Base-resolution profiling of active DNA demethylation using MAB-seq and caMAB-seq

Hao Wu[1–6], Xiaoji Wu[1–4,6] & Yi Zhang[1–4]

[1]Howard Hughes Medical Institute, Boston, Massachusetts, USA. [2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts, USA. [3]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [4]Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA. [5]Present address: Department of Genetics, the Epigenetics Program, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to Y.Z. (yzhang@genetics.med.harvard.edu).

Published online 12 May 2016; doi:10.1038/nprot.2016.069

A complete understanding of the function of the ten-eleven translocation (TET) family of dioxygenase-mediated DNA demethylation requires new methods to quantitatively map oxidized 5-methylcytosine (5mC) bases at high resolution. We have recently developed a methylase-assisted bisulfite sequencing (MAB-seq) method that allows base-resolution mapping of 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), two oxidized 5mC bases indicative of active DNA demethylation events. In standard bisulfite sequencing (BS-seq), unmodified C, 5fC and 5caC are read as thymine; thus 5fC and 5caC cannot be distinguished from C. In MAB-seq, unmodified C is enzymatically converted to 5mC, allowing direct mapping of rare modifications such as 5fC and 5caC. By combining MAB-seq with chemical reduction of 5fC to 5hmC, we also developed caMAB-seq, a method for direct 5caC mapping. Compared with subtraction-based mapping methods, MAB-seq and caMAB-seq require less sequencing effort and enable robust statistical calling of 5fC and/or 5caC. MAB-seq and caMAB-seq can be adapted to map 5fC/5caC at the whole-genome scale (WG-MAB-seq), within specific genomic regions enriched for enhancer-marking histone modifications (chromatin immunoprecipitation (ChIP)-MAB-seq), or at CpG-rich sequences (reduced-representation (RR)-MAB-seq) such as gene promoters. The full protocol, including DNA preparation, enzymatic treatment, library preparation and sequencing, can be completed within 6–8 d.

INTRODUCTION

DNA cytosine methylation is an evolutionarily conserved epigenetic modification, and it is indispensable for normal mammalian development[1,2]. Enzyme-catalyzed active DNA demethylation contributes substantially to dynamic regulation of the DNA methylome during development and in diseases. In mammals, active DNA demethylation (converting 5mC back to C) is initiated primarily by the enzymatic activity of the TET family of 5mC dioxygenases[3–5]. TET proteins convert 5mC into 5-hydroxymethylcytosine (5hmC)[6–8]. Further oxidative modification of 5hmC by TET proteins results in 5fC and 5caC[9,10], both of which can be efficiently excised by thymine DNA glycosylase (TDG) and restored to unmodified cytosines through the base-excision repair pathway (**Fig. 1**)[5,9,11]. Genetic analyses of mutant mice deficient in TET proteins indicated that these enzymes have essential roles in a wide range of biological processes, including gene regulation[12–18], embryonic development[19–21], stem cell differentiation[8,22,23], meiotic gene control[24], erasure of genomic imprinting[19,25], learning and memory[26–28], and cancer[29,30].

The TET/TDG-dependent active demethylation pathway involves the generation and excision repair of 5fC/5caC, and it may occur in both proliferating and postmitotic cells. Besides acting as intermediates of the active DNA demethylation pathway, oxidized methylcytosines may also possess unique regulatory functions[31,32]. Recent studies have identified potential reader proteins for oxidized methylcytosines, many of which are transcription factors, chromatin-modifying enzymes or proteins linked to DNA repair processes[33–35]. Furthermore, accumulated 5fC and 5caC within gene bodies may have a regulatory role in decreasing the elongation rate of RNA polymerase II (refs. 36,37). Finally, clustered 5fCpG sites have recently been shown to affect base-pairing and DNA structure *in vitro*[38,39], and they may thus affect DNA-templated processes by directly modulating DNA conformation. A better mechanistic understanding of these potential regulatory roles of 5fC and 5caC requires the ability to systematically map the genomic position and determine absolute levels of these oxidized methylcytosines in the mammalian genome. Here we present detailed protocols for the MAB-seq method that we recently developed for single-base-resolution mapping of 5fC and 5caC (ref. 40), as well as for caMAB-seq, a modified version of MAB-seq for direct 5caC mapping[40].

Development of MAB-seq

Identification of cytosines that are committed to active DNA demethylation requires quantitative measurement of TDG-mediated excision of 5fC/5caC at high resolution. In addition, because both 5fC and 5caC are excised by TDG, strand-specific preference of active DNA demethylation activity requires simultaneous mapping of 5fC and 5caC in a single experiment. Recent studies using either modification-specific antibodies or chemical tagging have shown that genomic regions that are enriched for cytosines undergoing TET/TDG-dependent active demethylation can be identified via analysis of the ectopic accumulation of 5fC/5caC in *Tdg*-depleted cells[41,42]. However, 5fC and 5caC maps generated by affinity-enrichment methods are of limited resolution (a few hundred base pairs). In addition, these 5fC and 5caC profiles represent only relative enrichment and lack strand distribution information. To circumvent these limitations, we have developed MAB-seq, a modified BS-seq strategy that allows simultaneous and quantitative mapping of both 5fC and 5caC at single-base resolution[40,43]. In traditional BS-seq, C, 5fC and 5caC are efficiently deaminated through sodium bisulfite treatment, and all are read as thymine (T) in subsequent sequencing experiments. In contrast, 5mC and 5hmC are resistant to this chemical conversion and are read as C. In MAB-seq, genomic DNA is first treated with the bacterial DNA CpG methyltransferase M.SssI, which is an enzyme (originally isolated from *Spiroplasma* sp. strain MQ1)

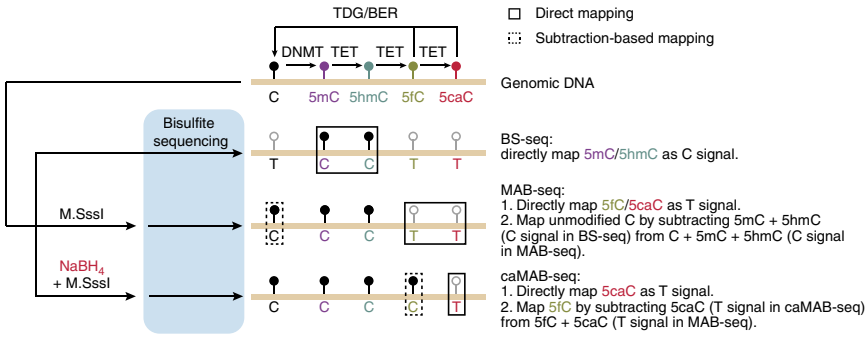

Figure 1 | Schematic diagram of BS-seq, MAB-seq and caMAB-seq. In mammalian cells, *de novo* and maintenance DNA methyltransferases (DNMTs) methylate unmodified cytosines (C) to generate 5-methylcytosines (5mC). The ten-eleven translocation (TET) family of DNA dioxygenases (TET1–3) is capable of iteratively oxidizing 5mC and its derivatives to generate three oxidized methylcytosines: 5hmC, 5fC and 5caC. Highly oxidized cytosine bases, 5fC and 5caC, are enzymatically excised by thymine DNA glycosylase (TDG), and the resulting abasic sites are repaired by the base-excision repair (BER) pathway to regenerate unmodified C, completing the DNA demethylation process (5mC to C). In standard bisulfite sequencing (BS-seq), 5mC and 5hmC are resistant to sodium-bisulfite-mediated deamination and are read as C in subsequent sequencing, whereas unmodified C, 5fC and 5caC are read as T. M.SssI exhibits robust methylase activity toward unmodified cytosines within CpGs. In MAB-seq, only 5fC and 5caC are read as T after genomic DNA is treated with M.SssI. In caMAB-seq, genomic DNA is first treated with sodium borohydride ($NaBH_4$) to reduce 5fC back to 5hmC, enabling 5caC to be directly mapped as T.

that efficiently methylates cytosines within CpG dinucleotides. Bisulfite conversion of M.SssI-treated DNA deaminates only 5fC and 5caC; originally unmodified C within CpGs becomes resistant to bisulfite conversion because of its conversion to 5mC. Subsequent sequencing will identify 5fC and 5caC as T, whereas C, 5mC and 5hmC are read as C (**Fig. 1**). We have also developed a method termed caMAB-seq (5caC methylase-assisted bisulfite sequencing) to directly map 5caC at single-base resolution. In this modified version of MAB-seq, 5fC is first reduced by $NaBH_4$ to 5hmC. Owing to the combination of $NaBH_4$ reduction with M.SssI treatment, 5caC is sequenced as T after bisulfite conversion, whereas C, 5mC, 5hmC and 5fC are read as C.

Application and limitations of MAB-seq

A major advantage of the MAB-seq method is that it allows for direct and simultaneous mapping of 5fC and 5caC, which requires less sequencing effort and enables quantitative detection of all TET/TDG-dependent DNA demethylation events in a single experiment at single-base resolution. MAB-seq is also amenable to both genome-scale and locus-specific analyses (**Fig. 2**). For instance, MAB-seq has been used to study both early embryos and mouse embryonic stem cells (ESCs), and it generated a whole-genome map of 5fC/5caC, as well as a quantitative profile of 5fC/5caC at specific loci[40,43,44]. The detection limit of MAB-seq is governed by several factors, including the error rate of the M.SssI methylase, the efficiency of bisulfite conversion of 5fC/5caC, the abundance of 5fC/5caC at the modified base, and overall sequencing depth. With the protocol described here, highly efficient conversion of C to 5mC and deamination of 5fC/5caC in genomic DNA can be achieved. Thus, sequencing depth is a major factor for sensitive and specific detection of 5fC/5caC by MAB-seq. However, MAB-seq is unable to distinguish 5fC/5caC from unmodified C within a non-CpG context because of the poor methylase activity of M.SssI toward C outside CpG dinucleotides. This limitation does not greatly affect the application of this technique for the following two reasons. First, whole-genome base-resolution mapping indicates that 5hmC is found almost exclusively in the CpG context (>99% in CpGs), even in mouse ESCs and neurons, in which non-CpG methylation is prevalent[45,46]. Second, TET proteins have a strong preference for 5mC within the CpG context compared with that in the non-CpG context[47,48]. Thus, 5mC within CpG dinucleotides is the primary target for TET proteins, and MAB-seq allows quantitative measurement of the abundance of 5fC/5caC within the CpG context.

Comparison of MAB-seq with other base-resolution 5fC or 5caC mapping methods

Both bisulfite-based and bisulfite-free mapping methods have been developed for mapping oxidized methylcytosines at single-base resolution (**Fig. 3**)[32]. Bisulfite-free base-resolution mapping methods, such as Pvu-Seal-seq[49] and fC-CET[50], currently

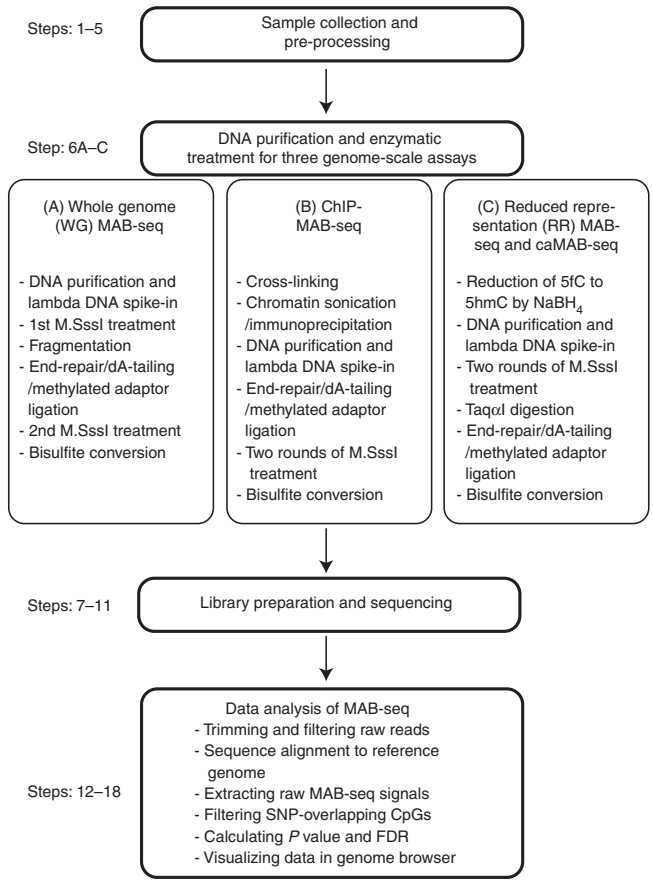

Figure 2 | Diagrammatic presentation of the workflow for MAB-seq and caMAB-seq.

