

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:

(a) obtaining, by a computer system, sequencing data derived from a plurality of cell-free nucleic acid molecules that ~~is~~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 at least about 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 about 10% of the one or more cell-free nucleic acid molecules comprises at least about 50% of the one or more cell-free nucleic acid molecules.

2- 10. The method of claim 1, wherein the at least about 10% of the one or more cell-free nucleic acid molecules comprises at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or about 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 about 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 of monitoring a progress of a condition of a subject, the method comprising:

_____ (a) determining a first state of the condition of the subject based on identification of a first set of one or more cell-free nucleic acid molecules from a first plurality of cell-free nucleic acid molecules that has been obtained or derived from the subject;

_____ (b) determining a second state of the condition of the subject based on identification of a second set of one or more cell-free nucleic acid molecules from a second plurality of cell-free nucleic acid molecules that has been obtained or derived from the subject,

_____ wherein the second plurality of cell-free nucleic acid molecules are obtained from the subject subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject; and

_____ (c) determining the progress of the condition based on the first state of the condition and the second state of the condition,

_____ 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;

_____ wherein the one or more cell-free nucleic acid molecules are captured from among the plurality of cell-free nucleic acid molecules 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;

_____ wherein the set of nucleic acid probes are designed to hybridize to at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or about 100% of (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.

24. The method of claim 23, wherein the progress of the condition is worsening of the condition.

25. The method of claim 23, wherein the progress of the condition is at least a partial remission of the condition.

26. The method of any one of claims 23-25, wherein a presence of the plurality of phased variants is indicative of the first state or the second state of the condition of the subject.

27. The method of any one of claims 23-26, wherein the second plurality of cell-free nucleic acid molecules has been obtained from the subject at least about 1 week, at least about 2 weeks, at least about 3 weeks, at least about 4 weeks, at least about 2 months, or at least about 3 months subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject.

28. The method of any one of claims 23-27, wherein the subject has been subjected to a treatment for the condition (i) prior to obtaining the second plurality of cell-free nucleic acid molecules from the subject and (ii) subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject.

29. The method of any one of claims 23-28, wherein the progress of the condition is indicative of minimal residual disease of the condition of the subject.

30. The method of any one of claims 23-29, wherein the progress of the condition is indicative of tumor burden or cancer burden of the subject.

31. The method of any one of claims 23-30, wherein at least 50% of the one or more cell-free nucleic acid molecules comprising a plurality of phased variants comprises a single nucleotide variant (SNV) that is at least 2 nucleotides away from an adjacent SNV.

32. The method of any one of claims 23-31, wherein the reference genomic sequence is derived from a sample of the subject.

33. The method of claim 32, wherein the sample is a healthy sample.

34. The method of any one of claims 23-33, 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 solid tumor, lymphoma, or blood tumor of the subject and (ii) sequencing data from a healthy cell of the subject or a healthy cohort.

35. The method of any one of claims 23-34, wherein the subject has been subjected to a treatment for the condition prior to (a).

36. The method of any one of claims 23-35, wherein the treatment comprises chemotherapy, radiotherapy, chemoradiotherapy, immunotherapy, adoptive cell therapy, hormone therapy, targeted drug therapy, surgery, transplant, or transfusion.

37. The method of any one of claims 23-36, wherein the plurality of cell-free nucleic acid molecules are derived from a bodily sample of the subject.

38. The method of claim 37, wherein the bodily sample comprises plasma, serum, or blood.

39. The method of any one of claims 23-38, wherein the subject is a mammal.

40. The method of any one of claims 23-39, wherein the subject is a human.

41. The method of any one of claims 23-40, wherein the condition comprises neoplasm, cancer, or tumor.

42. The method of claim 41, wherein the condition comprises a solid tumor.

43. The method of claim 42, wherein the condition comprises a lymphoma.

44. The method of claim 43, wherein the condition comprises a B-cell lymphoma.

45. The method of claim 44, wherein the condition comprises a sub-type of B-cell lymphoma selected from the group consisting of diffuse large B-cell lymphoma, follicular lymphoma, Burkitt lymphoma, and B-cell chronic lymphocytic leukemia.

46. 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 at least about 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.

