【Claims】

【Claim 1】 A method of detecting a numerical chromosomal abnormality, the method comprising:

(A) extracting nucleic acids from a biological sample and obtaining nucleic acid fragments to obtain sequence information therefrom;

(B) identifying positions of the nucleic acid fragments in a reference genome database based on the obtained sequence information (reads);

(C) grouping the sequence information (reads) into whole sequences, forward sequences, and reverse sequences;

(D) defining Representative Positions of the respective nucleic acid fragments using the grouped sequence information, and measuring the distance between the Representative Positions to calculate a fragment distance (FD) for each group; and

(E) calculating a fragment distance index (FDI) for the entire chromosomal region or each specific region to be analyzed based on the FD for each group calculated in step (D) and determining that there is a chromosomal abnormality when the FDI does not fall within a cutoff value range,

wherein the FD in step (D) is calculated as a distance between the Representative Position of an i^{th} nucleic acid

fragment and the Representative Position of at least one nucleic acid fragment selected from $i+1^{th}$ to n^{th} nucleic acid fragments among obtained n nucleic acid fragments,

wherein step (E) comprises:

(E-i) determining a representative FD (RepFD) for an entire chromosomal region or for each specific region;

(E-ii) calculating one or more selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of RepFD, reciprocals thereof and a combination thereof in a certain region in the sample, rather than the entire chromosomal region or specific genomic region, to derive a normalized factor;

(E-iii) calculating a representative FD ratio (RepFD ratio) based on Equation 1 below; and

Equation 1: RepFD ratio = RepFD Target genomic region / Normalized Factor

(E-iv) comparing the RepFD ratio of a normal reference group with that of the sample to calculate a fragment distance index (FDI),

wherein the representative FD (RepFD) of step (E-i) is at least one selected from the group consisting of a sum,

difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of FDs, and/or a reciprocal thereof,

wherein the nucleic acid fragment comprises at least 2 read pairs.

【Claim 2】 The method according to claim 1, wherein the step (A) comprises:

(A-i) obtaining nucleic acids from blood, semen, vaginal cells, hair, saliva, urine, oral cells, amniotic fluid containing placental cells or fetal cells, tissue cells, and a mixture thereof;

(A-ii) removing proteins, fats, and other residues from the collected nucleic acids using a salting-out method, a column chromatography method, or a bead method to obtain purified nucleic acids;

(A-iii) producing a single-end sequencing or pair-end sequencing library for the purified nucleic acids or nucleic acids randomly fragmented by an enzymatic digestion, pulverization, or hydroshear method;

(A-iv) reacting the produced library with a next-generation sequencer; and

(A-v) obtaining sequence information (reads) of the nucleic acids in the next-generation sequencer.

【Claim 3】 The method according to claim 1, wherein the Representative Position of the nucleic acid fragment is obtained by adding an arbitrary value to a median of the nucleic acid fragment or subtracting the arbitrary value therefrom,

wherein the median of the nucleic acid fragment is center position of the nucleic acid fragment.

【Claim 4】 The method according to claim 3, wherein, in paired-end sequencing, the Representative Position of the nucleic acid fragment is derived based on position values of forward and reverse reads.

【Claim 5】 The method according to claim 4, further comprising excluding nucleic acid fragments having a mapping quality score of reads below a cutoff value from calculation.

【Claim 6】 The method according to claim 1, wherein in single-end sequencing, the Representative Position of the nucleic acid fragment is derived based on one type of

position value of forward or reverse read.

【Claim 7】 The method according to claim 6, wherein an arbitrary value is added when a position value is derived based on sequence information aligned in the forward direction and the arbitrary value is subtracted when a position value is derived based on sequence information aligned in the reverse direction.

【Claim 8】 The method according to claim 3, wherein the arbitrary value is 30 to 70% of a mean length of the nucleic acid to be analyzed.

【Claim 9】 The method according to claim 3, wherein the arbitrary value is 0 to 5 kbp or 0 to 300% of a length of the nucleic acid fragment.