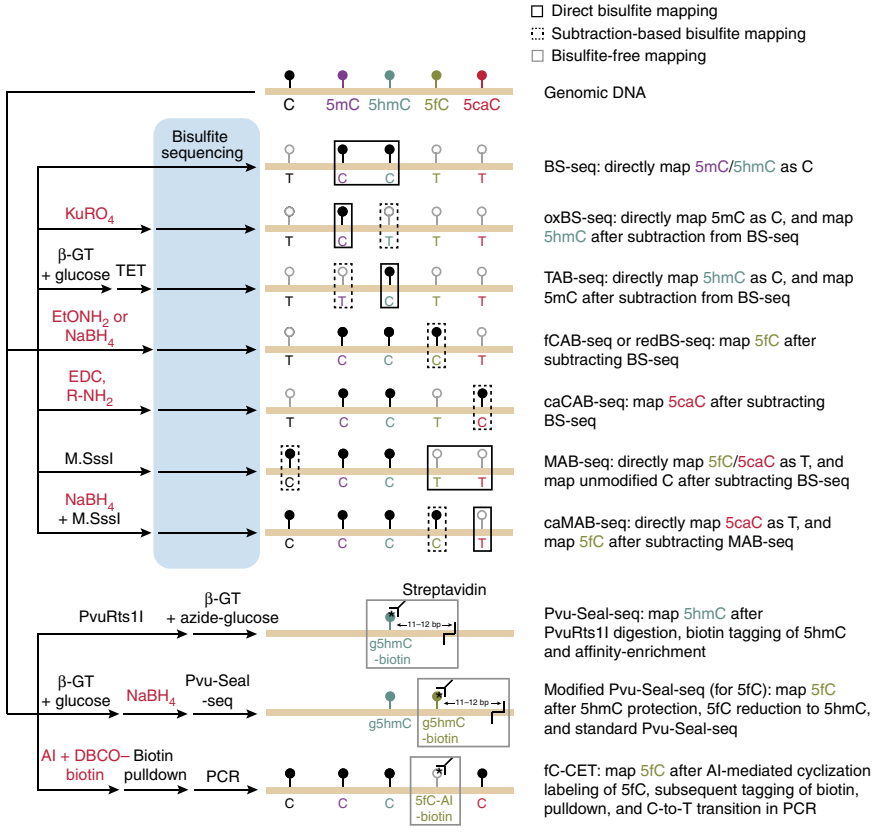

Figure 3 | Schematic diagram of base-resolution mapping methods for oxidized methylcytosines. Bisulfite sequencing (BS-seq)-based (upper panels) or bisulfite-free (lower panels) base-resolution mapping methods, coupled with various chemical (in red) and enzymatic (in black) treatments, have been developed to profile 5fC and 5caC. In direct bisulfite mapping, the position and abundance of specific oxidized methylcytosines (e.g., mapping of 5hmC by TAB-seq[46]) can be directly determined. In subtraction-based bisulfite mapping, determining the difference in signal between conventional BS-seq and modified BS-seq (e.g., mapping of 5hmC by subtracting signals of oxBS-seq[67] from those of standard BS-seq) is required to indirectly determine the position and abundance of oxidized methylcytosines. In two bisulfite-free mapping strategies, modification-sensitive restriction enzyme (5hmC-specific endonuclease PvuRts1I for Pvu-Seal-seq) or chemical (AI [azido derivative of 1,3-indandione] for fC-CET)-assisted tagging of oxidized methylcytosines is used to both enrich the genomic fragments containing methylcytosines (giving relative abundance) and identify the position of those cytosine variants. Adapted with permission from ref. 32.

provide only relative enrichment of cytosines marked by 5fC. Compared with bisulfite-free methods, a major advantage of bisulfite-based mapping methods (including MAB-seq and caMAB-seq) is their ability to determine both base-resolution location and absolute levels of 5fC or 5caC. Before MAB-seq[40,43], several bisulfite-based mapping methods were developed that use specific chemical treatment to protect 5fC (fCAB-seq[41] and redBS-seq[51]) or 5caC (caCAB-seq[52]) from bisulfite-induced deamination. However, these methods require subtraction of the C signals between standard BS-seq and those of modified BS-seq to determine the position and abundance of 5fC or 5caC, thereby doubling the required sequencing effort. In addition, MAB-seq is the only method that is capable of detecting 5fC and 5caC simultaneously, which is required for base-resolution analysis of the strand-specificity of TET/TDG-mediated active DNA demethylation events. Moreover, MAB-seq and caMAB-seq can be integrated to map 5fC and 5caC together (MAB-seq), 5fC individually (by subtracting caMAB-seq from MAB-seq) and 5caC individually (caMAB-seq) at single-base resolution. Thus, with two genome-scale mapping experiments (MAB-seq plus caMAB-seq), this integrated approach not only provides quantitative base-resolution mapping of active DNA demethylation events (generation and excision of 5fC/5caC) but also distinguishes 5fC from 5caC at base resolution.

Experimental design

Enzymatic treatment (Step 6A (iv, v, xi, xii) and Step 6C (v, vi)). Because incomplete methylation of unmodified CpGs will result in false positive 5fC/5caC signals, the conditions of the enzymatic treatment with M.SssI are critical to the success of MAB-seq. *S*-adenosylmethionine (SAM) is unstable at 37 °C and is sensitive to degradation at elevated pH (i.e., pH 7.5). Moreover, *S*-adenosylhomocysteine, the byproduct of the methylation reaction, binds more tightly to M.SssI than does SAM, and it greatly reduces the methylation reaction rate as time passes. Thus, it is important to ensure that only fresh SAM is used, as cofactors in the enzymatic reaction and the addition of more SAM after 2 h can increase the methylation rate. Because a high concentration of input DNA leads to the accumulation of high amounts of *S*-adenosylhomocysteine, we recommend that the concentration of input DNA does not exceed 20 ng/μl. M.SssI is more processive in the absence of magnesium and more distributive in the supplied reaction buffer (NEBuffer2 contains 10 mM $MgCl_2$). To maximize the methylation rate, we recommend treating genomic DNA in Mg^{2+}-free buffer for one round and in the supplied reaction buffer for the second round.

$NaBH_4$ reduction of 5fC to 5hmC (Step 6C(ii, iii)). Several studies have shown that $NaBH_4$ can efficiently reduce 5fC to 5hmC, which changes the behavior of 5fC during bisulfite sequencing and offers a strategy for 5fC mapping[40,41,49,51]. In caMAB-seq, a step of $NaBH_4$ reduction is added before M.SssI treatment. In our hands, this reduction step is compatible with M.SssI treatment and library preparation based on a reduced-representation bisulfite sequencing (RRBS) strategy[53]. We termed this method for genome-scale direct mapping of 5caC RR-caMAB-seq (Step 6C).

Bisulfite conversion of 5fC/5caC (Step 6A(xiv–xvi) and Step 6C(xiii)). Unlike unmodified C and 5caC (deamination rate >99%), converting 5fC to uracil by bisulfite treatment is less efficient[51]. We have optimized the bisulfite conversion conditions using different concentrations of sodium bisulfite (the Epitect Fast kit provides a much higher concentration of

TABLE 1 | Comparison of different bisulfite conversion kits and protocols

Bisulfite conversion kit	Time	5mC conversion rate (%)	5hmC conversion rate (%)	5fC conversion rate (%)	5caC conversion rate (%)
Epitect	10 h	2.2	3.2	83.8	99.6
Epitect	10 h + 5 h	3.5	3.7	NA	NA
Epitect Fast	20 min × 2	2.8	3.3	59.9	99.5
Epitect Fast	20 min × 4	5.1	4.4	76.5	NA
Epitect Fast	20 min × 6	7.5	5.5	84.4	NA
Epitect Fast	40 min × 4	7.6	5.5	84.2	NA

Note 1: numbers shown are the percentage of the indicated modification sequenced as T after bisulfite conversion.

Note 2: "Epitect 10 h" refers to the recommended protocol (the standard 5-h thermal program is performed twice).

Note 3: "Epitect 10 h + 5 h" refers to purifying the DNA after the 10-h recommended protocol and performing another round of bisulfite conversion using a standard 5-h thermal program.For the Epitect Fast kit, the standard thermal program is to perform two cycles of 95 °C for 5 min, 60 °C for 10 min. All four tested protocols are either elongating the 10 min at 60 °C to 20 min or 40 min or doing four or six cycles instead of two.

Note 4: there is a clear trend that harsher conditions lead to more side reactions such as unintended deamination of 5mC and 5hmC. The 5mC conversion rate is assessed using methylated lambda DNA. The 5hmC/5fC/5caC conversion rate is assessed using synthetic 38-mer oligos. The actual conversion rate may slightly differ from the measured conversion rate, as the methylated lambda DNA and synthetic oligos may not be perfect.

sodium bisulfite reagent than does the standard Epitect kit) and various incubation times. With the optimized bisulfite conversion protocol described in **Table 1** (standard Epitect + 10 h treatment), ~85% of 5fC can be deaminated with minimal conversion of 5mC or 5hmC (measured by methylated lambda DNA or modification-specific synthetic oligonucleotides).

Quality controls. To assess the methylation rate within the CpG context, we use unmethylated lambda DNA as a spike-in control in both genome-wide and locus-specific MAB-seq experiments (**Box 1**). To assess the behavior of oxidized bases (5hmC/5fC/5caC) in experiments for optimizing/validating reagents and conditions, we use modification-specific 38-mer synthetic double-stranded oligonucleotides containing nine CpG sites (**Box 2**). Validation of bisulfite treatment reagents via measurement of the deamination rate of unmodified cytosine can be performed along with MAB-seq experiments, and the bisulfite conversion rate of unmodified cytosine is ~99.5% under optimal experimental conditions.

Enrichment strategy (Step 6B and C). Whole-genome (WG) maps of 5fC and 5caC can be generated by the application of MAB-seq analysis to unenriched genomic DNA (WG-MAB-seq; Step 6A). However, WG-MAB-seq is sequencing-intensive, and it can be cost-prohibitive for samples with relatively low levels of 5fC/5caC. To reduce sequencing efforts, enrichment strategies can be integrated with the MAB-seq workflow to generate genome-scale maps. By combining ChIP with MAB-seq (ChIP-MAB-seq), one can examine 5fC/5caC abundance within a fraction of the genome where 5fC/5caC marks tend to be enriched (Step 6B)[40]. In addition to histone antibodies, DNA immunoprecipitation using antibody to 5fC and/or 5caC could also be used to first enrich DNA fragments containing these rare modified bases before MAB-seq analysis (H. Wu and Y. Zhang, unpublished observations). Coupling restriction-digestion-based RR-MAB-seq allows 5fC/5caC mapping within genomic regions containing CpG-rich sequences (mostly gene promoters and some repeat sequences) (Step 6C)[43]. It is important to note that the MspI enzyme used in standard RRBS only partially digests 5fC-containing C^CGG sites and completely fails to cut 5caCpGs, which will lead to underestimation of 5fC/5caC levels at the digestion sites[10]. Therefore, it is essential to choose a restriction enzyme (e.g., TaqαI) that efficiently cuts 5fC- and 5caC-modified CpGs. Finally, locus-specific MAB-seq or caMAB-seq analysis of regions of interest can be performed using appropriately designed PCR primers (**Box 3**).

Box 1 | Quality control of MAB-seq and caMAB-seq experiment ● TIMING 1 d

Lambda DNA spike-in for assessing M.SssI methylation and bisulfite conversion:

1. Before starting M.SssI treatment, spike 0.25% (wt/wt) unmethylated lambda DNA into sample genomic DNA.
2. Perform MAB-seq using the selected protocol (WG-MAB-seq, ChIP-MAB-seq, RR-MAB-seq or RR-caMAB-seq).
3. Align sequencing reads to lambda DNA genome. CpG sites are expected to be fully methylated (97–99%), whereas non-CpG sites are expected not to be efficiently methylated (0–1.5%).

