

CLAIMS

WHAT IS CLAIMED IS:

1. A DNA molecule, comprising:
 - a nucleic acid segment obtained or derived from a biological sample, wherein the nucleic acid segment is DNA;
 - a pair of error-correcting unique identifiers ligated to the nucleic acid segment to produce a ligation product, wherein the pair of error-correcting unique identifiers flanks the nucleic acid segment, wherein each of the pair of error-correcting unique identifiers is a DNA segment, wherein the pair of error-correcting unique identifiers collectively provides a unique identification of the nucleic acid segment against other nucleic acid segments represented in a set of sequencing reads; and
 - a pair of error-correcting dual index sample barcodes attached to the ligation product, wherein each of the pair of error-correcting dual index sample barcodes is a DNA segment, and wherein the pair of error-correcting dual index sample barcodes collectively provides a unique identification of the biological sample against other biological samples represented in a set of sequencing reads.
2. The DNA molecule of claim 1, wherein the nucleic acid segment is complementary DNA (cDNA).
3. The DNA molecule of claim 1 or 2, wherein the nucleic acid segment is obtained or derived from a cell-free DNA sample.
4. The DNA molecule according to any one of claims 1 to 3, wherein the pair of error-correcting dual index sample barcodes flank the ligation product.
5. A collection comprising a plurality of the DNA molecules according to any one of claims 1 to 4.
6. A method to prepare a DNA library for sequencing, the method comprising:
 - ligating onto a plurality of nucleic acid segments pairs of partial Y-adapters to flank each of the plurality of nucleic acid segments by a pair of partial Y-adapters, thereby producing a plurality of ligation products,

wherein each of the plurality of nucleic acid segments is DNA, and
wherein the plurality of nucleic acid segments is obtained or derived from a biological sample,

wherein each of the pair of partial Y-adapters comprises an error-correcting unique identifier and sequences for a primer to anneal in a grafting polymerase chain reaction, and

wherein the pair of error-correcting unique identifiers on each of the plurality of nucleic acid segments collectively provides a unique identification of the nucleic acid segments against other nucleic acid segments in the plurality of nucleic acid segments; and

grafting onto each of the plurality of ligation products a pair of error-correcting dual index sample barcodes to flank the ligation product by the error-correcting dual index sample barcodes, wherein the error-correcting dual index sample barcodes collectively provide a unique identification of the biological sample.

7. The method of claim 6, wherein the nucleic acid segment is complementary DNA (cDNA).

8. The method of claim 6 or 7, wherein the biological sample comprises a cell-free DNA sample.

9. The method according to any one of claims 6 to 8, wherein the error-correcting dual index sample barcodes collectively provide the unique identification of the biological sample against other biological samples represented in the DNA library.

10. A method to detect a neoplasm in an individual, the method comprising:
obtaining or having obtained cell-free nucleic acid sequencing reads of a plurality of cell-free nucleic acid molecules, wherein the plurality of cell-free nucleic acid molecules is obtained or derived from a first bodily sample of the individual;
obtaining or having obtained cell-derived nucleic acid sequencing reads of a plurality of cell-derived nucleic acid molecules, wherein the plurality of cell-derived nucleic acid molecules is obtained or derived from a second bodily sample of the individual;

identifying or having identified single nucleotide variants present in both the cell-free nucleic acid sequencing reads and the cell-derived nucleic acid sequencing reads;

determining or having determined, based at least in part on applying a first computational model to the identified single nucleotide variants, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm,

wherein the first computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a first set of individuals having a neoplasm and a second set of control individuals not having a neoplasm, and

wherein the first computational model integrates one or more of the following features: cell-derived DNA Bayesian background, cell-free DNA Bayesian background, germline depth, short fragment score 1, short fragment score 2, genomic start position and end position of the cell-free DNA molecule, and any combination thereof; and

detecting the neoplasm in the individual based at least in part on the determination that the cell-free nucleic acid sequencing reads contain the nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

11. The method of claim 10, wherein the first computational model further integrates one or more of the following features: transition/transversion, duplex support, pass outlier, mapping quality, cancer hotspot, UMI error corrected, Phred quality score, variant allele frequency (VAF %), mean barcode family size, variant position in a cell-free DNA molecule, polygenic risk score, nuclease motif, and any combination thereof.

