

【DESCRIPTION】

【Invention Title】

NOVEL METHOD CAPABLE OF DIFFERENTIATING FETAL SEX AND
FETAL SEX CHROMOSOME ABNORMALITY ON VARIOUS Next Generation
5 Sequencing PLATFORMS

【Technical Field】

[1] The present invention relates to a method for
aneuploidy analysis of fetal sex and fetal sex chromosomes
10 (for example, deletion and duplication of some regions in
the sex chromosome X and Y regions of fetal genes), and
discovery and verification thereof, using data obtained by
performing next-generation sequencing (NGS) from various
samples, wherein it is possible to obtain the same result
15 even though heterogeneous data from various platforms are
inputted.

[2]

【Background Art】

[3] From the 1970s until recently, diagnostic methods for
20 diseases related to chromosome numerical abnormality have
been continuously developed. In the past, amniotic
examination or chorionic villus examination was recommended
when the mother was 35 years or older. These techniques
are known to be invasive methods in which a fetal loss rate
25 is 1/100 to 1/300, accuracy is not high, and false positive

rate and false negative rate are high. 95% or more of pregnant women classified as a high risk group are in need of more accurate and safe diagnostic testing methods in order to give birth to healthy babies.

5 **[4]** In order to diagnose chromosome abnormalities and sex chromosome abnormalities, technologies with safety and high accuracy using next-generation sequencing analysis technology have been introduced from the late 2000s (Korean Patent Laid-open Patent Application No. 10-2010-7003969, No.
10 10-2012-7034453), wherein diagnosis of aneuploidy through sequencing of cell-free DNA (cfDNA) has advantages of high accuracy and low false negative rate, rather than traditional analysis methods.

[5] Recently, a non-invasive prenatal test (NIPT)
15 technology, which is an updated technology, is being carried out on the basis of the next-generation sequence analysis (NGS) (Bianchi DW et al., (2014) DNA Sequencing Versus Standard Prenatal Aneuploidy Screening. N. Engl. J. of Med. 370: 799-808., Sparks AB et al., (2012) Noninvasive
20 Prenatal Detection and Selective Analysis of Cell-Free DNA Obtained from Maternal Blood: Evaluation for Trisomy 21 and Trisomy 18. Am. J. Obstet. Gynecol. 206: 319.e1-9., Ehrich, M. et al., (2011) Noninvasive Detection of Fetal Trisomy 21 by Sequencing of DNA in Maternal Blood: A Study in a
25 Clinical Setting. Am. J. Obstet. Gynecol. 204: 205.e1-11.),

and a fast and efficient clinic analytical test is being conducted by using two platforms from Illumina and Ion Torrent companies. Sensitivity and specificity of the non-invasive prenatal test (NIPT) analysis using the Illumina platform at the beginning were about 98% (Lau, T.K. et al., (2014) Non-invasive Prenatal Testing for Fetal Chromosomal Abnormalities by Low-Coverage Whole Genome Sequencing of Maternal Plasma DNA: Review of 1982 Consecutive Cases in a Single Center. *Ultrasound Obstet. Gynecol.* 43: 254-264., Ashoor G et al., (2012) Chromosome-Selective Sequencing of Maternal Plasma Cell-Free DNA for First Trimester Detection of Trisomy 21 and Trisomy 18. *Am. J. Obstet. Gynecol.* 206: 322.e1-5.). However, recent studies have shown that aneuploidy diagnostic sensitivity of fetal chromosomes 13, 18, and 21 using the HiSeq (Illumina) and the Ion Proton (Ion Torrent) is higher than 99.5% (Liao, C., et al., (2014) Noninvasive Prenatal Diagnosis of Common Aneuploidies by Semiconductor Sequencing. *Proc. Natl. Acad. Sci. USA.* 111: 7415-7420.).

[6] A number of chromosome related clinic diseases are associated with the number of copies of sex chromosomes as well as autosomal changes. The most common sex chromosome aneuploidy (SCA) is Turner syndrome [45, X], XXX syndrome (Triple-X [47, XXX]), Klinefelter syndrome [47, XXY], and Super male syndrome ([47, XYY] syndrome). These diseases

are known to be have a relatively rare incidence, but have been reported that these diseases cause cumulative sex chromosome aneuploidy of about 0.3% (Morris, J.K., et al., Is The Prevalence of Klinefelter Syndrome Increasing; Eur. J. Hum. Genet. 2007; 16: 163- 70.)). In a study of 115 patients in the pediatric department of Kyungpook National University Hospital for 20 years, the frequency of sex chromosome abnormality syndrome was 52% in Klinefelter syndrome, 42% in Turner syndrome, 3% in XXX syndrome, and 3% in mixed gonadal dysplasia. Klinefelter syndrome has 97% of typical XXY type and 3% of mosaicism, Turner syndrome has 67% of typical X type, 23% of mosaicism, and 10% of structural abnormality of the X chromosome. XXX syndrome has 67% of XXX type and 33% of mosaicism, and the mixed gonadal dysplasia is all 45,X/46,XY (Seo, Hyun ji et al., a cytogenetic study in patients with sex chromosome abnormalities, Korea Journal of Pediatrics, Vol 48, No. 12, 2005).

