Cell Cycle 10:15, 2428-2436; August 1, 2011; © 2011 Landes Bioscience

Tet1 and 5-hydroxymethylation

A genome-wide view in mouse embryonic stem cells

Hao Wu[1,2] and Yi Zhang[3,4]

[1]Cardiovascular Research Center; Massachusetts General Hospital; Boston, MA USA; [2]Department of Stem Cell and Regenerative Biology; Harvard University; Cambridge, MA USA; [3]Howard Hughes Medical Institute; [4]Department of Biochemistry and Biophysics; Lineberger Comprehensive Cancer Center; University of North Carolina at Chapel Hill; Chapel Hill, NC USA

Inner cell mass (ICM) cells of a blastocyst, the source of embryonic stem (ES) cells, are characterized by their unique ability to give rise to all cell types in adult organisms. The epigenomes of germ cells and developing zygotes undergo extensive reprogramming to acquire such a pluripotent state. A major reprogramming event during early embryonic development is the erasure and subsequent re-establishment of patterns of methylation at the 5-position of cytosine (5mC). The recent demonstration that Ten-eleven translocation family proteins, Tet1-3 have the capacity to convert 5mC to 5-hydroxymethylcytosine (5hmC) raises the possibility that 5hmC may act as an distinct epigenetic state contributing to dynamic changes in DNA methylation and transcriptional regulation during embryonic development. In ES cells, Tet1 is highly expressed and 5hmC is present at relatively high levels compared to most differentiated cells, but the functional significance of Tet1 and 5hmC in these pluripotent cells are not clear. Recently, a flurry of papers that profile the distribution of Tet1 and/or 5hmC across the genome of mouse ES cells provide new insights into the role of Tet proteins and 5hmC in regulating expression of genes related to pluripotency and cellular differentiation. Through integrative analyses of datasets from different groups, we reveal the common Tet1 and 5hmC targets in undifferentiated mouse ES cells, which suggest that Tet1 may play a key role in orchestrating the balance between pluripotent and lineage committed states.

Key words: Tet1, 5-hydroxymethylcytosine, DNA methylation, embryonic stem cells, pluripotency, cell differentiation, polycomb repression, epigenomics

Submitted: 06/14/11

Accepted: 06/16/11

DOI: 10.4161/cc.10.15.16930

Correspondence to: Hao Wu and Yi Zhang; Email: haowu7@gmail.com and yi_zhang@med.unc.edu

Introduction

DNA cytosine methylation is essential for mammalian embryogenesis, and is dynamically regulated during early embryonic and germ cell development.[1] This epigenetic modification has been implicated in a variety of biological processes, including retrotransposon silencing, genomic imprinting, X chromosome inactivation and gene regulation. While DNA methylation is relatively stable in somatic cells and is thought to be primarily involved in long-term gene silencing, previous studies indicate that 5mC can be rapidly erased in zygotes independent of cell division, supporting the existence of enzymatic pathways that mediate active DNA demethylation.[2] Despite intensive efforts, enzymes that are responsible for active DNA demethylation in mammals have remained elusive until recently. The demonstration that human TET1 and mouse Tet proteins have the capacity to convert 5mC to 5hmC has raised the possibility that Tet proteins may be part of the enzymatic pathway for active DNA demethylation.[3,4]

TET proteins are mammalian homologs of the trypanosome base J binding proteins, JBP1 and JBP2, that contain catalytic motifs typical of Fe(II)- and 2-oxoglutarate (2OG)-dependent dioxygenases.[4,5] Through hydroxylation of the methyl group of 5mC, human and mouse Tet proteins can produce 5hmC.[3,4] While

