

CLAIMS

1. A method of detecting methylated DNA pattern in DNA in a biological sample, comprising:

5 (a) converting un-methylated cytosine of the DNA to uracil by deamination to thereby generate converted DNA;

(b) purifying the converted DNA from step (a);

(c) tagging a barcode sequence on the converted DNA, by performing a first PCR amplification using a plurality of forward and reverse primer pairs specific to the converted
10 sequence of target regions, wherein each forward primer of the plurality of forward and reverse primer pairs comprises a barcode sequence on its 5'end, wherein the barcode sequence of each forward primer is different from each other, wherein the target regions comprise one or more CpG sites;

(d) subjecting the tagged converted DNA from step (c) to a second PCR amplification with
15 universal indexed primers to thereby create a sequencing library with components required for multiplex sequencing;

(e) subjecting the sequencing library to multiplex sequencing on a next-generation sequencing platform;

(f) detecting the presence of the barcode sequence using Bioinformatics methods to count
20 and assign each DNA sequence from the next-generation sequencing to an original parental DNA molecule carrying the same barcode sequence, comprising:

(i) performing cluster reassignment of sequencing reads with the same barcode sequence to thereby generate barcode clusters wherein each barcode cluster contains reads from the same amplicon and with the same barcode sequence; and

25 (ii) performing consensus calling for each barcode cluster to thereby obtain consensus reads;

(g) reconstructing the methylated DNA pattern of the DNA by

(I) comparing the DNA sequence to a reference genome using a sequence alignment tool; and

30 (II) conducting variant analysis of the DNA sequence by comparing the consensus reads to the reference genome to detect the variations;

to thereby assess 1) the conversion efficiency of non-CpG cytosines to thymine as quality control, and 2) the frequency of methylation at each CpG cytosine across all consensus sequencing reads with distinct barcode sequences (i.e. reads that are derived from unique DNA molecules corresponding to a specific amplicon).

5

2. The method of claim 1, further comprising the following steps before step (a):

- 1) extracting the DNA from the biological sample;
- 2) treating the DNA using a first enzyme that oxidizes 5-methylcytosine and 5-hydroxymethylcytosine of the DNA to thereby protect the 5-methylcytosine and 5-hydroxymethylcytosine from deamination in step (a); and
- 3) purifying the DNA from step (2).

10

3. The method of claim 2, wherein in step (a), the un-methylated cytosine of the DNA is converted to uracil by deamination using a second enzyme to thereby generate converted DNA.

15

4. The method of claim 1, wherein the DNA is unextracted from the biological sample, and wherein in step (a), the un-methylated cytosine of the DNA is converted to uracil by deamination using bisulfite to thereby generate converted DNA.

20

5. The method of any one of the preceding claims, further comprising:

(h) performing a statistical modelling technique to thereby predict presence or absence of cancer; and

(i) performing a statistical modelling technique to thereby identify specific cancer types when the presence of cancer is predicted in step (h).

25

6. The method of claim 5, wherein the statistical modelling technique is selected from the group consisting of logistic regression, tree based classifiers and deep neural networks.

30

7. The method of claim 5 or claim 6, further comprising the following step prior to step (h):

analyzing methylated DNA pattern by capturing methylation pattern-specific information and generating new features and/or determining patterns within an amplicon as data input for statistical modelling techniques of step (h) and/or step (i).

5 8. The method of any one of the preceding claims, wherein the biological sample is selected from the group consisting of a liquid sample, a tissue sample, or a cell sample.

9. The method of claim 1, wherein the DNA is selected from the group consisting of cell-free DNA (cfDNA) and DNA encapsulated within tissues and/or cells.

10

10. The method of claim 2, wherein the first enzyme is a Ten-eleven translocation (TET) enzyme or an isoform thereof; and wherein optionally, the TET enzyme is selected from the group consisting of TET1 enzyme or an isoform thereof, TET2 enzyme or an isoform thereof, and TET3 enzyme or an isoform thereof.

15

11. The method of claim 3, wherein the second enzyme is a cytidine deaminase or an enzyme with cytidine deaminase properties;

wherein optionally, the cytidine deaminase is selected from the group consisting of APOBEC enzyme, CDA, and activation-induced cytidine deaminase; and

20

wherein optionally, the enzyme with cytidine deaminase properties is selected from the group consisting of M. SssI and M.HpaII.

12. The method of any one of the preceding claims, wherein the amount of DNA is at least 5 ng.

25

13. The method of any one of the preceding claims, wherein the barcode sequence is an oligonucleotide comprising 10 to 16 random nucleotides, or 10 to 15 random nucleotides, or 10 to 13 random nucleotides, or 10 random nucleotides, or 11 random nucleotides, or 12 random nucleotides, or 13 random nucleotides, or 14 random nucleotides, or 15 random nucleotides, or 16 random nucleotides; and

30

wherein optionally, the barcode sequence is an oligonucleotide having 10 random nucleotides.

14. The method of any one of the preceding claims, wherein the number of the forward and reverse primer pairs is at least 5; and wherein optionally, the number of the forward and reverse primer pairs is at least 159.

5 15. The method of any one of the preceding claims, wherein each forward and reverse primer pair covers a target region which comprises at least 1 CpG site.

16. The method of any one of the preceding claims, wherein the forward and reverse primer pairs comprise sequences as disclosed in Table 1.

