distance value X% and a Y-direction distance value Y% by evaluating the following formula:

$$X\% = ||0.5 - VAF| - x|/x$$

$$Y\% = |CN_t \times C_t - y|/|2 - y|$$

[0105] In step S144-5, the coefficient calculation unit 140 calculates an individual score using the X-direction distance value and the Y-direction distance value.

[0106] Specifically, the coefficient calculation unit 140 calculates an individual score $Score_n$ by evaluating the following formula. Note that m^2 signifies a square of m.

$$Score_n = X\%^2 + Y\%^2$$

[0107] In step S144-6, the coefficient calculation unit 140 determines whether an unselected target chromosome exists.

If an unselected target chromosome exists, the process proceeds to step S144-1.

If an unselected target chromosome does not exist, the process proceeds to step S144-7.

[0108] In step S144-7, the coefficient calculation unit 140 calculates the sum of the individual scores. The sum of the individual scores is the distance score.

[0109] Specifically, the coefficient calculation unit 140 calculates the distance score Score by evaluating the following formula. Note that $Score_n$ expresses an individual score of chromosome n.

$$Score = Score_1 + Score_{10} + Score_{19}$$

[0110] Back to Fig. 15, description continues from step S145-1.

In step S145-1, the coefficient calculation unit 140 compares the distance score with the minimum score. The initial value of the minimum score is the maximum value of a variable for a minimum score.

S149.

If adjustment of the reference coefficient is not to end, the process proceeds to step S148-1.

[0120] In step S149, the coefficient calculation unit 140 determines whether an unselected target gene exists.

If an unselected target gene exists, the process proceeds to step S147.

If an unselected target gene does not exist, the coefficient calculation process (S140) ends.

[0121] Back to Fig. 2, step S150 will be described.

In step S150, the copy-number calculation unit 150 calculates the copy number of each target gene in the cancer cell using the copy number of each target gene in a tumor sample and the correction coefficient.

[0122] A procedure of the copy-number calculation process (S150) will be described referring to Fig. 18.

In step S151, the copy-number calculation unit 150 selects one unselected target gene.

[0123] In step S152, the copy-number calculation unit 150 multiplies the tentative copy number of the target gene by the correction coefficient. The tentative copy number of the target gene is calculated in step S141-2 (see Fig. 15).

The copy number obtained by multiplying the tentative copy number of the target gene by the correction coefficient is the copy number of the target gene in the cancer cell, that is, the accurate copy number of the target gene.

[0124] Specifically, the copy-number calculation unit 150 calculates the copy number(CN) by evaluating the following formula. Note that C_{best} expresses a correction coefficient and that CN_t expresses a tentative copy number.

CLAIMS

[Claim 1] A copy-number measurement device comprising:

a position identification unit to map a plurality of tumor sample reads which are a plurality of reads obtained from a tumor sample involving a cancer cell, to a human genome sequence, and identify, for each target gene, a target position which is a genome position of a base, the genome position having changed with respect to the human genome sequence;

a frequency calculation unit to calculate a variant allele frequency for each target position of each target gene;

a distance calculation unit to calculate, for each target gene, a feature distance equivalent to a difference between a variant allele frequency corresponding to a peak density and a reference variant allele frequency in a density distribution indicating a density of a number of mapping reads with respect to the variant allele frequency, the number of mapping reads being a number of tumor sample reads mapped to respective target positions in the target gene;

a coefficient calculation unit to calculate a correction coefficient being used for correcting a copy number of each target gene in the tumor sample, using the feature distance of each target gene; and

a copy-number calculation unit to calculate a copy number of each target gene in the cancer cell using the copy number of each target gene in the tumor sample and the correction coefficient.

[Claim 2] The copy-number measurement device according to claim 1,

wherein the distance calculation unit generates a scatter graph indicating a relation between a variant allele frequency of each target position and the mapping read number of each target position; converts the scatter graph to a density distribution

graph; generates a correlation graph indicating a correlation between a lower area and a upper area, the lower area being, of the density distribution graph, a region expressing a variant allele frequency that is equal to or lower than the reference variant allele frequency, the upper area being, of the density distribution graph, a region expressing a variant allele frequency that is equal to or higher than the reference variant allele frequency; and calculates, as the feature distance, an absolute value of a difference between a variant allele frequency corresponding to a peak correlation value and the reference variant allele frequency, in the correlation graph.

[Claim 3] The copy-number measurement device according to claim 2, wherein the correlation graph indicates a correlation in density between a variant allele frequency in the lower area and a variant allele frequency in the upper area that are equal to each other regarding absolute values of differences thereof from the reference variant allele frequency.

[Claim 4] The copy-number measurement device according to any one of claims 1 to 3,

wherein the coefficient calculation unit calculates a value corresponding to a deviation amount between a relation graph and a measurement point, as the correction coefficient, the relation graph indicating a relation between the feature distance and a logarithmic value of a proportion of a copy number of a gene in a cancer cell to a copy number of a gene in a normal cell, the measurement point indicating a feature distance of a target gene, and a logarithmic value of a proportion of a copy number of the target gene in the tumor sample to a copy number of the target gene in a normal sample.

