

ENHANCEMENT OF CANCER SCREENING USING CELL-FREE VIRAL NUCLEIC ACIDS

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority from and is a non-provisional of U.S.

5 Provisional Application No. 62/537,328, entitled “Enhancement Of Cancer Screening Using Cell-Free Viral Nucleic Acids” filed July 26, 2017, the entire contents of which are herein incorporated by reference for all purposes.

BACKGROUND

[0002] The discovery that tumor cells release tumor-derived DNA into the blood stream has
10 sparked the development of non-invasive methods capable of determining the presence, location and/or type of tumor in a subject using cell-free samples (e.g., plasma). Many tumors can be treatable if detected early in their development. However, current methods can lack the sensitivity and/or specificity to detect a tumor at an early stage and can return a large number of false positive or false negative results. For example, certain viruses are associated with cancer,
15 but viral DNA can be detected in subjects that do not have cancer, thereby causing false positive results.

[0003] The sensitivity of a test can refer to the likelihood that a subject that is positive for a condition tests positive for the condition. The specificity of a test can refer to the likelihood that a subject that is negative for a condition tests negative for that condition. The problems
20 of sensitivity and specificity can be exaggerated in assays for the early detection of tumors, e.g., because samples on which such tumor detection methods are performed can have relatively small amounts of tumor-derived DNA and because the condition itself can have a relatively low prevalence among individuals tested in the early stage. Accordingly, there is a clinical need for methods having higher sensitivity and/or specificity for the detection of tumors.

25

SUMMARY

[0003a] In one aspect, provided herein is a method of analyzing a biological sample of a subject that is an animal, the biological sample including a mixture of cell-free DNA molecules from a

genome of the subject and from one or more other genomes, the method comprising: analyzing a plurality of cell-free DNA molecules from the biological sample, wherein analyzing one of the plurality of cell-free DNA molecules includes: identifying a location of the cell-free DNA molecule in a particular viral genome; and determining whether the cell-free DNA molecule is methylated at one or more sites of the particular viral genome of a virus; measuring one or more mixture methylation levels based on one or more amounts of the plurality of cell-free DNA molecules methylated at a set of one or more sites of the particular viral genome; comparing the one or more mixture methylation levels to one or more reference methylation levels determined from at least two cohorts of other subjects, wherein the at least two cohorts have different classifications associated with the particular viral genome, the different classifications including a first condition that is associated with the virus; and determining a first classification of whether the subject has the first condition based on the comparing; for each of a set of cell-free DNA molecules in a sample: measuring a size of the cell-free DNA molecule; and identifying the location of the cell-free DNA molecule in the particular viral genome, the sizes of the set of cell-free DNA molecules forming a size distribution, the sample being the biological sample or a different sample including a mixture of cell-free DNA molecules from the genome of the subject and from the one or more other genomes; determining a statistical value of the size distribution; comparing the statistical value to a reference size value determined from the at least two cohorts of other subjects; determining a second classification of whether the subject has the first condition based on the comparing of the statistical value to the reference size value; and determining a final classification using the first classification and the second classification.

[0003b] In another aspect, provided herein is a computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform operations of the methods as described herein.

[0003c] In another aspect, provided herein is a system comprising: the computer product as described herein; and one or more processors for executing instructions stored on the computer readable medium.

[0003d] In another aspect, provided herein is a system comprising means for the methods as described herein.

[0003e] In another aspect, provided herein is a system configured to perform the methods as described herein.

[0003f] In another aspect, provided herein is a system comprising modules that respectively perform the steps of the methods as described herein.

- 5 **[0004]** Examples provide systems, apparatuses, and methods for analyzing a biological sample of a subject, e.g., in the animal kingdom, such as a human. Cell-free DNA molecules in a mixture of the biological sample can be analyzed to detect viral DNA, e.g., by determining a

location in a particular viral genome. A methylation status of the viral DNA at one or more sites in the viral genome can be determined. Mixture methylation level(s) can be measured based on one or more amounts of the plurality of cell-free DNA molecules methylated at a set of site(s) of the particular viral genome. The mixture methylation level(s) can be determined in various ways,
5 e.g., as a percentage/density of cell-free DNA molecules that are methylated at a particular site or across multiple sites, and potentially across multiple regions, each including one or more sites.