[7] Recently, sex chromosome aneuploidy is detectable by using fetal cell DNA (cell-free fetal DNA, cffDNA) and NGS, but is detected by quantifying chromosome numbers with lower accuracy than autosomal aneuploidy. Further, sequence similarity between chromosomes X and Y and sequencing bias associated with GC composition are present, and similarity between chromosome Y and other chromosomes

is reduced at a signal-to-noise ratio, and the chromosome Y having a small size has many variations. Most of the cases are not easily detected, and always have a problem in classification of sex chromosome aneuploidy.

5 **[8]** Unlike the aneuploidy on chromosomes 13, 18, and 21, it is a global trend that accuracy of judgment for XO, XXX, and XXY is not high. In fact, in the cases of Verifi (Illumina) and NIFTY (BGI), the sensitivity on chromosomes 13, 18, and 21 is 98.15% or more. However, the Verifi has
10 95% of the sensitivity of Monosomy X for sex chromosomes, and for XXX/XXY/XYY, the amount of data is small, and thus accuracy is not represented. In the case of NIFTY, the sensitivity of Monosomy X is represented as 75% based on the modified internal chromosome Z-score rather than based
15 on the existing Z-score, and the remainder is not represented due to a small number of data. Even though the amount of data is small, a current algorithm does not have high accuracy, and thus it is necessary to develop an optimized new algorithm. In addition, current methods have
20 been developed on the basis of each specific platform, and thus it is not possible to apply data of other platforms and to perform normalization using two or more data.

[9] In addition, it is known that up to 4% of fetal DNA is able to be detected accurately which is similar to fetal
25 autosome aneuploidy (Chiu RW et al., Maternal Plasma DNA

Analysis With Massively Parallel Sequencing by Ligation for
Noninvasive Prenatal Diagnosis of Trisomy 21. Clin. Chem.
2009;56:459-63., Palomaki, G.E., et al., DNA Sequencing Of
Maternal Plasma To Detect Down Syndrome: An International
5 Clinical Validation Study. Genet Med. 2011;13:913-20.), and
thus better methods and classification of data sets are
needed.

[10] Therefore, in order to solve problems of gender
classification and sex chromosome aneuploidy, the present
10 inventors found that when LOESS normalization of a GC ratio
and mappability were performed, chrX and chrY values were
determined using percentages of chrX and chrY, the
reference line was adjusted by using internal chromosome
reference to determine final Z-score regardless of the
15 kinds of platform and data, it was possible to
differentiate sex and sex chromosome abnormality with high
accuracy and sensitivity, and completed the present
invention.

[11]

20 **[12] 【Summary of Invention】**

[13] An object of the present invention is to provide a
method of differentiating fetal sex and fetal sex
chromosome abnormality by simultaneously normalizing and
analyzing all inputted data at once regardless of the type
25 of sample and platform.

[14]

[15] In order to achieve the foregoing objects, the present invention provides a method of differentiating fetal sex and fetal sex chromosome abnormality including:

- 5 (a) performing next-generation sequencing (NGS) on two or more platforms by extracting DNA from maternal blood; (b) aligning sequences (reads) that are sequenced for each platform to a reference genome database; (c) extracting unique reads that are uniquely aligned on chromosome for
10 each platform and calculating a percent of unique reads (UR%) relative to the total reads; (d) calculating data for each platform at the same time, setting a cut-off value, and organizing the UR(y)% values; (e) classifying male and female using the organized UR(y)%; (f) calculating partial
15 reference chromosome (relative FGR) values using three or more autosomes that are similar in size to chromosomes X and Y; and (g) calculating Z-score using the partial reference chromosome (relative FGR) values, and calculating G score (Gscore) and M score (Mscore) based on the Z-score
20 to differentiate sex chromosome abnormality.

[16]

[Description of Drawings]

[17] FIG. 1 is a whole flow chart for diagnosis of sex chromosome aneuploidy of the present invention.

25 **[18]** FIG. 2 shows analysis results of sex chromosome

Unique Read% of samples analyzed using the HiSeq platform (Illumina). The left vertical axis represents UR(x)%, and the right vertical axis represents UR (y)%.

