Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells

Hao Wu,1,2,7 Ana C. D’Alessio,3,4 Shinsuke Ito,3,4 Zhibin Wang,5 Kairong Cui,6 Keji Zhao,6 Yi Eve Sun,1,2 and Yi Zhang3,4,8

1Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California 90095, USA; 2Department of Psychiatry and Biobehavioral Sciences, Intellectual Development and Disabilities Research Center, Semel Institute of Neuroscience and Human Behavior, University of California at Los Angeles, Los Angeles, California 90095, USA; 3Howard Hughes Medical Institute, University of North Carolina, Chapel Hill, North Carolina 27599, USA; 4Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA; 5Laboratory of Human Environmental Epigenomes, Department of Environmental Health Sciences, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA; 6Laboratory of Molecular Immunology, The National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Recent studies have demonstrated that the Ten-eleven translocation (Tet) family proteins can enzymatically convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). While 5mC has been studied extensively, little is known about the distribution and function of 5hmC. Here we present a genome-wide profile of 5hmC in mouse embryonic stem (ES) cells. A combined analysis of global 5hmC distribution and gene expression profile in wild-type and Tet1-depleted ES cells suggests that 5hmC is enriched at both gene bodies of actively transcribed genes and extended promoter regions of Polycomb-repressed developmental regulators. Thus, our study reveals the first genome-wide 5hmC distribution in pluripotent stem cells, and supports its dual function in regulating gene expression.

Supplemental material is available for this article.

Received January 25, 2011; revised version accepted February 22, 2011.

[Keywords: Tet1, 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), mouse embryonic stem cells, genome-wide 5hmC distribution, Polycomb repression]

7Present address: Cardiovascular Research Center, Massachusetts General Hospital, 185 Cambridge St., Boston, MA 02114, USA, and Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Ave., Cambridge, MA 02138, USA.

E-MAIL yi_zhang@med.unc.edu; FAX (919) 966-4330.

Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.2036011.

Mammalian genomes are chemically modified by DNA cytosine methylation, an inheritable epigenetic mark that is implicated in many biological and pathological processes, including gene regulation, genomic imprinting, X-chromosome inactivation, suppression of transposable elements, and tumorigenesis (Cedar and Bergman 2009; Ooi et al. 2009). Genome-wide studies of mammalian DNA methylation have shown that 5-methylcytosine (5mC) is widely distributed across the genome in a nonrandom manner [Weber et al. 2007; Fouse et al. 2008; Meissner et al. 2008; Lister et al. 2009]. In conjunction with other epigenetic modifications, 5mC can regulate accessibility of the DNA to transcription factors and chromatin regulators, thereby contributing to gene regulation and cellular differentiation.

Recent studies have uncovered 5-hydroxymethylcytosine (5hmC) as the sixth base of the genome, and that the Ten-eleven translocation (Tet) family proteins is responsible for the generation of 5hmC from 5mC in mammalian cells [Kriaucionis and Heintz 2009; Tahiliani et al. 2009; Ito et al. 2010]. This new discovery raises the possibility that 5hmC may function as another epigenetic mark by altering chromatin structure or contributing to the recruitment or exclusion of other DNA-binding proteins that affect transcription. Recent reports have shown that 5hmC is relatively enriched in several cell types, including mouse embryonic stem (ES) cells and certain neuronal cells [Kriaucionis and Heintz 2009; Tahiliani et al. 2009; Globisch et al. 2010; Szwagierczak et al. 2010]. Expression and functional analysis have further demonstrated that Tet1, the founding member of the Tet family, is highly expressed in mouse ES cells, and depletion of Tet1 results in impairment of ES cell self-renewal and maintenance [Ito et al. 2010]. Consistent with the essential role of Tet1 in ES cells, we showed recently that Tet1 occupies regulatory regions of both pluripotency-related genes and Polycomb group (PcG) protein-repressed developmental regulators [Wu et al. 2011]. However, little is known about the genomic distribution of 5hmC, dependence of 5hmC on Tet1 occupancy, and the regulatory function of 5hmC on transcription. Here we report the first genome-wide map of 5hmC occupancy in mouse ES cells. The comparison of 5hmC distribution with other epigenetic marks and global expression profile provides evidence for a role of 5hmC in both transcriptional activation and repression.

