genome of the subject and from one or more viral 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;; 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 viral genomes; determining a statistical value of the size distribution; performing a comparison, comprising: 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 wherein the one or more reference methylation levels depend on the statistical value of the size distribution; or comparing the statistical value to a reference size value determined from the at least two cohorts of other subjects, wherein the reference size value depends on a methylation level measured for the biological sample; and determining a classification of whether the subject has the first condition based on the comparison, wherein the first condition is a cancer, and the virus is a cancer-causing virus.

[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 performing 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.

[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

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 viral 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;

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 viral genomes;

determining a statistical value of the size distribution;

performing a comparison, comprising:

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 wherein the one or more reference methylation levels depend on the statistical value of the size distribution; or

comparing the statistical value to a reference size value determined from the at least two cohorts of other subjects, wherein the reference size value depends on a methylation level measured for the biological sample; and

determining a classification of whether the subject has the first condition based on the comparison, wherein the first condition is a cancer, and the virus is a cancer-causing virus.

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:

determining a second 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 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 wherein the second condition is infectious mononucleosis.

5. The method of any one of claims 1-4, wherein the first classification is that the subject does not have the first condition.

6. The method of any one of claims 1-5, wherein determining the first classification includes determining a level of the first condition.

7. The method of any one of claim 1-6, 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 any one of claims 1-7, 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.

9. The method of any one of claims 1-8, 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.

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.

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 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 any one of claims 1-13, 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 methylation level between a subject of one cohort and a subject of another cohort.

15. The method of any one of claims 1-13, 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.

16. The method of any one of claims 1-15, 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 any one of claims 1-16, further comprising:

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:

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 any one of claims 1-18, further comprising:

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 any one of claims 1-19, wherein identifying the location of the cell-free DNA molecule includes determining that the location corresponds to one of the set of one or more sites.

21. The method of any one of claims 1-20, 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 genome.

22. The method of any one of claims 1-21, 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 any one of claims 1-22, wherein the particular viral genome corresponds to Epstein-Barr virus, human papillomavirus, or hepatitis B virus.

24. The method of any one of claims 1-23, further comprising:

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 viral genomes;

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

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

25. The method of any one of claims 1-24, further comprising:

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. The method of any one of claims 1-25, wherein the first condition is nasopharyngeal carcinoma (NPC), natural killer (NK) T-cell lymphoma, Hodgkin lymphoma, gastric cancer, hepatocellular carcinoma (HCC), cervical cancer (CC), or head and neck squamous cell carcinoma (HNSCC).

27. The method of any one of claims 1-26, wherein the virus is Epstein-Barr virus (EBV), hepatitis C virus (HCV), human papillomavirus (HPV), or hepatitis B virus (HBV).

28. 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.

29. A system comprising:

the computer product of claim 28; and

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

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

31. A system configured to perform any of the above methods.

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