CLAIMS

1. A method for measuring methylation of a CpG site in a genome of a subject using cell-free DNA molecules, the method comprising:

analyzing a plurality of cell-free DNA molecules from a biological sample of the 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;

determining a first amount of cell-free DNA molecules ending at a first position within a window around the CpG site, the first position being between -1 to +1 position of the window; and

determining, using the first amount, a classification for methylation of the CpG site in the genome of the subject.

2. The method of claim 1, further comprising:

determining a second amount of cell-free DNA molecules ending at a second position within the window around the CpG site, the second position being different than the first position, wherein the classification is determined using the first amount and the second amount.

3. The method of claim 2, wherein the first position is the 0 position, and wherein the second position is at +1 or -1 from the CpG site.

4. The method of claim 2, wherein the window is at least -2 to +2 from the CpG site.

5. The method of claim 2, wherein determining the classification includes:

determining a separation value using the first amount and the second amount; and

comparing the separation value to a calibration value, wherein the calibration value is determined using cell-free DNA molecules that are from one or more calibration samples and that are located at CpG sites having known classifications.

6. The method of claim 5, wherein the classification is a quantitative value, and wherein comparing the separation value to the calibration value includes comparing the separation value to a calibration function.

7. The method of claim 6, wherein the quantitative value is a range that is 30% or less.

8. The method of claim 2, wherein the CpG site is a first CpG site and the classification is a first classification, and wherein the window includes a second CpG site and at least two positions other than the first CpG site and the second CpG site, and the method further comprising:

determining a third amount of cell-free DNA molecules ending at the second CpG site;

for each position of the at least two positions within the window:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts including the second amount, wherein the at least two positions include the first position;

generating a feature vector including the respective amounts, the first amount, and the third amount; and

inputting the feature vector into a machine learning model as part of determining the classification for the first CpG site and determining a second classification of the second CpG site, wherein the machine learning model is trained using cell-free DNA molecules located within a window around CpG sites having known classifications.

9. The method of claim 1, wherein the window includes at least two positions other than the first position, the method further comprising:

for each position of the at least two positions within the window:

determining a respective amount of cell-free DNA molecules ending at the position, wherein the at least two positions include the first position; and

comparing the first amount of cell-free DNA molecules ending at the CpG site to the respective amount of cell-free DNA molecules ending at the position as part of determining the classification.

10. The method of claim 1, wherein the window includes at least two positions other than the first position, the method further comprising:

for each position of the at least two positions within the window:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts;

generating a feature vector including the respective amounts and the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification, wherein the machine learning model is trained using cell-free DNA molecules located within a window around CpG sites having known classifications.

11. The method of claim 10, wherein the feature vector forms a matrix with each row corresponding to a base of a strand, and wherein a column includes a non-zero amount in the row corresponding to the base at a respective position.

12. The method of claim 10, wherein the machine learning model is a convolutional neural network.

13. The method of claim 10, wherein the first position and the at least two positions include all positions within the window, the window being at least +4 to -4 from the CpG site.

14. The method of claim 10, wherein the machine learning model uses a sequence context within the window.

15. The method of claim 14, wherein the machine learning model is trained for the sequence context within the window.

16. The method of claim 14, wherein the feature vector includes the sequence context.

17. The method of claim 1, wherein the classification for methylation indicates that the CpG site is in a hypermethylated state or a hypomethylated state for the cell-free DNA molecules at the CpG site, wherein the hypermethylated state indicates a methylation

density above a first threshold that is at least 70%, and wherein the hypomethylated state indicates the methylation density below a second threshold that is 30% or less.

18. The method of claim 1, wherein the first amount is normalized.

19. The method of claim 18, wherein the normalization uses a number of cell-free DNA molecules ending within a region including the CpG site.

20. The method of claim 18, wherein the normalization uses a number of cell-free DNA molecules covering the CpG site.

21. The method of claim 18, wherein the normalization uses an average or median depth of cell-free DNA molecules in a region including the CpG site.

22. The method of claim 1 or claim 2, further comprising:

determining, using the first amount and optionally the second amount, another classification for methylation of a different CpG site in the genome of the subject, the different CpG site being within 600 nt downstream from the 5' end of the CpG site.

