

ANALYSIS OF FRAGMENTATION PATTERNS OF CELL-FREE DNA

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority from U.S. Provisional Application Nos. 62/196,250 filed July 23, 2015 and 62/294,948 filed February 12, 2016 and from PCT

5 Application No. PCT/CN2016/073753 filed February 14, 2016, the entire contents of which are herein incorporated by reference for all purposes.

BACKGROUND

[0002] In previous studies, it was shown that plasma DNA mostly consists of short fragments of less than 200 bp (Lo et al. Sci Transl Med 2010; 2(61):61ra91). In the size
10 distribution of plasma DNA, a peak could be observed at 166 bp. In addition, it was observed that the sequenced tag density would vary with a periodicity of around 180 bp close to transcriptional start sites (TSSs) when maternal plasma DNA was sequenced (Fan et al. PNAS 2008;105:16266-71). These results are one set of evidence that the fragmentation of plasma DNA may not be a random process. However, the precise patterns of DNA
15 fragmentation in plasma, as well as the factors governing the patterns, have not been clear. Further, practical applications of using the DNA fragmentation have not been fully realized.

SUMMARY

[0002a] In one aspect, the present disclosure refers to a method of analyzing a biological sample, including a mixture of cell free DNA molecules from a plurality of tissues types that
20 includes a first tissue type, to determine a classification of a proportional contribution of the first tissue type in the mixture, the method comprising: identifying a first set of genomic positions at which ends of cell-free DNA molecules of the first tissue type occur at a rate above a threshold; analyzing, by a computer system, a first plurality of cell-free DNA molecules from the biological sample of a subject, wherein analyzing a cell-free DNA
25 molecule includes: determining a genomic position in a reference genome corresponding to at least one end of the cell-free DNA molecule; based on the analyzing of the first plurality of cell-free DNA molecules, determining that a first number of the first plurality of cell-free DNA molecules end within one of a plurality of windows, each window including at least one of the first set of genomic positions; computing a relative abundance of the first plurality of

cell-free DNA molecules ending within one of the plurality of windows by normalizing the first number of the first plurality of cell-free DNA molecules using a second number of cell-free DNA molecules, wherein the second number of cell-free DNA molecules includes cell-free DNA molecules ending at a second set of genomic positions outside of the plurality of windows including the first set of genomic positions; and determining the classification of the proportional contribution of the first tissue type by comparing the relative abundance to one or more calibration values determined from one or more calibration samples whose proportional contributions of the first tissue type are known.

[0002b] In another aspect, the present disclosure refers to a method of analyzing a biological sample, including a mixture of cell free DNA molecules from a plurality of tissues types that includes a first tissue type, to determine a classification of a proportional contribution of the first tissue type in the mixture, the method comprising: identifying at least one genomic region having a fragmentation pattern specific to the first tissue type; analyzing a plurality of cell-free DNA molecules from the biological sample, wherein analyzing a cell-free DNA molecule includes: determining a genomic position in a reference genome corresponding to at least one end of the cell-free DNA molecule; identifying a first set of first genomic positions, each first genomic position having a local minimum of ends of cell-free DNA molecules corresponding to the first genomic position; identifying a second set of second genomic positions, each second genomic position having a local maximum of ends of cell-free DNA molecules corresponding to the second genomic position; determining a first number of cell-free DNA molecules ending on any one of the first genomic positions in any one of the at least one genomic region; determining a second number of cell-free DNA molecules ending on any one of the second genomic positions in any one of the at least one genomic region; computing a separation value using the first number and the second number; and determining the classification of the proportional contribution of the first tissue type by comparing the separation value to one or more calibration values determined from one or more calibration samples whose proportional contributions of the first tissue type are known.