Results and Discussion

Genome-wide distribution of 5hmC in mouse ES cells

To test the specificity of 5hmC antibodies in immunoprecipitating unmethylated, methylated, and hydroxymethylated synthetic DNA, we optimized the amount of input DNA and found that affinity-purified polyclonal antibodies [Active Motif] for 5hmC specifically immunoprecipitated 5hmC-containing, but not 5mC- or C-containing DNA under denaturing conditions [Supplemental Fig. S1A]. We then tested the ability of both rabbit polyclonal [Active Motif] and rat monoclonal [Diagenode] antibodies in immunoprecipitation of genomic DNA at three Tet1-bound targets [Nanog, Tcl1, and Sox17] determined by
Tet1 chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq) (Wu et al. 2011). Quantitative PCR (qPCR) analysis indicated that both antibodies could readily immunoprecipitate 5hmC-modified genomic DNA, and 5hmC polyclonal antibodies showed slightly higher immunoprecipitation efficiency as compared with the monoclonal antibody [Supplemental Fig. S1B,C).

To determine global distribution of 5hmC, we further tested 5hmC antibody-based immunoprecipitation combined with chromosome-wide tiling microarrays [Supplemental Fig. S2A]. We found that both antibodies could consistently detect peaks of 5hmC at defined genomic regions, and that the 5hmC profile was different from that of 5mC at both gene and probe levels [Supplemental Fig. S2A–C], suggesting that the 5hmC antibody is specific. Since the majority of Tet1-binding sites in mouse ES cells are within nonrepetitive, gene-rich genomic regions [Wu et al. 2011], we mapped 5hmC distribution using whole-genome tiling microarrays that covered the entire nonrepetitive portion of the mouse genome. A total of 91,913 genomic regions enriched with 5hmC were identified with high confidence [Supplemental Table S1]. Nearly 60% of 5hmC peaks were found to be within gene bodies of annotated RefSeq genes [Fig. 1A], suggesting that 5hmC is also preferentially associated with gene-rich regions of the genome [Supplemental Fig. 2A]. We note that a recent study profiling the 5hmC distribution in mouse cerebellum using a different method has come to a similar conclusion [Song et al. 2011]. Further analysis showed that most 5hmC-enriched regions were associated with moderate CpG density [Fig. 1B]. Thus, 5hmC is enriched in both Tet1-activated and Tet1-repressed genes.

Figure 1. Genomic distribution of 5hmC in mouse ES cells. [A] Genomic distribution of 5hmC-enriched regions [log10 peak P-value > 2.3] relative to University of California at San Diego RefSeq genes [NCBI build 36]. The genome-wide 5hmC occupancy was determined by whole-genome tiling microarray analysis. [B] Proportion of 5hmC-enriched regions with different CpG densities. Note that 5hmC is enriched at genomic regions with moderate-density CpG dinucleotides.

5hmC is enriched in both Tet1-activated and Tet1-repressed genes

Previous genome-wide analyses have shown that distribution of DNA methylation in the genome may also be regulated by histone modifications [Ooi et al. 2007; Schlesinger et al. 2007; Mohn et al. 2008]. For example, H3K4me3 at proximal promoters generally regulates
5hmC enrichment (measured by %C0 modificiations at Tet1-bound genes, we cross-referenced relationship between Tet1, 5hmC, and major histone methylation (Ball et al. 2009). To further investigate the correlated with the presence of high levels of DNA methylation, is greatly depleted from binding sites of transcription factors related to pluripotency (e.g., Nanog, Sox2, and Oct4) (Chen et al. 2008). In contrast to a general depletion of 5mC at DNA–protein interaction sites, we observed a relative enrichment of 5mC toward the site of most DNA-binding proteins [Supplemental Fig. S5]. Previous analysis of DNA methylation in human ES cells using whole-genome bisulfite sequencing suggests that 5mC in a non-CpG context, but not CpG DNA methylation, is greatly depleted from binding sites of transcription factors related to pluripotency (Lister et al. 2009). Since bisulfite treatment cannot discriminate 5mC from 5hmC (Huang et al. 2010; Jin et al. 2010), bisulfite sequencing may overestimate the 5mC levels at these binding sites. Indeed, specific antibody-based immunoprecipitation analysis of 5mC and 5hmC in mouse ES cells indicated that 5mC was generally depleted from DNA–protein interaction sites, whereas 5hmC was relatively enriched at these sites [Supplemental Fig. S5].

