

Claims

1. A method for detecting the presence of a cancer and for identifying the cancer origin in a test subject, the method comprising:
 - a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
 - b) using the bisulfite treated cfDNA to prepare, for a genome of the species of the test subject,
 - (i) a first sequencing library for a plurality of specific target genomic regions, and
 - (ii) a second sequencing library, wherein the second sequencing library comprises the bisulfite treated cfDNA that fails to bind to the set of probes;
 - c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
 - d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for the plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of the plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the

second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from a cohort of a healthy subject; and
 e) responsive to inputting into a model each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model a categorical indication of a presence or absence of the cancer in the test subject, and
 in the case where the model determines presence of the cancer in the test subject, an origin of the cancer, wherein

the model is a composite model comprising four attribute models and a combination model,

each respective attribute model in the four attribute models produces a respective categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and

the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the respective categorical indication of the four attribute models.

2. The method of claim 1, wherein the plurality of specific target genomic regions comprises at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, or at least 500 or more cancer specific regions.
3. The method of claim 1 or 2, wherein the plurality of specific target genomic regions comprises between 400 and 500 cancer specific gene regions.
4. The method of claim 1, wherein the plurality of specific target genomic regions consists of between 17,500 and 18,500 CpG sites.
5. The method of claim 3, wherein the plurality of specific target genomic regions comprises at least five nucleic acid sequences selected from SEQ ID NOs: 1-450.

6. The method of claim 1, wherein the plurality of specific target genomic regions comprises at least 50 nucleic acid sequences selected from SEQ ID NOs: 1-450.
7. The method of claim 1, wherein the plurality of specific target genomic regions comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 1-450.
8. The method of claim 1, wherein the plurality of specific target genomic regions comprises at least 300 nucleic acid sequences selected from SEQ ID NOs: 1-450.
9. The method of claim 1, wherein each respective target genomic region in the plurality of specific target genomic regions encompasses a sequence selected from SEQ ID NOs: 1-450.
10. The method of any one of claims 2-5, wherein at least 20 respective cancer specific genomic regions in the plurality of cancer specific genomic regions encompass an oncogene and/or a tumor suppressor gene listed in Table 23.
11. The method of any one of claims 1-10, wherein the plurality of specific target genomic regions is captured by a set of DNA probes.
12. The method of claim 11, wherein the set of DNA probes comprises DNA fragments with a size ranging between 40 base-pair (bp) and 50 bp, between 51 bp and 60 bp, between 61 bp and 70 bp, between 71 bp and 80 bp, between 81 bp and 90 bp, between 91 bp and 100 bp, between 101 bp and 110 bp, between 111 bp and 120 bp, between 121 bp and 130 bp, between 131 bp and 140 bp, between 141 bp and 150 bp, between 151 bp and 160 bp, between 161 bp and 170 bp, between 171 bp and 180 bp, between 181 bp and 190 bp, or between 191 bp and 200 bp.
13. The method of claim 11 or 12, wherein the set DNA probes comprises DNA fragments

with a size ranging between 111 bp and 120 pb or between 121 bp and 130 bp.

14. The method of claim 12 or 13, wherein the set of DNA probes consists of between 400 DNA probes and 500 DNA probes, between 501 DNA probes and 1000 DNA probes, between 1001 DNA probes and 1500 DNA probes, between 1501 DNA probes and 2000 DNA probes, between 2001 DNA probes and 2100 DNA probes, between 2101 DNA probes and 2150 DNA probes, between 2151 DNA probes and 2200 DNA probes, between 2201 DNA probes and 2250 DNA probes, between 2251 DNA probes and 2300 DNA probes, between 2301 DNA probes and 2350 DNA probes, between 2351 DNA probes and 2400 DNA probes, between 2401 DNA probes and 2450 DNA probes, between 2451 DNA probes and 2500 DNA probes, between 2501 DNA probes and 3000 DNA probes, between 3001 DNA probes and 3500 DNA probes, or between 3501 DNA probes and 4000 DNA probes.
15. The method of any one of claims 11-14, wherein the set DNA probes consists of between 2201 DNA probes and 2250 DNA probes or between 2251 DNA probes and 2300 DNA probes.
16. The method of claim 15, wherein the set of DNA probes comprises at least 10 nucleic acid sequences selected from SEQ ID NOs: 451-2700.
17. The method of claim 15, wherein the set of DNA probes comprises at least 100 nucleic acid sequences selected from SEQ ID NOs: 451-2700.
18. The method of claim 15, wherein the set of DNA probes comprises at least 200 nucleic acid sequences selected from SEQ ID NOs: 451-2700.
19. The method of claim 1, wherein the first sequencing library is prepared for paired-end sequencing.

