3. W. Kim, S. Kook, D. J. Kim, C. Teodorof, W. K. Song, *J. Biol. Chem.* **279**, 8333 (2004).
4. V. Giambra et al., *Mol. Cell. Biol.* **28**, 6123 (2008).
5. F. E. Garrett et al., *Mol. Cell. Biol.* **25**, 1511 (2005).
6. W. A. Dunnick et al., *J. Exp. Med.* **206**, 2613 (2009).
7. M. Cogné et al., *Cell* **77**, 737 (1994).
8. J. P. Manis et al., *J. Exp. Med.* **188**, 1421 (1998).
9. A. G. Bébin et al., *J. Immunol.* **184**, 3710 (2010).
10. E. Pinaud et al., *Immunity* **15**, 187 (2001).
11. C. Vincent-Fabert et al., *Blood* **116**, 1895 (2010).
12. R. Wuerffel et al., *Immunity* **27**, 711 (2007).
13. Z. Ju et al., *J. Biol. Chem.* **282**, 35169 (2007).
14. H. Duan, H. Xiang, L. Ma, L. M. Boxer, *Oncogene* **27**, 6720 (2008).
15. M. Gostissa et al., *Nature* **462**, 803 (2009).
16. C. Chauveau, M. Cogné, *Nat. Genet.* **14**, 15 (1996).
17. C. Chauveau, E. Pinaud, M. Cogne, *Eur. J. Immunol.* **28**, 3048 (1998).
18. M. A. Sepulveda, F. E. Garrett, A. Price-Whelan, B. K. Birshtein, *Mol. Immunol.* **42**, 605 (2005).
19. E. Pinaud, C. Aupetit, C. Chauveau, M. Cogné, *Eur. J. Immunol.* **27**, 2981 (1997).
20. A. A. Khamlichi et al., *Blood* **103**, 3828 (2004).
21. R. Shinkura et al., *Nat. Immunol.* **4**, 435 (2003).
22. A. Yamane et al., *Nat. Immunol.* **12**, 62 (2011).
23. M. Liu et al., *Nature* **451**, 841 (2008).
24. J. Stavnezer, J. E. Guikema, C. E. Schrader, *Annu. Rev. Immunol.* **26**, 261 (2008).
25. S. Duchez et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3064 (2010).
26. T. K. Kim et al., *Nature* **465**, 182 (2010).

Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution

Michael J. Booth,[1]* Miguel R. Branco,[2,3]* Gabriella Ficz,[2] David Oxley,[4] Felix Krueger,[5] Wolf Reik,[2,3]† Shankar Balasubramanian[1,6,7]†

5-Methylcytosine can be converted to 5-hydroxymethylcytosine (5hmC) in mammalian DNA by the ten-eleven translocation (TET) enzymes. We introduce oxidative bisulfite sequencing (oxBS-Seq), the first method for quantitative mapping of 5hmC in genomic DNA at single-nucleotide resolution. Selective chemical oxidation of 5hmC to 5-formylcytosine (5fC) enables bisulfite conversion of 5fC to uracil. We demonstrate the utility of oxBS-Seq to map and quantify 5hmC at CpG islands (CGIs) in mouse embryonic stem (ES) cells and identify 800 5hmC-containing CGIs that have on average 3.3% hydroxymethylation. High levels of 5hmC were found in CGIs associated with transcriptional regulators and in long interspersed nuclear elements, suggesting that these regions might undergo epigenetic reprogramming in ES cells. Our results open new questions on 5hmC dynamics and sequence-specific targeting by TETs.

5-Methylcytosine (5mC) is an epigenetic DNA mark that plays important roles in gene silencing and genome stability and is found enriched at CpG dinucleotides (*1*). In metazoa, 5mC can be oxidized to 5-hydroxymethylcytosine (5hmC) by the ten-eleven translocation (TET) enzyme family (*2*, *3*). 5hmC may be an intermediate in active DNA demethylation but could also constitute an epigenetic mark per se (*4*). Levels of 5hmC in genomic DNA can be quantified with analytical methods (*2*, *5*, *6*) and mapped through the enrichment of 5hmC-containing DNA fragments that are then sequenced (*7*–*13*). Such approaches have relatively poor resolution and give only relative quantitative information. Single-nucleotide sequencing of 5mC has been performed by using bisulfite sequencing (BS-Seq), but this method cannot discriminate 5mC from 5hmC (*14*, *15*). Single-molecule real-time sequencing (SMRT) can detect derivatized 5hmC in genomic DNA (*16*). However, enrichment of 5hmC-containing DNA fragments is required, which causes loss of quantitative information (*16*). Furthermore, SMRT has a relatively high rate of sequencing errors (*17*), and the peak calling of modifications is imprecise (*16*). Protein and solid-state nanopores can resolve 5mC from 5hmC and have the potential to sequence unamplified DNA (*18*, *19*).