[19] FIG. 3 shows comparison between analysis results of sex chromosome Unique Read% of two experimental sets using the HiSeq platform (Illumina).

[20] FIG. 4 shows comparison between analysis results of sex chromosome Unique Read% of two experiment sets using Ion Proton platform (Ion torrent).

[21] FIG. 5 is a graph showing results of gender classification on the basis of cut-off values of 51 samples using the Illumina platform.

[22] FIG. 6 is a graph showing results of gender classification on the basis of the same cut-off values as in FIG. 5 in two different experimental sets analyzed using the Ion Proton platform (Ion Torrent).

[23] FIG. 7 is a graph showing Z-score values calculated by using only chromosome 7 (chr7) and chromosome 20 (chr20) as internal references.

[24] FIG. 8 is a graph showing Mscore calculated by using four internal references.

[25] FIG. 9 is a graph showing Gscore calculated by using four internal references.

[26] FIG. 10 is a graph showing two platforms (HiSeq of Illumina and Ion Proton of Ion torrent) from which PCR

duplication is removed.

[27] FIG. 11 is a graph showing percentages of unique reads on two platforms (HiSeq of Illumina and Ion Proton of Ion torrent).

5 **[28]** FIG. 12 is a flowchart showing a method of simultaneously performing sequencing on two platforms with one sample.

[29] FIG. 13 is a graph showing results of sex chromosome abnormality differentiated by Gscore when classifying
10 female in samples according to an exemplary embodiment of the present invention..

[30] FIG. 14 is a graph showing results of sex chromosome abnormality differentiated by Mscore when classifying male
15 in samples according to an exemplary embodiment of the present invention.

[31] FIG. 15 is a graph showing results of differentiation of sex chromosome abnormality using chrX and chrY by adding samples according to an exemplary embodiment of the present invention.

20 **[32]**

[33] **【Best Mode】**

[34] Unless defined otherwise, all the technical and scientific terms used herein have the same meanings as those generally understood by persons skilled in the art to

25 which the present invention pertains. In general,

nomenclature used in the present specification and experimental methods to be described below are well known and generally used in the present technical field.

[35]

5 **[36]** In the present invention, it was confirmed that when sequencing data generated from two or more NGS platforms were normalized and aligned based on the cut-off values, and then the unique reads % aligned on the sex chromosome using partial reference chromosome was used, it was
10 possible to differentiate fetal sex and sex chromosome abnormality with high accuracy and sensitivity.

[37] In other words, in an exemplary embodiment of the present invention, a method of sequencing genomic DNA (gDNA) extracted from maternal blood using two or more NGS
15 platforms, quantifying an amount of the unique reads for each chromosome to set a cut-off value based on a GC content and mappability, using chr7 to chr10 as partial reference chromosomes for quantification of chromosome X and chr19 to chr22 as partial reference chromosomes for
20 quantification of chromosome Y, and in the case of a male, calculating Mscore (Man-score) to determine normal (XY) and Klinefelter syndrome (XXY), and in the case of a female, calculating Gscore (Girl-score) to diagnose XXX syndrome (Triple-X, XXX) as well as Turner's syndrome (XO) and
25 normal (XX) (FIG. 1), has been developed.

[38] Therefore, in one aspect, the present invention relates to a method of differentiating fetal sex and fetal sex chromosome abnormality including: (a) performing next-generation sequencing (NGS) on two or more platforms by
5 extracting DNA from a maternal biological sample; (b) aligning sequences (reads) that are sequenced for each platform to a reference genome database; (c) extracting unique reads that are uniquely aligned on chromosome for each platform and calculating a percent of unique reads
10 (UR%) relative to the total reads; (d) calculating data for each platform at the same time, setting a cut-off value, and organizing the UR(y)% values; (e) classifying male and female using the organized UR(y)%; (f) calculating partial reference chromosome (relative FGR) values using three or
15 more autosomes that are similar in size to chromosomes X and Y; and (g) calculating Z-score using the relative FGR values, and calculating G score (Gscore) and M score (Mscore) based on the Z-score to differentiate sex chromosome abnormality.

20 **[39]**

[40] In the present invention, the biological sample may be selected from the group consisting of blood, plasma, serum, urine, and saliva.

[41] In the present invention, step (a) may include: (a)
25 separating a plasma by centrifuging the collected

biological sample; (b) extracting cfDNA from the separated plasma; (c) preparing a library using the extracted cfDNA; and (d) pooling the prepared library, and then decoding the base sequence using NGS equipment.

5 **[42]** The pooling the library of the present invention may be characterized by performing bead size selection in the case of the Illumina platform and performing E-gel size selection in the case of the Life technologies platform.