【Claim 10】 The method according to claim 1, wherein the representative FD (RepFD) of step (E-i) is a median, mean, or reciprocal of FDs.

【Claim 11】 The method according to claim 1, wherein the

certain region in the sample other than the entire chromosomal region or specific genomic region in step (E-ii) is selected using a method comprising:

a) randomly selecting a region other than an entire chromosomal region or a specific genomic region to be analyzed;

b) determining a representative RepFD of the genomic region selected in step a) with a pre-normalized factor (PNF);

c) calculating a representative FD ratio (RepFD ratio) based on Equation 2:

Equation 2: RepFD ratio = RepFD Target genomic region / PNF

d) calculating a coefficient of variance (SD/mean) of the RepFD ratio of a normal reference group; and

e) determining a genomic region having a smallest value among obtained coefficients of variance obtained by repeatedly performing steps a) to d) as the certain region in the sample, other than the entire chromosomal region or the specific genomic region.

【Claim 12】A computer-readable storage medium including an instruction configured to be executed by a processor for

detecting a numerical chromosomal abnormality through the following steps comprising:

(A) extracting nucleic acids from a biological sample and obtaining nucleic acid fragments to obtain sequence information;

(B) aligning the nucleic acid fragments to a reference genome database based on the obtained sequence information (reads);

(C) measuring the distance between the selected nucleic acid fragments to calculate a fragment distance (FD); and

(D) calculating a fragment distance index (FDI) over the entire chromosomal region or in each specific genomic region to be analyzed based on the FD calculated in step (C) and determining that there is a chromosomal abnormality when the FDI does not fall within a cutoff value range,

wherein the FD in step (D) is calculated as a distance between the Representative Position of an i^{th} nucleic acid fragment and the Representative Position of at least one nucleic acid fragment selected from $i+1^{th}$ to n^{th} nucleic acid fragments among obtained n nucleic acid fragments,

wherein step (E) comprises:

(E-i) determining a representative FD (RepFD) for an entire chromosomal region or for each specific region;

(E-ii) calculating one or more selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of RepFD, reciprocals thereof and a combination thereof in a certain region in the sample, rather than the entire chromosomal region or specific genomic region, to derive a normalized factor;

(E-iii) calculating a representative FD ratio (RepFD ratio) based on Equation 1 below; and

Equation 1: RepFD ratio = RepFD Target genomic region / Normalized Factor

(E-iv) comparing the RepFD ratio of a normal reference group with that of the sample to calculate a fragment distance index (FDI),

wherein the representative FD (RepFD) of step (E-i) is at least one selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of FDs, and/or a reciprocal thereof,

wherein the nucleic acid fragment comprises at least 2 read pairs.

【Claim 13】The method according to claim 3, wherein the calculated FD is a read distance (RD) when the arbitrary value is 50% of a mean length of the nucleic acid to be analyzed.

【Claim 14】A method for detecting a numerical chromosomal abnormality, the method comprising:

(A) extracting nucleic acids from a biological sample to obtain sequence information;

(B) aligning the obtained sequence information (reads) to a reference genome database;

(C) measuring a distance between the aligned reads in the aligned sequence information (reads) to calculate a read distance (RD); and

(D) calculating a read distance index (RDI) over an entire chromosomal region or in each specific genomic region to be analyzed based on the RD calculated in step (C), and determining that there is a chromosomal abnormality when the RDI does not fall within a cutoff value range,

wherein the RD is obtained by calculating the distance between the 5' or 3' end inside the i^{th} read and the 5' or 3' end inside at least one of the $i+1^{th}$ to n^{th} reads,

wherein the step (D) comprises:

(D-i) determining a representative RD (RepRD) for each entire chromosomal region or specific genomic region;

(D-ii) calculating one or more selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of RepRD in a certain region in the sample other than the entire chromosomal region or specific genomic region, and/or a reciprocal thereof, to derive a normalized factor;

(D-iii) calculating a representative RD ratio (RepRD ratio) based on Equation 10 below;

Equation 10: RepRD ratio = RepRD Target genomic region / Normalized Factor; and

(D-iv) comparing the RepRD ratio of a normal reference group with that of the sample to calculate a read distance index (RDI),

wherein the representative RD (RepRD) of step (D-i) is one or more selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of RD,

and/or reciprocals thereof.