23. A method for measuring methylation of a region in a genome of a subject, the method comprising:

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

determining an end sequence motif of at least one end of the cell-free DNA molecule, wherein an end of the cell-free DNA molecule has a first position at an outermost position, a second position that is next to the first position, and a third position that is next to the second position, wherein the plurality of cell-free DNA molecules are located in the region of a reference genome;

determining a first amount of a first set of one or more end sequence motifs, wherein:

the first set of one or more end sequence motifs have C at the first position and G at the second position, or

the first set of one or more end sequence motifs have C at the second position and G at the third position; and

determining a classification for methylation of the region in the genome of the subject by comparing the first amount to a calibration value, wherein the calibration value is determined using cell-free DNA molecules that are from one or more calibration samples and that are located at CpG sites having known classifications.

24. The method of claim 23, wherein the region is a portion of the genome of the subject, and wherein the plurality of cell-free DNA molecules are from one or more particular regions of the reference genome.

25. The method of claim 24, wherein the region includes a sequence variant or a copy number aberration determined from the plurality of cell-free DNA molecules, the method further comprising:

determining whether the sequence variant or the copy number aberration is from a tumor based on the classification for methylation of the region.

26. The method of claim 24, wherein at least one of the one or more particular regions corresponds to a particular genotype or haplotype comprising one or more alleles.

27. The method of claim 26, wherein the classification for methylation for the region is a first methylation level corresponding to a first haplotype of a female pregnant with a fetus, the method further comprising:

determining a second methylation level corresponding to a second haplotype of the female using a second amount of the first set of one or more end sequence motifs for cell-free DNA molecules from the second haplotype; and

determining an inherited haplotype of the fetus based on the first methylation level and the second methylation level.

28. The method of claim 26, further comprising:

determining a level of a pathology for the subject in the one or more particular regions based on the classification.

29. The method of claim 28, wherein the level of the pathology includes an abnormal level of a physiological process or a copy number aberration.

30. The method of claim 23, wherein the first set of one or more end sequence motifs have C at the first position and G at the second position, the method further comprising:

determining a second amount of a second set of one or more end sequence motifs that have C at the second position and G at the third position,

wherein determining the classification includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the calibration value.

31. The method of claim 30, wherein the first set of one or more end sequence motifs includes CGA, CGC, CGG, and CGT, and wherein the second set of one or more end sequence motifs includes ACG, CCG, GCG, and TCG.

32. The method of claim 23, further comprising:

determining an expected amount of the first set of one or more end sequence motifs based on a reference sequence of the reference genome in the region,

wherein determining the classification includes normalizing the first amount with the expected amount to obtain a normalized first amount that is compared to the calibration value.

33. The method of claim 23, further comprising:

determining a total amount of sequence motifs for the plurality of cell-free DNA molecules,

wherein determining the classification includes normalizing the first amount with the total amount to obtain a normalized first amount that is compared to the calibration value.

34. The method of claim 23, wherein the region is the reference genome.

35. The method of claim 23, wherein the region is 10 Mb or less of a particular chromosome.

36. The method of claim 23, wherein the first set of one or more end sequence motifs have C at the first position and G at the second position, the method further comprising:

determining a respective amount of each 3-mer end sequence motif that has a C at the first position and that has G at the second position, thereby determining respective amounts;

generating a feature vector including the respective amounts, which include the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification for methylation of the region in the genome of the subject, wherein the machine learning model is trained using cell-free DNA molecules that are from the one or more calibration samples.

37. The method of claim 36, wherein the feature vector includes respective amounts of all 3-mer end sequence motifs.

38. The method of claim 23, wherein analyzing a cell-free DNA molecule includes:

determining a location of the cell-free DNA molecule in the reference genome.

39. A method of analyzing a biological sample to determine a level of a pathology in the biological sample of a subject, the biological sample including cell-free DNA, the method comprising:

analyzing a plurality of cell-free DNA molecules from the biological sample to obtain sequence reads, wherein the sequence reads include ending sequences corresponding to ends of the plurality of cell-free DNA molecules;

for each of the plurality of cell-free DNA molecules, determining a sequence motif for each of one or more ends of the cell-free DNA molecule, wherein an end of a cell-free DNA molecule has a first position at an outermost position, a second position that is next to the first position, and a third position that is next to the second position;

determining a first amount of a first set of one or more end sequence motifs of the plurality of cell-free DNA molecules, wherein:

the first set of one or more end sequence motifs have C at the first position and G at the second position, or

the first set of one or more end sequence motifs have C at the second position and G at the third position; and

determining a classification of the level of the pathology for the subject based on a comparison of the first amount to a reference value, wherein the classification is determined using one or more end sequence motifs only selected from a group consisting of (1) end sequence motifs having C at the first position and G at the second position and (2) end sequence motifs having C at the second position and G at the third position.

