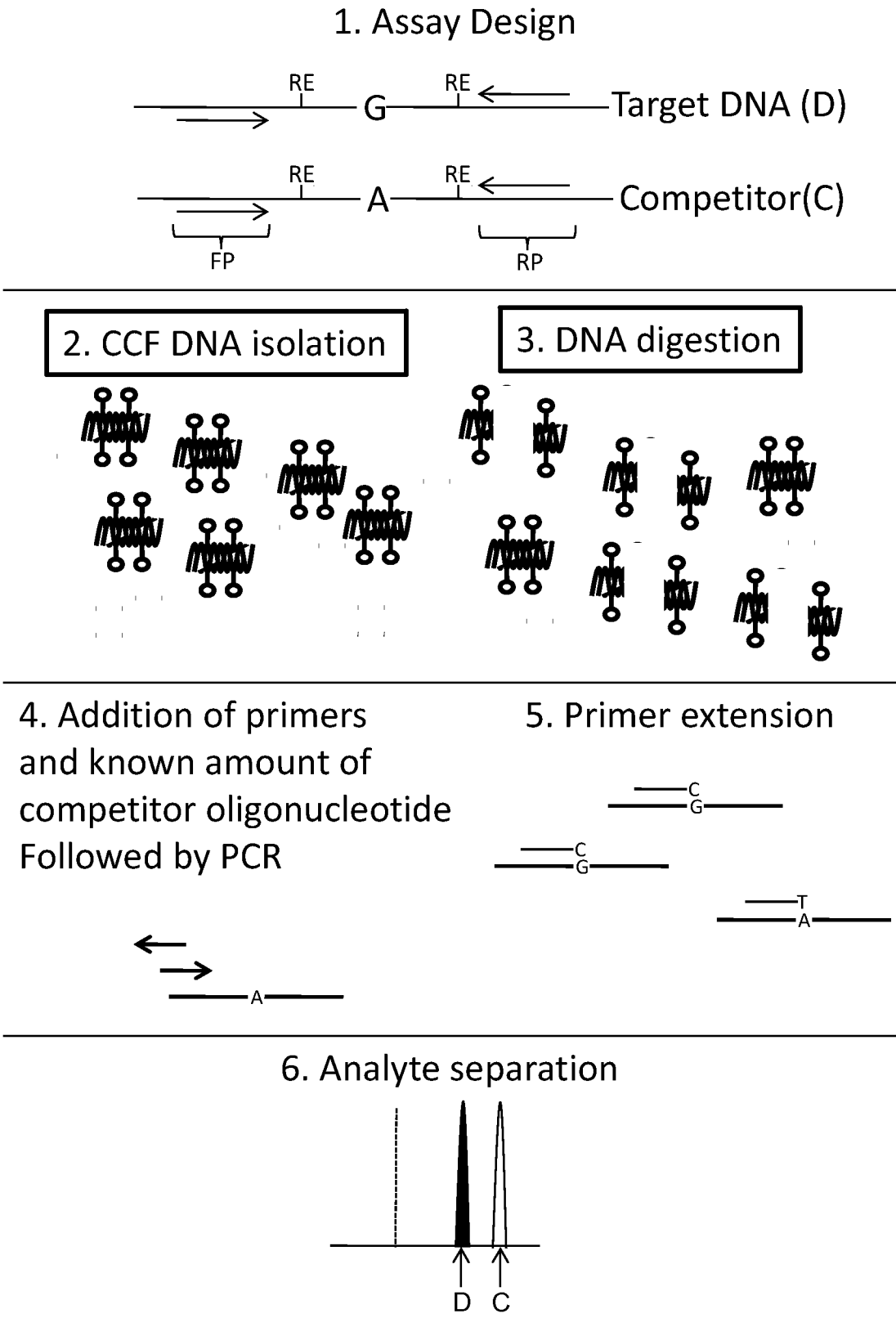
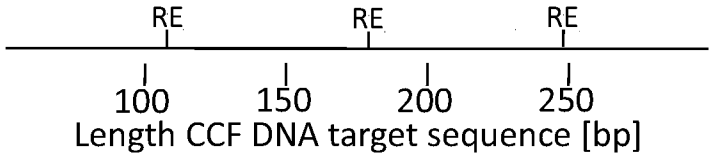
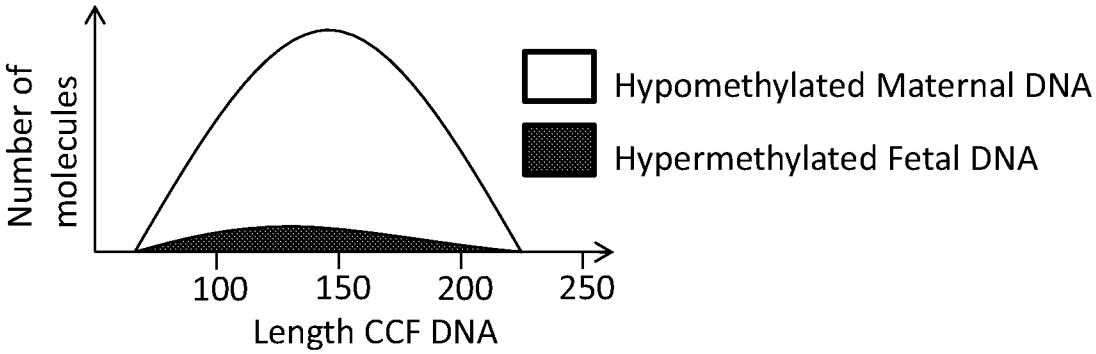

1. Assay Design
RE
RE
G
Target DNA (D)
RE
RE
A
Competitor(C)
FP
RP
2. CCF DNA isolation
3. DNA digestion
4. Addition of primers and known amount of competitor oligonucleotide Followed by PCR
A
5. Primer extension
C
G
C
G
T
A
6. Analyte separation
D
C

FIG. 148

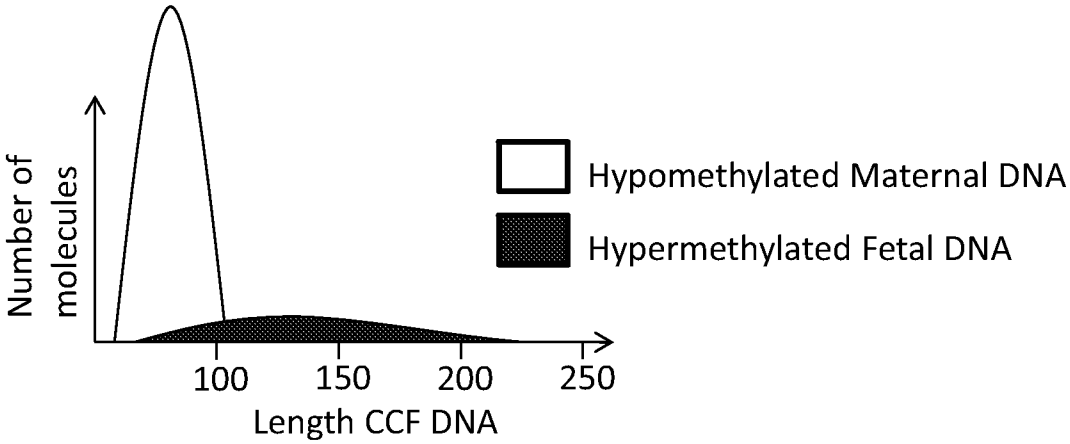
1. Selection of differentially methylated targets for specific DNA sequence capture



2. Distribution of CCF DNA after capture



3. Distribution of CCF DNA after digestion



4. Quantification of non-digested DNA molecules

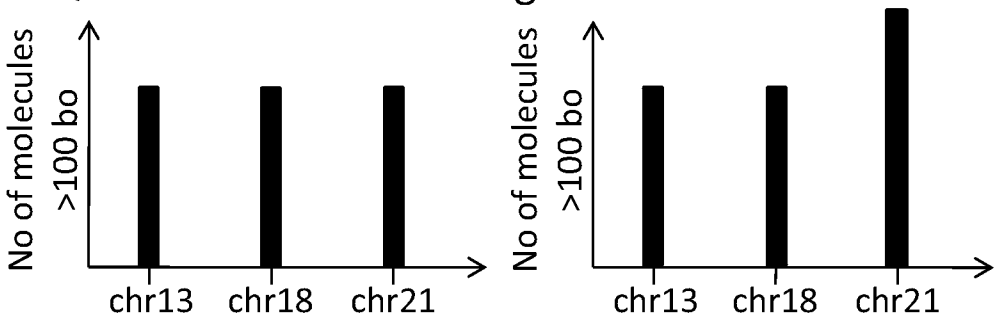


FIG. 149

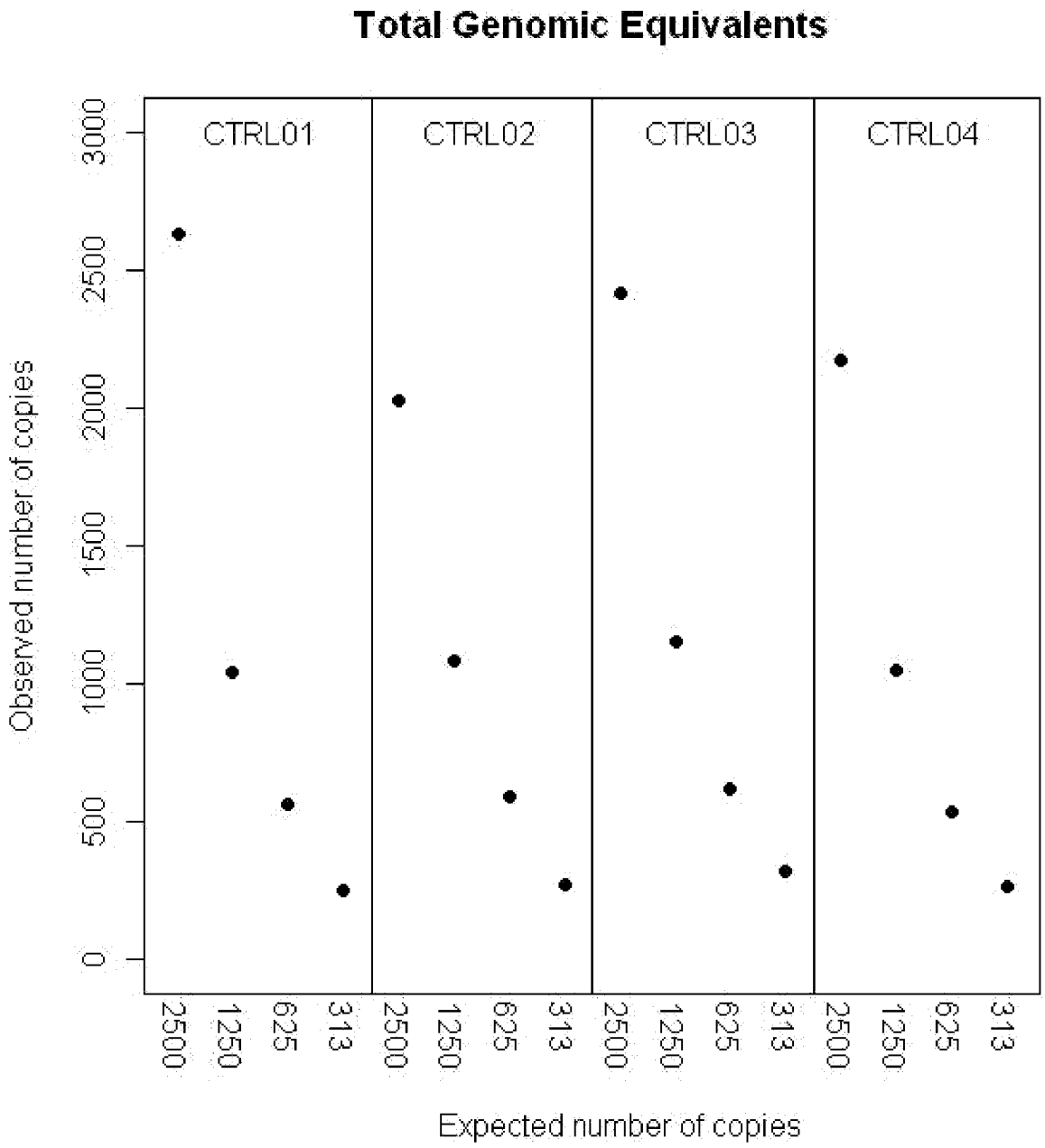

Total Genomic Equivalents
CTRL01
CTRL02
CTRL03
CTRL04
Observed number of copies
0
500
1000
1500
2000
2500
3000
2500
1250
625
313
Expected number of copies

FIG. 150A

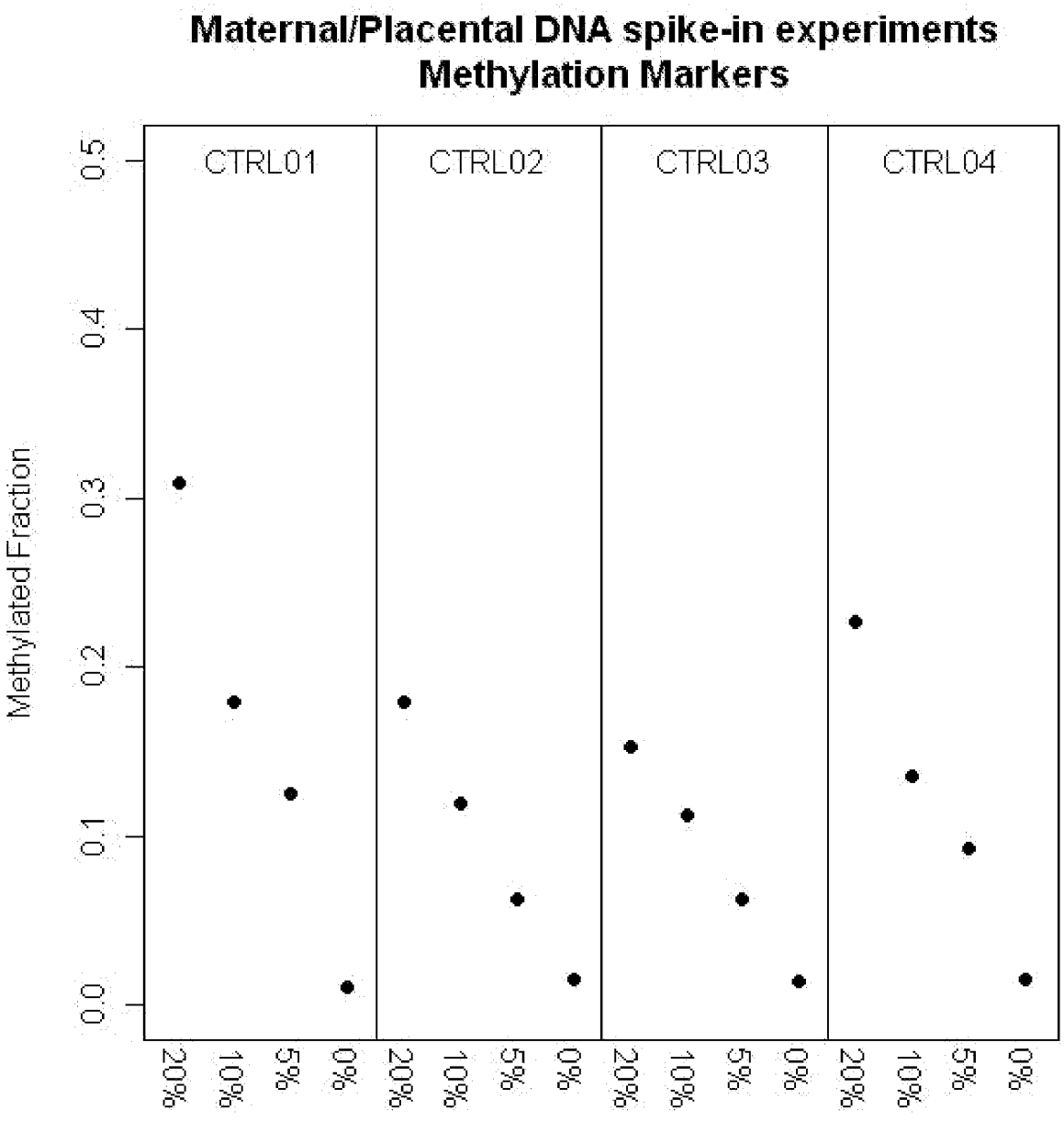

Maternal/Placental DNA spike-in experiments
Methylation Markers
Methylated Fraction
0.0
0.1
0.2
0.3
0.4
0.5
CTRL01
CTRL02
CTRL03
CTRL04
20%
10%
5%
0%

FIG. 150B

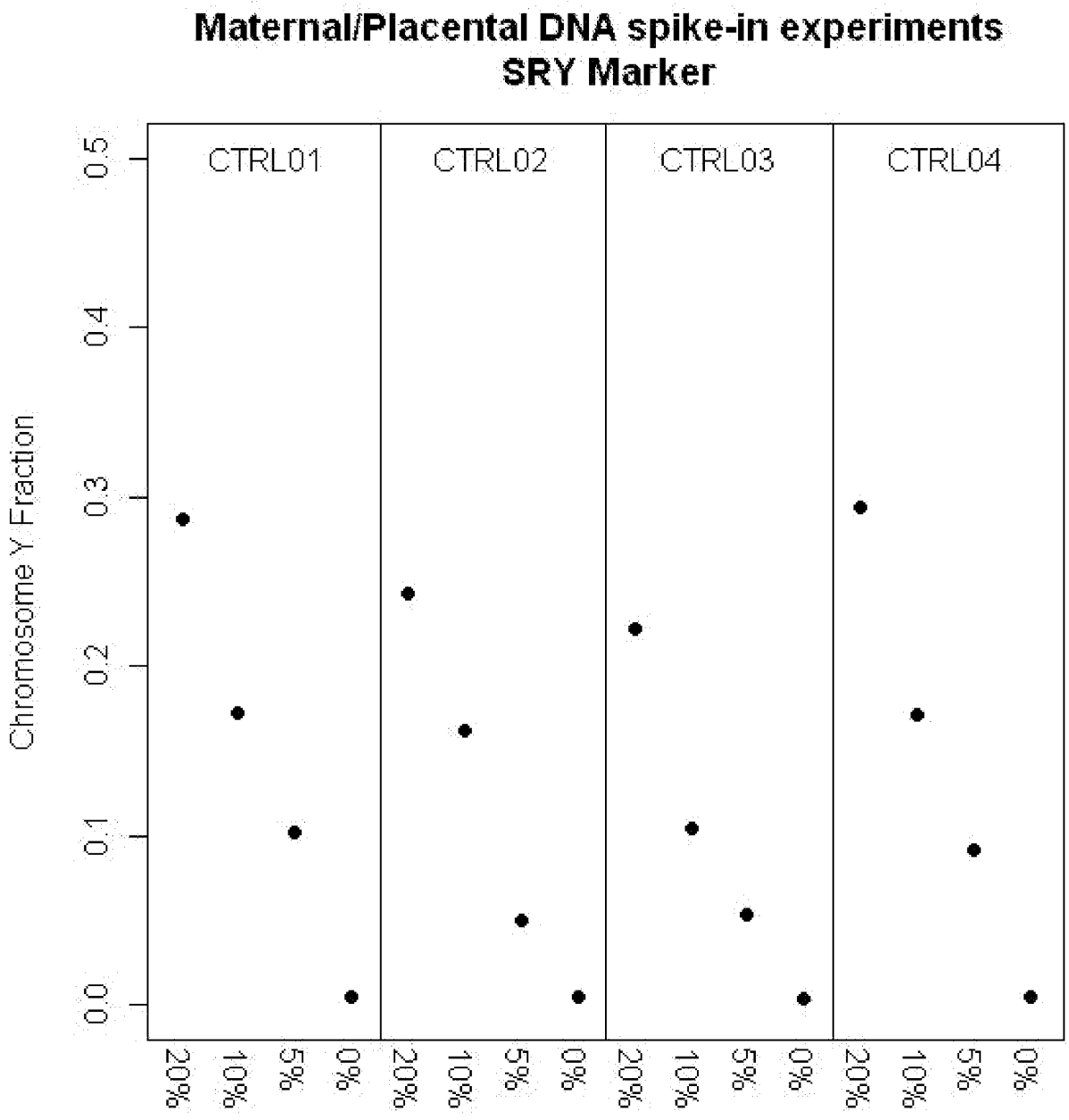

Maternal/Placental DNA spike-in experiments
SRY Marker
Chromosome Y Fraction
0.0
0.1
0.2
0.3
0.4
0.5
CTRL01
CTRL02
CTRL03
CTRL04
20%
10%
5%
0%