[0005] The mixture methylation level(s) can be compared to reference methylation level(s), e.g., determined from at least two cohorts of other subjects. The cohorts can have different classifications (including the first condition) associated with the particular viral genome. Other
10 cohort(s) can correspond to other condition(s). The comparison can be performed in a variety of ways, e.g., by forming a multidimensional point of N methylation levels and determining differences from N reference methylation levels. A first classification of whether the subject has the first condition can be determined based on the comparing.

[0006] These and other examples of the disclosure are described in detail below. For example,
15 other examples are directed to systems, devices, and computer readable media associated with methods described herein.

[0007] A better understanding of the nature and advantages of examples of the present disclosure may be gained with reference to the following detailed description and the accompanying drawings.

20 BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. 1 shows the design of capture probes for targeted bisulfite sequencing according to embodiments of the present disclosure.

[0009] FIG. 2 shows the methylation densities of CpG sites across the Epstein-Barr virus (EBV) genome in patients with infectious mononucleosis, nasopharyngeal carcinoma (NPC) and
25 Natural killer (NK)-T cell lymphoma according to embodiments of the present disclosure.

[0010] FIG. 3 shows the methylation profiles of plasma EBV DNA in a patient (AL038) from a screening cohort with early stage NPC (Stage I) according to embodiments of the present disclosure.

WHAT IS CLAIMED IS:

1. A method of analyzing a biological sample of a subject that is an animal, the biological sample including a mixture of cell-free DNA molecules from a genome of the subject and from one or more other genomes, the method comprising:

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

identifying a location of the cell-free DNA molecule in a particular viral genome; and

determining whether the cell-free DNA molecule is methylated at one or more sites of the particular viral genome of a virus;

10 measuring one or more mixture methylation levels based on one or more amounts of the plurality of cell-free DNA molecules methylated at a set of one or more sites of the particular viral genome;

comparing the one or more mixture methylation levels to one or more reference methylation levels determined from at least two cohorts of other subjects, wherein the at least two cohorts have different classifications associated with the particular viral genome, the different classifications including a first condition that is associated with the virus; and

15 determining a first classification of whether the subject has the first condition based on the comparing;

20 for each of a set of cell-free DNA molecules in a sample:

measuring a size of the cell-free DNA molecule; and

identifying the location of the cell-free DNA molecule in the particular viral genome, the sizes of the set of cell-free DNA molecules forming a size distribution, the sample being the biological sample or a different sample including a mixture of cell-free

25 DNA molecules from the genome of the subject and from the one or more other genomes;

determining a statistical value of the size distribution;

comparing the statistical value to a reference size value determined from the at least two cohorts of other subjects;

30 determining a second classification of whether the subject has the first condition based on the comparing of the statistical value to the reference size value; and

determining a final classification using the first classification and the second classification.

2. The method of claim 1, wherein the different classifications of the at least two cohorts further includes a second condition, the method further comprising:

5 determining another classification of whether the subject has the second condition based on the comparing.

3. The method of claim 2, wherein the one or more mixture methylation levels are compared to a plurality of reference methylation levels that include a first reference methylation level and a second reference methylation level, whether the first reference
10 methylation level is used to determine the first classification of whether the subject has the first condition, and wherein the second reference methylation level is used to determine the second classification of whether the subject has the second condition.

4. The method of claim 3, wherein the particular viral genome is of Epstein-Barr virus and the subject is a human, wherein the first condition is nasopharyngeal cancer, and
15 wherein the second condition is infectious mononucleosis.

5. The method of claim 1, wherein the first classification is that the subject does not have the first condition.

6. The method of claim 1, wherein determining the first classification includes determining a level of the first condition.

20 7. The method of claim 1, wherein the particular viral genome is of Epstein-Barr virus and the subject is a human, and wherein the first condition is nasopharyngeal cancer.

8. The method of claim 1, wherein the set of one or more sites includes at least two sites, and wherein the one or more mixture methylation levels is one mixture methylation level that is determined across the at least two sites.