3- 47. A method comprising:

(a) obtaining, by a computer system, sequencing data derived from a plurality of cell-free nucleic acid molecules that ~~is~~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

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

4- 48. A method comprising:

(a) obtaining sequencing data derived from a plurality of cell-free nucleic acid molecules that ~~is~~has been obtained or derived from a subject;

(b) processing the sequencing data to identify one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules with a limit of detection of less than about 1 out of 50,000 observations from the sequencing data; and

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

5. — ~~The method of claim 4, wherein the limit of detection of the identification step is less than about 1 out of 100,000, less than about 1 out of 500,000, less than about 1 out of 1,000,000, less than about 1 out of 1,500,000, or less than about 1 out of 2,000,000 observations from the sequencing data.~~

6. — ~~The method of any one of claims 4-5, wherein each of the one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a reference genomic sequence.~~

7. — ~~The method of claim 6, wherein a first phased variant of the plurality of phased variants and a second phased variant of the plurality of phased variants are separated by at least one nucleotide.~~

8. — ~~The method of any one of claims 4-7, wherein (a) to (c) are performed by a computer system.~~

9. — ~~The method of any one of preceding claims, wherein the sequencing data is generated based on nucleic acid amplification.~~

10. — ~~The method of any one of the preceding claims, wherein the sequencing data is generated based on polymerase chain reaction.~~

11. — ~~The method of any one of the preceding claims, wherein the sequencing data is generated based on amplicon sequencing.~~

12. — ~~The method of any one of the preceding claims, wherein the sequencing data is generated based on next-generation sequencing (NGS).~~

13. — ~~The method of any one of the preceding claims, wherein the sequencing data is generated based on non-hybridization-based NGS.~~

14. — ~~The method of any one of the preceding claims, wherein the sequencing data is generated~~

~~without use of molecular barcoding of at least a portion of the plurality of cell-free nucleic acid molecules.~~

~~15. The method of any one of the preceding claims, wherein the sequencing data is obtained without use of sample barcoding of at least a portion of the plurality of cell-free nucleic acid molecules.~~

~~16. The method of any one of the preceding claims, wherein the sequencing data is obtained without in silico removal or suppression of (i) background error or (ii) sequencing error.~~

~~17. A method of treating a condition of a subject, the method comprising:~~

~~—— (a) identifying the subject for treatment of the condition, wherein the subject has been determined to have the condition based on identification of one or more cell-free nucleic acid molecules from a plurality of cell-free nucleic acid molecules that is obtained or derived from the subject;~~

~~—— wherein each of the one or more cell-free nucleic acid molecules identified comprises a plurality of phased variants relative to a reference genomic sequence that are separated by at least one nucleotide, and~~

~~—— wherein a presence of the plurality of phased variants is indicative of the condition of the subject; and~~

~~—— (b) subjecting the subject to the treatment based on the identification in (a).~~

~~18.~~ 49. A method of monitoring a progress of a condition of a subject, the method comprising:

(a) determining a first state of the condition of the subject based on identification of a first set of one or more cell-free nucleic acid molecules from a first plurality of cell-free nucleic acid molecules that is has been obtained or derived from the subject;

(b) determining a second state of the condition of the subject based on identification of a second set of one or more cell-free nucleic acid molecules from a second plurality of cell-free nucleic acid molecules that is has been obtained or derived from the subject,

wherein the second plurality of cell-free nucleic acid molecules are obtained from

the subject subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject; and

(c) determining the progress of the condition based on the first state of the condition and the second state of the condition,

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.

~~19.—The method of claim 18, wherein the progress of the condition is worsening of the condition.~~

~~20.—The method of claim 18, wherein the progress of the condition is at least a partial remission of the condition.~~

~~21.—The method of any one of claims 18-20, wherein a presence of the plurality of phased variants is indicative of the first state or the second state of the condition of the subject.~~

~~22.—The method of any one of claims 18-21, wherein the second plurality of cell-free nucleic acid molecules is obtained from the subject at least about 1 week, at least about 2 weeks, at least about 3 weeks, at least about 4 weeks, at least about 2 months, or at least about 3 months subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject.~~

~~23.—The method of any one of claims 18-22, wherein the subject is subjected to a treatment for the condition (i) prior to obtaining the second plurality of cell-free nucleic acid molecules from the subject and (ii) subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject.~~

~~24.—The method of any one of claims 18-23, wherein the progress of the condition is indicative of minimal residual disease of the condition of the subject.~~

~~25.—The method of any one of claims 18-24, wherein the progress of the condition is indicative~~