FIG. 151A-B

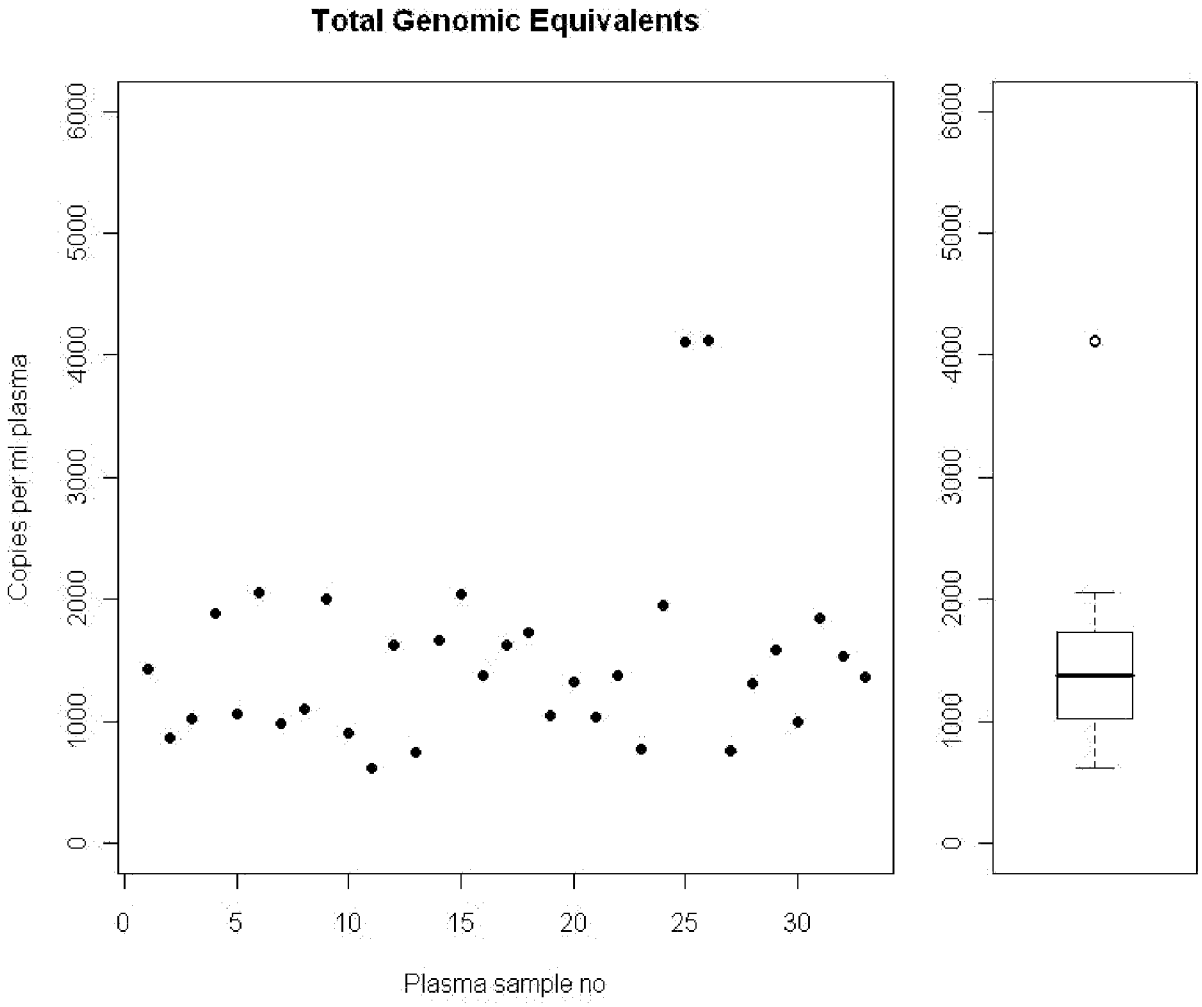


Figure 151A

Figure 151B

FIG. 152A

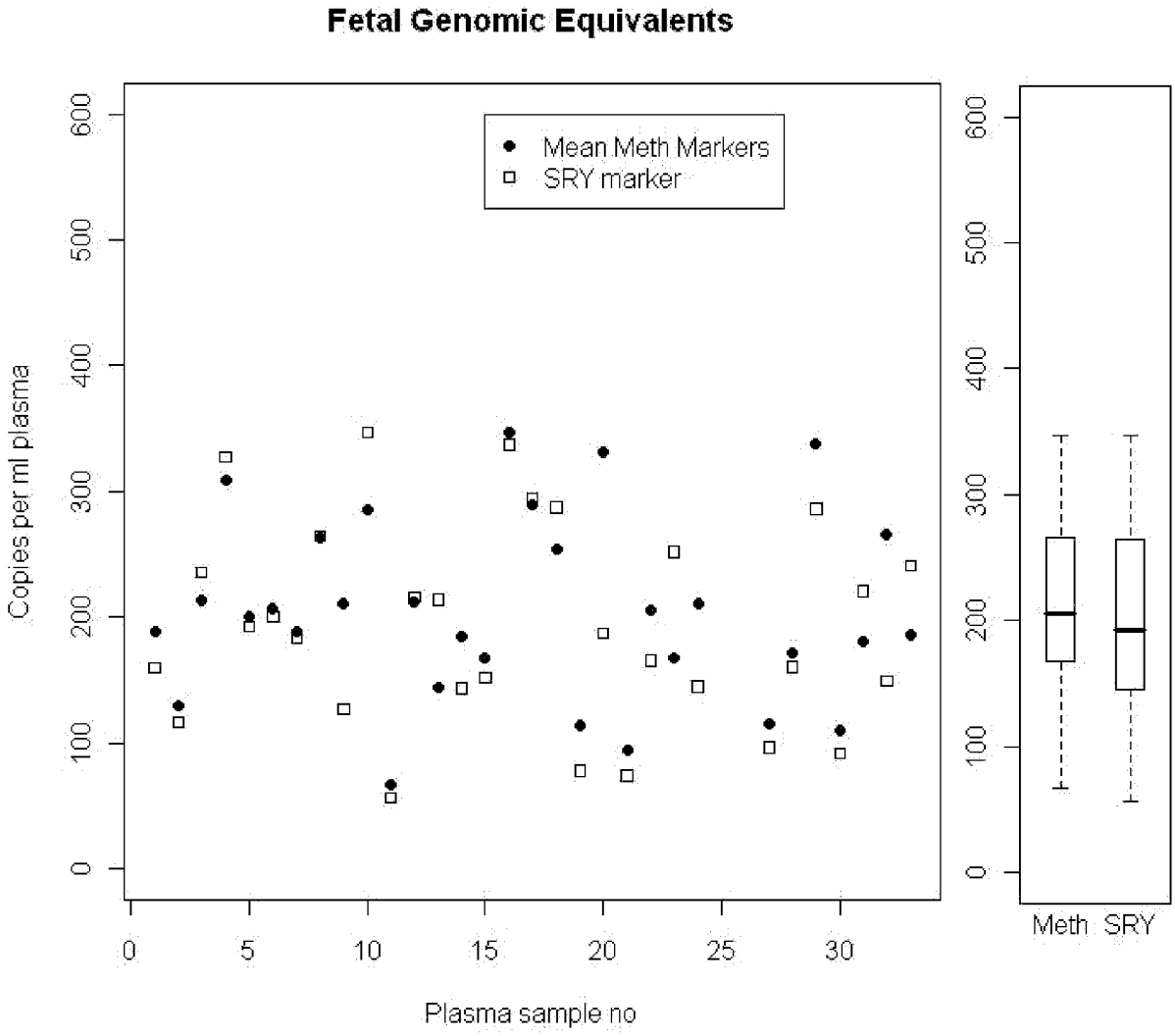


Figure 152A

Figure 152B

FIG. 153

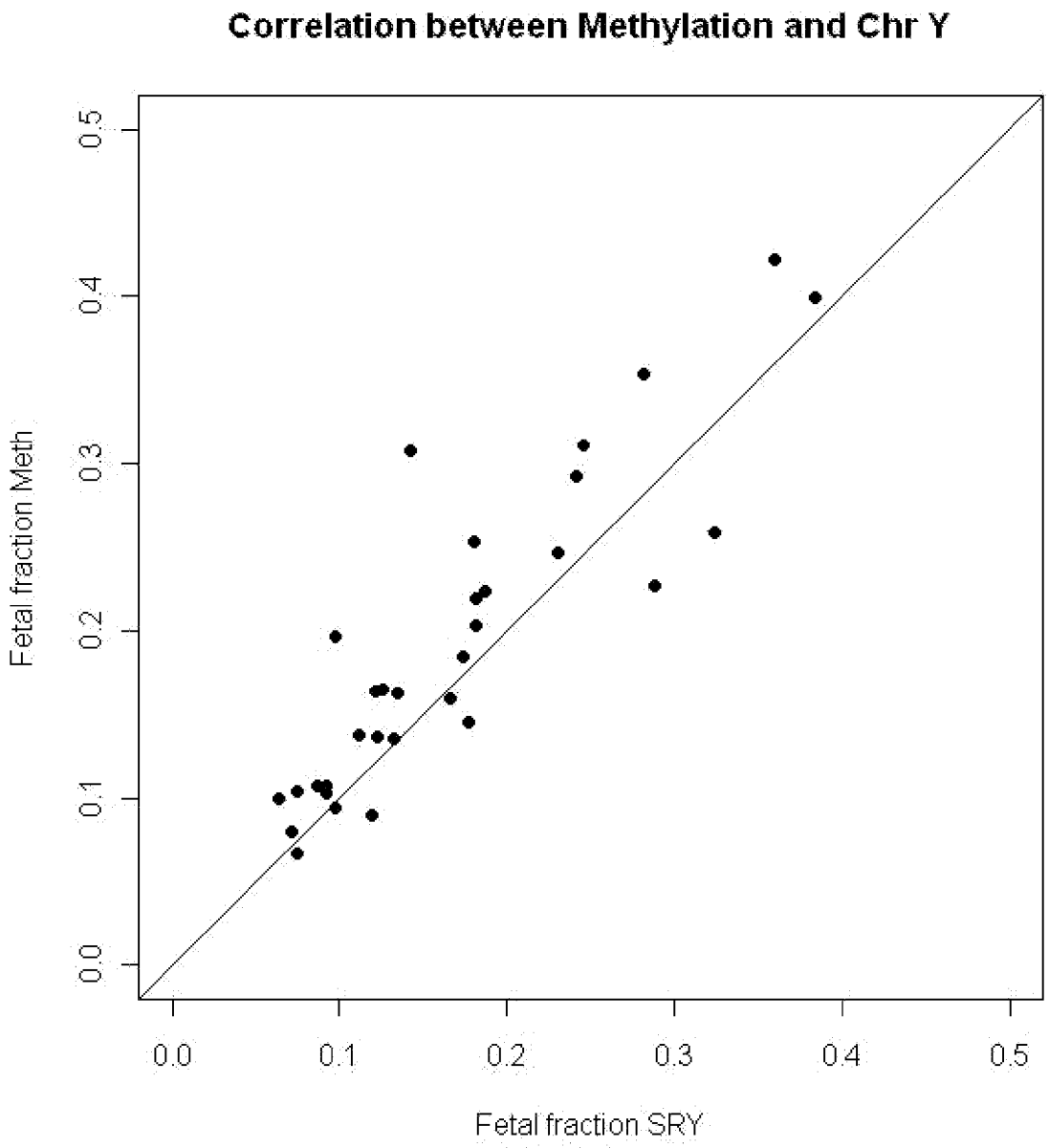

Correlation between Methylation and Chr Y
Fetal fraction Meth
Fetal fraction SRY
0.0
0.1
0.2
0.3
0.4
0.5

FIG. 154

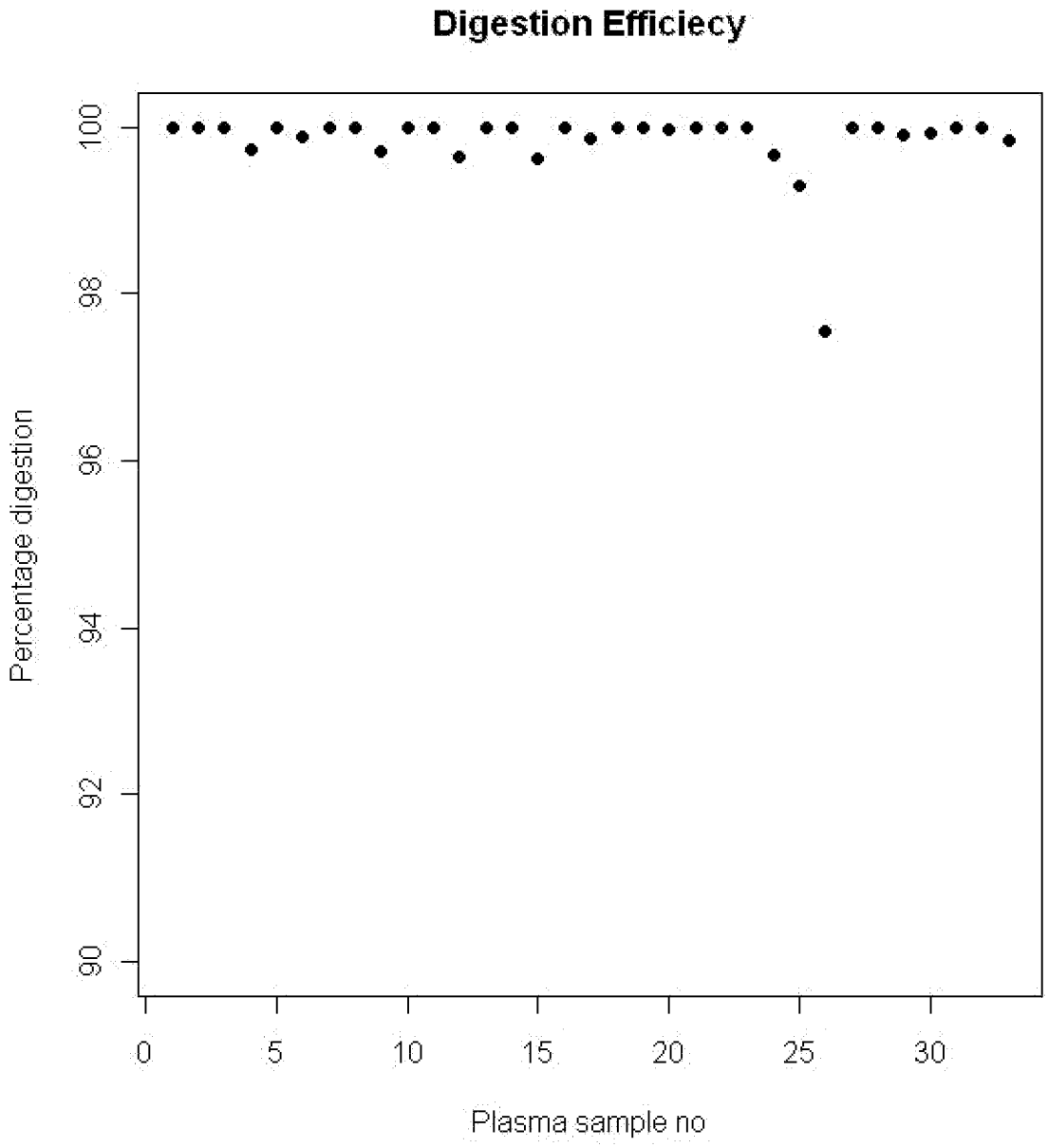

Digestion Efficiecy
Percentage digestion
100
98
96
94
92
90
0
5
10
15
20
25
30
Plasma sample no

FIG. 155

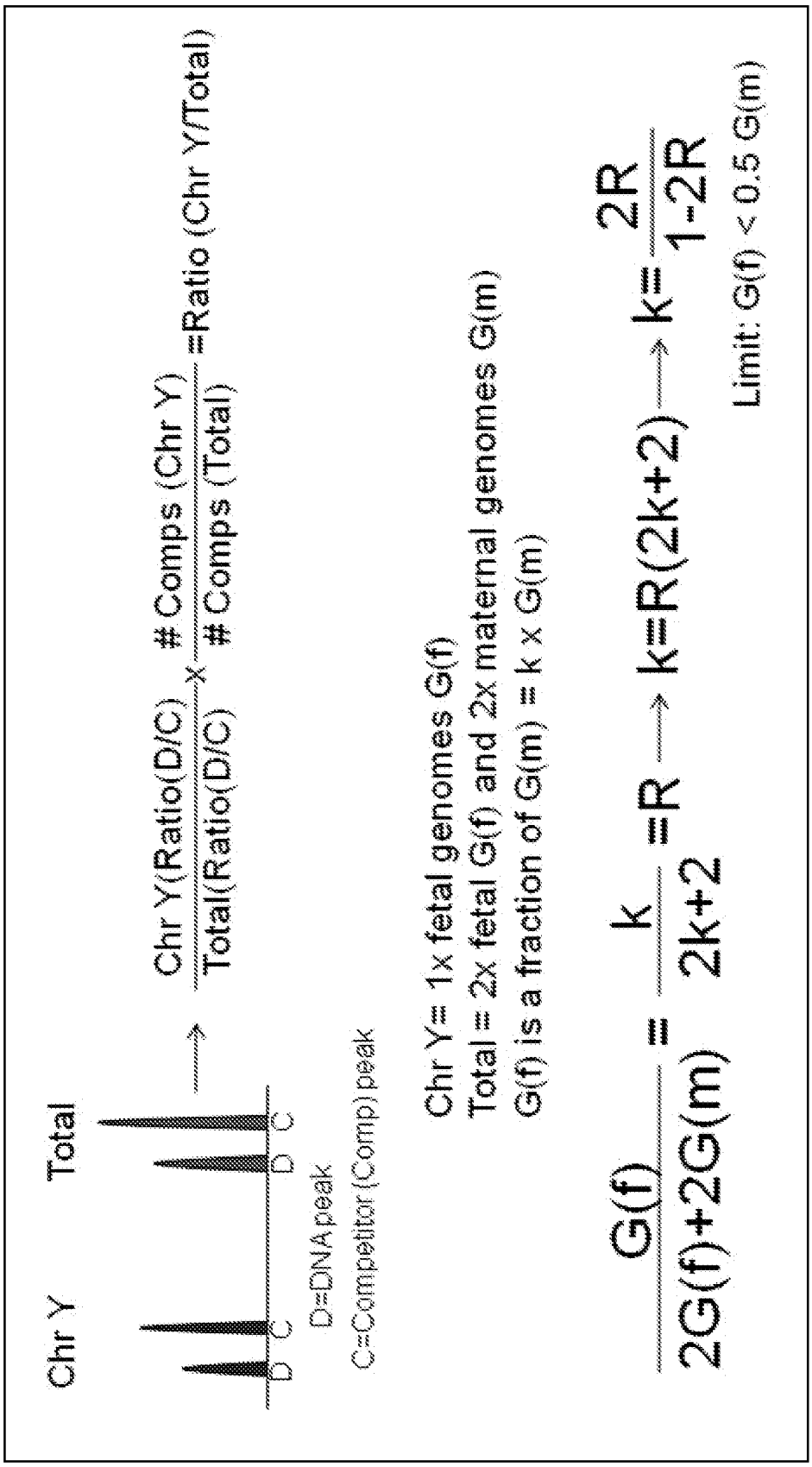
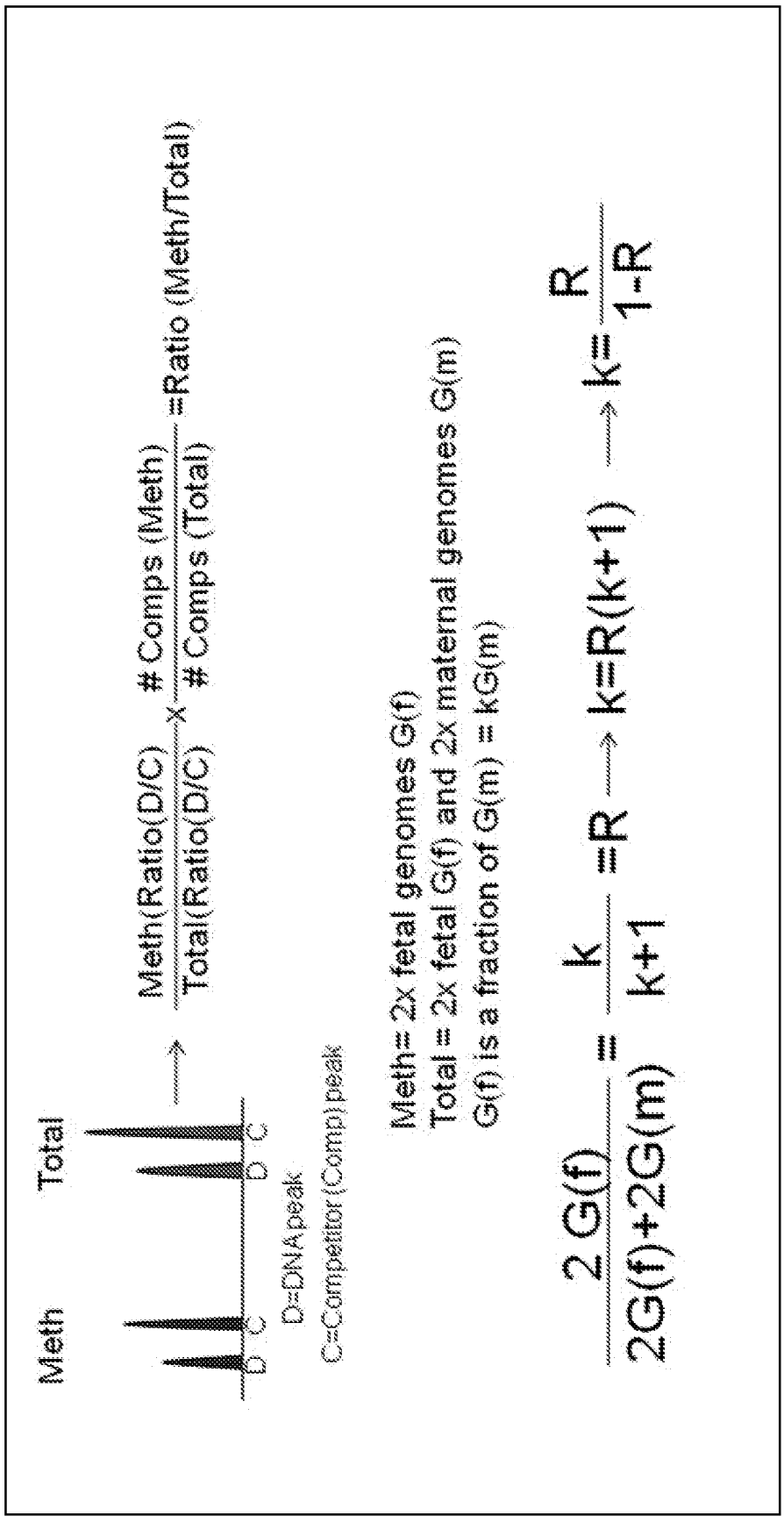

Chr Y
Total
D C
D C
D=DNA peak
C=Competitor (Comp) peak
Chr Y(Ratio(D/C)) / Total(Ratio(D/C)) x # Comps (Chr Y) / # Comps (Total) =Ratio (Chr Y/Total)
Chr Y= 1x fetal genomes G(f)
Total = 2x fetal G(f) and 2x maternal genomes G(m)
G(f) is a fraction of G(m) = k x G(m)
G(f) / 2G(f)+2G(m) = k / 2k+2 =R ⟶ k=R(2k+2) ⟶ k= 2R / 1-2R
Limit: G(f) < 0.5 G(m)

FIG. 156


Meth
Total
D C
D C
D=DNA peak
C=Competitor (Comp) peak
Meth(Ratio(D/C)) / Total(Ratio(D/C)) x # Comps (Meth) / # Comps (Total) =Ratio (Meth/Total)
Meth= 2x fetal genomes G(f)
Total = 2x fetal G(f) and 2x maternal genomes G(m)
G(f) is a fraction of G(m) = kG(m)
2 G(f) / 2G(f)+2G(m) = k / k+1 =R ⟶ k=R(k+1) ⟶ k= R / 1-R

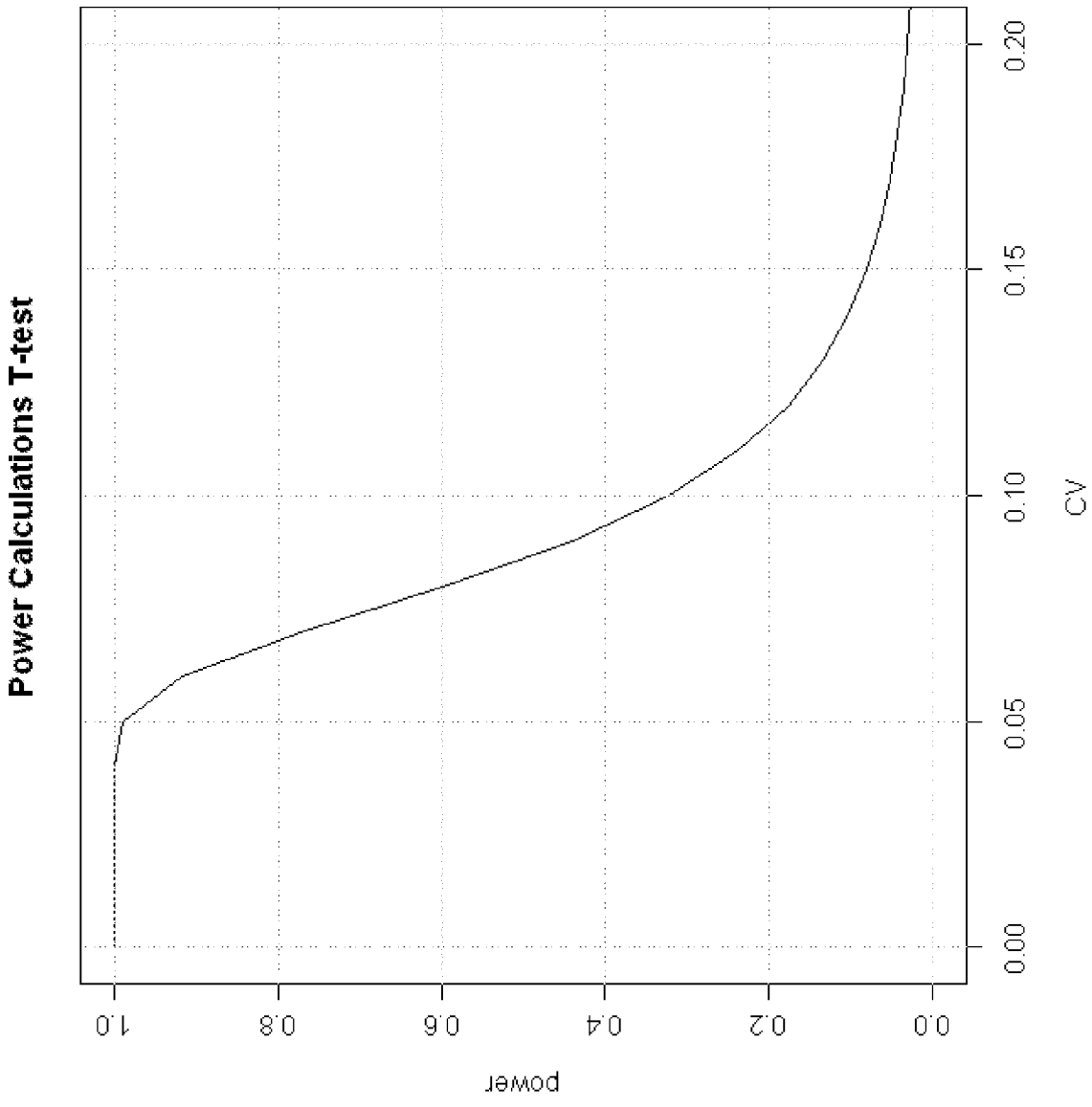

Power Calculations T-test
power
0.0
0.2
0.4
0.6
0.8
1.0
CV
0.00
0.05
0.10
0.15
0.20

