[0008] Bisulfite sequencing has also been used to detect cytosine methylation in RNA. Unlike other methods for detecting 5mC in RNA such as methylated-RNA-immunoprecipitation, RNA-bisulfite-sequencing (RNA-BS-seq) has the advantage of being able to determine of the extent of methylation of a specific C position in RNA. RNA-BS-seq, however, suffers from the same drawbacks described above for bisulfite sequencing of DNA. In particular, the reaction conditions can cause substantial degradation of RNA.

[0009] There is a need for a method for DNA methylation and hydroxymethylation analysis that is a mild reaction that can detect the modified cytosine (5mC and 5hmC) at base-resolution quantitatively without affecting the unmodified cytosine. Likewise, there is a need for a method for RNA methylation and hydroxymethylation analysis that employs mild reaction conditions and can detect the modified cytosine quantitatively at base resolution without affecting the unmodified cytosine.

SUMMARY

[0010] The present disclosure provides methods for identifying the location of one or more of 5-methylcytosine, 5- hydroxymethylcytosine, 5-carboxylcytosine and/or 5-formylcytosine in a nucleic acid. The methods described herein provide for DNA or RNA methylation and hydroxymethylation analysis involving mild reactions that detect the modified cytosine quantitatively with base-resolution without affecting the unmodified cytosine. Provided herein is a new method for identifying 5mC and 5hmC by combining TET oxidation and reduction by borane derivatives (e.g., pyridine borane and 2-picoline borane (pic-BH_3)), referred to herein as TAPS (**T**ET **A**ssisted **P**yridine borane **S**equencing) (Table 1). TAPS detects modifications directly with high sensitivity and specificity, without affecting unmodified cytosines, and can be adopted to detect other cytosine modifications. It is non-destructive, preserving RNA and DNA up to 10 kbs long. Compared with bisulfite sequencing, TAPS results in higher mapping rates, more even coverage and lower sequencing costs, enabling higher quality, more comprehensive and cheaper methylome analyses. Variations of this method that do not employ the oxidation step are used to identify 5fC and/or 5caC as described herein.

[0011] In one example, the present disclosure provides a method for identifying 5-methylcytosine (5mC) in a target nucleic acid comprising the steps of:

a. providing a nucleic acid sample comprising the target nucleic acid;
b. modifying the nucleic acid comprising the steps of:

i. adding a blocking group to the 5-hydroxymethylcytosine (5hmC) in the nucleic acid sample;

ii. converting the 5mC in the nucleic acid sample to 5-carboxylcytosine (5caC) and/or 5-formylcytosine (5fC); and

iii. converting the 5caC and/or 5fC to dihydrouracil (DHU) to provide a modified nucleic acid sample comprising a modified target nucleic acid; and

c. detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of a 5mC in the target nucleic acid.

[0012] In examples of the method for identifying 5mC in a target nucleic acid, the percentages of the T at each transition location provide a quantitative level of 5mC at each location in the target nucleic acid. In examples, the nucleic acid is DNA. In other examples, nucleic acid is RNA.

[0013] In another example, the present disclosure provides a method for identifying 5mC or 5hmC in a target nucleic acid comprising the steps of:

a. providing a nucleic acid sample comprising the target nucleic acid;
b. modifying the nucleic acid comprising the steps of:
   i. converting the 5mC and 5hmC in the nucleic acid sample to 5-carboxylcytosine (5caC) and/or 5fC; and
   ii. converting the 5caC and/or 5fC to DHU to provide a modified nucleic acid sample comprising a modified target nucleic acid; and
c. detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of either a 5mC or 5hmC in the target nucleic acid.

[0013a] In one aspect, the present disclosure refers to a method for identifying 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) in a target nucleic acid comprising the steps of: providing a nucleic acid sample comprising the target nucleic acid; modifying the target nucleic acid comprising the steps of: converting the 5mC and 5hmC in the nucleic acid sample to 5-carboxylcytosine (5caC) and/or 5-formylcytosine (5fC) by contacting the nucleic acid sample with a ten-eleven translocation (TET) enzyme so that one or more 5caC or 5fC residues are generated; and converting the 5caC and/or 5fC to dihydrouracil (DHU) by treating the target nucleic acid with a borane reducing agent to provide a modified nucleic acid sample comprising a modified target nucleic acid; and detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition or a cytosine (C) to DHU transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of either a 5mC or 5hmC in the target nucleic acid.

