CLAIMS

What is claimed is

1. A method for sequencing hydroxymethylated cell-free DNA (cfDNA),

(a) adding adaptor sequences onto the ends of cfDNA;

(b) incubating the adaptor-ligated cfDNA with a DNA β-glucosyltransferase and UDP glucose modified with a chemoselective group, thereby covalently labeling the hydroxymethylated DNA molecules in the cfDNA with the chemoselective group;

(c) linking a biotin moiety to the chemoselectively-modified cfDNA via a cycloaddition reaction;

(d) enriching for the biotinylated DNA molecules by binding the product of step (c) to a support that binds to biotin;

(e) amplifying the enriched DNA using primers that bind to the adaptors; and

(f) sequencing the amplified DNA to produce a plurality of sequence reads,

wherein the method does not comprise releasing the biotinylated DNA molecules from the support after step (d), prior to step (e).

2. A method of sample analysis, comprising:

(a) identifying hydroxymethylated sequences in a cell-free DNA (cfDNA) sample from a patient by:

(i) ligating adaptor sequences to the cfDNA, (ii) modifying 5-hydroxymethylcytosine (5hmC) residues in the cfDNA to contain a capture tag; (iii) enriching for cfDNA comprising hydroxymethylcytosine residues by linking the cfDNA to a support via the capture tag, thereby providing an enriched composition comprising support-bound hydroxymethylated cfDNA, (iv) amplifying the cfDNA molecules without releasing the hydroxymethylated cfDNA from the support, to provide amplified DNA, and (v) sequencing the amplified DNA to produce a plurality of sequence reads, and (vi) identifying hydroxymethylated sequences in the sequence reads;

(b) comparing the identified hydroxymethylated sequences with a set of signature hydroxymethylation sequences that are correlated with a phenotype; and

(c) providing a report indicating a correlation with the phenotype.

3. The method of claim 2, wherein the capture tag comprises a biotin moiety.

4. The method of claim 3, wherein the capture tag is added to 5-hydroxymethylcytosine residues in the cfDNA in step (a)(ii) by a method comprising: incubating the adaptor-ligated cfDNA with a DNA ß-glucosyltransferase and UDP glucose modified with a chemoselective group, thereby covalently binding the chemoselective group to 5hmC residues in the cfDNA; and reacting a biotin moiety with the chemoselective group to provide biotinylated, hydroxymethylated cfDNA.

5. The method of claim 4, wherein the cfDNA is amplified in step (a)(iv) using primers that bind to the adaptors.

6. The method of claim 5, further comprising, after step (a)(iii) and prior to step (a)(iv), washing the support and setting up an amplification reaction containing the support, without releasing the biotinylated DNA molecules from the support.

7. The method of claim 4, wherein the UDP glucose modified with a chemoselective group comprises UDP-6-N_3-Glu, the biotin moiety comprises dibenzocyclooctyne-modified biotin, and the support comprises avidin or streptavidin.

8. The method of claim 2, wherein the phenotype is a disease, condition, or clinical outcome.

9. The method of claim 2, wherein step (a)(vi) is quantitative.

10. The method of claim 9, wherein step (a) further comprises determining the hydroxymethylation level of each identified hydroxymethylated sequence ~~and~~ to provide a

first hydroxymethylation pattern comprising the hydroxymethylation level at each identified sequence.

11. The method of claim 10, further comprising obtaining a second sample of cfDNA from the patient at a different time point, and repeating step (a) with the second sample to generate a second hydroxymethylation pattern.

12. The method of claim 11, further comprising comparing the second hydroxymethylation pattern with the first hydroxymethylation pattern to determine any change in hydroxymethylation over time.

13. The method of claim 12, wherein the comparison results in a map of the changes in hydroxymethylation in the course of a disease, condition, or treatment of a disease or condition.

14. The method of claim 10, wherein the set of signature hydroxymethylation sequences comprises a map of target loci in a control hydroxymethylation profile, and step (b) comprises comparing the hydroxymethylation level for each identified hydroxymethylated sequence with the hydroxymethylation level at the corresponding target locus.