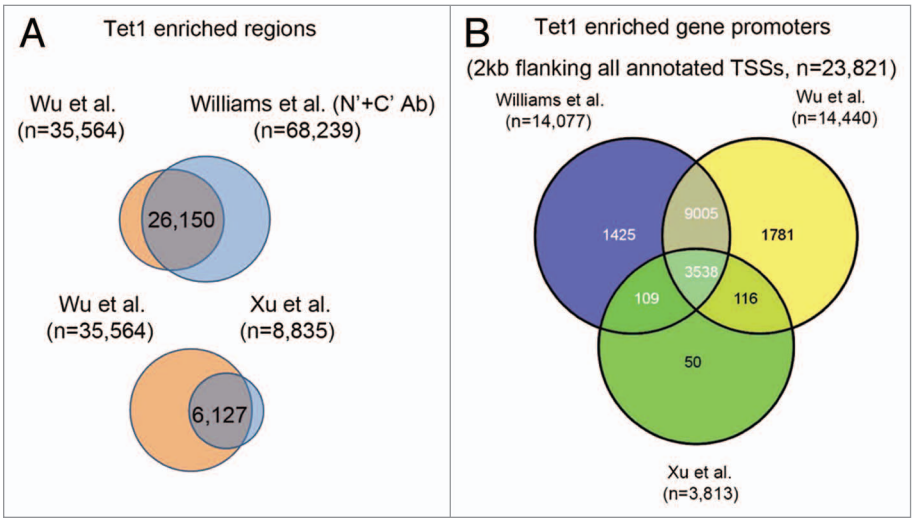

Figure 1. Comparison of Tet1 genomic binding profiles derived from different groups. (A) Pairwise comparison between lists of Tet1 binding sites identified by three independent studies. (B) Venn diagram showing the overlap of Tet1 enriched gene promoters [-2 kb to +2 kb relative to annotated transcription start sites (TSSs)] identified by three independent studies. Note that gene promoters identified by Williams et al. represent a combined list of N′ and C′ antibodies.

5hmC is present in many tissues and cell types, it is relatively enriched in ES cells and certain types of neurons.[4,6,7] Tet1 and Tet2 are highly expressed in mouse ES cells and are rapidly downregulated upon differentiation,[3,4,8] whereas Tet3 is expressed at very low levels in ES cells and upregulated during ES cell differentiation. Further analysis reveals that both Tet1 and Tet2 are downstream targets of key pluripotency factors.[8] In support of the notion that Tet proteins are critical components of a regulatory network in pluripotent stem cells, depletion of Tet proteins in mouse ES cells results in a partial loss of undifferentiated states and spontaneous differentiation towards trophoectoderm and endo/mesoderm lineages.[3,8]

To provide insight into the molecular function of Tet1 and 5hmC in pluripotent stem cells, we and others have recently undertaken genome-wide analyses of Tet1 and 5hmC distribution as well as Tet1-regulated genetic programs in mouse ES cells.[9-14] Through integrative analyses of datasets from different groups, here we reveal the common Tet1 and 5hmC targets in undifferentiated mouse ES cells. Also, we discuss the evidence that Tet1 may play a key role in orchestrating the balance between pluripotent and lineage committed states.

Genome-Wide Occupancy of Tet1 in Mouse Embryonic Stem Cells

To determine the genomic occupancy of Tet1 in ES cells, three groups independently generated specific antibodies against Tet1 and performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq).[11,13,14] Although different Tet1 antibodies were used and sequencing depth also varied between studies, pairwise comparison of the independently derived lists of Tet1 binding sites showed a high degree of overlap (**Fig. 1A**). For two studies with comparable sequencing depth of Tet1 ChIP-seq experiments, nearly 90% of Tet1 bound gene promoters (n = 12,543) are shared (Williams et al. n = 14,077 and Wu et al. n = 14,440) (**Figs. 1B and 3 and Table S1**). Visual inspection of representative Tet1 target genes also confirmed that Tet1 binding profiles determined by independent studies are highly similar (**Figs. 4 and 5**).

CpG-islands (CGIs) are DNA sequences with densely clustered CpG dinucleotides. In mammalian genomes, roughly 70% of gene promoters overlap with CGIs and are generally unmethylated, whereas the bulk genomic DNA is predominantly methylated.[15] Tet1 ChIP-seq analyses have uncovered a strong preference of Tet1 for CGI-containing gene promoters, probably due to the fact that Tet1 possesses a CXXC zinc-finger domain, which has a high affinity for nonmethylated CpG sequences.[16] The convergence of multiple CXXC domain containing epigenetic regulators such as Cfp1,[17] Kdm2a,[18] and Tet1 at CpG-rich sequences indicates that they may synergistically contribute to the establishment of a distinct chromatin environment at CGIs. In this scenario, Cfp1 confers trimethylation of lysine 4 on histone 3 (H3K4me3) by recruiting the H3K4me3 methyltransferase Setd1; Kdm2a binding results in the depletion of H3K36me2; Tet1 maintains a DNA hypomethylated state at CGIs. In support of this model, Tet1-bound CGIs are generally hypomethylated,[13] and depletion of Tet1 in mouse ES cells leads to an increase in 5mC levels at many CGIs.[11,13,14]