We observed the decarbonylation and deamination of 5-formylcytosine (5fC) to uracil (U) under bisulfite conditions that would leave 5mC unchanged (Fig. 1A and supplementary text). Thus, 5hmC sequencing would be possible if 5hmC could be selectively oxidized to 5fC and then converted to U in a two-step procedure (Fig. 1B). Whereas BS-Seq leads to both 5mC and 5hmC being detected as Cs, this "oxidative bisulfite" sequencing (oxBS-Seq) approach would yield Cs only at 5mC sites and therefore allow us to determine the amount of 5hmC at a particular nucleotide position by subtraction of this readout from a BS-Seq one (Fig. 1C).

Specific oxidation of 5hmC to 5fC (table S1) was achieved with potassium perruthenate ($KRuO_4$). In our reactivity studies on a synthetic 15-nucleotide oligomer single-stranded DNA (ssDNA) containing 5hmC, we established conditions under which $KRuO_4$ reacted specifically with the primary alcohol of 5hmC (Fig. 2A). Fifteen-nucleotide oligomer ssDNA that contained C or 5mC did not show any base-specific reactions with $KRuO_4$ (fig. S1, A and B). For 5hmC in DNA, we only observed the aldehyde (5fC) and not the carboxylic acid (*20*), even with a moderate excess of oxidant. The $KRuO_4$ oxidation can oxidize 5hmC in samples presented as double-stranded DNA (dsDNA), with an initial denaturing step before addition of the oxidant; this results in a quantitative conversion of 5hmC to 5fC (Fig. 2B).

To test the efficiency and selectivity of the oxidative bisulfite method, three synthetic dsDNAs containing either C, 5mC, or 5hmC were each oxidized with $KRuO_4$ and then subjected to a conventional bisulfite conversion protocol. Sanger sequencing revealed that 5mC residues did not convert to U, whereas both C and 5hmC residues did convert to U (fig. S2). Because Sanger sequencing is not quantitative, to gain a more accurate measure of the efficiency of transforming 5hmC to U, Illumina (San Diego, California) sequencing was carried out on the synthetic DNA containing 5hmC (122-nucleotide oligomer) after oxidative bisulfite treatment. An overall 5hmC-to-U conversion level of 94.5% was observed (Fig. 2C and fig. S14). The oxidative bisulfite protocol was also applied to a synthetic dsDNA that contained multiple 5hmC residues (135-nucleotide oligomer) in a range of different contexts that showed a similarly high conversion efficiency (94.7%) of 5hmC to U (Fig. 2C and fig. S14). Last, the $KRuO_4$ oxidation was carried out on genomic DNA and showed through mass spectrometry a quantitative conversion of 5hmC to