10 **[43]** In the present invention, step (c) may include: (a) removing PCR duplication and ambiguous reads in sequencing, aligning only perfectly matched reads, and extracting a read in which a GC content and mappability are more than or equal to the cut-off value in the aligned sequence; and (b) calculating a percentage (UR(x)%) of the read that uniquely
15 matches the chromosome X and a percentage (UR(y)%) of the read that uniquely matches the chromosome Y according to the following Equations, based on the extracted read:

[44] Equation 1) $UR(x)\% = \frac{\text{the number of chromosome X (chrX) reads}}{\text{total number of autosome reads}} \times 100$, and

20 **[45]** Equation 2) $UR(y)\% = \frac{\text{the number of chromosome Y (chrY) reads}}{\text{total number of autosome reads}} \times 100$.

[46] The removal of duplicated reads of the present invention may be performed by any method for removing duplication, but is preferably performed through LOESS
25 normalization.

[47] In the present invention, the step of analyzing the GC content and the mappability may be performed by dividing each region into 100 kb bin regions, wherein the cut-off value of the GC content may be 0.30 to 0.50, and the cut-off value of the mappability may be 20% or more.

[48]

[49] In the present invention, the normalization of UR(y)% is performed on the basis of the chromosome Y in order to perform accurate classification between the chromosomes XX and XY, wherein results of a single sample (FIG. 2) may be divided into Cut-Off criteria of UR(y)%, but when the normalization is performed on a mass sample of two or more samples or another platform sample, it is not possible to set the cut-off of UR(y)% (FIGS. 3 and 4). In order to not only obtain results of various samples in one laboratory and but also take advantage of many results that are already analyzed, it is necessary to normalize the results at various platforms and at different places at once and then to classify male and female according to the Cut-Off criteria, and thus it is characterized by performing normalization by correcting UR(y)% according to the following equation, the UR(y)% being calculated for each platform:

[50] Equation 3) $UR(y)\%_{new} = (UR(y)\%_{original} - (\text{mean value of the three smallest } UR(y)\% \text{ values}))$

[51] In the present invention, the cut-off value of (d) is preferably in the range of 0.02 to 0.06, and the most preferably 0.037.

[52] In the present invention, the gender classification
5 of (e) may be characterized in that when the organized UR(y)% value is more than or equal to the cut-off value, it is determined as a male, and when the organized UR(y)% value is smaller than the cut-off value, it is determined as a female.

10 **[53]**

[54] In the present invention, in order to detect sex chromosome aneuploidy, chromosomes chr19 to chr22 that are similar in size to chromosome Y are selected as internal chromosome references, chromosomes chr7 to chr10 that are
15 similar in size to chromosome X are selected as internal chromosome references, and a partial reference chromosome (relative FGR) is calculated (Lau, T.K., et.al, J. Matern. Fetal Neonatal Med. 2012 Aug;25(8):1370-4.). From the paper in which Chr7 and chr20 are only selected, incorrect
20 results are obtained since sequencing bias is so severe (FIG. 7), but it may be confirmed that a mean value of the respective values from the FGR(7) to the FGR(10) is calculated, and thus an error rate is significantly reduced (FIGS. 8 and 9).

25 **[55]** Therefore, in the present invention, step (f) may

include:

[56] (a) selecting autosomes (chr7, chr8, chr9, and chr10) that are similar in size to sex chromosome X (chrX) as internal reference chromosomes and calculating each FGR by

5 the following Equations:

[57] Equation 4) $FGR(7) = (UR(x)\% + UR(y)\%) / UR(7)\%$

[58] Equation 5) $FGR(8) = (UR(x)\% + UR(y)\%) / UR(8)\%$

[59] Equation 6) $FGR(9) = (UR(x)\% + UR(y)\%) / UR(9)\%$

[60] Equation 7) $FGR(10) = (UR(x)\% + UR(y)\%) / UR(10)\%$,

10 and

[61] (b) selecting autosomes (chr19, chr20, chr21, and chr22) that are similar in size to sex chromosome Y (chrY) as partial reference chromosomes and calculating each FGR by the following Equations:

15 **[62]** Equation 8) $FGR(19) = (UR(x)\% + UR(y)\%) / UR(19)\%$

[63] Equation 9) $FGR(20) = (UR(x)\% + UR(y)\%) / UR(20)\%$

[64] Equation 10) $FGR(21) = (UR(x)\% + UR(y)\%) / UR(21)\%$

[65] Equation 11) $FGR(22) = (UR(x)\% + UR(y)\%) / UR(22)\%$.