[0014] In examples of the method for identifying 5mC or 5hmC, the percentages of the T at each transition location provide a quantitative level of 5mC or 5hmC at each location in the target nucleic acid. In examples, the nucleic acid is DNA. In other examples, nucleic acid is RNA.

[0015] In another example, the disclosure provides a method for identifying 5mC and identifying 5hmC in a target nucleic acid comprising:

a. identifying 5mC in the target nucleic acid comprising the steps of:
   i. providing a first nucleic acid sample comprising the target nucleic acid;

ii. modifying the nucleic acid in the first sample comprising the steps of:
    1. adding a blocking group to the 5-hydroxymethylcytosine (5hmC) in the first nucleic acid sample;
    2. converting the 5mC in the first nucleic acid sample to 5caC and/or 5fC; and
    3. converting the 5caC and/or 5fC to DHU to provide a modified first DNA sample comprising a modified target nucleic acid;
iii. optionally amplifying the copy number of the modified target nucleic acid; and
iv. detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of a 5mC in the target nucleic acid.

b. identifying 5mC or 5hmC in the target nucleic acid comprising the steps of:
i. providing a second nucleic acid sample comprising the target nucleic acid;
ii. modifying the nucleic acid in the second sample comprising the steps of:
    1. converting the 5mC and 5hmC in the second nucleic acid sample to 5caC and/or 5fC; and
    2. converting the 5caC and/or 5fC to DHU to provide a modified second nucleic acid sample comprising a modified target nucleic acid;
iii. optionally amplifying the copy number of the modified target nucleic acid;
iv. detecting the sequence of the modified target nucleic acid from the second sample; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of either a 5mC or 5hmC in the target nucleic acid; and

c. comparing the results of steps (a) and (b), wherein a C to T transitions present in step (b) but not in step (a) provides the location of 5hmC in the target nucleic acid.

[0016] In examples for identifying 5mC and identifying 5hmC in a target nucleic acid, in step (a) the percentages of the T at each transition location provide a quantitative level of 5mC in the target nucleic acid; in step (b), the percentages of the T at each transition location provide a quantitative level of 5mC or 5hmC in the target nucleic acid; and in step (c) the differences in percentages for a C to T transition identified in step (b), but not in step (a) provides the quantitative level of a 5hmC at each location in the target nucleic acid. In examples, the nucleic acid is DNA. In other examples, nucleic acid is RNA.

[0017] In examples of the disclosure, the blocking group added to 5hmC in the nucleic acid sample is a sugar. In examples, the sugar is a naturally-occurring sugar or a modified sugar, for example glucose or a modified glucose. In examples of the disclosure, the blocking group is added to 5hmC by contacting the nucleic acid sample with UDP linked to a sugar, for example UDP-glucose or UDP linked to a modified glucose in the presence of a glucosyltransferase enzyme, for example, T4 bacteriophage β-glucosyltransferase (βGT) and T4 bacteriophage α-glucosyltransferase (αGT) and derivatives and analogs thereof.

[0018] In examples of the disclosure, the step of converting the 5mC in the nucleic acid sample to 5caC and/or 5fC and the step of converting the 5mC and 5hmC in the nucleic acid sample to 5caC and/or 5fC each comprises contacting the nucleic acid sample with a ten eleven translocation (TET) enzyme. In further examples, the TET enzyme is one or more of human TET1, TET2, and TET3; murine Tet1, Tet2, and Tet3; *Naegleria* TET (NgTET); *Coprinopsis cinerea* (CcTET) and derivatives or analogues thereof. In examples, the TET enzyme is NgTET.

[0019] In another example, the disclosure provides a method for identifying 5caC or 5fC in a target nucleic acid comprising the steps of:

a. providing a nucleic acid sample comprising the target nucleic acid;
b. converting the 5caC and 5fC to DHU to provide a modified nucleic acid sample comprising a modified target nucleic acid;
c. optionally amplifying the copy number of the modified target nucleic acid; and
d. detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of either a 5caC or 5fC in the target nucleic acid.