We next analyzed a set of genomic features defined by histone modifications or sequence-specific DNA-binding

Figure 2. Tet1 is required for maintaining 5hmC levels at defined genomic regions in ES cells. (A) Distribution of 5hmC relative to all annotated genes in ES cells. Averaged 5hmC enrichment (measured by –log10 peak P-value) in 200-base-pair (bp) bins upstream of/downstream from gene bodies or at 5% intervals within the gene body is shown along the transcription units from 5 kb upstream of TSSs to 5 kb downstream from the transcriptional end sites (TESs). Note that 5hmC levels are generally enriched in Tet1-bound genes as compared with Tet1-unbound genes. (B) Changes in 5hmC and 5mC enrichment (measured by –log10 peak P-value) in response to Tet1 knockdown are shown for Tet1-bound regions associated with different genomic features [gene body, intergenic region, and promoter (2 kb flanking TSSs)]. (C) Heat map representation of CpG islands and occupancy of Tet1, 5mC, and 5hmC in mouse ES cells at all Tet1-enriched regions [5 kb flanking the center of Tet1 peaks]. The heat map is rank-ordered by CpG density of genomic regions within 500 bp flanking the center of Tet1 peaks. The enrichment of 5hmC and 5mC was determined by whole-genome tiling microarrays. Tet1-bound regions at gene bodies, promoters, and intergenic regions are shown separately. The enrichment of Tet1 binding was determined previously by ChIP-seq analyses (Wu et al. 2011). All average binding was measured by –log10 [peak P-values] in 200-bp bins and are shown by color scale. The following color scales [white, no enrichment; blue, high enrichment] are used for 5hmC, 5mC, and Tet1, respectively: [0, 2], [0, 0.5], and [0, 50]. The presence of CpG islands is displayed in color [blue, present; white, absent]. (D) Tet1 occupancy and changes in 5hmC/5mC levels are shown for a group of representative Tet1 targets [Pcdha gene cluster on chr18] in control (Con) and Tet1 knockdown (Tet1 KD) ES cells. Tet1 ChIP-seq data in control knockdown (Con KD) and Tet1 knockdown (Tet1 KD) are shown in read counts, with the Y-axis floor set to 0.2 read per million reads. 5hmC and 5mC levels are shown as log2 ratios of immunoprecipitation/input (IP/input).

DNA methylation levels in a negative fashion [Weber et al. 2007; Meissner et al. 2008], whereas H3K36me3 within actively transcribed gene bodies is positively correlated with the presence of high levels of DNA methylation [Ball et al. 2009]. To further investigate the relationship between Tet1, 5hmC, and major histone modifications at Tet1-bound genes, we cross-referenced
proteins, including H3K4me3-enriched promoter regions, methylated H3K4 [H3K4me1]-enriched enhancers, CTCF-marked insulators, and H3K36me3-enriched transcribed regions [Mikkelsen et al. 2007; Chen et al. 2008; Meissner et al. 2008]. Except for H3K36me3-enriched regions, 5mC was, in general, depleted from these genomic features [Supplemental Fig. S5], consistent with the notion that DNA methylation negatively regulates most DNA–protein interactions. In contrast, average signal profiles of 5hmC showed a relative enrichment at promoters, enhancers, transcribed regions, and insulators [Supplemental Fig. S5]. Notably, the general enrichment of 5hmC at H3K4me3 peaks indicates that the absence of 5hmC at a subset of CpG-rich, H3K4me3-enriched proximal promoters is probably due to the existence of additional regulatory factors of Tet1 activity or proteins capable of rapidly converting 5hmC into unmethylated cytosine at these sites. Furthermore, the observed enrichment of 5hmC and concomitant depletion of 5mC at enhancer or insulator sequences may therefore contribute to maintaining a more accessible chromatin structure for binding of enhancer proteins (e.g., p300) and CTCF to these sites. Enrichment of both 5hmC and 5mC at actively transcribed regions marked by high levels of H3K36me3 suggests a transcriptional link between these two marks [Supplemental Fig. S5].

