CLAIMS

WHAT IS CLAIMED IS:

1. A method for providing hydroxymethylation state data of nucleic acids in a biological sample, the method comprising:
    a) obtaining the biological sample containing the nucleic acids;
    b) ligating oligonucleotide adapters to at least a portion of the nucleic acids in the biological sample, wherein the oligonucleotide adapters comprise 5hmC nucleotides, 5gmC nucleotides, 5caC nucleotides, 5cxmC nucleotides, or a combination thereof, thereby generating ligated nucleic acids;
    c) subjecting at least a portion of the ligated nucleic acids or a derivative thereof to a conversion condition that converts unmethylated and methylated cytosine nucleotides but not hydroxymethylated cytosine nucleotides of the ligated nucleic acids into uracil nucleotides, thereby generating converted nucleic acids; and
    d) sequencing at least a portion of the converted nucleic acids to obtain a nucleic acid sequence of the converted nucleic acids to provide the hydroxymethylation state data of the nucleic acids.

2. The method of claim 1, wherein the oligonucleotide adapters do not comprise cytosine nucleotides in flow cell binding regions or primer binding sites of the oligonucleotide adapters.

3. The method of claim 1, further comprising subjecting at least a portion of the ligated nucleic acids to glucosylation by β-glucosyltransferase (β-GT)/UDP-glucose to convert 5hmC nucleotides into 5gmC nucleotides after b) or prior to c).

4. The method of claim 1, wherein the conversion condition comprises bisulfite treatment, enzymatic treatment, or a combination thereof.

5. The method of claim 1, wherein the oligonucleotide adapters comprise 5hmC nucleotides.

6. The method of claim 1, wherein the oligonucleotide adapters comprise 5gmC nucleotides, 5caC nucleotides, 5cxmC nucleotides, or a combination thereof.

7. The method of claim 1, wherein the conversion condition comprises treatment with β-GT, a cytosine dioxygenase enzyme, carboxymethyltransferase, apolipoprotein B mRNA editing catalytic polypeptide-like protein (AID/APOBEC), or a combination thereof.

8. The method of claim 7, wherein the cytosine dioxygenase enzyme comprises ten eleven translocation protein 1 (TET1), ten eleven translocation protein 2 (TET2), ten eleven translocation protein 3 (TET3), or a functional variant thereof.

9. The method of claim 1, further comprising performing a sequence enrichment after b) or prior to c).

10. The method of claim 9, wherein the sequence enrichment comprises a target capture hybridization.

11. The method of claim 1, wherein at least a portion of the ligated nucleic acids are amplified prior to the sequencing.

12. The method of claim 1, wherein the oligonucleotide adapters are chemically synthesized using 5hmC phosphoramidites.

13. The method of claim 1, wherein the oligonucleotide adapters comprise 5gmC and 5caC nucleotides, wherein the oligonucleotide adapters are produced at least in part by synthesizing 5mC-containing oligonucleotides using phosphoramidite chemistry and enzymatically treating the 5mC-containing oligonucleotides with a TET enzyme and β-GT/UDP-glucose.

14. The method of claim 1, further comprising training a machine learning model to generate the hydroxymethylation profile using the hydroxymethylation state data.

15. The method of claim 1, further comprising training a machine learning model to generate a classifier using the hydroxymethylation state data.

16. The method of claim 1, further comprising aligning the nucleic acid sequence of the converted cfDNA to a reference nucleic acid sequence to determine the hydroxymethylation profile of the biological sample.

17. The method of claim 1, further comprising processing the hydroxymethylation state data using a machine learning model trained to be capable of distinguishing between healthy subjects and subjects with a cell proliferative disorder to provide an output value associated with a presence or a susceptibility of the cell proliferative disorder, thereby indicating the presence or the susceptibility of the cell proliferative disorder in the subject.

18. The method of claim 17, wherein the cell proliferative disorder comprises colorectal cancer, breast cancer, ovarian cancer, prostate cancer, lung cancer, pancreatic cancer, uterine cancer, liver cancer, esophagus cancer, stomach cancer, thyroid cancer, or bladder cancer.

19. The method of claim 17, wherein the machine learning model is tailored to detect the cell proliferative disorder at a pre-selected sensitivity and specificity.

20. The method of claim 17, wherein the machine learning model classifies the presence or the susceptibility of the cell proliferative disorder at a sensitivity of at least about 80%.

21. The method of claim 17, further comprising featurizing the hydroxymethylation state data and processing the featurized hydroxymethylation state data using a machine learning model that is trained to classify the biological sample into groups according to predesignated or preselected biological properties.

22. The method of claim 17, wherein the featurized hydroxymethylation state data correspond to properties of the nucleic acid sequence in the biological sample.

23. The method of claim 22, wherein the properties of the nucleic acid sequence are selected from presence or absence of pre-cancer, cancer or a stage of cancer, or a prognosis of cancer in the subject.

24. The method of claim 1, further comprising monitoring minimal residual disease in a subject previously treated for disease, wherein the monitoring comprises determining a hydroxymethylation profile as a baseline hydroxymethylation state, and further determining a hydroxymethylation profile at each of one or more pre-determined time points, wherein a change in hydroxymethylation profile from the baseline hydroxymethylation state indicates a change in the minimal residual disease status at the baseline hydroxymethylation state in the subject.

25. A method for detecting a cell proliferative disorder in a subject, the method comprising:

a) ligating oligonucleotide adapters to at least a portion of nucleic acids in a biological sample containing nucleic acids obtained from the subject, wherein the oligonucleotide adapters comprise 5hmC nucleotides, 5gmC nucleotides, 5caC nucleotides, 5cxmC nucleotides, or a combination thereof, thereby generating ligated nucleic acids;

b) subjecting at least a portion of the ligated nucleic acids to a conversion condition that converts unmethylated and methylated cytosine nucleotides in the ligated nucleic acids into uracil nucleotides, thereby generating converted nucleic acids;

c) sequencing at least a portion of the converted nucleic acids to obtain a nucleic acid sequence of the converted nucleic acids to provide hydroxymethylation state data of the nucleic acids; and

d) processing the hydroxymethylation state data using a machine learning model trained to be capable of distinguishing between healthy subjects and subjects with the cell proliferative disorder to provide an output value associated with a presence or a susceptibility of the cell proliferative disorder, thereby indicating the presence or the susceptibility of the cell proliferative disorder in the subject.