40. The method of claim 39, wherein analyzing the plurality of cell-free DNA molecules includes measuring a size of each of the plurality of cell-free DNA molecules, and wherein the first amount is determined for a first group of cell-free DNA molecules that are within a first size range.

41. The method of claim 40, wherein the first size range corresponds to sizes that are less than a size cutoff.

42. The method of claim 40, wherein the first size range corresponds to sizes that are greater than a first size cutoff.

43. The method of claim 42, further comprising:

determining a second amount of a second set of one or more end sequence motifs of the plurality of cell-free DNA molecules, wherein:

the second set of one or more end sequence motifs have C at the first position and G at the second position, or

the second set of one or more end sequence motifs have C at the second position and G at the third position,

wherein the second amount is determined for a second group of cell-free DNA molecules that are within a second size range corresponding to sizes that are less than second size cutoff, and wherein the classification is determined using the first amount and the second amount.

44. The method of claim 43, wherein the first amount is determined using the first set of one or more end sequence motifs having C at the first position and G at the second position, and wherein the second amount is determined using the second set of one or more end

sequence motifs having C at the first position and G at the second position, the method further comprising:

determining a third amount of the first group of cell-free DNA molecules having a third set of end sequence motifs with C at the second position and G at the third position, and

determining a fourth amount of the second group of cell-free DNA molecules having a fourth set of end sequence motifs with C at the second position and G at the third position, wherein the classification is further determined using the third amount and the fourth amount.

45. The method of claim 43, wherein the classification uses a ratio of the first amount and the second amount, uses a difference of the first amount and the second amount, or uses a machine learning model that receives the first amount and the second amount as separate inputs.

46. The method of claim 39, wherein the first amount is determined for a first group of cell-free DNA molecules that are each located within one or more regions that are each hypermethylated or hypomethylated for a particular tissue type, and wherein the pathology is of the particular tissue type.

47. The method of claim 39, wherein the plurality of cell-free DNA molecules are from a first region of a plurality of regions of a reference genome, and wherein the comparison of the first amount to the reference value indicates whether the first amount for the first region is aberrant, the method further comprising:

determining a number of the plurality of regions that are aberrant; and

comparing the number of aberrant regions to a threshold to determine the classification of the level of the pathology.

48. The method of claim 47, wherein, for each of the plurality of regions, a respective amount of the first set of one or more end sequence motifs is compared to a region-specific reference value.

49. The method of claim 39, wherein the plurality of cell-free DNA molecules are from one or more particular regions of a reference genome, and wherein the pathology is in the one or more particular regions.

50. The method of claim 49, wherein at least one of the one or more particular regions corresponds to a particular genotype or haplotype comprising one or more alleles.

51. The method of claim 49, wherein the level of the pathology includes a copy number aberration.

52. The method of claim 39, wherein the first set of one or more end sequence motifs includes CGN, where N includes any nucleotide at the third position.

53. The method of claim 39, wherein the first set of one or more end sequence motifs have C at the first position and G at the second position, the method further comprising:

determining a second amount of a second set of end sequence motifs that have C at the second position and G at the third position,

wherein determining the classification includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the reference value.

54. The method of claim 53, wherein the second set of end sequence motifs includes NCG, where N includes any nucleotide at the first position.

55. The method of claim 39, further comprising:

determining a total amount of sequence motifs for the plurality of cell-free DNA molecules,

wherein determining the classification includes normalizing the first amount with the total amount to obtain a normalized first amount that is compared to the reference value.

56. The method of claim 39, wherein the pathology is a cancer.

57. The method of claim 39, wherein the pathology is an abnormal level of a physiological process, wherein the physiological process is pregnancy, ageing, development, or growth.