[66]

20 **[67]**

[68] In the present invention, step (g) may include: (a) differentiating sex chromosome abnormality by the following steps when the sex of the fetus is determined to be female; and

25 **[69]** a-1) calculating Z-score for each chromosome in the

following Equations:

[70] Equation 12) $Zscore(7) = \{FGR(7) - (mean\ FGR(7))\} / (stdev\ FGR(7))$

[71] Equation 13) $Zscore(8) = \{FGR(8) - (mean\ FGR(8))\} / (stdev\ FGR(8))$

[72] Equation 14) $Zscore(9) = \{FGR(9) - (mean\ FGR(9))\} / (stdev\ FGR(9))$

[73] Equation 15) $Zscore(10) = \{FGR(10) - (mean\ FGR(10))\} / (stdev\ FGR(10))$

10 **[74]**

[75] a-2) calculating Gscore using the calculated Z-score by the following Equation; and

[76] Equation 16) $Gscore = \{Zscore(7) + Zscore(8) + Zscore(9) + Zscore(10)\} / 4$

15 **[77]**

[78] a-3) determining that the sex chromosome is XXX when the Gscore is more than or equal to 1.5, that the sex chromosome is XX when the Gscore is more than -3 and less than 1.5, and that the sex chromosome is XO when the Gscore

20 is less than or equal to -3;

[79]

[80] (b) differentiating sex chromosome abnormality by the following steps when the sex of the fetus is determined to be male;

25 **[81]** b-1) calculating Z-score for each chromosome in the

following Equations:

[82] Equation 17) $Zscore(19) = \{FGR(19) - (mean\ FGR(19))\} / (stdev\ FGR(19))$

[83] Equation 18) $Zscore(20) = \{FGR(20) - (mean\ FGR(20))\} / (stdev\ FGR(20))$

[84] Equation 19) $Zscore(21) = \{FGR(21) - (mean\ FGR(21))\} / (stdev\ FGR(21))$

[85] Equation 20) $Zscore(22) = \{FGR(22) - (mean\ FGR(22))\} / (stdev\ FGR(22))$

[86] b-2) calculating Mscore using the calculated Z-score by the following Equation; and

[87] Equation 21) $Mscore = \{Zscore(19) + Zscore(20) + Zscore(21) + Zscore(22)\} / 4$

[88]

[89] b-3) determining that the sex chromosome is XXY when the Mscore is more than or equal to 0.7, that the sex chromosome is XY when the Mscore is more than -3 and less than 0.7.

[90]

[91]

[92] The present invention also relates to a computer system including a computer readable medium having a plurality of instructions encrypted for controlling a computing system so as to perform a method of differentiating fetal sex and sex chromosome abnormality

using two or more NGS platforms from gDNA extracted from a maternal biological sample, wherein

[93] the biological sample contains cell free nucleic acid molecules of a fetus, and

5 **[94]** the differentiating includes: aligning sequences (reads) that are sequenced for each NGS platform to a reference genome database;

[95] extracting unique reads unique to the chromosome for each NGS platform and calculating a percent of unique reads
10 (UR%) relative to the total reads;

[96] calculating data for each NGS platform at the same time, setting a cut-off value, and organizing the UR(y)% values;

[97] classifying male and female using the organized
15 UR(y)%;

[98] calculating partial reference chromosomes (relative FGRs) using three or more autosomes that are similar in size to chromosomes X and Y; and

[99] calculating Z-score using the relative FGR values,
20 and calculating G score (Gscore) and M score (Mscore) based on the Z-score to differentiate sex chromosome abnormality.

[100]

[101] [Example]

[102] Hereinafter, the present invention will be described
25 in detail with reference to the following Examples.

However, the following examples are only for exemplifying the present invention, and it will be obvious to those skilled in the art that the scope of the present invention is not construed to be limited to these examples.

5 **[103]**

[104] <Example 1> Performing next-generation sequencing using maternal DNA

[105] The blood collected from the mother was separated using a centrifuge to extract the plasma, and then the
10 plasma was dissolved to prepare a library by extracting 30 ng or more of cfDNA. In the case of the Illumina platform, after combining adapters, pooling was performed using bead size selection, followed by sequencing. In the case of the Life technologies platform, pooling was performed using E-
15 gel size selection after combining adapters, followed by sequencing.

[106]

[107] <Example 2> Alignment of sequences obtained after sequencing to reference genome (human GRCh38)

20 **[108]** By using the latest version of GRCh38 (distributed on December 24, 2013) including only about 250 gaps, alignment of the sequences was performed to remove miscellaneous sequences. The minimum unique reads were extracted using the samtools view option as well as the general picard that
25 removes PCR duplication. In order to analyze data

generated from two or more platforms, distribution of initial file size and read length was analyzed (Table 1). Since only the unique reads in all regions could not be collected by general PCR duplication removal operation (FIG. 10), after other filter options were used, about 85% of unique reads were collected for HiSeq and about 76% of unique reads were collected for Proton (Table 2).