FIG. 157

FIG. 158

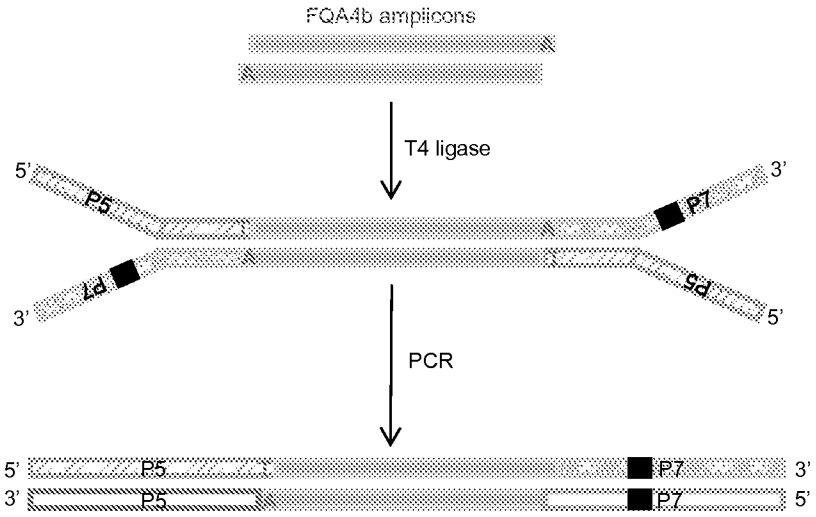


FIG. 159

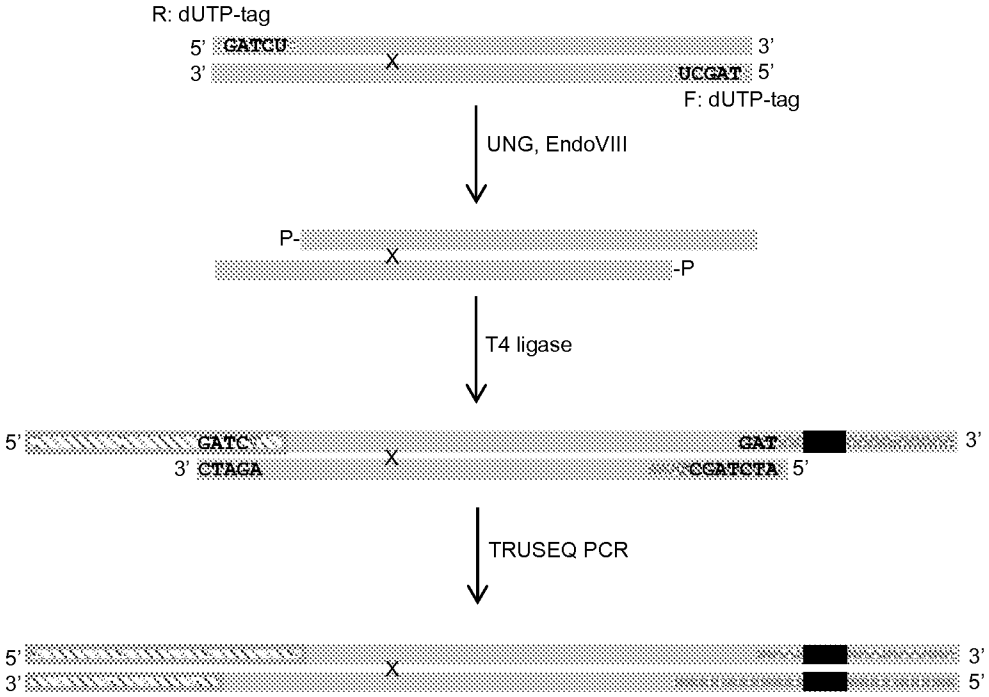
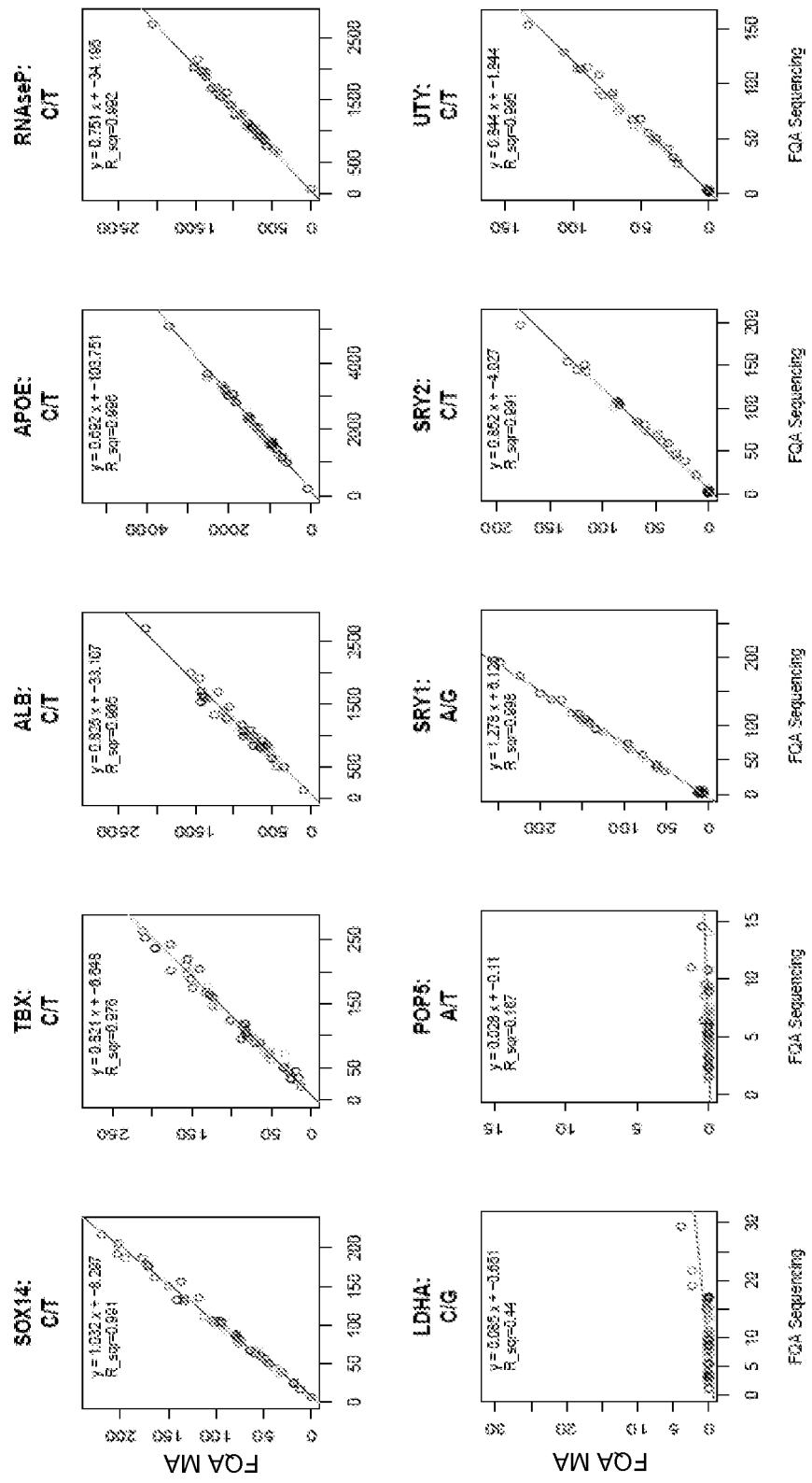

R: dUTP-tag
5' GATCU 3'
3' UCGAT 5'
F: dUTP-tag
UNG, EndoVIII
P-
-P
T4 ligase
5' GATC GAT 3'
3' CTAGA CCATCTA 5'
TRUSEQ PCR
5' 3'
3' 5'

FIG. 160


SOX14: C/T
TBX: C/T
ALB: C/T
APOE: C/T
RNAseP: C/T
LDHA: C/G
POP5: A/T
SRY1: A/G
SRY2: C/T
UTY: C/T
FQA MA
FQA Sequencing

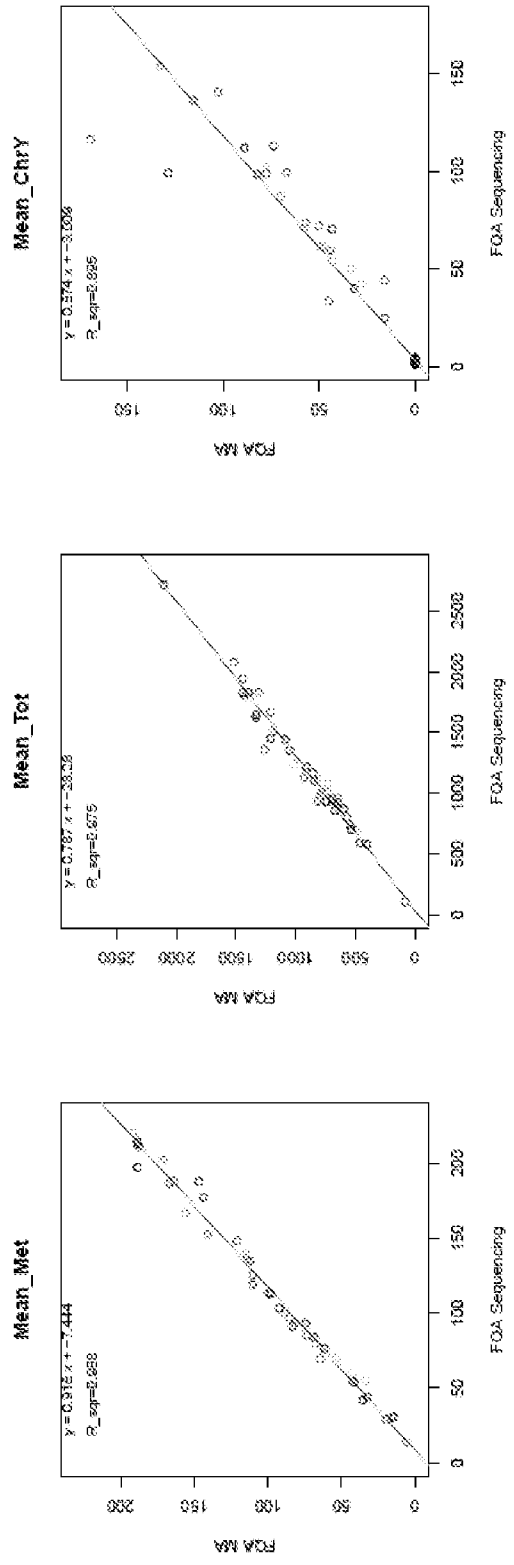

Mean_ChrY
y = 0.974 x + -3.936
R_sqr=0.995
FQA MA
FQA Sequencing
Mean_Tot
y = 0.787 x + -38.28
R_sqr=0.975
FQA MA
FQA Sequencing
Mean_Met
y = 0.912 x + -7.444
R_sqr=0.988
FQA MA
FQA Sequencing

FIG. 161

FIG. 162

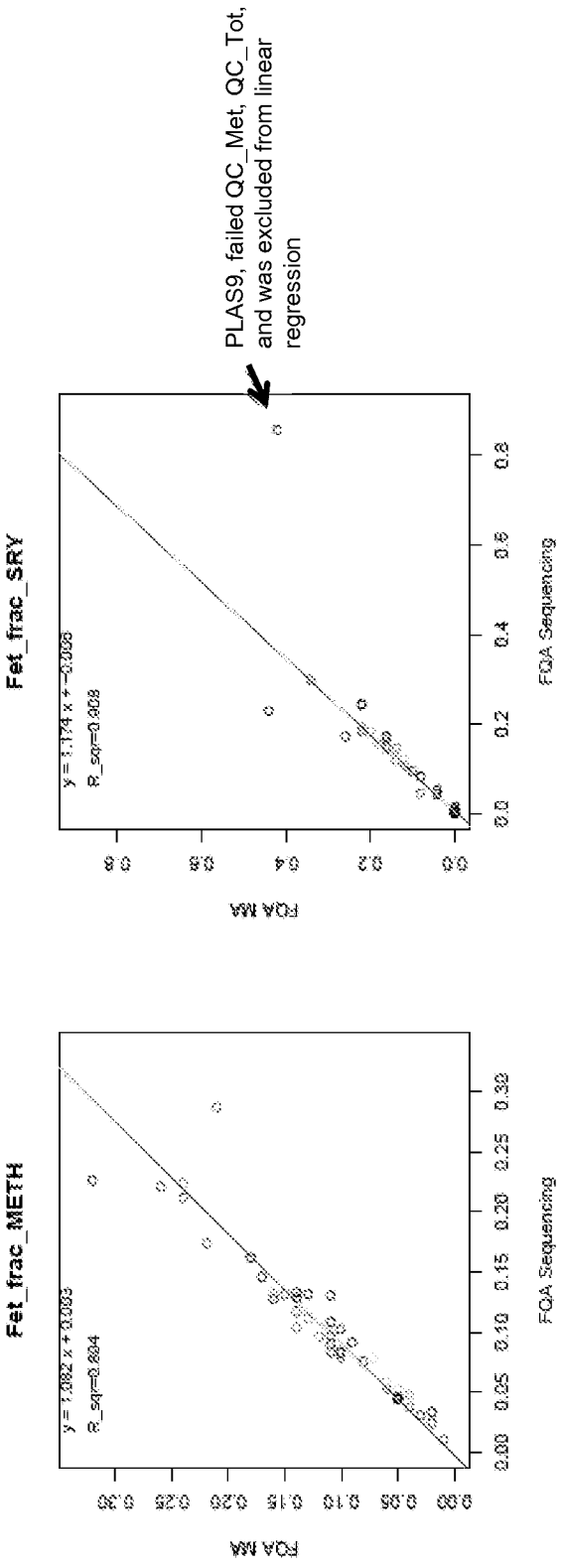
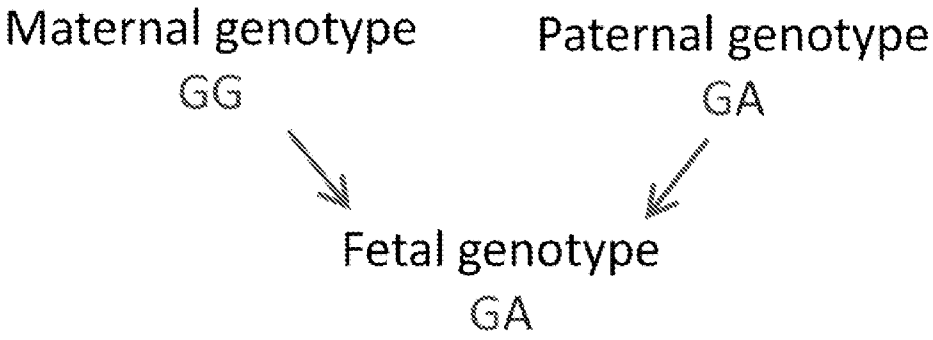

Fet_frac_METH
y = 1.082 x + 0.003
R_sq=0.894
FOA MA
FQA Sequencing
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Fet_frac_SRY
y = 1.174 x + −0.006
R_sq=0.908
FOA MA
FQA Sequencing
0.0
0.2
0.4
0.6
0.8
PLAS9, failed QC_Met, QC_Tot, and was excluded from linear regression

FIG. 163

0.24 likelihood of informative fetal/maternal genotype combination with SNP minor allele frequency of 0.4		Maternal Genotype		
		AA (p=0.36)	Aa (p=0.48)	aa (p=0.16)
Paternal Genotype	AA (p=0.36)	AA	AA Aa	**Aa (p=0.058)**
	Aa (p=0.48)	AA	AA Aa aa	**Aa (p=0.038)**
		Aa (p=0.086)		aa
	aa (p=0.16)	**Aa (p=0.058)**	Aa aa	aa

FIG. 164

Detected SNPs have high minor allele population frequency



CCF DNA from plasma of pregnant mother

Maternal CCF copies of GG = 900
Fetal CCF copies of GA = 100
Total CCF DNA copies = 1000

FIG. 165

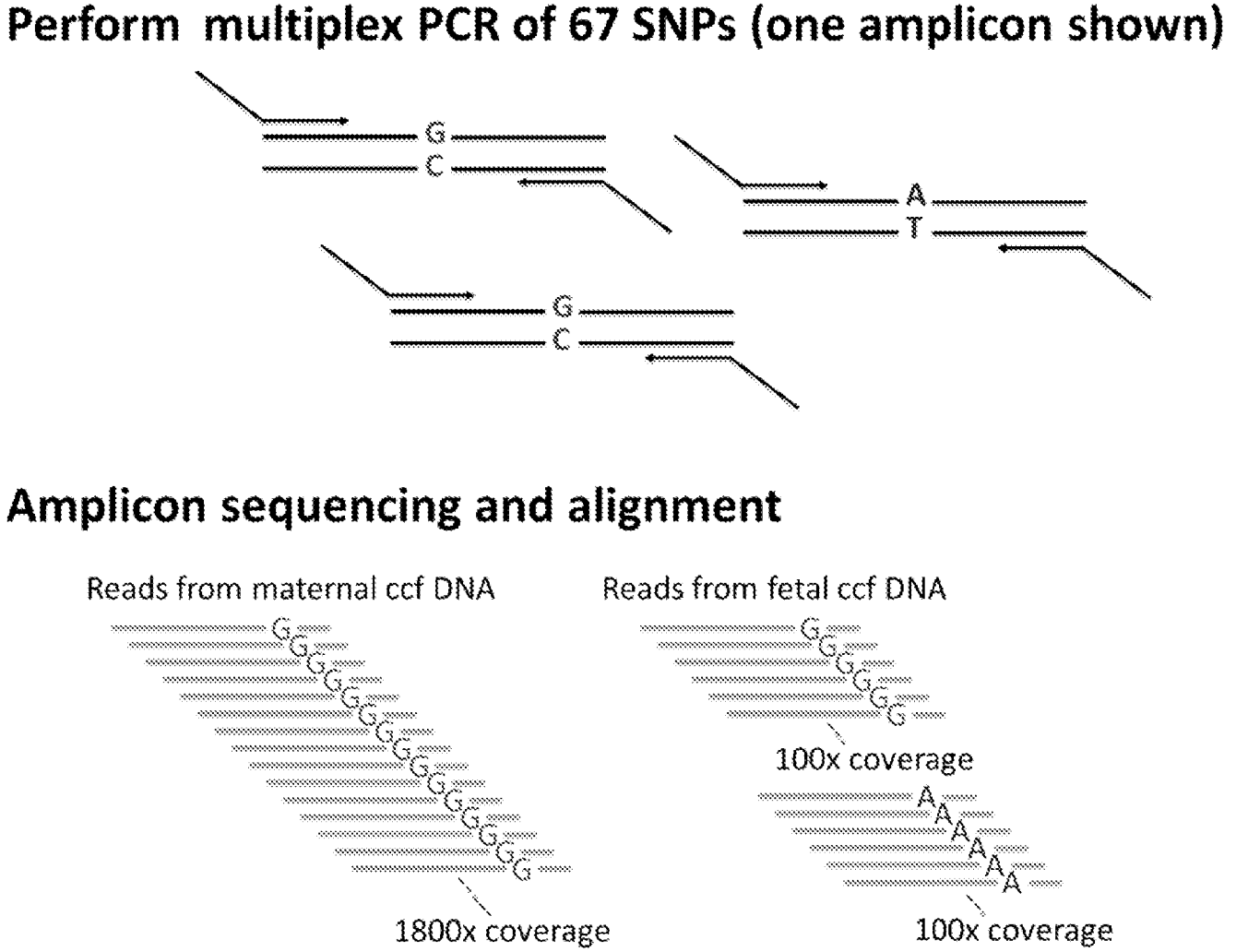

Perform multiplex PCR of 67 SNPs (one amplicon shown)
G
C
A
T
G
C
Amplicon sequencing and alignment
Reads from maternal ccf DNA
Reads from fetal ccf DNA
1800x coverage
100x coverage
100x coverage
Fetal fraction = # of reads alllele A / (# of reads allele A + # of reads allele G) x 2 = 10%

FIG. 166

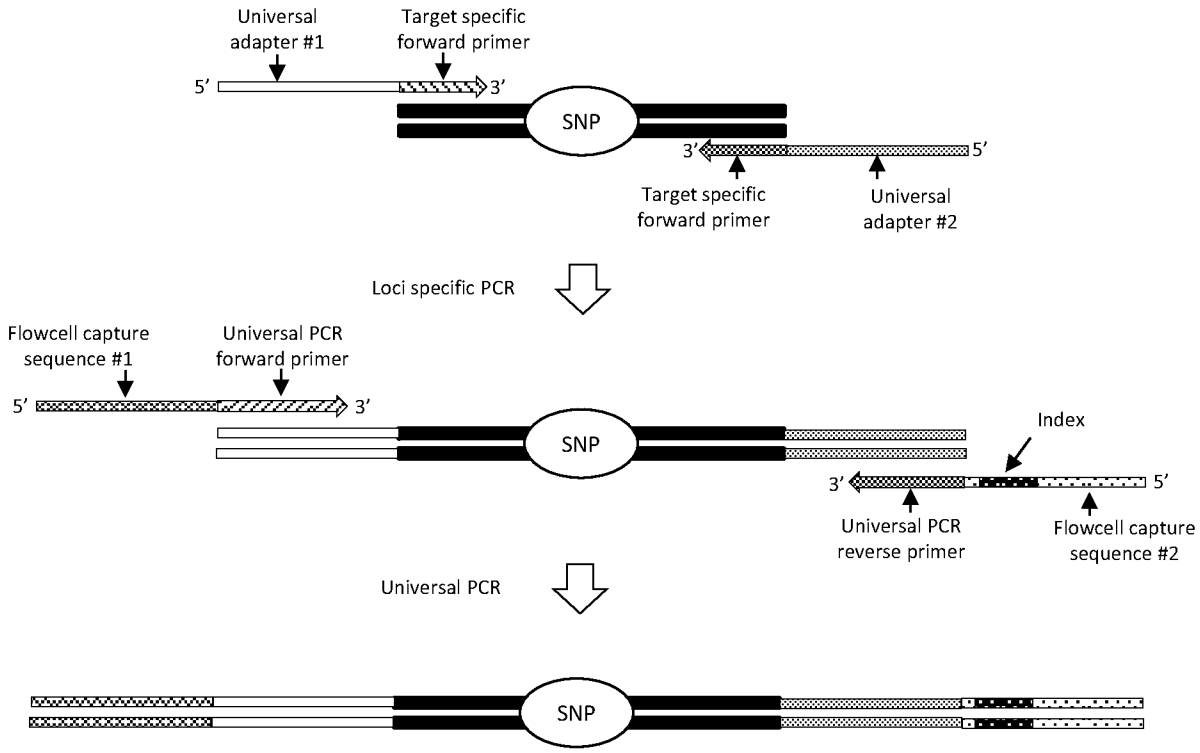

Universal adapter #1
Target specific forward primer
5'
3'
SNP
3'
5'
Target specific forward primer
Universal adapter #2
Loci specific PCR
Flowcell capture sequence #1
Universal PCR forward primer
5'
3'
SNP
Index
3'
5'
Universal PCR reverse primer
Flowcell capture sequence #2
Universal PCR
SNP

