

SUMMARY

[0007a] In one aspect, provided herein is a method of analyzing a biological sample of an organism, the biological sample comprising cell-free DNA originating from normal cells and potentially from cells associated with cancer, the method comprising: analyzing a plurality of cell-free DNA molecules from the biological sample, wherein analyzing each of the plurality of cell-free DNA molecules includes: determining a location of the cell-free DNA molecule in a genome of the organism; and determining whether the cell-free DNA molecule is methylated at one or more sites, wherein determining whether the cell-free DNA molecule is methylated at the one or more sites comprises performing methylation-aware sequencing, and wherein performing methylation-aware sequencing comprises performing methylation-aware massively parallel sequencing, the one or more sites of each of the plurality of cell-free DNA molecules providing a plurality of sites; for each site of the plurality of sites, individually counting a respective number of the cell-free DNA molecules at the site that are hypermethylated; and calculating a first methylation level using the respective numbers of cell-free DNA molecules hypermethylated at the plurality of sites.

[0007b] In one aspect, provided herein is a non-transitory computer readable medium comprising a plurality of instructions capable of execution by a computer system, that when so executed control the computer system to perform the method as disclosed herein.

[0008] Examples 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 examples of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

5

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.

10

[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 analyzing a biological sample of an organism, the biological sample comprising cell-free DNA originating from normal cells and potentially from cells associated with cancer, the method comprising:

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

determining a location of the cell-free DNA molecule in a genome of the organism; and

determining whether the cell-free DNA molecule is methylated at one or more sites, wherein determining whether the cell-free DNA molecule is methylated at the one or more sites comprises performing methylation-aware sequencing, and wherein performing methylation-aware sequencing comprises performing methylation-aware massively parallel sequencing, the one or more sites of each of the plurality of cell-free DNA molecules providing a plurality of sites;

for each site of the plurality of sites,

individually counting a respective number of the cell-free DNA molecules at the site that are hypermethylated; and

calculating a first methylation level using the respective numbers of cell-free DNA molecules hypermethylated at the plurality of sites;.

2. The method of claim 1, wherein performing methylation-aware sequencing comprises sequencing of at least 60,000 cell-free DNA molecules.

3. The method of claim 1, wherein performing methylation-aware sequencing generates between 39 million and 142 million reads.

4. The method of claim 1, wherein performing methylation-aware sequencing includes:

treating the cell-free DNA molecules with sodium bisulfite; and

performing the methylation-aware parallel sequencing on the treated cell-free DNA molecules.

5. The method of claim 4, wherein treating the cell-free DNA molecules with sodium bisulfite is part of Tet-assisted bisulfite conversion or oxidative bisulfite sequencing for a detection of 5-hydroxymethylcytosine.

6. The method of claim 1, further comprising determining a first classification of a level of cancer based on the first methylation level.

7. The method of claim 6, wherein determining the first classification of the level of cancer based on the first methylation level comprises:

comparing the first methylation level to a first cutoff value; and
determining the first classification of the level of cancer based on the comparison.

8. The method of claim 7, wherein the first classification indicates that cancer exists for the organism, the method further comprising

identifying a type of cancer associated with the organism by comparing the first methylation level to a corresponding value determined from other organisms, wherein at least two of the other organisms are identified as having different types of cancer.

9. The method of claim 7, wherein the first cutoff value is a specified distance from a reference methylation level established from a biological sample obtained from a healthy organism.

10. The method of claim 9, wherein the specified distance is a specified number of standard deviations from the reference methylation level.

11. The method of claim 7, wherein the first cutoff value is established from a reference methylation level determined from a previous biological sample of the organism obtained previous to the biological sample being tested.

12. The method of claim 7, wherein comparing the first methylation level to the first cutoff value includes:

determining a difference between the first methylation level and a reference methylation level; and

comparing the difference to a threshold corresponding to the first cutoff value.

13. The method of claim 7, further comprising:

determining a fractional concentration of tumor DNA in the biological sample;

and

calculating the first cutoff value based on the fractional concentration of tumor DNA in the biological sample.

14. The method of claim 7, further comprising:

measuring a size of the cell-free DNA molecules at the plurality of sites, thereby obtaining measured sizes; and

before comparing the first methylation level to the first cutoff value, normalizing the first methylation level using cell-free DNA molecules having a first size.

15. The method of claim 14, wherein the first size is a range of lengths.

16. The method of claim 14, wherein the cell-free DNA molecules having the first size are selected based on a physical separation that is dependent on size.

17. The method of claim 14, further comprising selecting the cell-free DNA molecules having the first size by:

performing paired-end massively parallel sequencing of the plurality of cell-free DNA molecules to obtain pairs of sequences for each of the cell-free DNA molecules;
determining a size of a cell-free DNA molecule by comparing the pair of sequences to a reference genome; and
selecting the cell-free DNA molecules having the first size.

18. The method of claim 14, wherein normalizing the first methylation level using the cell-free DNA molecules having the first size includes:

obtaining a functional relationship between size and methylation levels; and
using the functional relationship to normalize the first methylation level,
wherein the functional relationship provides scaling values corresponding to respective sizes.

19. The method of claim 18, further comprising:
computing an average size corresponding to the cell-free DNA molecules used to calculate the first methylation level; and
multiplying the first methylation level by a corresponding scaling value.

20. The method of claim 18, further comprising:
for each site of the plurality of sites:
for each of the cell-free DNA molecules located at the site:
obtaining a respective size of the cell-free DNA molecule at the site; and
using a scaling value corresponding to the respective size to normalize a contribution of the cell-free DNA molecule to the respective number of cell-free DNA molecules that are hypermethylated at the site.

21. The method of claim 7, wherein the plurality of sites includes CpG sites, wherein the CpG sites are organized into a plurality of CpG islands, each CpG island including more than one CpG site, wherein the first methylation level corresponds to a first CpG island.

22. The method of claim 21, further comprising:
for each CpG island of the plurality of CpG islands,
determining whether the CpG island is hypermethylated relative to a reference group of samples of other organisms by comparing a methylation level of the CpG island to a respective cutoff value, thereby determining hypermethylated CpG islands;
determining respective methylation densities for the hypermethylated CpG islands;
calculating a cumulative score from the respective methylation densities; and

comparing the cumulative score to a cumulative cutoff value to determine the first classification.

23. The method of claim 1, further comprising:
determining whether a fractional concentration of tumor DNA in the biological sample is greater than a minimum value; and
if the fractional concentration of tumor DNA is not greater than the minimum value, flagging the biological sample.

24. The method of claim 23, wherein the minimum value is determined based on an expected difference in methylation levels for a tumor relative to a reference methylation level.

25. The method of claim 1, wherein the plurality of sites are on a plurality of chromosomes.

26. The method of claim 1, wherein the plurality of sites are from disjointed regions separated from each other.

27. The method of claim 1, wherein analyzing the plurality of cell-free DNA molecules comprises analyzing at least 10 million sequence reads.

28. A non-transitory computer readable medium comprising a plurality of instructions capable of execution by a computer system, that when so executed control the computer system to perform the method of claim 1.