【Claim 15】The method according to claim 14, wherein step (A) is performed using a method comprising:

(A-i) obtaining nucleic acids from blood, semen, vaginal cells, hair, saliva, urine, oral cells, amniotic fluid containing placental cells or fetal cells, tissue cells, and a mixture thereof;

(A-ii) removing proteins, fats, and other residues from the collected nucleic acids using a salting-out method, a column chromatography method, or a bead method to obtain purified nucleic acids;

(A-iii) producing a single-end sequencing or pair-end sequencing library for the purified nucleic acids or nucleic acids randomly fragmented by an enzymatic digestion, pulverization, or hydroshear method;

(A-iv) reacting the produced library with a next-generation sequencer; and

(A-v) obtaining sequence information (reads) of the nucleic acids in the next-generation sequencer.

【Claim 16】The method according to claim 14, further comprising, prior to the step (C), grouping the aligned

reads according to an alignment direction.

【Claim 17】The method according to claim 14, wherein the RD in step (C) is calculated by a distance from a value of one of both ends of an i^{th} read and one or more reads selected from $i+1^{th}$ to n^{th} reads, among obtained n reads, to the value obtained by adding 50% of an average length of the nucleic acid thereto or subtracting 50% of the average length of the nucleic acid therefrom.

【Claim 18】 The method according to claim 14, wherein the representative RD (RepRD) of step (D-i) is a median, mean, or reciprocal of RDs.

【Claim 19】 The method according to claim 14, wherein the certain region in the sample other than the entire chromosome region or specific genomic region to be analyzed in step (D-ii) is selected using a method comprising:

a) randomly selecting a region other than the entire chromosome or specific genomic region to be analyzed;

b) determining a representative RepRD of the genomic region selected in step a) with a pre-normalized factor (PNF);

c) calculating a representative RD ratio (RepRD ratio) based on Equation 11:

Equation 11: RepRD ratio = RepRD Target genomic region / PNF;

d) calculating a coefficient of variance (SD/mean) of the RepRD ratio of a normal reference group; and

e) determining a genomic region having a smallest value among coefficients of variance obtained by repeatedly performing steps a) to d) as the certain region in the sample other than the entire chromosome region or the specific genomic region.

【Claim 20】A computer-readable storage medium including instructions configured to be executed by a processor for detecting a numerical chromosomal abnormality through the following steps comprising:

(A) extracting nucleic acids from a biological sample to obtain sequence information;

(B) aligning the obtained sequence information (reads) to a reference genome database;

(C) measuring the distance between the aligned reads among the selected sequence information (reads) to calculate a read distance (RD); and

(D) calculating a read distance index (RDI) over the entire chromosomal region or in each specific genomic region to be analyzed based on the RD calculated in step (C) and determining that there is a chromosomal abnormality when the RDI does not fall within a cutoff value range,

wherein the RD is obtained by calculating the distance between the 5' or 3' end inside the i^{th} read and the 5' or 3' end inside at least one of the $i+1^{th}$ to n^{th} reads,

wherein the step (D) comprises:

(D-i) determining a representative RD (RepRD) for each entire chromosomal region or specific genomic region;

(D-ii) calculating one or more selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of RepRD in a certain region in the sample other than the entire chromosomal region or specific genomic region, and/or a reciprocal thereof, to derive a normalized factor;

(D-iii) calculating a representative RD ratio (RepRD ratio) based on Equation 10 below;

Equation 10: RepRD ratio = RepRD Target genomic region / Normalized Factor; and

(D-iv) comparing the RepRD ratio of a normal reference group with that of the sample to calculate a read distance index (RDI) and,

wherein the representative RD (RepRD) of step (D-i) is one or more selected from the group consisting of a sum, difference, product, mean, log of product, log of sum, median, quantile, minimum, maximum, variance, standard deviation, median absolute deviation and coefficient of variance of RD, and/or reciprocals thereof.