***Tet1/2/3* triple-knockout (Tet TKO) cells or *Dnmt1/3a/3b* triple-knockout (Dnmt TKO) cells as negative controls for assessing false discovery rate (FDR):**

4. For each set of experiments, include one sample using Tet TKO or Dnmt TKO cells as a negative control, and perform the experiments in parallel with samples of interest. When these two cell lines are not available, other samples with low 5fC/5caC levels can serve as a negative control.
5. During data analysis, FDR for a sample of interest is calculated as (number of 5fC/5caC-modified CpG called in negative control)/(number of 5fC/5caC-modified CpG called in the sample of interest).

Box 2 | 5hmC/5fC/5caC-modified oligo spike-in for validating experimental conditions ● TIMING 2 d

1. Combine 250 pg of modified oligo with 250 ng of carrier genomic DNA.
2. When testing caMAB-seq, perform Step 6C(ii–ix). When testing MAB-seq, perform Step 6C(v–ix).
3. Perform bisulfite conversion according to the protocol in Step 6A(xiv–xvi).
4. Amplify the bisulfite-converted oligo using NEBNext multiplex oligo for Illumina, purify the amplified DNA and perform Illumina sequencing. A few thousand reads are sufficient for the purpose of quality control.
5. After sequencing data have been obtained, align the reads to the oligo template sequence. Non-CpG sites are expected not to be methylated, whereas 5hmC/5fC/5caC-modified CpG sites should behave as expected (**Table 1**).

Data analysis (Steps 12–18). For MAB-seq analysis, raw signals were calculated as the percentage of T/(C + T) at each CpG dinucleotide or genomic interval (**Box 4**). We counted the number of 'T' bases from MAB-seq reads as 5fC/5caC (denoted by N_T) and the number of 'C' bases as other forms of cytosines (C/5mC/5hmC; denoted by N_C).

There are potentially two major sources of false positive signals: (i) unmethylated CpG sites that were not methylated by M.SssI; and (ii) bisulfite-treatment-induced deamination of 5mC and 5hmC. To statistically account for both sources of false positive signals, we modeled these presumably stochastic events with a binomial distribution $X \sim B(N, p)$ (where random variable X follows binomial distribution B with parameters N and p, N is the sequencing coverage ($N_T + N_C$) at a given CpG site and p is the probability of detecting false positive signals, which is the sum of both the failure rate of M.SssI and the deamination rate of 5mC)[40]. By using unmethylated lambda DNA as an internal spike-in control, we can experimentally determine p for every experiment (in our published study, we used 2.04% as p, which was averaged from multiple genome-scale MAB-seq experiments[40]) and calculate the probability of whether raw MAB-seq signals ($N_T/(N_C + N_T)$) are significantly higher than would be expected by chance. A notable caveat for this statistical filtering strategy is that deamination of 5hmC cannot be directly accounted for. However, given that the 5hmC modification level within CpGs generally ranges from 10 to 30% and the deamination rate of 5hmC in our optimized

Box 3 | Locus-specific MAB-seq and caMAB-seq experiment ● TIMING 3 d

1. Spike 0.25% (wt/wt) lambda DNA into the genomic DNA of interest.
2. When testing caMAB-seq, perform Step 6C(ii–ix). When testing MAB-seq, perform Step 6C(v–ix).
3. Perform bisulfite conversion according to the protocol in Step 6A(xiv–xvi).
4. Amplify regions of interest by PCR using primers designed for bisulfite-converted genomic DNA. Programs such as methprimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) can facilitate the BS-PCR primer design. We recommend choosing primers that generate PCR amplicons between 200 and 500 bp. In addition, amplify a region from lambda DNA for spike-in controls. For each pair of primers, set up a 15-µl reaction using 2× KAPA HiFi Uracil+ HotStart ReadyMix. Using properly designed primers and sufficient starting material (we typically start with 250 ng of genomic DNA and use 1/20 of the bisulfite-converted DNA as a template for each BS-PCR reaction), 35–40 cycles of amplification should generate sufficient products for visualization on agarose gel.
5. Perform downstream analysis using either Sanger sequencing (option A) or Illumina sequencing (option B), depending on the experimental purpose and service availability.

(A) Sanger sequencing

(i) Clone individual PCR product with the Zero Blunt TOPO PCR cloning kit, according to the manufacturer's instructions.
(ii) For each amplified locus of a sample, pick individual colonies (<30 clones) for Sanger sequencing.

(B) Illumina sequencing

(i) Cast a 2% (wt/vol) agarose gel. Run 1 µl of BS-PCR products (diluted in 9 µl of water with 2 µl of 6× loading dye added) at 20 V/cm for 20 min. Include in one lane 10 µl of the quantitative DNA ladder. Image the gel using a standard gel imaging system. By comparing it with a ladder with known DNA mass, quantify the concentration of all BS-PCR amplicons on the basis of signal intensity using ImageJ or other gel quantification software.
(ii) Adjust the concentration of purified PCR products by adding nuclease-free water so that each PCR amplicon has the same molar concentration. On the basis of the concentrations of BS-PCR products, pool comparable amounts of PCR products for multiple genomic loci (from one sample) together.
(iii) Sonicate the mixture to a range between 100 and 200 bp. If the PCR amplicons are relatively small (from 200 to 350 bp), it is difficult to completely shear PCR products to smaller fragments. In this case, if more than 20% of the total PCR amplicons appear to be sonicated, proceed to the next step.
(iv) Purify the sonicated PCR products using the Qiagen MinElute PCR purification kit, and prepare sequencing libraries using the NEBNext DNA Library Prep Master Mix Set for Illumina.
(v) Perform Illumina sequencing using HiSeq 2500 or an equivalent sequencer.

6. Align sequencing results to amplicon sequences to quantify 5fC/5caC signals.

Box 4 | Identification of 5fC/5caC-enriched genomic intervals ● TIMING 1 d

1. Segment the genome into nonoverlapping 100-bp genomic bins. The bin information can be stored in bed format:

```
chr1      0      99
chr1    100     199
…
```

2. Identify the CpG sites covered at least five times in both the sample of interest and the negative control (common 5× CpG sites). Store the information in bed format.
3. Identify genomic bins with more than two common 5× CpG sites. Example command using coverageBed (from BEDTools):

```
/path-to-bedtools/coverageBed -a <common 5× CpG bed file> -b <100 bp
bin bed file> | awk -v OFS='\t' '{if ($4>=2) print $1, $2, $3, $4;}' >
<bin with at least two 5× CpG sites>
```

4. For each genomic bin with more than two common 5× CpG sites, sum up the C and T counts from all the common 5× CpG sites in this bin. MAB-seq or caMAB-seq signals of this bin are calculated as $(\text{Sum}(N_T)/(\text{Sum}(N_T) + \text{Sum}(N_C)))$.
5. Identify 5fC/5caC-modified bins and assess empirical FDR using the approach described in Step 18.

bisulfite conversion condition is quite small (3.2%), the chance of detecting deaminated 5hmC as 5fC/5caC is quite low (0.3–1%) at a given CpG site[40,46]. Indeed, when raw MAB-seq signals for called CpGs in *Tdg*-depleted cells are compared with those in wild-type cells (where 5hmC is present at similar levels), global MAB-seq signals in *Tdg* mutant cells are two- to threefold higher than in wild-type cells within 5fC/5caC-enriched genomic regions[40]. In addition, we have shown that >90% of 5hmC-modified CpGs are nonoverlapping with 5fC/5caC-modified sites in mouse ESCs[40]. Together, these observations indicate that 5hmC is unlikely to contribute significantly to the false positive signals in MAB-seq.

To estimate the empirical false discovery rate (FDR) in calling 5fC/5caC-modified CpGs, the steps above are repeated on MAB-seq signals of a negative control sample. For instance, we have successfully used genomic DNAs extracted from *Dnmt1/3a/3b*$^{-/-}$ or *Tet1/2/3*$^{-/-}$ mouse ESCs as negative control samples, as these cells do not contain any 5fC or 5caC modifications (**Box 1**). The FDR for a given *P* value cutoff is the number of called CpG sites (false positive signals) in negative controls divided by the number of called CpG sites (potential true signals) in the sample of interest. Please note that it is ideal to have a negative control processed side by side with the samples of interest for each batch of experiments, because this allows the estimation of FDR based on parallel experiments.

MATERIALS

REAGENTS

- Wild-type feeder-independent mouse ESC line E14TG2a (ATCC, cat. no. ATCC CRL-1821) ! CAUTION The cell lines used in your research should be regularly checked to ensure that they are authentic and that they are not infected with mycoplasma.
- Cells of interest. We have used Tdg-depleted[42] mouse ESCs, *Tet1/2/3* triple-knockout[54] mouse ESCs and *Dnmt1/3a/3b* triple-knockout[55] mouse ESCs (not commercially available)
- UltraPure DNase/RNase-free distilled water (Life Technologies, cat. no. 10977-023)
- Dulbecco's phosphate-buffered saline (DPBS) (Life Technologies, cat. no. 14190-250)
- Ethanol, absolute (200 proof; Fisher Scientific, cat. no. BP28184) ! CAUTION Ethanol is highly flammable. Handle it in a fume hood.
- Glycine (Sigma-Aldrich, cat. no. 50046)
- Formaldehyde (37% (vol/vol); Fisher Scientific, cat. no. BP531) ! CAUTION Formaldehyde is toxic if absorbed through skin, swallowed or inhaled. It is also flammable. Handle it in a fume hood with appropriate equipment.
- Triton X-100 (Sigma-Aldrich, cat. no. T8787) ! CAUTION Triton X-100 can cause skin and eye irritation. Handle it with caution.
- HEPES buffer solution (1 M; Sigma-Aldrich, cat. no. 83264)
- KOH (Sigma-Aldrich, cat. no. 484016) ! CAUTION KOH is corrosive if swallowed, and it causes skin and eye burn. Wear laboratory clothing (gloves, lab coat and goggles) when handling it.
- Sodium deoxycholate (Sigma-Aldrich, cat. no. D6750) ! CAUTION Sodium deoxycholate is harmful if swallowed, and it can lead to respiratory irritation. Handle it with caution.
- NP-40 Surfact-Amps detergent solution (10% (wt/vol); Life Technologies, cat. no. 28324)
- N-Lauroylsarcosine sodium salt solution (20% (wt/vol); Sigma-Aldrich, L7414)
- LiCl (Sigma-Aldrich, cat. no. L9650)
- NaCl (5 M; Life Technologies, cat. no. AM9760G)
- $NaBH_4$ (Sigma-Aldrich, cat. no. 480886) ! CAUTION $NaBH_4$ is corrosive. Hydrogen gas is produced when $NaBH_4$ reacts with water. Handle it with suitable equipment.
- Sodium acetate buffer solution (3 M, pH 5.2 ± 0.1; Sigma-Aldrich, cat. no. S7899)
- EDTA (0.5 M, pH 8.0; Life Technologies, cat. no. 15575-020)
- EGTA (0.5 M, pH 8.0; prepared from Sigma-Aldrich, cat. no. 03777)
- RNase A (10 mg/ml; Life Technologies, cat. no. EN0531)
- Anti-H3K4me1 (10 µg/ml; Abcam, cat. no. ab8895)
- TaqαI (20 U/µl; New England BioLabs, cat. no. R0149L)
- Klenow fragment (exo$^-$, 5 U/µl; Thermo Scientific, cat. no. EP0422)
- T4 DNA ligase (2,000 U/µl; New England BioLabs, cat. no. M0202M)
- CutSmart buffer (10×; New England BioLabs, cat. no. B7204S)
- ATP (100 mM; Thermo Scientific, cat. no. R0441)
- dNTP set 100 mM solutions (Life Technologies, cat. no. R0181)
- SPRIselect reagent kit (Beckman Coulter, cat. no. B23318)
- DNeasy blood & tissue kit (Qiagen, cat. no. 69504)
- Qiagen EpiTect DNA bisulfite kit (Qiagen, cat. no. 59104)
- QIAquick nucleotide removal kit (Qiagen, cat. no. 28304)
- MinElute PCR purification kit (Qiagen, cat. no. 28004)
- M.SssI (20 U/µl, New England BioLabs, M0226M)
- NEBuffer 2 (10×; New England BioLabs, cat. no. B7002S)
- *S*-adenosylmethionine (SAM; 32 mM; New England BioLabs, cat. no. B9003S) ▲ CRITICAL Always use SAM before its expiration date, and make aliquots to avoid multiple freeze–thaw cycles.
- KAPA HiFi Uracil+ HotStart ReadyMix (Kapa Biosystems, KK2801)

