

large number of whole-genome, base-resolution bisulfite sequencing data sets for many tissue and cell types. These public data sets, in combination with additional WGBS data generated, allowed full-genome characterization of local coupled CpG methylation across the largest set of human tissue types available to date to be performed, and annotate these blocks of co-methylated CpGs as a distinct set of genomic features.

[0005] DNA methylation is cell-type specific, and the pattern can be harnessed for deconvoluting the relative cell composition of heterogeneous samples, such as different white blood cells in whole blood (Houseman et al. (2016)), fetal components in maternal cell-free DNA (Sun et al. (2015)), or circulating tumor DNA in plasma (Sun et al. (2015)). Most of these recent efforts rely on the methylation level of individual CpG sites, and are fundamentally limited by the technical noise and sensitivity in measuring single CpG methylation. Very recently, Lehmann-Werman et al. demonstrated a superior sensitivity with multi-CpG haplotypes in detecting tissue-specific signatures in circulating DNA (Lehmann-Werman et al. (2016)). The markers in that study were discovered from Infinium 450k methylation array data, which represent only a very limited fraction of the human genome.

SUMMARY

[0006] Some examples described herein provide accurate tissue-of-origin mapping based on comparing the patterns and abundance of methylation haplotypes against a reference set of human reference tissues and provide accurate quantitative estimation of the cancer DNA fraction.

[0007] Some examples are described in the following numbered paragraphs:

1. In one aspect, the present disclosure refers to a method for detecting the presence of one or more nucleic acids indicative of a health condition, tissue of origin, germ layer of origin or organ of origin in a mixture of nucleic acids comprising: performing methylation analysis on a sample comprising a plurality of nucleic acids by sequencing the plurality of nucleic acids and identifying methylation haplotype blocks, wherein the methylation haplotype blocks comprise a plurality of methylation sites for which methylation status is coordinated; determining a methylation haplotype load (MHL) of the methylation haplotype blocks, wherein said MHL comprises the normalized fraction of methylated haplotypes at different lengths; and determining based on the MHL whether said sample includes a plurality of methylation haplotype blocks indicative of the presence one or more nucleic acids indicative of a health condition, tissue of origin, germ layer of origin, organ of origin or any combination thereof, ~~wherein said methylation haplotype blocks comprise a plurality of methylation sites for which the methylation status is coordinated.~~

12. The method of any one of Paragraphs 1-11, wherein said methylation haplotype blocks have an average size of 95bp.

13. The method of any one of Paragraphs 1-12, wherein said methylation haplotype blocks have a minimum of 3 CpGs per block.

5 14. The method of any one of Paragraphs 1-13 further comprising quantifying the level of said plurality of methylation haplotype blocks indicative of the presence of one or more nucleic acids indicative of a health condition, tissue of origin, germ layer of origin, organ of origin or any combination thereof in said sample.

10 15. The method of Paragraph 6 further comprising quantifying the level of said plurality of methylation haplotype blocks indicative of the presence of a tumor in said sample and quantifying the level of said plurality of methylation haplotype blocks indicative of the presence of one or more nucleic acids indicative of a normal tissue or normal organ corresponding to the tissue or organ of origin of said tumor in said sample.

15 16. In another aspect, the present disclosure refers to a method of identifying methylation haplotype blocks indicative of a cancer or tissue of origin in a plurality of nucleic acids comprising: sequencing the plurality of nucleic acids;

determining methylation haplotypes in ~~a the plurality of nucleic acids segments;~~
combining the methylation haplotypes and calculating methylation linkage disequilibrium on
the combined methylation haplotypes; ~~and~~ partitioning each segment into a plurality of
20 methylation haplotype blocks, wherein said methylation haplotype blocks comprise a plurality
of methylation sites for which the methylation status is coordinated; calculating the group
specific index (GSI) based on the normalized fraction of methylated haplotypes at different
methylation haplotype block lengths.

25 17. The method of Paragraph 16, wherein said methylation haplotypes are determined for a portion of a genome.

18. The method of any one of Paragraphs 16 and 17, wherein said methylation haplotypes are determined across a whole genome.

30 19. The method of any one of Paragraphs 16-18, wherein said methylation haplotype blocks are defined as the genomic region in which the r^2 value of two adjacent CpG sites is no less than 0.5.

20. The method of any one of Paragraphs 16-19, wherein said methylation haplotype blocks are identified in nucleic acids from a tumor tissue.

21. The method of any one of Paragraphs 16-20, wherein said methylation haplotype blocks are identified in nucleic acids from a known type of tissue.

WHAT IS CLAIMED IS:

1. A method for detecting the presence of one or more nucleic acids indicative of a health condition, tissue of origin, germ layer of origin or organ of origin in a mixture of nucleic acids comprising:

performing methylation analysis on a sample comprising a plurality of nucleic acids by sequencing the plurality of nucleic acids and identifying methylation haplotype blocks, wherein the methylation haplotype blocks comprise a plurality of methylation sites for which methylation status is coordinated;

determining a methylation haplotype load (MHL) of the methylation haplotype blocks, wherein said MHL comprises the normalized fraction of methylated haplotypes at different lengths; and

determining based on the MHL whether said sample includes a plurality of methylation haplotype blocks indicative of the presence of one or more nucleic acids indicative of a health condition, tissue of origin, germ layer of origin, organ of origin or any combination thereof, wherein said methylation haplotype blocks comprise a plurality of methylation sites for which the methylation status is coordinated.

2. The method of Claim 1, wherein said methylation analysis is performed on cell-free DNA.

3. The method of any one of Claims 1 and 2, wherein said methylation analysis is performed on cell-free DNA in a blood sample.