58. The method of claim 39, wherein the first set of one or more end sequence motifs have C at the first position and G at the second position, the method further comprising:

determining a respective amount of each 3-mer end sequence motif that has a C at the first position and that has G at the second position, thereby determining respective amounts;

generating a feature vector including the respective amounts, which include the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology for the subject, wherein the machine learning model is trained using cell-free DNA molecules in training samples having known classifications.

59. The method of claim 58, further comprising:

determining a respective additional amount of each 3-mer end sequence motif that has a C at the second position and that has G at the third position, thereby determining respective additional amounts, wherein the feature vector includes the respective additional amounts.

60. The method of claim 39, wherein the first set of one or more end sequence motifs have C at the second position and G at the third position, the method further comprising:

determining a respective amount of each 3-mer end sequence motif that has a C at the second position and that has G at the third position, thereby determining respective amounts;

generating a feature vector including the respective amounts, which include the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology for the subject, wherein the machine learning model is trained using cell-free DNA molecules in training samples having known classifications.

61. The method of claim 39, wherein the first set of one or more end sequence motifs is a plurality of end sequence motifs.

62. A method of analyzing a biological sample to determine a level of a pathology in the biological sample of a subject, the biological sample including cell-free DNA, the method comprising:

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

determining a location of the cell-free DNA molecule in a reference genome, wherein the plurality of cell-free DNA molecules are located in a first set of one or more regions of the reference genome, and wherein each of the first set of one or more regions is hypomethylated or each is hypermethylated in a first tissue type; and

determining an end sequence motif of at least one end of the cell-free DNA molecule, wherein an end of the cell-free DNA molecule has a first position at an outermost position, a second position that is next to the first position, and a third position that is next to the second position;

(b) determining a first amount of a first set of one or more end sequence motifs, wherein:

the first set of one or more end sequence motifs have C at the first position and G at the second position, or

the first set of one or more end sequence motifs have C at the second position and G at the third position; and

(c) determining a classification of the level of the pathology in the first tissue type for the subject based on a comparison of the first amount to a reference value.

63. The method of claim 62, wherein the first tissue type is a cancer tissue type.

64. The method of claim 63, wherein the biological sample is urine, and wherein the cancer tissued type is selected from bladder cancer, kidney cancer, and prostate cancer.

65. The method of claim 63, wherein the biological sample is plasma or serum, and wherein the cancer tissued type is selected from liver cancer, colon cancer, lung cancer, and breast cancer.

66. The method of claim 62, wherein the classification uses differentially-methylated regions for a plurality of tissue types, including the first tissue type, the method further comprising:

analyzing other sets of cell-free DNA molecules from the biological sample, wherein each set of cell-free DNA molecules are located in a respective set of one or more regions of the reference genome, each of the respective set of one or more regions being hypomethylated or hypermethylated in a respective tissue type of the plurality of tissue types.

67. The method of claim 66, wherein a particular set of one or more regions includes a plurality of regions of a plurality of region types, wherein a first region type is differentially-methylated relative to other tissues of the plurality of tissue types, and wherein a second region type is differentially-methylated relative to at least one other tissue of the plurality of tissue types but not all of the other tissues.

68. The method of claim 67, wherein another particular set of one or more regions are all hypomethylated or all hypermethylated in a cancer tissue type of a first tissue and a healthy first tissue.

69. The method of claim 66, wherein the classification of the level of the pathology uses a first machine learning model, and wherein amounts of the first set of one or more end sequence motifs are used to train one or more other machine learning models using training samples having the pathology in one or more other tissue types of the plurality of tissue types.

70. The method of claim 66, wherein the classification of the level of the pathology uses a multiclass machine learning model that provides a probability of each of the plurality of tissue types having the pathology.

71. The method of claim 66, wherein the plurality of tissue types includes cancer tissue types and non-cancer tissue types.

72. The method of claim 62, further comprising:

identifying a second set of CpG sites that are all hypomethylated or all hypermethylated across a plurality of tissue types in the reference genome; and

determining a second amount of cell-free DNA molecules ending around any one of the second set of CpG sites,

wherein determining the classification of the level of the pathology includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the reference value.