15. The method of claim 14, further comprising determining whether the hydroxymethylation level for each identified hydroxymethylation sequence is over-represented or under-represented relative to the hydroxymethylation level at the corresponding target locus.

16. The method of claim 15, wherein the method further comprises: (e) making a diagnosis, a treatment decision or a prognosis bases on results based on the identity of the hydroxymethylated sequences that are over-represented or under-represented relative to the corresponding target loci.

17. The method of claim 16, wherein the diagnosis, treatment decision or prognosis is a cancer diagnosis.

18. The method of claim 17, wherein the target loci include one or more of the following gene bodies: ABRACL, ADAMTS4, AGPG2, ALDH1A3, ALG lOB, AMOTLl, APCDDlL-ASl, ARL6IP6, ASPlB, ATP6V0A2, AUNIP, BAGE, C2orf62, C8orf22, CALCB, CC2D1B, CCDC33, CCNL2, CLDN15, COMMD6, CPLX2, CRP, CTRC, DACHl, DAZL, DDXllLl, DHRS3, DUSP26, DUSP28, EPN3, EPPIN-WPDC6, ETAAl, PAM96A, FENDRR, PLJ16779, PLJ31813, GBXl, GLP2R, GMCLlPl, GNPDA2, GPR26, GSTPl, HMOX2, HOXC5, IGSP9B, INSC, INSL4, IRF7, KIP16B, KIP20B, LARS, LDHD, LHX5, LINC00158, LINC00304, LOC100128946, LOC100131234, LOC100132287, LOC100506963, LOC100507250, LOC100507410, LOC255411, LOC729737, MAPP, NPAS4, NRADDP, P2RX2, PAIPl, PAXl, PODXL2, POU4P3, PSMGl, PTPN2, RAGl, RBM14-RBM4, RDHll, RPPL3, RNP122, RNP223, RNP34, SAMDll, SHISA2, SIGLEClO, SLAMP7, SLC25A46, SLC25A47, SLC9A3R2, SORD, SOX18, SPATA31El, SSR2, STXBP3, SYTl1, SYT2, TCEA3, THAP7-AS1, TMEM168, TMEM65, TMX2, TPM4, TPO, TRAMl, TTC24, UBQLN4, WASH7P, ZNF284, ZNF423, ZNF444, ZNF800, ZNF850, and ZRANB2.

19. The method of claim 18, wherein the adaptor sequences comprise a molecular barcode.

20. The method of claim 19, wherein the molecular barcode comprises a sample identifier sequence and a molecule identifier sequence.

21. The method of claim 2, wherein prior to step (a), a spike-in control composition is combined with sample.

22. The method of claim 21, wherein the spike-in control composition comprises three amplicons synthesized from a cocktail of dATP, dGTP, dTTP, and (1) dCTP, (2) dmCTP, or (3) dhmCTP and dCTP.

23. A kit for analyzing cfDNA, comprising:

DNA ß-glucosyltransferase;

UDP glucose modified with a chemoselective group;

an adaptor comprising at least one molecular barcode; and

a spiked-in control comprising three amplicons synthesized from a cocktail of dATP, dGTP, dTTP, and (1) dCTP, (2) dmCTP, or (3) dhmCTP and dCTP.

24. The kit of claim 23, wherein the at least one molecular barcode comprises a sample identifier sequence and a molecule identifier sequence.

25. A sample, comprising:

a pool of cell-free, adapter-ligated DNA molecules that are from a plurality of different sources and comprise one or more hydroxymethylcytosines modified to contain a capture tag, the DNA molecules further including molecular barcodes to indicate their source and allow sequences from different sources to be distinguished after analysis; and

a spiked-in control comprising three amplicons synthesized from a cocktail of dATP, dGTP, dTTP, and (1) dCTP, (2) dmCTP, or (3) dhmCTP and dCTP.