~~of tumor burden or cancer burden of the subject.~~

~~26. The method of any one of the preceding claims, wherein the one or more cell free nucleic acid molecules are captured from among the plurality of cell free nucleic acid molecules 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.~~

~~27. A method comprising:~~

~~—— (a) providing a mixture comprising (1) a set of nucleic acid probes and (2) a plurality of cell free nucleic acid molecules that is obtained or derived from a subject,~~

~~———— wherein an individual nucleic acid probe of the set of nucleic acid probes is designed to hybridize to at least a portion of a target cell free nucleic acid molecule comprising a plurality of phased variants relative to a reference genomic sequence that are separated by at least one nucleotide, and~~

~~———— wherein the individual nucleic acid probe comprises an activatable reporter agent, activation of the activatable reporter agent being selected from the group consisting of: (i) hybridization of the individual nucleic acid probe to the plurality of phased variants and (ii) dehybridization of at least a portion of the individual nucleic acid probe that has been hybridized to the plurality of phased variants;~~

~~—— (b) detecting the activatable reporter agent that is activated, 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 the plurality of phased variants; and~~

~~—— (c) analyzing the identified one or more cell free nucleic acid molecules to determine a condition of the subject.~~

~~28. A method comprising:~~

~~—— (a) providing a mixture comprising (1) a set of nucleic acid probes and (2) a plurality of cell free nucleic acid molecules that is obtained or derived from a subject,~~

~~———— wherein an individual nucleic acid probe of the set of nucleic acid probes is designed to hybridize to at least a portion of a target cell free nucleic acid molecule comprising a~~

plurality of phased variants relative to a reference genomic sequence, and

—wherein the individual nucleic acid probe comprises an activatable reporter agent, activation of the activatable reporter agent being selected from the group consisting of: (i) hybridization of the individual nucleic acid probe to the plurality of phased variants and (ii) dehybridization of at least a portion of the individual nucleic acid probe that has been hybridized to the plurality of phased variants;

—(b) detecting the activatable reporter agent that is activated, 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 the plurality of phased variants, wherein a limit of detection of the identification step is less than about 1 out of 50,000 cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules; and

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

29. —The method of claim 28, wherein the limit of detection of the identification step is less than about 1 out of 100,000, less than about 1 out of 500,000, less than about 1 out of 1,000,000, less than about 1 out of 1,500,000, or less than about 1 out of 2,000,000 cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules.

30. —The method of claim 28 or 29, wherein a first phased variant of the plurality of phased variants and a second phased variant of the plurality of phased variants are separated by at least one nucleotide.

31. —The method of any one of claims 27-30, wherein the activatable reporter agent is activated upon hybridization of the individual nucleic acid probe to the plurality of phased variants.

32. —The method of any one of claims 27-30, wherein the activatable reporter agent is activated upon dehybridization of at least a portion of the individual nucleic acid probe that has been hybridized to the plurality of phased variants.

33. —The method of any one of claims 27-32, further comprising mixing (1) the set of nucleic

acid probes and (2) the plurality of cell-free nucleic acid molecules.

34. — The method of any one of claims 27-33, wherein the activatable reporter agent is a fluorophore.

35. — The method of any one of the preceding claims, wherein analyzing the identified one or more cell-free nucleic acid molecules comprises analyzing (i) the identified one or more cell-free nucleic acid molecules and (ii) other cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules that do not comprise the plurality of phased variants as different variables.

36. — The method of any one of the preceding claims, 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 do not comprise the plurality of phased variants.

37. — The method of any one of the preceding claims, 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.

38. — The method of claim 37, wherein a ratio of (i) the number of the plurality of phased variants from the one or more cell-free nucleic acid molecules and (ii) a number of single-nucleotide variants (SNVs) from the one or more cell-free nucleic acid molecules is indicative of the condition of the subject.

39. — The method of any one of the preceding claims, 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.

40. — The method of claim 39, wherein the frequency is indicative of a diseased cell associated with the condition.

41. — The method of claim 40, wherein the condition is diffuse large B-cell lymphoma, and wherein the frequency is indicative of whether the one or more cell-free nucleic acid molecules are derived from germinal-center B-cell (GCB) or activated B-cell (ABC).

42. — The method of any one of the preceding claims, wherein genomic origin of the identified one or more cell-free nucleic acid molecules is indicative of the condition of the subject.