FIG. 167

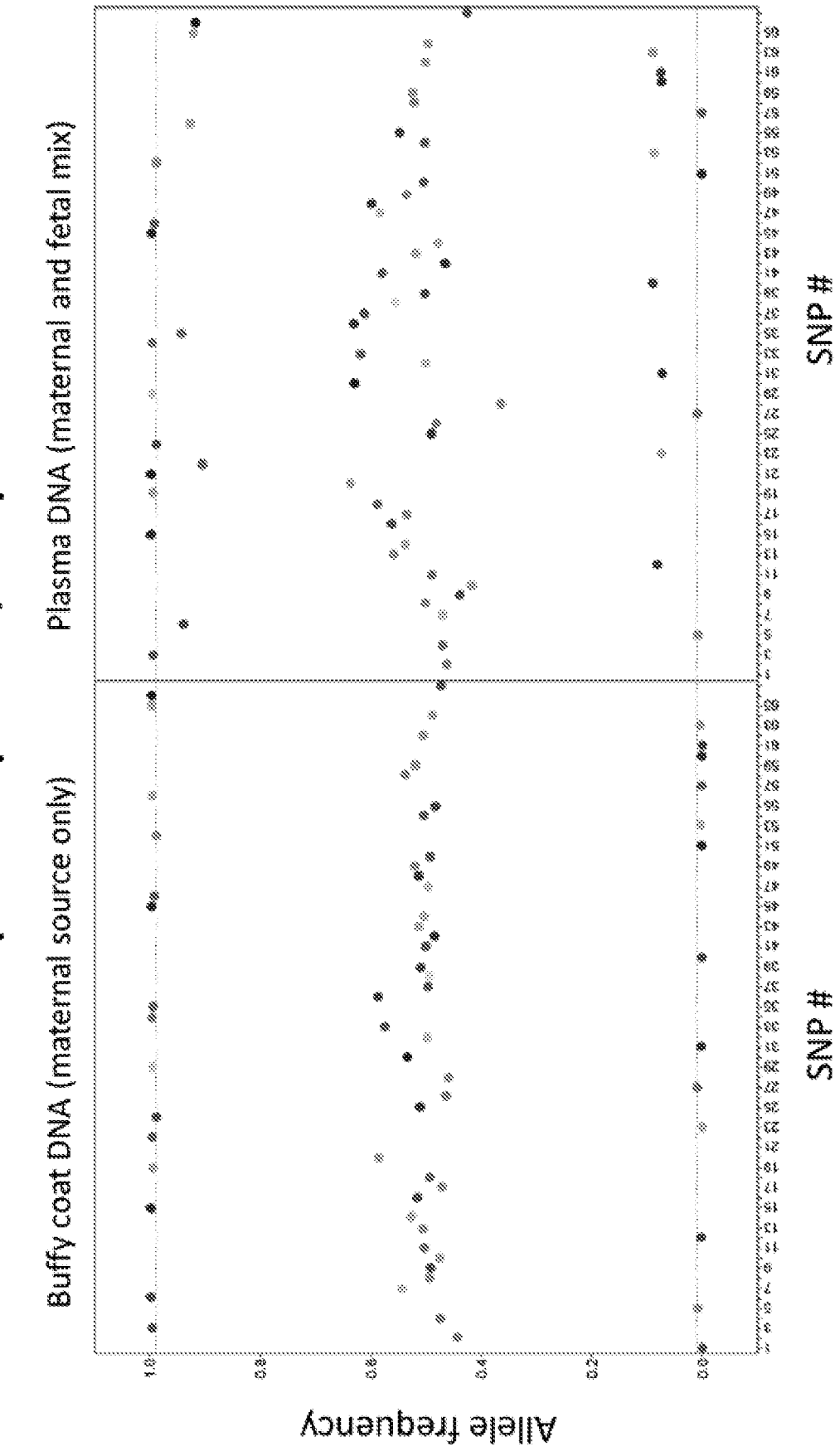

Allele frequencies per SNP, sample ID: 11
Buffy coat DNA (maternal source only)
Plasma DNA (maternal and fetal mix)
Allele frequency
SNP #
SNP #

FIG. 168

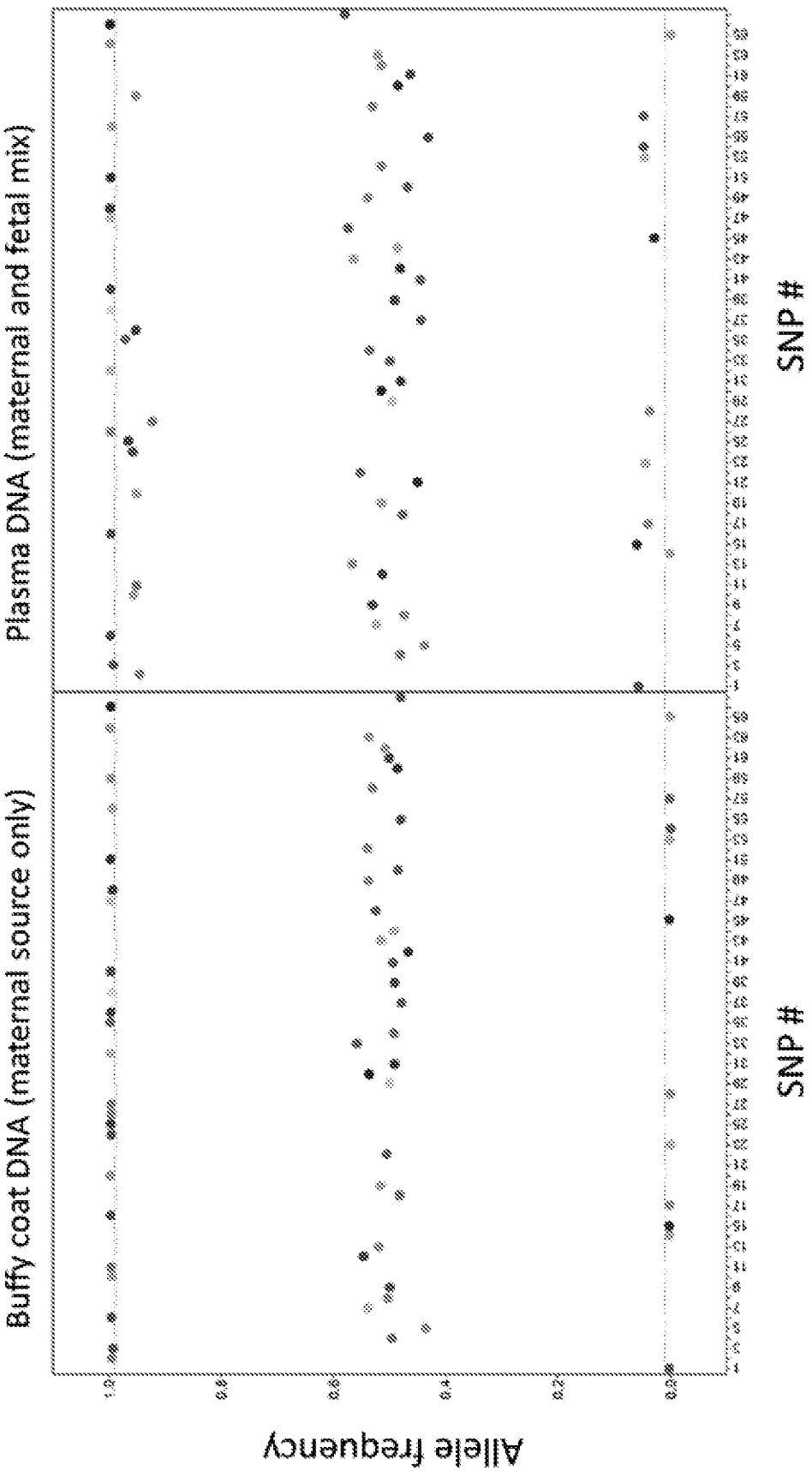

Allele frequencies per SNP, sample ID: 48
Buffy coat DNA (maternal source only)
Plasma DNA (maternal and fetal mix)
Allele frequency
SNP #
SNP #

FIG. 169

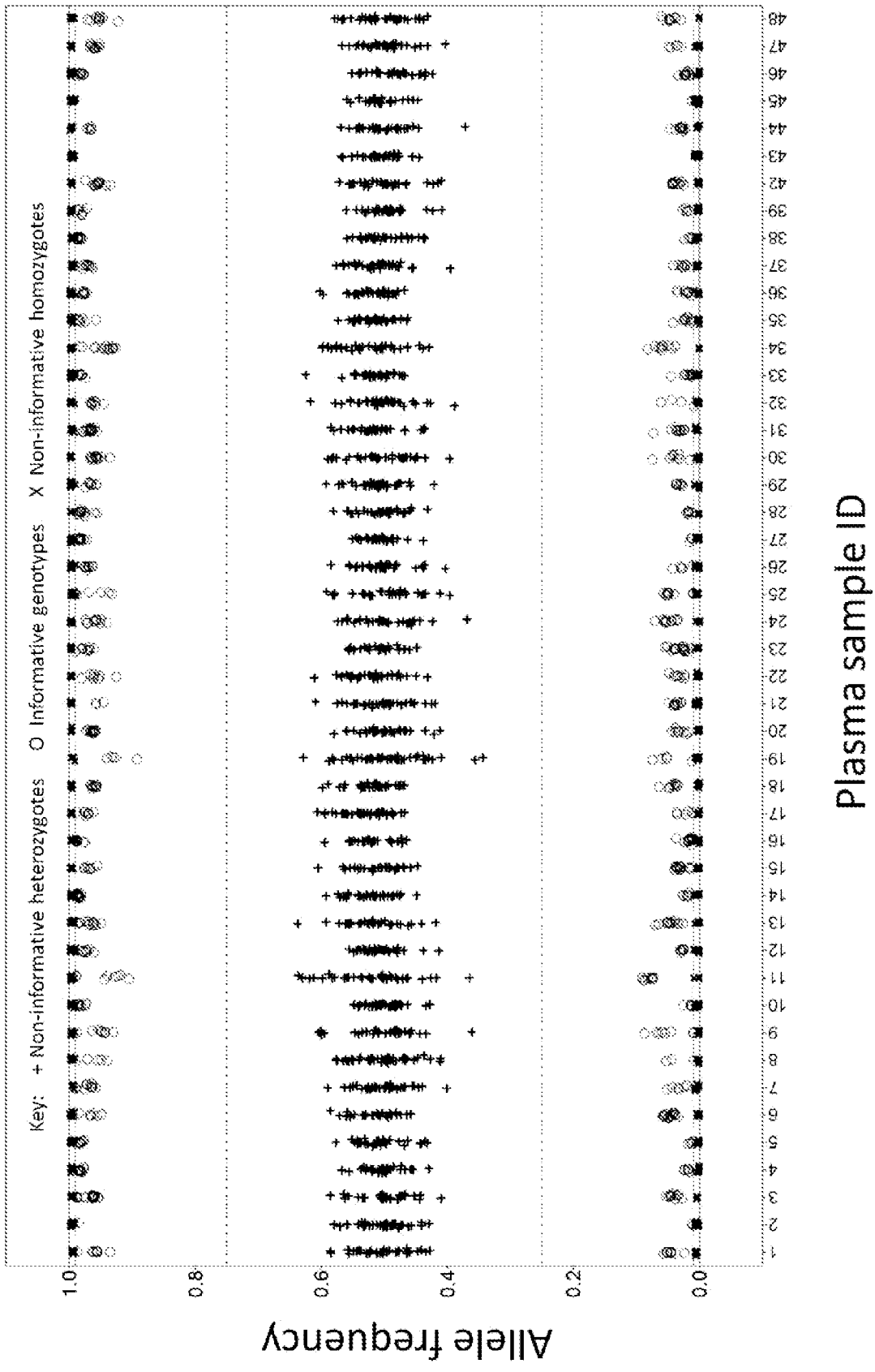

Allele frequencies per sample
Key: + Non-informative heterozygotes O Informative genotypes X Non-informative homozygotes
Allele frequency
1.0
0.8
0.6
0.4
0.2
0.0
Plasma sample ID
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
42
43
44
45
46
47
48

FIG. 170

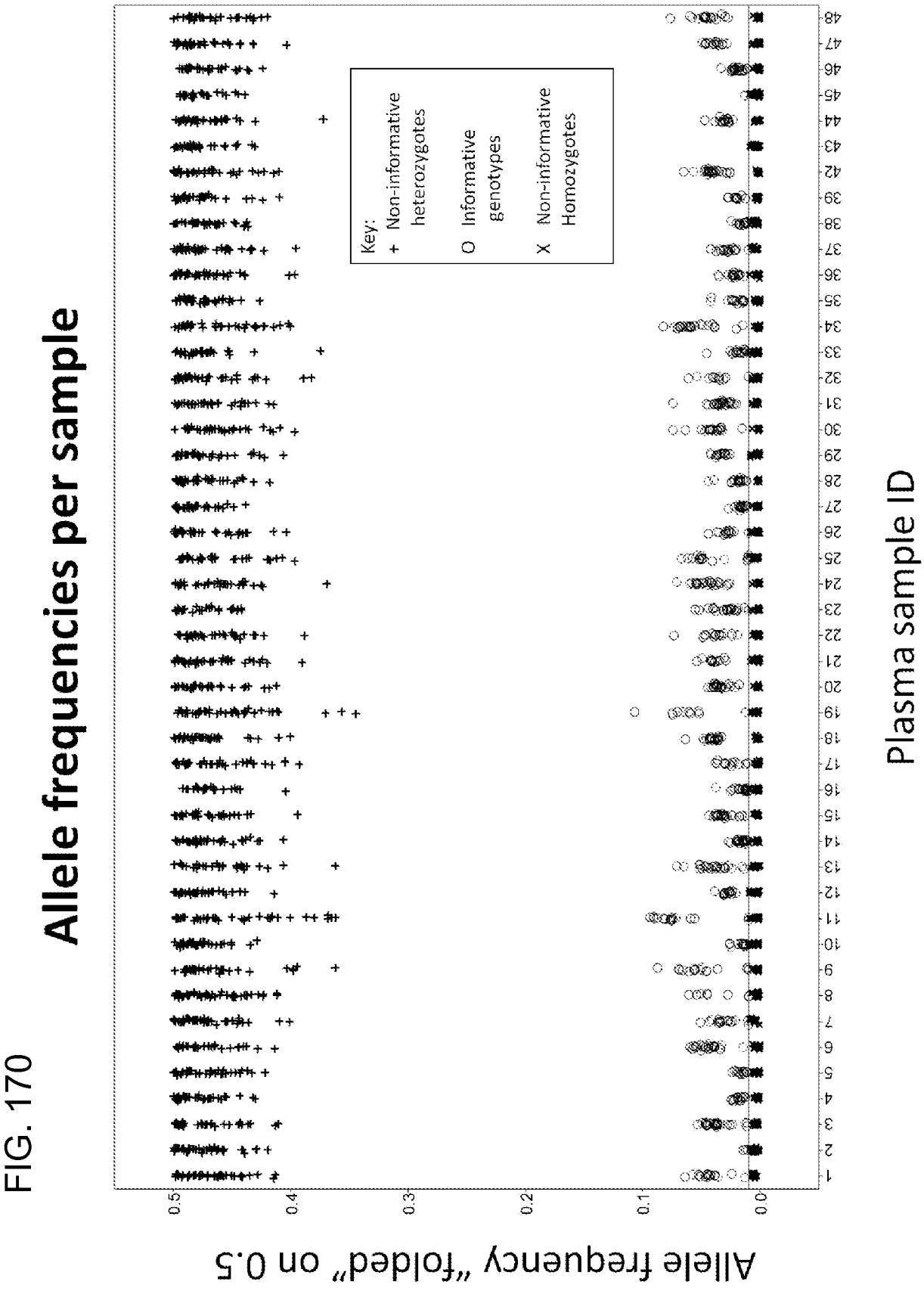

Allele frequencies per sample
Key:
+ Non-informative heterozygotes
O Informative genotypes
X Non-informative Homozygotes
Allele frequency "folded" on 0.5
0.0
0.1
0.2
0.3
0.4
0.5
Plasma sample ID
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
42
43
44
45
46
47
48

FIG. 171

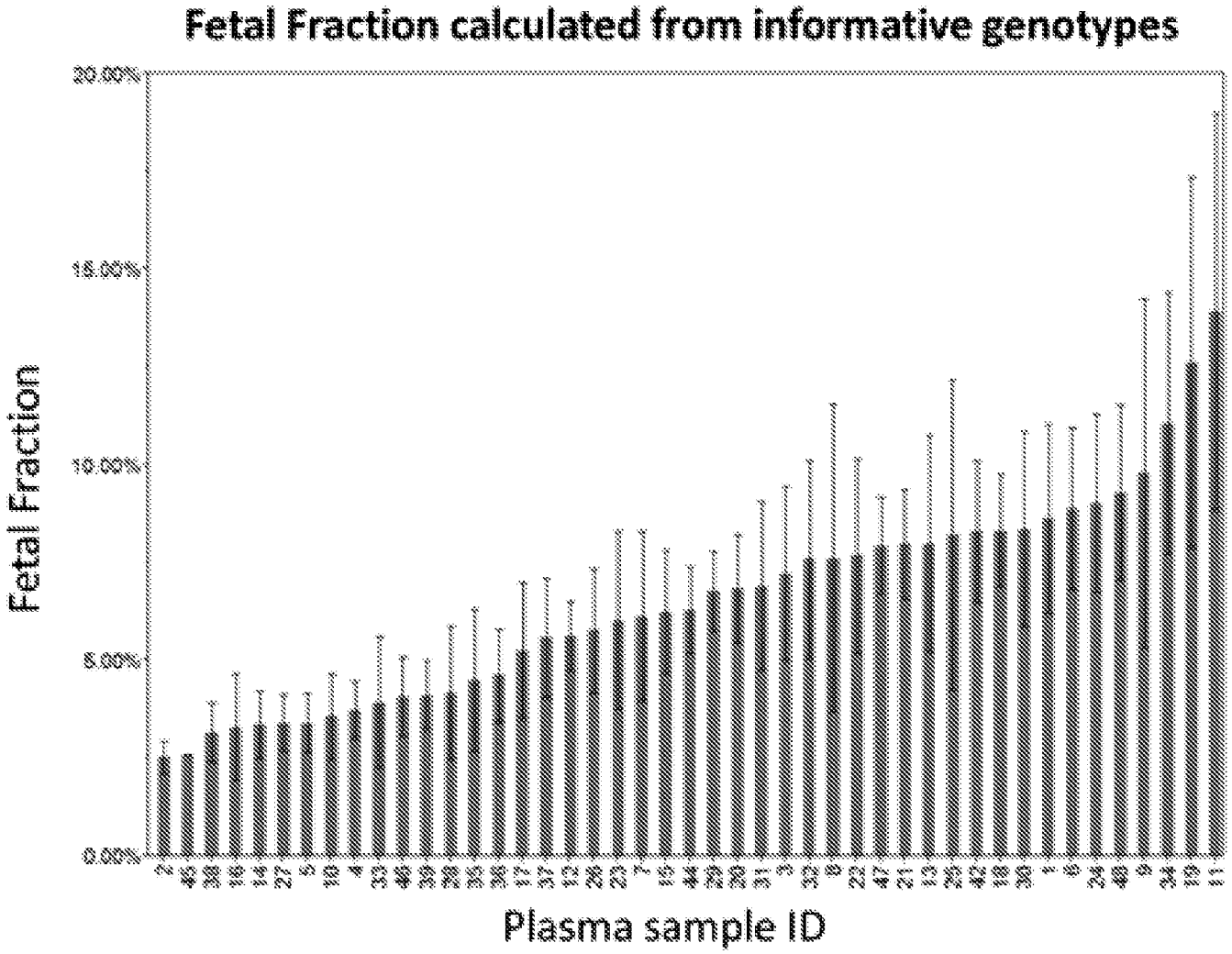

Fetal Fraction calculated from informative genotypes
Fetal Fraction
20.00%
15.00%
10.00%
5.00%
0.00%
Plasma sample ID

FIG. 172

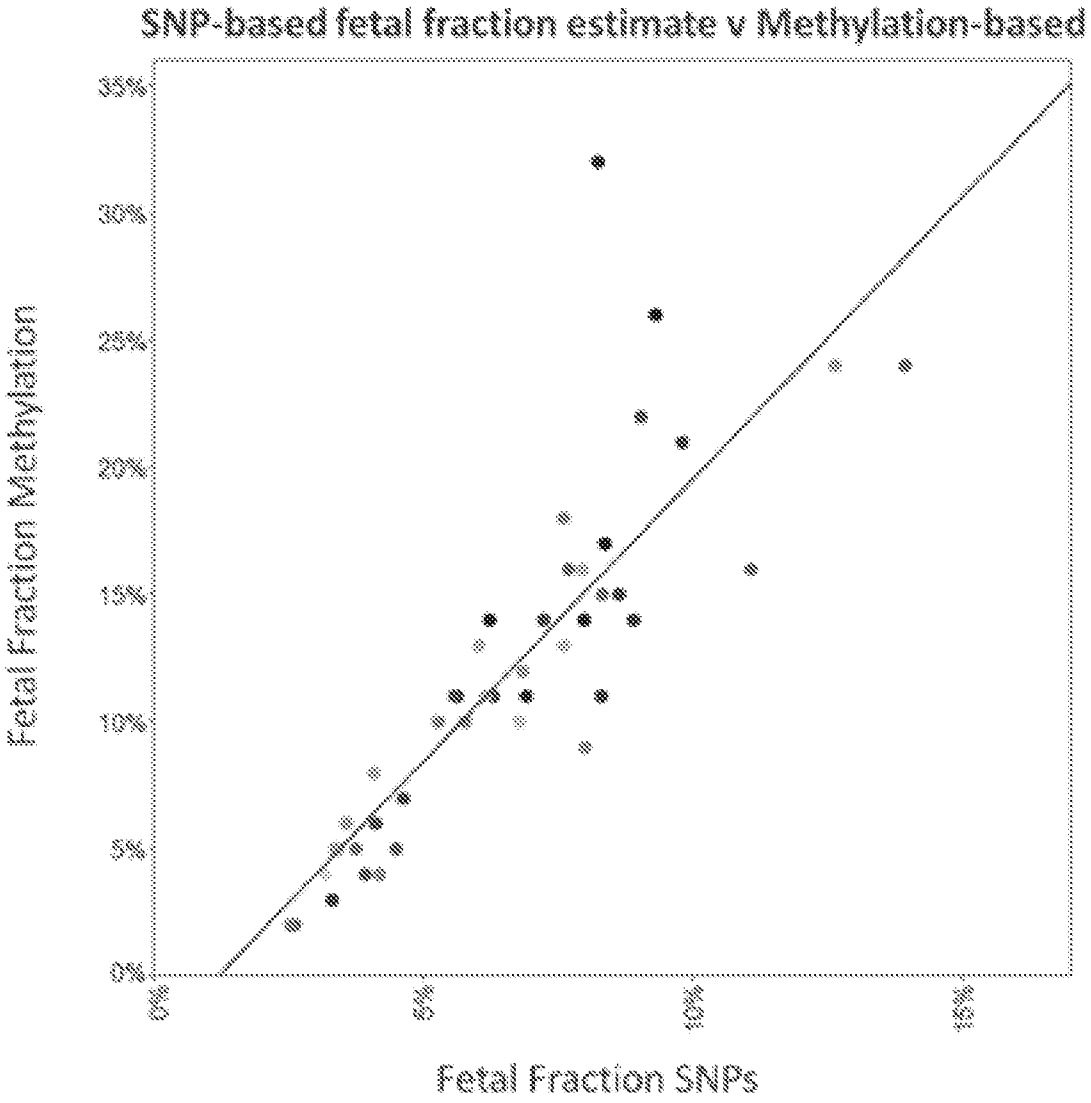

SNP-based fetal fraction estimate v Methylation-based
Fetal Fraction Methylation
35%
30%
25%
20%
15%
10%
5%
0%
0%
5%
10%
15%
Fetal Fraction SNPs