[1]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. [2]Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK. [3]Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK. [4]Proteomics Research Group, Babraham Institute, Cambridge CB22 3AT, UK. [5]Bioinformatics Group, Babraham Institute, Cambridge CB22 3AT, UK. [6]School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK. [7]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: wolf.reik@babraham.ac.uk (W.R.); sb10031@cam.ac.uk (S.B.)
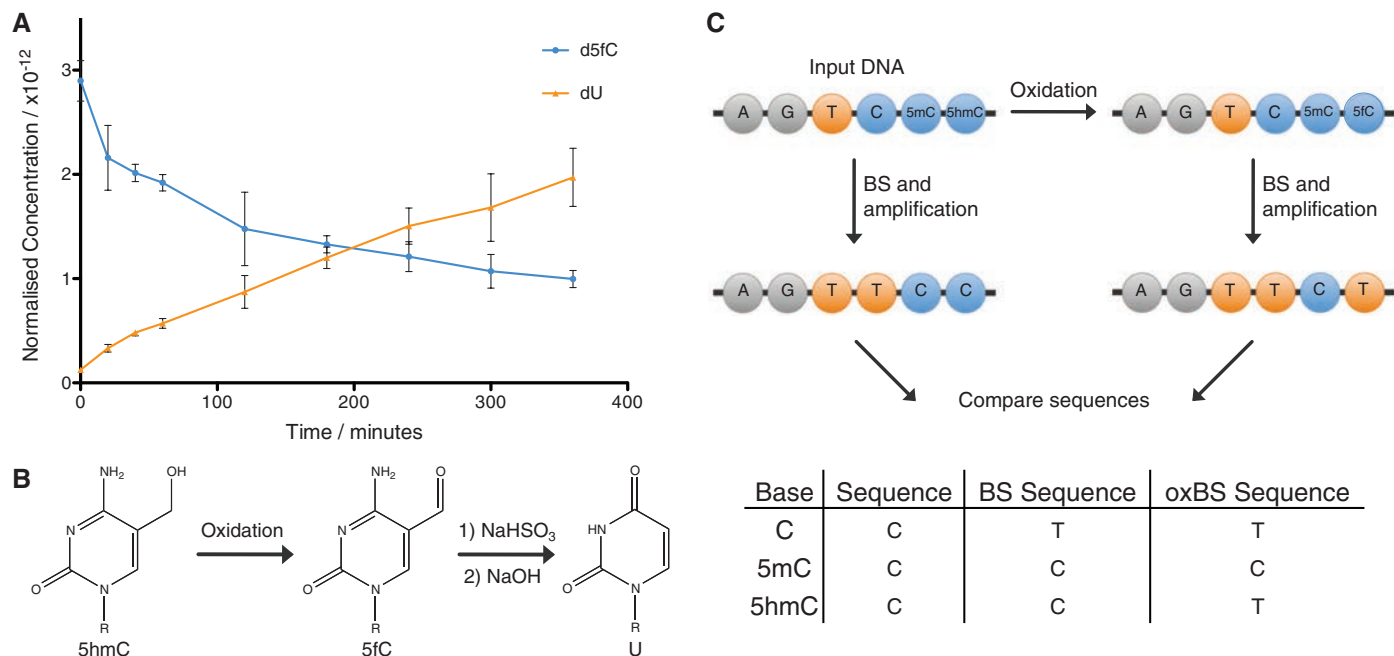
Fig. 1. A method for single-base resolution sequencing of 5hmC. (**A**) Reaction of 2′-deoxy-5-formylcytidine (d5fC) with NaHSO$_3$ (bisulfite) quenched by NaOH at different time points and then analyzed with high-performance liquid chromatography (HPLC). Data are mean ± SD of three replicates. (**B**) Oxidative bisulfite reaction scheme: oxidation of 5hmC to 5fC followed by bisulfite treatment and NaOH to convert 5fC to U. The R group is DNA. (**C**) Diagram and table outlining the BS-Seq and oxBS-Seq techniques.