12. The method of claim 10 or 11, wherein the first bodily sample and the second bodily sample are obtained or derived from a same blood sample, wherein the blood biopsy is separated into a cell-free fraction and cellular fraction, wherein the cell-free nucleic acid molecules are obtained or derived from the cell-free fraction, and wherein the cell-derived nucleic acid molecules are obtained or derived from the cellular fraction.

13. The method according to any one of claims 10 to 12, wherein one or more of the identified single nucleotide variants are removed from analysis, and wherein the removed single nucleotide variants include variants from clonal hematopoiesis genes.
14. The method according to any one of claims 10 to 13, wherein one or more of the identified single nucleotide variants are removed from analysis, and wherein the removed single nucleotide variants include variants present in the cell-derived nucleic acid sequencing reads.
15. The method according to any one of claims 10 to 14, further comprising:
identifying or having identified copy number variations present in both the cell-free nucleic acid sequencing reads and the cell-derived nucleic acid sequencing reads; and
determining or having determined, based at least in part on applying a second computational model to the identified copy number variations, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm,
wherein the second computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a third set of individuals having a neoplasm and a fourth set of control individuals not having a neoplasm, and
wherein the second computational model integrates one or more of the following features: a number of uniformly distributed genomic window regions, a number of GISTIC "hotspot" regions, the enrichment of GISTIC "hotspot" regions as compared to the uniform windows, and any combination thereof.
16. The method according to any one of claims 10 to 15, further comprising:
identifying or having identified genomic positions of a first nucleotide and a last nucleotide of each of a plurality of unique sequenced cell-free nucleic acid molecules within the cell-free nucleic acid sequencing reads;
determining or having determined a frequency of the identified genomic positions of the first nucleotide and the last nucleotide of each of the plurality of unique sequenced cell-free nucleic acid molecules; and

determining or having determined, based at least in part on applying a third computational model to the frequencies of the identified genomic positions of the first nucleotide and the last nucleotide of each of the plurality of unique sequenced cell-free nucleic acid molecules, that the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm, wherein the third computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a fifth set of individuals having a neoplasm and a sixth set of control individuals not having a neoplasm.

17. The method of claim 16, further comprising:
 - generating a confidence score from each of the first computational model, the second computational model, and the third computational model; and
 - integrating the confidence scores to generate a summarized score indicative of whether the individual has a neoplasm.
18. The method according to any one of claims 10 to 17, further comprising performing a clinical procedure on the individual based at least in part on the detected neoplasm.
19. The method according to any one of claims 10 to 18, further comprising treating the individual based at least in part on the detected neoplasm.
20. A method to detect a neoplasm in an individual, the method comprising:
 - obtaining or having obtained cell-free nucleic acid sequencing reads of a plurality of cell-free nucleic acid molecules, wherein the plurality of cell-free nucleic acid molecules is obtained or derived from a first bodily sample of the individual;
 - obtaining or having obtained cell-derived nucleic acid sequencing reads of a plurality of cell-derived nucleic acid molecules, wherein the plurality of cell-derived nucleic acid molecules is obtained or derived from a second bodily sample of the individual;
 - identifying or having identified copy number variations present in both the cell-free nucleic acid sequencing reads and the cell-derived nucleic acid sequencing reads; and

determining or having determined, based at least in part on applying a computational model to the identified copy number variations, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm,

wherein the computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a first set of individuals having a neoplasm and a second set of control individuals not having a neoplasm, and

wherein the computational model integrates one or more of the following features: a number of uniformly distributed genomic window regions, a number of GISTIC "hotspot" regions, the enrichment of GISTIC "hotspot" regions as compared to the uniform windows, and any combination thereof; and

detecting the neoplasm in the individual based at least in part on the determination that the cell-free nucleic acid sequencing data contains the nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