43. — The method of any one of the preceding claims, wherein the first and second phased variants are separated by at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 nucleotides.

44. — The method of any one of the preceding claims, wherein the first and second phased variants are separated by at most about 180, at most about 170, at most about 160, at most about 150, or at most about 140 nucleotides.

45. — The method of any one of the preceding claims, wherein at least about 10%, at least about 20%, at least about 30%, at least about 40%, or at least about 50% of the one or more cell-free nucleic acid molecules comprising a plurality of phased variants comprises a single-nucleotide variant (SNV) that is at least 2 nucleotides away from an adjacent SNV.

46. — The method of any one of the preceding claims, wherein the plurality of phased variants comprises at least 3, at least 4, at least 5, at least 10, at least 15, at least 20, or at least 25 phased variants within the same cell-free nucleic acid molecule.

47. — The method of any one of the preceding claims, wherein the one or more cell-free nucleic acid molecules identified comprises at least 2, at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, or at least 1,000 cell-free nucleic acid molecules.

48. — The method of any one of the preceding claims, wherein the reference genomic sequence is derived from a reference cohort.

49. — The method of claim 48, wherein the reference genomic sequence comprises a consensus sequence from the reference cohort.

50. — The method of claim 48, wherein the reference genomic sequence comprises at least a portion of hg19 human genome, hg18 genome, hg17 genome, hg16 genome, or hg38 genome.

51. — The method of any one of the preceding claims, wherein the reference genomic sequence is derived from a sample of the subject.

52. — The method of claim 51, wherein the sample is a healthy sample.

53. — The method of claim 52, wherein the sample comprises a healthy cell.

54. — The method of claim 53, wherein the healthy cell comprises a healthy leukocyte.

55. — The method of claim 51, wherein the sample is a diseased sample.

56. — The method of claim 55, wherein the diseased sample comprises a diseased cell.

57. — The method of claim 56, wherein the diseased cell comprises a tumor cell.

58. — The method of claim 55, wherein the diseased sample comprises a solid tumor.

59. — The method of any one of the preceding claims, 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 solid tumor, lymphoma, or blood tumor of the subject and (ii) sequencing data from a healthy cell of the subject or a healthy cohort.

60. — The method of claim 59, wherein the healthy cell is from the subject.

61. — The method of claim 59, wherein the healthy cell is from the healthy cohort.

62. — ~~The method of any one of the preceding claims, wherein the set of nucleic acid probes are designed to hybridize to at least a portion of sequences of genomic loci associated with the condition.~~

63. — ~~The method of claim 62, wherein the genomic loci associated with the condition are known to exhibit aberrant somatic hypermutation when the subject has the condition.~~

64. — ~~The method of any one of the preceding claims, wherein the set of nucleic acid probes are designed to hybridize to at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or about 100% of (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.~~

65. — ~~The method of any one of the preceding claims, wherein each nucleic acid probe of the set of nucleic acid probes has at least about 70%, at least about 80%, at least about 90% sequence identity, at least about 95% sequence identity, or about 100% sequence identity to a probe sequence selected from Table 6.~~

66. — ~~The method of any one of the preceding claims, wherein the set of nucleic acid probes comprises at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, or at least about 90% of probe sequences in Table 6.~~

67. — ~~The method of any one of the preceding claims, further comprising determining 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.~~

68. — ~~The method of claim 67, further comprising determining that the 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.~~

~~69.—The method of claim 68, wherein the statistical model analysis comprises a Monte Carlo statistical analysis.~~

~~70.—The method of any one of the preceding claims, further comprising monitoring a progress of the condition of the subject based on the identified one or more cell free nucleic acid molecules.~~

~~71.—The method of any one of the preceding claims, further comprising performing a different procedure to confirm the condition of the subject.~~

~~72.—The method of claim 71, wherein the different procedure comprises a blood test, genetic test, medical imaging, physical exam, or tissue biopsy.~~

~~73.—The method of any one of the preceding claims, further comprising determining a treatment for the condition of the subject based on the identified one or more cell free nucleic acid molecules.~~

~~74.—The method of any one of the preceding claims, wherein the subject has been subjected to a treatment for the condition prior to (a).~~

~~75.—The method of any one of the preceding claims, wherein the treatment comprises chemotherapy, radiotherapy, chemoradiotherapy, immunotherapy, adoptive cell therapy, hormone therapy, targeted drug therapy, surgery, transplant, transfusion, or medical surveillance.~~

