DNA molecules, as the libraries are denatured, diluted and diffused on the two-dimensional surface followed by bridge amplification (Head et al. Biotechniques. 2014;56:61-4).

~~BRIEF~~ SUMMARY

[0002a] In one aspect, the present disclosure refers to a method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

[0002b] In another aspect, the present disclosure refers to a computer product comprising a non-transitory computer readable medium storing instructions that, when executed, control a computing system to perform a method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

[0003] Methods and systems described herein involve using long cell-free DNA fragments to analyze a biological sample. Using these long cell-free DNA fragments allows for analysis not contemplated or not possible with shorter cell-free DNA fragments. The status of methylated CpG sites and single nucleotide polymorphisms (SNPs) is often used to analyze DNA fragments of a biological sample. A CpG site and a SNP are typically separated from the nearest CpG site

or SNP by hundreds or thousands of base pairs. The length of most of the cell-free DNA fragments in a biological sample is usually less than 200 bp. As a result, finding two or more consecutive CpG sites or SNPs on most cell-free DNA fragments is improbable or impossible. Cell-free DNA fragments longer than 200 bp, including those longer than 600 bp or 1 kb, may include multiple CpG sites and/or SNPs. The presence of multiple CpG sites and/or SNPs on long cell-free DNA fragments may allow for more efficient and/or accurate analysis than with short cell-free DNA fragments alone. The long cell-free DNA fragments can be used to identify a tissue of origin and/or to provide information on a fetus in a pregnant female. In addition, using long cell-free DNA fragments to accurately analyze samples from pregnant women is surprising as one would expect that such long cell-free DNA fragments are predominantly maternal in origin. One would not expect that long cell-free DNA fragments of fetal origin are present in sufficient amounts to provide information about the fetus.

[0004] Long cell-free DNA fragments with a SNP present may be used to determine the haplotype inherited by a fetus. Long cell-free DNA fragments, by having multiple CpG sites, may have a methylation pattern that indicates a tissue of origin. Additionally, trinucleotide repeats and other repeated sequences may be present on long cell-free DNA fragments. These repeats may be used to determine the likelihood of a genetic disorder in fetus or the paternity of a fetus. The amount of long cell-free DNA fragments may be used to determine gestational age. Similarly, the motifs at the end of long cell-free DNA fragments may also be used to determine gestational age. The long-cell free DNA fragments (including, for example, amounts, length

distribution, genomic locations, methylation status, etc. of such fragments) may be used to determine a pregnancy-associated disorder.

[0005] These and other ~~embodiments~~ examples of the disclosure are described in detail below. For example, other ~~embodiments~~ examples are directed to systems, devices, and computer readable media associated with methods described herein.

[0006] A better understanding of the nature and advantages of ~~embodiments~~ examples of the present disclosure may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIGS. 1A and 1B show the size distribution of cell-free DNA determined according to embodiments of the present invention. (A) 0–20 kb on a linear scale, (B) 0–20 kb on a logarithmic scale.

[0008] FIGS. 2A and 2B show the size distribution of cell-free DNA determined according to embodiments of the present invention. (A) 0–5 kb on a linear scale for the y-axis. (B) 0–5 kb on a logarithmic scale for the y-axis.

[0009] FIGS. 3A and 3B show the size distribution of cell-free DNA determined according to embodiments of the present invention. (A) 0–400 bp on a linear scale for the y-axis. (B) 0–400 bp on a logarithmic scale for the y-axis.

[0010] FIGS. 4A and 4B show the size distribution of cell-free DNA between fragments carrying shared alleles (Shared) and fetal-specific alleles (Fetal-specific) determined according to embodiments of the present invention. (A) 0–20 kb bp on a linear scale for the y-axis. (B) 0–20 kb on a logarithmic scale for the y-axis. The blue line indicates the fragments carrying shared alleles (predominant of maternal origin) and the red line indicates the fragments carrying fetal-specific alleles (of placental origin).

[0011] FIGS. 5A and 5B show the size distribution of cell-free DNA between fragments carrying shared alleles (Shared) and fetal-specific alleles (Fetal-specific) determined according to embodiments of the present invention. (A) 0–5 kb bp on a linear scale for the y-axis. (B) 0–5 kb

WHAT IS CLAIMED IS:

1. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value;

determining a sequence of the first haplotype and a sequence of the second haplotype from reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

2. The method of claim 1, wherein the cutoff value is 600 nt.

3. The method of claim 1, wherein the cutoff value is 1 knt.

4. The method of any one of claims 1 to 3, wherein determining the sequence of the first haplotype and the sequence of the second haplotype from the reads corresponding to the first set of cell-free DNA molecules comprises:

aligning reads corresponding to the first set of cell-free DNA molecules to a reference genome.

5. The method of claim 1, wherein determining the sequence of the first haplotype and the sequence of the second haplotype from the reads corresponding to the first set of cell-free DNA molecules comprises:

aligning a first subset of the reads to a second subset of the reads to identify a different allele at a locus in the reads,

determining that the first subset of the reads have a first allele at the locus,

determining that the second subset of the reads have a second allele at the locus,

determining that the first subset of the reads corresponds to the first haplotype, and

determining that the second subset of the reads corresponds to the second haplotype.

6. The method of any one of claims 1 to 5, wherein the parameter is a count of cell-free DNA molecules, a size profile of cell-free DNA molecules, or a methylation level of cell-free DNA molecules.

7. The method of claim 6, wherein:

the parameter is the count of cell-free DNA molecules, and

the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is greater than the second value.

8. The method of claim 6, wherein:

the parameter is the size profile of cell-free DNA molecules, and

the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is less than the second value, indicating that the second set of cell-free DNA molecules is characterized by a smaller size profile than the third set of cell-free DNA molecules.

9. The method of claim 6, wherein:

the parameter is the methylation level of cell-free DNA molecules, and the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is less than the second value.

10. The method of any one of claims 1 to 9, further comprising:

calculating a separation value using the first value and the second value;

comparing the separation value to a cutoff value; and

determining a likelihood of a fetal aneuploidy based on the comparison of the separation value to the cutoff value.

11. The method of claim 10, wherein:

the cutoff value is determined from reference samples from pregnant females with euploid fetuses,

the cutoff value is determined from reference samples from pregnant females with aneuploid fetuses, or

the cutoff value is calculated assuming an aneuploid fetus.

12. The method of any one of claims 1 to 11, further comprising:

identifying a number of repeats of a subsequence in a read of the reads corresponding to the first set of cell-free DNA molecules,

wherein:

determining the sequence of the first haplotype comprises determining the sequence of the first haplotype comprises the number of repeats of the subsequence.

13. The method of claim 12, wherein:

the repeats of the subsequence is associated with a repeat-associated disease, and

the method further comprises determining a likelihood of the fetus inheriting the repeat-associated disease.

14. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving sequence reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value; and

for a cell-free DNA molecule of the set of cell-free DNA molecules:

determining a methylation status at each site of a plurality of sites,

determining a methylation pattern, wherein:

the methylation pattern indicates a methylation status at each site of the plurality of sites, using one or more sequence reads corresponding to the cell-free DNA molecule,

comparing the methylation pattern to one or more reference patterns, wherein each of the one or more reference patterns is determined for a particular tissue type; and

determining a tissue of origin of the cell-free DNA molecule using the methylation pattern.

15. The method of claim 14, wherein the cutoff value is 600 nt.

16. The method of claim 14, wherein the cutoff value is 1 knt.

17. The method of any one of claims 14 to 16, further comprising determining the tissue of origin for each cell-free DNA molecule of the set of cell-free DNA molecules by:

determining the methylation status at each site of a plurality of respective sites, wherein the plurality of respective sites corresponds to the cell-free DNA molecule,

determining the methylation pattern, and

comparing the methylation pattern to at least one reference pattern of the one or more reference patterns.