FIG. 173

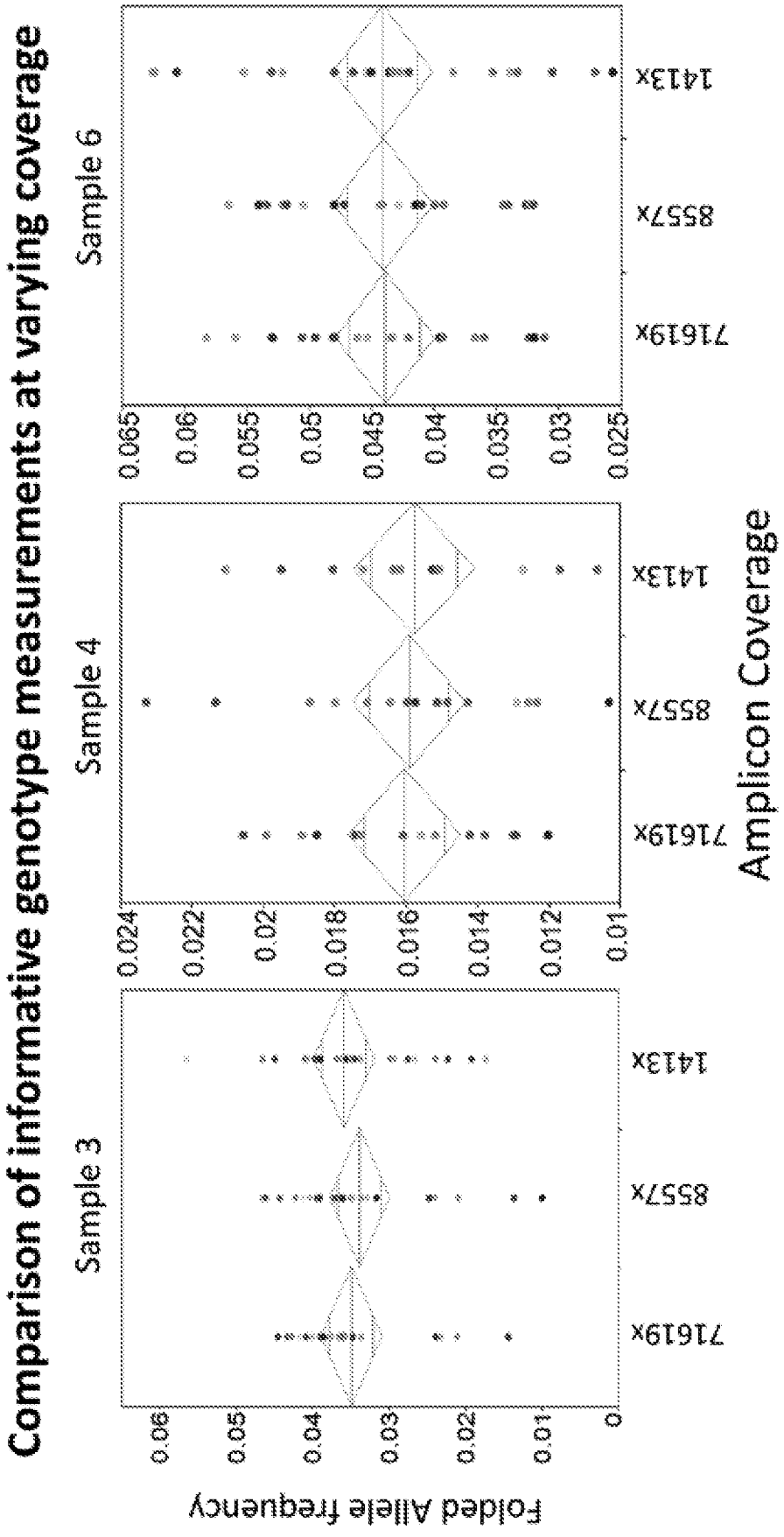
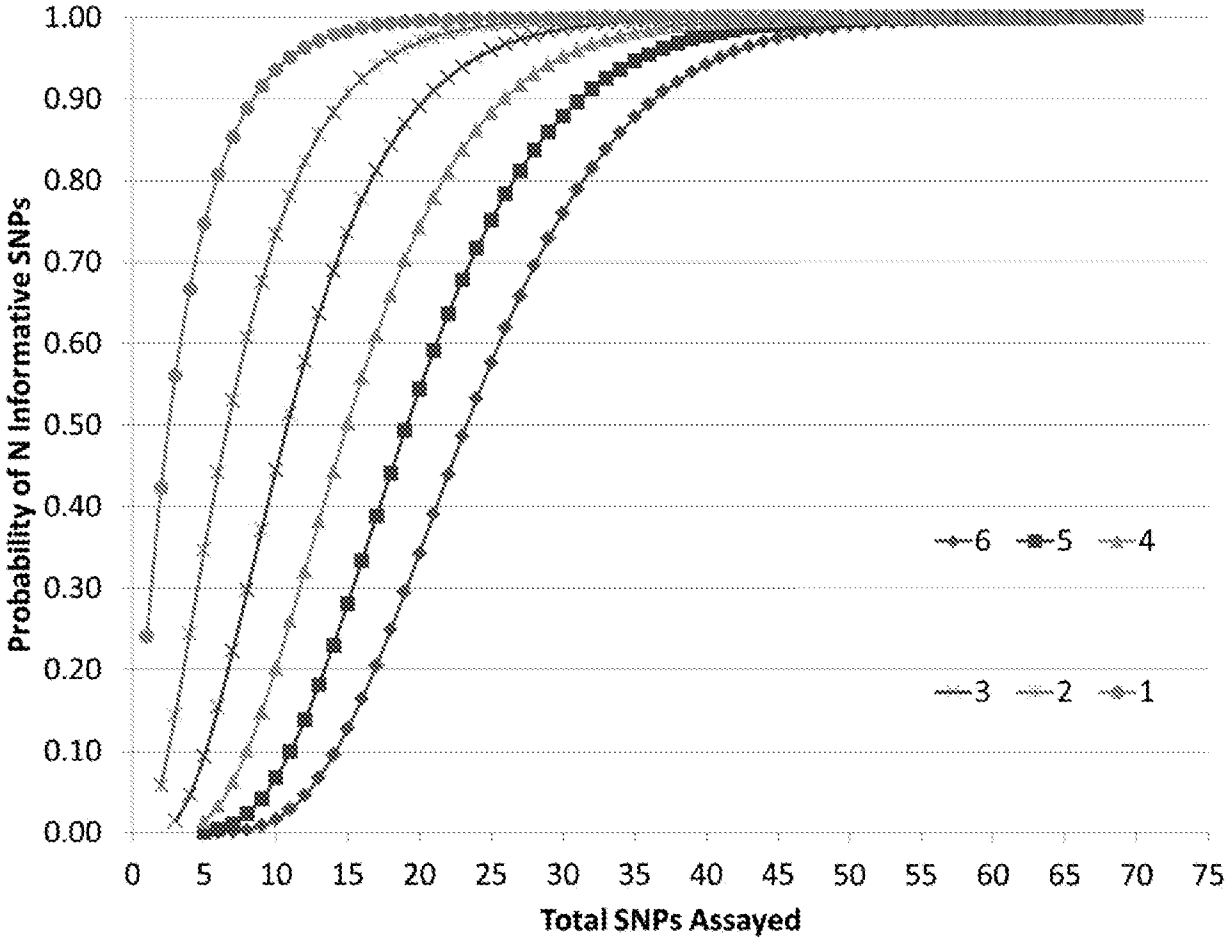

Comparison of informative genotype measurements at varying coverage
Sample 3
Sample 4
Sample 6
Folded Allele frequency
Amplicon Coverage
716.19x
855.7x
1413x
0
0.01
0.02
0.03
0.04
0.05
0.06
0.012
0.014
0.016
0.018
0.02
0.022
0.024
0.025
0.03
0.035
0.04
0.045
0.05
0.055
0.06
0.065

FIG. 174



Graph probabilities assume 0.4 MAF SNPs

FIG. 175

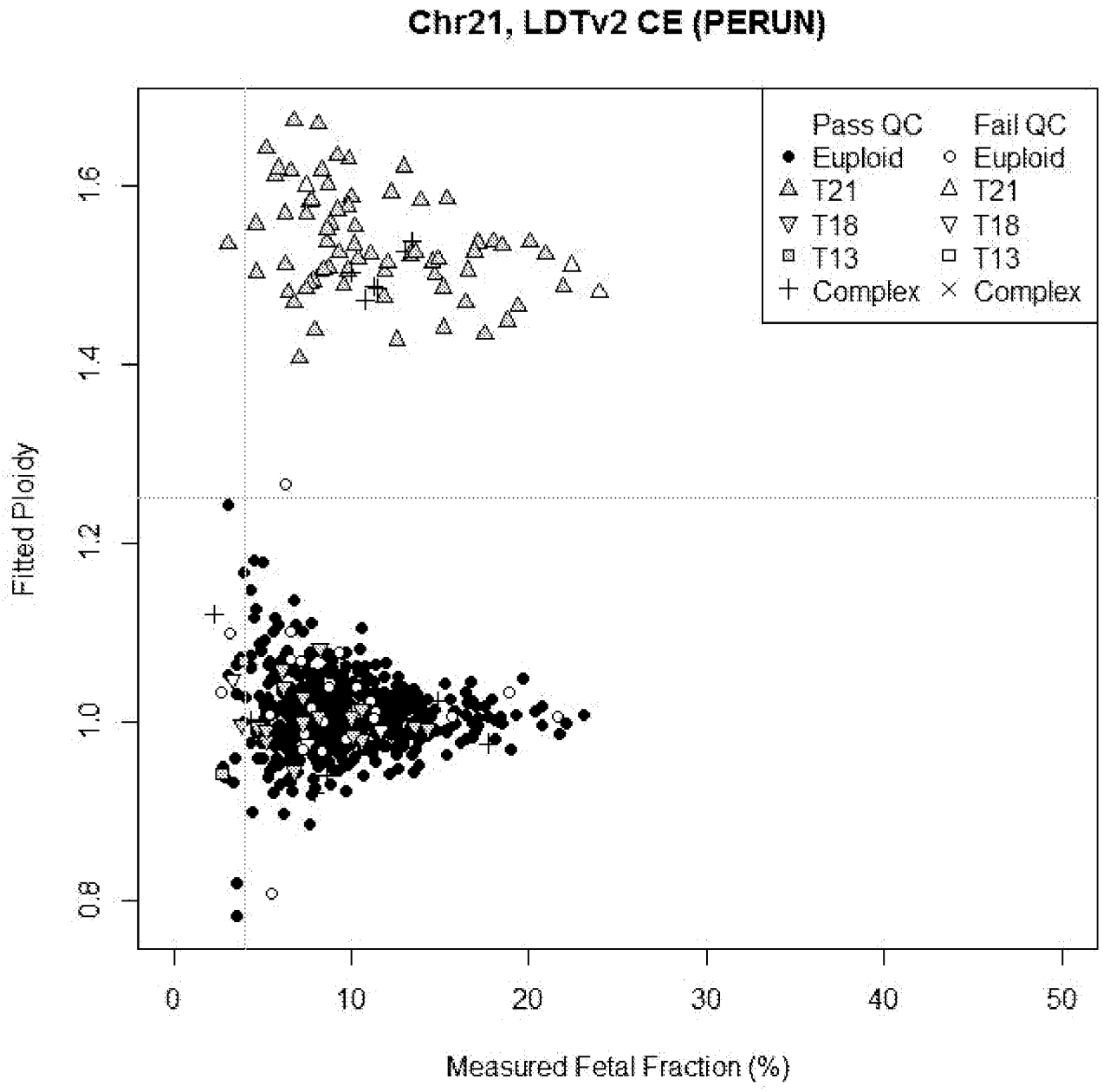

Chr21, LDTv2 CE (PERUN)
Pass QC
Euploid
T21
T18
T13
Complex
Fail QC
Euploid
T21
T18
T13
Complex
Fitted Ploidy
0.8
1.0
1.2
1.4
1.6
0
10
20
30
40
50
Measured Fetal Fraction (%)

FIG. 176

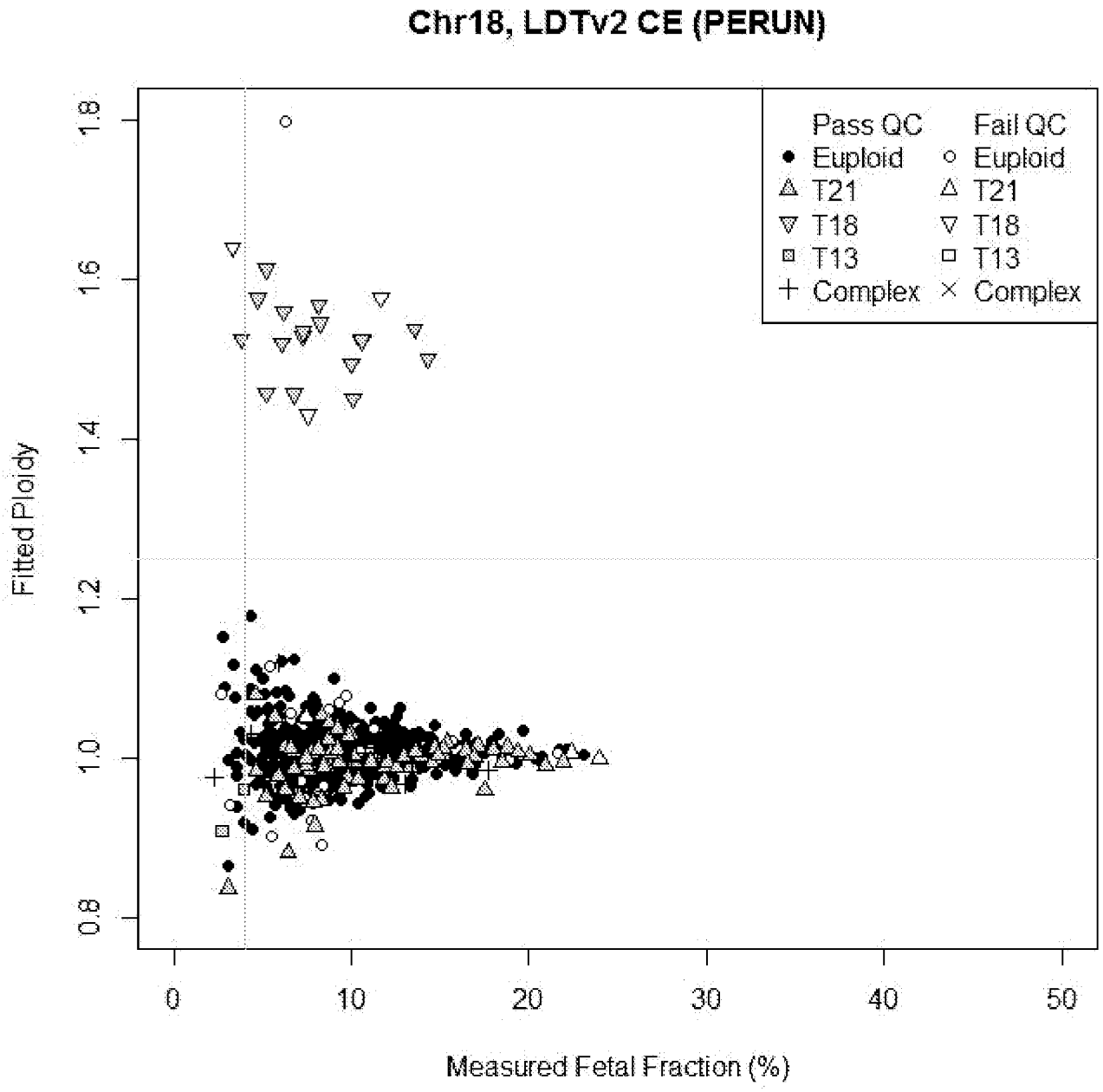
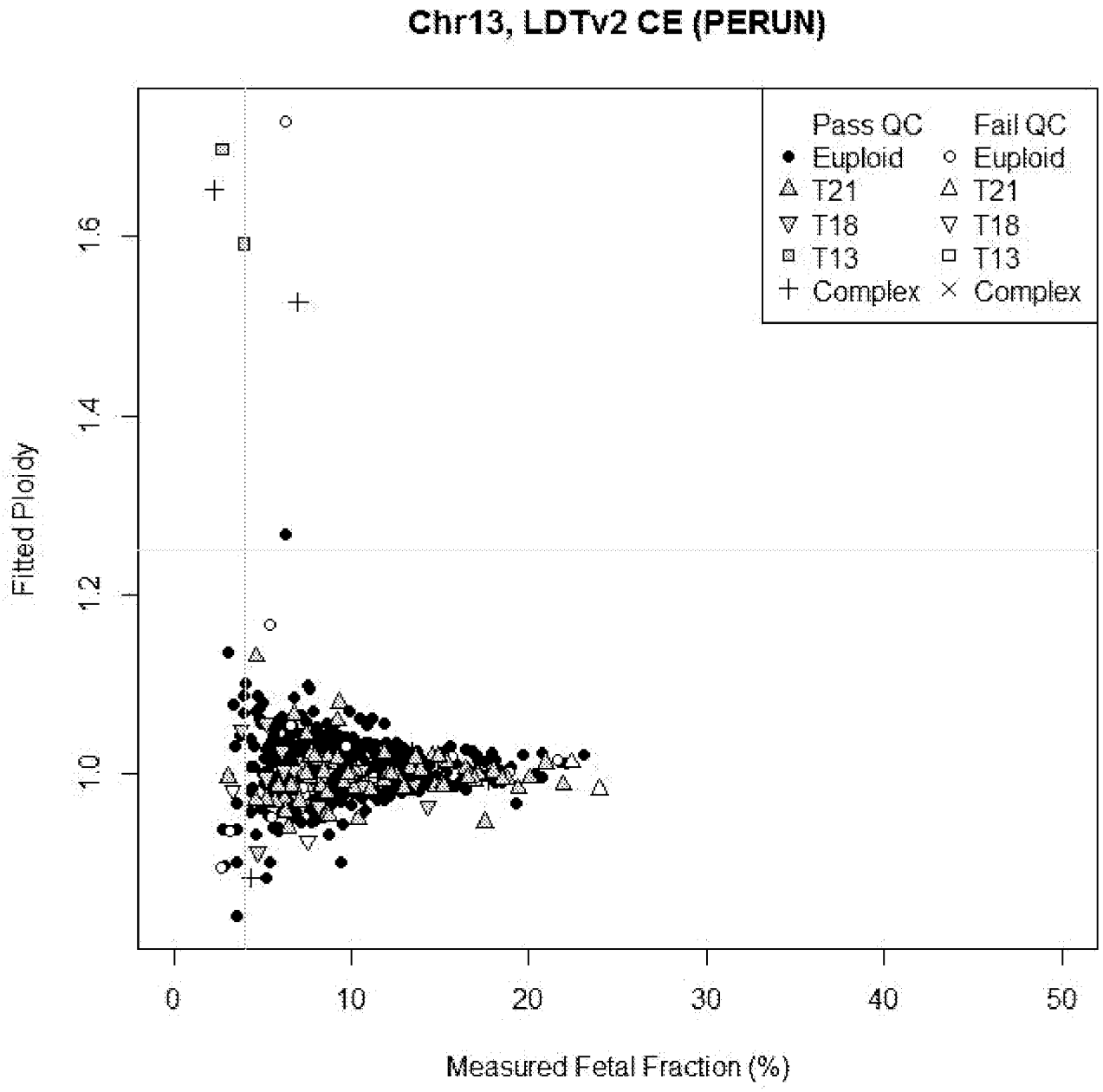

Chr18, LDTv2 CE (PERUN)
Pass QC
Euploid
T21
T18
T13
Complex
Fail QC
Euploid
T21
T18
T13
Complex
Fitted Ploidy
0.8
1.0
1.2
1.4
1.6
1.8
0
10
20
30
40
50
Measured Fetal Fraction (%)

FIG. 177


Chr13, LDTv2 CE (PERUN)
Pass QC
Euploid
T21
T18
T13
Complex
Fail QC
Euploid
T21
T18
T13
Complex
Fitted Ploidy
1.0
1.2
1.4
1.6
0
10
20
30
40
50
Measured Fetal Fraction (%)