- NEBNext Multiplex Oligos for Illumina (Index Primers Set 1; New England BioLabs, cat. no. E7335L)
- NEBNext Multiplex Oligos for Illumina (Index Primers Set 2; New England BioLabs, cat. no. E7500L)
- NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs, cat. no. E7370S)
- NEBNext DNA Library Prep Master Mix Set for Illumina (New England BioLabs, cat. no. E6040S)
- Custom methylated adaptors (asterisk denotes phosphorothioate bond; all cytosines are modified as 5mC; Integrated DNA Technologies): Forward: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3′, Reverse: 5′-/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTC-3′
- Custom 5hmC/5fC/5caC-modified oligo (X: 5hmC, 5fC or 5caC): Forward: 5′-AGCCXGXGCXGXGCXGGTXGAGXGGCXGCTCCXGCAGC-3′, Reverse: 5′-GCTGXGGGAGXGGCXGCTXGACXGGXGXGGXGXGGGCT-3′
  ▲ **CRITICAL** Only CpG sites of these oligos are modified, which is different from the methylated adaptor in which all cytosines are modified as 5mC.
- Unmethylated lambda DNA (Promega, cat. no. D1521)
- Agilent High Sensitivity DNA Kit (Agilent Technologies, cat. no. 5067-4626)
- cOmplete EDTA-free proteinase inhibitor (Roche, cat. no. 11873580001)
- Dynabeads Protein G for Immunoprecipitation (ThermoFisher Scientific, cat. no. 10004D)
- RNase A (Life Technologies, cat. no. 12091-201)
- Proteinase K (New England BioLabs, cat. no. P8107S)
- Glycogen (Roche Life Science, cat. no. 10901393001)
- 5 PRIME Phase Lock Gel heavy 2 ml (Fisher Scientific, cat. no. FP2302830)
- Phenol–chloroform–isoamyl alcohol mixture (Sigma-Aldrich, cat. no. 77617)
  ! **CAUTION** Phenol can cause skin and eye burn. Wear laboratory clothing (gloves, lab coat and goggles) and handle it with caution.
- Qubit dsDNA BR Assay Kit (Life Technologies, cat. no. Q32853)
- Qubit dsDNA HS Assay Kit (Life Technologies, cat. no. Q32854)
- Zero Blunt TOPO PCR cloning kit (Life Technologies, cat. no. K2800-20)
- Illumina sequencing reagents for HiSeq 2500 or equivalent sequencers

EQUIPMENT

- Eppendorf 0.2-ml PCR tube strips (Fisher Scientific, cat. no. E0030124286)
- Eppendorf RNA/DNA LoBind microcentrifuge tubes (1.5 ml; Sigma-Aldrich, cat. no. Z666548)
- Corning CentriStar centrifuge tubes (15 ml; Fisher Scientific, cat. no. 0553859A)
- Corning CentriStar centrifuge tubes (50 ml; Fisher Scientific, cat. no. 0553860)
- Agilent 2100 Bioanalyzer (Agilent, cat. no. G2939AA)
- Qubit 2.0 Fluorometer (Life Technologies, cat. no. Q32866)
- Qubit assay tubes (Life Technologies, cat. no. Q32856)
- Branson Sonifier 450 (Branson Ultrasonics)
- M220 Focused-ultrasonicator (Covaris)
- microTUBE-50 AFA Fiber Screw-Cap (Covaris, cat. no. 5201660)
- Filter tips (10, 20, 200 and 1,000 µl; Genesee Scientific)
- Serological pipette (1, 5, 10 and 25 ml; Genesee Scientific)
- C1000 thermal cycler (Bio-Rad, cat. no. 185-1096)
- Eppendorf Thermomixer R (Fisher Scientific, cat. no. 05-400-203)
- Vortex-Genie 2 (Scientific Industries, cat. no. SI-0286)
- Illumina sequencer (HiSeq 2500 or equivalent models)

Software

- Trim galore (version 0.3.7): http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- FastQC (version 0.10.1): http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
- Bismark (version 0.12.5)[56]: http://www.bioinformatics.babraham.ac.uk/projects/bismark/
- Bowtie2 (version 2.2.2)[57]: http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
- Picard (version 1.91): http://broadinstitute.github.io/picard/
- Samtools (version 0.1.19)[58]: http://samtools.sourceforge.net
- IGV genome browser[59]: download from https://www.broadinstitute.org/igv/
- BEDTools (version 2.19.0)[60]: http://bedtools.readthedocs.org/en/latest/
- BisSNP (version 0.82.2)[61]: http://people.csail.mit.edu/dnaase/bissnp2011/

REAGENT SETUP

Cross-linking buffer (11×) ChIP cross-linking buffer consists of 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 11% (vol/vol) formaldehyde. Freshly prepare the buffer for each experiment.

Blocking solution Mix 0.5% (wt/vol) BSA in 1× DPBS. Freshly prepare the solution for each experiment.

ChIP lysis buffer Combine 10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% (wt/vol) sodium deoxycholate, 0.5% (wt/vol) *N*-lauroylsarcosine and 0.1% (wt/vol) SDS. This solution can be stored at room temperature (22–25 °C) for 1 year.

RIPA buffer Make ChIP RIPA washing buffer with the following composition: 50 mM HEPES-KOH (pH 7.5), 500 mM LiCl, 1 mM EDTA, 1% (wt/vol) NP-40 and 0.7% (wt/vol) Na-deoxycholate. This solution can be stored at 4 °C for 1 year.

TE buffer This buffer consists of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA. This solution can be stored at room temperature for 1 year.

Elution buffer The ChIP elution buffer consists of 50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% (wt/vol) SDS. This solution can be stored at room temperature for 1 year.

Mg^{2+}-free M.SssI buffer (10×) Make 10× homemade M.SssI buffer with the following composition: 100 mM Tris-HCl (pH 8.0), 500 mM NaCl and 100 mM EDTA. Dilute it to 1× final concentration when used in reaction. Compared with the supplied NEBuffer 2, this Mg^{2+}-free buffer may help M.SssI work in a more processive manner. This buffer can be stored at 4 °C for at least 6 months.

dNTP mix for RR-MAB-seq and RR-caMAB-seq Make dNTP mix stock solution with the following composition: 10 mM dATP, 1 mM dGTP and 1 mM dCTP. Dilute the stock solution ten times in nuclease-free water to a working solution (1 mM dATP, 0.1 mM dCTP and 0.1 mM dGTP) immediately before use, and add the indicated volume to achieve the final concentration. The stock solution can be stored at −20 °C for at least 6 months. ▲ **CRITICAL** TaqαI digestion generates 5′ CG overhang (5′-T^CGA-3′, where '^' denotes the cutting site); therefore, dCTP and dGTP are required for end repair. The high concentration of dATP is for 3′ dA-tailing after end repair (both processes are taking place in a single reaction). However, when other restriction enzymes generating different overhangs are used, the dNTP composition needs to be adjusted accordingly. Make the working solution each time before use, and do not store it.

Methylated adaptors Dilute the forward and reverse oligos to 100 µM with nuclease-free water. Set up the following reaction to anneal the forward and reverse oligos: 10 µl of forward oligo, 10 µl of reverse oligo, 2.5 µl of water and 2.5 µl of NEBuffer 2. The thermal cycle for annealing is as follows: 94 °C for 10 min; slowly ramp down to 60 °C and hold for 10 min (1% ramp if using an Eppendorf thermal cycler, or similar rate (for example, 0.1 °C per second) if using other thermal cyclers); and then slowly ramp down to 4 °C and hold (using the same ramping rate as in the previous step). This will generate a 40 µM annealed adaptor. Dilute the annealed adaptor to 15 µM, make aliquots and store them at −20 °C. For RR-MAB-seq or RR-caMAB-seq adaptor ligation, dilute the adaptor stock (15 µM) 20 times with water to a working solution (0.75 µM), and add the indicated volume. ▲ **CRITICAL** Use LoBind tubes for all procedures to avoid loss of oligos during the preparation and storage, and do not stock diluted oligos.

5hmC/5fC/5caC-modified oligo Dilute the forward and reverse modified oligos (ssDNA) with nuclease-free water to 1 µg/µl. Set up the following reaction to anneal the forward and reverse oligos: 4 µl of forward oligo, 4 µl of reverse oligo, 1 µl of water and 1 µl of NEBuffer 2. Perform oligo annealing using the same thermal program for annealing methylated adaptor oligos described above (the annealing reaction can be scaled down proportionally), which will give rise to 0.8 µg/µl 38-bp dsDNA. Take 2 µl (1.6 µg) for end repair and dA-tailing by NEBNext DNA Library Prep Master Mix Set for Illumina according to the manufacturer's instruction, with two modifications: the reaction is scaled down to 60% (60 µl for end repair and 30 µl for dA-tailings), and purification after end repair and dA-tailings is performed using the QIAquick nucleotide removal kit. Use the Qubit dsDNA HS kit to measure the concentration of the recovered 38-bp dsDNA, and ligate it to the methylated adaptor with a DNA:adaptor molar ratio of 1:10 (the ligation reaction is set up in 30 µl using the quick T4 ligase provided in the library preparation kit). Purify the adaptor-ligated oligos with the QIAquick nucleotide removal kit, and adjust the concentration to 1 ng/µl for storage at −20 °C. ▲ **CRITICAL** Use LoBind tubes for all procedures to avoid loss of oligos during the preparation and storage, and do not stock diluted oligos.

Sodium borohydride aqueous solution for caMAB-seq Before each experiment, make 1 M sodium borohydride solution by dissolving sodium borohydride in nuclease-free water. ▲ **CRITICAL** Sodium borohydride is unstable in aqueous solution, so the solution should be made freshly before each use.

PROCEDURE

Cell culture and vitamin C treatment ● TIMING 4–5 d

▲ CRITICAL Steps 1–5 can be readily adapted for cell types other than mouse ESCs.

1| Culture wild-type (e.g., E14TG2a), *Tdg*-depleted[42] and negative control (*Tet1/2/3* triple-knockout[54] or *Dnmt1/3a/3b* triple-knockout[55]) mouse ESCs in feeder-free conditions on 0.1% (wt/vol) gelatin-coated six-well plates or 10-cm dishes. Passage the cells every 2–3 d, and change the culture medium daily.

2| To better mimic the metabolic milieu *in vivo*, maintain mouse ESCs in culture medium containing 100 μg/ml (final concentration) of vitamin C (VC) for 60 h.

▲ CRITICAL STEP Because VC is present at a relatively high level in both embryonic and adult mouse tissues[62] and positively regulates catalytic activity of TET enzymes[63,64], it is recommended that a physiologically relevant amount of VC be supplemented to the cell culture medium either transiently or long-term.

3| At ~70% confluency, aspirate the medium and wash the cells with 1× DPBS. Aspirate DPBS and add sufficient 0.05% (wt/vol) trypsin to the culture dish (3 ml for each 10-cm dish). Incubate the mixture at 37 °C for 5 min.

4| After the cells have detached, add DMEM with 20% (vol/vol) FBS to inactivate the trypsin (6 ml of 10-cm dish), gently pipette to dissociate the cells and collect the cell suspension in a 15-ml Falcon conical tube.

5| Centrifuge the cell suspension at 500*g* for 5 min at room temperature, and aspirate the supernatant and discard it.

DNA purification and M.SssI treatment

6| Prepare genomic DNA for different scales of analysis, and perform enzymatic treatment using CpG DNA methyltransferase M.SssI to protect unmodified CpGs in the genome from bisulfite conversion. Below are protocols for base-resolution mapping of 5fC/5caC using whole genomic DNA (option A: WG-MAB-seq), chromatin-immunoprecipitated DNA (option B: ChIP-MAB-seq) and enzyme-digested genomic DNA (option C: RR-MAB-seq). WG-MAB-seq (option A) provides an unbiased view of 5fC/5caC distribution across the entire genome. ChIP-MAB-seq (option B) using antibody to specific histone marks (e.g., histone 3 lysine 4 monomethylation (H3K4me1) or H3K27me3) can preferentially enrich genomic regions that contain relatively high levels of 5fC/5caC (e.g., poised promoters and active enhancers), and thus it requires less sequencing effort than WG-MAB-seq. Enzymatic-digestion-based RR-MAB-seq (option C) provides an approach to enrich CpG-rich promoters and repeat sequences. In option C, we also describe procedures combining caMAB-seq with the enzymatic-digestion-based strategy for genome-scale 5caC mapping (RR-caMAB-seq).