In addition to the presence of the transcriptionally active mark H3K4me3 at nearly all nonmethylated CGIs[15] where Tet1 is highly enriched in mouse ES cells, a subset of CGI promoters, particularly those of developmentally regulated transcription factors, are simultaneously associated with repressive histone mark H3K27me3.[19] Polycomb repression complex 2 (PRC2) catalyzes H3K27me3 at these CGI promoters and is required for repression of these genes.[20] Bivalent CGI promoters, which possess H3K4me3 as well as H3K27me3, account for approximately 20% of CpG-rich promoters in mouse ES cells.[19] Interestingly, comparing Tet1 bound genes to the list of bivalent genes indicates that a significant fraction of bivalent CGI promoters are directly bound by Tet1.[11,13] In fact, a majority of Ezh2 and Suz12 (two core subunits of PRC2) co-bound sites overlap with Tet1 binding sites.[13] Therefore, Tet1 can bind to both actively transcribed H3K4me3-only genes and PRC2-repressed CpG-rich genes (**Fig. 3**). Consequently, Tet1 binding alone cannot predict whether a gene is active or silenced.

While there is a strong positive correlation between Tet1 occupancy and CpG-density, Tet1 also binds to a subset of actively transcribed CpG-poor gene promoters, including pluripotency-associated factors such as *Nanog*, *Tcl1* and *Esrrb*[13]