18. The method of claim 17, further comprising:

determining an amount of cell-free DNA molecules corresponding to each tissue of origin, and

determining a fractional contribution of the tissue of origin in the biological sample using the amount of cell-free DNA molecules corresponding to each tissue of origin.

19. The method of any one of claims 14 to 18, wherein measuring the sizes of the plurality of cell-free DNA molecules comprises:

aligning the sequence reads to a reference genome.

20. The method of any one of claims 14 to 18, wherein measuring sizes of the plurality of cell-free DNA molecules comprises:

full-length sequencing of the plurality of cell-free DNA molecules, and

counting the number of nucleotides in each cell-free DNA molecule of the plurality of cell-free DNA molecules.

21. The method of claim 14 or 17, wherein measuring the sizes of the plurality of cell-free DNA molecules comprises:

physically separating the plurality of cell-free DNA molecules from the biological sample from other cell-free DNA molecules in the biological sample, wherein the other cell-free DNA molecules have sizes less than the cutoff value.

22. The method of any one of claims 14 to 21, wherein a reference pattern of the one or more reference patterns is determined by:

measuring a methylation density at each reference site of a plurality of reference sites using DNA molecules from a reference tissue,

comparing the methylation density at each reference site of the plurality of reference sites to one or more threshold methylation densities, and

identifying each reference site of the plurality of reference sites as methylated, unmethylated, or non-informative based on comparing the methylation density to the one or more threshold methylation densities, wherein the plurality of sites is the plurality of reference sites that are identified as methylated or unmethylated.

23. The method of any one of claims 14 to 22, wherein the tissue of origin is the placenta.

24. The method of any one of claims 14 to 22, wherein the tissue of origin is fetal or maternal.

25. The method of claim 24, wherein:

the tissue of origin is fetal,

the method further comprising:

aligning a sequence read of the sequence reads to a first region of a reference genome, the first region comprising a plurality of sites corresponding to alleles, the plurality of sites including a threshold number of sites,

determining a first haplotype using the respective allele present at each site of the plurality of sites,

comparing the first haplotype to a second haplotype corresponding to a male subject, and

determining a classification of a likelihood that the male subject being the father of the fetus using the comparison.

26. The method of claim 24, wherein:

the tissue of origin is fetal,

the method further comprising:

aligning a sequence read of the sequence reads to a first region of a reference genome, the first region comprising a first plurality of sites corresponding to alleles, the plurality of sites including a threshold number of sites,

comparing the allele at each site of the plurality of sites to an allele at the corresponding site in the genome of a male subject, and

determining a classification of a likelihood that the male subject being the father of the fetus using the comparison.

27. The method of claim 24, further comprising:

for each cell-free DNA molecule of the set of cell-free DNA molecules:

aligning the sequence read corresponding to the cell-free DNA molecule to a reference genome,

identifying the sequence read as corresponding to a haplotype present in the female,

determining the tissue of origin as fetal using the methylation pattern, and

determining the haplotype to be a maternally inherited fetal haplotype.

28. The method of claim 27, further comprising:

identifying the haplotype as carrying a disease-causing genetic mutation or variation, and

classifying that the fetus is likely to have the disease caused by the genetic mutation or variation.

29. The method of claim 28, wherein identifying the haplotype as carrying the disease-causing genetic mutation comprises:

identifying the genetic mutation or variation in a first sequence read,

measuring a first methylation level in a second sequence read corresponding to a first genomic location within a first distance of the first sequence read, and

measuring a second methylation level in a third sequence read corresponding to a second genomic location within a second distance of the first sequence read, wherein:

the first methylation level and the second methylation level are associated with the genetic mutation.

30. The method of claim 24, further comprising:

for each cell-free DNA molecule of the set of cell-free DNA molecules:

aligning the sequence read corresponding to the cell-free DNA molecule to a reference genome,

identifying the sequence read as corresponding to a region, wherein the region is determined by:

receiving a plurality of fetal sequence reads corresponding to a plurality of fetal DNA molecules from fetal tissue,

~~receiving a plurality of maternal sequence reads corresponding to a plurality of maternal DNA molecules,~~

~~determining a fetal methylation status at each methylation site of a plurality of methylation sites within the region for each fetal sequence read of the plurality of fetal sequence reads,~~

~~determining a maternal methylation status at each methylation site of the plurality of methylation sites for each maternal sequence read of the plurality of maternal sequence reads,~~

~~determining value of a parameter characterizing an amount of sites where the fetal methylation status differs from the maternal methylation status,~~

~~comparing the value of the parameter to a threshold value, and~~

~~determining the value of the parameter exceeds the threshold value.~~

~~31. The method of any one of claims 14 to 28, wherein the cutoff value is at least 500 nt.~~

~~32. The method of any one of claims 14 to 31, wherein determining the tissue of origin of the cell-free DNA molecule comprises inputting the methylation pattern into a machine learning model, the model trained by:~~

~~receiving a plurality of training methylation patterns, each training methylation pattern having a methylation status at one or more sites of the plurality of sites, each training methylation pattern determined from a DNA molecule from a known tissue,~~

~~storing a plurality of training samples, each training sample including one of the plurality of training methylation patterns and a label indicating the known tissue corresponding to the training methylation pattern, and~~

~~optimizing, using the plurality of training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels when the plurality of training methylation patterns is input to the model, wherein an output of the model specifies a tissue corresponding to an input methylation pattern.~~

33. The method of claim 32, wherein the machine learning model comprises convolution neural networks (CNN), linear regression, logistic regression, deep recurrent neural network, Bayes's classifier, hidden Markov model (HMM), linear discriminant analysis (LDA), k-means clustering, density-based spatial clustering of applications with noise (DBSCAN), random forest algorithm, or support vector machine (SVM)

34. The method of claim 32, wherein each DNA molecule from the known tissue is cellular DNA.

35. The method of claim 32 or 34, wherein the parameters of the model comprise a first parameter indicating whether one site of the plurality of sites has the same methylation status as another site of the plurality of sites.

36. The method of any one of claims 32 to 35, wherein the parameters of the model comprise a second parameter indicating a distance between sites of the plurality of sites.

37. The method of any one of claims 14 to 31, wherein a reference pattern of the one or more reference pattern corresponds to a reference tissue,

the method further comprising determining the tissue of origin to be the reference tissue when the methylation pattern matches the reference pattern.

38. The method of any one of claims 14 to 37, wherein the plurality of sites comprise at least 5 CpG sites.

39. The method of any one of claims 14 to 31, wherein determining the tissue of origin using the methylation pattern comprises:

determining a similarity score by comparing the methylation pattern with a first reference methylation pattern from a first reference tissue of a plurality of reference tissues;

comparing the similarity score with a threshold value; and

determining the tissue of origin to be the first reference tissue when the similarity score exceeds the threshold value.

40. The method of claim 39, wherein:

the similarity score is a first similarity score,

the method further comprising:

calculating the threshold value by:

determining a second similarity score by comparing the methylation pattern with a second reference methylation pattern from a second reference tissue of the plurality of reference tissues, the first reference tissue and the second reference tissue being different tissues, the threshold value being the second similarity score.

41. The method of claim 39 or 40, wherein:

the first reference methylation pattern comprises a first subset of sites having at least a first probability of being methylated for the first reference tissue,

the first reference methylation pattern comprises a second subset of sites having at most a second probability of being methylated for the first reference tissue, and

determining the similarity score comprises:

increasing the similarity score when a site of the plurality of sites is methylated and the site of the plurality of sites is in the first subset of sites, and

decreasing the similarity score when a site of the plurality of sites is methylated and the site of the plurality of sites is in the second subset of sites.