(A) WG-MAB-seq ● TIMING 3 d

(i) Wash the cell pellet with 1 ml of 1× DPBS, centrifuge it at 1,000 r.p.m. for 5 min at room temperature, and aspirate the supernatant and discard it.

■ PAUSE POINT Snap-freeze the cell pellet in liquid nitrogen; the pellet may be stored in a –80 °C freezer for at least 1 month.

(ii) Extract genomic DNA from frozen cell pellets with the DNeasy blood & tissue kit and elute purified genomic DNA in 100 μl of nuclease-free water according to the manufacturer's instructions.

(iii) Add 2.5 ng (0.25% (wt/wt)) of unmethylated lambda DNA to 1 μg of genomic DNA.

(iv) Set up the first round of the M.SssI treatment reaction as follows. Add each component in the order listed below. Mix them well and incubate the reaction at 37 °C for 4 h.

Component	Volume (μl)	Final concentration
Genomic DNA in nuclease-free water	41.5	
Mg^{2+}-free M.SssI buffer (10×)	5.0	1×
S-adenosylmethionine (32 mM)	1.0	0.64 mM
M.SssI methylase (20 U/μl)	2.0	0.8 U/μl
Total	50	

▲ CRITICAL STEP Use fresh SAM.

? TROUBLESHOOTING

(v) Add an additional 0.5 μl of M.SssI and 1 μl of SAM to the 50-μl reaction. Incubate the mixture at 37 °C for another 4 h.

(vi) Fragment M.SssI-treated genomic DNA (in 50 μl) to an average size of 300–600 bp with Covaris M220 (20% duty factor, 200 cycles per burst, 80 s × 2).

(vii) Purify sheared DNA with solid phase reversible immobilization (SPRI) beads (1.2×). Specifically, vortex the SPRI beads until they are well dispersed. Add 60 µl of SPRI beads to 50 µl of sheared genomic DNA and mix well by repeated pipetting. Incubate the mixture at room temperature for 10 min, and then place the tubes on the magnetic rack until the solution becomes clear (at least 5 min). Remove and discard the supernatant carefully without disturbing the beads. Add 200 µl of freshly prepared 80% (vol/vol) ethanol and wait for 30 s; carefully remove the supernatant. Repeat the 80% (vol/vol) ethanol wash step. Let the tubes stand at room temperature for ~10 min with the lids open until the beads become dry. Resuspend the beads with 60 µl of nuclease-free water, mix well by repeated pipetting and incubate the mixture at room temperature for 2 min. Place the tube on the magnetic rack and wait for at least 5 min until the solution becomes clear. Transfer the supernatant to new tubes carefully to avoid disturbing the beads.
▲ **CRITICAL STEP** The size range of sheared genomic DNA can be examined using the Agilent Bioanalyzer, and it should be 150–800 bp.

(viii) Perform end-repair/dA-tailing of sheared DNA with the NEBNext Ultra DNA library prep kit from Illumina. Set up the end-repair/dA-tailing reaction as follows and perform the following cycle on the thermocycler: first at 20 °C for 30 min, then at 65 °C for 30 min and finally at 4 °C for temporary storage.

Component	Volume (µl)
Purified DNA from Step 6A(vii)	55.5
End-repair reaction buffer (10×)	6.5
End prep enzyme mix	3
Total	65

(ix) Perform adaptor ligation reaction with the NEBNext Ultra DNA Library Prep kit from Illumina. Set up the ligation reaction as follows, and perform the following cycle on the thermocycler: first at 20 °C for 15 min, and then at 4 °C.

Component	Volume (µl)
End-repaired/dA-tailed DNA from Step 6A(viii)	65
Blunt/TA ligase mix	15
Ligation enhancer	1
Methylated adaptors (15 µM)	2.5
Total	83.5

(x) Purify methylated-adaptor-ligated DNA with SPRI beads (1.0×) and elute with 44 µl of nuclease-free water.

(xi) Set up the second round of M.SssI treatment reaction as follows, and add each component in the order of the list below. Mix well and incubate the reaction at 37 °C for 4 h.

Component	Volume (µl)	Final concentration
Genomic DNA in nuclease-free water from Step 6A(x)	41.5	
NEBuffer 2 (10×)	5.0	1×
S-adenosylmethionine (32 mM)	1.0	0.64 mM
M.SssI methylase (20 U/µl)	2.0	0.8 U/µl
Total	50	

(xii) Add an additional 0.5 µl of M.SssI and 1 µl of SAM to the 50-µl reaction. Incubate the mixture at 37 °C for another 4 h.

(xiii) Purify sheared DNA with SPRI beads (1.2×) and elute with 22 µl of nuclease-free water.

(xiv) Perform bisulfite conversion of the purified DNA with the Qiagen EpiTect bisulfite kit. Prepare the bisulfite reactions in 200-µl PCR tubes according to the instructions below, and add each component in the order listed.

Component	Volume (µl)
M.SssI-treated DNA	20
Bisulfite conversion mix	85
DNA protection buffer	35
Total	140

? TROUBLESHOOTING

(xv) Perform the bisulfite conversion reaction on a thermal cycler using the following thermal cycling program.

Step	Temperature (°C)	Time (min)
(1) Denaturation	95	5
(2) Incubation	60	25
(3) Denaturation	95	5
(4) Incubation	60	85
(5) Denaturation	95	5
(6) Incubation	60	175
Repeat steps 1–6 one more time (two cycles of steps 1–6 in total)	—	—
(7) Hold	20	Indefinite

(xvi) Purify bisulfite-converted DNA according to the manufacturer's instructions for the Qiagen EpiTect bisulfite kit, and elute the DNA with 20 µl of EB buffer.

▲ CRITICAL STEP Add carrier RNA provided in the Qiagen EpiTect bisulfite kit to the supplied BL buffer to enhance the recovery rate (3.5 µl of carrier RNA plus 350 µl of BL buffer per sample). To prepare the carrier RNA solution, add 310 µl of RNase-free water to the lyophilized carrier RNA (310 µg per vial) to obtain a 1 µg/µl solution. Dissolve the carrier RNA thoroughly by vortexing.

▲ CRITICAL STEP Elute with prewarmed (50 °C) EB buffer to increase the yield.

? TROUBLESHOOTING

(B) Chromatin immunoprecipitation for genome-scale MAB-seq ● TIMING 5 d

▲ CRITICAL STEP Step 6B(i–xiv) describes the protocol of chromatin immunoprecipitation used for the published H3K4me1-MAB-seq analysis[40]. The protocol may need to be optimized for other antibodies.

(i) Resuspend the cell pellet (~10 million cells from Step 5) in 10 ml of 1× DPBS. Add 1 ml of cross-linking buffer and incubate it for 10 min at room temperature on a rotating platform. Quench the cross-linking reaction with 125 mM glycine for 5 min.

! CAUTION Formaldehyde is harmful if inhaled or absorbed through the skin. Waste should be disposed of according to local regulations.

(ii) After two washes with 10 ml of cold 1× DPBS, centrifuge the cell suspension at 1,000*g* and 4 °C for 5 min.

■ PAUSE POINT Snap-freeze the cell pellet in liquid nitrogen; the pellet may be stored in a –80 °C freezer for at least 1 month.

(iii) Prepare 100 µl of protein G Dynabeads for each ChIP sample. Wash the beads with 1 ml of blocking solution by inverting the tubes several times, and then place the tubes on the magnetic rack to collect the beads. Resuspend Dynabeads in 250 µl of blocking solution and add 10 µg of anti-H3K4me1. Immobilize the antibody with beads for at least 6 h at 4 °C. After incubation, wash the beads with 1 ml of blocking solution twice and resuspend the beads in 50 µl of blocking solution.

(iv) Lyse the cell pellet from Step 6B(ii) with 1 ml of ChIP lysis buffer, and incubate it on ice for 10 min.

▲ CRITICAL STEP Add freshly prepared proteinase inhibitor (50× stock) to ChIP lysis buffer.

(v) Sonicate chromatin using a microtip in a container of ice–water mixture. We use the Branson sonifier 450 with 30 cycles of 10-s ON/30-s OFF pulses at 20% power amplitude. Pellet cell debris at 16,000*g* at 4 °C for 10 min. Transfer the supernatant to a new tube and discard the pellet.

! CAUTION Always wear ear protection when you are operating the sonicator.

▲ CRITICAL STEP The size range of sheared fragments should be 200–1,000 bp in length. We recommend optimizing the sonication conditions before the actual ChIP experiment. It should be noted that different sonifiers and cell types

require separate optimizations. To experimentally estimate the fragment size range, add 150 µl of elution buffer to 50 µl of sheared chromatin and reverse-cross-link the sample by incubating the tube at 65 °C overnight. Purify the DNA (Step 6B(ix–xiv)) and examine the DNA fragment size by standard agarose gel electrophoresis or using Agilent Bioanalyzer.

(vi) Measure chromatin concentration using NanoDrop (concentration (µg/ml) is estimated by (OD_{260} × 50 × dilution factor). Dilute the chromatin to the concentration of 500 µg/ml in ChIP lysis buffer and aliquot 500 µg of chromatin (in 1 ml) for each ChIP sample.

(vii) Add 100 µl of 10% (vol/vol) Triton X-100 to 500 µg chromatin (in 1 ml) immediately before immunoprecipitation. Perform immunoprecipitation overnight at 4 °C by adding 50 µl of antibody-conjugated Dynabeads from Step 6B(iii).
▲ **CRITICAL STEP** Optimal immunoprecipitation conditions need to be separately determined for other antibodies.

(viii) Wash DNA–protein complexes with 1 ml of RIPA buffer five times and with 1 ml of TE–50 mM NaCl once. To wash the sample, first invert the tubes several times and then place the sample tubes on a rocking platform for 2 min. Place tubes on a magnetic rack for 1 min to collect beads each time (until the solution becomes clear). After the final wash, centrifuge it at 1,000*g* for 3 min at 4 °C to pellet the magnetic beads, and discard TE–50 mM NaCl buffer.

(ix) Add 210 µl of elution buffer to the bead pellet and incubate it at 65 °C for 15 min on a Thermomixer (shaking at a speed of 1,000 r.p.m.). After incubation, spin down the beads at 16,000*g* for 1 min at room temperature. Transfer 200 µl of supernatant and reverse-cross-link DNA–protein complexes by incubating the sample at 65 °C overnight on a Thermomixer.

(x) Add 200 µl of TE to each tube of reverse-cross-linked IP sample. Add 8 µl of 10 mg/ml RNaseA (0.2 mg/ml final concentration) and incubate it at 37 °C for 2 h. Add 7 µl of 300 mM $CaCl_2$ and 4 µl of 20 mg/ml proteinase K (0.2 mg/ml final concentration) sequentially and incubate it at 55 °C for 30 min.

(xi) Add 400 µl of phenol:chloroform:isoamyl (25:24:1) alcohol to each tube and mix the sample on a vortex mixer. Prepare one phase lock gel tube for each ChIP sample by spinning the tube at 16,000*g* at room temperature for 30 s. Transfer 800 µl of sample to the phase lock tube. Spin the sample in a centrifuge at 16,000*g* for 5 min at room temperature.
! **CAUTION** Phenol and chloroform are both toxic and should be used only in a fume hood.

(xii) Transfer the aqueous layer (~400 µl) to a new 1.5-ml LoBind microcentrifuge tube. Add 16 µl of 5M NaCl (200 mM final concentration), 1.5 µl of 20 µg/µl glycogen (30 µg total) and 880 µl of EtOH. Mix and cool the mixture for 30 min at –80 °C.

(xiii) Centrifuge the mixture at 16,000*g* for 10 min at 4 °C, discard the supernatant and wash the pellet with 1 ml of cold 70% (vol/vol) EtOH.

(xiv) Dry the pellet at 37 °C for 5–10 min until residual EtOH is completely removed. Thoroughly resuspend each pellet in 30 µl of EB buffer.
■ **PAUSE POINT** The eluate can be stored at –20 °C for at least 1 month.