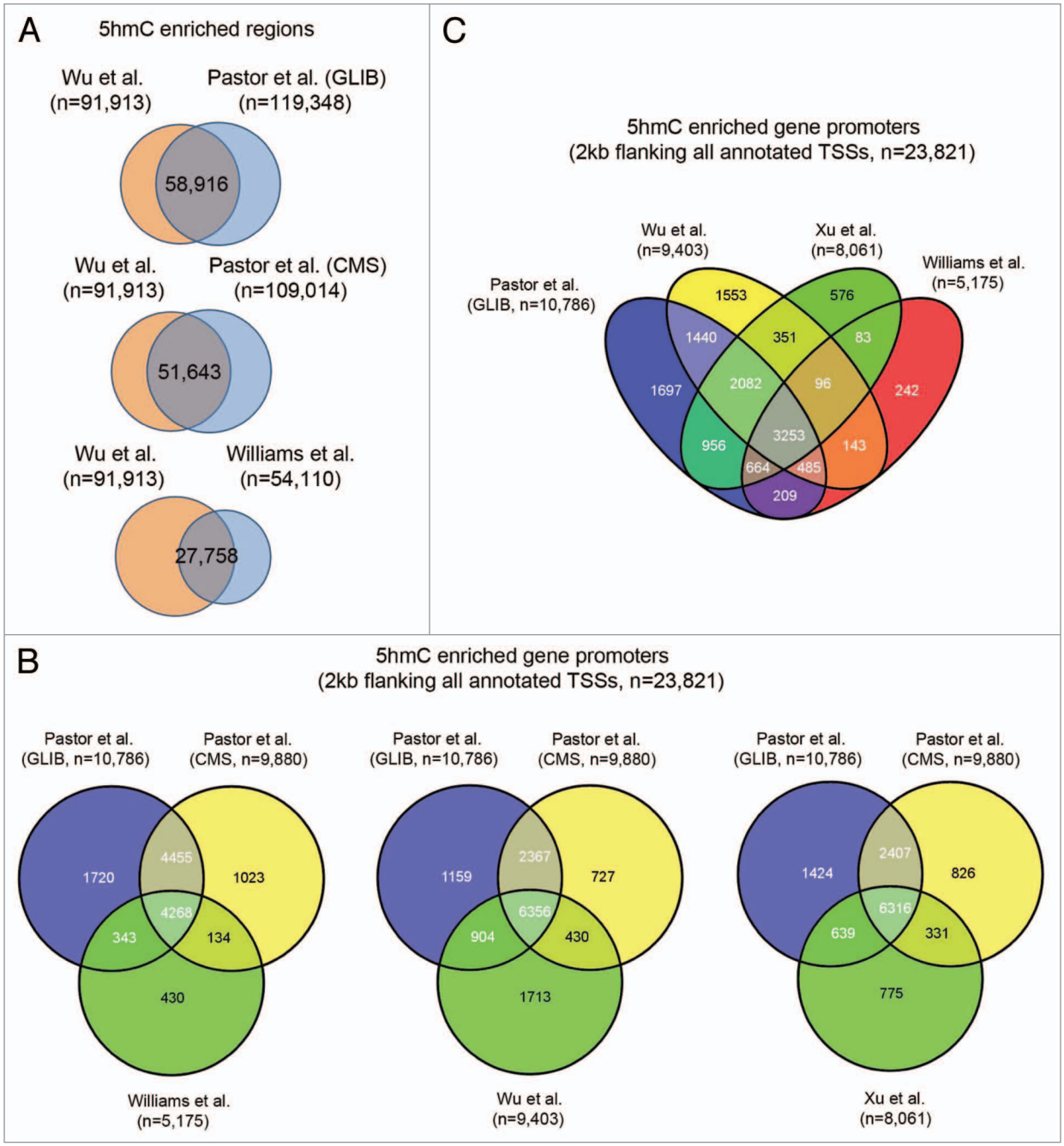

Figure 2. Comparison of 5hmC enriched genomic loci and gene promoters identified by different groups. (A) Pairwise comparison between lists of 5hmC enriched regions identified by four independent methods (Wu et al./Xu et al./Ficz et al.: 5hmC antibodies from Active Motif; Williams et al.: customized 5hmC antibodies; Pastor et al. GLIB and CMS). (B) Venn diagram showing the overlap of 5hmC enriched gene promoters (-2 kb to +2 kb relative to annotated transcription start sites) identified by 5hmC antibody-based immunpreciptation methods (Williams/Wu/Xu) and alternative methods developed by Pastor et al. (C) Venn diagram showing the overlap of 5hmC enriched gene promoters (-2 kb to +2 kb relative to annotated transcription start sites) identified by four independent studies.

(Fig. 4). On these gene promoters, Tet1 plays a crucial role in promoting a transcriptionally active state by maintaining a nonmethylated promoter status, thereby allowing efficient recruitment of transcription factors and basal transcriptional machinery.

Genome-Wide Distribution of 5hmC in Mouse Embryonic Stem Cells

Since the standard bisulphite sequencing technique cannot discriminate 5mC from 5hmC,[21,22] real-time sequencing or chemical modification strategies have recently been developed to quantify locus-specific or global levels of 5hmC in various tissues.[7,23-25] To determine the localization of 5hmC in the genome of mouse ES cells, four groups performed 5hmC immunoprecipitation using antibodies specifically

