

# Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine

Miao Yu<sup>1</sup>, Gary C Hon<sup>2</sup>, Keith E Szulwach<sup>3</sup>, Chun-Xiao Song<sup>1</sup>, Peng Jin<sup>3</sup>, Bing Ren<sup>2</sup> & Chuan He<sup>1</sup>

<sup>1</sup>Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois, USA. <sup>2</sup>Department of Cellular and Molecular Medicine, San Diego School of Medicine, Ludwig Institute for Cancer Research, University of California, La Jolla, California, USA. <sup>3</sup>Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA. Correspondence should be addressed to C.H. ([chuanhe@uchicago.edu](mailto:chuanhe@uchicago.edu)).

Published online 29 November 2012; doi:10.1038/nprot.2012.137

**A complete understanding of the potential function of 5-hydroxymethylcytosine (5-hmC), a DNA cytosine modification in mammalian cells, requires an accurate single-base resolution sequencing method. Here we describe a modified bisulfite-sequencing method, Tet-assisted bisulfite sequencing (TAB-seq), which can identify 5-hmC at single-base resolution, as well as determine its abundance at each modification site. This protocol involves  $\beta$ -glucosyltransferase ( $\beta$ -GT)-mediated protection of 5-hmC (glucosylation) and recombinant mouse Tet1(mTet1)-mediated oxidation of 5-methylcytosine (5-mC) to 5-carboxylcytosine (5-caC). After the subsequent bisulfite treatment and PCR amplification, both cytosine and 5-caC (derived from 5-mC) are converted to thymine (T), whereas 5-hmC reads as C. The treated genomic DNA is suitable for both whole-genome and locus-specific sequencing. The entire procedure (which does not include data analysis) can be completed in 14 d for whole-genome sequencing or 7 d for locus-specific sequencing.**

## INTRODUCTION

5-Hydroxymethylcytosine was first observed in mammals in 1972, but it was not given much attention until recently<sup>1</sup>. In 2009, 5-hmC was found to exist in relatively high abundance in Purkinje neurons and embryonic stem cells (ESCs), and was produced specifically through 5-mC oxidation catalyzed by the Tet family of proteins<sup>2,3</sup>. 5-hmC is thought to be an intermediate in an active demethylation process and may have direct roles in gene expression, as the modified base itself cannot be recognized by most 5-mC-binding proteins<sup>3–7</sup>. With the development and application of more sensitive detection technologies, 5-hmC has been found to be present at different levels in the genomes of various cell types or tissues<sup>8–11</sup>. Genome-wide profiling of 5-hmC further indicates potential regulatory roles of 5-hmC in ESC regulation, myelopoiesis, zygote development and neurodevelopment, thus suggesting that it may serve as an epigenetic mark<sup>12–20</sup>.

After the discovery of 5-hmC, several groups independently reported further oxidization of 5-hmC to 5-formylcytosine (5-fC) and 5-caC catalyzed by Tet proteins<sup>21–23</sup>. Both 5-caC and 5-fC can be recognized and excised by thymine DNA glycosylase (TDG) and then converted back to cytosine through the base excision repair pathway<sup>21,24,25</sup>. This newly discovered active demethylation pathway again suggests that 5-hmC is an intermediate of demethylation. 5-hmC accumulates to high abundance in certain brain tissues, implying functional roles other than as an intermediate in demethylation. Determination of the exact location and relative abundance of 5-hmC will be crucial in order to fully unveil the biology associated with this base modification. We describe here a detailed protocol for the TAB-seq method that we recently published for single-base resolution sequencing of 5-hmC<sup>26</sup>.

## Development of the protocol

Traditional bisulfite sequencing, which has been widely used to detect 5-mC at single-base resolution, cannot differentiate 5-mC from 5-hmC, as both resist deamination during the treatment of DNA with sodium bisulfite<sup>7,27</sup>. The protocol described here overcomes this limitation by selectively converting 5-mC

to 5-caC in two steps (Fig. 1): protection of 5-hmC through glucosylation and mTet1-mediated oxidation of 5-mC to 5-caC. After subsequent bisulfite conversion, the protected  $\beta$ -glucosyl-5-hydroxymethylcytosine (5-gmC; from 5-hmC) is sequenced as C, whereas 5-caC and C read as T, enabling single-base resolution sequencing of 5-hmC<sup>26</sup>.

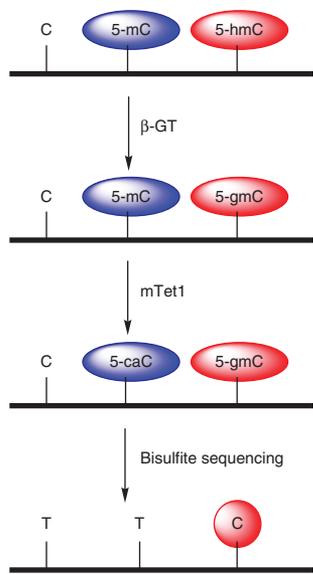
In the first step,  $\beta$ -GT, a T4 bacteriophage protein, is used to transfer a glucose to the hydroxyl group of 5-hmC and generate 5-gmC<sup>28,29</sup>. This  $\beta$ -GT-catalyzed glucosylation is highly selective and efficient with either natural or chemically modified uridine diphosphate (UDP)-glucose<sup>11,30</sup>. Several groups, including ours, have used this selective glucosylation reaction of 5-hmC for the enrichment of 5-hmC-containing genomic DNA fragments<sup>11,31–33</sup>.

5-Methylcytosine can be converted to 5-caC by Tet proteins, which is eventually read as T in bisulfite sequencing. 5-fC, which can be partially converted to T under standard bisulfite treatment, can also be oxidized by Tet proteins to 5-caC<sup>34</sup>. Thus, only protected 5-gmC will read as C in TAB-seq. Most reagents in the protocol are readily available. Active mTet1 is now commercially available (Wisegen) and expression as well as purification procedures for wild-type  $\beta$ -GT and the active domain of mTet1 can be followed as reported<sup>11,13,26</sup>. We also provide a detailed protocol for producing and purifying a recombinant mTet1 protein (Box 1).

## Applications of the method and limitations

TAB-seq is amenable to both whole-genome sequencing and locus-specific sequencing. This method has recently been used to produce genome-wide 5-hmC maps at base resolution in human and mouse ESCs<sup>26</sup>. Although we have not tested this method combined with reduced representation bisulfite sequencing (RRBS), we believe that the TAB method is compatible with RRBS<sup>35</sup>. In TAB-seq, the detection limit is governed by the conversion rate of 5-mC, protection efficiency of 5-hmC, abundance of 5-hmC at the modification site and sequencing depths<sup>26</sup>. With the protocol described here, highly efficient conversion of 5-mC to T (above 96%) in genomic DNA can be achieved, with at least 90% of the 5-hmC protected from conversion.

## PROTOCOL



**Figure 1** | Overview of Tet-assisted bisulfite sequencing (TAB-seq). 5-hmC is protected specifically by  $\beta$ -GT to generate 5-gmC, followed by oxidation of 5-mC to 5-caC by mTet1. Only 5-gmC is read as C after bisulfite treatment and PCR amplification.

