

WHAT IS CLAIMED IS:

1. A method for detecting chromosome aneuploidy, which is characterized in that the method includes the following steps of:

high-throughput sequencing of peripheral blood cell-free DNA from a pregnant woman to be tested to produce sequencing data comprising all of the chromosomes;

calculating coverage statistics for all of the chromosomes with the sequencing data by segmenting the chromosomes into windows so as to produce a pre-correction coverage for each chromosome;

performing a Z-test on the number of unique sequences in each window of the pregnant woman to produce a Z_{CNV} value and then locating chromosomal fragments with copy number variations of the pregnant woman on the basis of the magnitude of the Z_{CNV} value; wherein the chromosomal fragments with copy number variations of the pregnant woman are 300 Kb or more in length and have Z_{CNV} values of the chromosome fragments greater than or equal to 4 or less than or equal to -4 in 80% or more of the total windows within the fragments which are 300 Kb or more in length,

correcting the pre-correction coverage of each chromosome by utilizing the impact of the fragments with copy number variations of the pregnant woman on the pre-correction coverage of each chromosome to produce the corrected coverage for each chromosome; and

performing a Z-test for each chromosome by using the corrected coverage of each chromosome to obtain a Z_{aneu} value, and determining whether the chromosome has an aneuploidy based on whether the absolute value of Z_{aneu} is greater than or equal to 3; wherein when the absolute value of Z_{aneu} is greater than or equal to 3, then it is determined that the chromosome has an aneuploidy;

wherein the impact of the fragments with copy number variations of the pregnant woman to be tested on the pre-correction coverage of each chromosome is represented by a parameter α ,

when the fetus inherits the fragments with copy number variations from the mother, the parameter α is calculated as formula (1):

$$\alpha = \frac{(m-n) \cdot 2 + n \cdot cn}{m \cdot 2} \dots\dots\dots (1)$$

when the fetus does not inherit the fragments with copy number variations from the mother, the parameter α is calculated as formula (2):

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots\dots\dots (2)$$

in formula (1) and formula (2), m represents the effective length of the chromosome in which the fragments with copy number variations occur, in the unit of Mb; and n represents the length of the fragments with copy number variations of the pregnant woman to be tested, in the unit of Mb; cn represents the copy number of the fragments with copy number variations found in the pregnant woman to be tested;

in formula (2), f represents the concentration of the cell-free fetal DNA existing in the peripheral blood cell-free DNA of the pregnant woman to be tested, and the concentration f of the cell-free fetal DNA is assumed to be less than 50%;

correcting the pre-correction coverage of each chromosome by using $x' = \frac{\hat{x}}{\alpha}$,

wherein $\hat{x}$ represents the pre-correction coverage of each chromosome and x' represents the corrected coverage of each chromosome.

2. The method according to claim 1, wherein the coverage statistics is calculated by segmenting all of the chromosomes in the sequencing data into windows with equal sizes so as to produce the pre-correction coverage of each chromosome.

3. The method according to claim 2, wherein the length of each window is 100 Kb and the overlapping ratio between two adjacent windows is 50%.

4. The method according to claim 1, wherein the step of performing a Z-test on the number of unique sequences in each window of the pregnant woman to be tested to produce the Z_{CNV} value and

then locating chromosomal fragments with the copy number variations of the pregnant woman to be tested on the basis of the magnitude of the Z_{CNV} value further includes the steps of:

counting the number of the unique sequences in each window according to the sequencing depth of each sequence in the sequencing data;

calculating the number of the unique sequences in each window according to the GC content and the mapping rate of each chromosome to obtain the pre-correction coverage of the number of the unique sequences in each window; and

normalizing the pre-correction coverage of the number of the unique sequences in each window to obtain the Z_{CNV} value of the number of the unique sequences in each window and determining whether the pregnant woman to be tested has the chromosomal fragments with copy number variations on the basis of the magnitude of the Z_{CNV} value;

if there are fragments which are 300 Kb or more in length in the sequencing data, and within the fragments which are 300 Kb or more in length, the Z_{CNV} values of the numbers of the unique sequences in 80% or more of the total windows are greater than or equal to 4 or less than or equal to -4, then the fragments which are 300 Kb or more in length are determined to be the fragments with copy number variations of the pregnant woman to be tested.

5. The method according to claim 1, for the step of performing a Z-test for each chromosome by using the corrected coverage of each chromosome to obtain the Z_{aneu} value, the Z_{aneu} value is calculated as:

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

wherein $\bar{x}$ represents the pre-correction coverage obtained by a known negative sample population according to a LOESS algorithm; s represents the standard error of $(x' - \bar{x})$ in the negative sample population.

