

CLAIMS

What is claimed is:

1. A method of sequencing hydroxymethylated cell-free DNA (cfDNA), comprising:
adding an affinity tag to only hydroxymethylated DNA molecules in a sample of cfDNA;
enriching for the DNA molecules that are tagged with the affinity tag by binding to a support; and
sequencing the enriched DNA molecules.
2. The method of claim 1, wherein the method comprises:
 - (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.
3. The method of claim 2, wherein the method does not comprise releasing the biotinylated DNA molecules from the support after step (d), prior to step (e).
4. The method of claim 3, wherein step (e) comprises:
 - i. washing the support of (d) after the biotinylated DNA molecules have bound to the support; and then
 - ii. setting up an amplification reaction containing the support, without releasing the biotinylated DNA molecules from the support.

5. The method of any of claims 2-4, wherein the adding adaptor sequences step (a) comprises ligating a Y or hairpin adaptor onto the ends of the cfDNA.
6. The method of any of claims 2-5, wherein the cycloaddition reaction is between an azido group and an alkynyl group.
7. The method of any of claims 2-6, wherein the UDP glucose modified with chemoselective group of step (b) is UDP-6-N₃-Glu.
8. The method of any of claims 2-7, wherein the biotin moiety of step (c) is dibenzocyclooctyne-modified biotin.
9. The method of any of claims 2-8, wherein the DNA β -glucosyltransferase of step (b) is T4 DNA β -glucosyltransferase.
10. The method of any of claims 2-9, wherein the biotin moiety of step (c) comprises biotin.
11. The method of any of claims 2-10, wherein the support that binds to biotin in step (d) comprises streptavidin beads.
12. The method of claim 11, wherein the beads are magnetic.
13. The method of any prior claim, wherein the method quantitatively determines which sequences are hydroxymethylated in the cfDNA.
14. A method for identifying a hydroxymethylation pattern that correlates with a phenotype, comprising:
 - (a) performing the method of any of claims 1-13 on a plurality of cfDNA samples, wherein the cfDNA samples are isolated from patients having a known phenotype, thereby determining which sequences are hydroxymethylated in cfDNA from each of the patients; and
 - (b) identifying a hydroxymethylation signature that is correlated with the phenotype.

15. The method of claim 14, wherein the phenotype is a disease, condition or clinical outcome.
16. The method of claim 14, wherein the hydroxymethylation signature is diagnostic, prognostic or theranostic.
17. A method of sample analysis, comprising:
 - (a) identifying, using the method of any of claims 1-13, sequences that are hydroxymethylated in cfDNA of a patient;
 - (b) comparing the sequences identified in step (a) to a set of signature sequences that are correlated with a phenotype;
 - (c) providing a report indicating a correlation with the phenotype.
18. The method of claim 17, further comprising making a diagnosis, prognosis or recommending a therapy based on the results of the comparison of (b).
19. A method of sample analysis, comprising:
 - (a) determining, using the method of any of claims 1-13, which sequences are hydroxymethylated in a first sample of cfDNA and which sequences are hydroxymethylated in the second sample of cfDNA, wherein the first and second samples of cfDNA are obtained from the same subject at two different time points; and
 - (b) comparing the hydroxymethylation pattern for the first sample to the hydroxymethylation pattern for the second sample to determine if there has been a change in hydroxymethylation over time.
20. The method of claim 19, wherein the determining step (a) is quantitative.
21. The method of claim 20, wherein the comparing step (b) comprises comparing the level of hydroxymethylation of one or more selected sequences.
22. The method of claim 19, wherein the comparison of (b) results in a map of the changes in hydroxymethylation in the course of a disease, condition, or a treatment of a disease or condition.

23. A composition comprising circulating cell-free DNA, wherein the hydroxymethylcytosines residues in the DNA are modified to contain a capture tag.
24. The composition of claim 23, wherein the circulating cell-free DNA is in double-stranded form.
25. The composition of any of claims 23-24, wherein the capture tag is a biotin moiety.
26. The composition of any of claims 23-24, wherein the capture tag is a chemoselective group .
27. The composition of any of claims 23-26, further comprises a β -glucosyltransferase and a UDP glucose modified with a chemoselective group.
28. The composition of any of claims 23-27, wherein the cell-free hydroxymethylated DNA is adaptor-ligated.
29. The composition of any of claims 23-28, wherein at least 10% of the nucleic acid molecules in the composition comprise one or more hydroxymethylcytosines that are modified to contain a capture tag.
30. The composition of any of claims 23-29, wherein the composition further comprises a support, wherein the support and circulating cell-free DNA are linked to one another via the capture tag.
31. The composition of any of claim 29, wherein the composition further comprises copies of the cell-free hydroxymethylated DNA that have been made by PCR.
32. The composition of claim 30 or 31, wherein the support and the cell-free DNA are linked to each other by a covalent bond.
33. The composition of claim 30 or 31, wherein the support and the circulating cell-free DNA are linked to each other by a non-covalent bond.