(vii) Measure the concentration of ChIP DNA with the Qubit dsDNA HS assay kit and add 0.25% (wt/wt) unmethylated lambda DNA to ChIP DNA.
▲ **CRITICAL STEP** The expected yield from the H3K4me1 ChIP experiment described here is 100–200 ng, and thus 0.25–0.5 ng of unmethylated lambda DNA should be added to H3K4me1 ChIP DNA as internal controls for M.SssI treatment.

(xv) End-repair/dA-tailing immunoprecipitated DNA as in Step 6A(viii).

(xvi) Perform methylated adaptor ligation as in Step 6A(ix, x).
▲ **CRITICAL STEP** The concentration of the methylated adaptor should be adjusted according to the approximate concentration of ChIP DNA. For instance, we use tenfold less adaptor for H3K4me1 ChIP DNA in this step (compared with WG-MAB-seq).

(xvii) Perform the first round of M.SssI treatment as in Step 6A(iv, v).

(xviii) Purify sheared DNA with SPRI beads (1.2×) and elute it with 44 µl of nuclease-free water.

(xix) Perform the second round of M.SssI treatment as in Step 6A(xi, xii).

(xx) Purify sheared DNA with SPRI beads (1.2×) and elute it with 22 µl of nuclease-free water.

(xxi) Perform bisulfite conversion of M.SssI-treated, methylated-adaptor-ligated ChIP DNA using the Qiagen Epitect bisulfite kit as in Step 6A(xiv–xvi).

(C) Enzymatic-digestion-based RR-caMAB-seq and RR-MAB-seq ● TIMING 3 d

▲ **CRITICAL STEP** Step 6C(ii–iv) is specifically for caMAB-seq. For the standard MAB-seq method, skip Step 6C(ii–iv) and start from Step 6C(v) using 250 ng of genomic DNA mixed with 0.625 ng of unmethylated lambda DNA (0.25% (wt/wt) spike-in).

(i) Purify genomic DNA from cultured cells as in Step 6A(i, ii).
▲ **CRITICAL STEP** For RR-MAB-seq-based simultaneous mapping of 5fC/5caC, jump to Step 6C(v).

(ii) Add 5 µl of freshly prepared sodium borohydride aqueous solution (1 M) to 250 ng of genomic DNA mixed with 0.625 ng of unmethylated lambda DNA (0.25% (wt/wt) spike-in) in 15 µl of nuclease-free water. Incubate the reaction at room temperature in darkness for 1 h. During the 1-h incubation, briefly spin the reaction tube and open the lid to release the pressure every 15 min.

(iii) Quench the reaction by slowly adding (drop by drop) 10 µl of sodium acetate (0.75 M, pH 5) to the 20-µl reaction from Step 6C(ii). Incubate until no bubbles are observed (~1.5 h).
▲ **CRITICAL STEP** The reaction needs to be quenched completely to prevent any interference with downstream steps. Check every 15 min until no more gas is being released.

(iv) Add 370 µl of nuclease-free water to 30 µl of sample from Step 6C(iii). Purify the DNA by phenol:chloroform:isoamyl (25:24:1) alcohol extraction and ethanol precipitation according to the procedure described in Step 6B(xi–xiii). Dry the pellet at 37 °C for 5–10 min until residual EtOH is completely removed. Thoroughly resuspend each pellet in 45 µl of EB buffer.

(v) Set up a 50-µl M.SssI treatment reaction as follows, and incubate it at 37 °C for 4 h.

Component	Volume (µl)	Final concentration
DNA in nuclease-free water from Step 6C(i) or 6C(iv)	43	
Mg^{2+}-free M.SssI buffer (10×)	5	1×
S-adenosylmethionine (32 mM)	1	0.64 mM
M.SssI methylase (20 U/µl)	1	0.4 U/µl
Total	50	

▲ **CRITICAL STEP** Step 6C(v–xiii) are common steps for RR-MAB-seq and RR-caMAB-seq.
▲ **CRITICAL STEP** Use fresh SAM.

(vi) Add 1 µl of M.SssI and 1 µl of SAM to the 50-µl reaction. Incubate the mixture at 37 °C for another 4 h.

(vii) Heat-inactivate the reaction at 65 °C for 20 min.

(viii) Add 350 µl of nuclease-free water to ~50 µl of sample from Step 6C(vii). Purify the DNA by phenol:chloroform:isoamyl (25:24:1) alcohol extraction and ethanol precipitation according to the procedure described in Step 6C(iv). Dissolve the DNA pellet in 45 µl of nuclease-free water.

(ix) Repeat the procedure described in Step 6C(v–viii) for the second round of M.SssI treatment, except replacing Mg^{2+}-free M.SssI buffer with NEBuffer 2 (10×).

(x) Set up the TaqαI digestion reaction as follows using 1 ng of purified M.SssI-treated DNA diluted in water to a volume of 15.7 µl. Incubate the mixture at 65 °C for 3 h, and then subject to heat inactivation at 80 °C for 20 min.

Component	Volume (µl)	Final amount/concentration
DNA	15.7	1 ng
NEB Cutsmart buffer	1.8	1×
TaqαI (20 U/µl)	0.5	0.556 U/µl
Total	18	

(xi) Set up an end-repair reaction by adding 2 µl of end-prep mix (prepared as follows) to the 18-µl mixture from Step 6C(x). Incubate the mixture at 37 °C for 40 min, and then subject to heat inactivation at 75 °C for 15 min.

End-prep mix (2 µl per reaction)	Volume (µl)	Final concentration
Klenow fragment exo- (5 U/µl)	1.0	0.25 U/µl
NEB Cutsmart buffer	0.2	1×
dNTP mix for RRBS (1 mM dATP, 0.1 mM dCTP, 0.1 mM dGTP)	0.8	40 µM dATP, 4 µM dCTP, 4 µM dGTP (in 20 µl)
Total	2 µl (total reaction volume: 20 µl)	

▲ **CRITICAL STEP** Dilute dNTP mix stock (10 mM dATP, 1 mM dCTP and 1 mM dGTP) ten times in nuclease-free water to a working concentration immediately before use.

(xii) Set up an adaptor-ligation reaction by first adding all components of the ligation mix listed below, except T4 ligase (4 µl), to the 20-µl mixture from Step 6C(xi). Mix the reaction by gently tapping the tube. Then add 1 µl of T4 ligase and immediately mix the reaction by gentle tapping. Incubate the mixture at room temperature for 10 min, at 16 °C for 3 h and at 4 °C overnight, and then subject to heat inactivation at 65 °C for 20 min.

Ligation mix (5 µl per reaction)	Volume (µl)	Final concentration
Nuclease-free water	2.25	
T4 ligase (2,000 U/µl)	1.0	80 U/µl
NEB Cutsmart buffer	0.5	1×
ATP (100 mM)	0.25	1 mM (in 25 µl)
Methylated adaptor (0.75 µM)	1.0	30 nM (in 25 µl)
Total	5 (total reaction volume: 25 µl)	

▲ **CRITICAL STEP** Dilute methylated adaptor stock solution (15 µM) 20 times in nuclease-free water to a working solution (0.75 µM) immediately before use. Discard the remaining working solution.

▲ **CRITICAL STEP** To avoid potential ligation between methylated adaptors (which leads to unwanted adaptor dimers in the final library), we recommend adding the ligation mix in a step-wise manner as described above and mixing the reaction immediately after the T4 ligase is added.

(xiii) Add 10 µl of nuclease-free water to the 25-µl ligation reaction mixture from Step 6C(xii). Perform bisulfite conversion and DNA purification using the Qiagen Epitect bisulfite kit as in Step 6A(xiv–xvi) with minor modifications in reaction composition as below:

Component	Volume (µl)
M.SssI-treated DNA	35
Bisulfite conversion mix	85
DNA protection buffer	20
Total	140

Library preparation and quality control ● TIMING 2–3 d

7| Prepare the library amplification PCR reaction as follows. Add each component in the order listed.

Component	Volume (µl)
Nuclease-free water	3.5
Bisulfite-converted DNA	19.0
NEBNext universal primer (10 µM)	1.25
NEBNext index primer (10 µM)	1.25
KAPA HiFi Uracil+ ReadyMix (2×)	25
Total	50

▲ **CRITICAL STEP** KAPA HiFi Uracil+ ReadyMix must be used in this PCR reaction because this enzyme is insensitive to the presence of uracil in the DNA template (uracil usually causes DNA polymerase to stall). Alternatively, PfuTurbo Cx hotstart DNA polymerase (Agilent), which is also resistant to uracil stalling, can be used in this step.

▲ **CRITICAL STEP** NEBNext Universal and Index primers are from the NEBNext Multiplex Oligos for Illumina kits. Depending on the version of the kit, the primer concentration may be either 25 µM (older version) or 10 µM (lastest version). Our PCR reaction composition is calculated using the latest version of the kits. Adjust the composition accordingly if 25 µM primers are to be used.

? TROUBLESHOOTING

8| Perform library preparative PCR using the thermal cycling program below.
▲ **CRITICAL STEP** The total number of PCR cycles needs to be adjusted according to the initial DNA amount. WG-MAB-seq (from Step 6A) typically needs 5–6 cycles of PCR amplification; ChIP-MAB-seq (from Step 6B) usually requires 10–12 cycles; and RR-MAB-seq (from Step 6C) generally requires 15–17 cycles.

Step	Temperature (°C)	Time
(1) Denature	95	2 min
(2) Denature	98	20 s
(3) Anneal	63	30 s
(4) Extend	72	1 min
Repeat steps 2–4 for 5–17 cycles	—	—
(5) Extend	72	5 min
(6) Hold	4	Indefinite

? TROUBLESHOOTING

9| Purify library preparative PCR products with SPRI beads using option A for DNA from WG-MAB-Seq (Step 6A) and ChIP-MAB-Seq (Step 6B) or using option B for DNA from RR-MAB-Seq or RR-caMAB-seq (Step 6C).

(A) WG-MAB-seq or ChIP-MAB-seq library

(i) Purify the amplified library with 60 µl of SPRI beads (1.2×) according to the procedure described in Step 6A(vii) and elute with 15 µl of nuclease-free water.

(B) RR-MAB-seq or RR-caMAB-seq library

(i) Perform size selection of the amplified library with SPRI beads using a double size selection strategy. First, add 30 µl (0.60×) of SPRI beads to 50-µl PCR reactions. Incubate the mixture at room temperature for 10 min, and then place the tubes on the magnetic rack for 10 min (until the solution becomes clear). Transfer the supernatant to a new tube and discard the beads. Next, add 40 µl (1.4×) of SPRI beads to 80 µl of supernatant and place the tubes on the magnetic rack for 10 min. Discard the supernatant and wash the beads with 200 µl of freshly prepared 80% (vol/vol) ethanol twice. Let the beads air-dry at room temperature for 10 min. Elute size-selected final libraries with 15 µl of nuclease-free water.

10| Quantify the concentration of the final libraries (from Step 9) with a Qubit fluorometer and the dsDNA HS assay kit. Assess the size distribution of the final libraries with Agilent Bioanalyzer 2100 and high-sensitivity DNA kit (**Fig. 4**). Calculate the molar concentration (nM) of the final libraries on the basis of the average size (in base pairs, measured by Bioanalyzer) and library concentration (in ng/µl, measured by Qubit) using the following formula:

$$\text{Sample concentration (nM)} = (\text{library concentration (ng/}\mu\text{L)}/1000)/(\text{Average fragment size (bp)} \times 660) \times 10^9$$

? TROUBLESHOOTING

11| Sequence the libraries using Illumina HiSeq 2000/2500 or NextSeq 500 sequencers with a final concentration at ~80% of that used in regular sequencing experiments.