6. An apparatus for detecting chromosome aneuploidy, which is characterized in that the apparatus includes the following modules:

a sequencing data detection module: for high-throughput sequencing of peripheral blood cell-free DNA from a pregnant woman to be tested to produce the sequencing data comprising all of the chromosomes;

a first coverage calculation module: for calculating a coverage statistics of all of the chromosomes with the sequencing data by segmenting the chromosomes into windows so as to produce a pre-correction coverage for each chromosome;

a Z_{CNV} value calculation module: for calculating the Z_{CNV} value on the number of unique sequences in each window of the pregnant woman;

fragments with copy number variations search module: for searching the fragments that are 300 Kb or more in length in the sequencing data and which have the Z_{CNV} values of the chromosome fragments greater than or equal to 4 or less than or equal to -4 in 80% or more of the total windows;

fragments with copy number variations determination module: for determining fragments in the sequencing data that are 300 Kb or more in length and which have Z_{CNV} values of the chromosome fragments greater than or equal to 4 or less than or equal to -4 in 80% or more of the total windows as the fragments with copy number variations of the pregnant woman;

a first α calculation module: for calculating a parameter α according to the formula (1) in the case where the fetus inherits the fragments with copy number variations from the mother, wherein the parameter α represents the impact of the fragments with copy number variations of the pregnant woman on the pre-correction coverage of each chromosome;

$$\alpha = \frac{(m-n) \cdot 2 + n \cdot cn}{m \cdot 2} \dots\dots\dots (1)$$

in formula (1), m represents the effective length of the chromosome in which the fragments with copy number variations occurs, in the unit of Mb; and n represents the length of the fragments with copy number variations of the pregnant woman, in the unit of Mb; cn represents the copy number of the fragments with copy number variations found in the pregnant woman;

a second α calculation module: for calculating a parameter α according to formula (2) in the case where the fetus does not inherit the fragments with copy number variations from the mother, wherein the parameter α is calculated according to formula (2)

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots\dots\dots (2)$$

in formula (2), m represents the effective length of the chromosome in which the fragments with copy number variations occurs, in the unit of Mb; and n represents the length of the fragments with copy number variations of the pregnant woman, in the unit of Mb; cn represents the copy number of the fragments with copy number variations found in the pregnant woman; f represents the concentration of the cell-free fetal DNA existing in the peripheral blood cell-free DNA of the pregnant woman, and the concentration f of the cell-free fetal DNA is assumed to be less than 50%;

a correction module: for correcting the pre-correction coverage of each chromosome by

using: $x' = \frac{\hat{x}}{\alpha}$ to produce the corrected coverage of each chromosome; wherein $\hat{x}$ represents the pre-correction coverage of each chromosome and x' represents the corrected coverage of each chromosome;

a second coverage calculation module: for calculating the Z_{aneu} value of each chromosome by using the corrected coverage of each chromosome;

Z_{aneu} value determination module: for determining whether the absolute Z_{aneu} value is greater than or equal to 3;

a chromosome aneuploidy confirmation module: for confirming the chromosome has aneuploidy in the case where the absolute Z_{aneu} value is greater than or equal to 3.

7. The apparatus according to claim 6, wherein the first coverage calculation module includes:

a chromosome window segmentation sub-module: for segmenting all of the chromosomes in the sequencing data into windows with equal size;

a first coverage calculation sub-module: for calculating the coverage statistics in the form of windows with equal size to produce the pre-correction coverage of each chromosome.

8. The apparatus according to claim 7, wherein the length of each window in the chromosome window segmentation sub-module is 100Kb, and the overlapping ratio between two adjacent windows is 50%.

9. The apparatus according to claim 6, wherein the Z_{CNV} value calculation module includes:

a unique sequence counting unit: for counting the number of the unique sequences in each window according to the sequencing depth of each sequence in the sequencing data;

a unique sequence coverage calculation unit: for calculating the number of the unique sequences in each window according to the GC content and the mapping rate of each chromosome to obtain the pre-correction coverage of the number of the unique sequences in each window; and

a unique sequence Z_{CNV} value calculation unit: for normalizing the pre-correction coverage of the number of the unique sequences in each window to obtain the Z_{CNV} value of the number of the unique sequences in each window.

10. The apparatus according to claim 6, wherein in the second coverage calculation module, the Z_{aneu} value is calculated as:

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

wherein $\bar{x}$ is the pre-correction coverage obtained by a known negative sample population according to a LOESS algorithm; s represents the standard error of $(x' - \bar{x})$ in the negative sample population.