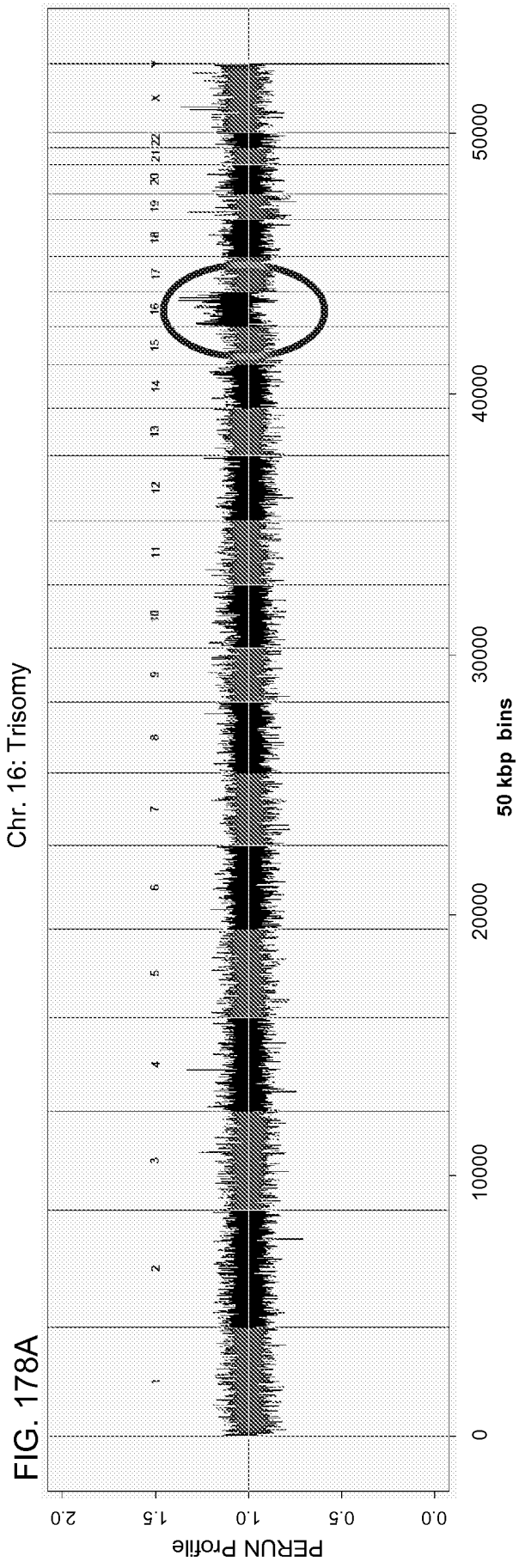


FIG. 178A

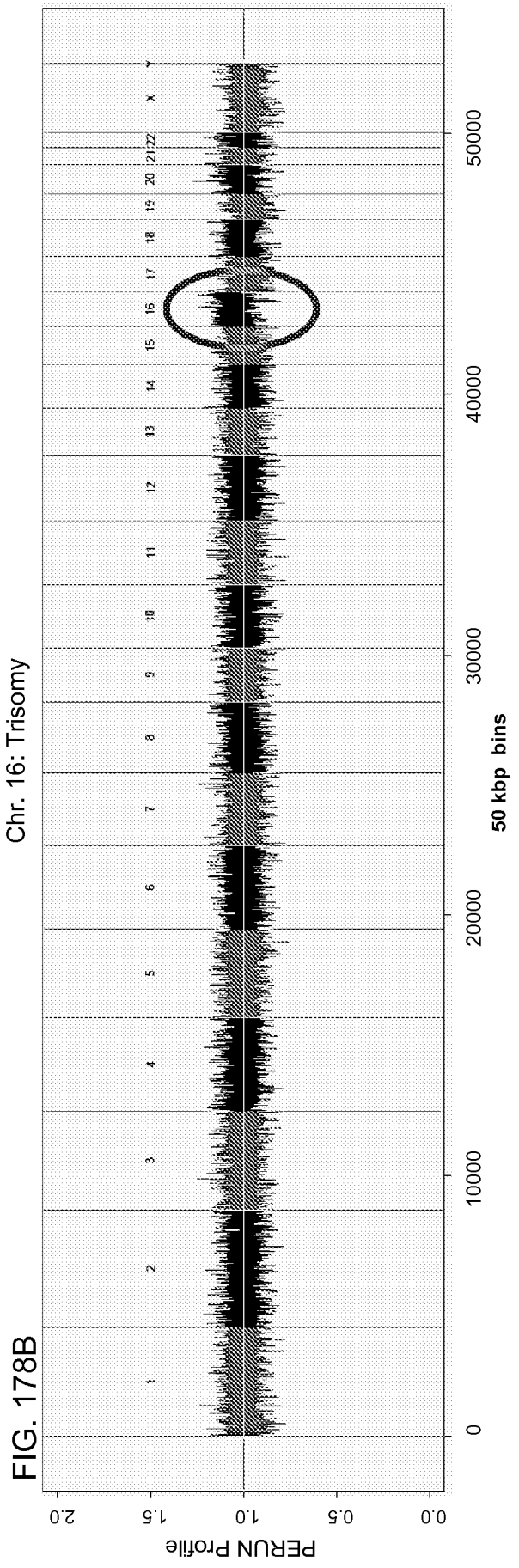


FIG. 178B

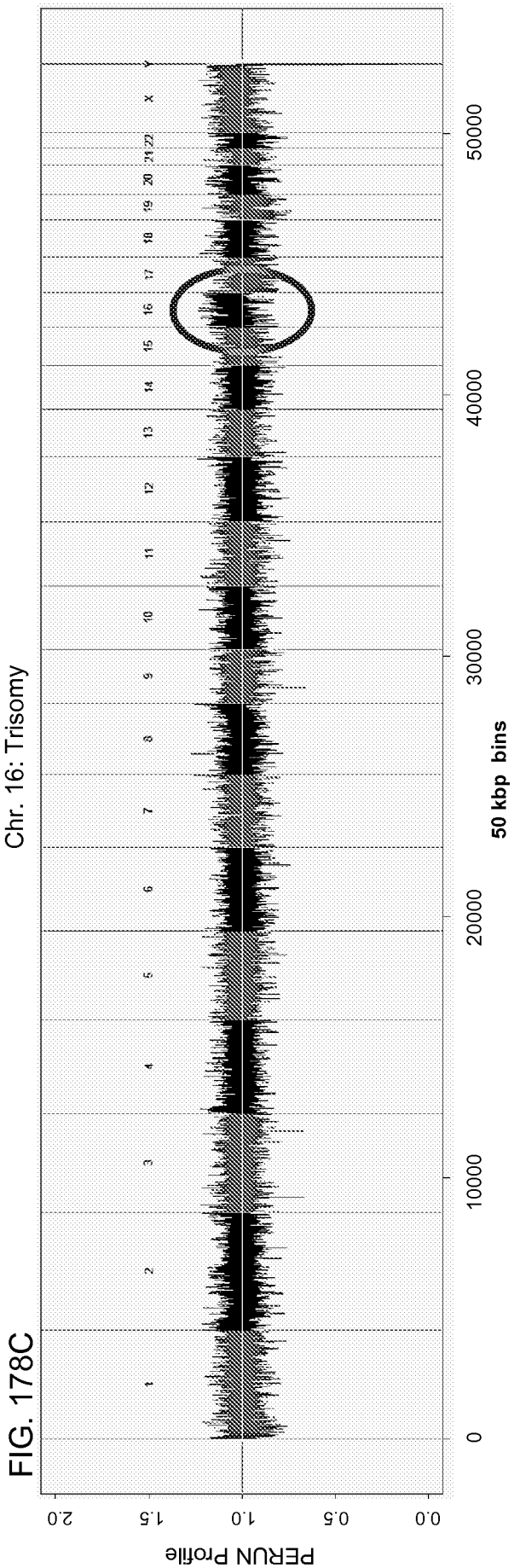


FIG. 178C

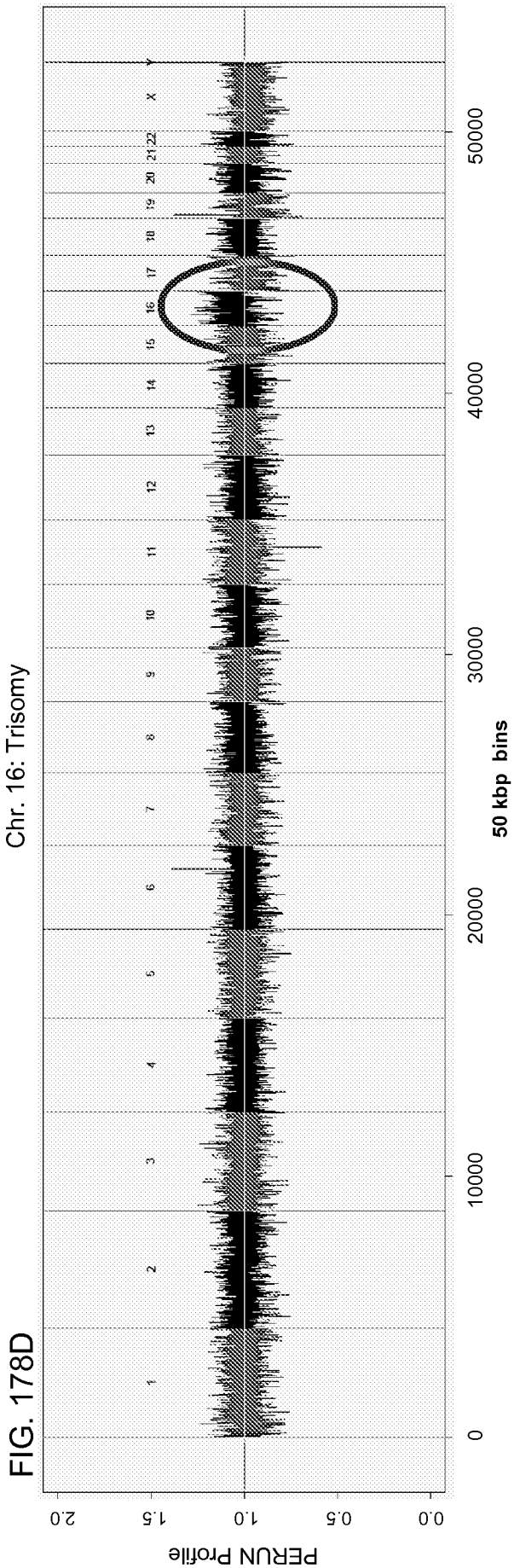


FIG. 178D

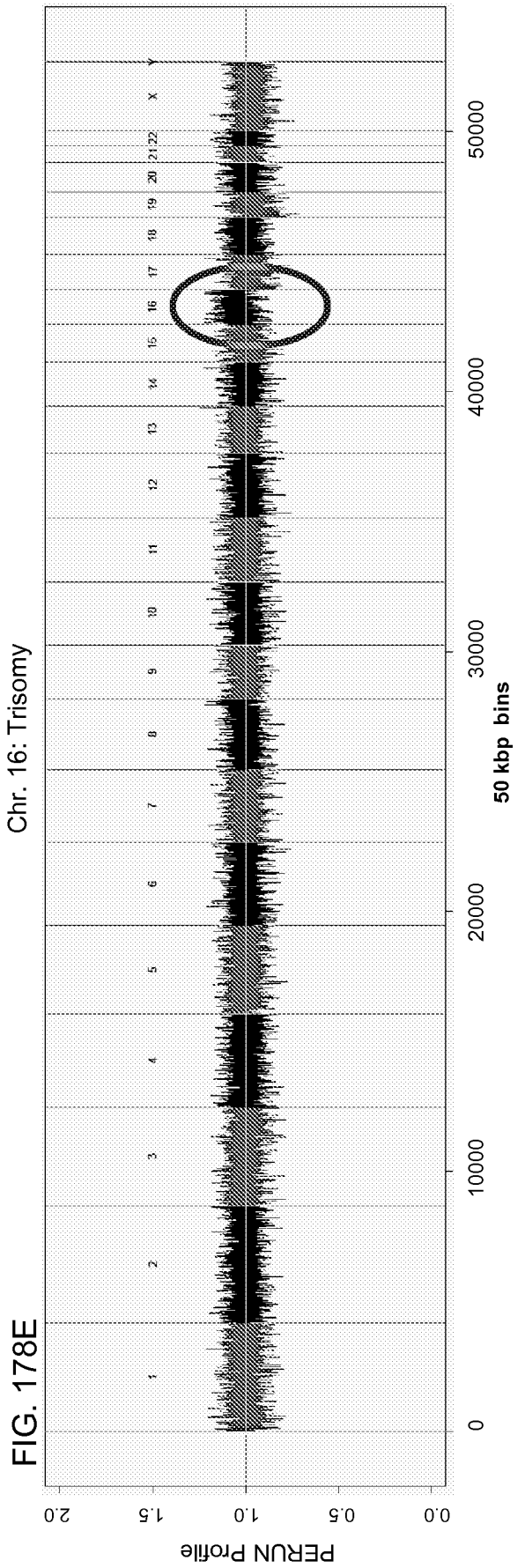


FIG. 178E

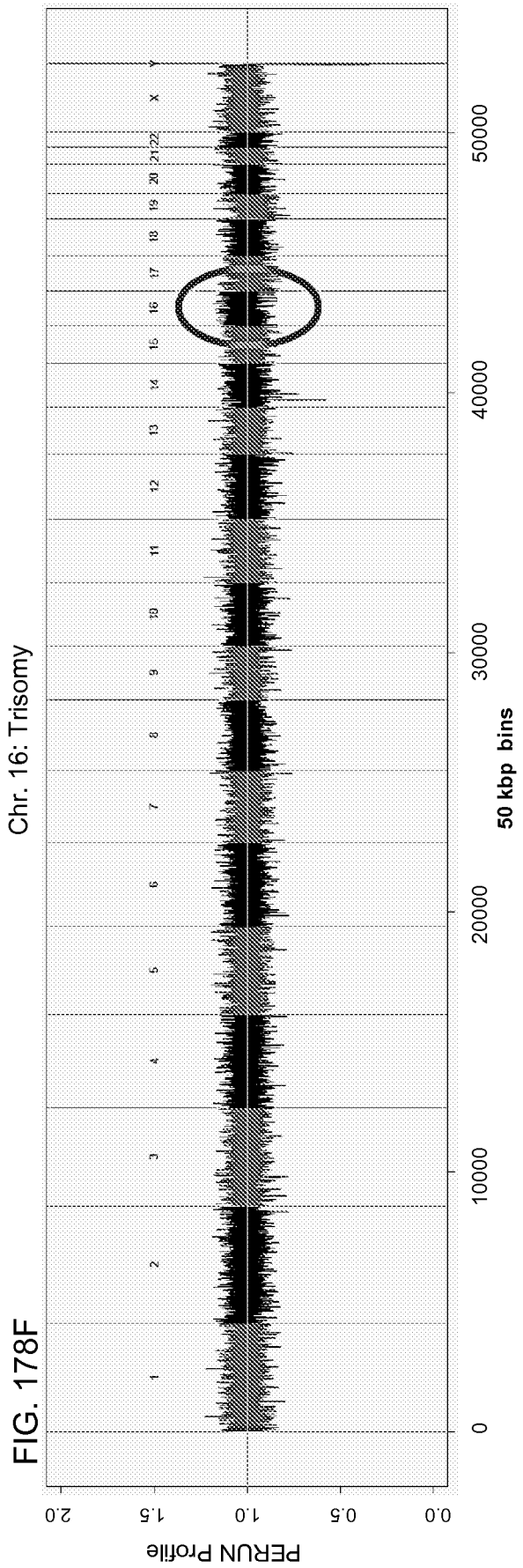


FIG. 178F

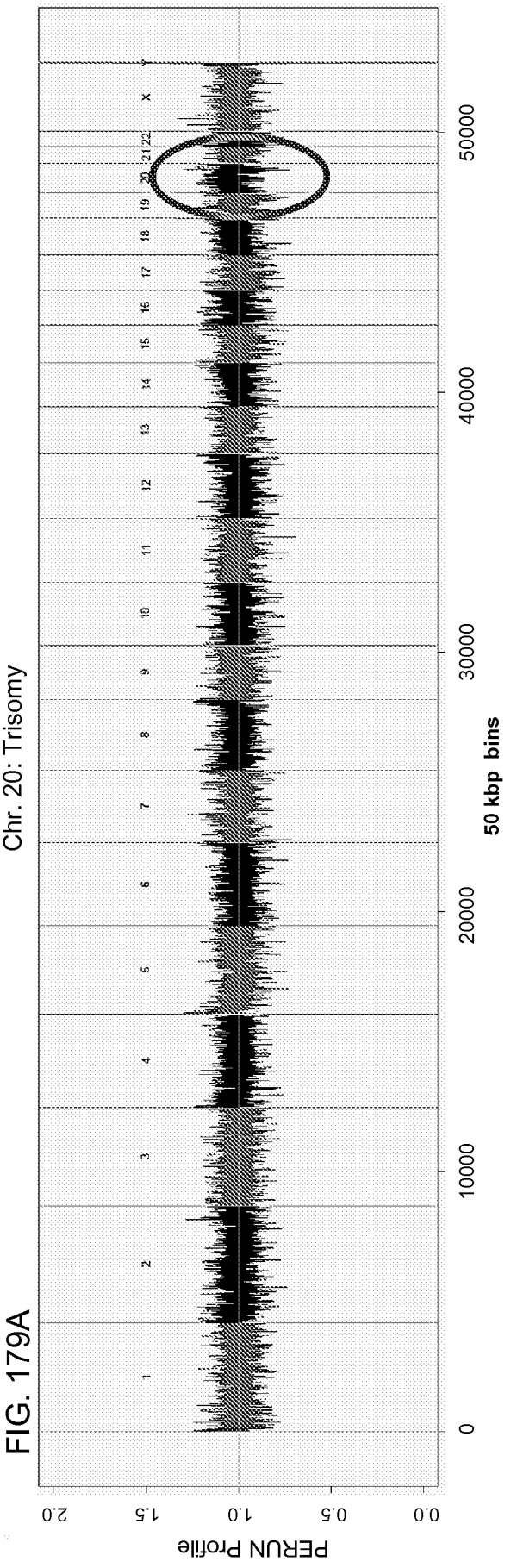


FIG. 179A

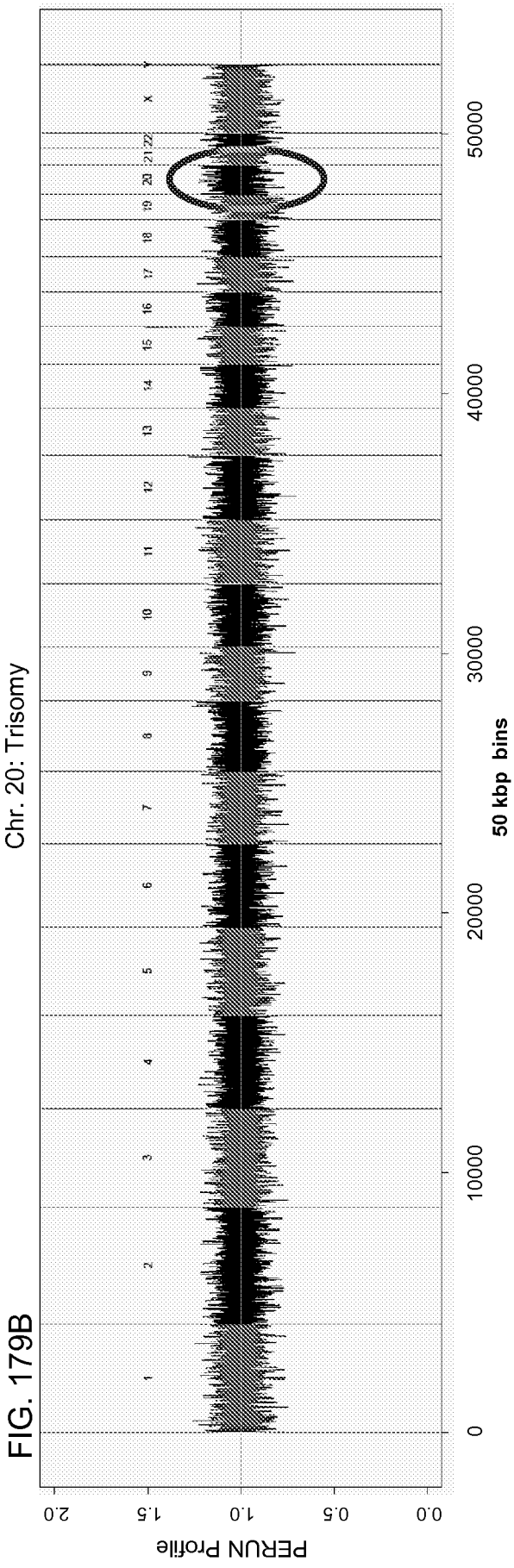


FIG. 179B

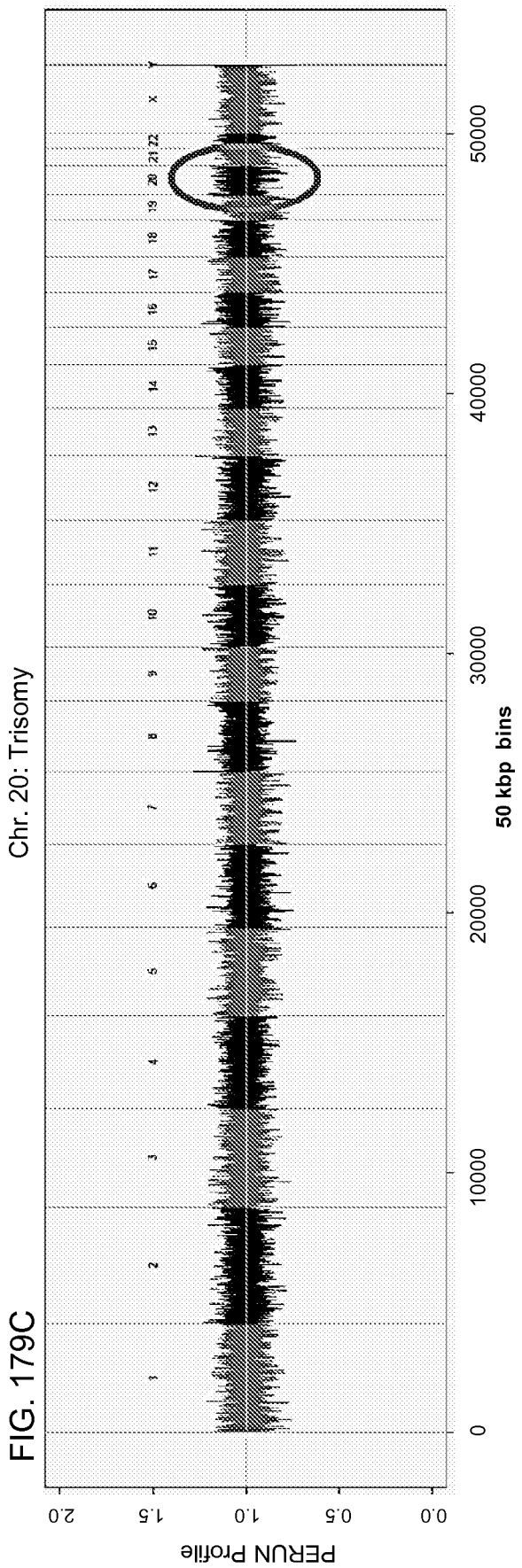


FIG. 179C