21. A method for detecting a neoplasm in an individual, the method comprising: obtaining or having obtained cell-free nucleic acid sequencing reads of a plurality of cell-free nucleic acid molecules, wherein the plurality of cell-free nucleic acid molecules is obtained or derived from a bodily sample of the individual;

identifying or having identified genomic positions of a first nucleotide and a last nucleotide of each of a plurality of unique sequenced cell-free nucleic acid molecules within the sequencing reads;

determining or having determined a frequency of the identified genomic positions of the first nucleotide and the last nucleotide of each of the plurality of unique sequenced cell-free nucleic acid molecules; and

determining or having determined, based at least in part on applying a computational model to the frequencies of the identified genomic positions of the first nucleotide and the last nucleotide of each of the plurality of unique sequenced cell-free nucleic acid molecules, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm, wherein the computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a first set of

individuals having a neoplasm and a second set of control individuals not having a neoplasm; and

detecting the neoplasm in the individual based at least in part on the determination that the cell-free nucleic acid sequencing data contains the nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

22. The method according to any one of claims 10 to 21, further comprising determining at least one of a first set of quantitative measures of cfDNA molecules starting at each of a plurality of genomic positions and a second set of quantitative measures of cfDNA molecules ending at each of a plurality of genomic positions; and analyzing the at least one of the first set of quantitative measures or the second set of quantitative measures to detect the neoplasm.

23. The method of claim 22, further comprising analyzing the at least one of the first set of quantitative measures and the second set of quantitative measures using a trained machine learning classifier to detect the neoplasm.

24. The method of claim 22 or 23, further comprising analyzing the at least one of the first set of quantitative measures and the second set of quantitative measures to determine a tumor variant allele frequency of the neoplasm.

25. The method according to any one of claims 22 to 24, further comprising analyzing the at least one of the first set of quantitative measures and the second set of quantitative measures to determine a metabolic tumor volume of the neoplasm.

26. The method according to any one of claims 22 to 25, further comprising detecting the neoplasm with an AUC of at least about 0.80.

27. A method for detecting a neoplasm in an individual, the method comprising:
obtaining or having obtained cell-free nucleic acid sequencing reads of a plurality of cell-free nucleic acid molecules, wherein the plurality of cell-free nucleic acid molecules is obtained or derived from a bodily sample of the individual;
identifying or having identified a fragment length of each of a plurality of unique sequenced cell-free nucleic acid molecules within the sequencing reads;

selecting a subset of the sequencing reads corresponding to cell-free nucleic acid molecules of the plurality of unique sequenced cell-free nucleic acid molecules having a fragment length indicative of a sub-mononucleosomal fragment or a sub-disomal fragment;

analyzing the subset of the sequencing reads to determine a frequency of the identified fragment lengths indicative of the sub-mononucleosomal fragments or the sub-disomal fragments; and

determining or having determined, based at least in part on applying a computational model to the identified fragment lengths indicative of the sub-mononucleosomal fragments or the sub-disomal fragments, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm, wherein the computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a first set of individuals having a neoplasm and a second set of control individuals not having a neoplasm; and

detecting the neoplasm in the individual based at least in part on the determination that the cell-free nucleic acid sequencing data contains the nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

28. The method of claim 27, wherein a fragment length of less than 160 base pairs (bp) is indicative of the sub-mononucleosomal fragments.

29. The method of claim 27 or 28, wherein a fragment length of between 230 bp and 310 bp is indicative of the sub-disomal fragments.

30. The method according to any one of claims 27 to 29, wherein the plurality of cell-free nucleic acid molecules are obtained at least in part by performing a size selection of nucleic acid molecules of the bodily sample of the individual to enrich for at least one of the sub-mononucleosomal fragments and the sub-disomal fragments.