73. The method of claim 62, wherein the first set of one or more end sequence motifs define a first set of one or more CpG sites within the first set of one or more regions, the method further comprising:

for each position of at least two positions within a window around any one of the first set of one or more CpG sites:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts;

generating a feature vector including the respective amounts and the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology in the first tissue type for the subject, wherein the machine learning model is trained using cell-free DNA molecules from training samples having known classifications.

74. The method of claim 73, wherein the first set of one or more CpG sites are all hypomethylated for the first tissue type, and wherein the feature vector is a first feature vector, the method further comprising:

generating a second feature vector using a second set of one or more CpG sites that are all hypermethylated; and

inputting the second feature vector and the first feature vector into the machine learning model as part of determining the classification of the level of the pathology in the first tissue type for the subject.

75. The method of claim 62, wherein the first set of one or more end sequence motifs define a first set of CpG sites within the first set of one or more regions, and wherein determining the first amount includes:

determining respective amounts of cell-free DNA molecules ending at each of the first set of CpG sites, wherein determining classification of the level of the pathology includes comparing each of the respective amounts to a respective reference value.

76. The method of claim 75, wherein comparing each of the respective amounts to the respective reference value is performed by inputting the respective amounts as part of a feature vector into a machine learning model.

77. The method of claim 76, further comprising, for each CpG site of the first set of one or more CpG sites:

for each position of at least two positions within a window around the CpG site:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts; and

including the respective amounts for the at least two positions and the respective amount for the CpG site in the feature vector.

78. The method of claim 62, wherein the first set of one or more regions is a plurality of regions.

79. The method of claim 62, wherein the first set of one or more end sequence motifs includes CGN, where N includes any nucleotide at the third position.

80. The method of claim 62, further comprising:

determining a second amount of a second set of end sequence motifs that have C at the second position and G at the third position,

wherein determining the classification includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the reference value.

81. The method of claim 80, wherein the second set of end sequence motifs includes NCG, where N includes any nucleotide at the first position.

82. The method of claim 62, further comprising:

determining a total amount of sequence motifs for the plurality of cell-free DNA molecules,

wherein determining the classification includes normalizing the first amount with the total amount to obtain a normalized first amount that is compared to the reference value.

83. The method of claim 62, wherein the pathology is a cancer.

84. The method of claim 62, wherein the pathology is an abnormal level of a physiological process, wherein the physiological process is pregnancy, ageing, development, or growth.

85. The method of claim 62, further comprising:

repeating (a), (b), and (c) for one or more additional sets of one or more regions of the reference genome, and wherein the one or more additional sets are all hypomethylated or all hypermethylated in one or more other tissue types, thereby determining a type of a cancer for the subject.

86. The method of claim 62, wherein the first set of one or more end sequence motifs includes all sequence motifs having C at the first position and G at the second position, the method further comprising:

determining a respective amount of each 3-mer end sequence motif that has a C at the first position and that has G at the second position, thereby determining respective amounts;

generating a feature vector including the respective amounts, which include the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology in the first tissue type for the subject, wherein the machine learning model is trained using cell-free DNA molecules in training samples having known classifications.

87. The method of claim 86, further comprising:

determining a respective additional amount of each 3-mer end sequence motif that has a C at the second position and that has G at the third position, thereby determining respective additional amounts, wherein the feature vector includes the respective additional amounts.

88. The method of claim 86, wherein the feature vector includes respective amounts of all 3-mer end sequence motifs.

89. The method of claim 62, wherein the first set of one or more end sequence motifs includes all sequence motifs having C at the second position and G at the third position, the method further comprising:

determining a respective amount of each 3-mer end sequence motif that has a C at the second position and that has G at the third position, thereby determining respective amounts;

generating a feature vector including the respective amounts, which include the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology for the subject, wherein the machine learning model is trained using cell-free DNA molecules in training samples having known classifications.

90. The method of claim 86 or claim 89, wherein the classification uses differentially-methylated regions for a plurality of tissue types, including the first tissue type, the method further comprising:

analyzing other sets of cell-free DNA molecules from the biological sample, wherein each set of cell-free DNA molecules are located in a respective set of one or more regions of the reference genome, each of the respective set of one or more regions being hypomethylated or hypermethylated in a respective tissue type of the plurality of tissue types,

wherein the feature vector forms a matrix with each row corresponding to one of the respective sets of one or more regions of the reference genome, and wherein a column corresponding to the respective amount of each 3-mer end sequence motif of the first set, and wherein the machine learning model is a convolutional neural network.