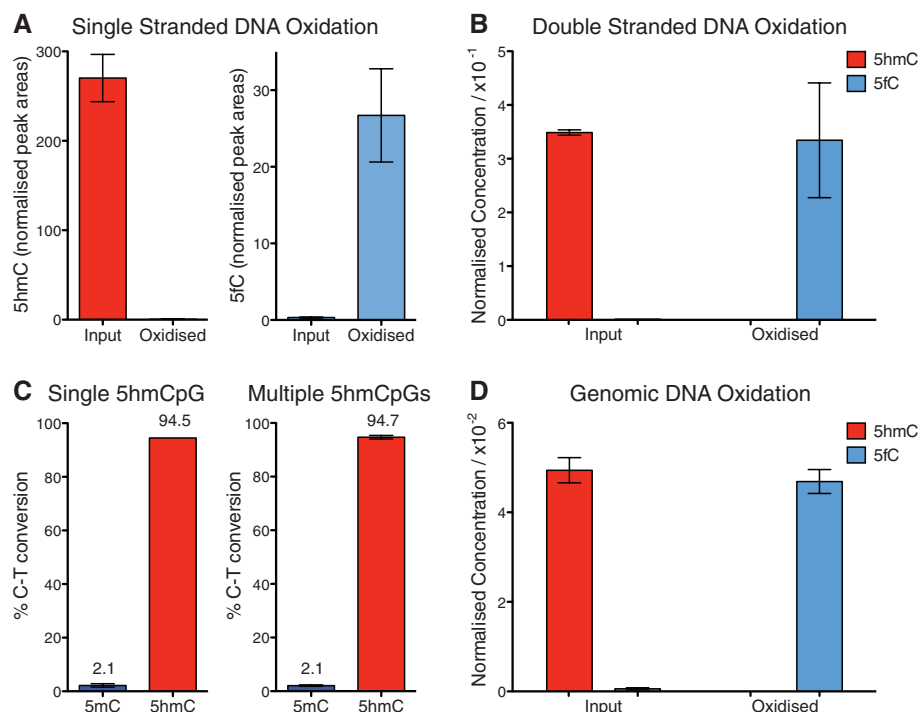
Fig. 2. Quantification of 5hmC oxidation. (**A**) Levels of 5hmC and 5fC (normalized to T) in a 15-nucleotide oligomer ssDNA oligonucleotide before and after KRuO$_4$ oxidation, measured with mass spectrometry. (**B**) Levels of 5hmC and 5fC (normalized to 5mC) in a 135-nucleotide oligomer dsDNA fragment before and after KRuO$_4$ oxidation. (**C**) C-to-T conversion levels as determined by means of Illumina sequencing of two dsDNA fragments containing either a single 5hmCpG (122-nucleotide oligomer) or multiple 5hmCpGs (135-nucleotide oligomer) after oxidative bisulfite treatment. 5mC was also present in these strands. (**D**) Levels of 5hmC and 5fC (normalized to 5mC in primer sequence) in ES cell DNA measured before and after oxidation. Data are mean ± SD.