31. A method for detecting a neoplasm in an individual, the method comprising:
obtaining or having obtained cell-free nucleic acid sequencing reads of a plurality of cell-free nucleic acid molecules, wherein the plurality of cell-free nucleic acid molecules is obtained or derived from a bodily sample of the individual;

analyzing the sequencing reads to determine a variant allele frequency (VAF) of a plurality of single nucleotide variants (SNVs); and

determining or having determined, based at least in part on applying a computational model to the determined VAFs of the plurality of SNVs, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm, wherein the computational model is constructed utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a first set of individuals having a neoplasm and a second set of control individuals not having a neoplasm; and

detecting the neoplasm in the individual based at least in part on the determination that the cell-free nucleic acid sequencing data contains the nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

32. The method of claim 31, further comprising determining a mean value of the determined VAFs across the plurality of SNVs; and determining or having determined, based at least in part on applying the computational model to the determined mean value, whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

33. The method of claim 31 or 32, further comprising comparing the determined mean value of the determined VAFs across the plurality of SNVs to a reference value, to determine whether the cell-free nucleic acid sequencing reads contain nucleotides indicative of cell-free nucleic acid molecules derived from a neoplasm.

34. The method according to any one of claims 31 to 33, further comprising determining a metabolic tumor volume of the detected neoplasm.

35. The method according to any one of claims 31 to 34, further comprising determining a stage of the detected neoplasm.

36. The method according to any one of claims 31 to 35, further comprising determining a likelihood of recurrence of the detected neoplasm.

37. The method according to any one of claims 10 to 36, wherein the plurality of cell-free nucleic acid molecules is enriched from the bodily sample of the individual using a set of capture bait molecules,
- wherein the set of capture bait molecules are configured to selectively hybridize to sequences that are at least partially complementary to at least one sequence of the set of capture bait molecules,
- wherein the set of capture bait molecules are configured to selectively hybridize to sequences that are at least partially complementary to at least one genomic locus selected from the group of genomic loci in Table 1.
38. The method according to any one of claims 10 to 37, wherein the neoplasm comprises lung cancer.
39. A bait set for hybridization capture, the bait set comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, or 5000 different polynucleotide-containing probes, wherein the polynucleotide-containing probes are, collectively, configured to hybridize to cfDNA derived from at least 5% of the genomic regions set forth in Table 1.
40. The bait set of claim 39, wherein each of the polynucleotide-containing probes has a nucleic acid sequence that is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, or 300 nucleotides in length.
41. The bait set of claim 39 or 40, wherein each of the polynucleotide-containing probes has a nucleic acid sequence of no more than 300, 280, 260, 240, 220, 200, 180, 160, 140, 120, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, or 2 nucleotides in length.
42. The bait set of any one of claims 39 to 41, wherein each of the polynucleotide-containing probes is conjugated to an affinity moiety.
43. The bait set of claim 42, wherein the affinity moiety comprises biotin.

44. The bait set of any one of claims 39 to 43, wherein the polynucleotides probes are, collectively, configured to hybridize to cfDNA derived from at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% of the genomic regions set forth in Table 1.

45. The bait set of any one of claims 39 to 44, wherein an entirety of polynucleotide probes in the bait set are configured to hybridize to cfDNA molecules derived from at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% of the genomic regions set forth in Table 1.

46. A mixture comprising:
DNA from a biological sample; and
the bait set of any one of claims 39 to 45.

47. A method of performing capture hybridization, the method comprising:
obtaining a plurality of DNA molecules derived from a cell-free DNA source; and
mixing a fraction of the plurality of DNA molecules with a set of capture bait molecules, wherein the set of capture bait molecules are configured to selectively hybridize to DNA molecules that are at least partially complementary to at least one sequence of the set of capture bait molecules,
wherein the set of capture bait molecules are configured to selectively hybridize to DNA molecules comprising sequences that comprise at least a portion of a genomic locus selected from the group of genomic loci in Table 1.

48. The method of claim 47, wherein the portion of the genomic locus comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, or 300 consecutive nucleotides of the genomic locus.

49. The method of claim 47 or 48, wherein the fraction is at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the plurality of DNA molecules.

50. The method of claim 47, further comprising optimizing a molar ratio of the fraction of the plurality of DNA molecules and the set of capture bait molecules to yield an optimal recovery of a total number of unique molecules or to yield an optimal recovery of a total number of duplexed cell-free DNA molecules in which both strands of the sourced cell-free DNA duplex are sequenced, wherein the molar ratio is at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%.