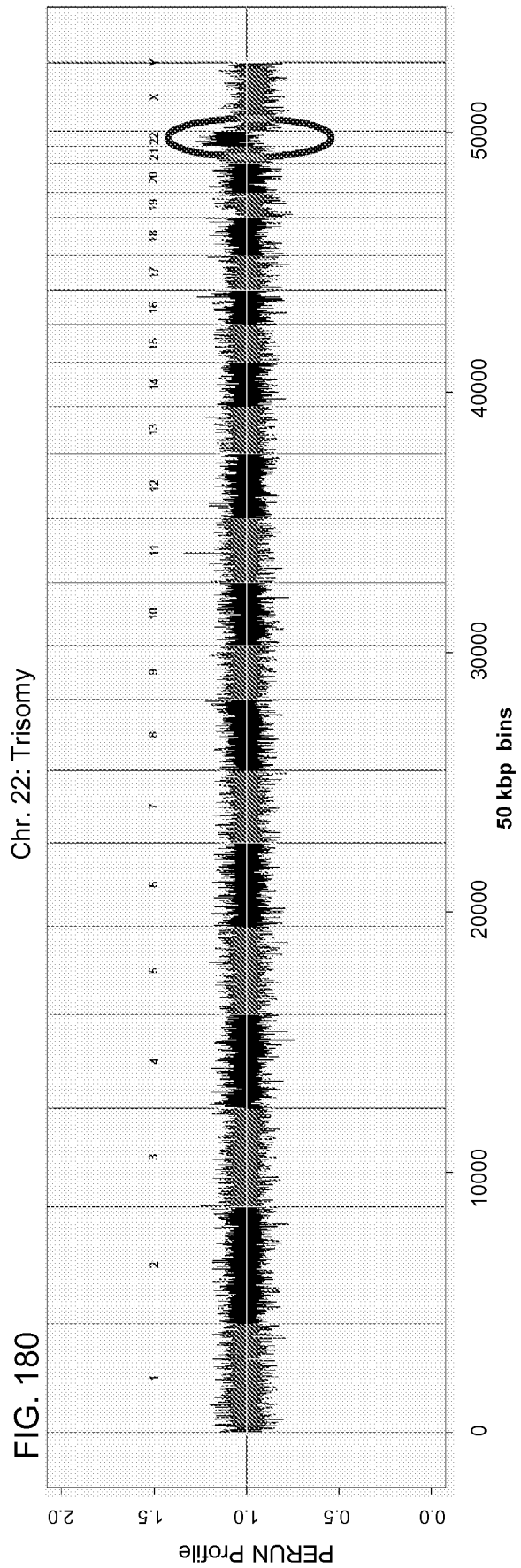


FIG. 180

FIG. 181A

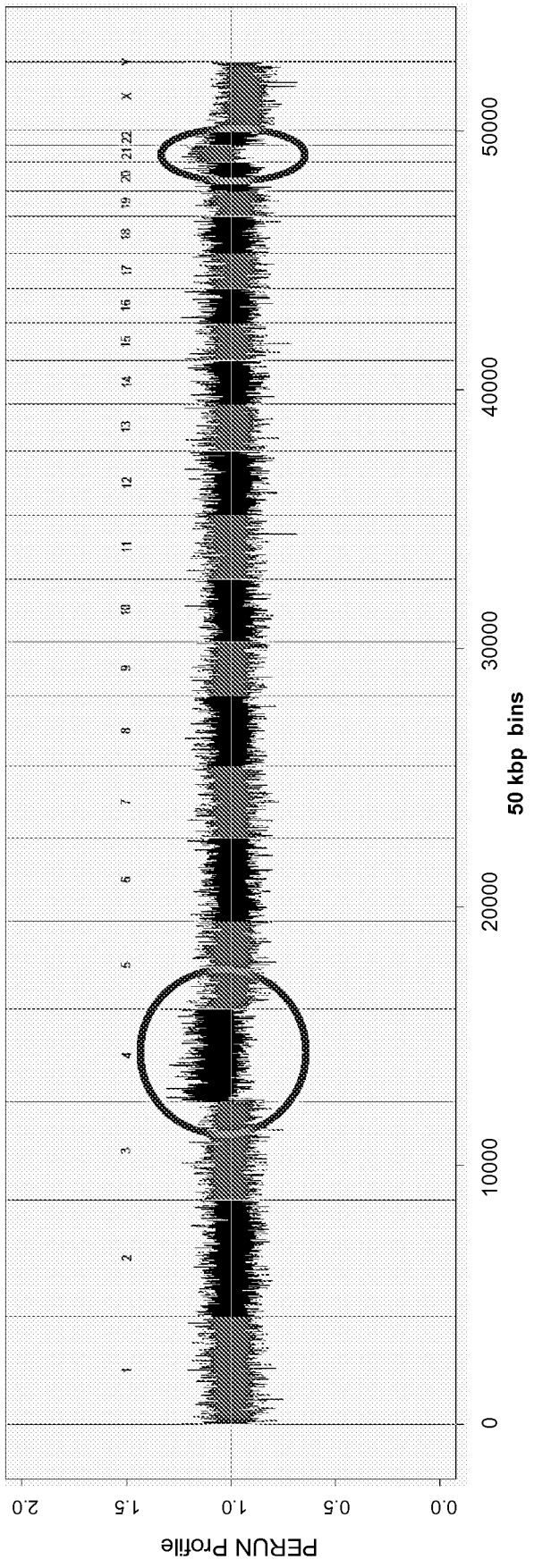
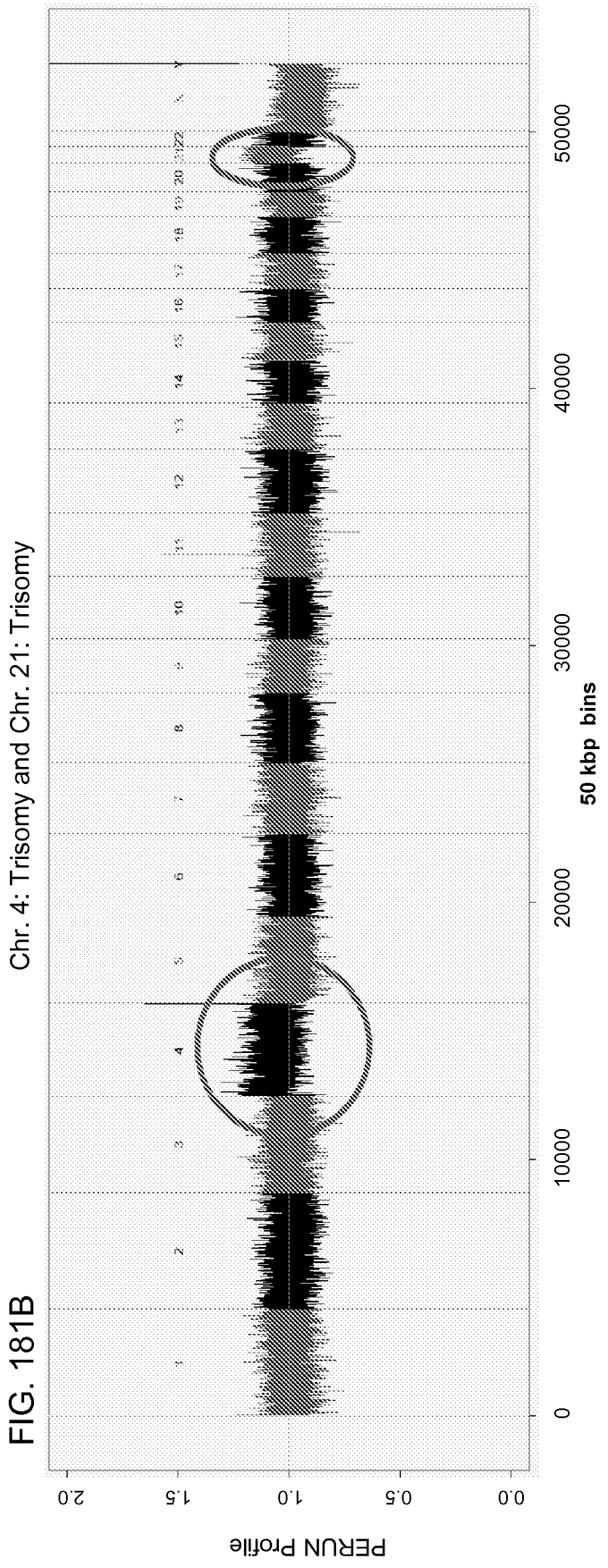


FIG. 181B



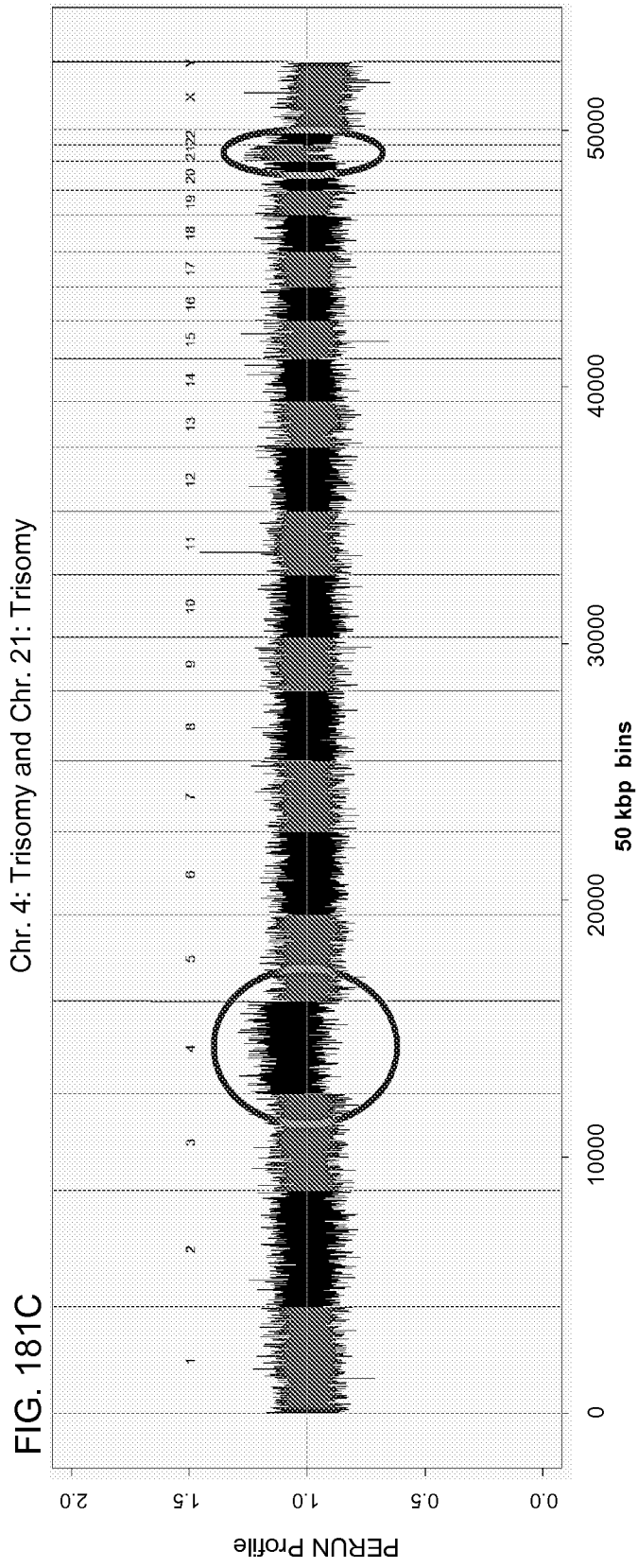


FIG. 181C

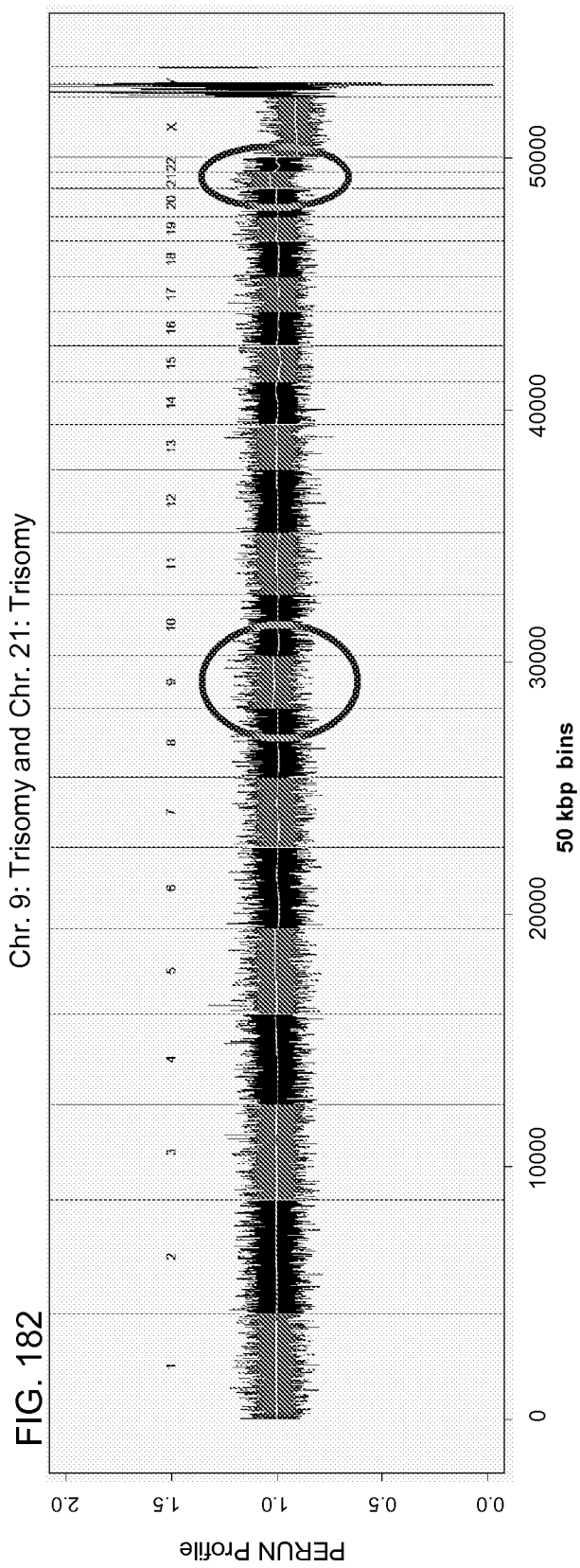


FIG. 182

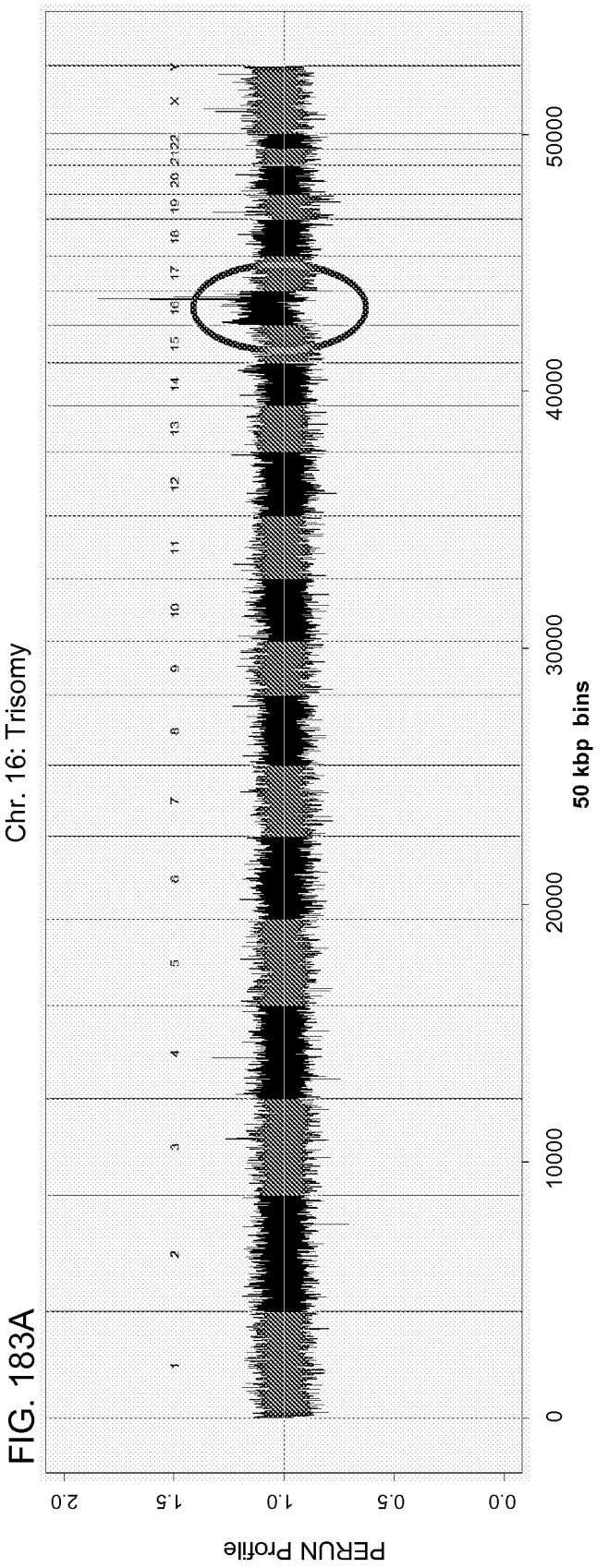


FIG. 183A

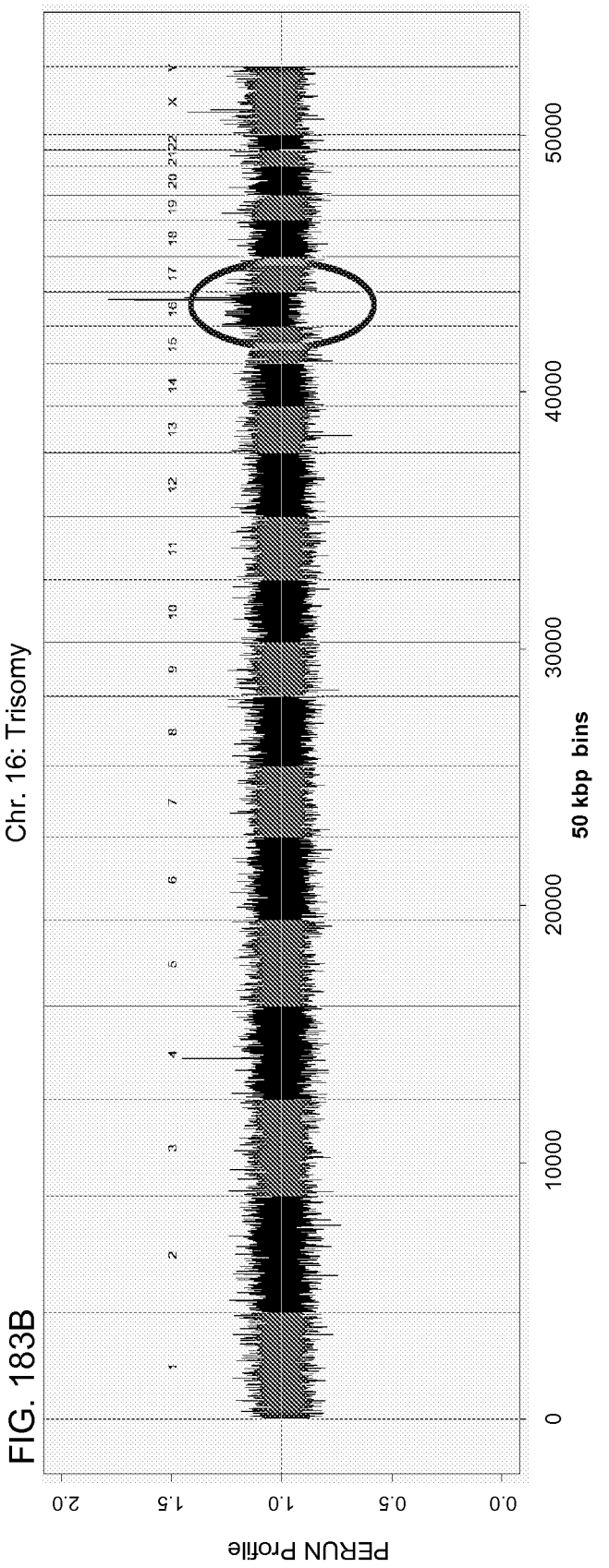


FIG. 183B

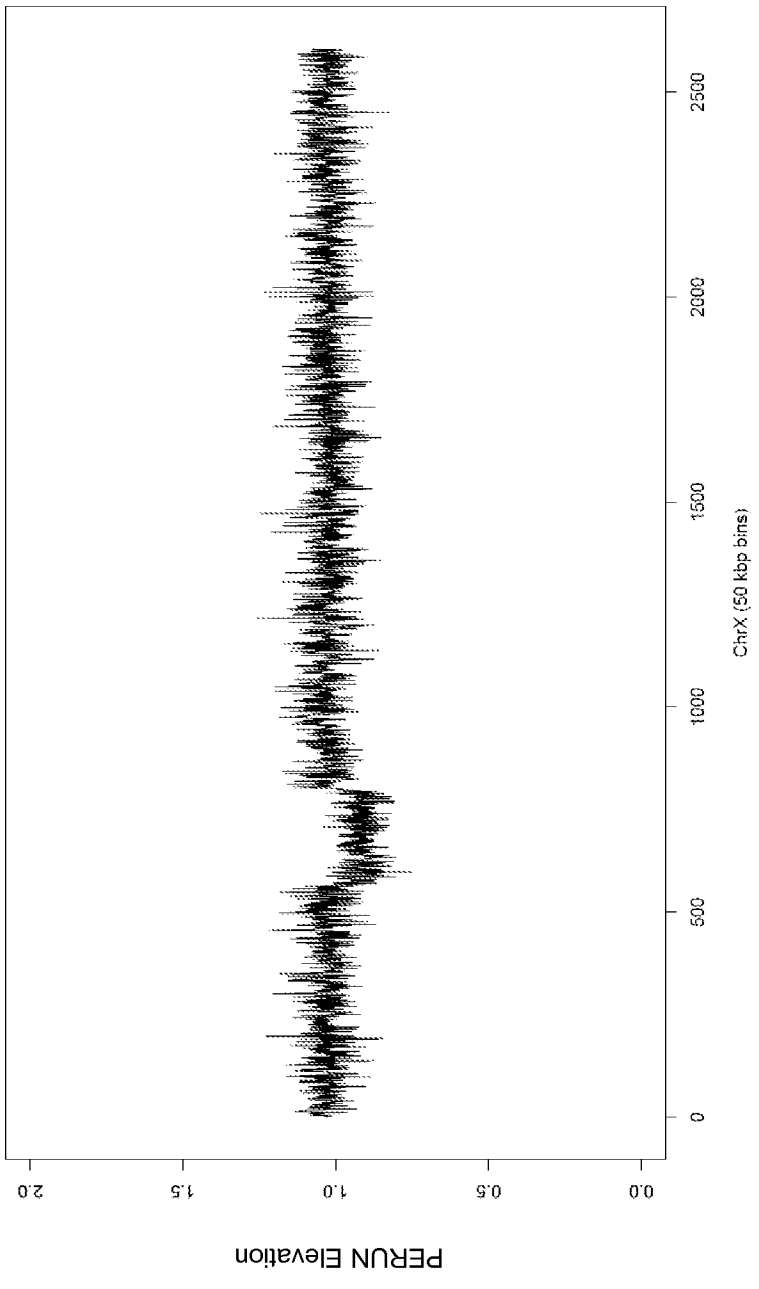

PERUN Elevation
0.0
0.5
1.0
1.5
2.0
0
500
1000
1500
2000
2500
ChrX (50 kbp bins)