42. The method of claim 39 or 40, wherein:

the first reference methylation pattern comprises the plurality of sites, with each site of the plurality of sites characterized by a probability of being methylated and a probability of being unmethylated for the first reference tissue,

the similarity score is determined by:

for each site of the plurality of sites:

determining the probability in the reference tissue corresponding to the methylation status of the site in the cell-free DNA molecule,

calculating a product of the plurality of probabilities, the product being the similarity score.

43. The method of claim 42, wherein the probability is determined using a beta distribution.

44. The method of any one of claims 14 to 43, further comprising:

sequencing the plurality of cell-free DNA molecules to obtain the sequence reads, and

determining a methylation status of the site by measuring a characteristic corresponding to a nucleotide of the site and nucleotides neighboring the site.

45. The method of any one of claims 14 to 44, wherein sizes of the plurality of cell-free DNA molecules comprise a number of CpG sites.

46. The method of any one of claims 14 to 45, wherein at least one site of the plurality of sites is methylated.

47. The method of any one of claims 14 to 46, wherein two sites of the plurality of sites are separated by at least 160 nt.

48. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including cell-free DNA molecules from the fetus and the female, the method comprising:

receiving a first sequence read corresponding to a cell-free DNA molecule of the cell-free DNA molecules;

aligning the first sequence read to a region of a reference genome, the region known to potentially include repeats of a subsequence;

identifying a number of repeats of the subsequence in the first sequence read corresponding to the cell-free DNA molecule;

comparing the number of repeats of the subsequence to a threshold number; and

determining a classification of a likelihood of the fetus having a genetic disorder using the comparison of the number of repeats to the threshold number.

49. The method of claim 48, wherein determining the classification of the likelihood of the fetus having the genetic disorder comprises:

~~determining that the fetus is likely to have the genetic disorder when the number of repeats exceeds the threshold number.~~

~~50. The method of claim 48 or 49, wherein the threshold number is 55 or more.~~

~~51. The method of any one of claims 48 to 50, wherein the genetic disorder is fragile X syndrome.~~

~~52. The method of any one of claims 48 to 51, wherein the subsequence is a trinucleotide sequence.~~

~~53. The method of any one of claims 48 to 52, wherein the cell-free DNA molecules have a length greater than a cutoff value.~~

~~54. The method of claim 53, wherein the cutoff value is 600 nt.~~

~~55. The method of claim 53, wherein the cutoff value is 1 knt.~~

~~56. The method of any one of claims 48 to 54, further comprising determining that the cell-free DNA molecule is of fetal origin.~~

~~57. The method of claim 56, wherein:~~

~~the number of repeats of the subsequence in the first sequence read is a first number of repeats of the subsequence,~~

~~determining that the cell-free DNA molecule is of fetal origin comprises:~~

~~receiving a second sequence read corresponding to a cell-free DNA molecule of maternal origin obtained from a buffy coat or a sample of the female before pregnancy,~~

~~aligning the second sequence read to the region of the reference genome,~~

~~identifying a second number of repeats of the subsequence in the second sequence read, and~~

~~determining that the second number of repeats is less than the first number of repeats.~~

~~58. The method of claim 56, wherein:~~

~~determining that the cell-free DNA molecule is of fetal origin comprises:~~

~~determining a methylation level of the cell-free DNA molecule using the methylated and unmethylated sites of the cell-free DNA molecule, and~~

~~comparing the methylation level to a reference level.~~

~~59. The method of claim 58, further comprising determining the methylation level exceeds the reference level.~~

~~60. The method of claim 56, wherein:~~

~~determining that the cell-free DNA molecule is of fetal origin comprises:~~

~~determining a methylation pattern of a plurality of sites of the cell-free molecule,~~

~~determining a similarity score by comparing the methylation pattern to a reference pattern from a maternal or fetal tissue, and~~

~~comparing the similarity score to one or more threshold values.~~

~~61. The method of claim 48, further comprising:~~

~~receiving a plurality of sequence reads corresponding to the cell-free DNA molecules,~~

~~aligning the plurality of sequence reads to a plurality of regions of the reference genome, the plurality of regions known to potentially include repeats of subsequences,~~

~~identifying numbers of repeats of the subsequences in the plurality of sequence reads;~~

~~comparing the numbers of repeats of the subsequences to a plurality of threshold numbers; and~~

~~for each of a plurality of genetic disorders, determining a classification of a likelihood of the fetus having the respective genetic disorder using the comparison to a threshold number of the plurality of threshold numbers.~~

62. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including cell-free DNA molecules from the fetus and the female, the method comprising:

receiving a first sequence read corresponding to a cell-free DNA molecule of the cell-free DNA molecules;

aligning the first sequence read to a first region of a reference genome;

identifying a first number of repeats of a first subsequence in the first sequence read corresponding to the cell-free DNA molecule;

analyzing sequence data obtained from a male subject to determine whether a second number of repeats of the first subsequence is present in the first region; and

determining a classification of a likelihood of the male subject being the father of the fetus using the determination of whether the second number of repeats of the first subsequence is present.

63. The method of claim 62, further comprising:

determining that the cell-free DNA molecule is of fetal origin.

64. The method of claim 62 or 63, wherein the first subsequence comprises an allele.

65. The method of any one of claims 62 to 64, wherein:

the classification is that the male subject is likely the father when the second number of repeats of the first subsequence is determined to be present, or

the classification is that the male subject is likely not the father when the second number of repeats of the first subsequence is determined to be not present.

66. The method of any one of claims 62 to 65, further comprising:

comparing the first number of repeats with the second number of repeats, wherein:

determining the classification of the likelihood of the male subject being the father comprises:

using the comparison of the first number of repeats with the second number of repeats, and

the classification is that the male subject is likely the father when the first number of repeats is within a threshold value of the second number of repeats.

67. The method of any one of claims 62 to 66, wherein:

the cell-free DNA molecule is a first cell-free DNA molecule;

the method further comprising:

receiving a second sequence read corresponding to a second cell-free DNA molecule of the cell-free DNA molecules;

aligning the second sequence read to a second region of the reference genome;

identifying a first number of repeats of a second subsequence in the second sequence read corresponding to the second cell-free DNA molecule;

analyzing the sequence data obtained from the male subject to determine whether a second number of repeats of the second subsequence is present in the second region;

wherein:

determining the classification of the likelihood of the male subject being the father of the fetus further comprises using the determination of whether the second number of repeats of the second subsequence is present in the second region.

68. The method of any one of claims 62 to 67, wherein the cell-free DNA molecules have a size greater than a cutoff value.

69. The method of claim 68, wherein the cell-free DNA molecules have a size greater than 600 nt.

70. The method of claim 68, wherein the cell-free DNA molecules have a size greater than 1 knt.

71. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

measuring sizes of the plurality of cell-free DNA molecules;

measuring a first amount of cell-free DNA molecules having sizes greater than a cutoff value;

generating a value of a normalized parameter using the first amount;

comparing the value of the normalized parameter to one or more calibration data points, wherein each calibration data point specifies a gestational age corresponding to a calibration value of the normalized parameter, and wherein the one or more calibration data points are determined from a plurality of calibration samples with known gestational ages and including cell-free DNA molecules having sizes greater than the cutoff value; and

determining a gestational age using the comparison.

72. The method of claim 71, further comprising:

determining a reference gestational age of the fetus using an ultrasound or the date of the last menstrual period of the female,

comparing the gestational age to the reference gestational age, and

determining a classification of a likelihood of a pregnancy-associated disorder using the comparison of the gestational age to the reference gestational age.