Thus, sensitivity and specificity of 5-hmC detection by TAB-seq depends on sequencing depth. A cytosine base with less-abundant 5-hmC modification (i.e., <5%) will require more sequencing depth than a base with a higher level of 5-hmC. With RRBS or locus-specific sequencing, a better sensitivity of 5-hmC detection may be achieved owing to higher sequencing depth at selected bases.

### Comparison with other methods

**Affinity-based methods.** The affinity-based methods have been widely used to enrich the 5-hmC-containing fragments and profile

5-hmC distribution in the genome. There are two main strategies that were developed previously. One is antibody based, wherein antibodies against 5-hmC<sup>12,36–38</sup> or cytosine 5-methylenesulphonate, the product of 5-hmC after bisulfite treatment<sup>31</sup>, were used. The other is  $\beta$ -GT based and involves one of the following three approaches to achieve selective modification of 5-hmC for affinity purification: (i) an azide-modified glucose is transferred onto 5-hmC, followed by selective chemical labeling to attach a biotin tag<sup>11</sup>; (ii) a natural glucose is transferred, followed by periodate oxidation and biotinylation<sup>31</sup>; or (iii) a protein (JBP1) that specifically recognizes and binds to 5-gmC is used to enrich glucosylated 5-hmC<sup>32,33</sup>. However, affinity-based methods can neither detect 5-hmC at single-base resolution nor quantify its abundance at the modification site. In the antibody-based approach, recovery of hydroxymethylated fragments can be affected by the density of 5-hmC, especially in 5-hmC immunoprecipitation (hMeDIP), which uses antibodies that recognize 5-hmC<sup>31</sup>. The regions with high 5-hmC density may be overrepresented, whereas the regions with low 5-hmC density may be underrepresented.

**Single-molecule real-time sequencing.** Single-molecule real-time sequencing can identify modified bases on the basis of the different polymerase passing rates at and around the base. Although this technology is capable of detecting 5-mC and 5-hmC modifications directly, its application is limited by low sensitivity and low throughput<sup>39</sup>. In early 2012, we modified our previous 5-hmC labeling method and combined it with single-molecule real-time (SMRT) DNA sequencing<sup>40</sup>. With larger kinetic signature, increased 5-hmC abundance and reduced amount of DNA to sequence, SMRT sequencing can be applied to detect 5-hmC in genomic DNA at single-base resolution. However, the quantitative information about 5-hmC at each modification site is lost during enrichment. The throughput of the method needs to be improved for the sequencing of large genomes.

## Box 1 | Expression and purification of recombinant mTet1 ● TIMING 9–10 d

1. Generate recombinant bacmid and produce recombinant baculovirus P1, P2 and P3 with pFastBac dual-mTet1 using the Bac-to-Bac baculovirus expression system (Invitrogen) according to the manufacturer's instructions.
2. Infect High Five cells by adding 1 ml of P3 baculovirus stock to each culture dish (150 mm  $\times$  2.5 mm) with 20 ml of fresh medium for Tet1 expression and incubating at 27 °C for 48 h.
3. Collect cells from 20 dishes by centrifuging at 3,800g for 10 min at 4 °C. Wash the cells with mTet1 cell lysis buffer and resuspend the cells in 30 ml of mTet1 cell lysis buffer.
- ▲ **CRITICAL STEP** During the purification, all the steps should be performed on ice or at 4 °C.
- **PAUSE POINT** Cells can be frozen in liquid nitrogen and stored at –80 °C for at least 3 months.
4. Break the cells by flowing them through the homogenizer twice.
5. Centrifuge the cell lysate at 30,000g for 40 min at 4 °C.
6. Add 3 ml of Anti-flag M2 affinity gel to the Econo-Column and wash it with at least 60 ml of mTet1 cell lysis buffer using gravity.
7. Load the clarified cell lysate to the equilibrated Anti-flag M2 affinity gel and then resuspend. Incubate at 4 °C for 10 min.
8. Let the cell lysate flow through using gravity at the rate of 1 drop per s.
9. Load the flow-through back to the gel and repeat steps 7 and 8 once.
10. Wash the gel with at least 200 ml of mTet1 cell lysis buffer using gravity.
11. Add 1 mM TCEP, 1 mM PMSF, 1  $\mu$ g ml<sup>-1</sup> leupeptin and 1  $\mu$ g ml<sup>-1</sup> pepstatin to the 3 $\times$  Flag peptide (0.2 mg ml<sup>-1</sup>).
12. Resuspend the gel with 3 ml of 3 $\times$  Flag peptide and elute the protein with a flow rate of 1 drop per s after incubating for 10 min.
13. Repeat step 12 twice more.
14. Concentrate the eluate to 2 ml and then apply it to the 120 ml Superdex-200 column equilibrated with mTet1 gel filtration running buffer.
15. Collect the peak fraction and concentrate mTet1 with an Amicon Ultra-4 filter (MWCO 30,000 Da) to around 3 mg ml<sup>-1</sup>.
16. Add glycerol to purified mTet1 to 30% (vol/vol) and aliquot. Freeze the protein in liquid nitrogen and store it at –80 °C.
- ▲ **CRITICAL STEP** Storage of mTet1 above –80 °C may lead to decreased 5-mC oxidation efficiency.

## Box 2 | Assaying the activity of recombinant mTet1 ● TIMING 4–5 d

1. Prepare the mTet1 oxidation reactions as follows. Add each component in the order listed.

Component	Amount	Final concentration
Genomic DNA of interest	500 ng	10 ng $\mu\text{l}^{-1}$
Methylated $\lambda$ -DNA (from Step 3 of the main PROCEDURE)	2.5 ng	
Tet oxidation reagent 1	3.5 $\mu\text{l}$	
Tet oxidation reagent 2	15 $\mu\text{l}$	
mTet1 protein (3 mg $\text{ml}^{-1}$ )	5 $\mu\text{l}$	0.3 $\mu\text{g } \mu\text{l}^{-1}$
Milli-Q water	to 50 $\mu\text{l}$	
Final volume	50 $\mu\text{l}$	

2. Mix well and allow the reaction to proceed at 37 °C for 80 min.

3. Add 1  $\mu\text{l}$  of proteinase K (20 mg  $\text{ml}^{-1}$ ) to the reaction mixture and incubate at 50 °C for 1 h.

4. Clean up the DNA with Micro Bio-Spin 30 columns and then apply it to the Qiagen QIAquick nucleotide removal kit—follow the manufacturer’s instructions for both. Elute the DNA in 30  $\mu\text{l}$  of EB buffer (included in the Qiagen kit).

■ **PAUSE POINT** The oxidized DNA can be stored at –20 °C for several weeks.

5. Apply the treated genomic DNA from step 4 to the Qiagen Epitect bisulfite kit according to the manufacturer’s instructions. Other bisulfite conversion kits can also be used.