~~76.—The method of any one of the preceding claims, wherein the plurality of cell free nucleic acid molecules comprise a plurality of cell free deoxyribonucleic acid (DNA) molecules.~~

~~77.—The method of any one of the preceding claims, wherein the condition comprises a disease.~~

~~78. The method of any one of the preceding claims, wherein the plurality of cell-free nucleic acid molecules are derived from a bodily sample of the subject.~~

~~79. The method of claim 78, wherein the bodily sample comprises plasma, serum, blood, cerebrospinal fluid, lymph fluid, saliva, urine, or stool.~~

~~80. The method of any one of the preceding claims, wherein the subject is a mammal.~~

~~81. The method of any one of the preceding claims, wherein the subject is a human.~~

~~82. The method of any one of the preceding claims, wherein the condition comprises neoplasm, cancer, or tumor.~~

~~83. The method of claim 82, wherein the condition comprises a solid tumor.~~

~~84. The method of claim 82, wherein the condition comprises a lymphoma.~~

~~85. The method of claim 84, wherein the condition comprises a B-cell lymphoma.~~

~~86. The method of claim 85, wherein the condition comprises a sub-type of B-cell lymphoma selected from the group consisting of diffuse large B-cell lymphoma, follicular lymphoma, Burkitt lymphoma, and B-cell chronic lymphocytic leukemia.~~

~~87. The method of any one of the preceding claims, wherein the plurality of phased variants have been previously identified as tumor-derived from sequencing a prior tumor sample or cell-free nucleic acid sample.~~

~~88.~~ 50. A composition comprising a bait set comprising a set of nucleic acid probes designed to capture cell-free DNA molecules derived from at least about 5% of genomic regions set forth in (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.

89.— ~~The composition of claim 88, wherein the set of nucleic acid probes are designed to pull down cell free DNA molecules derived from at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or about 100% of the genomic regions set forth in (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.~~

90.— ~~The composition of any one of claims 88-89, wherein the set of nucleic acid probes are designed to capture the one or more cell free DNA molecules derived from at most about 10%, at most about 20%, at most about 30%, at most about 40%, at most about 50%, at most about 60%, at most about 70%, at most about 80%, at most about 90%, or about 100% of the genomic regions set forth in (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.~~

91.— ~~The composition of any one of claims 88-90, wherein the bait set comprises at most 5, at most 10, at most 50, at most 100, at most 500, at most 1000, or at most 2000 nucleic acid probes.~~

92.— ~~The composition of any one of claims 88-91, wherein an individual nucleic acid probe of the set of nucleic acid probes comprises a pull down tag.~~

93.— ~~The composition of any one of claims 88-92, wherein the pull down tag comprises a nucleic acid barcode.~~

94.— ~~The composition of any one of claims 88-93, wherein the pull down tag comprises biotin.~~

95.— ~~The composition of any one of claims 88-94, wherein each of the cell free DNA molecules is between about 100 nucleotides and about 180 nucleotides in length.~~

96.— ~~The composition of any one of claims 88-95, wherein the genomic regions are associated with a condition.~~

~~97. — The composition of any one of claims 88-96, wherein the genomic regions exhibit aberrant somatic hypermutation when a subject has the condition.~~

~~98. — The composition of any one of claims 88-97, wherein the condition comprises a B-cell lymphoma.~~

~~99. — The composition of claim 98, wherein the condition comprises a sub-type of B-cell lymphoma selected from the group consisting of diffuse large B-cell lymphoma, follicular lymphoma, Burkitt lymphoma, and B-cell chronic lymphocytic leukemia.~~

~~100. — The composition of any one of claims 88-99, further comprising a plurality of cell-free DNA molecules obtained or derived from a subject.~~

~~101. — A method to perform a clinical procedure on an individual, the method comprising:~~

~~—— obtaining or having obtained a targeted sequencing result of a collection of cell-free nucleic acid molecules,~~

~~—— wherein the collection of cell-free nucleic acid molecules are sourced from a liquid or waste biopsy of an individual, and~~

~~—— wherein the targeting sequencing is performed utilizing nucleic acid probes to pull down sequences of genomic loci known to experience aberrant somatic hypermutation in a B-cell cancer;~~