51. The method of claim 47, further comprising using an *in silico* simulation of the capture hybridization to determine the fraction of the plurality of DNA molecules that is mixed with the set of capture bait molecules, wherein the fraction is no more than about 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5%.

52. The method of any one of claims 47 to 51, wherein the set of capture bait molecules comprises the bait set of any one of claims 39 to 45.

53. A DNA molecule, comprising:
a nucleic acid molecule sourced from a biological sample, wherein the nucleic acid molecule is DNA or cDNA;
a pair of error-correcting unique identifiers that flank the nucleic acid molecule, wherein the error-correcting unique identifiers are each DNA molecules and the combination of the error-correcting unique identifiers provides an identification of the nucleic acid molecule in a sequencing result; and
a pair of error-correcting dual index sample barcodes, wherein the error-correcting unique identifiers are each DNA molecules and the combination of the error-correcting unique identifiers provides an identification of the biological sample in a sequencing result.

54. A method to prepare DNA library for sequencing, comprising:
ligating onto a collection of nucleic acid molecules pairs of partial Y-adapters such that each nucleic acid molecule is flanked by a pair of partial Y-adapters,
wherein each nucleic acid molecule is DNA or cDNA and the collection of nucleic acid molecules is sourced from a biological sample,

wherein each partial Y-adapter includes an error-correcting unique identifier and sequences for a primer to anneal in a grafting polymerase chain reaction, and

wherein each flanking combination of the two error-correcting unique identifiers on each nucleic acid molecule identifies the ligation of the pair of partial Y-apters to that nucleic acid molecule; and

grafting onto each ligation product a pair of error-correcting dual index sample barcodes and the sequence of a universal primer such that the ligation product is flanked by the error-correcting dual index sample barcodes and the sequence of the universal primer, wherein the combination of error-correcting dual index sample barcodes identifies the collection of nucleic acid molecules.

55. A method to mitigate nucleotide transversions that arise during sequencing library preparation, comprising:

performing sequence library preparation with an reactive oxygen species scavenger or enzyme in the reaction mixture.

56. The method of claim 55, wherein the sequence capture reaction is performed with the reactive oxygen species scavenger hypotaurine in the reaction mixture.

57. The method of claim 55 or 56, wherein the reactive oxygen species scavenger is glutathione, hypotaurine, or sodium sulfite; and wherein the enzyme is uracil-DNA glycosylase (UDG), Formamidopyrimidine [fapy]-DNA glycosylase (FPG), or catalase enzyme.

58. A method to perform a clinical procedure on an individual, the method comprising:

obtaining or having obtained a sequencing result of a collection of cell-free nucleic acid molecules, wherein the collection of cell-free nucleic acid molecules are sourced from a first biopsy of an individual;

obtaining or having obtained a sequencing result of a collection of cell-derived nucleic acid molecules, wherein the collection of cell-derived nucleic acid molecules are sourced from a second biopsy of the individual;

identifying or having identified single nucleotide variants within both the cell-free nucleic acid sequencing result and the cell-derived nucleic acid sequencing result;

determining or having determined, utilizing a first computational model and the identified single nucleotide variants, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm,

wherein the first computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm, and

wherein the first computational model integrates one or more of the following features: cell-derived DNA Bayesian background, cfDNA Bayesian background, germline depth, short fragment score 1, short fragment score 2, or genomic start and end position of the cfDNA molecule; and

performing a clinical procedure on the individual based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

59. The method of claim 58, wherein the first computational model further integrates one or more of the following features: transition/transversion, duplex support, pass outlier, mapping quality, cancer hotspot, UMI error corrected, Phred quality score, variant allele frequency (VAF %), mean barcode family size, variant position in a cfDNA molecule, polygenic risk score, or nuclease motif.

60. The method of claim 58 or 59, wherein the first biopsy and the second biopsy are the same blood biopsy and the blood biopsy is separated into a cell-free fraction and cellular fraction, and wherein the cell-free fraction is used to source cell-free nucleic acid molecules and the cellular fraction is used to source cell-derived nucleic acid molecules.