[0002c] In another aspect, the present disclosure refers to a method of analyzing a biological sample, including a mixture of cell free DNA molecules from a plurality of tissues types that includes a first tissue type, to determine a genotype of the first tissue type, the first tissue type potentially having a different genotype than other tissue types of the plurality of tissue types, the method comprising: identifying a first genomic position at which ends of cell-free DNA

molecules of the first tissue type occur at a rate above a threshold; analyzing, by a computer system, a first plurality of cell-free DNA molecules from the biological sample of a subject, wherein analyzing a cell-free DNA molecule includes: determining a genomic position in a reference genome corresponding to at least one end of the cell-free DNA molecule; based on the analyzing of the first plurality of cell-free DNA molecules, identifying a set of cell-free DNA molecules that end at the first genomic position; for each of the set of cell-free DNA molecules: determining a corresponding base occurring at the first genomic position, thereby determining corresponding bases at the first genomic position; determining the genotype of the first tissue type at the first genomic position using the corresponding bases occurring at the first genomic position in the set of cell-free DNA molecules.

[0002d] In another aspect, the present disclosure refers to a method of analyzing a biological sample, including a mixture of cell free DNA molecules from a plurality of tissues types that includes a first tissue type, the method comprising: analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample of a subject, each of the plurality of cell-free DNA molecules having a left end and a right end, wherein analyzing a cell-free DNA molecule includes: determining a left ending position in a reference genome corresponding to the left end of the cell-free DNA molecule; determining a right ending position in the reference genome corresponding to the right end of the cell-free DNA molecule; identifying a left set of left genomic positions, each having a local maximum of left ends of the plurality of cell-free DNA molecules corresponding to one of the left set of left genomic positions; identifying a right set of right genomic positions, each having a local maximum of right ends of the plurality of cell-free DNA molecules corresponding to one of the right set of right genomic positions; identifying a first set of genomic positions as being specific to the first tissue type by: comparing left genomic positions of the left set to right genomic positions of the right set to identify the first set of genomic positions where a distance from a left genomic position to a nearest right genomic position is greater than a first threshold distance, the first threshold distance being at least 5 genomic positions in the reference genome.

[0002e] In another aspect, the present disclosure refers to a computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform an operation of any of the methods above.

[0002f] In another aspect, the present disclosure refers to a system comprising: the computer product as disclosed herein; and one or more processors for executing instructions stored on the computer readable medium.

5 **[0002g]** In another aspect, the present disclosure refers to a system comprising means for performing any of the methods above.

[0002h] In another aspect, the present disclosure refers to a system configured to perform any of the above methods.

[0002i] In another aspect, the present disclosure refers to a system comprising modules that respectively perform the steps of any of the above methods.

10 **[0003]** Various examples are directed to applications (e.g., diagnostic applications) of the analysis of the fragmentation patterns of cell-free DNA, e.g., plasma DNA and serum DNA. Examples of one application can determine a classification of a proportional contribution of a particular tissue type in a mixture of cell-free DNA from different tissue types. For example, specific percentages, range of percentages, or whether the proportional contribution is above
15 a specified percentage can be determined as a classification. In one example, preferred ending positions for the particular tissue type can be identified, and a relative abundance of cell-free DNA molecules ending on the preferred ending positions can be used to provide the classification of the proportional contribution. In another example, an amplitude in a fragmentation pattern (e.g., number of cell-free DNA molecules ending at a genomic
20 position) in a region specific to the particular tissue type can be used.

[0004] Examples of another application can determine a genotype of a particular tissue type in a mixture of cell-free DNA from different tissue types. In one example, preferred ending positions for the particular tissue type can be identified, and the genotype can be determined using cell-free DNA molecules ending on the preferred ending positions.

5 [0005] Examples of another application can identify preferred ending positions by comparing a local maximum for left ends of cell-free DNA molecules to a local maximum for right ends of cell-free DNA molecules. Preferred ending positions can be identified when corresponding local maximum are sufficiently separated. Further, amounts of cell-free DNA molecules ending on a local maximum for left/right end can be compared to an amount of
10 cell-free DNA molecules for a local maximum with low separation to determine a proportional contribution of a tissue type.

[0006] Other examples are directed to systems, portable consumer devices, and computer readable media associated with methods described herein.

[0007] A better understanding of the nature and advantages of embodiments of the present
15 invention may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. 1 shows an illustrative example for the definition of intact probability (P_I) according to embodiments of the present invention.

20 [0009] FIGS. 2A and 2B shows variation in P_I across a segment on chromosome 6 using 25 as the value of z , according to embodiments of the present invention.