6. Use the table below as a guide to prepare 50- $\mu\text{l}$  PCRs.

Component	Amount ( $\mu\text{l}$ )	Final concentration
PfuTurbo $C_x$ reaction buffer (10 $\times$ )	5	1 $\times$
dNTP mix (25 mM each)	0.4	200 $\mu\text{M}$ each
5-mC_test_F (10 $\mu\text{M}$ )	1	0.2 $\mu\text{M}$
5-mC_test_R (10 $\mu\text{M}$ )	1	0.2 $\mu\text{M}$
Bisulfite-treated DNA (from step 5)	2	
PfuTurbo $C_x$ DNA polymerase (2.5 U $\mu\text{l}^{-1}$ )	1	2.5 U per 50 $\mu\text{l}$ PCR
Milli-Q water	to 50	
Final volume	50	

7. Set up and run the PCR program as follows:

Cycle number	Denature	Anneal	Extend
1	95 °C, 2 min		
2–41	95 °C, 30 s	57 °C, 30 s	72 °C, 1 min
42			72 °C, 10 min

8. Verify PCR amplification by running 10  $\mu\text{l}$  of the PCR on a 1.5% (wt/vol) agarose gel for 30 min at 100 V. There should be a single, clear band of ~300 bp.

9. TOPO-clone 2  $\mu\text{l}$  of the reaction using Zero Blunt TOPO PCR cloning kit according to the manufacturer’s instructions; isolate at least 20 clones with the Qiagen spin miniprep kit for Sanger sequencing.

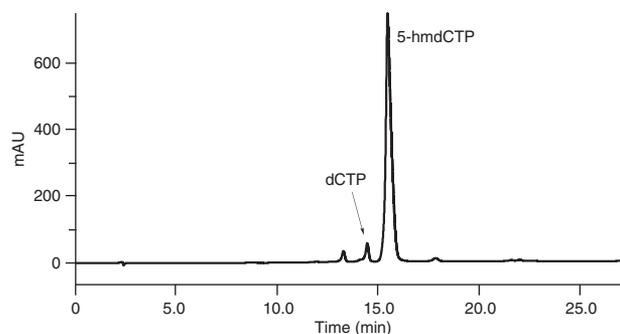
10. Calculate the 5-mC conversion rate. The ideal conversion rate of 5-mC at CpG positions is >96%. The conversion rate of 5-mC = number of ‘T’ reads at CpG cytosines/(number of non-CpG cytosines  $\times$  number of clones sequenced).

**Oxidative bisulfite sequencing.** Oxidative bisulfite sequencing (oxBS-seq), which can discriminate 5-mC from 5-hmC, was recently reported<sup>34</sup>. In this modified bisulfite-sequencing method,  $\text{KRuO}_4$  selectively oxidizes 5-hmC to 5-fC at high efficiency, followed by conversion to T in subsequent bisulfite treatment and PCR amplification. A comparison of the results of oxBS-seq with those of standard bisulfite sequencing allows for the quantitative sequencing of both 5-mC and 5-hmC at single-base resolution. Application of oxBS-seq requires multiple rounds of bisulfite treatment to fully deaminate 5-fC, and chemical oxidation may cause extensive oxidative DNA damage. Compared with oxBS-seq, TAB-seq gives direct reads of 5-hmC, and the treatment procedure incurs less DNA damage<sup>34</sup>. However, TAB-seq requires highly active Tet protein for efficient conversion of 5-mC to 5-caC.

### Experimental design

**Purification of the active domain of mTet1.** As incomplete oxidation of 5-mC will result in false-positive 5-hmC signals, the

availability of highly active mTet1 protein is crucial to TAB-seq. The expression and purification procedures of recombinant mTet1 are described in **Box 1**. To ensure high activity of the recombinant mTet1, all steps must be performed at 4 °C or on ice during



**Figure 2** | HPLC analysis of commercial 5-hmdCTP. The commercial 5-hmdCTP contains 4–5% of dCTP. mAU, milliabsorbance units.

### Box 3 | Purification of commercial 5-hmdCTP with HPLC ● TIMING 2 d

1. After pump washing, equilibrate the C18 reversed-phase LC column (4.6 × 150 mm) with 100% HPLC buffer B (100% CH<sub>3</sub>CN) with a flow rate of 1 ml min<sup>-1</sup> for 30 min and then with 100% HPLC buffer A (0.1 M TEAA) for 30 min.
2. Inject 4 pmol of commercial 5-hmdCTP and run the program. (Start from 0% buffer B and increase to 6% buffer B in 30 min.)
- ▲ **CRITICAL STEP** If semipreparative or preparative C18 column is used, a larger amount of 5-hmdCTP can be applied.
3. Collect the peak of 5-hmdCTP as indicated in **Figure 2**.
4. Repeat Step 3 twice more, combine the 5-hmdCTP fraction and dry it with a lyophilizer.
5. Dissolve the purified 5-hmdCTP in 10 μl of Milli-Q water and measure the concentration of 5-hmdCTP ( $\epsilon$  at 275 nm = 7.7 E × mmol<sup>-1</sup> × cm<sup>-1</sup>).

purification. Aliquots of mTet1 proteins are stored at -80 °C before use. Improper storage or multiple freeze-thaw cycles (more than twice) may result in decreased oxidation activity. The protocol to test the activity of purified mTet1 is described in **Box 2**; it is strongly recommended to carry out the activity test on each batch of newly purified recombinant mTet1 before applying it to TAB-seq.

**Spike-in control.** For both locus-specific and whole-genome sequencing, two key parameters exist for an accurate estimation of 5-hmC abundance besides the conversion rate of unmodified cytosine to uracil: the oxidation efficiency of 5-mC to 5-caC and the protection efficiency of 5-hmC. Although nonprotected 5-hmC can result in an underestimation of 5-hmC abundance, nonconversion of unmodified C and 5-mC will result in false-positive 5-hmC signals, and should therefore be determined in each experiment. With sufficiently high C and mC conversion rates and 5-hmC protection rates, the abundance of 5-hmC can be quantified from the frequency with which C is read compared with T at a given genomic position in any sequencing experiment. To assess the conversion rates in genomic DNA, controls containing 5-mC and 5-hmC need to be spiked in before treatment. Such controls should be of sufficient complexity and contain modified cytosines in various sequence contexts (i.e., multiple CpGs throughout). For genome-wide sequencing, spike-in DNA should span at least 1 kb of the sequence, such that subsequent random fragmentation by sonication and sequencing can distinguish PCR duplicates. Furthermore, after bisulfite conversion, spike-in DNA should not align to a bisulfite-converted target genome. In practice, we find that DNA from the lambda phage and the pUC19 plasmid work well as spike-ins for mouse and human samples.

For the 5-mC control, DNA can either be selectively methylated at CpG sites using CpG methyltransferase or amplified with 5-mdCTP. However, the CpG-methylated control is recommended, as the frequent neighboring 5-mC generated by PCR may lead to the underestimation of the oxidation efficiency.