61. The method according to any one of claims 58 to 60, wherein a number of identified variants are removed from analysis, and wherein the removed variants include variants from clonal hematopoiesis genes or somatic mutations in other non-malignant tissue types.

62. The method of according to any one of claims 58 to 61, wherein a number of identified variants are removed from analysis, and wherein the removed variants include variants present in the cell-derived nucleic acid sequencing result.

63. The method of according to any one of claims 58 to 62, further comprising:
identifying or having identified copy number variations within both the cell-free nucleic acid sequencing result and the cell-derived nucleic acid sequencing result; and
determining or having determined, utilizing a second computational model and the identified copy number variations, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm,

wherein the second computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm, and

wherein the second computational model integrates one or more of the following features: the number of uniformly distributed genomic window regions, the number of GISTIC “hotspot” regions, and the enrichment of GISTIC “hotspot” regions as compared to the uniform windows;

wherein the performing the clinical procedure on the individual is based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

64. The method of claim according to any one of claims 58 to 63, further comprising:
identifying or having identified the genomic position of the first and last nucleotide of each unique sequenced cell-free nucleic acid molecule within both the cell-free nucleic acid sequencing result;

determining or having determined, the frequency of identified genomic positions of the first and last nucleotide of each unique sequenced cell-free nucleic acid; and

determining or having determined, utilizing a third computational model and the frequency of the identified genomic positions of the first and last nucleotide of each unique sequenced cell-free nucleic acid molecule, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm, wherein the third computational model is built utilizing cell-free nucleic acid sequencing data and cell-

derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm;

wherein the performing the clinical procedure on the individual is based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

65. The method of claim 64, further comprising:

generating a confidence score from each the first computational model, the second computational model, and the third computational model; and

integrating the confidence scores to generate a summarized score indicating that the individual has a neoplasm;

wherein the performing the clinical procedure on the individual is based upon the summarized score indicating that the individual has a neoplasm.

66. A method to treat an individual for a neoplasm, the method comprising:

obtaining or having obtained a sequencing result of a collection of cell-free nucleic acid molecules, wherein the collection of cell-free nucleic acid molecules are sourced from a first biopsy of an individual;

obtaining or having obtained a sequencing result of a collection of cell-derived nucleic acid molecules, wherein the collection of cell-derived nucleic acid molecules are sourced from a second biopsy of the individual;

identifying or having identified single nucleotide variants within both the cell-free nucleic acid sequencing result and the cell-derived nucleic acid sequencing result;

determining or having determined, utilizing a first computational model and the identified single nucleotide variants, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm,

wherein the first computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm, and

wherein the first computational model integrates one or more of the following features: cell-derived DNA Bayesian background, cfDNA Bayesian background, germline depth, short fragment score 1, short fragment score 2, or genomic start and end position of the cfDNA molecule; and

treating the individual based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

67. The method of claim 66, wherein the first computational model further integrates one or more of the following features: transition/transversion, duplex support, pass outlier, mapping quality, cancer hotspot, UMI error corrected, Phred quality score, variant allele frequency (VAF %), mean barcode family size, variant position in a cfDNA molecule, polygenic risk score, or nuclease motif.

68. The method of claim 66 or 67, wherein the first biopsy and the second biopsy are the same blood biopsy and the blood biopsy is separated into a cell-free fraction and cellular fraction, and wherein the cell-free fraction is used to source cell-free nucleic acid molecules and the cellular fraction is used to source cell-derived nucleic acid molecules.

69. The method according to any one of claims 66 to 68, wherein a number of identified variants are removed from analysis, and wherein the removed variants include variants from clonal hematopoiesis genes.

70. The method according to any one of claims 66 to 69, wherein a number of identified variants are removed from analysis, and wherein the removed variants include variants present in the cell-derived nucleic acid sequencing result.