91. The method of claim 62, wherein the first set of one or more end sequence motifs is a plurality of end sequence motifs.

92. The method of claim 62, wherein the classification is determined using one or more end sequence motifs only selected from a group consisting of (1) end sequence motifs having C at the first position and G at the second position and (2) end sequence motifs having C at the second position and G at the third position.

93. A method of analyzing a biological sample, including a mixture of cell-free DNA molecules, to determine a level of a pathology in a subject from which the biological sample is obtained, the mixture including cell-free DNA molecules from the subject and from a virus, the method comprising:

(a) analyzing a plurality of viral cell-free DNA molecules from the biological sample of the subject, wherein analyzing each of a group of viral cell-free DNA molecules of the virus includes:

determining a location of the viral cell-free DNA molecule in a reference genome of the virus, wherein the group of viral cell-free DNA molecules are located in a first set of one or more CpG sites of the reference genome; and

determining an end sequence motif of at least one end of the viral cell-free DNA molecule, thereby determining end sequence motifs, wherein an end of the viral cell-free DNA molecule has a first position at an outermost position, a second position that is next to the first position, and a third position that is next to the second position;

(b) determining, using the end sequence motifs, a first amount of the group of viral cell-free DNA molecules of the virus that end at any one or more of -1 to +1 positions of any one of the first set of one or more CpG sites; and

(c) determining a classification of the level of the pathology for the subject based on a comparison of the first amount to a reference value.

94. The method of claim 93, wherein each of the first set of one or more CpG sites have a methylation level within a specified range in subjects having the pathology.

95. The method of claim 93, wherein each of the first set of one or more CpG sites are identified using a first training set of samples from subjects having the pathology and a second training set of samples from subjects not having the pathology, wherein the first amount

at each of the first set of one or more CpG sites differs by at least a specified percentage between the first training set and the second training set.

96. The method of claim 93, wherein each of the first set of one or more CpG sites is differentially-methylated in subjects having the pathology relative to subjects not having the pathology.

97. The method of claim 93, wherein the first amount is of a first set of one or more end sequence motifs that have C at the first position and that have G at the second position.

98. The method of claim 97, wherein the first set of one or more end sequence motifs includes CGN, where N includes any nucleotide at the third position.

99. The method of claim 97, further comprising:

determining a second amount of a second set of end sequence motifs that have C at the second position and G at the third position,

wherein determining the classification includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the reference value.

100. The method of claim 99, wherein the second set of end sequence motifs includes NCG, where N includes any nucleotide at the first position.

101. The method of claim 93, wherein the first amount is of a first set of one or more end sequence motifs that have C at the second position and G at the third position.

102. The method of claim 93, wherein the first amount is a sum of the viral cell-free DNA molecules that end at a -1 or +1 position of the first set of one or more CpG sites.

103. The method of claim 102, wherein the first set of one or more CpG sites is each hypomethylated.

104. The method of claim 93, further comprising:

for each position of at least two positions within a window around any one the first set of one or more CpG sites:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts;

generating a feature vector including the respective amounts and the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology, wherein the machine learning model is trained using cell-free DNA molecules from training samples having known classifications.

105. The method of claim 104, wherein each of the first set of one or more CpG sites is hypomethylated in subjects having the pathology.

106. The method of claim 105, wherein the feature vector is a first feature vector, further comprising:

generating a second feature vector using a second set of one or more CpG sites that are all hypermethylated; and

inputting the second feature vector and the first feature vector into the machine learning model as part of determining the classification of the level of the pathology.

107. The method of claim 93, wherein the first set of one or more CpG sites is a first set of CpG sites, wherein determining the first amount includes:

determining respective amounts of cell-free DNA molecules ending at each of the first set of CpG sites, and

wherein determining classification of the level of the pathology includes comparing each of the respective amounts to a respective reference value.

108. The method of claim 107, wherein comparing each of the respective amounts to the respective reference value is performed by inputting the respective amounts as part of a feature vector into a machine learning model.

