

SUMMARY

[0007a] In one aspect, the present invention provides a method of determining a type of cancer by analyzing cell-free DNA molecules from a biological sample of an organism, the method comprising:

- 5 (a) obtaining sequence reads from cell-free DNA molecules from a biological sample, wherein the sequence reads comprise a methylation status of the cell-free DNA molecules;
- (b) analyzing the sequence reads to determine, at single nucleotide resolution, methylation statuses for a plurality of sites in the cell-free DNA molecules;
- 10 (c) determining a methylation profile from the methylation statuses for the plurality of sites; and
- (d) determining a type of cancer based at least in part on the methylation profile.

[0007b] In another aspect, the present invention provides a method of analyzing a biological sample from a subject, the method comprising:

- 15 (a) obtaining sequence reads for cell-free DNA molecules from a biological sample of a subject, wherein the sequence reads include methylation statuses for the cell-free DNA molecules at single nucleotide resolution;
- (b) analyzing the sequence reads to determine a methylation profile for a plurality of sites based on the methylation statuses for the plurality of sites; and
- 20 (c) determining a tissue source for at least a portion of the cell-free DNA molecules from the biological sample based, at least in part, on the methylation profile.

[0007c] In another aspect, the present invention provides a method of estimating a methylation level of DNA in a biological sample of an organism, the method comprising:

- for each size of a plurality of sizes:
- 25 measuring an amount of a plurality of DNA fragments from the biological sample corresponding to the size;
- calculating, with a computer system, a first size value of a first parameter based on the amounts of DNA fragments at multiple sizes, the first parameter providing a statistical measure of a size profile of DNA fragments in the biological sample;
- 30 comparing the first size value to a reference size value, the reference size value calculated from DNA fragments of a reference sample; and
- estimating the methylation level based on the comparison.

[0007d] In yet another aspect, the present invention provides a computer product comprising a computer readable medium storing a plurality of instructions that when executed controls a computer system to perform the method of the present disclosure.

[0007e] In yet another aspect, the present invention provides a computer product including processors that are programmed to implement any of the methods of of the present disclosure.

[0008] Embodiments provide systems, methods, and apparatuses for determining and using methylation profiles of various tissues and samples. Examples are provided. A methylation profile can be deduced for fetal/tumor tissue based on a comparison of plasma methylation (or other sample with cell-free DNA, e.g., urine, saliva, genital washings) to a methylation profile of the mother/patient. A methylation profile can be determined for fetal/tumor tissue using tissue-specific alleles to identify DNA from the fetus/tumor when the sample has a mixture of DNA. A methylation profile can be used to determine copy number variations in genome of a fetus/tumor. Methylation markers for a fetus have been identified via various techniques. The methylation profile can be determined by determining a size parameter of a size distribution of DNA fragments, where reference values for the size parameter can be used to determine methylation levels.

[0009] Additionally, a methylation level can be used to determine a level of cancer. In the context of cancer, the measurement of the methylomic changes in plasma can allow one to detect the cancer (e.g. for screening purposes), for monitoring (e.g. to detect response following anti-cancer treatment; and to detect cancer relapse) and for prognostication (e.g. for measuring the load of cancer cells in the body or for staging purposes or for assessing the chance of death from disease or disease progression or metastatic processes).

[0010] A better understanding of the nature and advantages of embodiments of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] FIG. 1A shows a table 100 of sequencing results for maternal blood, placenta, and maternal plasma according to embodiments of the present invention.

[0012] FIG. 1B shows methylation density in 1-Mb windows of sequenced samples according to embodiments of the present invention.

[0013] FIGS. 2A-2C show plots of the beta-values against the methylation indices: (A) Maternal blood cells, (B) Chorionic villus sample, (C) Term placental tissue.

WHAT IS CLAIMED IS:

1. A method of determining a type of cancer by analyzing cell-free DNA molecules from a biological sample of an organism, the method comprising:
 - (a) obtaining sequence reads from cell-free DNA molecules from a biological
5 sample, wherein the sequence reads comprise a methylation status of the cell-free DNA molecules;
 - (b) analyzing the sequence reads to determine, at single nucleotide resolution, methylation statuses for a plurality of sites in the cell-free DNA molecules;
 - (c) determining a methylation profile from the methylation statuses for the
10 plurality of sites; and
 - (d) determining a type of cancer based at least in part on the methylation profile.
2. The method of claim 1, wherein the biological sample is selected from a group consisting of blood, plasma, and serum.
3. The method of claim 1, wherein the biological sample is selected from a
15 group consisting of urine, vaginal fluid, uterine or vaginal flushing fluids, plural fluid, ascitic fluid, cerebrospinal fluid, saliva, sweat, tears, sputum, bronchoalveolar lavage, fluid, and stool.
4. The method of any one of claims 1–3, wherein the type of cancer comprises an organ in which the cancer has developed.
5. The method of any one of claims 1–3, wherein the type of cancer
20 comprises a tissue of origin of the cancer.
6. The method of claim 5, wherein the tissue of origin is selected from the group consisting of brain, bone, lung, heart, kidney, liver, breast, colon, prostate, nasopharynx, stomach, testes, skin, ovaries, pancreas, uterus, and lymph node.
7. The method of any one of claims 1–6, wherein the type of cancer is
25 selected from the group consisting of lung cancer, breast cancer, colorectal cancer, prostate cancer, nasopharyngeal cancer, gastric cancer, testicular cancer, skin cancer, cancer affecting the nervous system, bone cancer, ovarian cancer, liver cancer, hematologic malignancies, pancreatic cancer, endometriocarcinoma, and kidney cancer.

8. The method of any one of claims 1–7, further comprising sequencing the cell-free DNA molecules to obtain the sequence reads.

9. The method of claim 8, wherein the sequencing comprises methylation-aware sequencing.

5 10. The method of claim 9, wherein the methylation-aware sequencing comprises bisulfite sequencing.

11. The method of claim 10, wherein the bisulfite sequencing comprises genome-wide bisulfite sequencing.

10 12. The method of any one of claims 9–11, wherein the methylation-aware sequencing comprises paired-end massively parallel sequencing or whole genome sequencing.

13. The method of any one of claims 8–12, further comprising enriching the cell-free DNA molecules before the sequencing.

14. The method of claim 13, wherein the enriching comprises use of hybridization probes, polymerase chain reaction amplification, or solid phase hybridization.

15 15. The method of any one of claims 1–14, wherein the plurality of sites comprise one or more CpG sites.

16. The method of any one of claims 1–15, wherein the plurality of sites comprise one or more CpG sites that are organized into one or more CpG islands.

20 17. The method of claim 16, wherein the determining the methylation profile for the plurality of sites comprises calculating a methylation density for the one or more CpG islands using the methylation statuses determined from the sequence reads at the one or more CpG islands.

25 18. The method of any one of claims 1–17, wherein the determining the methylation profile for the plurality of sites comprises calculating a methylation density for each site of the plurality of sites using the methylation statuses determined from the sequence reads at the site.

19. The method of any one of claims 1–18, wherein the analyzing further comprises determining locations of the cell-free DNA molecules in a genome

20. The method of claim 19, wherein the determining the methylation profile comprises determining one or more methylation levels in one or more genomic regions using the
5 methylation statuses determined from the sequence reads of the cell-free DNA molecules located in the one or more genomic regions.

21. The method of claim 20, wherein the one or more genomic regions comprise a plurality of genomic bins specified in the genome.

22. The method of claim 21, wherein the plurality of genomic bins have sizes
10 selected from a group consisting of 50 kb, 100 kb, 200 kb, 500 kb, and 1 Mb.

23. The method of claim 21 or claim 22, wherein the plurality of genomic bins have different sizes.

24. The method of any one of claims 21–23, further comprising selecting a group of genomic bins by comparing a methylation level of each of the genomic bins to a cutoff
15 value.

25. The method of claim 24, further comprising determining the type of cancer based at least in part on the methylation levels of the selected group of genomic bins.

26. The method of claim 25, wherein the cutoff value is determined from a subject known to have cancer.

20 27. The method of claim 25, wherein the cutoff value is determined from a subject without cancer.

28. The method of any one of claims 1–27, wherein the determining the type of cancer comprises comparing the methylation profile to a reference methylation profile.

25 29. The method of claim 28, wherein the reference methylation profile corresponds to a healthy subject.

30. The method of claim 28, wherein the reference methylation profile is obtained from at least one sample obtained from a subject known to have cancer.

31. The method of any one of claims 1–30, further comprising determining presence or absence of a copy number aberration based on the methylation profile.

32. The method of any one of claims 1–31, further comprising determining one or more methylation levels at one or more genomic sites in the cell-free DNA molecules.

5 33. The method of any one of claims 1–32, wherein determining the methylation profile comprises determining one or more methylation levels for one or more CpG islands.