71. The method according to any one of claims 66 to 70, further comprising:
identifying or having identified copy number variations within both the cell-free nucleic acid sequencing result and the cell-derived nucleic acid sequencing result;
determining or having determined, utilizing a second computational model and the identified copy number variations, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm,

wherein the second computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm, and

wherein the second computational model integrates one or more of the following features: the number of uniformly distributed genomic window regions, the number of GISTIC “hotspot” regions, and the enrichment of GISTIC “hotspot” regions as compared to the uniform windows;

wherein the treating the individual is based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

72. The method according to any one of claims 66 to 71, further comprising:

identifying or having identified the genomic position of the first and last nucleotide of each unique sequenced cell-free nucleic acid molecule within both the cell-free nucleic acid sequencing result;

determining or having determined, the frequency of identified genomic positions of the first and last nucleotide of each unique sequenced cell-free nucleic acid; and

determining or having determined, utilizing a third computational model and the frequency of the identified genomic positions of the first and last nucleotide of each unique sequenced cell-free nucleic acid molecule, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm, wherein the third computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm;

wherein the treating the individual is based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

73. The method of claim 72, further comprising:

generating a confidence score from each the first computational model, the second computational model, and the third computational model; and

integrating the confidence scores to generate a summarized score that indicates that the individual has a neoplasm;

wherein the treating the individual is based upon the summarized score indicating that the individual has a neoplasm.

74. A method of performing capture hybridization on a sequencing library, the method comprising:

obtaining a sequencing library comprising a plurality of unique sequencing molecules derived from cell-free DNA source;

mixing a fraction of the sequencing library with capture baits molecules to hybridize and pull down particular sequences recognized by the capture baits, wherein the fraction of sequencing library is at least 10%, 25%, or 50%.

75. The method of claim 74, wherein the molar ratio of the fraction of the sequencing library and the capture baits is optimized to yield an optimal recovery of total unique molecules or to yield an optimal recovery of the total duplexed cell-free DNA molecules in which both strands of the sourced cell-free DNA duplex are sequenced.

76. The method of claim 74 or 75, wherein an *in silico* simulation of the capture hybridization is utilized to determine the fraction of sequencing library that is mixed with capture baits.

77. A method to treat an individual for a neoplasm, the method comprising:

obtaining or having obtained a sequencing result of a collection of cell-free nucleic acid molecules, wherein the collection of cell-free nucleic acid molecules are sourced from a first biopsy of an individual;

obtaining or having obtained a sequencing result of a collection of cell-derived nucleic acid molecules, wherein the collection of cell-derived nucleic acid molecules are sourced from a second biopsy of the individual;

identifying or having identified copy number variations within both the cell-free nucleic acid sequencing result and the cell-derived nucleic acid sequencing result;

determining or having determined, utilizing a computational model and the identified copy number variations, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm,

wherein the second computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm, and

wherein the second computational model integrates one or more of the following features: the number of uniformly distributed genomic window regions, the number of GISTIC “hotspot” regions, and the enrichment of GISTIC “hotspot” regions as compared to the uniform windows; and treating the individual based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.

78. A method to treat an individual for a neoplasm, the method comprising:
- obtaining or having obtained a sequencing result of a collection of cell-free nucleic acid molecules, wherein the collection of cell-free nucleic acid molecules are sourced from a biopsy of an individual;
 - identifying or having identified the genomic position of the first and last nucleotide of each unique sequenced cell-free nucleic acid molecule within both the cell-free nucleic acid sequencing result;
 - determining or having determined, the frequency of identified genomic positions of the first and last nucleotide of each unique sequenced cell-free nucleic acid;
 - determining or having determined, utilizing a computational model and the frequency of the identified genomic positions of the first and last nucleotide of each unique sequenced cell-free nucleic acid molecule, that the cell-free nucleic acid sequencing result contains nucleotides derived from a neoplasm, wherein the third computational model is built utilizing cell-free nucleic acid sequencing data and cell-derived nucleic acid sequencing data from a cohort of individuals having a neoplasm and a cohort of control individuals not having a neoplasm; and
 - treating the individual based upon determining that the cell-free nucleic acid sequencing result contains nucleic acid sequences derived from a neoplasm.