109. The method of claim 108, further comprising, for each CpG site of the first set of one or more CpG sites:

for each position of at least two positions within a window around the CpG site:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts; and

including the respective amounts for the at least two positions and the respective amount for the CpG site in the feature vector.

110. The method of claim 93, further comprising:

determining a total amount of sequence motifs for the plurality of viral cell-free DNA molecules,

wherein determining the classification includes normalizing the first amount with the total amount to obtain a normalized first amount that is compared to the reference value.

111. The method of claim 93, wherein the pathology is a cancer.

112. The method of claim 93, further comprising:

determining a respective amount of each 3-mer end sequence motif that has a C at the first position and that has G at the second position, thereby determining respective amounts;

generating a feature vector including the respective amounts, which include the first amount; and

inputting the feature vector into a machine learning model as part of determining the classification of the level of the pathology, wherein the machine learning model is trained using cell-free DNA molecules in training samples having known classifications.

113. The method of claim 112, further comprising:

determining a respective additional amount of each 3-mer end sequence motif that has a C at the second position and that has G at the third position, thereby determining respective additional amounts, wherein the feature vector includes the respective additional amounts.

114. The method of claim 112, wherein the feature vector includes respective amounts of all 3-mer end sequence motifs.

115. The method of claim 93, wherein analyzing the plurality of viral cell-free DNA molecules includes determining a size of each of the plurality of viral cell-free DNA molecules, the method further comprising:

determining a statistical value of a size distribution of the plurality of viral cell-free DNA molecules, and

wherein determining the classification of the level of the pathology for the subject further includes comparing the statistical value to a size reference value.

116. The method of claim 93 or claim 115, further comprising:

determining an amount of the plurality of viral cell-free DNA molecules, and

wherein determining the classification of the level of the pathology for the subject further includes comparing the amount to a count reference value.

117. A method for measuring a fractional concentration of DNA from a first tissue type in a biological sample of a subject, the biological sample comprising cell-free DNA, the method comprising:

analyzing a plurality of cell-free DNA molecules from the biological sample of the 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;

identifying a first set of CpG sites that all have a first differential methylation for the first tissue type in the reference genome, wherein the first set of CpG sites includes one or more CpG sites;

determining a first amount of cell-free DNA molecules ending at one of 0 or -1 position of any one of the first set of CpG sites; and

determining the fractional concentration of DNA from the first tissue type in the biological sample by comparing the first amount to a calibration value, wherein the calibration value is determined from one or more calibration samples having known fractional concentrations of DNA from the first tissue type.

118. The method of claim 117, wherein the first differential methylation is that all of the first set of CpG sites are hypomethylated, or wherein the first differential methylation is that all of the first set of CpG sites are hypermethylated.

119. The method of claim 118, further comprising:

identifying a second set of CpG sites that are all hypomethylated or all hypermethylated across a plurality of tissue types in the reference genome; and

determining a second amount of cell-free DNA molecules ending within a window around any one of the second set of CpG sites,

wherein determining the fractional concentration includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the calibration value.

120. The method of claim 118, further comprising:

identifying a second set of CpG sites that are all hypomethylated or all hypermethylated for a second tissue type in the reference genome, wherein the second set of CpG sites includes one or more CpG sites;

determining a second amount of cell-free DNA molecules ending at any one of the second set of CpG sites; and

determining a change in a fractional concentration of DNA from the second tissue type in the biological sample by comparing the second amount to another calibration value, wherein the other calibration value is determined from one or more other calibration samples.

121. The method of claim 117, wherein determining the first amount includes:

determining respective amounts of cell-free DNA molecules ending at the 0 position of each of the first set of CpG sites, and

wherein determining the fractional concentration of DNA from the first tissue type includes comparing each of the respective amounts to a respective calibration value.

122. The method of claim 121, wherein comparing each of the respective amounts to the respective calibration value is performed by inputting the respective amounts as part of a feature vector into a machine learning model.

123. The method of claim 122, further comprising, for each CpG site of the one or more CpG sites:

for each position of at least two positions within a window around the CpG site:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts; and

including the respective amounts for the at least two positions and the respective amount for the CpG site in the feature vector.