[0010] FIG. 3 shows the illustration of the synchronous variation of P_I for maternally and fetally-derived DNA in maternal plasma.

[0011] FIG. 4 shows an illustration of asynchronous variation of P_I for maternally and
25 fetally derived DNA in maternal plasma.

[0012] FIG. 5 is a flowchart showing an analysis on whether maternal and fetal DNA molecules are synchronous in the variation in P_I .

WHAT IS CLAIMED IS:

1. A method of analyzing a biological sample, including a mixture of cell-free DNA molecules from a plurality of tissues types that includes a first tissue type, to determine a classification of a proportional contribution of the first tissue type in the mixture, the method comprising:

identifying a first set of genomic positions at which ends of cell-free DNA molecules of the first tissue type occur at a rate above a threshold;

analyzing, by a computer system, a first plurality of cell-free DNA molecules from the biological sample of a subject, wherein analyzing a cell-free DNA molecule includes:

determining a genomic position in a reference genome corresponding to at least one end of the cell-free DNA molecule;

based on the analyzing of the first plurality of cell-free DNA molecules, determining that a first number of the first plurality of cell-free DNA molecules end within one of a plurality of windows, each window including at least one of the first set of genomic positions;

computing a relative abundance of the first plurality of cell-free DNA molecules ending within one of the plurality of windows by normalizing the first number of the first plurality of cell-free DNA molecules using a second number of cell-free DNA molecules, wherein the second number of cell-free DNA molecules includes cell-free DNA molecules ending at a second set of genomic positions outside of the plurality of windows including the first set of genomic positions; and

determining the classification of the proportional contribution of the first tissue type by comparing the relative abundance to one or more calibration values determined from one or more calibration samples whose proportional contributions of the first tissue type are known.

2. The method of claim 1, wherein identifying the first set of genomic positions includes:

analyzing, by the computer system, a second plurality of cell-free DNA molecules from at least one first additional sample to identify ending positions of the second plurality of cell-free DNA molecules, wherein the at least one first additional sample is known to include the first tissue type and is of a same sample type as the biological sample;

for each genomic window of a plurality of genomic windows:
 computing a corresponding number of the second plurality of cell-free DNA molecules ending on the genomic window; and
 comparing the corresponding number to a reference value to determine whether the rate of cell-free DNA molecules ending on one or more genomic positions within the genomic window is above the threshold.

3. The method of claim 2, wherein a first genomic window of the plurality of genomic windows has a width greater than one genomic position, and wherein each of the genomic positions within the first genomic window are identified as having the rate of cell-free DNA molecules ending on the genomic position be above the threshold when the corresponding number exceeds the reference value.

4. The method of claim 2, wherein the first set of genomic positions have the highest N values for the corresponding numbers, wherein N is at least 10,000.

5. The method of claim 2, further comprising:
 determining a size of each of the second plurality of cell-free DNA molecules, wherein identifying the first set of genomic positions further includes:

 determining a first statistical value of a size distribution of cell-free DNA molecules of the second plurality of cell-free DNA molecules ending within a first genomic window determined to have the rate above the threshold;

 comparing the first statistical value to a size threshold; and

 excluding the first genomic window from the first set of genomic positions when the first statistical value does not exceed the size threshold.

6. The method of claim 2, wherein the one or more calibration samples include the at least one first additional sample.

7. The method of claim 6, further comprising:
 for each of the one or more calibration samples:
 measuring a corresponding proportional contribution of the first tissue type; and
 determining a corresponding relative abundance using the corresponding numbers of the second plurality of cell-free DNA molecules ending within the plurality of windows corresponding to the first set of genomic positions, thereby obtaining a

calibration data point, wherein each calibration data point specifies the measured proportional contribution of the first tissue type for the at least one first additional sample and the corresponding relative abundance.

8. The method of claim 7, wherein the one or more calibration data points are a plurality of calibration data points that form a calibration function that approximates the plurality of calibration data points.

9. The method of claim 2, wherein each genomic position of the first set of genomic positions has at least a specified number of cell-free DNA molecules of the second plurality of cell-free DNA molecules ending on the genomic position.