[109]

[110] [Table 1]

File Size of 10M read(unit: Giga byte)

SampleID	77	91	64	73	92	72	76	62	68	67	83	74	86	85	71	66	87	63	65	70	69	84
Illumina	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
Proton	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.8	2.8	2.8	2.8	2.8	2.7	2.7	2.6	2.6

[111]

[112] Table 2]

Mappability(%) of 10M read(unit: %)

	SampleID	62	63	64	65	66	67	68	69	70	71	72	73	74	76	77	83	84	85	86	87	91	92	Avg
HiSeq	HRemDep	96	96	95	96	96	96	95	96	96	96	96	96	96	94	95	96	96	96	95	96	95	96	96
	HUnique	86	86	86	86	85	85	86	85	85	86	85	85	83	84	85	85	86	86	86	86	85	86	85
Proton	PRemDep	81	84	84	77	82	83	82	83	84	83	82	80	83	84	83	80	81	82	84	83	84	82	82
	PUnique	76	79	79	71	76	78	75	74	76	78	77	75	76	78	77	75	72	75	78	77	79	77	76

[113]

[114] <Example 3> Data maintenance through normalization of multisamples and multiplatform data

[115] Each chromosome was divided into 100kb bin regions. Then, for the GC content, regions other than 0.30 <= GC content <= 0.50 were removed, and for the mappability, only

regions having 20% or more of mappability were selected to remove the PCR duplication and noise region. Only mean values of Z-score were used to observe the overall change. In order to establish the criteria of the bias (FIGS. 3 and 4) of experimental results obtained from difference between Illumina and Proton platforms and difference in each experiment set and location or environment, a mean value of the three smallest values of UR(y)% was obtained and the mean value was subtracted from the original value, thereby confirming that degree of variation of the value was significantly reduced by the certain cut-off value (FIGS. 5 and 6).

[116]

[117] <Example 4> Gender classification using maintained data values and GMScore calculation

[118] The mean value of the same three smallest values from results performed under different experimental conditions using different platforms was set as the reference line. Then, the cut-off value of the percent of unique reads UR(y)% of chromosome Y was set in the range of 0.02 to 0.04 to classify male and female. When the sex is a female (XX), XO and XXX were differentiated by using cases that were out of normal distribution. When the sex is a male (XY), data that were out of the normal XY distribution was judged as XXY.

[119] When regarding the $UR(x)\%$ which was the percent of unique reads of chrX, the partial reference chromosome, i.e., relative FGR, was calculated using chr7 having a similar size, and the respective Z-scores were calculated (FIG. 7), and regarding the $UR(y)\%$ which was the percent of unique reads of chrY, the partial reference chromosome, i.e., relative FGR, was also calculated using chr20 having a similar size, and the respective Z-scores were calculated (FIG. 7), the distribution degree of values was highly biased. Thus, in order to obtain the mean value of chromosomes having similar sizes, it was confirmed that the mean value of chr7, chr8, chr9, and chr10, which were similar in size to chrX, was obtained to calculate the partial reference chromosome (relative FGR) and Z-score, thereby obtaining Mscore which was the value closest to the standard value of chrX (FIG. 9). As in chrY, according to the same principle as chrX, it was confirmed that the mean value of chr19, chr20, chr21, and chr22 was obtained to calculate the relative FGR and Z-score, thereby obtaining Mscore, thereby detecting the degree of change of chrY the most accurately (FIG. 8).

[120]

[121] <Example 5> Final diagnosis result by GMscore

[122] By using the methods of Examples 1 to 4, it was confirmed that 94 normal persons, 1 person with XXX

genotype, and 1 person with XXY genotype (Table 3) could be clearly differentiated and all the cut-off values could be satisfied. Then, diagnosis was confirmed by amniocentesis. Even though a small sample was used, it was more difficult to obtain the sample as compared to abnormality of homologous chromosome and there was provided the score that could be clearly differentiated from the normal ratio, and thus it could be confirmed that the method was able to sufficiently judge the chromosome abnormality by only GMscore (FIGS. 13 and 14). XX could be differentiated from XXX by Gscore, and it could be predicted that XO was also able to be judged by using the cut-off value of -3 or less. Further, it was confirmed that the distribution of XXY could be sufficiently differentiated from XY by Mscore.

