

SUMMARY

[0007a] In one aspect, the present invention provides a computer-implemented method of determining a type of cancer by analyzing cell-free DNA molecules from a biological sample of a subject, 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 at one or more sites of each of the cell-free DNA molecules;
- (b) analyzing the sequence reads to determine a number of the cell-free DNA molecules methylated at each of a plurality of sites, and using the numbers to obtain, at single
10 nucleotide resolution, methylation statuses for the plurality of sites in the cell-free DNA molecules;
- (c) determining a methylation profile from the methylation statuses for the plurality of sites; and
- (d) determining that the subject has a first type of cancer based at least in part on
15 the methylation profile, wherein the determining the first type of cancer comprises comparing the methylation profile to reference methylation profiles determined from methylation statuses of other subjects, wherein at least two of the other subjects are identified as having different types of cancer, including the first type of cancer.

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

- (a) obtaining sequence reads for cell-free DNA molecules from a biological sample of a subject, wherein the sequence reads include methylation statuses at a plurality of sites for the cell-free DNA molecules at single nucleotide resolution;
- (b) analyzing the sequence reads to determine a number of the cell-free DNA
25 molecules methylated at each of the plurality of sites based on the methylation statuses, and using the numbers to determine a methylation profile for the plurality of sites; and
- (c) determining whether the methylation profile corresponds to a first tissue source for at least a portion of the cell-free DNA molecules from the biological sample, wherein the determining comprises comparing the methylation profile to reference methylation profiles
30 corresponding to different tissue sources, including the first tissue source.

WHAT IS CLAIMED IS:

1. A computer-implemented method of determining a type of cancer by analyzing cell-free DNA molecules from a biological sample of a subject, the method comprising:
 - (a) obtaining sequence reads from cell-free DNA molecules from a biological sample, wherein the sequence reads comprise a methylation status at one or more sites of each of the cell-free DNA molecules;
 - (b) analyzing the sequence reads to determine a number of the cell-free DNA molecules methylated at each of a plurality of sites, and using the numbers to obtain, at single nucleotide resolution, methylation statuses for the plurality of sites in the cell-free DNA molecules;
 - (c) determining a methylation profile from the methylation statuses for the plurality of sites; and
 - (d) determining that the subject has a first type of cancer based at least in part on the methylation profile, wherein the determining the first type of cancer comprises comparing the methylation profile to reference methylation profiles determined from methylation statuses of other subjects, wherein at least two of the other subjects are identified as having different types of cancer, including the first type of cancer.
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 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 determining that the subject has the first type of cancer comprises determining an organ in which the cancer has developed.
5. The method of any one of claims 1–3, wherein determining that the subject has the first type of cancer comprises determining 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 first 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.

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.

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.

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

16. The method of claim 15, wherein the CpG sites are organized into a plurality of CpG islands.

17. The method of claim 16, wherein the determining the methylation profile for the plurality of sites comprises, for each of the plurality of CpG islands, determining a number of sequence reads showing methylation at CpG sites in the CpG island.

18. The method of any one of claims 1–17, wherein the determining the methylation profile for the plurality of sites comprises, for each site of the plurality of sites, determining a number of the sequence reads showing methylation at the site and determining a total number of sequence reads at the plurality of sites.

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 a number of sequence reads of the cell-free DNA molecules showing methylation at sites in 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 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, wherein determining the methylation profile comprises selecting a group of genomic bins by comparing a methylation level of each of the genomic bins to a cutoff value.

25. The method of claim 24, wherein the first type of cancer is determined 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.

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 reference methylation profiles are obtained from reference samples comprising cell-free DNA molecules.

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

30. The method of claim 28, wherein the reference samples comprise a plurality of reference samples corresponding to respective subjects, and wherein comparing the methylation profile to the reference methylation profiles comprises:

clustering the reference methylation profiles determined from the plurality of reference samples to obtain one or more clusters of reference methylation profiles; and

comparing the methylation profile to the one or more clusters to determine the first type of cancer associated with the methylation profile.

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.

33. The method of claim 16, wherein determining the methylation profile comprises determining one or more methylation levels for the plurality of CpG islands.

34. The method of claim 33, wherein the plurality of CpG islands include at least 110 CpG islands.

35. A computer-implemented method of analyzing a biological sample from a subject, the method comprising:

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

(b) analyzing the sequence reads to determine a number of the cell-free DNA molecules methylated at each of the plurality of sites based on the methylation statuses, and using the numbers to determine a methylation profile for the plurality of sites; and

(c) determining whether the methylation profile corresponds to a first tissue source for at least a portion of the cell-free DNA molecules from the biological sample, wherein the determining comprises comparing the methylation profile to reference methylation profiles corresponding to different tissue sources, including the first tissue source.

36. The method of claim 35, 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.

37. The method of claim 36, 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.

38. The method of claim 35, further comprising sequencing the cell-free DNA molecules to obtain the sequence reads.

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

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

41. The method of claim 40, wherein the plurality of sites comprise CpG sites.

42. The method of any one of claims 1 to 41, wherein the methylation profile comprises a pattern of CpG island hypermethylation.

43. The method of claim 42, wherein the pattern of CpG island hypermethylation corresponds to a pattern of CpG island hypermethylation for at least 110 CpG sites.

44. The method of any of claims 1 to 43, wherein obtaining sequence reads for cell-free DNA molecules includes obtaining the sequence reads from massively parallel sequencing of bisulfite converted cell-free DNA molecules from the biological sample.

45. 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 any one of claims 1–7, 15-36, and 44.

46. A computer system including processors that are programmed to implement any of the methods of any of claims 1–7, 15-36, and 44.