10. The method of claim 2, wherein the reference value is an expected number of cell-free DNA molecules ending within the genomic window according to a probability distribution and an average length of cell-free DNA molecules in the at least one first additional sample.

11. The method of claim 10, wherein the probability distribution is a Poisson distribution, and wherein determining whether the rate of cell-free DNA molecules ending on one or more genomic positions within the genomic window is above the threshold includes:

determining a corresponding p-value using the corresponding number and the expected number, wherein the threshold corresponds to a cutoff p-value, the corresponding p-value being less than the cutoff p-value indicating that the rate of cell-free DNA molecules ending within the genomic window is above the threshold.

12. The method of claim 2, wherein the genomic positions whose rate of the second plurality of cell-free DNA molecules ending on the genomic positions is above the threshold comprises a first superset, and wherein identifying the first set of genomic positions further includes:

analyzing, by the computer system, a third plurality of cell-free DNA molecules from at least one second additional sample identified as having a reduced amount of the first tissue type to identify a second superset of the third plurality of cell-free DNA molecules ending on the genomic position is above the threshold; and

identifying the first set of genomic positions as including the genomic positions that are in the first superset and that are not in the second superset.

13. The method of claim 2, wherein the reference value includes a measured number of cell-free DNA molecules ending within the genomic window, the measured number determined from a third plurality of cell-free DNA molecules of at least one second additional sample identified as not having the first tissue type.

14. The method of claim 13, further comprising:
determining a size of each of the second plurality of cell-free DNA molecules, wherein identifying the first set of genomic positions further includes:

determining a first statistical value of a first size distribution of cell-free DNA molecules of the second plurality of cell-free DNA molecules ending on a first genomic position determined to have the rate above the threshold;

determining a second statistical value of a second size distribution of cell-free DNA molecules of the third plurality of cell-free DNA molecules ending on one or more second genomic positions determined to have the rate above the threshold;

comparing the first statistical value to the second statistical value; and

excluding the first genomic position from the first set of genomic positions when the first statistical value does not exceed the second statistical value by at least a specified amount to indicate that the first size distribution is smaller than the second size distribution.

15. The method of claim 13, wherein comparing the corresponding number to the reference value includes:

computing a first ratio of the corresponding number and a third number of the second plurality of cell-free DNA molecules covering the genomic window; and

comparing the first ratio to the reference value, the reference value including a reference ratio of the measured number of reads ending within the genomic window and a fourth number of the third plurality of cell-free DNA molecules covering the genomic window and not ending within the genomic window.

16. The method of claim 15, wherein the third number of the second plurality of cell-free DNA molecules do not end within the genomic window.

17. The method of claim 15, wherein determining whether the rate of cell-free DNA molecules ending within the genomic window is above the threshold includes:

determining whether the first ratio is greater than a multiplicative factor times the reference ratio.

18. The method of claim 2, wherein the sample type of the biological sample and the at least one first additional sample is selected from a group consisting of plasma, serum, cerebrospinal fluid, and urine.

19. The method of claim 2, wherein the genomic window is a genomic position, and wherein the first tissue type has a plurality of first tissue-specific alleles, and wherein computing the corresponding number of the second plurality of cell-free DNA molecules ending on the genomic position includes:

identifying whether the cell-free DNA molecule ending on the genomic position includes at least one of the plurality of first tissue-specific alleles;

including the cell-free DNA molecule in the corresponding number when the cell-free DNA molecule includes a first tissue-specific allele; and

not including the cell-free DNA molecule in the corresponding number when the cell-free DNA molecule does not include a first tissue-specific allele.

20. The method of claim 1, wherein the first tissue type has a plurality of first tissue-specific alleles in at least one additional sample, and wherein the first set of genomic positions are determined using cell-free DNA molecules of the least one additional sample that include at least one of the plurality of first tissue-specific alleles.

21. The method of claim 20, wherein the second set of genomic positions are such that ends of cell-free DNA molecules of a second tissue type occur at a rate above the threshold in the at least one additional sample, wherein the second tissue type has a plurality of second tissue-specific alleles in the at least one additional sample, and wherein the second set of genomic positions are determined using cell-free DNA molecules of the at least one additional sample that include at least one of the plurality of second tissue-specific alleles.