5hmC has different distribution profiles at active and repressed genes

To investigate a potential role of 5hmC in transcriptional regulation, we examined the relationship between 5hmC distribution and the global gene expression profile. This analysis showed that 5hmC was relatively enriched within intragenic regions of genes transcribed at high and medium levels [Fig. 4A, blue and red, Supplemental Fig. S6, blue and red], as well as promoter regions of transcriptionally inactive genes [Fig. 4A, green; Supplemental Fig. S6, green]. Further analysis of 5hmC distribution on Tet1-bound genes ranked by their expression levels also supported a potential role of 5hmC in both transcriptional activation and repression [Fig. 4B]. Interestingly, 5hmC was enriched at promoters of both Tet1/PRC2-cobound [Supplemental Fig. S7, blue] and Tet1-only targets that were expressed at low levels [Supplemental Fig. S7, purple] in mouse ES cells, suggesting that promoter 5hmC may function as a general repressive mark [Supplemental Fig. S7]. To investigate further how 5hmC distribution may contribute to Tet1-dependent gene expression, we compared the 5hmC profiles between control and Tet1-depleted ES cells on Tet1-repressed and Tet1-activated targets. We found that 5hmC levels were decreased at both groups of Tet1 targets [Supplemental Fig. S8]. A decrease in 5hmC was more pronounced at promoter regions and the 5' end of intragenic regions on Tet1-repressed targets, whereas a depletion of intragenic 5hmC was evident for Tet1-activated targets. Taken together, these results indicate that, similar to 5mC, 5hmC may play a complex role in

Figure 3. 5hmC is enriched in both repressed [bivalent, Tet1/PRC2-cobound] and actively transcribed [Tet1-only] genes. (A) Heat map representation of genomic regions with enriched 5hmC, binding profile of Tet1, and Ezh2, and major histone modifications (H3K4me3, H3K27me3, and H3K36me3) [Mikkelsen et al. 2007] in mouse ES cells at all Tet1 target genes [5 kb flanking TSSs]. The heat map is rank-ordered from genes with the highest H3K27me3 enrichment to no H3K27me3 within 5-kb genomic regions flanking TSSs. The enrichment of 5hmC and 5mC was determined by whole-genome tiling microarrays. The enrichment of Tet1, H3K4me3, H3K27me3, and H3K36me3 binding was determined previously by ChIP-seq analyses [Mikkelsen et al. 2007; Wu et al. 2011]. All average binding was measured by −log10 [peak P-values] in 200-bp bins and is shown by color scale. The following color scales [white, no enrichment; blue, high enrichment] are used for 5hmC/5mC, Tet1/H3K27me3/H3K36me3, and H3K4me3, respectively: [0, 2], [0, 50], and [0, 100]. The presence of CpG islands is displayed in color (blue, present; white, absent). (B) Average distribution profiles of 5hmC enrichment are shown for Tet1/PRC2-cobound targets, Tet1-only targets, and nontargets. Averaged expression levels of these three groups of genes are shown in the bottom panels (measured by log2 values of expression microarray signals). (C) Shown are profiles of Tet1 [Wu et al. 2011], 5hmC, Ezh2 [Ku et al. 2008], RNA polymerase II [Seila et al. 2008], and major histone modification [Mikkelsen et al. 2007] occupancy at two representative Tet1 targets: a Tet1/PRC2-cobound target (Lhx2), and a Tet1-only target (Rest promoter). ChIP-seq data in mouse ES cells are shown in read counts, with the Y-axis floor set to 0.2 read per million reads.

Wu et al.
transcriptional regulation, depending on its location (Wu et al. 2010). Our analysis of 5hmC distribution in mouse ES cells suggests that promoter and gene body 5hmC may preferentially contribute to gene repression and activation, respectively.