20. The method of any one of claims 1-19, wherein the plurality of specific target genomic regions have a different methylation percentage between the test subject and the cohort of healthy subjects.
21. The method of any one of claims 1-20, wherein the plurality of specific target genomic regions have a methylation percentage higher in the test subject as compared to the cohort of healthy subjects.
22. The method of claim 21, wherein the methylation in the test subject is about two fold higher than the methylation in the cohort of healthy subjects.
23. The method of claim 1, wherein the genome comprises 22 chromosomes.
24. The method of claim 1, wherein the second sequencing library comprises universal adapter sequences.
25. The method of claim 1, the method further comprising converting the second sequencing library into cfDNA sequencing library spheres for genomic sequencing.
26. The method of claim 25, wherein the genomic sequencing comprises rolling circle sequencing or MGI-DNBseq sequencing.
27. The method of claim 1, wherein the analysis of the sequencing results from (d)(ii)-(d)(iv) is performed by measuring non-duplicating fragments in the genome.
28. The method of claim 27, wherein the methylation density for the genome in (d)(ii) is determined for each respective second bin region, in a plurality of second bins, wherein the plurality of second bins consists of between 2500 second bins and 3000 second bins.

29. The method of claim 28, wherein each respective second bin region consists of between 800,000 nucleotides and 1,200,000 nucleotides.
30. The method of claim 28 or 29, wherein the measuring of the methylation density identifies respective second bin regions in the plurality of second bin regions that are differentially methylated between the test subject and the cohort of healthy subjects.
31. The method of claim 30, wherein the methylation density in each respective second bin region is evaluated based on a Z score value.
32. The method of claim 1, wherein the plurality of first bins is between 2500 first bins and 3000 first bins.
33. The method of claim 32, wherein each first bin in the plurality of first bins represents a different between 800,000 nucleotides and 1,200,000 nucleotides of the genome.
34. The method of claim 1, wherein the measuring of respective copy number of cfDNA identifies a subset of first bins in the plurality of first bins with variation in the number of copies of DNA per bin between the test subject and the cohort of healthy subjects.
35. The method of claim 34, wherein the variation in the number of copies of DNA between the test subject and the cohort of healthy subjects in each first bin is evaluated based on a Z score value.
36. The method of claim 35, wherein the Z score identifies regions of instability in the genome.
37. The method of claim 1, wherein the measuring of the fragment size pattern distribution of cfDNA across the genome comprises determining a fragment size pattern distribution in each third bin in a plurality of third bins, wherein the plurality of third bins consists of

between 500 third bins and 600 third bins.

38. The method of claim 37, wherein each respective third bin in the plurality of third bins represents a different between 4.5 million nucleotides (4.5 megabases) and 5.5 million nucleotides (5.5 megabases) of the genome.
39. The method of claim 37 or 38, wherein the measuring of the fragment size pattern distribution of cfDNA identifies a subset of third bins in the plurality of third bins with a variation in the fragment size pattern distribution of cfDNA per bin between the test subject and the cohort of healthy subjects.
40. The method of claim 39, wherein the variation in the fragment size pattern distribution of the cfDNA in each third bin in the plurality of third bins is evaluated based on cfDNA fragment length ratio (RF) value.
41. The method of claim 40, wherein the RF value identifies presence of cancer, wherein cfDNA fragment length released from tumor cells from the test subject is shorter than cfDNA fragment length released by cells of the cohort of healthy subjects.
42. The method of any one of claims 1-41, wherein the cohort of healthy subjects consists of between 5 and 50 healthy subjects, between 5 and 100 healthy subjects, between 5 and 1000 healthy subjects, between 5 and 5000 healthy subjects, between 50 and 500 healthy subjects, between 50 and 1000 healthy subjects, between 50 and 5000 healthy subjects, between 100 and 500 healthy subjects, between 100 and 1000 healthy subjects, between 100 and 5000 healthy subjects, between 500 and 1000 healthy subjects, or between 500 and 5000 healthy subjects, or more.
43. The method of any one of claims 1-42, wherein the liquid biopsy sample comprises a body fluid, blood, or plasma.

44. The method of any one of claims 1-43, wherein the origin of the cancer comprises colorectal cancer (CRC), liver cancer, lung cancer, breast cancer, or gastric cancer.
45. The method of any one of claims 1-44, wherein the subject is a human.
46. The method of claim 1, wherein the combination model is a logistic regression combined linear model of the four attribute models, in which each of the four attribute models is independently assigned a different probability weight.
47. The method of any one of claims 1-46, wherein the combination model comprises at least 100 parameters.
48. The method of any one of claims 1-47, wherein the combination model comprises a logistic regression, a deep neural network, a fully connected neural network, a convolutional neural network, a graph based neural network, or a support vector machine.
49. The method of claim 48, wherein the combination model is a deep neural network, and wherein the deep neural network further specifies a tissue for cancer origin when the combination model determines presence of the cancer in the test subject, an origin of the cancer.
50. A method for monitoring likelihood of cancer recurrence in a subject previously treated for cancer, the method comprising:
- a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
 - b) using the bisulfite treated cfDNA to prepare, for a genome of the species of the test subject,
 - (i) a first sequencing library for a plurality of specific target genomic regions, and
 - (ii) a second sequencing library, wherein the second sequencing library comprises

the bisulfite treated cfDNA that fails to bind to the set of probes;

c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;

d) analyzing the corresponding first and second plurality of sequencing results by measuring:

- i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from the cohort of a healthy subject; and
- e) responsive to inputting into a model each of the analyzed sequencing results

from (d)(i)-(d)(iv), receiving as output from the model a categorical indication of a presence or absence of the cancer in the test subject, and

in the case where the model determines presence of the cancer in the test subject, an origin of the cancer, wherein

the model is a composite model comprising four attribute models and a combination model,

each respective attribute model in the four attribute models produces a respective categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and

the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the respective categorical indication of the four attribute models,

wherein the detection of a cancer is indicative of cancer recurrence and need of resuming treatment to the subject.

51. A method for assessing the efficacy of a cancer treatment in a subject suffering from cancer, the method comprising:

- a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
- b) using the bisulfite treated cfDNA to prepare, for a genome of the species of the test subject,
 - (i) a first sequencing library for a plurality of specific target genomic regions, and
 - (ii) a second sequencing library, wherein the second sequencing library comprises the bisulfite treated cfDNA that fails to bind to the set of probes;
- c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
- d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome

determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from the cohort of healthy subjects;

- iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from the cohort of healthy subjects, and
- iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from a cohort of a healthy subject; and

e) responsive to inputting into a model each of the analyzed sequencing results from (d)(i)-(d)(iv), receiving as output from the model a categorical indication of a presence or absence of the cancer in the test subject, and

in the case where the model determines presence of the cancer in the test subject, an origin of the cancer, wherein

the model is a composite model comprising four attribute models and a combination model,

each respective attribute model in the four attribute models produces a respective categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and

the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the respective categorical indication of the four attribute models,

wherein the detection of a cancer is indicative of efficacy of treatment and need of continuing, modifying or discontinuing treatment of the subject.

52. A method for determining the need to treat a subject and the type of treatment that is the most efficacious for the subject, the method comprising:

- a) bisulfite treating cell free DNA (cfDNA) from a liquid biopsy sample of the test subject;
- b) using the bisulfite treated cfDNA to prepare, for a genome of the species of the test subject,
 - (i) a first sequencing library for a plurality of specific target genomic regions, and
 - (ii) a second sequencing library, wherein the second sequencing library comprises the bisulfite treated cfDNA that fails to bind to the set of probes;
- c) sequencing the prepared first and second sequencing libraries, thereby producing a corresponding first and second plurality of sequencing results;
- d) analyzing the corresponding first and second plurality of sequencing results by measuring:
 - i. a plurality of site specific methylation densities, using the first plurality of sequencing results, for the plurality of specific target genomic regions of the test subject relative to a plurality of site specific methylation densities determined using a plurality of sequencing results for the plurality of specific target genomic regions in a plurality of liquid biopsies obtained from a cohort of healthy subjects;
 - ii. a methylation density for the genome, using the second plurality of sequencing results, of the test subject relative a methylation density for the genome determined from a plurality of genome wide sequencing results for a plurality of liquid biopsies obtained from the cohort of healthy subjects;
 - iii. a respective copy number of cfDNA in a plurality of first bins across the genome, using the second plurality of sequencing results, of the test subject relative to a respective copy number of cfDNA in the plurality of first bins across the genome determined using a plurality of genome wide sequencing results of a plurality of liquid biopsies obtained from the cohort of healthy subjects, and
 - iv. a fragment size pattern distribution of cfDNA across the genome, using the second plurality of sequence results, of the test subject relative to a fragment size distribution of cfDNA determined using a plurality of genome sequencing results for a plurality of liquid biopsies obtained from the cohort of a healthy subject; and
 - e) responsive to inputting into a model each of the analyzed sequencing results

from (d)(i)-(d)(iv), receiving as output from the model a categorical indication of a presence or absence of the cancer in the test subject, and
in the case where the model determines presence of the cancer in the test subject, an origin of the cancer, wherein

the model is a composite model comprising four attribute models and a combination model,

each respective attribute model in the four attribute models produces a respective categorical classification upon input of a different one of the analyzed sequencing results from (d)(i)-(d)(iv), and

the combination model combines the respective categorical indication of the presence or absence of cancer in the test subject of each attribute model in the four attribute models by a weighted combination of the respective categorical indication of the four attribute models,

wherein the detection of a cancer and the identification of the cancer origin are indicative of the need to treat the subject and the type of treatment that is the most efficacious given the tumor origin.

53. The method of claim 4, wherein the plurality of specific target genomic regions consists of 18,000 CpG sites.

54. The method of claim 1, wherein bisulfite treated cfDNA in step a) is further treated with Human Cot1 DNA, or a xGen Universal Blocker, to reduce the rate of nonspecific capture by the set of probes.

55. The method of claim 1, wherein the plurality of specific target genomic regions comprises at least 75 or more cancer specific regions.