22. The method of claim 21, wherein the at least one additional sample is from a pregnant female, and wherein the first tissue type is fetal tissue and the second tissue type is maternal tissue.

23. The method of claim 21, wherein genomic positions at which ends of cell-free DNA molecules having a shared allele between the first tissue type and the second tissue type occur at a second rate above the threshold are excluded from the first set of genomic positions and excluded from the second set of genomic positions.

24. The method of claim 1, wherein the relative abundance includes a ratio of the first number and the second number.

25. The method of claim 1, wherein the plurality of windows have a width of one genomic position, and wherein the relative abundance is computed by:

for each genomic position of the first set of genomic positions:

computing a corresponding number of the first plurality of cell-free DNA molecules ending on the genomic position as part of determining that the first number of the first plurality of cell-free DNA molecules end on any one of the first set of genomic positions;

computing a third number of the first plurality of cell-free DNA molecules covering the genomic position and not ending on the genomic position as part of determining the second number of cell-free DNA molecules; and

computing a first ratio of the corresponding number and the third number;

and

computing a mean of the first ratios as the relative abundance.

26. The method of claim 1, wherein the relative abundance is computed by:

for each genomic position of the first set of genomic positions:

computing a corresponding number of the first plurality of cell-free DNA molecules ending within a first window including the genomic position as part of determining that the first number of the first plurality of cell-free DNA molecules end within one of the plurality of windows;

computing a third number of the first plurality of cell-free DNA molecules ending within a second window including the genomic position, the second window larger than the first window; and

computing a first ratio of the corresponding number and the third number;
and

computing a mean of the first ratios as the relative abundance.

27. The method of claim 1, wherein the second set of genomic positions and the first set of genomic positions do not overlap.

28. The method of claim 1, wherein the second set of genomic positions includes all genomic positions corresponding to an end of at least one of the first plurality of cell-free DNA molecules.

29. The method of claim 1, wherein analyzing one or more of the cell-free DNA molecules includes determining both genomic positions corresponding to both ends of the cell-free DNA molecule.

30. The method of claim 1, wherein the classification of the proportional contribution corresponds to a range above a specified percentage.

31. The method of claim 1, wherein the first tissue type is a tumor.

32. The method of claim 31, wherein the classification is selected from a group consisting of: an amount of tumor tissue in the subject, a size of the tumor in the subject, a stage of the tumor in the subject, a tumor load in the subject, and presence of tumor metastasis in the subject.

33. The method of claim 1, wherein the one or more calibration samples are from the subject and are obtained at a different time than the biological sample.

34. The method of claim 1, further comprising:
obtaining template DNA molecules from the biological sample to be analyzed;
preparing a sequencing library of analyzable DNA molecules using the template DNA molecules, the preparing of the sequencing library of analyzable DNA molecules not including a step of DNA amplification of the template DNA molecules; and

sequencing the sequencing library of analyzable DNA molecules to obtain a plurality of sequence reads corresponding to the first plurality of cell-free DNA molecules, wherein analyzing the first plurality of cell-free DNA molecules includes: receiving, at the computer system, the plurality of sequence reads; and aligning, by the computer system, the plurality of sequence reads to the reference genome to determine genomic positions for the plurality of sequence reads.

35. The method of claim 1, further comprising providing a therapeutic intervention based on the classification or performing imaging of the subject based on the classification.

36. The method of claim 1, wherein the first set of genomic positions comprises between 600 and 10,000 genomic positions.

37. A method of analyzing a biological sample, including a mixture of cell-free DNA molecules from a plurality of tissues types that includes a first tissue type, to determine a classification of a proportional contribution of the first tissue type in the mixture, the method comprising:

identifying at least one genomic region having a fragmentation pattern specific to the first tissue type;

analyzing a plurality of cell-free DNA molecules from the biological sample, wherein analyzing a cell-free DNA molecule includes:

determining a genomic position in a reference genome corresponding to at least one end of the cell-free DNA molecule;

identifying a first set of first genomic positions, each first genomic position having a local minimum of ends of cell-free DNA molecules corresponding to the first genomic position;

identifying a second set of second genomic positions, each second genomic position having a local maximum of ends of cell-free DNA molecules corresponding to the second genomic position;

determining a first number of cell-free DNA molecules ending on any one of the first genomic positions in any one of the at least one genomic region;

determining a second number of cell-free DNA molecules ending on any one of the second genomic positions in any one of the at least one genomic region;

computing a separation value using the first number and the second number;
and

determining the classification of the proportional contribution of the first tissue type by comparing the separation value to one or more calibration values determined from one or more calibration samples whose proportional contributions of the first tissue type are known.