FIG. 184A

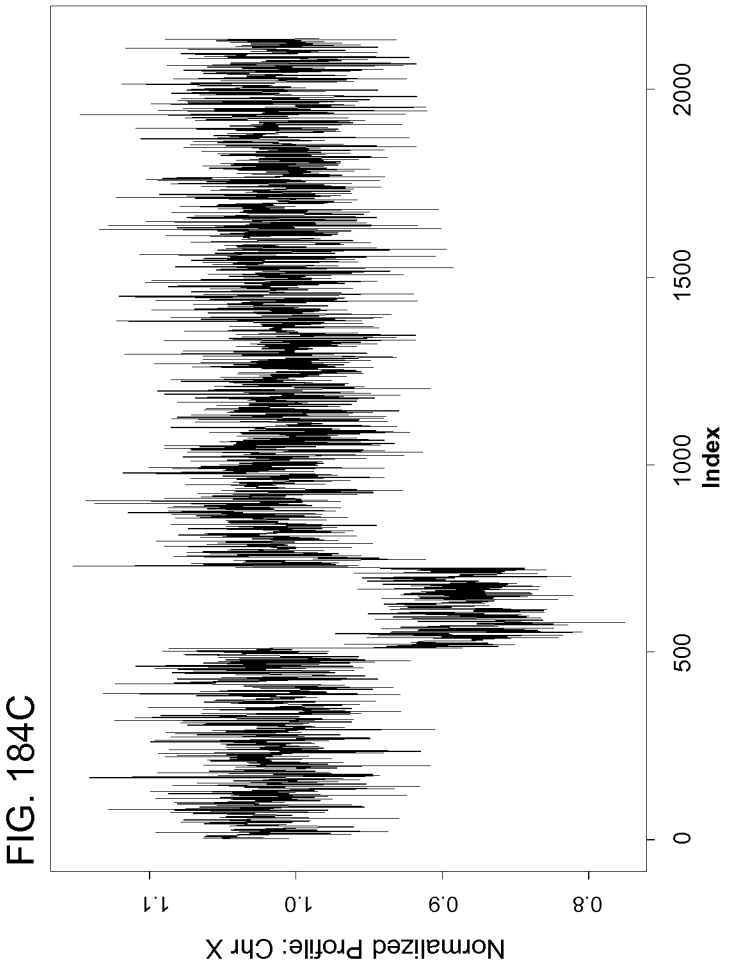


FIG. 184C

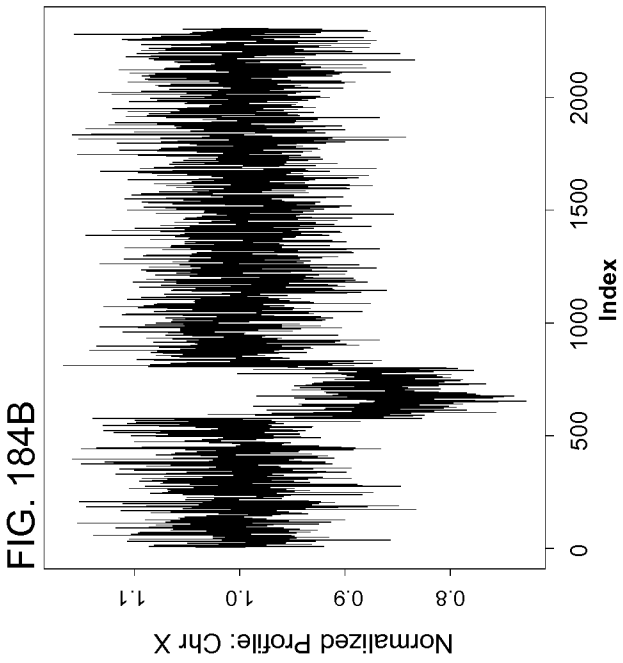


FIG. 184B

FIG. 185

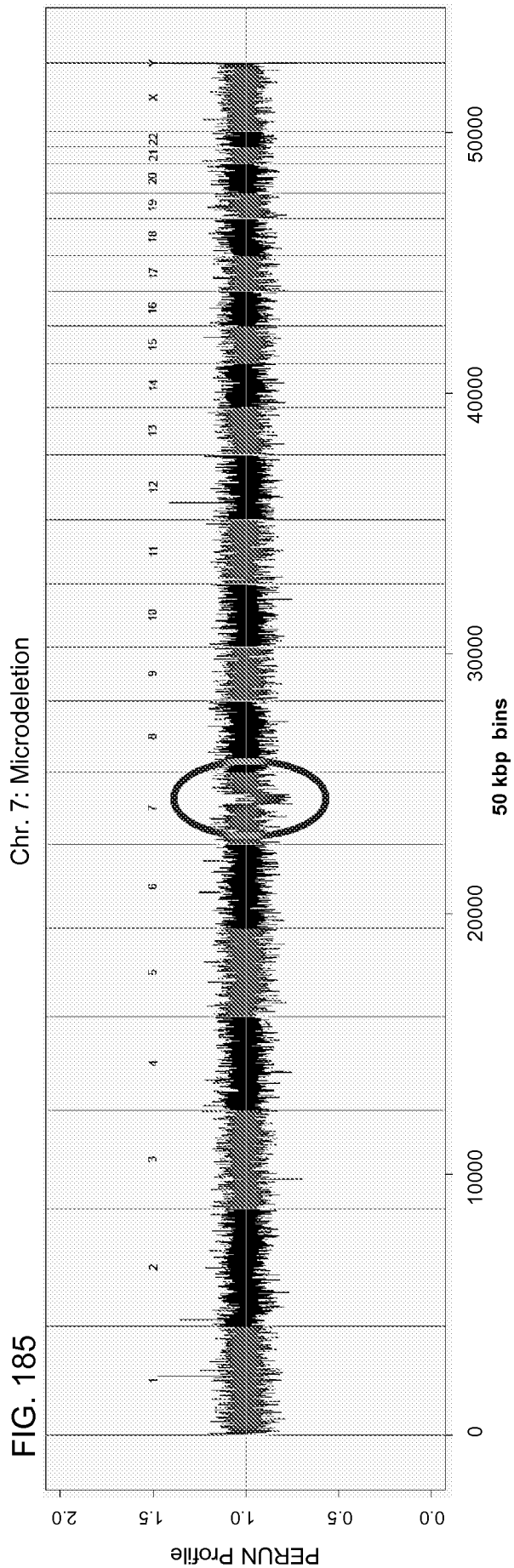
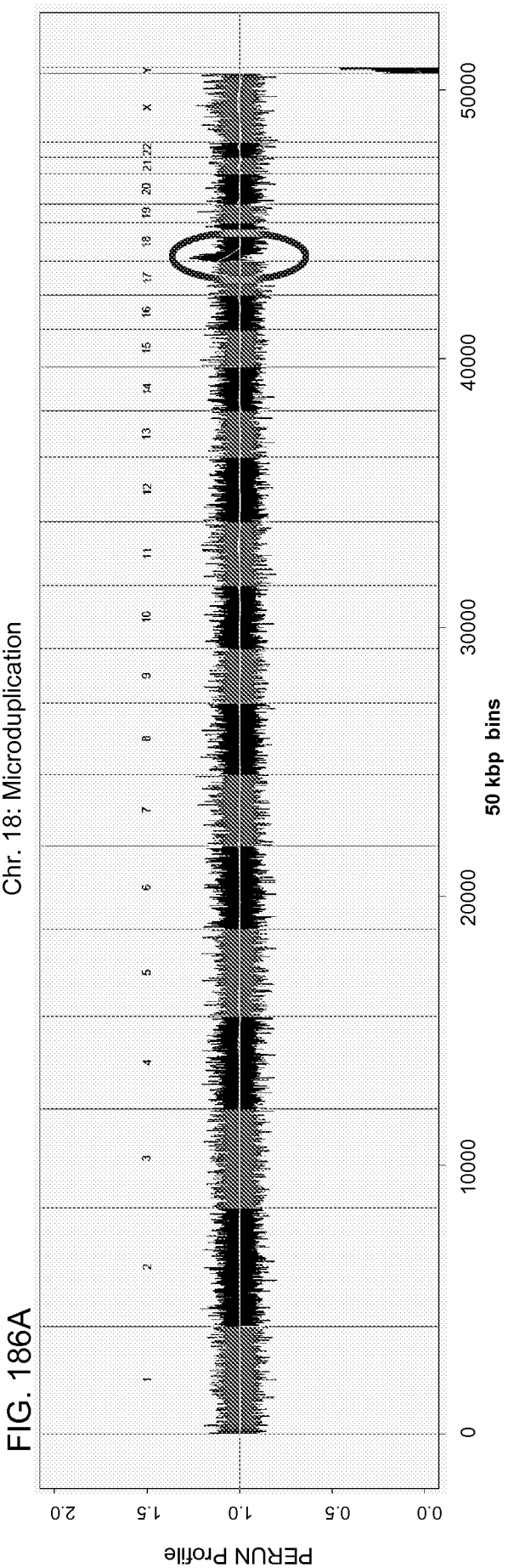


FIG. 186A



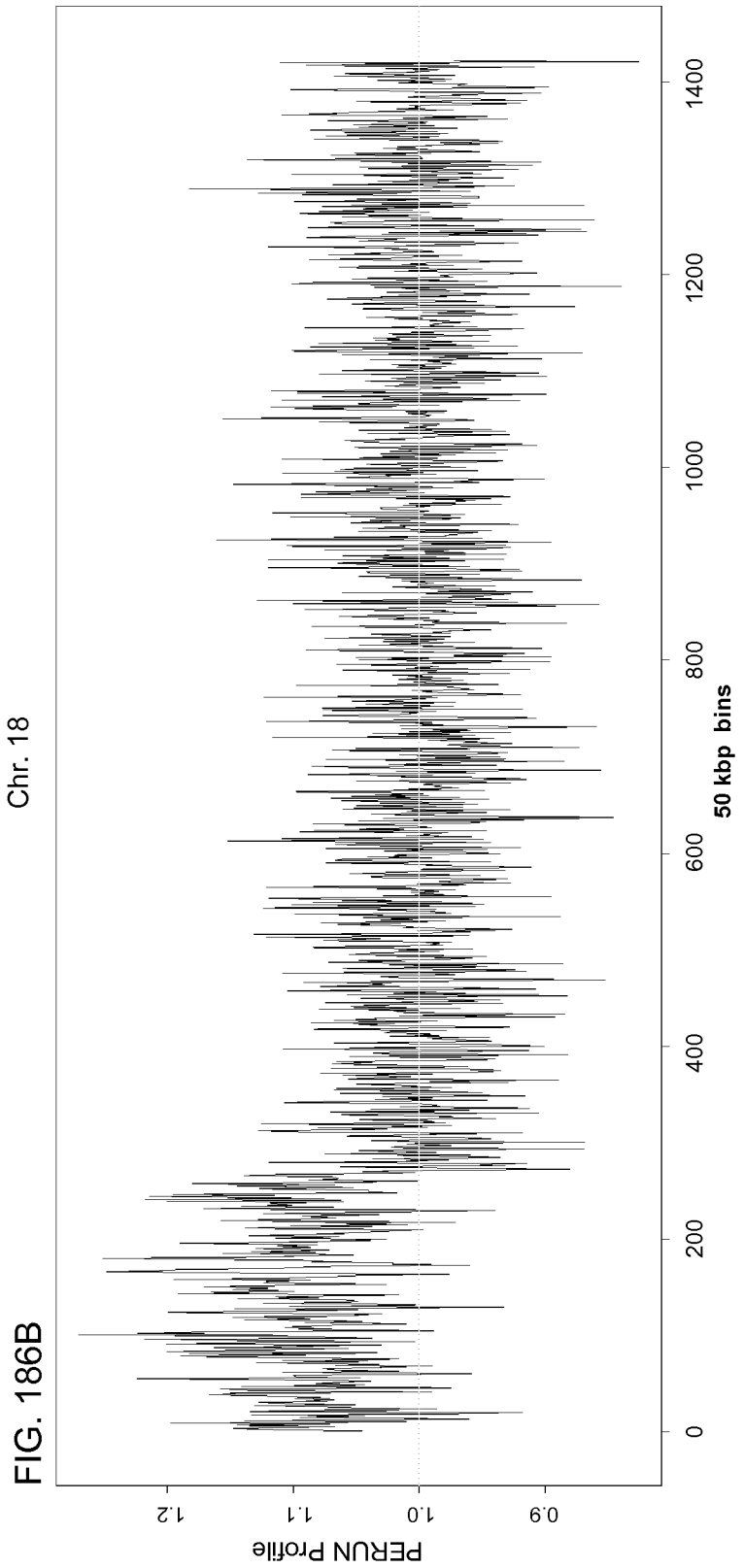

Chr. 18
PERUN Profile
0.9
1.0
1.1
1.2
0
200
400
600
800
1000
1200
1400
50 kbp bins

FIG. 186B

FIG. 187

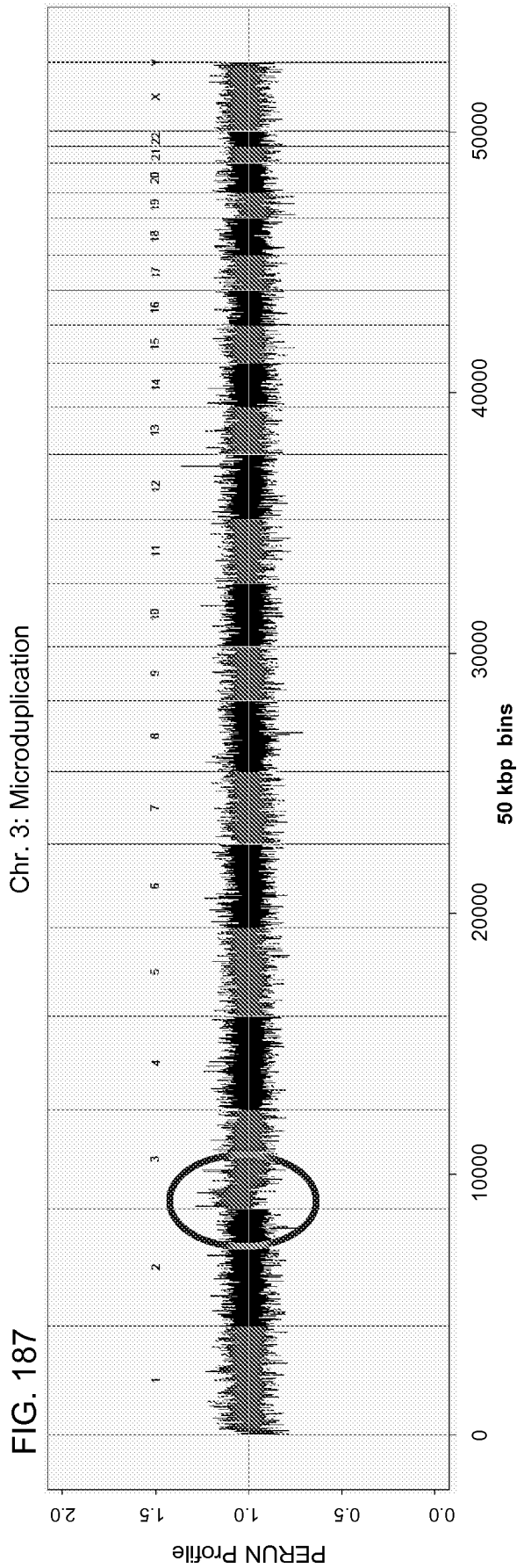
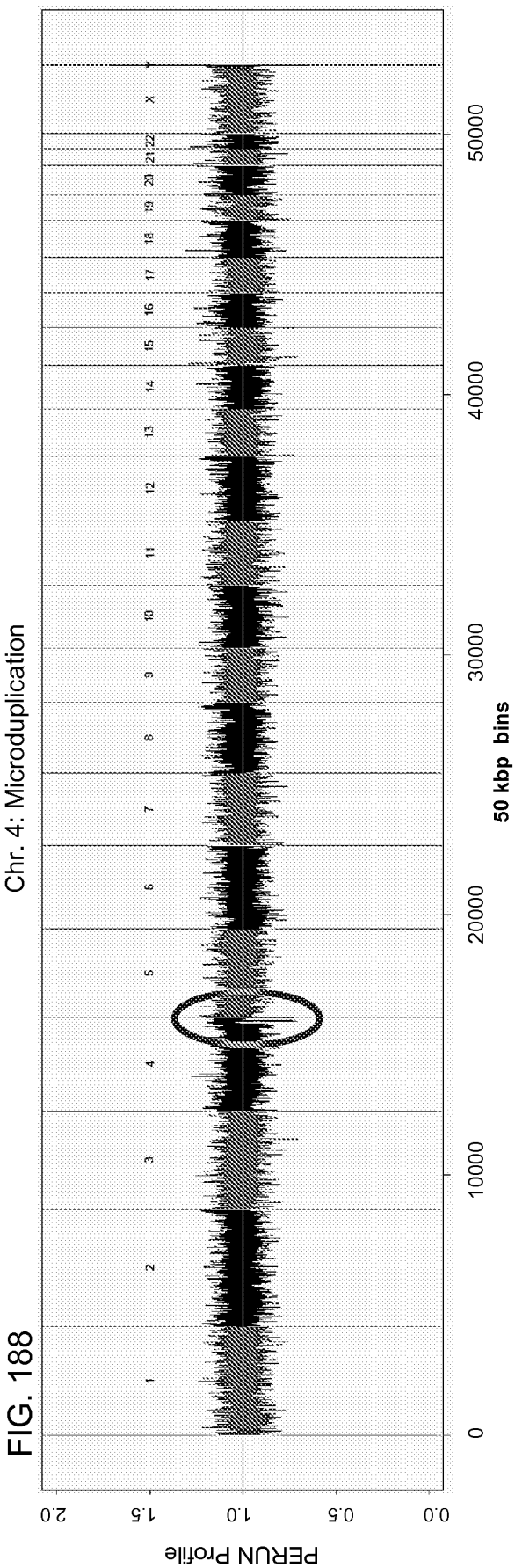


FIG. 188



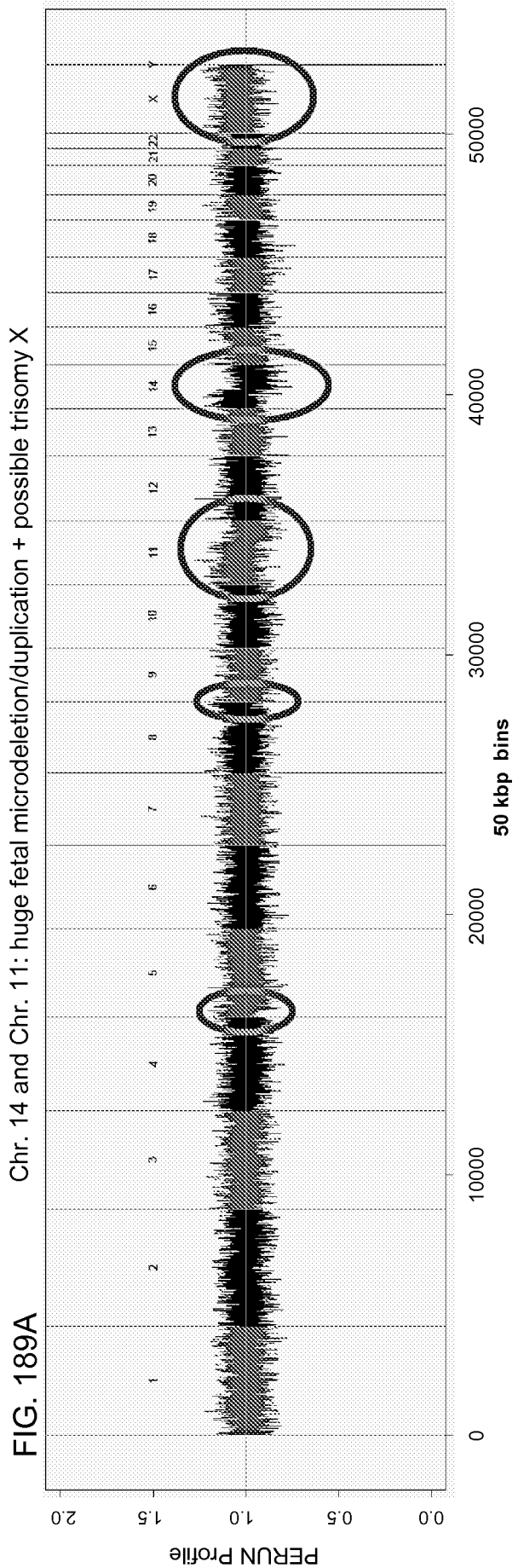


FIG. 189A

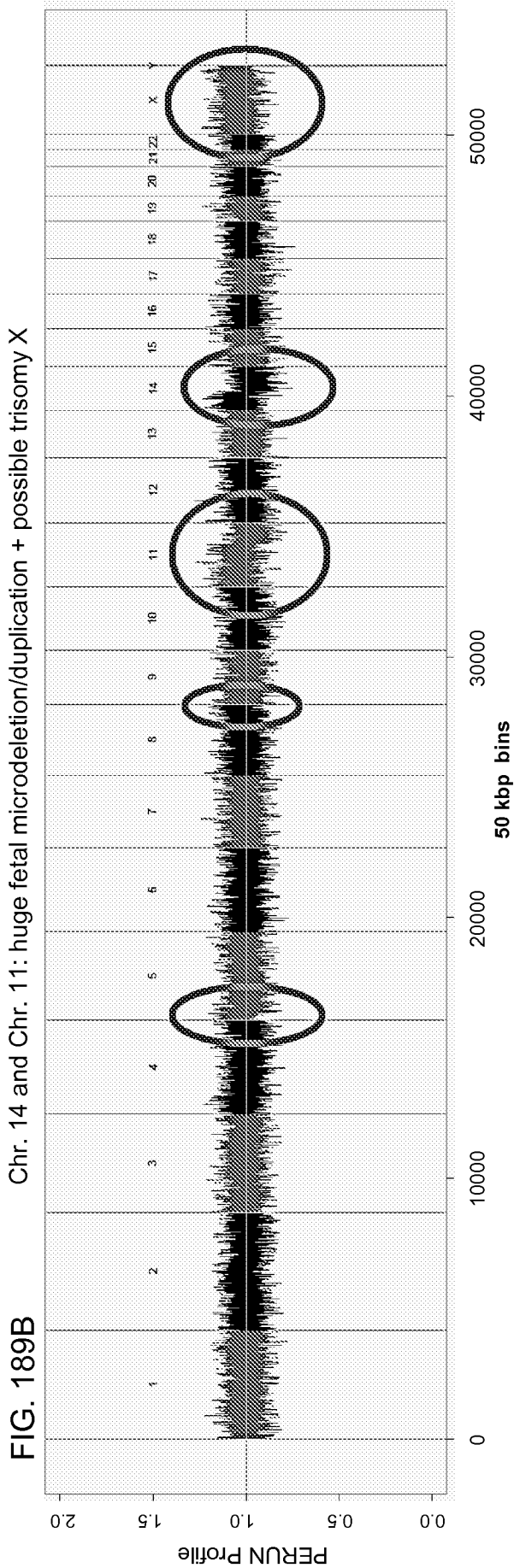
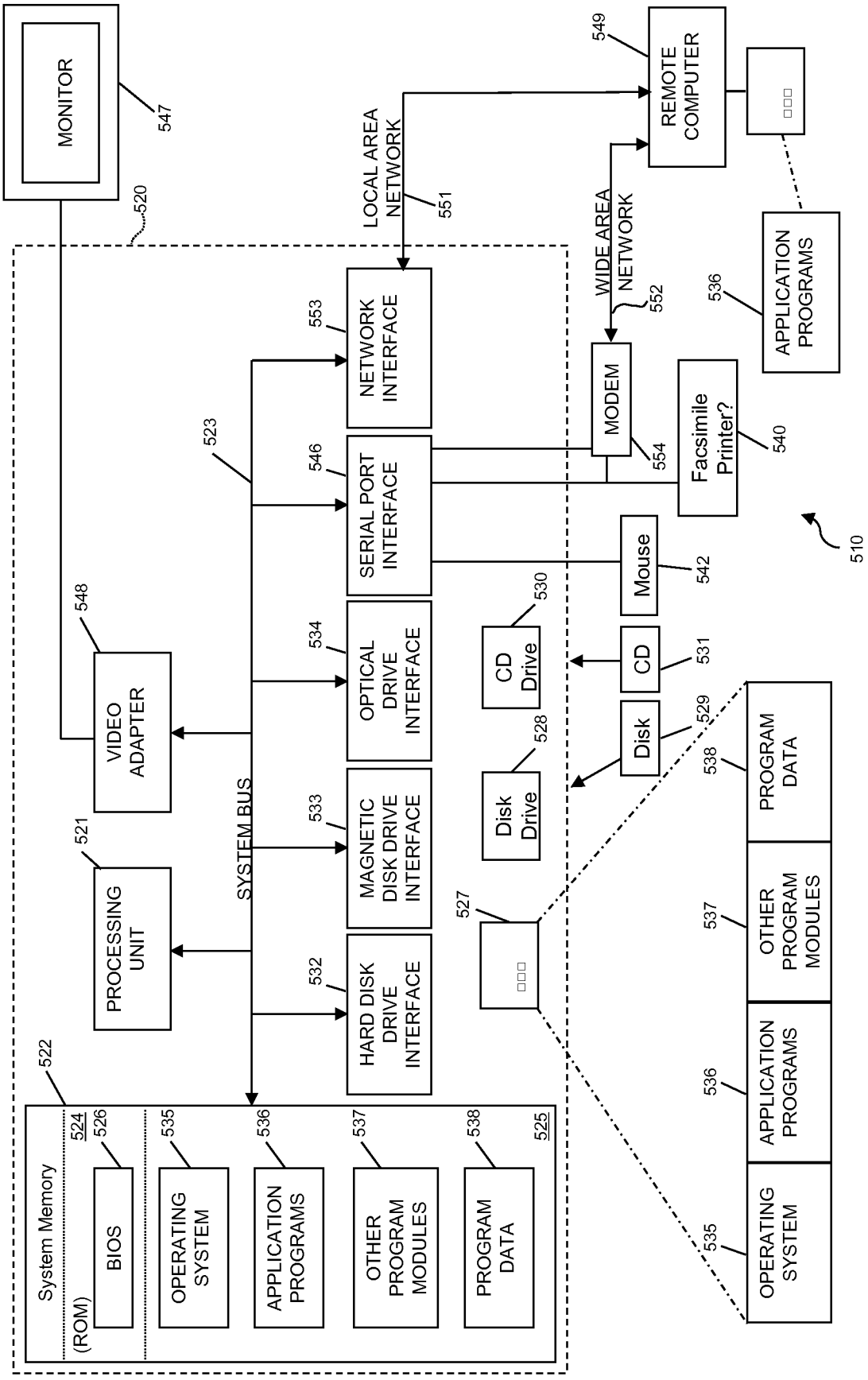


FIG. 189B

FIG. 190


MONITOR
547
520
LOCAL AREA NETWORK
551
549
REMOTE COMPUTER
WIDE AREA NETWORK
552
536
APPLICATION PROGRAMS
553
NETWORK INTERFACE
523
546
SERIAL PORT INTERFACE
MODEM
554
Facsimile Printer?
540
510
548
VIDEO ADAPTER
534
OPTICAL DRIVE INTERFACE
530
CD Drive
Mouse
542
531
CD
529
Disk
528
Disk Drive
521
PROCESSING UNIT
SYSTEM BUS
533
MAGNETIC DISK DRIVE INTERFACE
527
532
HARD DISK DRIVE INTERFACE
538
PROGRAM DATA
537
OTHER PROGRAM MODULES
536
APPLICATION PROGRAMS
535
OPERATING SYSTEM
522
System Memory
524
(ROM)
526
BIOS
535
OPERATING SYSTEM
536
APPLICATION PROGRAMS
537
OTHER PROGRAM MODULES
538
PROGRAM DATA
525