73. The method of claim 71, further comprising:

determining a first subsequence corresponding to at least one end of the cell-free DNA molecules having sizes greater than the cutoff value,

wherein:

the first amount is of cell-free DNA molecules having a size greater than the cutoff value and having the first subsequence at one or more ends of the respective cell-free DNA molecule.

74. The method of claim 73, wherein the first subsequence is 1, 2, 3, or 4 nucleotides.

75. The method of claim 73 or 74, wherein generating the value of the normalized parameter comprises:

(a) normalizing the first amount by a total amount of cell-free DNA molecules having a size greater than the cutoff value;

(b) normalizing the first amount by a second amount of cell-free DNA molecules having a size greater than the cutoff value and ending on a second subsequence, the second subsequence being different than the first subsequence, or

(c) normalizing the first amount by a third amount of cell-free DNA molecules having a size less than the cutoff value.

76. The method of any one of claims 71 to 75, further comprising receiving sequence reads corresponding to the plurality of cell-free DNA molecules.

77. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

measuring sizes of the plurality of cell-free DNA molecules;

measuring a first amount of cell-free DNA molecules having sizes greater than a cutoff value;

generating a first value of a normalized parameter using the first amount;

obtaining a second value corresponding to an expected value of the normalized parameter for a healthy pregnancy, wherein the second value is dependent on a gestational age of the fetus;

determining a deviation between the first value of the normalized parameter and the second value of the normalized parameter; and

determining a classification of a likelihood of a pregnancy-associated disorder using the deviation.

78. The method of claim 77, wherein obtaining the second value comprises:

obtaining the second value from a calibration table relating measurements of pregnant females with calibration values of the normalized parameter, wherein the calibration table is generated by:

obtaining a first table relating gestational ages with the measurements of pregnant female subjects,

obtaining a second table relating gestational ages with calibration values of the normalized parameter, and

creating the calibration table relating the measurements with the calibration values from the first table and the second table.

79. The method of claim 78, wherein the measurements of the pregnant female subjects are the time since the last menstrual period.

80. The method of claim 78, wherein the measurements of the pregnant female subjects are characteristics of images of the pregnant female subjects.

81. The method of claim 80, wherein characteristics of the image comprise length, size, appearance, or anatomy of a fetus of the female subject.

82. The method of any one of claims 72 to 81, wherein the pregnancy-associated disorder comprises preeclampsia, intrauterine growth restriction, invasive placentation, pre-term birth, hemolytic disease of the newborn, placental insufficiency, hydrops fetalis, fetal malformation, hemolysis, elevated liver enzymes, and a low platelet count (HELLP) syndrome, or systemic lupus erythematosus.

83. The method of any one of claims 71 to 82, wherein the cutoff value is 600 nt or more.

84. The method of any one of claims 71 to 82, wherein the cutoff value is 1,000 nt or more.

85. The method of any one of claims 71 to 84, wherein the first amount is a number or a frequency.

86. The method of any one of claims 71 to 85, wherein generating the value of the normalized parameter using the first amount comprises:

measuring a second amount of cell-free DNA molecules including sizes less than the cutoff value; and

calculating a ratio of the first amount and the second amount.

87. The method of claim 86, wherein:

the cutoff value is a first cutoff value,

a second cutoff value is less than the first cutoff value, and

the second amount comprises cell-free DNA molecules having sizes less than the second cutoff value or the second amount comprises all cell-free DNA molecules in the plurality of cell-free DNA molecules.

88. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

measuring sizes of the plurality of cell-free DNA molecules;

identifying a set of cell-free DNA molecules having sizes greater than a cutoff value;

generating a value of an end motif parameter using a first amount, wherein generating the value of the end motif parameter comprises:

measuring the first amount of cell-free DNA molecules in the set having a first subsequence at one or more ends of the cell-free DNA molecules in the set;

comparing the value of the end motif parameter to a threshold value; and

determining a classification of a likelihood of a pregnancy-associated disorder using the comparison.

89. The method of claim 88, the method further comprising:

measuring a second amount of cell-free DNA molecules having a subsequence different from the first subsequence at one or more ends of the cell-free DNA molecules, and

wherein:

generating the value of the end motif parameter comprises using a ratio of the first amount and the second amount.

90. The method of claim 88, wherein the first subsequence is 1, 2, 3, or 4 nucleotides in length.

91. The method of claim 90, wherein the first subsequence comprises the last nucleotide at the end of the respective cell-free DNA molecule.

92. The method of claim 88, wherein:

the threshold value is a first threshold value, and

the end motif parameter is a first end motif parameter,

the method further comprising:

measuring a second amount of cell-free DNA molecules having a second subsequence different from the first subsequence at one or more ends of the cell-free DNA molecules,

generating a value of a second end motif parameter using the third amount, and

comparing the value of the second end motif parameter to a second threshold value;

wherein:

determining the classification of the likelihood of the pregnancy-associated disorder uses the comparison of the value of the second end motif parameter to the second threshold value, wherein the pregnancy-associated disorder is likely when the value of the first end motif parameter exceeds the first threshold value and the value of the second end motif parameter exceeds the second threshold value.

93. The method of claim 88, wherein the first amount of cell-free DNA molecules comprises cell-free DNA molecules determined to be from a tissue of origin.

94. The method of claim 88, wherein:

the threshold value is a first threshold value, and

the set of cell-free DNA molecules is a first set of cell-free DNA molecules,

the method further comprising:

identifying a second set of cell-free DNA molecules having sizes in a first size range, the first size range including sizes greater than the cutoff value,

generating a value of a size parameter using a second amount of cell-free DNA molecules in the second set, and

comparing the value of the size parameter to a second threshold value,

wherein determining the classification of the likelihood of the pregnancy-associated disorder comprises using the comparison of the value of the size parameter to the second threshold value.

95. The method of any one of claims 88 to 94, wherein the cutoff value is 600 nt.

96. The method of any one of claims 88 to 94, wherein the cutoff value is 1,000 nt.

97. A method of analyzing a biological sample of a pregnant organism, the biological sample including a plurality of cell-free nucleic acid molecules, the method comprising:

sequencing the plurality of cell-free nucleic acid molecules, wherein over 20% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 200 nt.

98. The method of claim 97, wherein sequencing is by a single molecule, real-time technique.

99. The method of claim 97 or 98, wherein:

over 11% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 400 nt,

over 10% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 500 nt,

over 8% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 600 nt,

over 6% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 1 knt,

over 3% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 2 knt,

~~over 1% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 3 knt,~~

~~at least 0.9% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 4 knt, or~~

~~at least 0.04% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 10 knt.~~

~~100. The method of any one of claims 97 to 99, wherein the plurality of cell-free nucleic acid molecules comprise at least 100 cell-free nucleic acid molecules.~~

~~101. The method of any one of claims 97 to 100, wherein the plurality of cell-free nucleic acid molecules are from a plurality of different genomic regions.~~

~~102. The method of any one of claims 97 to 101, wherein the sequencing results in reads that are used in any one of claims 1 to 94.~~

~~103. The method of any one of claims 97 to 101, wherein the sequencing results in reads,~~

~~the method further comprising:~~

~~using the reads to determine a fetal aneuploidy, an aberration, a genetic mutation or variation, or an inheritance of a parental haplotype.~~

~~104. The method of any one of claims 1 to 103, wherein:~~

~~the plurality of cell-free DNA molecules is enriched for sizes greater than or equal to the cutoff value relative to the biological sample, wherein over 20% of the cell-free nucleic acid molecules in the biological sample have sizes greater than 200 nt.~~