38. The method of claim 37, wherein the first set of first genomic positions includes multiple genomic positions, wherein the second set of second genomic positions includes multiple genomic positions,

wherein determining the first number of cell-free DNA molecules includes determining a first amount of cell-free DNA molecules ending on each first genomic position, thereby determining a plurality of first amounts,

wherein determining the second number of cell-free DNA molecules includes determining a second amount of cell-free DNA molecules ending on each second genomic position, thereby determining a plurality of second amounts, and

wherein computing the separation value includes:

determining a plurality of separate ratios, each separate ratio of one of the plurality of first amounts and one of the plurality of second amounts, and

determining the separation value using the plurality of separate ratios.

39. The method of claim 37, wherein the at least one genomic region includes one or more DNase hypersensitivity sites.

40. The method of claim 37, wherein each of the at least one genomic region having a fragmentation pattern specific to the first tissue type includes one or more first tissue-specific alleles in at least one additional sample.

41. The method of claim 37, wherein the at least one genomic region includes one or more ATAC-seq or micrococcal nuclease sites.

42. The method of claim 37, wherein the cell-free DNA molecules aligned to one genomic position of the first set of first genomic positions extend a specified number of nucleotides to both sides of the one genomic position.

43. The method of claim 42, wherein the specified number is between 10 and 80 nucleotides.

44. The method of claim 37, wherein identifying the first set of first genomic positions includes:

for each of a plurality of genomic positions:

determining a first amount of cell-free DNA molecules that are located at the genomic position and extend a specified number of nucleotides to both sides of the genomic position;

determining a second amount of cell-free DNA molecules that are located at the genomic position; and

determining a ratio of the first amount and the second amount; and

identifying a plurality of local minima and a plurality of local maxima in the ratios.

45. The method of claim 37, wherein the mixture is plasma or serum.

46. The method of claim 37, wherein the plurality of cell-free DNA molecules is at least 1,000 cell-free DNA molecules.

47. The method of claim 44, wherein, for a given genomic position of the plurality of genomic positions, the second amount corresponds to a total number of the cell-free DNA molecules aligning to the given genomic position.

48. A method of analyzing a biological sample, including a mixture of cell-free DNA molecules from a plurality of tissues types that includes a first tissue type, to determine a genotype of the first tissue type, the first tissue type potentially having a different genotype than other tissue types of the plurality of tissue types, the method comprising:

identifying a first genomic position at which ends of cell-free DNA molecules of the first tissue type occur at a rate above a threshold;

analyzing, by a computer system, a first plurality of cell-free DNA molecules from the biological sample of a subject, wherein analyzing a cell-free DNA molecule includes:

determining a genomic position in a reference genome corresponding to at least one end of the cell-free DNA molecule;

based on the analyzing of the first plurality of cell-free DNA molecules,
identifying a set of cell-free DNA molecules that end at the first genomic position;
for each of the set of cell-free DNA molecules:
determining a corresponding base occurring at the first genomic position,
thereby determining corresponding bases at the first genomic position;
determining the genotype of the first tissue type at the first genomic position
using the corresponding bases occurring at the first genomic position in the set of cell-free
DNA molecules.

49. The method of claim 48, further comprising:
filtering the set of cell-free DNA molecules to exclude or modify a weighting
of at least one of the cell-free DNA molecules that end at the first genomic position, wherein
the genotype is determined using a filtered set of cell-free DNA molecules.

50. The method of claim 49, wherein the filtering uses at last one of: a size
of a cell-free DNA molecule, a methylation status of the cell-free DNA molecule at one or
more positions, and whether the cell-free DNA molecule covers one or more other genomic
position at which ends of cell-free DNA molecules of the first tissue type occur at the rate
above the threshold.