In summary, our studies have presented a genome-wide map of 5hmC in mouse ES cells. Systematic comparison of 5hmC distribution, Tet1 occupancy, and major histone modifications indicate that 5hmC may be involved in establishing and maintaining chromatin structure for both actively transcribed genes and PcG-repressed developmental regulators. We also provide initial evidence indicating that 5hmC may contribute to both transcriptional activation and repression in a context-dependent manner. Collectively, these results and the demonstration of a simple antibody-based approach in genome-wide 5hmC mapping have set the stage for further understanding the functions of Tet family proteins and 5hmC in development and disease.

Materials and methods

Genome-wide and locus-specific 5hmC analysis (hMeDIP)

To immunoprecipitate 5hmC, genomic DNA was sequenzially digested with proteinase K and RNase A and purified by phenol-chloroform extraction. Purified genomic DNA was sonicated to 200–1000 base pairs (bp) and heat-denatured (10 min, 95°C). An aliquot of sonicated genomic DNA was saved as input. Five micrograms of fragmented genomic DNA was immunoprecipitated with 5 µg of rat 5hmC Ab (Diagenode, catalog no. MAAb-633HMC) or rabbit 5hmC Ab (Active Motif, catalog no. 39791) overnight at 4°C in a final volume of 500 µL of immunoprecipitation buffer (10 mM sodium phosphate at pH 7.0, 140 mM NaCl, 0.05% Triton X-100). The DNA–antibody mixture was incubated with 30 µL of protein G Dynabeads (Invitrogen) for 2 h at 4°C and washed three times with 1 mL of immunoprecipitation buffer. The beads were then treated with proteinase K for at least 3 h at 55°C, and the methylated DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. For whole-genome DNA tiling microarray analysis, immunoprecipitated 5hmC-containing DNA from control or Tet1-depleted ES cells was cohybridized with input DNA to mouse whole-genome tiling microarrays ( NimbleGen). Locus-specific hMeDIP–qPCR was performed similarly using nondenatured genomic DNA, and immunoprecipitated DNA was analyzed on an ABI 7300 system (Applied Biosystems) using SYBR Green (Invitrogen). Primer sequences are listed in Supplemental Table S2.

To evaluate the immunoprecipitation efficiency of 5hmC antibodies (Active Motif) for synthetic DNA (949 bp, Zymo Research), 25 pg of unmethylated, methylated, or hydroxymethylated DNA was diluted in 480 µL of 1× TE buffer. DNA was heat-denatured for 10 min at 95°C, and chilled for 5 min on ice. 5hmC immunoprecipitation was performed as described above. qPCR was carried out with primers H/me-1-F (AGGTGGAGG AACGCTCATGTC) and H/me-1-R (ATAAACCGAACC GCTACACC).

Whole-genome tiling microarray analysis

For whole-genome DNA tiling microarray analysis of 5hmC distribution, immunoprecipitated and input DNA was prepared from both control and Tet1-depleted ES cells and amplified using a whole-genome amplification kit (Sigma). Probe labeling, amplification, hybridization, data extraction, and analysis were performed as described previously (Wu et al. 2011).

For identification of probes associated with significant levels of 5hmC, a nonparametric one-sided Kolmogorov-Smirnov (KS) test was used. Briefly, from the scaled log2 ratio data, a fixed-length window (750 bp) was placed around each consecutive probe, and the one-sided KS test was applied to determine whether the probes were drawn from a significantly more positive distribution of intensity log ratios than those in the rest of the array. The resulting score for each probe was the [−log10 P-value] from the windowed KS test around that probe. Peak data files were generated from the P-value data files using NimbleScan version 2.5. Peaks within 500 bp of each other were merged. For calculating the absolute 5mC levels in control knockdown and Tet1 knockdown ES cells, the MEDME program (Pelizzola et al. 2008) was used to correct the nonlinear relationship between microarray signals and genomic CpG density. To visualize 5mC and 5hmC distributions in the Cisgenome browser (ji et al. 2008), probe-level smoothing distributions in the Cisgenome browser (jī et al. 2008) were applied to microarray data (Pelizzola et al. 2008).