~~105. The method of claim 104, further comprising:~~

~~enriching for the plurality of cell-free DNA molecules using electrophoresis.~~

~~106. The method of claim 104, further comprising:~~

~~enriching for the plurality of cell-free DNA molecules using magnetic beads to selectively bind cell-free DNA molecules based on size.~~

~~107. The method of claim 104, further comprising:~~

~~enriching for the plurality of cell-free DNA molecules using hybridization, immunoprecipitation, amplification or CRISPR.~~

~~108. The method of any one of claims 105 to 107, wherein enriching is for sizes greater than 600 nt, 700 nt, 800 nt, 900 nt, or 1 knt.~~

~~109. The method of any one of claims 1 to 103, wherein the plurality of cell-free DNA molecules is enriched for a methylation profile relative to the biological sample, the method further comprising:~~

~~enriching for the plurality of cell-free DNA molecules using immunoprecipitation.~~

~~110. A computer program product comprising instructions that when executed control a computing system to perform the method of any one of the above claims.~~

~~111. A computer readable storage medium comprising the computer program product of claim 110.~~

~~112. A computing system comprising the computer program product of claim 111.~~

~~AMENDED CLAIMS~~

~~Received by the International Bureau on 15 June 2021 (15.06.21)~~

1. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype ~~from~~ using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

2. The method of claim 1, wherein the cutoff value is 600 nt.

3. The method of claim 1, wherein the cutoff value is 1 knt.

4. The method of any one of claims 1 to 3, wherein determining the sequence of the first haplotype and the sequence of the second haplotype from the reads corresponding to the first set of cell-free DNA molecules comprises:

aligning reads corresponding to the first set of cell-free DNA molecules to a reference genome.

5. The method of claim 1, wherein determining the sequence of the first haplotype and the sequence of the second haplotype from the reads corresponding to the first set of cell-free DNA molecules comprises:

aligning a first subset of the reads to a second subset of the reads to identify a different allele at a locus in the reads,

determining that the first subset of the reads have a first allele at the locus,

determining that the second subset of the reads have a second allele at the locus,

determining that the first subset of the reads corresponds to the first haplotype, and

determining that the second subset of the reads corresponds to the second haplotype.

6. The method of any one of claims 1 to 5, wherein the parameter is a count of cell-free DNA molecules, a size profile of cell-free DNA molecules, or a methylation level of cell-free DNA molecules.

7. The method of claim 6, wherein:

the parameter is the count of cell-free DNA molecules, and

the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is greater than the second value.

8. The method of claim 6, wherein:

the parameter is the size profile of cell-free DNA molecules, and

the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is less than the second value, indicating that the second set of cell-free DNA molecules is characterized by a smaller size profile than the third set of cell-free DNA molecules.

9. The method of claim 6, wherein:

the parameter is the methylation level of cell-free DNA molecules, and the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is less than the second value.

10. The method of any one of claims 1 to 9, wherein the cutoff value is a first cutoff value, the method further comprising:

calculating a separation value using the first value and the second value;

comparing the separation value to a second cutoff value; and

determining a likelihood of a fetal aneuploidy based on the comparison of the separation value to the second cutoff value.

11. The method of claim 10, wherein:

the second cutoff value is determined from reference samples from pregnant females with euploid fetuses,

the second cutoff value is determined from reference samples from pregnant females with aneuploid fetuses, or

the second cutoff value is calculated assuming an aneuploid fetus.

12. The method of any one of claims 1 to 11, further comprising:

identifying a number of repeats of a subsequence in a read of the reads corresponding to the first set of cell-free DNA molecules,

wherein:

determining the sequence of the first haplotype comprises determining the number of repeats of the subsequence.

13. The method of claim 12, wherein:

the number of repeats of the subsequence is associated with a repeat-associated disease, and

the method further comprises determining a likelihood of the fetus inheriting the repeat-associated disease.

14. The method of claim 1, wherein:

the plurality of cell-free DNA molecules is enriched for sizes greater than or equal to the cutoff value relative to the biological sample, wherein over 20% of the cell-free DNA molecules in the biological sample have sizes greater than 200 nt.

15. The method of claim 14, further comprising:

enriching for the plurality of cell-free DNA molecules using electrophoresis.

16. The method of claim 14, further comprising:

enriching for the plurality of cell-free DNA molecules using magnetic beads to selectively bind cell-free DNA molecules based on size.

17. The method of claim 14, further comprising:

enriching for the plurality of cell-free DNA molecules using hybridization, immunoprecipitation, amplification, or CRISPR.

18. The method of claim 16, wherein enriching is for sizes greater than 600 nt, 700 nt, 800 nt, 900 nt, or 1 knt.

19. The method of claim 1, wherein the plurality of cell-free DNA molecules is enriched for a methylation profile relative to the biological sample,

the method further comprising:

enriching for the plurality of cell-free DNA molecules using immunoprecipitation.

20. A computer product comprising a non-transitory computer readable medium storing instructions that, when executed, control a computing system to perform a method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the

biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

21. The method of claim 1, wherein the reads corresponding to the plurality of cell-free DNA molecules are obtained by single molecule sequencing.

22. The method of claim 21, wherein the single molecule sequencing comprises optical monitoring of a DNA polymerase incorporating a new base into a complementary strand of a cell-free DNA molecule of the plurality of cell-free DNA molecules.

14. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving sequence reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value; and

for a cell-free DNA molecule of the set of cell-free DNA molecules:

determining a methylation status at each site of a plurality of sites,

determining a methylation pattern, wherein:

the methylation pattern indicates a methylation status at each site of the plurality of sites, using one or more sequence reads corresponding to the cell-free DNA molecule,

comparing the methylation pattern to one or more reference patterns, wherein each of the one or more reference patterns is determined for a particular tissue type; and

determining a tissue of origin of the cell-free DNA molecule using the methylation pattern.

15. The method of claim 14, wherein the cutoff value is 600 nt.

16. The method of claim 14, wherein the cutoff value is 1 knt.

17. The method of any one of claims 14 to 16, further comprising determining the tissue of origin for each cell-free DNA molecule of the set of cell-free DNA molecules by:

determining the methylation status at each site of a plurality of respective sites, wherein the plurality of respective sites corresponds to the cell-free DNA molecule,

determining the methylation pattern, and

comparing the methylation pattern to at least one reference pattern of the one or more reference patterns.

~~18. The method of claim 17, further comprising:~~

~~determining an amount of cell-free DNA molecules corresponding to each tissue of origin, and~~

~~determining a fractional contribution of the tissue of origin in the biological sample using the amount of cell-free DNA molecules corresponding to each tissue of origin.~~

~~19. The method of any one of claims 14 to 18, wherein measuring the sizes of the plurality of cell-free DNA molecules comprises:~~

~~aligning the sequence reads to a reference genome.~~

~~20. The method of any one of claims 14 to 18, wherein measuring sizes of the plurality of cell-free DNA molecules comprises:~~

~~full-length sequencing of the plurality of cell-free DNA molecules, and~~

~~counting the number of nucleotides in each cell-free DNA molecule of the plurality of cell-free DNA molecules.~~

~~21. The method of claim 14 or 17, wherein measuring the sizes of the plurality of cell-free DNA molecules comprises:~~

~~physically separating the plurality of cell-free DNA molecules from the biological sample from other cell-free DNA molecules in the biological sample, wherein the other cell-free DNA molecules have sizes less than the cutoff value.~~

~~22. The method of any one of claims 14 to 21, wherein a reference pattern of the one or more reference patterns is determined by:~~

~~measuring a methylation density at each reference site of a plurality of reference sites using DNA molecules from a reference tissue,~~

~~comparing the methylation density at each reference site of the plurality of reference sites to one or more threshold methylation densities, and~~

~~identifying each reference site of the plurality of reference sites as methylated, unmethylated, or non-informative based on comparing the methylation density to the one or more threshold methylation densities, wherein the plurality of sites is the plurality of reference sites that are identified as methylated or unmethylated.~~

23. The method of any one of claims 14 to 22, wherein the tissue of origin is the placenta.

24. The method of any one of claims 14 to 22, wherein the tissue of origin is fetal or maternal.

25. The method of claim 24, wherein:

the tissue of origin is fetal,

the method further comprising:

aligning a sequence read of the sequence reads to a first region of a reference genome, the first region comprising a plurality of sites corresponding to alleles, the plurality of sites including a threshold number of sites,

determining a first haplotype using the respective allele present at each site of the plurality of sites,

comparing the first haplotype to a second haplotype corresponding to a male subject, and

determining a classification of a likelihood that the male subject being the father of the fetus using the comparison.