Data analysis of MAB-seq ● TIMING 2 d

12| Analyze data using option A for WG-MAB-seq or ChIP-MAB-seq, or using option B for RR-MAB-seq or RR-caMAB-seq

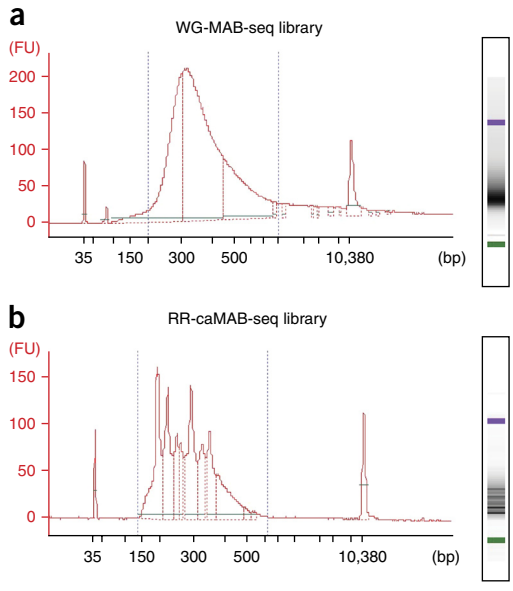

Figure 4 | Bioanalyzer electropherogram of sequencing libraries of MAB-seq/caMAB-seq. (**a**) Representative example of final library size distribution obtained from WG-MAB-seq experiments. (**b**) Representative example of final library size distribution obtained from TaqαI-digested RR-caMAB-seq experiments. Sharp peaks are from digestion of repetitive elements. FU, fluorescence units.

(A) WG-MAB-seq and ChIP-MAB-seq

(i) Trim low-quality bases and contaminating adaptor sequences using the Trim Galore program. Example command:

```
trim_galore --fastqc --gzip --length 36 <raw.fastq.gz>
```

(B) RR-MAB-seq or RR-caMAB-seq

(i) Trim raw sequencing reads for Illumina adaptor sequences, low-quality bases and experimentally introduced 3′ CG using Trim Galore. Example command:

```
trim_galore --fastqc --gzip --three_prime_clip_R1 2 --length 36
<raw.fastq.gz>
```

▲ **CRITICAL STEP** During the end repair following TaqαI digestion, an unmodified CG is experimentally introduced at the 3′ end of the fragments, and it will be sequenced as TG after bisulfite conversion. To ensure that this artificial CG does not interfere with the analysis, after adaptors and low-quality bases are trimmed, an additional two bases at the 3′ end are trimmed from all reads.

▲ **CRITICAL STEP** In this case, we keep only reads that are no less than 36 bp in length after trimming. This cutoff can be adjusted for sequencing libraries with different read lengths.

13| Align trimmed reads to the bisulfite-converted reference genome using the Bismark program. The reference genome sequences (e.g., mm9.fasta plus lambda.fasta) need to be first bisulfite-converted and indexed *in silico*. Example command:

```
/path-to-bismark/bismark_genome_preparation --bowtie2 –path_to_bowtie
<bowtie2 folder> --verbose <genome folder>
```

Once the bisulfite-converted reference genome index for Bismark/Bowtie2 has been generated, use the following command line to map trimmed reads:

```
/path-to-bismark/bismark --bowtie2 --path_to_bowtie <bowtie2_folder> -
-samtools_path <samtools folder> --fastq -p 8 --gzip --bam <genome
folder> <trimmed.fastq.gz>
```

This command will generate a BAM file containing all the alignment output.

14| Use SortSam.jar and MarkDuplicate.jar in the Picard toolkit to sort the mapped reads, and remove PCR duplicates, respectively. Example command:

```
java -Xmx32g -jar /path-to-picard-tools-1.91/SortSam.jar \
INPUT=<bismark-aligned.bam> \
OUTPUT=<sorted.bam> \
SORT_ORDER=coordinate CREATE_INDEX=true

java -Xmx32g -jar /path-to-picard-tools-1.91/MarkDuplicates.jar
ASSUME_SORTED=true REMOVE_DUPLICATES=true \
INPUT=<sorted.bam> \
OUTPUT=<deduplicated.bam> \
METRICS_FILE=MarkDuplicates.metrics.txt \
VALIDATION_STRINGENCY=SILENT CREATE_INDEX=true
```

These commands will generate a sorted BAM file containing only monoclonal reads.

▲ **CRITICAL STEP** Because restriction digestion instead of random shearing is used for fragmenting genomic DNA in RRBS-based experiments, skip this step for RR-MAB-seq or RR-caMAB-seq analysis.

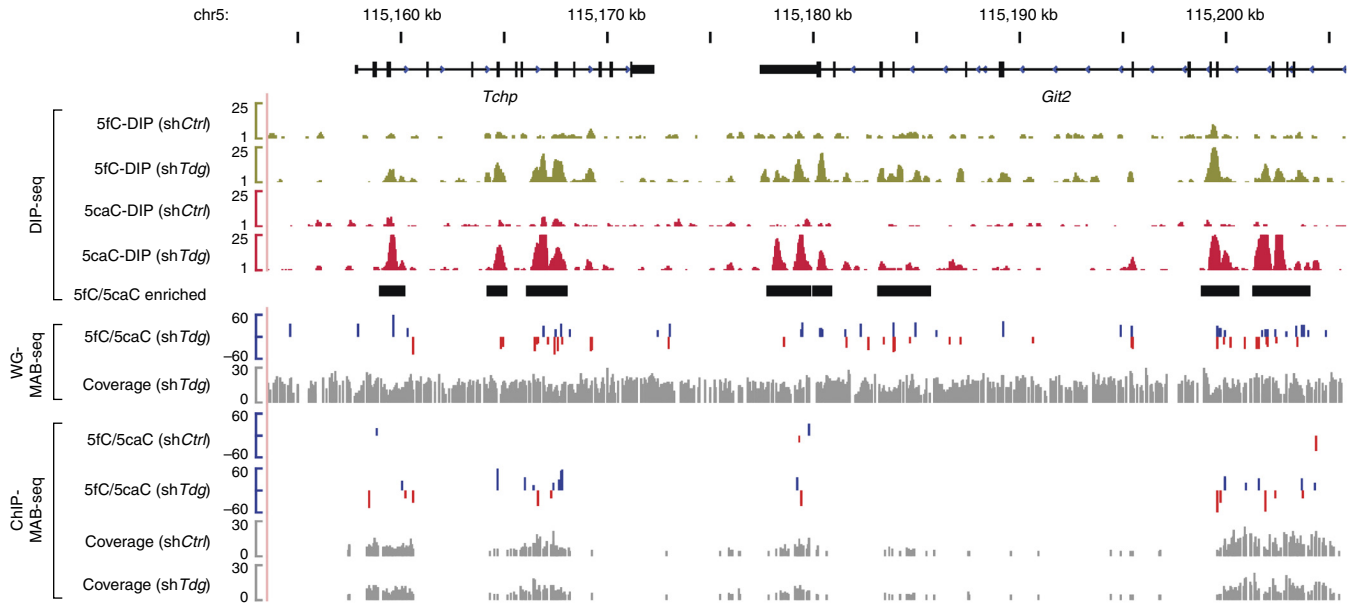

Figure 5 | Base-resolution 5fC/5caC maps and affinity-enrichment-based 5fC/5caC maps at the *Tchp–Git2* locus in mouse ESCs. For WG-MAB-seq and ChIP-MAB-seq data sets, positive values (blue) indicate CpGs on the Watson strand, whereas negative values (red) indicate CpGs on the Crick strand (vertical axis limits are –60% to +60%). Only CpGs sequenced to a depth of ≥5 and associated with a statistically significant level of 5fC/5caC (FDR = 5%) are shown. Sequencing coverage for WG-MAB-seq and ChIP-MAB-seq is shown in gray in separate tracks (vertical axis limits are 0–30 reads). Highlighted by black horizontal bars are 5fC/5caC-enriched genomic regions identified by DNA immunoprecipitation followed by sequencing (DIP-seq). The vertical axis limits for DIP-seq data sets are 1–25 (in rp10m: reads per 10 million reads). sh*Ctrl*, control knockdown mouse ESCs; sh*Tdg*, *Tdg*-knockdown mouse ESCs cells. Adapted with permission from ref. 40.

15| Use the Bismark methylation extractor program to extract 5fC/5caC modification information for each CpG site. For MAB-seq, the 5fC + 5caC level at a CpG site equals $N_T/(N_T + N_C)$. For caMAB-seq, the 5caC level at a CpG site equals $N_T/(N_T + N_C)$. Example command:

```
/path-to-bismark/bismark_methylation_extractor --single-end --gzip --

bedGraph --counts --genome_folder <genome sequence folder> --output <output

folder> <deduplicated.bam>
```

This command will generate a compressed bedgraph file, <bismark.cov.gz>, which can be used to extract raw MAB-seq signals and sequencing coverage information for each CpG site in the genome. This bedgraph file contains six columns: <chromosome>, <start position>, <end position>, <percentage of C>, <number of C> and <number of T>.
▲ **CRITICAL STEP** Because the samples are spiked in with unmethylated lambda DNA, the M.SssI methylase efficiency can be estimated from the methylation level within CpG dinucleotides of lambda DNA.
▲ **CRITICAL STEP** For downstream analysis, we recommend using CpG sites with at least 5× coverage, because some low-coverage CpG sites arise from wrong alignment or other unwanted situations.

16| Visualize raw MAB-seq signals in the IGV genome browser. Example command:

```
zcat <bismark.cov.gz> | awk 'BEGIN{OFS="\t"} {coverage=$5+$6; if

(coverage >= 5) print $1, $2, $3,100-$4}' | gzip > <bismark.bedgraph.gz>

igvtools tile <bismark.bedgraph.gz> <bismark.bedgraph.tdf> mm9
```

This command will generate a .tdf file for visualizing raw MAB-seq signals in the IGV browser (**Fig. 5**)[40].

17| Mouse strain or human tissue/cell-line-specific single-nucleotide polymorphisms (SNPs; e.g., C-to-T mutations) overlapping with CpG dinucleotides in the reference genome can affect the calculation of MAB-seq signals ($N_T/(N_T + N_C)$) by introducing false positive or false negative signals. Known SNPs can be used to exclude the affected CpG sites from

downstream analysis. Obtain SNP information from Sanger Institute (https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1303)[65] and create a bed file summarizing all the genomic CpG sites containing SNPs (at C, G or both). Exclude these SNP-containing CpG sites using the intersectBed command from the BEDTools. Example command:

```
/path-to-bedtools/intersectBed -a <Unfiltered bed file> -b <SNP-
containing CpG bed file> -v > <Filtered bed file>
```

▲ **CRITICAL STEP** If WG-MAB-seq or ChIP-MAB-seq data sets are of sufficient sequencing depth (e.g., at least ten reads per strand), *de novo* SNPs overlapping with annotated CpG sites in the reference genome can be identified and removed with the BisSNP program. In the case of RR-MAB-seq and RR-caMAB-seq, only a small proportion of the CpG sites are covered in both positive and negative strands. Thus, *de novo* SNP detection, which relies on sequencing information obtained from both strands, will not remove all the SNPs.

▲ **CRITICAL STEP** In addition to SNP filtering, we recommend removing CpG sites with unusually high levels (for instance, CpGs in 80% reads are sequenced as TpGs) of raw MAB-seq signals ($N_T/(N_T + N_C)$) in both the negative control (e.g., *Dnmt1/3a/3b*-knockout mouse ESCs) and the sample of interest. It is possible that these annotated CpG sites present in the reference genome could be mutated (e.g., C-to-T) in the sample.

18| To identify CpG sites modified by a significant level of 5fC/5caC, we use the binomial distribution (N as the sequencing coverage ($N_T + N_C$) and p as the probability of detecting false positive signals (the sum of the M.SssI methylase failure rate and the deamination rate of 5mC)) to assess the probability of observing N_T or greater by chance. Example R command:

```
apply(df, 1, function(x) diff(pbinom(c(x[1],x[2]), size=x[2], prob=p)))
```

df: a data frame (d*f*) containing two columns: the first one (x[1]) is the number of T reads and the second one (x[2]) is the total number of reads (C + T)).
p: the probability of detecting false positive signals (we used 2.04% in our published study[40]).

To estimate the empirical FDR of calling 5fC/5caC-modified CpGs, calculate the above statistics at each of the common CpGs in negative control samples (*Dnmt1/3a/3b*-knockout or *Tet1/2/3*-knockout cells) in which true 5fC/5caC signals are absent. The empirical FDR for a given P value cutoff is the number of called CpGs in negative controls divided by the number detected in the sample of interest. For RR-MAB-seq and RR-caMAB-seq analysis, 5fC/5caC-modified sites identified using the binomial-distribution-based P value cutoff approach can be further filtered by a numeric cutoff (for example, $N_T/(N_T + N_C)$ should be at least 10%). In our hands, applying this additional numeric filter to RR-MAB-seq and RR-caMAB-seq data sets allows comparatively low FDR without compromising the detection sensitivity. For data sets with low sequencing depth or from cell types with relatively low levels of 5fC/5caC, analyzing genomic bins (divide genome into 100-bp bins) instead of single nucleotides may identify 5fC/5caC-marked regions with increased accuracy and sensitivity (**Box 4**). This is probably because true 5fC/5caC signals tend to cluster together, whereas false positive signals tend to disperse randomly in the genome (at least in mouse ESCs). Identification of 5fC/5caC-enriched genomic regions via a binning strategy is conceptually similar to the identification of differentially methylated regions in WGBS or RRBS analysis. Therefore, other tests (e.g., Fisher exact test) can be used as alternatives to the binomial-distribution-based test.