[0020] In examples of the method for identifying 5caC or 5fC in a target nucleic acid, the percentages of the T at each transition location provide a quantitative level for 5caC or 5fC at each location in the target nucleic acid.

[0021] In another example, the disclosure provides a method for identifying 5caC in a target nucleic acid comprising the steps of:

a. providing a nucleic acid sample comprising the target nucleic acid;
b. adding a blocking group to the 5fC in the nucleic acid sample;
c. converting the 5caC to DHU to provide a modified nucleic acid sample comprising a modified target nucleic acid;

d. optionally amplifying the copy number of the modified target nucleic acid; and
e. determining the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of a 5caC in the target nucleic acid.

[0022] In examples of the method for identifying 5caC in a target nucleic acid, the percentages of the T at each transition location provide a quantitative level for 5caC at each location in the target nucleic acid. In examples, the nucleic acid is DNA. In other examples, nucleic acid is RNA.

[0023] In examples of the disclosure, adding a blocking group to the 5fC in the nucleic acid sample comprises contacting the nucleic acid with an aldehyde reactive compound including, for example, hydroxylamine derivatives (such as O-ethylhydroxylamine), hydrazine derivatives, and hydrazide derivatives.

[0024] In another example, the disclosure provides a method for identifying 5fC in a target nucleic acid comprising the steps of:

a. providing a nucleic acid sample comprising the target nucleic acid;
b. adding a blocking group to the 5caC in the nucleic acid sample
c. converting the 5fC to DHU to provide a modified nucleic acid sample comprising a modified target nucleic acid;
d. optionally amplifying the copy number of the modified target nucleic acid;
e. detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of a 5fC in the target nucleic acid.

[0025] In examples of the method for identifying 5fC in a target nucleic acid, the percentages of the T at each transition location provide a quantitative level for 5fC at each location in the target nucleic acid. In examples, the nucleic acid (sample and target) is DNA. In other examples, nucleic acid (sample and target) is RNA.

[0025a] In one aspect, the present disclosure refers to a method for chemically modifying a nucleic acid sample comprising the steps of: providing a nucleic acid sample comprising 5-carboxylcytosine (5caC) and/or 5-formylcytosine (5fC); and converting the 5caC and/or 5fC to dihydrouracil (DHU) by treating the nucleic acid with a borane reducing agent to provide a modified nucleic acid sample comprising a modified nucleic acid.

[0026] In examples, the step of adding a blocking group to the 5caC in the nucleic acid sample comprises contacting the nucleic acid sample with a carboxylic acid derivatization reagent, including, for example, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and (ii) an amine (such as ethylamine), hydrazine, or hydroxylamine compound.

[0027] In examples of the disclosure, the methods above further comprise the step of amplifying the copy number of the modified target nucleic acid. In examples, this amplification step is performed prior to the step of detecting the sequence of the modified target nucleic acid. The step of amplifying the copy number when the modified target nucleic acid is DNA may be accomplished by performing the polymerase chain reaction (PCR), primer extension, and/or cloning. When the modified target nucleic acid is RNA, the step of amplifying the copy number may be accomplished by RT-PCR using oligo(dT) primer (for mRNA), random primers, and/or gene specific primers.

[0028] In examples of the disclosure, the DNA sample comprises picogram quantities of DNA. In examples of the disclosure, the DNA sample comprises about 1 pg to about 900 pg DNA, about 1 pg to about 500 pg DNA, about 1 pg to about 100 pg DNA, about 1 pg to about 50 pg DNA, about 1 to about 10 pg, DNA, less than about 200 pg, less than about 100 pg DNA, less than about 50 pg DNA, less than about 20 pg DNA, and less than about 5 pg DNA. In other examples of the disclosure, the DNA sample comprises nanogram quantities of DNA. In examples of the disclosure, the DNA sample contains about 1 to about 500 ng of DNA, about 1 to about 200 ng of DNA, about 1 to about 100 ng of DNA, about 1 to about 50 ng of DNA, about 1 ng to about 10 ng of DNA, about 1 ng to about 5 ng of DNA, less than about 100 ng of DNA, less than about 50 ng of DNA less than about 5 ng of DNA, or less that about 2 ng of DNA. In examples of the disclosure, the DNA sample comprises circulating cell-free DNA (cfDNA). In examples of the disclosure the DNA sample comprises microgram quantities of DNA.