Consequently, this analysis of the methylation level of 5mC and 5hmC in mouse ES cells should be further improved by combining a method that can distinguish 5mC from 5hmC, such as mass spectrometry or high-resolution melting analysis.

 Constructs and antibodies

All of the constructs and antibodies used in this study were described previously [Ito et al. 2010, Wu et al. 2011].
ChIP-seq, gene expression profiling, and data analysis

ChIP-seq experiments and data analysis for Tet1 were described previously (Wu et al. 2011). ChIP-seq data sets of H3K4me3, H3K27me3, H3K36me3 (Mikkelsen et al. 2007), Ezh2 (Ku et al. 2008), and RNA polymerase II (Seila et al. 2008) were obtained from previous publications and reanalyzed in MACS using identical parameters (except statistical cutoff was set to P-value < 10⁻⁵). ChIP-seq sequencing read counts for each ChIP-seq experiments were binned into 400-bp windows at 100-bp steps along the genome and visualized in the Cisgenome browser (Ji et al. 2008). To assign ChIP-seq enriched regions to genes, RefSeq genes were downloaded from the UCSC Table Browser (May 2010). For all data sets, genes with enriched regions within 5 kb of their TSSs were called bound. Gene expression profiling analysis of control and Tet1-depleted mouse ES cells was carried out using the Affymetrix GeneChip Mouse Genome 430 2.0 array. Tet1 ChIP-seq and gene expression microarray data have been deposited in the Gene Expression Omnibus under accession number GSE26833 (Wu et al. 2011).

Mouse ES cell culture, Tet knockdown, and qPCR

Mouse E14Tg2A ES cells were cultured in feeder-free conditions (Ito et al. 2010). Control and Tet1 knockdown cell preparation and qPCR verification were described previously (Wu et al. 2011).

Acknowledgments

We thank Brian Abraham and Iouri Chepelev for help with data transfer, Jinzhao Wang for FACS sorting, and Susan Wu for critical reading of the manuscript. This work was supported by NIH grant GM68804 (to Y.Z.), and support to the Division of Intramural Research Program of National Heart, Lung, and Blood Institute from the NIH (to K.Z.). S.L. is a research fellow of the Japanese Society for the Promotion of Science. Y.Z. is an Investigator of the Howard Hughes Medical Institute.

References

Ball MP, Li JB, Gao Y, Lee YH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM. 2009. Targeted and genome-scale strategies reveal gene body methylation signatures in human cells. Nat Biotechnol 27: 361–368.

Cedar H, Bergman Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10: 295–304.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orolov YL, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133: 1106–1117.

Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell 2: 160–169.

Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Briel M, Carell T. 2010. Tissue distribution of 5-hydroxymethylcytosine in bisulfite sequencing. Mol Cell 38: 755–766.

Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS ONE 5: e13567. doi: 10.1371/journal.pone.0013567.

Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS ONE 5: e13567. doi: 10.1371/journal.pone.0013567.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Nigo QM, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Svavlenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al. 2008. Genomic-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 583–590.

Mohr F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 30: 755–766.

Ooi SK, Qi C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, et al. 2007. Dnmt3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448: 714–717.

Ooi SK, O’Donnell AH, Bestor TH. 2009. Mammalian cytosine methylation at a glance. J Cell Sci 122: 2787–2791.

Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, Habalan M. 2008. MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray-derived MeDIP-enrichment. Genome Res 18: 1652–1659.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 39: 232–236.

Seila AC, Calabrese JM, Levine SS, Yeow GW, Rahl PB, Flynn RA, Young KP, Sharp PA. 2008. Divergent transcription from active promoters. Science 322: 1849–1851.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, et al. 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29: 68–72.

Szewczureczak A, Bultmann S, Schmidt CS, Spada F, Leonardirt H. 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 38: e181. doi: 10.1093/nar/gkp684.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324: 930–935.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39: 457–466.

Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, Li E, Zhang Y, Sun YE. 2010. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science 329: 444–448.

Wu H, D’Alessio AC, Ito S, Xie K, Wang Z, Cui K, Zhao K, Sun Y, Zhang Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature doi: 10.1038/nature09934.