[123]

[124] [Table 3]

Number of Samples

	Female	Male
Normal	49	45
XXX	1	
XXY		1
Total	50	46

[125]

[126] <Example 6> Diagnosis of sex in additional sample groups

[127] By using the method disclosed in Example 5, it was confirmed that the sample groups (data of 3 persons with XO genotype, 1 person with XYY genotype, and 121 normal persons) shown in Table 4 were analyzed to differentiate sex and chromosome abnormality with a probability of 100% (FIG. 15).

[128]

[129] [Table 4]

	Female	Male
Normal	101	114
XXX	1	
XO	3	
XXY		1
XYY		1
Total	105	116

[130]

10 **[131]** The present invention has been described in detail based on particular features thereof, and it is obvious to those skilled in the art that these specific technologies are merely preferable embodiments and thus the scope of the present invention is not limited to the embodiments.

15 Therefore, the substantial scope of the present invention is defined by the accompanying claims and equivalent thereof.

[132]

【Industrial Applicability】

[133] The method of differentiating fetal sex and sex chromosome abnormality according to the present invention
5 may not only increase the accuracy of gender classification using next-generation sequencing (NGS) but also increase the detection accuracy of XO, XXX, XXY, etc., which are difficult to be detected, and thus commercial use of sex chromosome together with the detection of autosome
10 abnormality non-invasively may be increased. Therefore, the method of the present invention is useful for prenatal diagnosis capable of determining occurrence of malformation due to abnormality of fetal sex chromosomes in a fetus at an early stage.

15

【CLAIMS】

1. A method of differentiating fetal sex and fetal sex chromosome abnormality including: (a) performing next-generation sequencing (NGS) on two or more platforms by
5 extracting DNA from maternal blood;

(b) aligning sequences (reads) that are sequenced for each platform to a reference genome database;

(c) extracting unique reads that are uniquely aligned on chromosome for each platform and calculating a percent
10 of unique reads (UR%) relative to the total reads;

(d) calculating data for each platform at the same time, setting a cut-off value, and organizing the UR(y)% values by normalization by correcting UR(y)% according to the following equation, the UR(y)% is calculated for each
15 platform:

$$\text{UR}(y)\% \text{ new} = (\text{UR}(y)\% \text{ original} - (\text{mean value of the three smallest UR}(y)\% \text{ values});$$

(e) classifying male and female using the organized UR(y)%;

20 (f) calculating partial reference chromosome (relative FGR) values using three or more autosomes that are similar in size to chromosomes X and Y; and

(g) calculating Z-score using the partial reference chromosome (relative FGR) values, and calculating G score
25 (Gscore) and M score (Mscore) based on the Z-score to

differentiate sex chromosome abnormality.

2. The method of claim 1, wherein the step (a) comprising the steps of:

- 5 (a) separating a plasma by centrifuging the collected biological sample;
- (b) extracting cfDNA from the separated plasma;
- (c) preparing a library using the extracted cfDNA; and
- (d) pooling the prepared library, and then decoding
- 10 the base sequence using NGS equipment.

3. The method of claim 1, wherein the step (c) comprising the steps of:

- (a) removing PCR duplication and ambiguous reads in
- 15 sequencing, aligning only perfectly matched reads, and extracting a read in which a GC content and mappability are more than or equal to the cut-off value in the aligned sequence; and

- (b) calculating a percentage (UR(x)%) of the read that
- 20 uniquely matches the chromosome X and a percentage (UR(y)%) of the read that uniquely matches the chromosome Y according to the following Equations, based on the extracted read:

Equation 1) $UR(x)\% = \frac{\text{the number of chromosome X (chrX) reads}}{\text{total number of autosome reads}} \times 100$, and

25

Equation 2) $UR(y)\% = \frac{\text{the number of chromosome Y (chrY) reads}}{\text{total number of autosome reads}} \times 100$.

4. The method of claim 3, wherein the step (a) is performed
5 by dividing each region into 100 kb bin regions, wherein
the cut-off value of the GC content is 0.30 to 0.50, and
the cut-off value of the mappability is 20% or more.

10 5. The method of claim 1, wherein the cut-off value of step
(d) is 0.02 ~ 0.06.

6. The method of claim 1, wherein the gender classification
of (e) is characterized in that when the organized $UR(y)\%$
15 value is more than or equal to the cut-off value, it is
determined as a male, and when the organized $UR(y)\%$ value
is smaller than the cut-off value, it is determined as a
female.