5fC (Fig. 2D), with no detectable degradation of C (fig. S1C). Thus, the oxidative bisulfite protocol specifically converts 5hmC to U in DNA, leaving C and 5mC unchanged, enabling quantitative, single-nucleotide-resolution sequencing on widely available platforms.

We then used oxBS-Seq to quantitatively map 5hmC at high resolution in the genomic DNA of mouse embryonic stem (ES) cells. We chose to combine oxidative bisulfite with reduced representation bisulfite sequencing (RRBS) (21), which allows deep, selective sequencing of a fraction of the genome that is highly enriched for CpG islands (CGIs). We generated RRBS and oxidative RRBS (oxRRBS) data sets, achieving an average sequencing depth of ~120 reads per CpG, which when pooled yielded an average of ~3300 methylation calls per CGI (fig. S3). After applying depth and breadth cutoffs (supplementary materials, materials and methods), 55% (12,660) of all CGIs (22) were covered in our data sets.

To identify 5hmC-containing CGIs, we tested for differences between the RRBS and oxRRBS data sets using stringent criteria, yielding a false discovery rate of 3.7% (supplementary materials, materials and methods). We identified 800 5hmC-containing CGIs, which had an average of 3.3% (range of 0.2 to 18.5%) CpG hydroxymethylation (Fig. 3, A and B). We also identified 4577 5mC-containing CGIs averaging 8.1% CpG methylation (Fig. 3B). We carried out sequencing on an independent biological duplicate sample of the same ES cell line but at a different passage
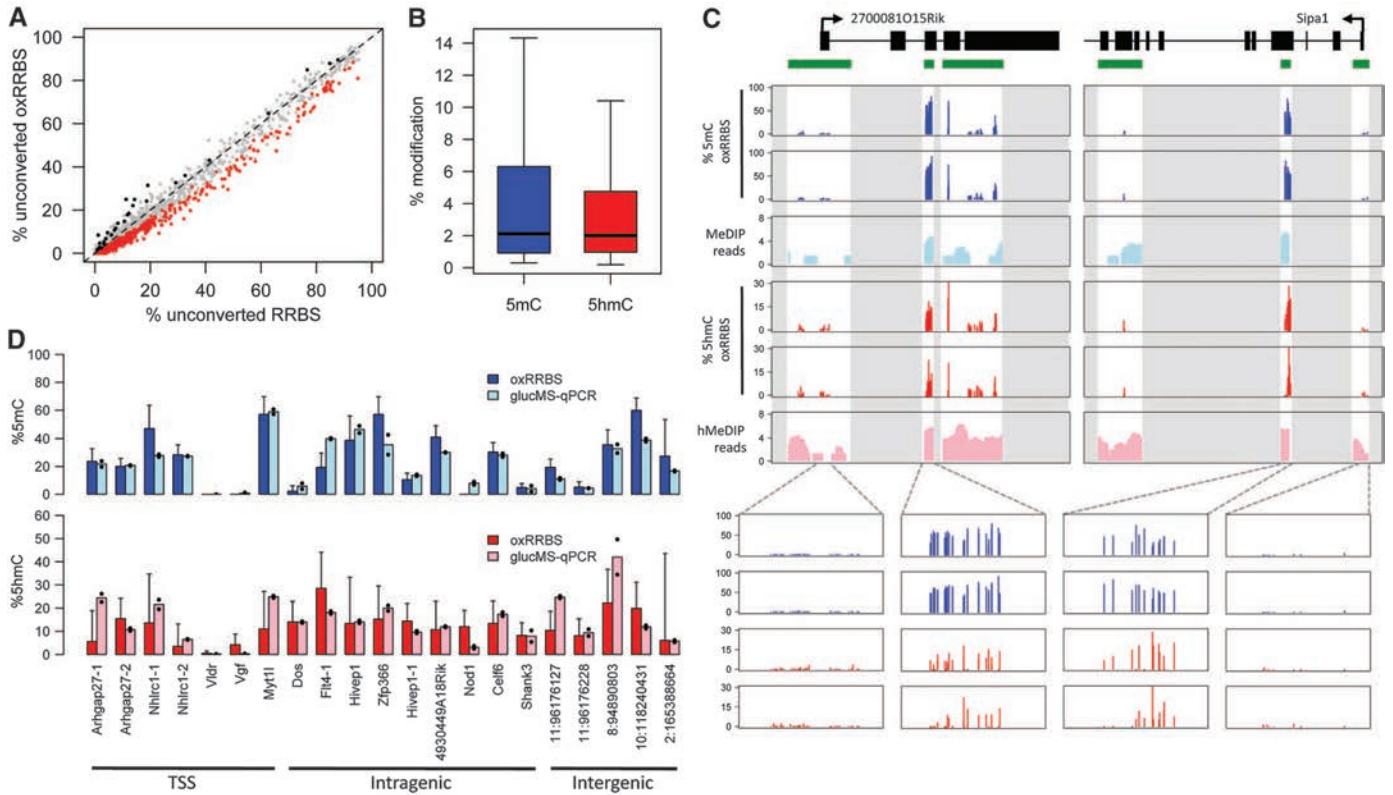
Fig. 3. Quantification of 5mC and 5hmC levels at CGIs by means of oxRRBS. (**A**) Fraction of unconverted cytosines per CGI; 5hmC-containing CGIs (red) have a statistically significant lower fraction in the oxRRBS data set; a false discovery rate of 3.7% was estimated from the CGIs with the opposite pattern (black). (**B**) 5mC and 5hmC levels within CGIs with significant levels of the respective modification. (**C**) Examples of genomic RRBS and oxRRBS profiles overlapped with (h)MeDIP-Seq profiles (7). Green bars represent CGIs; data outside CGIs were masked (gray areas). Each bar in the oxRRBS tracks represents a single CpG (in either DNA strand). (**D**) 5mC and 5hmC levels at selected MspI sites were validated through glucMS-qPCR. OxRRBS data are percentage ± 95% confidence interval. Mean glucMS-qPCR values are shown, with the black dots representing individual replicates.

number, which according to mass spectrometry had reduced levels of 5hmC (0.10 versus 0.16% of all Cs), and consistently we found fewer 5hmC-containing CGIs (supplementary text). 5hmC-containing CGIs present in both samples showed good quantitative reproducibility (fig. S5). In non-CpG contexts, we found very few CGIs (71) with levels of 5mC above the bisulfite conversion error (0.2%) (fig. S9) and no CGIs with detectable levels of 5hmC.