11. A kit for detecting the chromosome aneuploidy, which is characterized in that the kit includes :

the detection reagents and a detection device: for high-throughput sequencing of peripheral blood cell-free DNA from a pregnant woman to be tested to produce the sequencing data containing all the chromosomes;

a first coverage calculation device: for calculating a coverage statistics of all of the chromosomes with the sequencing data by segmenting the chromosomes into windows so as to produce a pre-correction coverage for each chromosome;

a Z_{CNV} value calculation device: for performing a Z-test on the number of unique sequences in each window of the pregnant woman to be tested to obtain the Z_{CNV} value;

fragments with copy number variations search device: for searching the fragments in the sequencing data that are 300 Kb or more in length and which have the Z_{CNV} values of the chromosome fragments greater than or equal to 4 or less than or equal to -4 in 80% or more of the total windows;

fragments with copy number variations determination device: for obtaining the fragments with copy number variations of the pregnant woman to be tested on the basis of the magnitude of the Z_{CNV} value;

a first α calculation device: for calculating a parameter α according to the formula (1) in the case where the fetus inherits the fragments with copy number variations from the mother, wherein the parameter α represents the impact of the fragments with copy number variations of the pregnant woman to be tested on the pre-correction coverage of each chromosome;

$$\alpha = \frac{(m-n) \cdot 2 + n \cdot cn}{m \cdot 2} \dots\dots\dots (1)$$

m represents the effective length of the chromosome in which the fragments with copy number variations occurs, in the unit of Mb; and n represents the length of the fragments with copy number variations of the pregnant woman to be tested, in the unit of Mb; cn represents the copy number of the fragments with copy number variations found in the pregnant woman;

a second α calculation device: for calculating a parameter α according to formula (2) in the case where the fetus does not inherit the fragments with copy number variations from the mother, wherein the parameter α is calculated according to formula (2)

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots\dots\dots (2)$$

m represents the effective length of the chromosome in which the fragments with copy number variations occurs, in the unit of Mb; and n represents the length of the fragments with copy number variations of the pregnant woman, in the unit of Mb; cn represents the copy number of the fragments with copy number variations found in the pregnant woman; f represents the concentration of the cell-free fetal DNA existing in the peripheral blood cell-free DNA of the pregnant woman, and the concentration f of the cell-free fetal DNA is assumed to be less than 50%;

a correction device: for correcting the pre-correction coverage of each chromosome by using:

$$x' = \frac{\hat{x}}{\alpha}$$

to produce the corrected coverage of each chromosome; wherein $\hat{x}$ represents the pre-correction coverage of each chromosome and x' represents the corrected coverage of each chromosome;

a second coverage calculation device: for calculating the Z_{aneu} value of each chromosome by using the corrected coverage of each chromosome;

Z_{aneu} value determination device: for determining whether the absolute Z_{aneu} value is greater than or equal to 3;

a chromosome aneuploidy confirmation device: for confirming the chromosome has aneuploidy in the case where the absolute Z_{aneu} value is greater than or equal to 3.

12. The kit according to claim 11, wherein the first coverage calculation device includes:

a chromosome window segmentation component: for segmenting all of the chromosomes in the sequencing data into windows with equal size;

a first coverage calculation component: for calculating the coverage statistics in the form of windows with equal size to produce the pre-correction coverage of each chromosome.

13. The kit according to claim 12, wherein the length of each window in the chromosome window segmentation component is 100Kb, and the overlapping ratio between two adjacent windows is 50%.

14. The kit according to claim 11, wherein the Z_{CNV} value calculation device includes:

a unique sequence counting component: for counting the number of the unique sequences in each window according to the sequencing depth of each sequence in the sequencing data;

a unique sequence coverage calculation component: for calculating the number of the unique sequences in each window according to the GC content and the mapping rate of each chromosome to obtain the pre-correction coverage of the number of the unique sequences in each window; and

a unique sequence Z_{CNV} value calculation component: for normalizing the pre-correction coverage of the number of the unique sequences in each window to obtain the Z_{CNV} value of the number of the unique sequences in each window.

15. The kit according to claim 11, wherein in the second coverage calculation device, the Z_{aneu} value is calculated as:

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

wherein $\bar{x}$ is the pre-correction coverage obtained by a known negative sample population according to a LOESS algorithm; s represents the standard error of $(x' - \bar{x})$ in the negative sample population.