20 7. The method of claim 1, wherein the step (f) comprising
the steps of:

(a) selecting autosomes (chr7, chr8, chr9, and chr10)
that are similar in size to sex chromosome X (chrX) as
internal reference chromosomes and calculating each FGR by
25 the following Equations:

Equation 4) $FGR(7) = (UR(x)\% + UR(y)\%) / UR(7)\%$

Equation 5) $FGR(8) = (UR(x)\% + UR(y)\%) / UR(8)\%$

Equation 6) $FGR(9) = (UR(x)\% + UR(y)\%) / UR(9)\%$

Equation 7) $FGR(10) = (UR(x)\% + UR(y)\%) / UR(10)\%$, and

5 (b) selecting autosomes (chr19, chr20, chr21, and chr22) that are similar in size to sex chromosome Y (chrY) as partial reference chromosomes and calculating each FGR by the following Equations:

Equation 8) $FGR(19) = (UR(x)\% + UR(y)\%) / UR(19)\%$

10 Equation 9) $FGR(20) = (UR(x)\% + UR(y)\%) / UR(20)\%$

Equation 10) $FGR(21) = (UR(x)\% + UR(y)\%) / UR(21)\%$

Equation 11) $FGR(22) = (UR(x)\% + UR(y)\%) / UR(22)\%$.

8. The method of claim 1, wherein the step (g) comprising
15 the steps of:

(a) differentiating sex chromosome abnormality by the following steps when the sex of the fetus is determined to be female;

a-1) calculating Z-score for each chromosome in the
20 following Equations:

Equation 12) $Zscore(7) = \{FGR(7) - (mean\ FGR(7))\} / (stdev\ FGR(7))$

Equation 13) $Zscore(8) = \{FGR(8) - (mean\ FGR(8))\} / (stdev\ FGR(8))$

25 Equation 14) $Zscore(9) = \{FGR(9) - (mean\ FGR(9))\} /$

(stdev FGR(9))

Equation 15) $Zscore(10) = \{FGR(10) - (mean\ FGR(10))\} /$
 (stdev FGR(10))

a-2) calculating Gscore using the calculated Z-score
 5 by the following Equation; and

Equation 16) $Gscore = \{Zscore(7) + Zscore(8) +$
 $Zscore(9) + Zscore(10)\} / 4$

a-3) determining that the sex chromosome is XXX when
 the Gscore is more than or equal to 1.5, that the sex
 10 chromosome is XX when the Gscore is more than -3 and less
 than 1.5, and that the sex chromosome is XO when the Gscore
 is less than or equal to -3;

(b) differentiating sex chromosome abnormality by the
 following steps when the sex of the fetus is determined to
 15 be male;

b-1) calculating Z-score for each chromosome in the
 following Equations:

Equation 17) $Zscore(19) = \{FGR(19) - (mean\ FGR(19))\} /$
 (stdev FGR(19))

20 Equation 18) $Zscore(20) = \{FGR(20) - (mean\ FGR(20))\} /$
 (stdev FGR(20))

Equation 19) $Zscore(21) = \{FGR(21) - (mean\ FGR(21))\} /$
 (stdev FGR(21))

Equation 20) $Zscore(22) = \{FGR(22) - (mean\ FGR(22))\} /$
 25 (stdev FGR(22))

b-2) calculating Mscore using the calculated Z-score by the following Equation; and

Equation 21)
$$\text{Mscore} = \{ \text{Zscore}(19) + \text{Zscore}(20) + \text{Zscore}(21) + \text{Zscore}(22) \} / 4$$

5 b-3) determining that the sex chromosome is XXY when the Mscore is more than or equal to 0.7, that the sex chromosome is XY when the Mscore is more than -3 and less than 0.7.

10 9. The method of claim 1, wherein the biological sample is selected from the group consisting of blood, plasma, serum, urine, and saliva.

15 10. A computer system including a computer readable medium having a plurality of instructions encrypted for controlling a computing system so as to perform a method of differentiating fetal sex and sex chromosome abnormality using two or more NGS platforms from gDNA extracted from a maternal biological sample, wherein

20 the biological sample contains cell free nucleic acid molecules of a fetus, and

the method of differentiating fetal sex and sex chromosome abnormality includes:

aligning sequences (reads) that are sequenced for each
25 NGS platform to a reference genome database;

extracting unique reads unique to the chromosome for each NGS platform and calculating a percent of unique reads (UR%) relative to the total reads;

calculating data for each NGS platform at the same time, setting
5 a cut-off value, and organizing the UR(y)% values by normalization by correcting UR(y)% according to the following equation, the UR(y)% is calculated for each platform:

$$\text{UR(y)\% new} = (\text{UR(y)\% original} - (\text{mean value of the three smallest UR(y)\% values}));$$

10 classifying male and female using the organized UR(y)%;

calculating partial reference chromosomes (relative FGRs) using three or more autosomes that are similar in size to chromosomes X and Y; and

15 calculating Z-score using the relative FGR values, and calculating G score (Gscore) and M score (Mscore) based on the Z-score to differentiate sex chromosome abnormality.