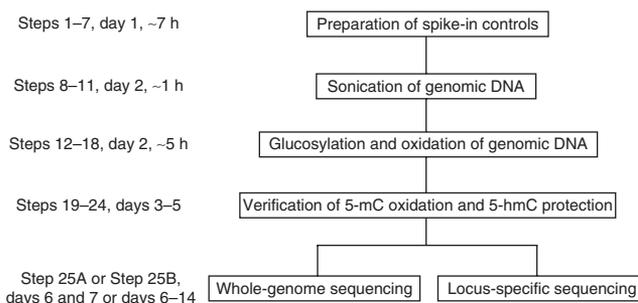
For the 5-hmC control, there is no enzyme that can selectively generate 100% 5-hmC from C or 5-mC; therefore, besides synthesizing long oligonucleotides containing 5-hmC at required positions, the easiest and most cost-effective way to generate DNA longer than 1 kb with multiple 5-hmC sites may be through PCR amplification with 5-hydroxymethyl dCTP (5-hmdCTP). With this method, each C position is supposed to be 100% 5-hmC. However, we have found that many commercial 5-hmdCTPs contain contaminant dCTP, which will result in the underestimation of protection efficiency because unmodified cytosine will display as 'T' in TAB-seq (**Fig. 2**). Purifying the commercial 5-hmdCTP with HPLC offers one solution to this problem and takes about 2 d (**Box 3**). However, it should

be noted that it may be difficult to amplify fragments larger than 2 kb with purified 5-hmdCTP. Alternatively, if experiments are run alongside conventional bisulfite treatments, dCTP contamination can be adjusted for by direct measurement of bisulfite-converted cytosine in the 5-hmC spike-in control. Typically, >90% protection efficiency can be achieved for 5-hmC protection after taking contamination into account.

In this protocol, we use methylated λ-DNA as C and 5-mC control in both genome-wide and locus-specific sequencing. For the 5-hmC control, a PCR product of 1.64 kb from a pUC19 vector (5-hmC control 1; generated with 5-hmC\_F\_1 and 5-hmC\_R\_1 primers) is used for the estimation of 5-hmC protection efficiency only in genome-wide sequencing. The 290-bp control (5-hmC control 2; generated with 5-hmC\_F\_2 and 5-hmC\_R\_2 primers), which is relatively easier to amplify and clone after bisulfite treatment, is used for the verification of 5-hmC protection efficiency after β-GT and mTet1 treatment with TOPO cloning. In both 5-hmC spike-in controls, all Cs are 5-hmC except for those in the primer sequence.

Verification of 5-mC conversion and 5-hmC protection on spike-in controls with TOPO cloning is strongly recommended after β-GT and mTet1 treatment but the procedure before proceeding with large-scale sequencing. However, it may be simplified if no quantitative conversion or protection rate is required for locus-specific sequencing.

**Data analysis.** Data analysis for TAB-seq and traditional bisulfite sequencing are nearly identical. At each genomic locus, the estimated abundance of 5-hmC ( $A_{5\text{-hmC}}$ ) is measured as the number of cytosine base calls divided by the total (C + T) sequencing depth at the locus. For genome-wide analysis, only good-quality base calls (Phred score of 20 or greater) are considered. To correct for the 5-hmC protection rate ( $r_{5\text{-hmC}}$ ) not being 100% efficient, the absolute abundance of 5-hmC is calculated as  $E_{5\text{hmC}} = A_{5\text{hmC}} / r_{5\text{hmC}}$ .



**Figure 3** | Workflow for TAB-seq.

**TABLE 1** | Details of the oligonucleotides used in the protocol.

Name	Sequence (5'→3')	Amplicon size	Use
5-hmC_F_1	GCAGATTGACTGAGAGTGC	1.64 kb	Generation of 5-hmC spike-in control 1
5-hmC_R_1	TGCTGATAAATCTGGAGCCG		
5-hmC_F_2	TCACTCATGGTTATGGCAGC	290 bp	Generation of 5-hmC spike-in control 2
5-hmC_R_2	GATGCTGAAGATCAGTTGGG		
5-mC_test_F	TTGGGTTATGTAAGTTGATTTATG	296 bp	Assay the activity of recombinant mTet1 and test the conversion rate of 5-mC to 5-caC
5-mC_test_R	CACCTACTTACTAAAATTACACC		
5-hmC_test_F	TTATTATGGTTATGGTAGTA	290 bp	Test the protection efficiency of 5-hmC
5-hmC_test_R	AATACTAAAATCAATTAATGC		

The resolution of TAB-seq to detect 5-hmC is crucially dependent on sequencing depth. For example, the median abundance of 5-hmC at 5-hmC sites in H1 ESCs is just under 20%, and detecting base-resolution 5-hmC at this level requires sequencing at a depth of ~25 times per cytosine, or ~50 times the haploid genome size. For cells with higher levels of 5-hmC, less sequencing is required; we estimate that a depth of ~15 times per cytosine is sufficient to detect base-resolution 5-hmC with an abundance of 30%.

If base-resolution precision is not necessary, much less sequencing is required. We recommend sequencing reads with a length of at least 100 bp. Please see genome-wide bisulfite sequencing methods for additional details<sup>26,41</sup>.

The entire workflow for the method is shown in **Figure 3**. The details of all oligonucleotides used in this protocol are listed in **Table 1**. **Supplementary Note 1** shows the sequence of the 5-hmC spike-in controls.

## MATERIALS

### REAGENTS

- Genomic DNA sample (3 µg of genomic DNA purified with a genomic DNA purification kit (Wizard, cat. no. A1120 or similar))
- Milli-Q water
- Recombinant mTet1 (3 mg ml<sup>-1</sup>; **Box 1** for expression and purification procedures)
- pFastBac dual-mTet1 (available on request from the corresponding author; sequence in **Supplementary Note 2**)
- Bac-to-Bac baculovirus expression system (Invitrogen, cat. no. 10359-016)
- High Five cells (Invitrogen, cat. no. B855-02)
- Grace's insect medium, supplemented (Invitrogen, cat. no. 11605102)
- Medium for mTet1 expression (Reagent Setup)
- Penicillin-streptomycin, liquid (Invitrogen, cat. no. 15140-163)
- FBS (Gemini Bio, cat. no. GBP-900198)
- Tris hydrochloride (Tris-HCl; 20 mM, pH 8.0; Fisher BioReagents, cat. no. BP1531)
- Tris (2-carboxyethyl) phosphine hydrochloride (TCEP; Sigma-Aldrich, cat. no. C4706)
- Phenylmethylsulfonyl fluoride (PMSF; Chem-impex international, cat. no. 00632)
- Leupeptin hemisulfate salt (Sigma-Aldrich, cat. no. L2884)
- Pepstatin A (Fisher BioReagents, cat. no. BP2671-100)
- Anti-flag M2 affinity gel (Sigma-Aldrich, cat. no. A2220)
- 3× Flag peptide (Sigma-Aldrich, cat. no. F4799; dissolved in 150 mM NaCl at 0.2 mg ml<sup>-1</sup>)
- mTet1 cell lysis buffer (Reagent Setup)
- mTet1 gel filtration running buffer (Reagent Setup)
- Unmethylated λ-DNA (Promega, cat. no. D1521)
- SssI CpG methyltransferase (NEB, cat. no. M0226M; kit contains NEBuffer 2 and SAM)
- QIAEX II gel extraction kit (Qiagen, cat. no. 20021)
- ZymoTaq DNA polymerase (Zymo Research, cat. no. E2001)
- Hydroxymethyl dCTP (hmdCTP; Bionline, cat. no. BIO-39046)