Genes associated with 5mC-containing CGIs included *Dazl*, which is known to be methylated in ES cells (fig. S7) (23). Similarly, we found that *Zfp64* and *Ecat1* had significant levels of 5hmC (7). Genes with >5% 5hmC at transcription start site (TSS) CGIs were associated with gene ontology terms related to transcription factor activity—and in particular were enriched in developmentally relevant genes encoding for Homeobox-containing proteins (such as *Irx4*, *Gbx1*, and *Hoxc4*). To validate our method, we quantified 5hmC and 5mC levels at 21 CGIs containing MspI restriction sites by means of glucosylation-coupled methylation-sensitive quantitative polymerase chain reaction (glucMS-qPCR) (Fig. 3D) (24). We found a good correlation between the quantification with oxRRBS and glucMS-qPCR [correlation coefficient (r) = 0.86,
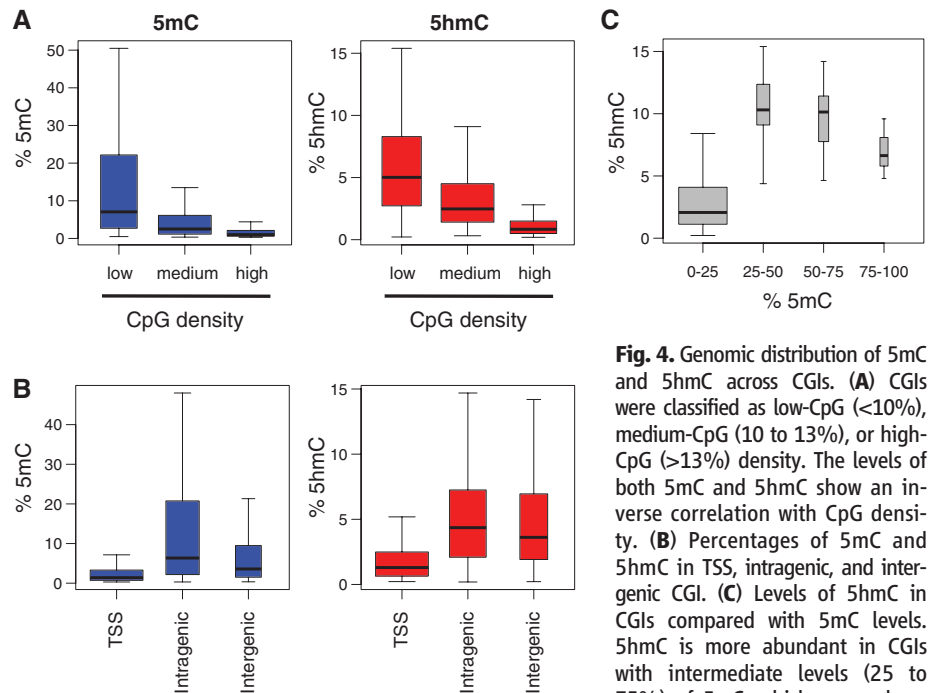

Fig. 4. Genomic distribution of 5mC and 5hmC across CGIs. (**A**) CGIs were classified as low-CpG (<10%), medium-CpG (10 to 13%), or high-CpG (>13%) density. The levels of both 5mC and 5hmC show an inverse correlation with CpG density. (**B**) Percentages of 5mC and 5hmC in TSS, intragenic, and intergenic CGI. (**C**) Levels of 5hmC in CGIs compared with 5mC levels. 5hmC is more abundant in CGIs with intermediate levels (25 to 75%) of 5mC, which are perhaps more epigenetically plastic. For all boxplots, the width of the box is proportional to the amount of data within that group.

$P = 5 \times 10^{-7}$ and $r = 0.52$, $P = 0.01$ for 5mC and 5hmC, respectively], showing that oxRRBS reliably measures 5hmC at individual CpGs. We also found a good correlation between oxRRBS and our previously published (hydroxy)methylated DNA immunoprecipitation sequencing [(h)MeDIP-Seq] data sets (fig. S8) (7).

Across CGIs, both 5mC and 5hmC levels are inversely correlated with CpG density, and intragenic and intergenic CGIs contain higher levels of either modification than those overlapping TSSs (Fig. 4, A and B, and fig. S6) (13, 22). TET1 is enriched at TSSs, and thus, a high turnover of 5mC and 5hmC that would keep the steady-state levels low at these sites has been suggested (9). Non-TSS CGIs, however, appear to accumulate substantial amounts of both marks, suggesting reduced turnover in these regions. We find that the highest levels of 5hmC are found at CGIs with intermediate levels (25 to 75%) of 5mC (Fig. 4C and fig. S6). Although low-5mC CGIs have reduced potential for 5hmC generation and/or are subjected to a high turnover, high-5mC CGIs are perhaps protected from extensive TET-mediated oxidation, thus stabilizing methylation. Intermediate-5mC CGIs are therefore potentially more epigenetically plastic, given the relatively high abundance of both marks.