4. The method of any one of Claims 1-3, wherein said plurality of methylation haplotype blocks comprises at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least 500 or more than 500 methylation haplotype blocks.

5. The method of any one of Claims 1-4, wherein said health condition is a tumor.

6. The method of Claim 5, further comprising determining whether said sample includes a plurality of methylation haplotype blocks indicative of the presence of one or more nucleic acids indicative of a normal tissue or normal organ corresponding to the tissue or organ of origin of said tumor.

7. The method of any one of Claims 1-6, wherein said health condition is fetal aneuploidy.

8. The method of any one of Claims 1-7, wherein said sample is a blood sample.
9. The method of any one of Claims 1-8, further comprising quantitating the level of said one or more nucleic acids indicative of a health condition, tissue of origin, organ of origin or any combination thereof in said sample.
10. The method of any one of Claims 1-9, wherein said methylation analysis is performed using a technique selected from the group consisting of bisulfite methylation analysis, reduced representation bisulfite sequencing, WGBS, BSPP, micro-droplet PCR, selector probe based methods, and MeDiP.
11. The method of any one of Claims 1-10, ~~further comprising determining a methylation haplotype load for each methylation haplotype block, wherein said methylation haplotype load comprises the normalized fraction of methylated haplotypes at different lengths~~wherein the presence of the one or more nucleic acids indicative of the health condition, tissue of origin, germ layer of origin or organ of origin is confirmed when the MHL of a particular methylation haplotype block in the mixture is greater than 0.4 and the MHL of the particular methylation haplotype block is less than 0.05 in a corresponding mixture which does not contain the one or more nucleic acids indicative of the health condition, tissue of origin, germ layer of origin or organ of origin.
12. The method of any one of Claims 1-~~10~~11, further comprising determining a unmethylated haplotype load for each methylation haplotype block, wherein said unmethylated haplotype load comprises the normalized fraction of unmethylated haplotypes at different lengths.
13. The method of any one of Claims 1-~~12~~11, wherein said methylation haplotype blocks have an average size of 95bp.
14. The method of any one of Claims 1-~~13~~12, wherein said methylation haplotype blocks have a minimum of 3 CpGs per block.
15. The method of any one of Claims 1-~~14~~13, further comprising quantifying the level of said plurality of methylation haplotype blocks indicative of the presence of one or more nucleic acids indicative of a health condition, tissue of origin, germ layer of origin, organ of origin or any combination thereof in said sample.
16. The method of Claim 6, further comprising quantifying the level of said plurality of methylation haplotype blocks indicative of the presence of a tumor in said sample and quantifying the level of said plurality of methylation haplotype blocks indicative of the

presence of one or more nucleic acids indicative of a normal tissue or normal organ corresponding to the tissue or organ of origin of said tumor in said sample

17. A method of identifying methylation haplotype blocks indicative of a cancer or tissue of origin in a plurality of nucleic acids comprising:

sequencing the plurality of nucleic acids;

determining methylation haplotypes in ~~a~~the plurality of nucleic acids segments;

combining the methylation haplotypes and calculating methylation linkage disequilibrium on the combined methylation haplotypes; ~~and~~

partitioning each segment into a plurality of methylation haplotype blocks, wherein said methylation haplotype blocks comprise a plurality of methylation sites for which the methylation status is coordinated;

calculating the group specific index (GSI) based on the normalized fraction of methylated haplotypes at different methylation haplotype block lengths.

18. The method of Claim 17, wherein said methylation haplotypes are determined for a portion of a genome.

19. The method of any one of Claims 17 and 18, wherein said methylation haplotypes are determined across a whole genome.

20. The method of any one of Claims 17-19, wherein said methylation haplotype blocks are defined as the genomic region in which the r^2 value of two adjacent CpG sites is no less than 0.5.

21. The method of any one of Claims 17-20, wherein said methylation haplotype blocks are identified in nucleic acids from a tumor tissue.

22. The method of any one of Claims 17-21, wherein said methylation haplotype blocks are identified in nucleic acids from a known type of tissue.

23. The method of any one of Claims 17-22, wherein said methylation haplotype blocks are identified in nucleic acids from a fetus.

24. The method of any one of Claims 17-23, wherein said methylation haplotype blocks are identified in nucleic acids from an embryonic stem cell.

25. The method of any one of Claims 17-24, wherein said methylation haplotype blocks are identified in nucleic acids from a known germ layer.

26. The method of any one of Claims 17-25, wherein said methylation haplotype blocks have an average size of 95bp.

27. The method of any one of Claims 17-26, wherein said methylation haplotype blocks have a minimum of 3 CpGs per block.

28. The method of any one of Claims 17-27, wherein said methylation haplotype blocks are identified in nucleic acid regions from a Whole Genome Bisulfite Sequencing analysis.

29. The method of any one of Claims 17-28, wherein said methylation haplotype blocks are identified in data sets from methylation analysis of ENCODE cell lines or tissue samples.

~~30. The method of any one of Claims 17-29, wherein said methylation haplotype blocks are identified in data sets from methylation analysis of Infinium HumanMethylation450 BeadChip (HM450K).~~

~~31.~~30. The method of any one of Claims 17-~~29~~30, further comprising calculating the pairwise correlation coefficient of adjacent CpG methylation levels across different sample sets for block partitioning.

~~32.~~31. The method of any one of Claims 17-~~30~~31, further comprising determining methylation haplotype load for each methylation haplotype block, wherein said methylation haplotype load comprises the normalized fraction of methylated haplotypes at different lengths.

~~33.~~32. The method of any one of Claims 17-31, further comprising determining unmethylated haplotype load for each methylation haplotype block, wherein said unmethylated haplotype load comprises the normalized fraction of unmethylated haplotypes at different lengths.