- Ultrapure dATP (USB, cat. no. 77102)
- Ultrapure dGTP (USB, cat. no. 77106)
- Ultrapure dTTP (USB, cat. no. 77108)
- 5-hmC dNTP mix, 10 mM (Reagent Setup)
- 5-hmC\_F\_1, 5-hmC\_R\_1, 5-hmC\_F\_2, and 5-hmC\_R\_2 primers (synthesized by Operon; **Table 1**)
- pUC 19 vector (NEB, cat. no. N3041S; dilute in H<sub>2</sub>O to 1 ng µl<sup>-1</sup>)
- Agarose (Denville, cat. no. CA3510-8)
- Ethidium bromide (Acros, cat. no. 170960050; dissolved in H<sub>2</sub>O at 10 mg ml<sup>-1</sup>) **! CAUTION** Ethidium bromide is toxic and suspected of causing genetic defects if inhaled; use personal protective equipment when handling.
- QIAquick PCR purification kit (Qiagen, cat. no. 28104)
- MinElute gel extraction kit (Qiagen, cat. no. 28004)
- HEPES (500 mM and 1 M, pH 8.0; GenScript, cat. no. C01621)
- MgCl<sub>2</sub>·6H<sub>2</sub>O (Fluka, cat. no. 63064; dissolved in H<sub>2</sub>O at 250 mM)
- β-GT protection buffer 10× (Reagent Setup)
- T4 Phage β-GT (T4-βGT; NEB, cat. no. M0357S)
- UDP-glucose (Wako, cat. no. EPL1144; dissolved in H<sub>2</sub>O at 10 mM)
- **▲ CRITICAL** UDP-glucose should be stored at -20 °C.
- Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O (Alfa Aesar, cat. no. 13448; dissolved in H<sub>2</sub>O at 1.5 mM) **! CAUTION** It causes gastrointestinal injury and skin, eye and respiratory irritation. Use personal protective equipment when handling.
- NaCl (Fisher BioReagents, cat. no. BP358-10; dissolved in H<sub>2</sub>O at 2 M)
- L-Ascorbic acid (Sigma-Aldrich, cat. no. 255564; dissolved in H<sub>2</sub>O at 40 mM)
- DTT (Roche Diagnostics, cat. no. 93889320; dissolved in H<sub>2</sub>O at 50 mM)
- Disodium salt ATP (Fisher BioReagents, cat. no. BP413-25; dissolved in H<sub>2</sub>O at 24 mM)
- α-Ketoglutaric acid disodium salt hydrate (α-KG; Sigma-Aldrich, cat. no. K3752; dissolved in H<sub>2</sub>O at 20 mM)
- Tet oxidation reagent 1 (Reagent Setup)
- Tet oxidation reagent 2 (Reagent Setup)
- Micro Bio-Spin 30 columns (Bio-Rad, cat. no. 7326203)
- Proteinase K (Fermentas, cat. no. EO0491)



## PROTOCOL

- QIAquick gel extraction kit (Qiagen, cat. no. 28704)
- QIAquick nucleotide removal kit (Qiagen, cat. no. 28304)
- 5-mC\_test\_F, 5-mC\_test\_R, 5-hmC\_test\_F, 5-hmC\_test\_R (synthesized by Operon; **Table 1**)
- QIAprep spin miniprep kit (Qiagen, cat. no. 27104)
- Zero Blunt TOPO PCR cloning kit (Invitrogen, cat. no. K2800-20)
- End-It DNA end-repair kit (Epicentre, cat. no. ER81050).
- Klenow fragment (3'→5' exo-; NEB, cat. no. M0212L)
- Quick ligation kit (NEB, cat. no. M2200L)
- TruSeq DNA sample preparation kit (Illumina, cat. no. FC-121-2001)
- MethylCode bisulfite conversion kit (Invitrogen, cat. no. MECOV-50)
- PfuTurbo C<sub>x</sub> hotstart DNA polymerase (Agilent, cat. no. 600410)
- Acetonitrile (CH<sub>3</sub>CN; Fisher Scientific, cat. no. A9984)
- Triethylammonium acetate, 1 M solution (TEAA; Calbiochem, cat. no. 625718)
- Liquid nitrogen
- S-Adenosylmethionine

### EQUIPMENT

- Napco 8000 water-jacketed CO<sub>2</sub> incubator (Fisher Scientific, cat. no. 15-497-001)
- Optima L-90K ultracentrifuge (Beckman Coulter, part no. 365670)
- Sorvall RC 3BP Plus centrifuge (Sorvall, cat. no. 75007565)
- Centrifuge tubes (Beckman Coulter, cat. no. 355655)
- Centrifuge tubes (50 ml; Genesee Scientific, cat. no. 21-108)
- Homogenizer (Avestin, model no. EmulsiFlex-C3)
- Tissue culture dish (Fisher Scientific, cat. no. 08-772-6)
- Econo-Column chromatography column (Bio-Rad, cat. no. 737-4252)
- AKTA FPLC chromatography system (GE Healthcare, cat. no. 18-1900-26)
- HiLoad 16/60 Superdex 200 pg (GE Healthcare, cat. no. 17-1069-01)
- Amicon Ultra-4 centrifugal filter (MWCO 30,000 Da; Millipore, cat. no. UFC803024)
- Water bath (VWR Scientific, model no. 1202)
- Centrifuge tubes (1.5 ml; ITI, cat. no. B71063)
- Manual pipette (Mettler Toledo, cat. no. SL-10XLS (0.5–10 µl); cat. no. SL-100XLS (10–100 µl); cat. no. SL-1000XLS (100–1,000 µl))
- Pipette tips (Denville, cat. no. p2200 (10 µl); cat. no. p2201 (200 µl); VWR, cat. no. 89079-470 (1250 µl))
- Tabletop centrifuge (Denville, model no. 2600)

- Thermal cycler (Bio-Rad; model no. C1000)
- PCR tubes (8 strip; DOT, cat. no. 203-8PCR)
- PCR tube caps (8 strip; DOT, cat. no. 204-8PCR)
- Covaris system (Covaris, cat. no. S2/E210)
- Thermomixer R (Eppendorf, cat. no. 022670107)
- Dark Reader transilluminator (Clare Chemical Research, cat. no. DR88X)
- Illumina HiSeq 2000 (Illumina, cat. no. SY-401-1001)
- Acclaim 120 C18 reversed-phase LC column (DIONEX, product no. 059148)
- HPLC system (Varian, model no. 210)

### REAGENT SETUP

**Medium for mTet1 expression** Add 50 ml of FBS and 5 ml of penicillin-streptomycin to 500 ml of supplemented Grace's insect medium. The medium should be freshly prepared before use.

**mTet1 cell lysis buffer** Mix 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM TCEP, 1 mM PMSF, 1 µg ml<sup>-1</sup> leupeptin and 1 µg ml<sup>-1</sup> pepstatin. TCEP, PMSF, leupeptin and pepstatin should be added immediately before use. The cell lysis buffer without TCEP, PMSF, leupeptin or pepstatin can be stored for 6 months at 4 °C.

**mTet1 gel filtration running buffer** Mix 20 mM HEPES (pH 8.0), 150 mM NaCl and 1 mM DTT. DTT should be added immediately before use. The GF running buffer without DTT can be stored for 6 months at 4 °C.