26. The method of claim 24, wherein:

the tissue of origin is fetal,

the method further comprising:

aligning a sequence read of the sequence reads to a first region of a reference genome, the first region comprising a first plurality of sites corresponding to alleles, the plurality of sites including a threshold number of sites,

comparing the allele at each site of the plurality of sites to an allele at the corresponding site in the genome of a male subject, and

determining a classification of a likelihood that the male subject being the father of the fetus using the comparison.

27. The method of claim 24, further comprising:

for each cell-free DNA molecule of the set of cell-free DNA molecules:

aligning the sequence read corresponding to the cell-free DNA molecule to a reference genome,

identifying the sequence read as corresponding to a haplotype present in the female,

determining the tissue of origin as fetal using the methylation pattern, and

determining the haplotype to be a maternally inherited fetal haplotype.

28. The method of claim 27, further comprising:

identifying the haplotype as carrying a disease-causing genetic mutation or variation, and

classifying that the fetus is likely to have the disease caused by the genetic mutation or variation.

29. The method of claim 28, wherein identifying the haplotype as carrying the disease-causing genetic mutation comprises:

identifying the genetic mutation or variation in a first sequence read,

measuring a first methylation level in a second sequence read corresponding to a first genomic location within a first distance of the first sequence read, and

measuring a second methylation level in a third sequence read corresponding to a second genomic location within a second distance of the first sequence read, wherein:

the first methylation level and the second methylation level are associated with the genetic mutation.

30. The method of claim 24, further comprising:

for each cell-free DNA molecule of the set of cell-free DNA molecules:

aligning the sequence read corresponding to the cell-free DNA molecule to a reference genome,

identifying the sequence read as corresponding to a region, wherein the region is determined by:

receiving a plurality of fetal sequence reads corresponding to a plurality of fetal DNA molecules from fetal tissue,

~~receiving a plurality of maternal sequence reads corresponding to a plurality of maternal DNA molecules,~~

~~determining a fetal methylation status at each methylation site of a plurality of methylation sites within the region for each fetal sequence read of the plurality of fetal sequence reads,~~

~~determining a maternal methylation status at each methylation site of the plurality of methylation sites for each maternal sequence read of the plurality of maternal sequence reads,~~

~~determining value of a parameter characterizing an amount of sites where the fetal methylation status differs from the maternal methylation status,~~

~~comparing the value of the parameter to a threshold value, and~~

~~determining the value of the parameter exceeds the threshold value.~~

~~31. The method of any one of claims 14 to 28, wherein the cutoff value is at least 500 nt.~~

~~32. The method of any one of claims 14 to 31, wherein determining the tissue of origin of the cell-free DNA molecule comprises inputting the methylation pattern into a machine learning model, the model trained by:~~

~~receiving a plurality of training methylation patterns, each training methylation pattern having a methylation status at one or more sites of the plurality of sites, each training methylation pattern determined from a DNA molecule from a known tissue,~~

~~storing a plurality of training samples, each training sample including one of the plurality of training methylation patterns and a label indicating the known tissue corresponding to the training methylation pattern, and~~

~~optimizing, using the plurality of training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels when the plurality of training methylation patterns is input to the model, wherein an output of the model specifies a tissue corresponding to an input methylation pattern.~~

33. The method of claim 32, wherein the machine learning model comprises convolution neural networks (CNN), linear regression, logistic regression, deep recurrent neural network, Bayes's classifier, hidden Markov model (HMM), linear discriminant analysis (LDA), k-means clustering, density-based spatial clustering of applications with noise (DBSCAN), random forest algorithm, or support vector machine (SVM)

34. The method of claim 32, wherein each DNA molecule from the known tissue is cellular DNA.

35. The method of claim 32 or 34, wherein the parameters of the model comprise a first parameter indicating whether one site of the plurality of sites has the same methylation status as another site of the plurality of sites.

36. The method of any one of claims 32 to 35, wherein the parameters of the model comprise a second parameter indicating a distance between sites of the plurality of sites.

37. The method of any one of claims 14 to 31, wherein a reference pattern of the one or more reference pattern corresponds to a reference tissue,

the method further comprising determining the tissue of origin to be the reference tissue when the methylation pattern matches the reference pattern.

38. The method of any one of claims 14 to 37, wherein the plurality of sites comprise at least 5 CpG sites.

39. The method of any one of claims 14 to 31, wherein determining the tissue of origin using the methylation pattern comprises:

determining a similarity score by comparing the methylation pattern with a first reference methylation pattern from a first reference tissue of a plurality of reference tissues;

comparing the similarity score with a threshold value; and

determining the tissue of origin to be the first reference tissue when the similarity score exceeds the threshold value.

40. The method of claim 39, wherein:

the similarity score is a first similarity score,

the method further comprising:

calculating the threshold value by:

determining a second similarity score by comparing the methylation pattern with a second reference methylation pattern from a second reference tissue of the plurality of reference tissues, the first reference tissue and the second reference tissue being different tissues, the threshold value being the second similarity score.

41. The method of claim 39 or 40, wherein:

the first reference methylation pattern comprises a first subset of sites having at least a first probability of being methylated for the first reference tissue,

the first reference methylation pattern comprises a second subset of sites having at most a second probability of being methylated for the first reference tissue, and

determining the similarity score comprises:

increasing the similarity score when a site of the plurality of sites is methylated and the site of the plurality of sites is in the first subset of sites, and

decreasing the similarity score when a site of the plurality of sites is methylated and the site of the plurality of sites is in the second subset of sites.

42. The method of claim 39 or 40, wherein:

the first reference methylation pattern comprises the plurality of sites, with each site of the plurality of sites characterized by a probability of being methylated and a probability of being unmethylated for the first reference tissue,

the similarity score is determined by:

for each site of the plurality of sites:

determining the probability in the reference tissue corresponding to the methylation status of the site in the cell-free DNA molecule,

calculating a product of the plurality of probabilities, the product being the similarity score.