34. A method of analyzing a biological sample from a subject, the method comprising:

10 (a) obtaining sequence reads for cell-free DNA molecules from a biological sample of a subject, wherein the sequence reads include methylation statuses for the cell-free DNA molecules at single nucleotide resolution;

 (b) analyzing the sequence reads to determine a methylation profile for a plurality of sites based on the methylation statuses for the plurality of sites; and

15 (c) determining a tissue source for at least a portion of the cell-free DNA molecules from the biological sample based, at least in part, on the methylation profile.

35. The method of claim 34, wherein at least a portion of the cell-free DNA molecules are cancer-derived molecules, the method further comprising determining a type of cancer of the subject based, at least in part, on the methylation profile.

20 36. The method of claim 35, wherein the type of cancer is selected from the group consisting of lung cancer, breast cancer, colorectal cancer, prostate cancer, nasopharyngeal cancer, gastric cancer, testicular cancer, skin cancer, cancer affecting the nervous system, bone cancer, ovarian cancer, liver cancer, hematologic malignancies, pancreatic cancer, endometriocarcinoma, and kidney cancer.

25 37. The method of claim 34, further comprising sequencing the cell-free DNA molecules to obtain the sequence reads.

38. The method of claim 37, wherein the sequencing comprises methylation-aware sequencing.

39. The method of claim 38, wherein the methylation-aware sequencing comprises bisulfite sequencing.

40. The method of claim 39, wherein the plurality of sites comprise one or more CpG sites.

5

41. A method of estimating a methylation level of DNA in a biological sample of an organism, the method comprising:

for each size of a plurality of sizes:

10 measuring an amount of a plurality of DNA fragments from the biological sample corresponding to the size;

calculating, with a computer system, a first size value of a first parameter based on the amounts of DNA fragments at multiple sizes, the first parameter providing a statistical measure of a size profile of DNA fragments in the biological sample;

15 comparing the first size value to a reference size value, the reference size value calculated from DNA fragments of a reference sample; and
estimating the methylation level based on the comparison.

42. The method of claim 41, wherein the first parameter is a percentage of DNA fragments of a particular size relative to all of the measured DNA fragments.

20 43. The method of claim 41, wherein the first parameter is a ratio of a first amount of DNA fragments corresponding to a first size and a second amount of DNA fragments corresponding to a second size.

44. The method of claim 41, wherein the plurality of sizes correspond to different ranges of sizes.

25 45. The method of claim 41, wherein measuring the amount of the plurality of DNA fragments from the biological sample corresponding to the size is determined based on a physical separation that is dependent on size.

46. The method of claim 41, wherein measuring amounts of DNA fragments corresponding to the sizes includes:

for each of the plurality of DNA fragments from the biological sample:

30 measuring a size of the DNA fragment,

wherein measuring a size of a DNA fragment includes:

performing paired-end sequencing of the DNA fragment to obtain paired sequence reads;

aligning the paired sequence reads to a reference genome; and

5 using the aligned locations to determine the size of the DNA fragment.

47. The method of claim 46, wherein the methylation level corresponds to a particular region by determining the first size value only from DNA fragments that align to the particular region.

48. The method of claim 41, wherein measuring amounts of DNA fragments
10 corresponding to the sizes includes using electrophoresis.

49. The method of claim 41, wherein the reference sample is obtained from the organism at an earlier time than the biological sample, and wherein the estimated methylation level is a relative level corresponding to an amount of increase or decrease of the first size value relative to the reference size value.

15 50. The method of claim 49, wherein the organism is identified as having cancer at the earlier time, and wherein the biological sample is obtained after treatment.

51. The method of claim 49, wherein the organism is identified as having cancer at the earlier time, and wherein the increase of the first size value relative to the reference size value is used to determine a progression of the cancer.

20 52. The method of claim 41, wherein the reference size value is part of a calibration curve, and wherein comparing the first size value to the reference size value includes identifying a point on the calibration curve corresponding to the first size value.

53. The method of claim 41, wherein the methylation level is a methylation density.

25 54. The method of claim 53, wherein the methylation density is for CpG sites.

55. The method of claim 41, wherein the biological sample is of a female subject pregnant with a fetus.

56. The method of claim 41, wherein the biological sample includes nucleic acid molecules originating from normal cells and potentially from cells associated with cancer.

57. A computer product comprising a computer readable medium storing a plurality of instructions that when executed controls a computer system to perform the method
5 of any one of claims 1–7, 15-36, and 40-56.

58. A computer system including processors that are programmed to implement any of the methods of any of claims 1–7, 15-36, and 40-56.