[0029] In examples of the disclosure, the step of converting the 5caC and/or 5fC to DHU comprises contacting the nucleic acid sample with a reducing agent including, for example, pyridine borane, 2-picoline borane (pic-BH_3), borane, sodium borohydride, sodium cyanoborohydride, and sodium triacetoxyborohydride. In a preferred example, the reducing agent is pic-BH_3 and/or pyridine borane.

[0030] In examples of the disclosure, the step of determining the sequence of the modified target nucleic acid comprises chain termination sequencing, microarray, high-throughput sequencing, and restriction enzyme analysis.

BRIEF DESCRIPTION OF THE FIGURES

[0031] **Fig. 1. Borane-containing compounds screening.** Borane-containing compounds were screened for conversion of 5caC to DHU in an 11mer oligonucleotide ("oligo"), with conversion rate estimated by MALDI. 2-picoline borane (pic-borane), borane

Claims:

1. A method for identifying 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) in a target nucleic acid comprising the steps of:

providing a nucleic acid sample comprising the target nucleic acid;

modifying the target nucleic acid comprising the steps of:

converting the 5mC and 5hmC in the nucleic acid sample to 5-carboxylcytosine (5caC) and/or 5-formylcytosine (5fC) by contacting the nucleic acid sample with a ten-eleven translocation (TET) enzyme so that one or more 5caC or 5fC residues are generated; and

converting the 5caC and/or 5fC to dihydrouracil (DHU) by treating the target nucleic acid with a borane reducing agent to provide a modified nucleic acid sample comprising a modified target nucleic acid; and

detecting the sequence of the modified target nucleic acid;

wherein a cytosine (C) to thymine (T) transition or a cytosine (C) to DHU transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of either a 5mC or 5hmC in the target nucleic acid.

2. The method of claim 1, wherein the borane reducing agent is 2-picoline borane.

3. The method of claim 1, wherein the step of detecting the sequence of the modified target nucleic acid comprises one or more of chain termination sequencing, microarray, high-throughput sequencing, and restriction enzyme analysis.

4. The method of claim 1, wherein the TET enzyme is selected from the group consisting of human TET1, TET2, and TET3; murine Tet1, Tet2, and Tet3; *Naegleria* TET (NgTET); and *Coprinopsis cinerea* (CcTET).

5. The method of claim 1, further comprising a step of blocking one or more modified cytosines.

6. The method of claim 5, wherein the step of blocking comprises adding a sugar to a 5hmC.

7. The method of claim 1, wherein the method further comprises a step of amplifying the copy number of one or more nucleic acid sequences.

8. A method for chemically modifying a nucleic acid sample comprising the steps of:

providing a nucleic acid sample comprising 5-carboxylcytosine (5caC) and/or 5-formylcytosine (5fC); and

converting the 5caC and/or 5fC to dihydrouracil (DHU) by treating the nucleic acid with a borane reducing agent to provide a modified nucleic acid sample comprising a modified nucleic acid.

9. The method of claim 8, wherein the borane reducing agent is selected from the group consisting of 2-picoline borane (pic-BH3), borane, sodium borohydride, sodium cyanoborohydride, and sodium triacetoxyborohydride.

10. The method of claim 8, wherein the borane reducing agent is 2-picoline borane.

11. The method of claim 8, further comprising the step of detecting the sequence of the modified nucleic acid.

12. The method of claim 11, wherein the step of detecting the sequence of the modified nucleic acid comprises one or more of chain termination sequencing, microarray, high-throughput sequencing, and restriction enzyme analysis.

13. The method of claim 11, wherein the sequencing step provides a quantitative level of one or more cytosine modifications in the nucleic acid sample.

14. The method of claim 8, wherein the method further comprises a step of amplifying the copy number of one or more nucleic acid sequences.