? TROUBLESHOOTING

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

TABLE 2 | Troubleshooting table

Step	Problem	Possible reason	Solution
6A(iv)	Low M.SssI methylation efficiency	Poor-quality reagents (e.g., expired SAM)	Use fresh SAM and make SAM aliquots stored at −20 °C to avoid excessive rounds of freezing/thawing
		Too much DNA is used	Methyl transfer reaction generates SAH, a potent inhibitor of M.SssI. Therefore, reducing the amount of input DNA will help to reduce the generation of SAH and increase M.SssI efficiency. In addition, when input DNA is sufficient, perform multiple rounds of M.SssI treatment (with DNA purification through phenol:chloroform:isoamyl (25:24:1) alcohol extraction and ethanol precipitation in between, following the procedure described in Step 6B(xi–xiii))

(continued)

TABLE 2 | Troubleshooting table (continued)

Step	Problem	Possible reason	Solution
6A(xiv,xvi)	Incomplete/ excessive bisulfite conversion	Poor-quality reagents	If the Epitect bisulfite kit is used, the dissolved bisulfite mix can be stored at –20 °C and used within a month; however, avoid freezing/thawing more than once. DNA protection buffer and buffer BD should be stored at 4 °C and, ideally, replaced every 6 months
		Nonoptimal protocols when using other bisulfite conversion kits	When using kits other than the Qiagen Epitect bisulfite kit, test the protocol first using synthetic oligos and lambda DNA to make sure that 5fC is efficiently converted and that 5mC and 5hmC are not wrongly deaminated
	Poor yield of DNA after bisulfite conversion	DNA degradation during bisulfite conversion	Purify the DNA shortly (within 6 h) after the thermal program is completed to avoid any potential DNA degradation
		DNA loss during purification	Use fresh DNA protection buffer (green color and less than 6 months old)
			Carrier RNA should be added freshly to buffer BL before use to facilitate DNA recovery, and dissolved carrier RNA should be stored at –20 °C. Use EB warmed to 50 °C to facilitate elution from the column.
7 and 8	Low PCR amplification efficiency of the library	Poor DNA yield/ quality before PCR amplification	Use LoBind tubes when dealing with small amounts of DNA, and avoid unnecessary pipetting. Use glycogen to facilitate DNA precipitation during PCI purification. Once you start, try to avoid pausing for a long time (<12 h) between steps
		Poor quality of PCR reagents	Make aliquots of PCR reagents and avoid excessive cycles of freezing/thawing
10	Abnormal size/shape of library after PCR amplification	Insufficient/excessive sonication of DNA (for WG-MAB-seq and ChIP-MAB-seq)	Adjust sonication conditions correspondingly for different amounts of input DNA/different equipment/different enrichment methods
		Poor DNA yield/quality (for RR-MAB-seq)	For RR-MAB-seq library preparation, avoid pausing for a long time (<6 h) once end preparation is started

● TIMING

Steps 1–5, cell culture and sample harvest collection: 4–5 d
Step 6A, DNA purification and enzymatic treatment for WG-MAB-seq: 3 d
Step 6B, DNA purification and enzymatic treatment for ChIP-MAB-seq: 5 d
Step 6C, DNA purification and enzymatic treatment for RR-MAB-seq: 3 d
Steps 7–10, library preparation PCR and purification of final library: 1 d
Steps 10 and 11, quality control and high-throughput DNA sequencing: 2 d
Steps 12–18, data analysis for MAB-seq data: 2 d
Box 1, quality control of MAB-seq and caMAB-seq experiment: 1 d
Box 2, oxidized methylcytosine synthetic oligo spike-in experiment: 2 d
Box 3, locus-specific MAB-seq and caMAB-seq experiment: 3 d
Box 4, identification of 5fC/5caC-enriched genomic intervals: 1 d

ANTICIPATED RESULTS

Sequencing library preparation

For ChIP-MAB-seq and WG-MAB-seq, the expected yield after SPRI bead purification is 2–5 ng/μl in 15 μl, if the PCR conditions are optimized. For RR-MAB-seq and RR-caMAB-seq, the expected yield after SPRIselect purification is 3–5 ng/μl in 15 μl. The expected size range of an ideal library is 200–500 bp for ChIP-MAB-seq and for WG-MAB-seq, with an average size of 350–400 bp (**Fig. 4a**). For RR-MAB-seq and RR-caMAB-seq based on TaqαI digestion, the expected size range of an ideal library is 150–650 bp (this range can be different if other enzymes are used), with an average size of 250–350 bp (**Fig. 4b**).

Sequencing coverage for different genome-scale MAB-seq experiments

For WG-MAB-seq analysis of mouse genome, we have sequenced ~300–400 million reads (100–150 bp single-end) per library on HiSeq 2500 or NextSeq 500. After adaptor sequences and low-quality bases have been trimmed, roughly 80% of reads can

be uniquely aligned to the mouse genome. When two biologically independent sequencing libraries are combined (~500 million uniquely mapped and monoclonal reads), >95% of CpG dyads (out of a total of ~21 million CpG dyads) in the mouse genome can be covered with an average of 28× coverage per CpG dyad (combine both strands). Under this sequencing coverage, ~34 million CpGs (~80% of a total of ~42 million CpGs) can be covered by at least five times per strand. For ChIP-MAB-seq, we typically sequence ~100 million reads (100 bp single-end) per sample. Approximately 80% of reads can be uniquely mapped to the mouse genome after quality filtering. Under this sequencing depth, 3–4 million CpG dyads can be covered by at least five times per CpG dyad. For RR-MAB-seq or RR-caMAB-seq analysis, we usually sequence ~30 million reads (100 bp single-end) per library. After adaptor trimming and filtering of low-quality or short reads, ~45% of the reads can be mapped uniquely to the reference genome using the parameters specified in the data analysis part, giving rise to ~12 million uniquely mapped reads. Under this sequencing depth, ~1 million CpG sites are covered by at least five times per CpG (on each strand).

Statistical calling of 5fC/5caC-modified CpGs

In our published study[40], WG-MAB-seq analysis of *Tdg*-depleted and VC-treated (60 h) mouse ESCs identified 675,325 5fC/5caC-modified CpGs (2.7% out of 24,872,637 CpGs with coverage of ≥10, using a stringent *P* value cutoff leading to an FDR <5%). With a similar FDR cutoff (<5%) and sequencing depth filtering (≥10 per CpG dyads), H3K4me1-MAB-seq analysis identified 127,576 (7.6% out of 1,670,036 CpG dyads with $N \geq 10$) 5fC/5caC-modified CpG dyads in VC-treated, *Tdg*-depleted mouse ESCs (sh*Tdg* + VC). The higher percentage of 5fC/5caC-modified CpG identified in H3K4me1-marked genomic domains is expected, as 5fC and 5caC are preferentially enriched in these gene regulatory regions. Choosing the *P* value cutoff partly depends on the purpose of the downstream analyses. For a biological sample with much less 5fC/5caC (e.g., wild-type mouse ESCs), a much higher sequencing depth is required to distinguish true 5fC/5caC signals from background. To partially circumvent the requirement of ultra-high sequencing depth, one can analyze MAB-seq data sets by combining individual CpG sites into genomic bins (e.g., 100-bp intervals) and by using statistical calling with a less stringent FDR cutoff, which may provide a lower-resolution map of 5fC/5caC-modified CpGs in cell types with lower levels of 5fC/5caC. Similarly, RR-MAB-seq data sets tend to generate a higher FDR at a given *P* value for the following two reasons. First, PCR duplicates cannot be removed when using a standard RRBS strategy, which increases the likelihood of overly amplifying random background signals. Second, RRBS preferentially covers GC-rich regions that are generally depleted of 5fC/5caC (at least in mouse ESCs). These potential problems can be partly solved by adopting a unique molecule identifier (UMI)-based RRBS approach[66] to remove PCR duplicates, and by using other restriction enzymes that cover more 5fC/5caC-modified regions.

ACKNOWLEDGMENTS H.W. was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research and is currently supported by the National Human Genome Research Institute (R00HG007982). X.W. was supported by the China Scholarship Council. Y.Z. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS H.W. and X.W. performed experiments and carried out data analysis. H.W., X.W. and Y.Z. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
2. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
3. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68 (2014).
4. Pastor, W.A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
5. Kohli, R.M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).
6. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
7. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
8. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
9. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
10. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
11. 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).
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. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393 (2011).
14. Pastor, W.A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
15. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
16. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
17. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**, 451–464 (2011).
18. Wu, H. & Zhang, Y. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. *Cell Cycle* **10** 2428–2436 (2011).
19. Dawlaty, M.M. *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310–323 (2013).
20. Kang, J. *et al.* Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc. Natl. Acad. Sci. USA* **112**, E4236–4245 (2015).

21. Gu, T.P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477** 610 (2011).
22. 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).
23. 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).
24. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
25. Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460–464 (2013).
26. Zhang, R.R. *et al.* Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* **13**, 237–245 (2013).
27. Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109–1122 (2013).
28. Kaas, G.A. *et al.* TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* **79**, 1086–1093 (2013).
29. Cimmino, L., Abdel-Wahab, O., Levine, R.L. & Aifantis, I. TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell* **9**, 193–204 (2011).
30. Ko, M. *et al.* TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol. Rev.* **263**, 6–21 (2015).
31. Song, C.X. & He, C. Potential functional roles of DNA demethylation intermediates. *Trends Biochem. Sci.* **38**, 480–484 (2013).
32. Wu, H. & Zhang, Y. Charting oxidized methylcytosines at base resolution. *Nat. Struct. Mol. Biol.* **22**, 656–661 (2015).
33. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
34. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
35. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).
36. Kellinger, M.W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
37. Wang, L. *et al.* Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* **523**, 621–625 (2015).
38. Raiber, E.A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).
39. Szulik, M.W. *et al.* Differential stabilities and sequence-dependent base pair opening dynamics of Watson-Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochemistry* **54**, 1294–1305 (2015).
40. Wu, H., Wu, X., Shen, L. & Zhang, Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **32**, 1231–1240 (2014).
41. Song, C.-X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
42. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013).
43. Neri, F. *et al.* Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.* **10**, 674–683 (2015).
44. Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**, 447–458 (2014).
45. Lister, R. *et al.* Global epigenomic reconfiguration during Mammalian brain development. *Science* **341**, 1237905 (2013).
46. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
47. Hu, L. *et al.* Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545–1555 (2013).
48. Hashimoto, H. *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395 (2014).
49. Sun, Z. *et al.* A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Mol. Cell* **57**, 750–761 (2015).
50. Xia, B. *et al.* Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Methods* **12**, 1047–1050 (2015).
51. Booth, M.J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **6**, 435–440 (2014).
52. Lu, X. *et al.* Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Res.* **25**, 386–389 (2015).
53. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
54. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* **28**, 2103–2119 (2014).
55. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
56. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
57. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
58. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
60. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
61. Liu, Y., Siegmund, K.D., Laird, P.W. & Berman, B.P. Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol.* **13**, R61 (2012).
62. Sotiriou, S. *et al.* Ascorbic-acid transporter Slc23a1 is essential for vitamin C transport into the brain and for perinatal survival. *Nat. Med.* **8**, 514–517 (2002).
63. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).
64. Yin, R. *et al.* Ascorbic Acid Enhances Tet-Mediated 5-Methylcytosine Oxidation and Promotes DNA Demethylation in Mammals. *J. Am. Chem. Soc.* **135**, 10396–10403 (2013).
65. Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
66. Wang, K. *et al.* Q-RRBS: a quantitative reduced representation bisulfite sequencing method for single-cell methylome analyses. *Epigenetics* **10**, 775–783 (2015).
67. Booth, M.J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).