**5-hmC dNTP mix, 10 mM** Mix 10 mM ultrapure dATP, dTTP, dGTP and 5-hmC dCTP. It can be stored for 3 months at -20 °C.

**βGT protection buffer, 10×** Mix 500 mM HEPES (pH 8.0) and 250 mM MgCl<sub>2</sub>. This buffer can be stored for 1 year at room temperature (20–25 °C).

**Tet oxidation reagent 1** Add 14.7 mg of Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O to 1 ml of H<sub>2</sub>O and then make a 24-fold dilution to a concentration of 1.5 mM.

▲ **CRITICAL** Tet oxidation reagent 1 can be stored at -80 °C for 2 months.

Storage of Tet oxidation reagent 1 at room temperature can lead to reduced 5-mC oxidation efficiency.

**Tet oxidation reagent 2** Mix 333 mM NaCl, 167 mM HEPES (pH 8.0), 4 mM ATP, 8.3 mM DTT, 3.3 mM α-KG and 6.7 mM L-ascorbic acid. ▲ **CRITICAL** Tet oxidation reagent 2 can be stored at -80 °C for 2 months. Storage of Tet oxidation reagent 2 at room temperature will lead to reduced 5-mC oxidation efficiency.

## PROCEDURE

### Creating methylated λ-DNA (5-mC and C spike-in control) ● **TIMING 3 h**

1| Set up an *in vitro* methylation reaction as follows. Add each component in the order listed.

Component	Amount (µl)	Final concentration
Milli-Q water	44	
NEbuffer 2 (10×	10	1×
S-adenosylmethionine (32 mM)	2	0.64 mM
Unmethylated λ-DNA (450 µg ml <sup>-1</sup> )	40	180 ng µl <sup>-1</sup>
SssI methylase (20 U µl <sup>-1</sup> )	4	0.8 U µl <sup>-1</sup>
Final volume	100	

2| Mix well and allow the reaction to proceed at 37 °C for 2 h, followed by heating at 65 °C for 20 min to stop the reaction.

▲ **CRITICAL STEP** Use fresh S-adenosylmethionine.

3| Clean up the DNA with the QIAEX II gel extraction kit and elute methylated DNA in 20 µl of Milli-Q water by following the manufacturer's instructions.

**Creating 5-hmC spike-in control** ● **TIMING 4 h**

4| Use the table below as a guide to prepare two 50- $\mu$ l PCRs: one for 5-hmC control 1 (tube 1) and one for 5-hmC control 2 (tube 2).

Component	Amount ( $\mu$ l)		Final concentration
	Tube 1	Tube 2	
Reaction buffer (2 $\times$ )	25	25	1 $\times$
5-hmC dNTP mix (10 mM each)	1	1	200 $\mu$ M each
5-hmC_F_1 (10 $\mu$ M)	2	—	0.4 $\mu$ M
5-hmC_R_1 (10 $\mu$ M)	2	—	0.4 $\mu$ M
5-hmC_F_2 (10 $\mu$ M)	—	2	0.4 $\mu$ M
5-hmC_R_2 (10 $\mu$ M)	—	2	0.4 $\mu$ M
pUC 19 vector (1 ng $\mu$ l <sup>-1</sup> )	1		1 ng per 50 $\mu$ l
ZymoTaq DNA Polymerase (5 U $\mu$ l <sup>-1</sup> )	0.4	0.4	2 U per 50 $\mu$ l PCR
Milli-Q water	to 50	to 50	
Final volume	50	50	

5| Set up and run the PCR program as follows:

Cycle number	Denature	Anneal	Extend
1	95 °C, 10 min		
2–41	95 °C, 30 s	57 °C, 30 s	72 °C, 1 min for tube 2; 72 °C, 1.5 min for tube 1
42			72 °C, 7 min

6| Verify PCR amplification by running 10  $\mu$ l of the PCR on 1.5% (wt/vol) agarose gel for 30 min at 100 V. There should be a single, clear band of ~1.6 kb (5-hmC control 1; tube 1) or ~300 bp (5-hmC control 2; tube 2).

7| Run the remaining 40  $\mu$ l of each PCR on a 1.5% (wt/vol) agarose gel for 30 min at 100 V and perform a gel extraction with the Qiagen QIAquick gel extraction kit according to the manufacturer's instructions. Elute the DNA in 40  $\mu$ l of Milli-Q water.

**? TROUBLESHOOTING**

**Sonication of genomic DNA** ● **TIMING 1 h**

8| For genome-wide sequencing, add 5 ng (0.5%) of 5-mC/C spike-in control (from Step 3) and 2.5 ng (0.25%) of 5-hmC control 1 (from tube 1 at Step 7, amplified with 5-hmC\_F\_1 and 5-hmC\_R\_1 primers) to each 1  $\mu$ g of genomic DNA. For locus-specific sequencing, add 5 ng (0.5%) of 5-mC/C spike-in control (from Step 3).

9| Sonicate the genomic DNA in 120  $\mu$ l of EB buffer to the desired size range. We typically sonicate for 55 s using a Covaris system, with a 10% duty cycle, intensity set at 4 and cycle per burst set at 200 in order to achieve the size range of 200–500 bp.

▲ **CRITICAL STEP** Because of platform and machine variability, the precise sonication parameters should be tested before sample preparation. Increasing or decreasing the length of sonication will decrease and increase, respectively, the size distribution of fragments. The desired size range depends on the downstream sequencing parameters.

▲ **CRITICAL STEP** Always analyze DNA fragmentation by running an aliquot on an agarose gel before proceeding to Step 10.



## PROTOCOL

**10|** Add 2.5 ng (0.25%) of 5-hmC control 2 (~300 bp from tube 2 at Step 7, amplified with 5-hmC\_F\_2 and 5-hmC\_R\_2 primers) after sonication to each 1 µg of genomic DNA for both genome-wide sequencing and locus-specific sequencing.

**11|** Purify the sonicated genomic DNA with the Qiagen MinElute PCR purification kit according to the manufacturer's instructions; elute the DNA in 10 µl of Milli-Q water.

▲ **CRITICAL STEP** Fragments with a size of <70 bp will be eliminated after purification.

■ **PAUSE POINT** The sonicated genomic DNA can be stored at -20 °C for several months.

### β-GT glucosylation ● TIMING 2 h

**12|** Prepare the β-GT reactions as follows. (1 µg is sufficient for an 8–9 loci test; 3 µg is recommended for whole-genome sequencing.)

Component	Amount	Final concentration
Sheared DNA from Step 11	1 or 3 µg	50 or 150 ng µl <sup>-1</sup>
UDP-glucose (10 mM)	1 µl	200 µM
β-GT protection buffer (10×)	2 µl	1×
T4-βGT (40 µM)	0.5 µl	1 µM
Milli-Q water	to 20 µl	
Final volume	20 µl	

**13|** Mix well and allow the reaction to proceed at 37 °C for 1 h.

▲ **CRITICAL STEP** If commercial T4-βGT is used, follow the protocol from NEB.

**14|** Clean up the DNA using a Qiagen QIAquick nucleotide removal kit according to the manufacturer's instructions and elute the DNA in 30 µl of Milli-Q water.