25 9. The method of claim 1, wherein the one or more mixture methylation levels include N mixture methylation levels, N being an integer greater than one, wherein the set of one or more sites includes at least two sites, and wherein the comparing includes:

measuring differences between the N mixture methylation levels and N reference methylation levels; and

determining whether the subject belongs to one of the at least two cohorts using the differences.

5 10. The method of claim 9, wherein determining whether the subject belongs to one of the at least two cohorts using the differences includes performing a hierarchical clustering analysis.

11. The method of claim 9, wherein each of the N mixture methylation levels is measured for one of a plurality of predetermined regions.

10 12. The method of claim 11, wherein the plurality of predetermined regions are of a same size and span the particular viral genome, and wherein the same size is between 50 bases and 1,000 bases.

13. The method of claim 11, wherein each of the plurality of predetermined regions satisfies one or more criteria including (1) a difference in methylation level among
15 multiple subjects of a same cohort and/or (2) a difference in methylation level between a subject of one cohort and a subject of another cohort.

14. The method of claim 1, wherein the set of one or more sites reside in a plurality of regions that each satisfies one or more criteria including (1) a difference in methylation level among multiple subjects of a same cohort and/or (2) a difference in
20 methylation level between a subject of one cohort and a subject of another cohort.

15. The method of claim 1, wherein the set of one or more sites satisfies one or more criteria including (1) a difference in methylation level among multiple subjects of a same cohort and/or (2) a difference in methylation level between a subject of one cohort and a subject of another cohort.

25 16. The method of claim 1, wherein comparing the one or more mixture methylation levels to the one or more reference methylation levels determined from at least two cohorts of other subjects comprises:

inputting the one or more mixture methylation levels into a machine learning model that was trained using the one or more reference methylation levels determined from the at least two cohorts of other subjects.

17. The method of claim 1, further comprising:

5 for each site of the set of one or more sites:

determining a respective number of DNA molecules that are methylated at the site, thereby determining the one or more amounts of the plurality of cell-free DNA molecules methylated at the set of one or more sites of the particular viral genome.

18. The method of claim 17, further comprising:

10 performing methylation-aware sequencing of the plurality of cell-free DNA molecules to obtain sequence reads; and

aligning the sequence reads to the particular viral genome to determine the respective number of DNA molecules that are methylated at each site of the set of one or more sites.

19. The method of claim 1, further comprising:

15 performing a methylation-aware assay of the plurality of cell-free DNA molecules as part of determining locations of the plurality of cell-free DNA molecules and whether the plurality of cell-free DNA molecules are methylated at the set of one or more sites.

20. The method of claim 1, wherein identifying the location of the cell-free

20 DNA molecule includes determining that the location corresponds to one of the set of one or more sites.

21. The method of claim 1, wherein a group of the plurality of cell-free DNA molecules are analyzed collectively to determine the one or more amounts of the plurality of cell-free DNA molecules methylated at the set of one or more sites of the particular viral

25 genome.

22. The method of claim 1, wherein the plurality of cell-free DNA molecules includes at least 10 cell-free DNA molecules located in the particular viral genome.

23. The method of claim 1, wherein the particular viral genome corresponds to Epstein-Barr virus, human papillomavirus, or hepatitis B virus.

24. The method of claim 1, further comprising:

5 determining an amount of cell-free DNA molecules derived from the particular viral genome in a sample, the sample being the biological sample or a different sample including a mixture of cell-free DNA molecules from the genome of the subject and from the one or more other genomes;

comparing the amount to a reference value determined from the at least two cohorts of other subjects;

10 determining a third classification of whether the subject has the first condition based on the comparing of the amount to the reference value; and

determining the final classification using the first classification, the second classification, and the third classification.

25. The method of claim 1, further comprising:

15 responsive to the first classification being that the subject has the first condition, providing a treatment to the subject to improve the first condition.

26. A computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform operations of any of the methods above.

20 27. A system comprising:
the computer product of claim 26; and
one or more processors for executing instructions stored on the computer readable medium.

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

25 29. A system configured to perform any of the above methods.

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