[Claim 5] The copy-number measurement device according to any one of claims 1 to 4, comprising

a content ratio calculation unit to calculate a content ratio of the cancer cell in

the tumor sample based on a copy number of each target gene in the cancer cell.

[Claim 6] The copy-number measurement device according to claim 5,

wherein the content ratio calculation unit calculates a content ratio candidate using a copy number in the cancer cell for each target gene, and determines the content ratio of the cancer cell in the tumor sample based on the content ratio candidate of each target gene.

[Claim 7] The copy-number measurement device according to any one of claims 1 to 6,

wherein the tumor sample is a sample of a brain tumor, and

wherein the target gene is at least one of ATRX, IDH1, IDH2, TP53, TERT, BRAF, PDGFRA, MET, EGFR, BRSK1, EHD2, AKT2, TP73, NMNAT1, TGFBR3, and PTEN.

[Claim 8] A copy-number measurement program to cause a computer to function as:

a position identification unit to map a plurality of tumor sample reads which are a plurality of reads obtained from a tumor sample involving a cancer cell, to a human genome sequence, and identify, for each target gene, a target position, which is a genome position of a base, the genome position having changed with respect to the human genome sequence;

a frequency calculation unit to calculate a variant allele frequency for each target position of each target gene;

a distance calculation unit to calculate, for each target gene, a feature distance equivalent to a difference between a variant allele frequency corresponding to a peak density and a reference variant allele frequency in a density distribution indicating a density of a number of mapping reads with respect to the variant allele frequency, the

number of mapping reads being a number of tumor sample reads mapped to respective target positions in the target gene;

a coefficient calculation unit to calculate a correction coefficient being used for correcting a copy number of each target gene in the tumor sample, using the feature distance of each target gene; and

a copy-number calculation unit to calculate a copy number of each target gene in the cancer cell using the copy number of each target gene in the tumor sample and the correction coefficient.

[Claim 9] The copy-number measurement program according to claim 8,

wherein the distance calculation unit generates a scatter graph indicating a relation between a variant allele frequency of each target position and the number of mapping reads of each target position; converts the scatter graph to a density distribution graph; generates a correlation graph indicating a correlation between a lower area and a upper area, the lower area being, of the density distribution graph, a region expressing a variant allele frequency that is equal to or lower than the reference variant allele frequency, the upper area being, of the density distribution graph, a region expressing a variant allele frequency that is equal to or higher than the reference variant allele frequency; and calculates, as the feature distance, an absolute value of a difference between a variant allele frequency corresponding to a peak correlation value and the reference variant allele frequency in the correlation graph.

[Claim 10] The copy-number measurement program according to claim 9,

wherein the correlation graph indicates a correlation in density between a variant allele frequency in the lower area and a variant allele frequency in the upper area that are equal to each other regarding absolute values of differences thereof from the reference variant allele frequency.

[Claim 11] The copy-number measurement program according to any one of claims 8 to 10,

wherein the coefficient calculation unit calculates a value corresponding to a deviation amount between a relation graph and a measurement point, as the correction coefficient, the relation graph indicating a relation between the feature distance and a logarithmic value of a proportion of a copy number of a gene in a cancer cell to a copy number of a gene in a normal cell, the measurement point indicating a feature distance of a target gene, and a logarithmic value of a proportion of a copy number of the target gene in the tumor sample to a copy number of the target gene in a normal sample.

[Claim 12] The copy-number measurement program according to any one of claims 8 to 11, comprising

a content ratio calculation unit to calculate a content ratio of the cancer cell in the tumor sample based on a copy number of each target gene in the cancer cell.

[Claim 13] The copy-number measurement program according to claim 12,

wherein the content ratio calculation unit calculates a content ratio candidate using a copy number in the cancer cell for each target gene, and determines the content ratio of the cancer cell in the tumor sample based on the content ratio candidate of each target gene.

[Claim 14] The copy-number measurement program according to any one of claims 8 to 13,

wherein the tumor sample is a sample of a brain tumor, and

wherein the target gene is at least one of ATRX, IDH1, IDH2, TP53, TERT, BRAF, PDGFRA, MET, EGFR, BRSK1, EHD2, AKT2, TP73, NMNAT1, TGFBR3, and PTEN.

[Claim 15] A copy-number measurement method comprising:

by a position identification unit, mapping a plurality of tumor sample reads which are a plurality of reads obtained from a tumor sample involving a cancer cell to a human genome sequence, and identifying, for each target gene, a target position which is a genome position of a base, the genome position having changed with respect to the human genome sequence;

by a frequency calculation unit, calculating a variant allele frequency for each target position of each target gene;

by a distance calculation unit, calculating, for each target gene, a feature distance equivalent to a difference between a variant allele frequency corresponding to a peak density and a reference variant allele frequency in a density distribution indicating a density of a number of mapping reads with respect to the variant allele frequency, the number of mapping reads being a number of tumor sample reads mapped to respective target positions in the target gene;

by a coefficient calculation unit, calculating a correction coefficient being used for correcting a copy number of each target gene in the tumor sample, using the feature distance of each target gene; and

by a copy-number calculation unit, calculating a copy number of each target gene in the cancer cell using the copy number of each target gene in the tumor sample and the correction coefficient.