~~—— identifying or having identified a plurality of variants in phase within the cell-free nucleic acid sequencing result;~~

~~—— determining or having determined, utilizing a statistical model and the identified phased variants, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm; and~~

~~—— performing a clinical procedure on the individual to confirm the presence of the B-cell cancer, based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences likely derived from the B-cell cancer.~~

~~102.— The method of claim 101, wherein the biopsy is one of blood, serum, cerebrospinal fluid, lymph fluid, urine, or stool.~~

~~103.— The method of claim 101, wherein the genomic loci are selected from (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.~~

~~104.— The method of claim 101, wherein the sequences of the nucleic acid probes are selected from Table 6.~~

~~105.— The method of claim 101, wherein the clinical procedure is a blood test, medical imaging, or a physical exam.~~

~~106.— A method to treat an individual for a B-cell cancer, the method comprising:
—— obtaining or having obtained a targeted sequencing result of a collection of cell-free nucleic acid molecules;
—— wherein the collection of cell-free nucleic acid molecules are sourced from a liquid or waste biopsy of an individual, and
—— wherein the targeting sequencing is performed utilizing nucleic acid probes to pull down sequences of genomic loci known to experience aberrant somatic hypermutation in a B-cell cancer;
—— identifying or having identified a plurality of variants in phase within the cell-free nucleic acid sequencing result;
—— determining or having determined, utilizing a statistical model and the identified phased variants, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm; and
—— treating the individual to curtail the B-cell cancer, based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from the B-cell cancer.~~

~~107.— The method of claim 106, wherein the biopsy is one of blood, serum, cerebrospinal fluid, lymph fluid, urine or stool.~~

~~108.— The method of claim 106, wherein the genomic loci are selected from (i) the genomic regions identified in Table 1, (ii) the genomic regions identified in Table 3, or (iii) the genomic regions identified to have a plurality of phased variants in Table 3.~~

~~109.— The method of claim 106, wherein the sequences of the nucleic acid probes are selected from Table 6.~~

~~110.— The method of claim 106, wherein the treatment is chemotherapy, radiotherapy, immunotherapy, hormone therapy, targeted drug therapy, or medical surveillance.~~

~~111.— A method to detect cancerous minimal residual disease in an individual and to treat the individual for a cancer, the method comprising:~~

~~—— obtaining or having obtained a targeted sequencing result of a collection of cell free nucleic acid molecules,~~

~~—— wherein the collection of cell free nucleic acid molecules are sourced from a liquid or waste biopsy of an individual,~~

~~—— wherein the liquid or waste biopsy is sourced after a series of treatments in order to detect minimal residual disease, and~~

~~—— wherein the targeting sequencing is performed utilizing nucleic acid probes to pull down sequences of genomic loci determined to contain a plurality of variants in phase, as determined by a prior sequencing result on a prior biopsy derived from the cancer;~~

~~—— identifying or having identified at least one set of the plurality of variants in phase within the cell free nucleic acid sequencing result; and~~

~~—— treating the individual to curtail the cancer, based upon determining that the cell free nucleic acid sequencing result contains nucleic acid sequences derived from the cancer.~~

~~112.— The method of claim 111, wherein the liquid or waste biopsy is one of blood, serum, cerebrospinal fluid, lymph fluid, urine or stool.~~

~~113.— The method of claim 111, wherein the treatment is chemotherapy, radiotherapy,~~

~~immunotherapy, hormone therapy, targeted drug therapy, or medical surveillance.~~

~~414.~~ 51. 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.

~~415.~~ 52. 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.

53. A method for preparing an enriched library of nucleic acids for sequencing, the method comprising:

a. identifying a patient-specific set of six or more genomic regions that have been shown to harbor a plurality of phased variants that are within 170 base pairs of each other;

b. obtaining a personalized set of nucleic acid probes that have been designed to enrich for phased variant-containing nucleic acids from the six or more genomic regions;

c. preparing cell-free nucleic acids from the patient for sequencing; and

d. selectively enriching the prepared cell-free nucleic acids from the patient for phased variant-containing nucleic acids from the six or more genomic regions via hybridization capture or amplification to form an enriched library of nucleic acids for sequencing;

e. wherein identifying the patient-specific set of six or more genomic regions that have been shown to harbor a plurality of phased variants comprises comparing sequencing data from a either a diseased sample from the patient or a cell-free DNA sample from the patient to sequencing data from a non-diseased sample from the patient;

f. wherein at least 10% of phased variant-containing nucleic acids from the library of nucleic acids have a first phased variant and a second phased variant that are separated from each other by at least one nucleotide.