■ **PAUSE POINT** The glucosylated DNA can be stored at -20 °C for several weeks.

### mTet1 oxidation ● TIMING 3 h

**15|** Prepare the mTet1 oxidation reactions as follows. Add each component in the order listed.

Component	Amount	Final concentration
Glucosylated DNA from Step 14	500 ng	10 ng µl <sup>-1</sup>
Tet oxidation reagent 1	3.5 µl	
Tet oxidation reagent 2	15 µl	
mTet1 protein (3 mg ml <sup>-1</sup> )	5 µl	0.3 µg µl <sup>-1</sup>
Milli-Q water	to 50 µl	
Final volume	50 µl	

▲ **CRITICAL STEP** For genome-wide sequencing, perform at least three reactions in parallel (1.5 µg of glucosylated DNA input in total) to obtain enough DNA for library generation. If more than 2% of the whole genome is 5-mC in the genomic DNA of interest, increase the concentration of mTet1 accordingly.

**16|** Mix well and allow the reaction to proceed at 37 °C for 80 min.

**17|** Add 1 µl of proteinase K (20 mg ml<sup>-1</sup>) to the reaction mixture and incubate at 50 °C for 1 h.

18| Clean up the DNA with Micro Bio-Spin 30 columns and then use a Qiagen QIAquick nucleotide removal kit. Follow the manufacturer's instructions in each case. Elute the DNA in 30  $\mu\text{l}$  of EB buffer.

■ **PAUSE POINT** The oxidized DNA can be stored at  $-20\text{ }^{\circ}\text{C}$  for several weeks.

**Verification of 5-mC conversion and 5-hmC protection** ● **TIMING 1–3 d**

19| Apply 50 ng of treated genomic DNA (from Step 18) to the Qiagen Epitect bisulfite kit according to the manufacturer's instructions. Other bisulfite conversion kits can also be used.

20| Use the table below as a guide to prepare two 50- $\mu\text{l}$  PCRs: one for the C and 5-mC conversion test and one for the 5-hmC protection test.

Component	Amount ( $\mu\text{l}$ )		Final concentration
	C and 5-mC test	5-hmC test	
PfuTurbo C <sub>x</sub> reaction buffer (10 $\times$ )	5	5	1 $\times$
dNTP mix (25 mM each)	0.4	0.4	200 $\mu\text{M}$ each
5-mC_test_F (10 $\mu\text{M}$ )	1	—	0.2 $\mu\text{M}$
5-mC_test_R (10 $\mu\text{M}$ )	1	—	0.2 $\mu\text{M}$
5-hmC_test_F (10 $\mu\text{M}$ )	—	1	0.2 $\mu\text{M}$
5-hmC_test_R (10 $\mu\text{M}$ )	—	1	0.2 $\mu\text{M}$
Bisulfite-treated DNA (from Step 19)	2	2	
PfuTurbo C <sub>x</sub> DNA Polymerase (2.5 U $\mu\text{l}^{-1}$ )	1	1	2.5 U per 50 $\mu\text{l}$ PCR
Milli-Q water	to 50	to 50	
Final volume	50	50	

21| Set up and run the PCR program as follows:

Cycle number	Denature	Anneal	Extend
1	95 $^{\circ}\text{C}$ , 2 min		
2–41	95 $^{\circ}\text{C}$ , 30 s	57 $^{\circ}\text{C}$ , 30 s for 5-mC test 45 $^{\circ}\text{C}$ , 30 s for 5-hmC test	72 $^{\circ}\text{C}$ , 1 min
42			72 $^{\circ}\text{C}$ , 10 min

22| Verify PCR amplifications by running 10  $\mu\text{l}$  of each PCR on a 1.5% (wt/vol) agarose gel for 30 min at 100 V. There should be a single, clear band of  $\sim 300$  bp for both the 5-mC and 5-hmC test reactions.

23| TOPO-clone 2  $\mu\text{l}$  of each reaction using the Zero Blunt TOPO PCR cloning kit) according to the manufacturer's instructions. Isolate 20 clones (for 5-mC control test) or 10 clones (for 5-hmC control 2 test) with the Qiagen spin miniprep kit (follow the manufacturer's instructions) for Sanger sequencing.

▲ **CRITICAL STEP** If no quantitative conversion efficiency or protection efficiency is needed, skip Steps 23 and 24. Clean up the DNA from Step 22 by using the Qiagen PCR purification kit and then apply it to Sanger sequencing (Step 25).

24| Calculate the conversion rate of unmodified C, the conversion rate of 5-mC and the 5-hmC protection rate. The conversion rate of unmodified C from non-CpG cytosines in 5-mC/C spike-in control is usually over 99%. The ideal conversion rate of 5-mC

## PROTOCOL

is >96% and the protection efficiency of 5-hmC is >80% in the 5hmC spike-in control. The conversion rate of unmodified C = number of 'T' reads at non-CpG cytosines/(number of non-CpG cytosines × number of clones sequenced). The conversion rate of 5-mC = number of 'T' reads at CpG cytosines/(number of non-CpG cytosines × number of clones sequenced). The protection efficiency of 5-hmC = number of 'C' reads at cytosines/(number of cytosines × number of clones sequenced).

**▲ CRITICAL STEP** Do not take the cytosines of the primer sequence into consideration when doing the calculation.

### ? TROUBLESHOOTING

#### Sequencing of genomic DNA

**25|** Analyze the treated genomic DNA by locus-specific sequencing (option A) or whole-genome sequencing (option B).

##### (A) Locus-specific sequencing ● TIMING 12 h

(i) Apply ~200 ng of treated genomic DNA (from Step 18) to the Qiagen Epiectect bisulfite kit by following the manufacturer's instructions. Other bisulfite conversion kits can also be used.

**■ PAUSE POINT** The bisulfite-treated DNA can be stored at -20 °C for several weeks.

(ii) Use the table below as a guide to prepare 50- $\mu$ l PCRs with primers designed to detect 5-hmC at the loci of interest.

Component	Amount ( $\mu$ l)	Final concentration
PfuTurbo C <sub>x</sub> reaction buffer (10 $\times$ )	5	1 $\times$
dNTP mix (25 mM each)	0.4	200 $\mu$ M
Forward primer (10 $\mu$ M)	1	0.2 $\mu$ M
Reverse primer (10 $\mu$ M)	1	0.2 $\mu$ M
Bisulfite-treated DNA (from Step 25A(i))	2	
PfuTurbo C <sub>x</sub> DNA polymerase (2.5 U $\mu$ l <sup>-1</sup> )	1	2.5 U per 50 $\mu$ l PCR
Milli-Q water	to 50	
Final volume	50	

(iii) Set up and run the PCR program as follows:

Cycle number	Denature	Anneal	Extend
1	95 °C, 2 min		
2–41	95 °C, 30 s	Primer dependent	72 °C, 1 min
42			72 °C, 10 min

(iv) Verify PCR amplification by running 10  $\mu$ l of the PCR on a 1.5% (wt/vol) agarose gel for 30 min at 100 V. If a single, clear band of the desired size is seen, purify the PCR product with the Qiagen PCR cleanup kit or the gel extraction kit, according to the manufacturer's instructions.