51. The method of claim 50, wherein the filtering assigns a weight to the
cell-free DNA molecule corresponding to a likelihood that the cell-free DNA molecule is
from the first tissue type, the method further comprising:
determining a weighted sum for each of a plurality of bases; and
determining a percentage contribution for each of the plurality of bases using
the weighted sums, wherein the genotype is determined using the percentage contributions.

52. The method of claim 48, wherein determining the genotype of the first
tissue type at the first genomic position includes:
determining a percentage contribution for each of a plurality of bases; and
comparing each of the percentage contributions to one or more cutoff values.

53. The method of claim 52, wherein a first cutoff value of the one or more
cutoff values corresponds to a homozygous genotype of a first base when the percentage
contribution of the first base is above the first cutoff value.

54. The method of claim 52, wherein a first cutoff value and a second cutoff value of the one or more cutoff values correspond to a heterozygous genotype for a first base and a second base when the percentage contributions of the first base and the second base are above the first cutoff value and below the second cutoff value.

55. The method of claim 48, wherein the first tissue type corresponds to a tumor.

56. The method of claim 48, wherein the first tissue type corresponds to a fetus, and wherein the subject is pregnant with the fetus.

57. A method of analyzing a biological sample, including a mixture of cell-free DNA molecules from a plurality of tissues types that includes a first tissue type, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample of a subject, each of the plurality of cell-free DNA molecules having a left end and a right end, wherein analyzing a cell-free DNA molecule includes:

determining a left ending position in a reference genome corresponding to the left end of the cell-free DNA molecule;

determining a right ending position in the reference genome corresponding to the right end of the cell-free DNA molecule;

identifying a left set of left genomic positions, each having a local maximum of left ends of the plurality of cell-free DNA molecules corresponding to one of the left set of left genomic positions;

identifying a right set of right genomic positions, each having a local maximum of right ends of the plurality of cell-free DNA molecules corresponding to one of the right set of right genomic positions;

identifying a first set of genomic positions as being specific to the first tissue type by:

comparing left genomic positions of the left set to right genomic positions of the right set to identify the first set of genomic positions where a distance from a left genomic position to a nearest right genomic position is greater than a first threshold distance, the first threshold distance being at least 5 genomic positions in the reference genome.

58. The method of claim 57, further comprising:

identifying a second set of genomic positions by:

comparing left genomic positions of the left set to right genomic positions of the right set to identify the second set of genomic positions where the distance from the left genomic position to the nearest right genomic position is less than a second threshold distance;

determining a separation value using a first number of the plurality of cell-free DNA molecules ending at one of the left set of left genomic positions and a second number of the plurality of cell-free DNA molecules ending at one of the right set of right genomic positions; and

determining a classification of a proportional contribution of the first tissue type by comparing the separation value to one or more calibration values determined from one or more calibration samples whose proportional contributions of the first tissue type are known.

59. The method of claim 58, wherein determining the separation value includes:

identifying pairs of the first set of genomic positions and the second set of genomic positions;

for each of the pairs:

determining a first amount of cell-free DNA molecules ending at a first genomic position of the pair; and

determining a second amount of cell-free DNA molecules ending at a second genomic position of the pair,

wherein the first amounts of cell-free DNA molecules correspond to the first number of the plurality of cell-free DNA molecules and the second amounts of cell-free DNA molecules correspond to the second number of the plurality of cell-free DNA molecules.

60. The method of claim 59, wherein determining the separation value includes:

for each of the pairs:

determining a ratio including the first amount and the second amount; and
determining the separation value from the ratios.

61. The method of claim 59, wherein the pairs of the first set of genomic positions and the second set of genomic positions are nearest to each other.

62. The method of claim 58, wherein the second threshold distance is less than 5 genomic positions in the reference genome.

63. The method of claim 57, wherein the first set of genomic positions include both left genomic positions and right genomic positions.

64. A computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform an operation of any of the methods above.

65. A system comprising:
the computer product of claim 64; and
one or more processors for executing instructions stored on the computer readable medium.

66. A system comprising means for performing any of the methods above.

67. A system configured to perform any of the above methods.

68. A system comprising modules that respectively perform the steps of any of the above methods.