124. The method of claim 117, wherein the first set of CpG sites is a plurality of CpG sites.

125. The method of claim 117, wherein the first amount is of cell-free DNA molecules ending at the 0 position, the method further comprising:

determining a second amount of cell-free DNA molecules ending at a first position within a window around any one the first set of CpG sites,

wherein determining the fractional concentration includes normalizing the first amount with the second amount to obtain a normalized first amount that is compared to the calibration value.

126. The method of claim 125, wherein the first position is at +1 or -1 from a CpG site.

127. The method of claim 125, wherein the window is at least -2 to +2 from a CpG site.

128. The method of claim 125, further comprising:

for each position of at least two positions within the window:

determining a respective amount of cell-free DNA molecules ending at the position, wherein the at least two positions include the first position; and

comparing the first amount of cell-free DNA molecules ending at the CpG site to the respective amount of cell-free DNA molecules ending at the position as part of determining the fractional concentration.

129. The method of claim 125, further comprising:

for each position of at least two positions within the window:

determining a respective amount of cell-free DNA molecules ending at the position, thereby determining respective amounts including the second amount, wherein the at least two positions include the first position;

generating a feature vector including the respective amounts and the first amount; and

inputting the feature vector into a machine learning model as part of determining the fractional concentration, wherein the machine learning model is trained using cell-free DNA molecules from the one or more calibration samples having the known fractional concentrations of DNA from the first tissue type.

130. The method of claim 129, wherein the first set of one or more CpG sites are all hypomethylated for the first tissue type in the reference genome, and wherein the feature vector is a first feature vector, the method further comprising:

generating a second feature vector using a second set of one or more CpG sites that are all hypermethylated; and

inputting the second feature vector and the first feature vector into the machine learning model as part of determining the fractional concentration.

131. The method of claim 129, wherein the at least two positions include all positions within the window, the window being at least +4 to -4 from a CpG site.

132. The method of claim 129, wherein the machine learning model is support vector regression.

133. The method of claim 117, wherein the first amount is normalized.

134. The method of claim 133, wherein the normalization uses a number of cell-free DNA molecules ending within a region including a CpG site.

135. The method of claim 133, wherein the normalization uses a number of cell-free DNA molecules covering a CpG site.

136. The method of claim 133, wherein the normalization uses an average or median depth of cell-free DNA molecules in a region including a CpG site.

137. The method of claim 117, wherein the fractional concentration of DNA is a range that is 30% or less.

138. The method of claim 117, wherein the first amount is determined of cell-free DNA molecules ending at the 0 position of any one of the first set of CpG sites.

139. A method for measuring a fractional concentration of DNA from a first tissue type in a biological sample of a subject, the biological sample comprising cell-free DNA, the method comprising:

analyzing a plurality of cell-free DNA molecules from the biological sample of the 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; and

determining an end sequence motif of at least one end of the cell-free DNA molecule, wherein an end of the cell-free DNA molecule has a first position at an outermost position, a second position that is next to the first position, and a third position that is next to the second position;

identifying a first set of CpG sites that are all hypomethylated or hypermethylated for the first tissue type in the reference genome, wherein the first set of CpG sites includes one or more CpG sites;

determining a first amount of a first set of one or more end sequence motifs, wherein:

the first set of one or more end sequence motifs have C at the first position and G at the second position, or

the first set of one or more end sequence motifs have C at the second position and G at the third position; and

determining the fractional concentration of DNA from the first tissue type in the biological sample by comparing the first amount to a calibration value, wherein the calibration value is determined from one or more calibration samples having known fractional concentrations of DNA from the first tissue type.

140. The method of any one of the preceding claims, wherein analyzing the plurality of cell-free DNA molecules includes sequencing the plurality of cell-free DNA molecules.

141. The method of any one of the preceding claims, wherein analyzing the plurality of cell-free DNA molecules includes using PCR.

142. The method of claim 141, wherein the PCR targets sequences in a repeat region.

143. The method of any one of the preceding claims, wherein analyzing the plurality of cell-free DNA molecules includes analyzing at least 1,000 cell-free DNA molecules.

144. A computer product comprising a non-transitory computer readable medium storing a plurality of instructions that, when executed, cause a computer system to perform the method of any one of the preceding claims.

145. A system comprising:

the computer product of claim 144; and

one or more processors for executing instructions stored on the computer readable medium.

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

147. A system comprising one or more processors configured to perform any of the above methods.

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