~~43. The method of claim 42, wherein the probability is determined using a beta distribution.~~

~~44. The method of any one of claims 14 to 43, further comprising:~~

~~sequencing the plurality of cell-free DNA molecules to obtain the sequence reads, and~~

~~determining a methylation status of the site by measuring a characteristic corresponding to a nucleotide of the site and nucleotides neighboring the site.~~

~~45. The method of any one of claims 14 to 44, wherein sizes of the plurality of cell-free DNA molecules comprise a number of CpG sites.~~

~~46. The method of any one of claims 14 to 45, wherein at least one site of the plurality of sites is methylated.~~

~~47. The method of any one of claims 14 to 46, wherein two sites of the plurality of sites are separated by at least 160 nt.~~

~~48. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including cell-free DNA molecules from the fetus and the female, the method comprising:~~

~~receiving a first sequence read corresponding to a cell-free DNA molecule of the cell-free DNA molecules;~~

~~aligning the first sequence read to a region of a reference genome, the region known to potentially include repeats of a subsequence;~~

~~identifying a number of repeats of the subsequence in the first sequence read corresponding to the cell-free DNA molecule;~~

~~comparing the number of repeats of the subsequence to a threshold number; and~~

~~determining a classification of a likelihood of the fetus having a genetic disorder using the comparison of the number of repeats to the threshold number.~~

~~49. The method of claim 48, wherein determining the classification of the likelihood of the fetus having the genetic disorder comprises:~~

~~determining that the fetus is likely to have the genetic disorder when the number of repeats exceeds the threshold number.~~

~~50. The method of claim 48 or 49, wherein the threshold number is 55 or more.~~

~~51. The method of any one of claims 48 to 50, wherein the genetic disorder is fragile X syndrome.~~

~~52. The method of any one of claims 48 to 51, wherein the subsequence is a trinucleotide sequence.~~

~~53. The method of any one of claims 48 to 52, wherein the cell-free DNA molecules have a length greater than a cutoff value.~~

~~54. The method of claim 53, wherein the cutoff value is 600 nt.~~

~~55. The method of claim 53, wherein the cutoff value is 1 knt.~~

~~56. The method of any one of claims 48 to 54, further comprising determining that the cell-free DNA molecule is of fetal origin.~~

~~57. The method of claim 56, wherein:~~

~~the number of repeats of the subsequence in the first sequence read is a first number of repeats of the subsequence,~~

~~determining that the cell-free DNA molecule is of fetal origin comprises:~~

~~receiving a second sequence read corresponding to a cell-free DNA molecule of maternal origin obtained from a buffy coat or a sample of the female before pregnancy,~~

~~aligning the second sequence read to the region of the reference genome,~~

~~identifying a second number of repeats of the subsequence in the second sequence read, and~~

~~determining that the second number of repeats is less than the first number of repeats.~~

~~58. The method of claim 56, wherein:~~

~~determining that the cell-free DNA molecule is of fetal origin comprises:~~

~~determining a methylation level of the cell-free DNA molecule using the methylated and unmethylated sites of the cell-free DNA molecule, and~~

~~comparing the methylation level to a reference level.~~

~~59. The method of claim 58, further comprising determining the methylation level exceeds the reference level.~~

~~60. The method of claim 56, wherein:~~

~~determining that the cell-free DNA molecule is of fetal origin comprises:~~

~~determining a methylation pattern of a plurality of sites of the cell-free molecule,~~

~~determining a similarity score by comparing the methylation pattern to a reference pattern from a maternal or fetal tissue, and~~

~~comparing the similarity score to one or more threshold values.~~

~~61. The method of claim 48, further comprising:~~

~~receiving a plurality of sequence reads corresponding to the cell-free DNA molecules,~~

~~aligning the plurality of sequence reads to a plurality of regions of the reference genome, the plurality of regions known to potentially include repeats of subsequences,~~

~~identifying numbers of repeats of the subsequences in the plurality of sequence reads;~~

~~comparing the numbers of repeats of the subsequences to a plurality of threshold numbers; and~~

~~for each of a plurality of genetic disorders, determining a classification of a likelihood of the fetus having the respective genetic disorder using the comparison to a threshold number of the plurality of threshold numbers.~~

62. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including cell-free DNA molecules from the fetus and the female, the method comprising:

receiving a first sequence read corresponding to a cell-free DNA molecule of the cell-free DNA molecules;

aligning the first sequence read to a first region of a reference genome;

identifying a first number of repeats of a first subsequence in the first sequence read corresponding to the cell-free DNA molecule;

analyzing sequence data obtained from a male subject to determine whether a second number of repeats of the first subsequence is present in the first region; and

determining a classification of a likelihood of the male subject being the father of the fetus using the determination of whether the second number of repeats of the first subsequence is present.

63. The method of claim 62, further comprising:

determining that the cell-free DNA molecule is of fetal origin.

64. The method of claim 62 or 63, wherein the first subsequence comprises an allele.

65. The method of any one of claims 62 to 64, wherein:

the classification is that the male subject is likely the father when the second number of repeats of the first subsequence is determined to be present, or

the classification is that the male subject is likely not the father when the second number of repeats of the first subsequence is determined to be not present.

66. The method of any one of claims 62 to 65, further comprising:

comparing the first number of repeats with the second number of repeats, wherein:

determining the classification of the likelihood of the male subject being the father comprises:

using the comparison of the first number of repeats with the second number of repeats, and

the classification is that the male subject is likely the father when the first number of repeats is within a threshold value of the second number of repeats.

67. The method of any one of claims 62 to 66, wherein:

the cell-free DNA molecule is a first cell-free DNA molecule;

the method further comprising:

receiving a second sequence read corresponding to a second cell-free DNA molecule of the cell-free DNA molecules;

aligning the second sequence read to a second region of the reference genome;

identifying a first number of repeats of a second subsequence in the second sequence read corresponding to the second cell-free DNA molecule;

analyzing the sequence data obtained from the male subject to determine whether a second number of repeats of the second subsequence is present in the second region;

wherein:

determining the classification of the likelihood of the male subject being the father of the fetus further comprises using the determination of whether the second number of repeats of the second subsequence is present in the second region.

68. The method of any one of claims 62 to 67, wherein the cell-free DNA molecules have a size greater than a cutoff value.

69. The method of claim 68, wherein the cell-free DNA molecules have a size greater than 600 nt.

70. The method of claim 68, wherein the cell-free DNA molecules have a size greater than 1 knt.

71. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

measuring sizes of the plurality of cell-free DNA molecules;

measuring a first amount of cell-free DNA molecules having sizes greater than a cutoff value;

generating a value of a normalized parameter using the first amount;

comparing the value of the normalized parameter to one or more calibration data points, wherein each calibration data point specifies a gestational age corresponding to a calibration value of the normalized parameter, and wherein the one or more calibration data points are determined from a plurality of calibration samples with known gestational ages and including cell-free DNA molecules having sizes greater than the cutoff value; and

determining a gestational age using the comparison.

72. The method of claim 71, further comprising:

determining a reference gestational age of the fetus using an ultrasound or the date of the last menstrual period of the female,

comparing the gestational age to the reference gestational age, and

determining a classification of a likelihood of a pregnancy-associated disorder using the comparison of the gestational age to the reference gestational age.

73. The method of claim 71, further comprising:

determining a first subsequence corresponding to at least one end of the cell-free DNA molecules having sizes greater than the cutoff value,

wherein:

the first amount is of cell-free DNA molecules having a size greater than the cutoff value and having the first subsequence at one or more ends of the respective cell-free DNA molecule.

74. The method of claim 73, wherein the first subsequence is 1, 2, 3, or 4 nucleotides.

75. The method of claim 73 or 74, wherein generating the value of the normalized parameter comprises:

(a) normalizing the first amount by a total amount of cell-free DNA molecules having a size greater than the cutoff value;

(b) normalizing the first amount by a second amount of cell-free DNA molecules having a size greater than the cutoff value and ending on a second subsequence, the second subsequence being different than the first subsequence, or

(c) normalizing the first amount by a third amount of cell-free DNA molecules having a size less than the cutoff value.

76. The method of any one of claims 71 to 75, further comprising receiving sequence reads corresponding to the plurality of cell-free DNA molecules.

77. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

measuring sizes of the plurality of cell-free DNA molecules;

measuring a first amount of cell-free DNA molecules having sizes greater than a cutoff value;

generating a first value of a normalized parameter using the first amount;

obtaining a second value corresponding to an expected value of the normalized parameter for a healthy pregnancy, wherein the second value is dependent on a gestational age of the fetus;

determining a deviation between the first value of the normalized parameter and the second value of the normalized parameter; and

determining a classification of a likelihood of a pregnancy-associated disorder using the deviation.

78. The method of claim 77, wherein obtaining the second value comprises:

obtaining the second value from a calibration table relating measurements of pregnant females with calibration values of the normalized parameter, wherein the calibration table is generated by:

obtaining a first table relating gestational ages with the measurements of pregnant female subjects,

obtaining a second table relating gestational ages with calibration values of the normalized parameter, and

creating the calibration table relating the measurements with the calibration values from the first table and the second table.

79. The method of claim 78, wherein the measurements of the pregnant female subjects are the time since the last menstrual period.

80. The method of claim 78, wherein the measurements of the pregnant female subjects are characteristics of images of the pregnant female subjects.

81. The method of claim 80, wherein characteristics of the image comprise length, size, appearance, or anatomy of a fetus of the female subject.

82. The method of any one of claims 72 to 81, wherein the pregnancy-associated disorder comprises preeclampsia, intrauterine growth restriction, invasive placentation, pre-term birth, hemolytic disease of the newborn, placental insufficiency, hydrops fetalis, fetal malformation, hemolysis, elevated liver enzymes, and a low platelet count (HELLP) syndrome, or systemic lupus erythematosus.

83. The method of any one of claims 71 to 82, wherein the cutoff value is 600 nt or more.

84. The method of any one of claims 71 to 82, wherein the cutoff value is 1,000 nt or more.

85. The method of any one of claims 71 to 84, wherein the first amount is a number or a frequency.

86. The method of any one of claims 71 to 85, wherein generating the value of the normalized parameter using the first amount comprises:

measuring a second amount of cell-free DNA molecules including sizes less than the cutoff value; and

calculating a ratio of the first amount and the second amount.

87. The method of claim 86, wherein:

the cutoff value is a first cutoff value,

a second cutoff value is less than the first cutoff value, and

the second amount comprises cell-free DNA molecules having sizes less than the second cutoff value or the second amount comprises all cell-free DNA molecules in the plurality of cell-free DNA molecules.

88. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

measuring sizes of the plurality of cell-free DNA molecules;

identifying a set of cell-free DNA molecules having sizes greater than a cutoff value;

generating a value of an end motif parameter using a first amount, wherein generating the value of the end motif parameter comprises:

measuring the first amount of cell-free DNA molecules in the set having a first subsequence at one or more ends of the cell-free DNA molecules in the set;

comparing the value of the end motif parameter to a threshold value; and

determining a classification of a likelihood of a pregnancy-associated disorder using the comparison.

89. The method of claim 88, the method further comprising:

measuring a second amount of cell-free DNA molecules having a subsequence different from the first subsequence at one or more ends of the cell-free DNA molecules, and

wherein:

generating the value of the end motif parameter comprises using a ratio of the first amount and the second amount.

90. The method of claim 88, wherein the first subsequence is 1, 2, 3, or 4 nucleotides in length.

91. The method of claim 90, wherein the first subsequence comprises the last nucleotide at the end of the respective cell-free DNA molecule.

92. The method of claim 88, wherein:

the threshold value is a first threshold value, and

the end motif parameter is a first end motif parameter,

the method further comprising:

measuring a second amount of cell-free DNA molecules having a second subsequence different from the first subsequence at one or more ends of the cell-free DNA molecules,

generating a value of a second end motif parameter using the third amount, and

comparing the value of the second end motif parameter to a second threshold value;

wherein:

determining the classification of the likelihood of the pregnancy-associated disorder uses the comparison of the value of the second end motif parameter to the second threshold value, wherein the pregnancy-associated disorder is likely when the value of the first end motif parameter exceeds the first threshold value and the value of the second end motif parameter exceeds the second threshold value.

93. The method of claim 88, wherein the first amount of cell-free DNA molecules comprises cell-free DNA molecules determined to be from a tissue of origin.

94. The method of claim 88, wherein:

the threshold value is a first threshold value, and

the set of cell-free DNA molecules is a first set of cell-free DNA molecules,

the method further comprising:

identifying a second set of cell-free DNA molecules having sizes in a first size range, the first size range including sizes greater than the cutoff value,

generating a value of a size parameter using a second amount of cell-free DNA molecules in the second set, and

comparing the value of the size parameter to a second threshold value,

wherein determining the classification of the likelihood of the pregnancy-associated disorder comprises using the comparison of the value of the size parameter to the second threshold value.

95. The method of any one of claims 88 to 94, wherein the cutoff value is 600 nt.

96. The method of any one of claims 88 to 94, wherein the cutoff value is 1,000 nt.

97. A method of analyzing a biological sample of a pregnant organism, the biological sample including a plurality of cell-free nucleic acid molecules, the method comprising:

sequencing the plurality of cell-free nucleic acid molecules, wherein over 20% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 200 nt.

98. The method of claim 97, wherein sequencing is by a single molecule, real-time technique.

99. The method of claim 97 or 98, wherein:

over 11% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 400 nt,

over 10% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 500 nt,

over 8% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 600 nt,

over 6% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 1 knt,

over 3% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 2 knt,

~~over 1% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 3 knt,~~

~~at least 0.9% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 4 knt, or~~

~~at least 0.04% of the plurality of the cell-free nucleic acid molecules sequenced have lengths greater than 10 knt.~~

~~100. The method of any one of claims 97 to 99, wherein the plurality of cell-free nucleic acid molecules comprise at least 100 cell-free nucleic acid molecules.~~

~~101. The method of any one of claims 97 to 100, wherein the plurality of cell-free nucleic acid molecules are from a plurality of different genomic regions.~~

~~102. The method of any one of claims 97 to 101, wherein the sequencing results in reads that are used in any one of claims 1 to 94.~~

~~103. The method of any one of claims 97 to 101, wherein the sequencing results in reads,~~

~~the method further comprising:~~

~~using the reads to determine a fetal aneuploidy, an aberration, a genetic mutation or variation, or an inheritance of a parental haplotype.~~

~~104. The method of any one of claims 1 to 103, wherein:~~

~~the plurality of cell-free DNA molecules is enriched for sizes greater than or equal to the cutoff value relative to the biological sample, wherein over 20% of the cell-free nucleic acid molecules in the biological sample have sizes greater than 200 nt.~~

~~105. The method of claim 104, further comprising:~~

~~enriching for the plurality of cell-free DNA molecules using electrophoresis.~~

~~106. The method of claim 104, further comprising:~~

~~enriching for the plurality of cell-free DNA molecules using magnetic beads to selectively bind cell-free DNA molecules based on size.~~

~~107. The method of claim 104, further comprising:~~

enriching for the plurality of cell-free DNA molecules using hybridization, immunoprecipitation, amplification or CRISPR.

108. The method of any one of claims 105 to 107, wherein enriching is for sizes greater than 600 nt, 700 nt, 800 nt, 900 nt, or 1 knt.

109. The method of any one of claims 1 to 103, wherein the plurality of cell-free DNA molecules is enriched for a methylation profile relative to the biological sample,

the method further comprising:

enriching for the plurality of cell-free DNA molecules using immunoprecipitation.

110. A computer program product comprising instructions that when executed control a computing system to perform the method of any one of the above claims.

111. A computer readable storage medium comprising the computer program product of claim 110.

112. A computing system comprising the computer program product of claim 111.