54. A method for preparing an enriched library of nucleic acids for sequencing, the method comprising:

_____ a. identifying a patient-specific set of three or more genomic regions that have been shown to harbor a plurality of phased variants that are within 170 base pairs of each other;
_____ b. preparing cell-free nucleic acids from the patient for sequencing; and
_____ c. selectively enriching the prepared cell-free nucleic acids from the patient for phased variant-containing nucleic acids from the three or more genomic regions to form an enriched library of nucleic acids for sequencing;
_____ d. wherein at least 10% of phased variant-containing nucleic acids from the enriched library of nucleic acids have a first phased variant and a second phased variant that are separated from each other by at least one nucleotide.

55. A method for detection or monitoring of a disease, the method comprising:

_____ a. identifying a patient-specific set of three or more genomic regions that have been shown to harbor a plurality of phased variants that are within 170 base pairs of each other;
_____ b. preparing cell-free nucleic acids from the patient for sequencing;
_____ c. selectively enriching the prepared cell-free nucleic acids from the patient for phased variant-containing nucleic acids from the three or more genomic regions to form an enriched library of nucleic acids for sequencing;
_____ d. sequencing the enriched library of nucleic acids to produce sequencing data; and
_____ e. analyzing the sequencing data to detect or monitor a disease of the patient;
_____ f. wherein at least 10% of phased variant-containing nucleic acids from the library of nucleic acids have a first phased variant and a second phased variant that are separated from each other by at least one nucleotide.

56. A method for identifying a subset of phased variant-containing nucleic acids molecules from a subject, the method comprising:

_____ (a) obtaining, by a computer system, sequencing data derived from a plurality of nucleic acid molecules from a subject, wherein the plurality nucleic acid molecules from the subject are cell-free nucleic acid molecules or nucleic acids molecules from tumor tissue;
_____ (b) processing, by the computer system, the sequencing data to identify a subset of phased-variant containing nucleic acid molecules from the plurality of nucleic acid molecules, wherein each nucleic acid molecule of the subset of phased-variant containing 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, and wherein at least about 10% of the subset of 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, wherein processing the sequence data comprises:

- _____ (i) aligning the sequencing data to the reference genomic sequence;
- _____ (ii) identifying phased variant-containing nucleic acid molecules from the plurality of nucleic acid molecules based on alignment of the sequencing data to the reference genomic sequence; and
- _____ (iii) filtering out phased variant-containing nucleic acid molecules in which at least one variant of the plurality of phased variants is observed in a subject-matched non-tumor sample or a population of non-tumor samples that are not from the subject.

57. A method comprising:

- _____ a. obtaining a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;
- _____ b. enriching the plurality of cell-free nucleic acid molecules for genomic regions associated with a set of phased variants to form a sequencing library, wherein the sequencing library comprises at least 150,000 evaluable sequencing fragments;
- _____ c. obtaining sequencing data from the sequencing library for each of the at least 150,000 evaluable sequencing fragments;
- _____ d. computer processing the sequencing data for each of the at least 150,000 evaluable sequencing fragments to align the sequencing data for each of the at least 150,000 evaluable sequencing fragments to a human reference genome; and
- _____ e. identifying one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the identified one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a sequence from the human reference genome, 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

f. determining (i) the number of evaluable sequencing fragments processed in (d) and (ii) the number of the one or more cell-free nucleic acid molecules that comprise the plurality of phased variants identified in (e);

g. determining a condition of the subject based at least in part on the plurality of phased variants;

h. wherein an evaluable sequencing fragment is a nucleic acid fragment that covers a region of the genome that comprises two or more predefined mutations from the set of phased variants, wherein the predefined mutations occur within 170 bp of each other in genomic space.

58. A method comprising:

a. obtaining a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;

b. enriching the plurality of cell-free nucleic acid molecules for genomic regions associated with at least 5% of the genomic regions set forth in Table 1 to form a sequencing library, wherein the sequencing library comprises at least 150,000 evaluable sequencing fragments;

c. obtaining sequencing data from the sequencing library for each of the at least 150,000 evaluable sequencing fragments;

d. computer processing the sequencing data for each of the at least 150,000 evaluable sequencing fragments to align the sequencing data for each of the at least 150,000 evaluable sequencing fragments to a human reference genome; and

e. identifying one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the identified one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a sequence from the human reference genome, 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

f. determining a condition of the subject based at least in part on the plurality of phased variants;

g. wherein an evaluable sequencing fragment is a nucleic acid fragment that covers a region of the genome that comprises two or more predefined mutations from a set of phased variants, wherein the predefined mutations occur within 170 bp of each other in genomic space.