10

17. The method of any one of the preceding claims, wherein the forward primer in the primer pair comprises one or more degenerate bases, and/or the reverse primer in the primer pair comprises one or more degenerate bases; wherein optionally, the degenerate base is selected from the group consisting of C, T, A and G.

15

18. The method of claim 17, wherein each primer of the primer pair comprises 1, 2, or 3 degenerate bases; and wherein optionally, each primer of the primer pair has one degenerate base.

20 19. The method of any one of claims 1-18, wherein the primer pair does not comprise a degenerate base.

20. The method of any one of the preceding claims, wherein the first PCR amplification comprises a number of PCR cycles selected from the group consisting of 2, 3, 4 and 5 PCR
25 cycles.

21. The method of claim 1, wherein the universal indexed primers comprise:

a forward primer comprising the sequence of
AATGATACGGCGACCACCGAGATCTACACCTAGCGCTTACACTCTTTCCCTACAC
30 GACGCTCTTCCGATC*T; and

a reverse primer comprising the sequence of

CAAGCAGAAGACGGCATAACGAGATAAACCGCGGGTGACTGGAGTTCAGACGTGT
GCTCTTCCGATC*T.

- 5 22. The method of claim 1, wherein the methylated DNA pattern is reconstructed using a Bioinformatics analysis method selected from the group consisting of bwa-meth, Bismark, MethylDackel, bisulfite-treated reads analysis tools (BRAT), methyQA, mrsFAST, BSMAP, VerJInxer, RMAP-bs, MethylCoder, BS-seeker2, and Bison.
- 10 23. A kit for detecting methylated DNA pattern in DNA in a biological sample according to the method of claim 1, comprising:
- (a) a first enzyme capable of oxidizing 5-methylcytosine and 5-hydroxymethylcytosine of the DNA;
 - (b) a second enzyme capable of converting un-methylated cytosine of the DNA to uracil by
 - 15 deamination;
 - (c) a plurality of forward and reverse primer pairs specific to the converted sequence of target regions, wherein each forward primer of the plurality of forward and reverse primer pairs comprises a barcode sequence on its 5' end, wherein the barcode sequence of each forward primer is different from each other, wherein the target regions comprise one or more
 - 20 CpG sites;
 - (d) a plurality of universal indexed primers for creating the sequencing library;
 - (e) a first DNA polymerase capable of amplifying DNA with uracil bases, for amplification of converted DNA;
 - (f) a reagent capable of removing excess primers;
 - 25 (g) a second DNA polymerase capable of amplifying DNA, for creating the sequencing library; and
 - (h) sodium bisulfite.
24. The kit of claim 23, wherein the first enzyme is selected from the group consisting of
- 30 a Ten-eleven translocation (TET) enzyme or an isoform thereof; and wherein optionally, the TET enzyme is selected from the group consisting of TET1 enzyme or an isoform thereof, TET2 enzyme or an isoform thereof, and TET3 enzyme or an isoform thereof.

25. The kit of claim 23, wherein the second enzyme is a cytidine deaminase or an enzyme with cytidine deaminase properties;

wherein optionally, the cytidine deaminase is selected from the group consisting of APOBEC enzyme, CDA, and activation-induced cytidine deaminase; and

5 wherein optionally, the enzyme with cytidine deaminase properties is selected from the group consisting of M. SssI and M.HpaII.

26. The kit of claim 23, wherein the reagent is selected from the group consisting of paramagnetic beads and single-strand exonucleases; wherein optionally the paramagnetic
10 beads are selected from the group consisting of AMPure XP beads, SPRI beads, and Dynabeads.

27. The kit of claim 23, wherein the barcode sequence is an oligonucleotide comprising 10 to 16 random nucleotides, or 10 to 15 random nucleotides, or 10 to 13 random
15 nucleotides, or 10 random nucleotides, or 11 random nucleotides, or 12 random nucleotides, or 13 random nucleotides, or 14 random nucleotides, or 15 random nucleotides, or 16 random nucleotides; and wherein optionally, the barcode sequence is an oligonucleotide having 10 random nucleotides.

20 28. The kit of claim 23, wherein the number of the forward and reverse primer pairs is at least 5; and wherein optionally, the number of the forward and reverse primer pairs is at least 159.

29. The kit of claim 23, wherein each forward and reverse primer pair covers at least 1
25 CpG site.

30. The kit of claim 23, wherein the forward and reverse primer pairs comprises sequences as disclosed in Table 1.

30 31. The kit of claim 23, wherein the forward primer in the primer pair comprises one or more degenerate bases, and/or the reverse primer in the primer pair comprises one or more

degenerate bases; and wherein optionally, the degenerate base is selected from the group consisting of C, T, A and G.

32. The kit of claim 31, wherein each primer of the primer pair comprises 1, 2, or 3 degenerate bases; and wherein optionally, each primer of the primer pair has one degenerate base.

33. The kit of claim 23, wherein the primer pair does not comprise a degenerate base.

34. The kit of claim 23, wherein the universal indexed primers comprise:
a forward primer comprising the sequence of
AATGATACGGCGACCACCGAGATCTACACCTAGCGCTACACTCTTTCCCTACAC
GACGCTCTTCCGATC*T; and
a reverse primer comprising the sequence of
CAAGCAGAAGACGGCATACGAGATAAACCGCGGTGACTGGAGTTCAGACGTGT
GCTCTTCCGATC*T