34. The composition of claim 31, wherein the support is linked to streptavidin and the capture agent is biotin.
35. A method comprising:
- (a) obtaining a sample comprising circulating cell-free DNA;
 - (b) enriching for the hydroxymethylated DNA in the sample; and
 - (c) independently quantifying the amount of nucleic acids in the enriched hydroxymethylated DNA that map to each of one or more target loci.
36. The method of claim 35, wherein the method further comprises:
- (d) determining whether one or more nucleic acid sequences in the enriched hydroxymethylated DNA are over-represented or under represented in the enriched hydroxymethylated DNA.
37. The method of claim 36, wherein the method further comprises:
- (e) making a diagnosis, a treatment decision or a prognosis bases on results based on the identity of the nucleic acids that are over-represented or under represented in the enriched hydroxymethylated DNA.
38. The method of claim 37, wherein the diagnosis, treatment decision or prognosis, wherein the diagnosis, treatment decision or prognosis is a cancer diagnosis.
39. The method of any of claims 35-38, wherein the target loci include one or more of the following gene bodies: ABRACL, ADAMTS4, AGFG2, ALDH1A3, ALG10B, AMOTL1, APCDD1L-AS1, ARL6IP6, ASF1B, ATP6V0A2, AUNIP, BAGE, C2orf62, C8orf22, CALCB, CC2D1B, CCDC33, CCNL2, CLDN15, COMMD6, CPLX2, CRP, CTRC, DACH1, DAZL, DDX11L1, DHRS3, DUSP26, DUSP28, EPN3, EPPIN-WFDC6, ETAA1, FAM96A, FENDRR, FLJ16779, FLJ31813, GBX1, GLP2R, GMCL1P1, GNPDA2, GPR26, GSTP1, HMOX2, HOXC5, IGSF9B, INSC, INSL4, IRF7, KIF16B, KIF20B, LARS, LDHD, LHX5, LINC00158, LINC00304, LOC100128946, LOC100131234, LOC100132287, LOC100506963, LOC100507250, LOC100507410, LOC255411, LOC729737, MAFF, NPAS4, NRADDP, P2RX2, PAIP1, PAX1, PODXL2, POU4F3, PSMG1, PTPN2, RAG1, RBM14-RBM4, RDH11, RFPL3, RNF122, RNF223, RNF34, SAMD11, SHISA2, SIGLEC10, SLAMF7, SLC25A46, SLC25A47, SLC9A3R2, SORD, SOX18, SPATA31E1,

SSR2, STXBP3, SYT11, SYT2, TCEA3, THAP7-AS1, TMEM168, TMEM65, TMX2, TPM4, TPO, TRAM1, TTC24, UBQLN4, WASH7P, ZNF284, ZNF423, ZNF444, ZNF800, ZNF850, and ZRANB2.

40. The method of any of claims 35-39, wherein the target loci include one or more of the following intervals in the hg19 reference genome: chr1:114670001-114672000, chr1:169422001-169424000, chr1:198222001-198224000, chr1:239846001-239848000, chr1:24806001-24808000, chr1:3234001-3236000, chr1:37824001-37826000, chr1:59248001-59250000, chr1:63972001-63974000, chr1:67584001-67586000, chr1:77664001-77666000, chr2:133888001-133890000, chr2:137676001-137678000, chr2:154460001-154462000, chr2:200922001-200924000, chr2:213134001-213136000, chr2:219148001-219150000, chr2:41780001-41782000, chr2:49900001-49902000, chr3:107894001-107896000, chr3:108506001-108508000, chr3:137070001-137072000, chr3:17352001-17354000, chr3:23318001-23320000, chr3:87312001-87314000, chr3:93728001-93730000, chr4:39342001-39344000, chr4:90790001-90792000, chr5:103492001-103494000, chr5:39530001-39532000, chr5:83076001-83078000, chr6:122406001-122408000, chr6:129198001-129200000, chr6:156800001-156802000, chr6:157286001-157288000, chr6:45304001-45306000, chr7:11020001-11022000, chr7:13364001-13366000, chr8:42934001-42936000, chr8:53686001-53688000, chr8:69672001-69674000, chr9:3496001-3498000 and chr9:88044001-88046000.

41. The method of any of claims 38-40, wherein cancer diagnosis includes an indication of the tissue-type of the cancer.

42. The method of any of claims 35-41, wherein step (c) is done by sequencing, digital PCR or hybridization to an array.

43. The method of any of claims 35-42, wherein the determining step is done relative to a control, wherein the control comprises one or more control sequences in:

- the enriched hydroxymethylated DNA;
- the sample of (a);
- the sample of (a), after the hydroxymethylated DNA has been removed in step (b); or
- a different sample.

44. A method of sample analysis comprising:
attaching labels to DNA molecules that comprise one or more hydroxymethylcytosine and methylcytosine nucleotides in a sample of cfDNA, wherein the hydroxymethylcytosine nucleotides are labeled with a first fluorophore and the methylcytosine nucleotides are labeled with a second fluorophore that is distinguishable from the first label, to produce a labeled sample; and
hybridizing the labeled sample to an array comprising at least one probe for a gene or interval of Tables 10A, 10B, 11A and 11B.
45. The method of claim 44, wherein the array comprises top strand probes and bottom strand probes.
46. A method of sample analysis comprising:
attaching labels to DNA molecules that comprise one or more hydroxymethylcytosine and methylcytosine nucleotides in a sample of cfDNA, wherein the hydroxymethylcytosine nucleotides are labeled with a first capture tag and the methylcytosine nucleotides are labeled with a second capture tag that is different to the first capture, to produce a labeled sample; and
enriching for the DNA molecules that are labeled; and
sequencing the enriched DNA molecules.
47. The method of claim 46, wherein the method comprises separately enriching the DNA molecules that comprise one or more hydroxymethylcytosines and the DNA molecules that comprise one or more methylcytosine nucleotides.
48. A kit for analyzing cfDNA, comprising:
a DNA β -glucosyltransferase
UDP glucose modified with a chemoselective group; and
an adaptor comprising a molecular barcode.
49. The kit of claim 48, wherein the adaptor is Y or hairpin adaptor.
50. The kit of claim 48 or 49, further comprising a biotin moiety, wherein the biotin moiety is reactive with the chemoselective group.