

CLAIMS

1. A method comprising:

(a) obtaining, by a computer system, sequencing data derived from a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;

(b) processing, by the computer system, the sequencing data to identify one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a reference genomic sequence, wherein each of the plurality of phased variants are separated from each other by no more than 170 nucleotides, wherein at least 10% of the one or more cell-free nucleic acid molecules comprises a first phased variant of the plurality of phased variants and a second phased variant of the plurality of phased variants that are separated by at least one nucleotide; and

(c) analyzing, by the computer system, the identified one or more cell-free nucleic acid molecules to determine a condition of the subject, wherein the condition comprises a cancer which is not a hematological malignancy.

2. The method of claim 1, further comprising separating, in silico, (i) at least a portion of the identified one or more cell-free nucleic acid molecules from (ii) one or more other cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules that are not identified to comprise the plurality of phased variants.

3. The method of claim 2, further comprising analyzing, by the computer system, (i) and (ii) as different variables.

4. The method of any one of claims 1–3, wherein the analyzing of the identified one or more cell-free nucleic acid molecules is not based on other cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules that are not identified to comprise the plurality of phased variants.

5. The method of any one of claims 1–4, wherein the sequencing data has been obtained without in silico removal or suppression of (i) background error or (ii) sequencing error.

6. The method of any one of claims 1–5, wherein a number of the plurality of phased

variants from the identified one or more cell-free nucleic acid molecules is indicative of the condition of the subject.

7. The method of claim 6, wherein a ratio of (i) the number of the plurality of phased variants from the identified one or more cell-free nucleic acid molecules and (ii) a number of single nucleotide variants (SNVs) from the identified one or more cell-free nucleic acid molecules is indicative of the condition of the subject.

8. The method of any one of claims 1–7, wherein a frequency of the plurality of phased variants in the identified one or more cell-free nucleic acid molecules is indicative of the condition of the subject.

9. The method of claim 1, wherein the at least 10% of the one or more cell-free nucleic acid molecules comprises at least 50% of the one or more cell-free nucleic acid molecules.

10. The method of claim 9, wherein the at least 10% of the one or more cell-free nucleic acid molecules comprises at least 100% of the one or more cell-free nucleic acid molecules.

11. The method of any one of claims 1–10, wherein the first and second phased variants are separated by at least 2 nucleotides.

12. The method of any one of claims 1–11, wherein the first phased variant and the second phased variant are separated by at most 160 nucleotides.

13. The method of any one of claims 1–12, wherein the reference genome sequence is derived from a reference cohort and comprises at least a portion of hg19 human genome, hg18 genome, hg17 genome, hg16 genome, or hg38 genome.

14. The method of any one of claims 1–12, wherein the reference genomic sequence is derived from a sample of the subject.

15. The method of any one of claims 1–12, wherein the reference genomic sequence is derived from a healthy cell of the subject.

16. The method of any one of claims 1–15, wherein the plurality of cell-free nucleic acid molecules comprise a plurality of cell-free deoxyribonucleic acid (DNA) molecules.

17. The method of any one of claims 1–16, wherein the condition comprises a solid tumor.
18. The method of any one of claims 1–17, wherein the cancer is not a lymphoma.
19. The method of any one of claims 1-18, wherein the cancer is a lung cancer, a colorectal cancer, a gastrointestinal cancer, an esophageal cancer or a breast cancer.
20. The method of any one of claims 1–19, further comprising determining, by the computer, that the subject has the condition or determining a degree or status of the condition of the subject, based on the identified one or more cell-free nucleic acid molecules comprising the plurality of phased variants, wherein the identified one or more cell-free nucleic acid molecules are derived from a sample associated with the condition, based on performing a statistical model analysis of the identified one or more cell-free nucleic acid molecules.
21. The method of any one of claims 1 to 20, wherein the plurality of cell-free nucleic acid molecules from which sequencing data is derived are captured with a set of nucleic acid probes, wherein the set of nucleic acid probes is configured to hybridize to at least a portion of cell-free nucleic acid molecules comprising one or more genomic regions associated with the condition.
22. The method of claim 21, wherein the set of nucleic acid probes is designed based on the plurality of phased variants that are identified by comparing (i) sequencing data from a tumor sample of the subject and (ii) sequencing data from a healthy cell of the subject or a healthy cohort.
23. A method comprising:
 - (a) obtaining, by a computer system, sequencing data derived from a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;
 - (b) processing, by the computer system, the sequencing data to identify one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a reference genomic sequence, wherein each of the plurality of phased variants are separated from each other by no more than 170 nucleotides, wherein at least 10% of the one or more cell-free nucleic acid molecules comprises a first phased variant of the plurality of phased

variants and a second phased variant of the plurality of phased variants that are separated by at least one nucleotide; and

(c) analyzing, by the computer system, the identified one or more cell-free nucleic acid molecules to determine a condition of the subject.

24. A method comprising:

(a) obtaining, by a computer system, sequencing data derived from a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;

(b) processing, by the computer system, the sequencing data to identify one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a reference genomic sequence that are separated by at least one nucleotide, and wherein each of the plurality of phased variants are separated from each other by no more than 170 nucleotides; and

(c) analyzing, by the computer system, the identified one or more cell-free nucleic acid molecules to determine a condition of the subject.

25. A computer program product comprising a non-transitory computer-readable medium having computer-executable code encoded therein, the computer-executable code adapted to be executed to implement a method according to any one of the preceding claims.

26. A system comprising one or more computer processors and computer memory coupled thereto, wherein the computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements a method according to any one of the preceding claims.

27. A method comprising:

a. obtaining, by a computer system, sequencing data for at least 1,000 cell-free DNA molecules from a subject;

b. processing, by the computer system, the sequencing data to identify one or more cell-free DNA molecules of the at least 1,000 cell-free DNA molecules, wherein identifying the one or more cell-free DNA molecules comprises aligning reads corresponding to each of the at

least 1,000 cell-free DNA molecules to a reference genome, wherein each of the one or more cell-free DNA molecules comprises a plurality of phased variants relative to a sequence from the reference genome, wherein each of the plurality of phased variants are separated from each other by no more than 170 nucleotides, wherein at least 10% of the one or more cell-free DNA molecules comprises a first phased variant of the plurality of phased variants and a second phased variant of the plurality of phased variants that are separated by at least one nucleotide; and

c. analyzing, by the computer system, the identified one or more cell-free nucleic acid molecules to determine a cancer or a cancer state of the subject.