59. A method comprising:

a. obtaining a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;

b. enriching the plurality of cell-free nucleic acid molecules for genomic regions associated with at least 5% of the genomic regions set forth in Table 2 to form a sequencing library, wherein the sequencing library comprises at least 150,000 evaluable sequencing fragments;

c. obtaining sequencing data from the sequencing library for each of the at least 150,000 evaluable sequencing fragments;

d. computer processing the sequencing data for each of the at least 150,000 evaluable sequencing fragments to align the sequencing data for each of the at least 150,000 evaluable sequencing fragments to a human reference genome; and

e. identifying one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the identified one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a sequence from the human reference genome, 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

f. determining a condition of the subject based at least in part on the plurality of phased variants;

g. wherein an evaluable sequencing fragment is a nucleic acid fragment that covers a region of the genome that comprises two or more predefined mutations from a set of phased variants, wherein the predefined mutations occur within 170 bp of each other in genomic space.

60. A method comprising:

a. obtaining a plurality of cell-free nucleic acid molecules that has been obtained or derived from a subject;

b. enriching the plurality of cell-free nucleic acid molecules for genomic regions associated with at least 5% of the genomic regions set forth in Table 3 to form a sequencing library, wherein the sequencing library comprises at least 150,000 evaluable sequencing fragments;

c. obtaining sequencing data from the sequencing library for each of the at least 150,000

evaluable sequencing fragments;

d. computer processing the sequencing data for each of the at least 150,000 evaluable sequencing fragments to align the sequencing data for each of the at least 150,000 evaluable sequencing fragments to a human reference genome; and

e. identifying one or more cell-free nucleic acid molecules of the plurality of cell-free nucleic acid molecules, wherein each of the identified one or more cell-free nucleic acid molecules comprises a plurality of phased variants relative to a sequence from the human reference genome, 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

f. determining a condition of the subject based at least in part on the plurality of phased variants;

g. wherein an evaluable sequencing fragment is a nucleic acid fragment that covers a region of the genome that comprises two or more predefined mutations from a set of phased variants, wherein the predefined mutations occur within 170 bp of each other in genomic space.

61. 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 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.

62. A method of monitoring a progress of a condition of a subject, the method comprising:

(a) determining a first state of the condition of the subject based on identification of a first set of one or more cell-free nucleic acid molecules from a first plurality of cell-free nucleic acid molecules that has been obtained or derived from the subject;

(b) determining a second state of the condition of the subject based on identification of a second set of one or more cell-free nucleic acid molecules from a second plurality of cell-free nucleic acid molecules that has been obtained or derived from the subject,

wherein the second plurality of cell-free nucleic acid molecules are obtained from the subject subsequent to obtaining the first plurality of cell-free nucleic acid molecules from the subject; and

(c) determining the progress of the condition based on the first state of the condition and the second state of the condition;

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.

63. A method of monitoring progress of a condition, the method comprising:

(a) obtaining, by a computer system, sequencing data for a plurality of DNA molecules obtained from a subject at a first timepoint;

(b) processing, by the computer system, the sequencing data for the plurality of DNA molecules obtained from the subject at the first timepoint to identify and determine genomic locations of a plurality of disease-derived phased-variant containing DNA molecules, wherein the identity and genomic locations of the disease-derived DNA molecules are determined based on alignment of reads corresponding to each of the plurality of DNA molecules to a reference genomic sequence;

(c) obtaining, by the computer system, sequencing data for a plurality of cell-free DNA molecules obtained from the subject at a second timepoint that is subsequent to the first timepoint; and

(d) aligning, by the computer system, the sequencing data for the plurality of cell-free DNA molecules obtained from the subject at the second timepoint to the genomic locations of the plurality of disease-derived phased-variant containing DNA identified in (b) to assess an extent to which the cell-free DNA molecules from the subject at the second timepoint comprise disease-

derived phased-variant containing cell-free DNA molecules.