(v) Apply the purified PCR products to Sanger sequencing or TOPO-clone 2  $\mu$ l of the reaction using the Zero Blunt TOPO PCR cloning kit by following the manufacturer's instructions for sequencing individual amplicons. The presence of C represents the positions of 5-hmC.

### ? TROUBLESHOOTING

##### (B) Whole-genome sequencing ● TIMING ~9 d

(i) Take 500 ng to 1  $\mu$ g of treated genomic DNA (from Step 18) for library construction according to Illumina's protocol for 'whole-genome bisulfite sequencing for methylation analysis'. ([https://shell.cgrb.oregonstate.edu/sites/default/files/Files/Docs/Illumina/prep/WGBS\\_for\\_Methylation\\_Analysis\\_Guide\\_15021861\\_A.pdf](https://shell.cgrb.oregonstate.edu/sites/default/files/Files/Docs/Illumina/prep/WGBS_for_Methylation_Analysis_Guide_15021861_A.pdf)).

(ii) Sequence the generated libraries using the Illumina Hi-seq 2000 platform according to the manufacturer's instructions.

### ? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
7	Little or no PCR product recovered	Gel purification failure	Ensure that the pH of buffer QG (in QIAquick gel extraction kit) is correct (<7.5). If not, add 10 µl of 3 M NaOAc (pH 5.0)
24	Low conversion rate of 5-mC to 5-caC	Inactive mTet	Ensure that recombinant mTet is active ( <b>Box 2</b> ) and aliquot into 10 µl before freezing. Avoid repeated freeze-thaw cycle
		Old Tet oxidation reagent 1 or 2	Store Tet oxidation reagents 1 and 2 properly (−80 °C) or freshly prepare before each assay
	Low protection rate of 5-hmC	Old UDP-glucose	Ensure that the UDP-glucose is stored properly
25A(v)	Little or no PCR product detected	Improper primers or thermal cycle	Design different primers or optimize the thermal cycle
		Not enough template	Increase the volume of template used in Step 25A(ii)

● TIMING

Steps 1–7, preparation of C, 5-mC and 5-hmC spike-in controls: ~7 h, day 1  
 Steps 8–11, sonication of genomic DNA: ~1 h, day 2  
 Steps 12–18, treatment of genomic DNA: ~5 h, day 2  
 Steps 19–24, verification of 5-mC oxidation and 5-hmC protection: 1–3 d, days 3–5  
 Step 25A, locus-specific sequencing: 12 h, days 6 and 7  
 Step 25B, whole-genome sequencing: ~9 d  
**Box 1**, expression and purification of recombinant mTet1: 9–10 d  
**Box 2**, assaying the activity of the purified recombinant mTet1: 4–5 d  
**Box 3**, purification of commercial 5-hmdCTP: 2 d

ANTICIPATED RESULTS

For genome-wide sequencing, the 5-hmC protection rate on the spike-in control should be >80% before adjusting for contaminants. The 5-mC oxidation rate (5-mC to 5-caC) on the spike-in control is the single most important parameter in TAB-seq, as it defines the lowest statistically obtainable abundance of 5-hmC. Because of the rarity of 5-hmC in most cell types, this value should not be <95%.

For locus-specific sequencing, genuine 5-hmC will display partially as C (i.e., not all C sites will be modified) in Sanger sequencing results depending on the abundance of 5-hmC at that site. The relative abundance of 5-hmC at the modification site can be further estimated by TOPO cloning.

Note: Supplementary information is available in the [online version of the paper](#).

**ACKNOWLEDGMENTS** This study was supported by the US National Institutes of Health (GM071440 and HG006827 to C.H., U01 ES017166 to B.R., NS079625 and HD073162 to P.J.), a Catalyst Award (to C.H.) from the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust, the Ludwig Institute for Cancer Research (to B.R.) and the Emory Genetics Discovery Fund (to P.J.).

**AUTHOR CONTRIBUTIONS** M.Y., C.-X.S. and C.H. conceived the original idea. M.Y., C.-X.S. and C.H. designed the experiment with the help from B.R. and P.J.; M.Y. performed treatment of genomic DNA; M.Y., G.C.H. and K.E.S. performed locus-specific sequencing; and G.C.H. and K.E.S. performed genome-wide sequencing. M.Y. and C.H. drafted the manuscript, and all the authors participated in writing and editing the manuscript.

**COMPETING FINANCIAL INTERESTS** The authors declare competing financial interests: details are available in the [online version of the paper](#).

Published online at <http://www.nature.com/doi/10.1038/nprot.2012.137>. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126**, 781–790 (1972).
2. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
3. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
4. Valinluck, V. *et al.* Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **32**, 4100–4108 (2004).
5. Valinluck, V. & Sowers, L.C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).



6. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841–4849 (2012).
7. Jin, S.G., Kadam, S. & Pfeifer, G.P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38**, e125 (2010).
8. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
9. Munzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
10. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
11. Song, C.X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72 (2011).
12. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25**, 679–684 (2011).
13. Ito, S. *et al.* Role of Tet proteins in 5-mC to 5-hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
14. Dawlaty, M.M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166–175 (2011).
15. Gu, T.P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
16. Iqbal, K., Jin, S.G., Pfeifer, G.P. & Szabo, P.E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. USA* **108**, 3642–3647 (2011).
17. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
18. Koh, K.P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200–213 (2011).
19. Szulwach, K.E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
20. Szulwach, K.E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011).
21. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
22. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
23. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
24. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
25. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **8**, 328–330 (2012).
26. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
27. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5**, e8888 (2010).
28. Josse, J. & Kornberg, A. Glucosylation of deoxyribonucleic acid. III.  $\alpha$ - and  $\beta$ -glucosyl transferases from T4-infected *Escherichia coli*. *J. Biol. Chem.* **237**, 1968–1976 (1962).
29. Lariviere, L. & Morera, S. Structural evidence of a passive base-flipping mechanism for  $\beta$ -glucosyltransferase. *J. Biol. Chem.* **279**, 34715–34720 (2004).
30. Song, C.X. *et al.* Detection of 5-hydroxymethylcytosine in DNA by transferring a keto-glucose by using T4 phage  $\beta$ -glucosyltransferase. *Chembiochem* **12**, 1682–1685 (2011).
31. Pastor, W.A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
32. Robertson, A.B. *et al.* A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **39**, e55 (2011).
33. Robertson, A.B., Dahl, J.A., Ougland, R. & Klungland, A. Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nat. Protoc.* **7**, 340–350 (2012).
34. Booth, M.J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).
35. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
36. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
37. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
38. Xu, Y. *et al.* Genome-wide regulation of 5-hmC, 5-mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell.* **42**, 451–464 (2011).
39. Flusberg, B.A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
40. Song, C.X. *et al.* Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat. Methods* **9**, 75–77 (2012).
41. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).



陈相军  
运营总监

艾德科技（北京）有限公司  
A&D Technology Corporation

地址：北京市昌平区中关村生命科学园路东60米  
邮编：102206  
电话：010-52406250 手机：15311218870  
网址：www.aderr.com Q Q：1951545998  
E-MAIL: tech@aderr.com

开户银行：中国银行北京奥运村支行  
银行账号：3428 5708 7342  
公司税号：1101 1457 3239 692