Most TSS CGIs (98%) have less than 10% 5mC, as well as low 5hmC, and these are associated with higher transcription levels than average (fig. S10). Within this narrow window, we find a mild negative correlation between transcription and both 5mC and 5hmC levels (fig. S10). At higher 5mC levels, there are insufficient CGIs to obtain a statistically significant result, and it remains possible that here the epigenetic balance between 5mC and 5hmC plays an important transcriptional role, as we previously suggested (7).

Last, we quantified 5mC and 5hmC levels at two classes of retrotransposons [long interspersed nuclear element–1 (LINE1) and intracisternal A-particle (IAP)] using two approaches: aligning the oxRRBS reads to the respective consensus sequences and combining oxidative bisulfite with MassARRAY technology (Sequenom, San Diego, California) (fig. S11). We find that LINE1 elements display a considerable amount of 5hmC (approximately 5%), as previously suggested through (h)MeDIP-Seq (7). IAPs, on the other hand, have low or no 5hmC. Because LINE1 elements are reprogrammed during preimplantation development whereas IAPs are resistant to this process (25), this suggests a possible involvement of 5hmC in the demethylation of specific repeat classes.

The oxBS-Seq method reliably maps and quantifies both 5mC and 5hmC at the single-nucleotide level. Owing to the fundamental mechanism of oxBS-Seq, the approach is compatible with any sequencing platform. In ES cells, we found that in CGIs 5hmC is exclusive to CpG dinucleotides and that it accumulates at intragenic, low-CpG-density CGIs, which tend to have intermediate levels of 5mC and may be particularly epigenetically plastic.

References and Notes

1. A. M. Deaton, A. Bird, *Genes Dev.* **25**, 1010 (2011).
2. M. Tahiliani et al., *Science* **324**, 930 (2009).
3. S. Ito et al., *Nature* **466**, 1129 (2010).
4. M. R. Branco, G. Ficz, W. Reik, *Nat. Rev. Genet.* **13**, 7 (2012).
5. S. Kriaucionis, N. Heintz, *Science* **324**, 929 (2009).
6. M. Münzel et al., *Angew. Chem. Int. Ed.* **49**, 5375 (2010).
7. G. Ficz et al., *Nature* **473**, 398 (2011).
8. W. A. Pastor et al., *Nature* **473**, 394 (2011).
9. H. Wu et al., *Genes Dev.* **25**, 679 (2011).
10. S. G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **39**, 5015 (2011).
11. C. X. Song et al., *Nat. Biotechnol.* **29**, 68 (2011).
12. K. Williams et al., *Nature* **473**, 343 (2011).
13. Y. Xu et al., *Mol. Cell* **42**, 451 (2011).
14. Y. Huang et al., *PLoS ONE* **5**, e8888 (2010).
15. C. Nestor, A. Ruzov, R. Meehan, D. Dunican, *Biotechniques* **48**, 317 (2010).
16. C. X. Song et al., *Nat. Methods* **9**, 75 (2012).
17. J. Eid et al., *Science* **323**, 133 (2009).
18. E. V. Wallace et al., *Chem. Commun. (Camb.)* **46**, 8195 (2010).
19. M. Wanunu et al., *J. Am. Chem. Soc.* **133**, 486 (2010).
20. G. Green, W. P. Griffith, D. M. Hollinshead, S. V. Ley, M. Schroder, *J. Chem. Soc. Perkin Trans. 1* **1**, 681 (1984).
21. A. Meissner et al., *Nature* **454**, 766 (2008).
22. R. S. Illingworth et al., *PLoS Genet.* **6**, e1001134 (2010).
23. J. Borgel et al., *Nat. Genet.* **42**, 1093 (2010).
24. S. M. Kinney et al., *J. Biol. Chem.* **286**, 24685 (2011).
25. N. Lane et al., *Genesis* **35**, 88 (2003).

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1220671/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 and S2
References (26–40)

Downloaded from www.sciencemag.org on July 7, 2015