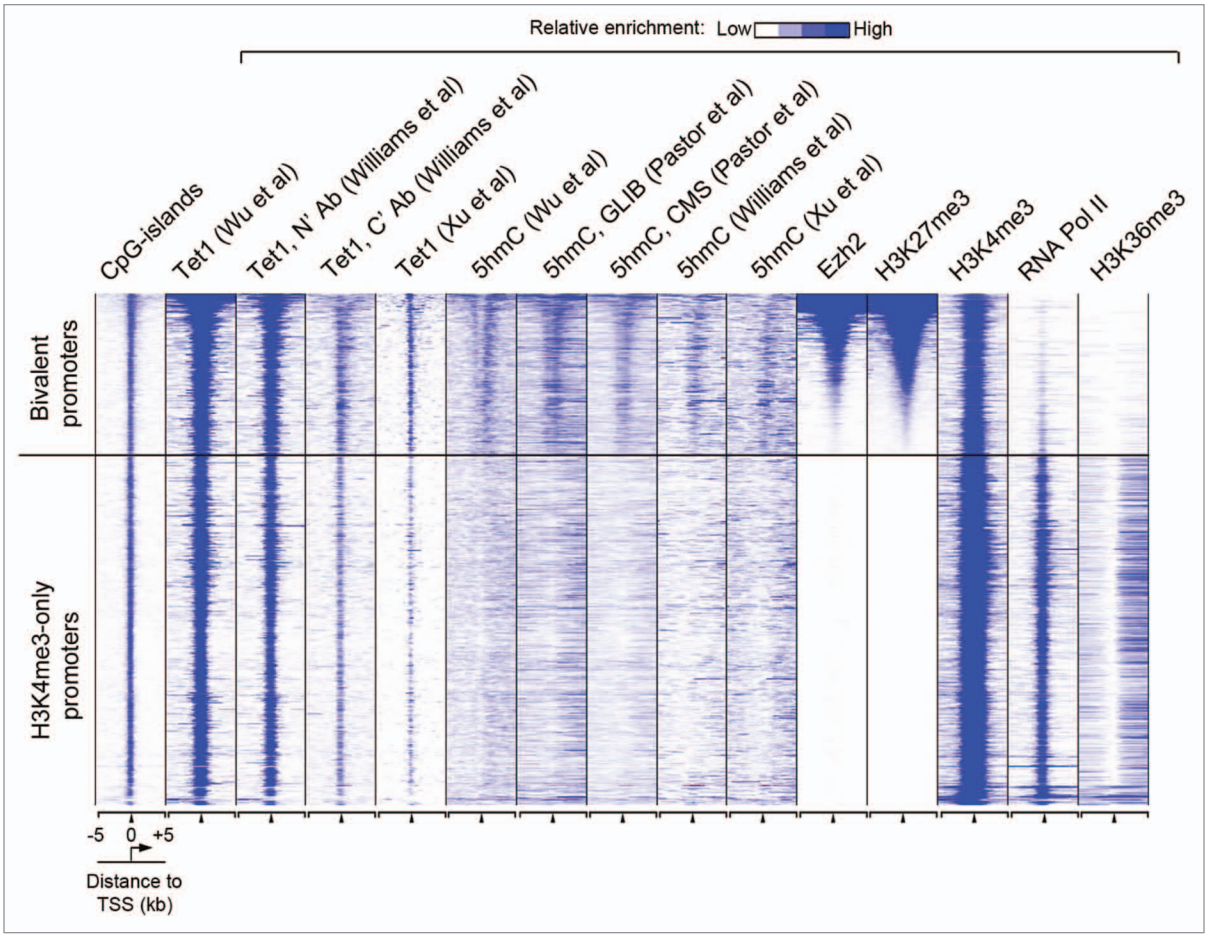

Figure 3. Enrichment of Tet1 and 5hmC at both repressed (bivalent) and actively transcribed (H3K4me3-only) genes. Heatmap representation of CGIs, binding profiles of Tet1 (Wu: GSE26833; Williams: GSE24843; Xu: GSE28500), 5hmC (Wu: GSE27613; Pastor: GSE28682; Williams: GSE24843; Xu: GSE28500) and other epigenetic regulators/marks [Ezh2 (GSE13084), H3K27me3, H3K4me3, H3K36me3 (GSE11074) and RNA Pol II (GSE12680)] in mouse ES cells in two groups of genes (5-kb flanking TSSs of bivalent and H3K4me3-only genes). The heatmap is rank-ordered from genes with highest H3K27me3 enrichment to no enrichment within 5-kb genomic regions flanking TSSs. The following color scales were used for Tet1 (Wu and Williams)/Ezh2/H3K27me3, Tet1 (Xu), 5hmC (Wu), 5hmC (GLIB, CMS, Williams and Xu) and H3K4me3/RNA Pol II/H3K36me3, respectively: (0, 200), (0, 50), (0, 2), (0, 100) and (0, 1000).

recognizing 5hmC.[9,11,12,14] Pastor and colleagues developed two alternative methods for precipitation of 5hmC.[10] The first approach, termed GLIB (glucosylation, periodate oxidation, biotinylation) uses a series of enzymatic and chemical steps to add two biotin molecules to each 5hmC; the second method uses antisera against cytosine 5-methylenesulphonate (CMS) which is produced when 5hmC reacts with sodium bisulphite. Both methods may possess enhanced sensitivity for detection of 5hmC sparsely distributed in genomic DNA.[10] These new methods have allowed for the analysis of genome-wide 5hmC distribution.

Comparative analysis of independently derived lists of 5hmC enriched sites show a substantial overlap between the different studies (**Fig. 2A**). For example, 64% of 5hmC enriched regions identified by a commercially available antibody (used by Wu et al., Ficz et al. and Xu et al.) are also independently confirmed by the GLIB method (**Fig. 2A and B**). 5hmC enriched gene promoters identified by different methods also largely overlap (7,260 targets are shared between Wu et al. n = 9,403 and Pastor et al. GLIB, n = 10,786) (**Fig. 2B and C and Table S2**). Interestingly, 5hmC modification is widely distributed across gene-rich regions, but is generally not present at repetitive sequences.[9,11,12,14] In addition, detailed sequencing analysis of strand specificity of immunoprecipitated 5hmC-containing genomic DNA indicates that strand-specific 5-hydroxymethylation is prevalent in the non-CpG context.[9]

5hmC signals obtained in wild-type ES cells were largely absent in *Dnmt* triple knockout (TKO) mouse ES cells, confirming that 5hmC is derived from the pre-existing 5mC.[9,11] In addition, 5hmC overlaps extensively with 5mC within H3K36me3-marked transcribed regions, particularly at exons.[10,12] However, many 5hmC enriched regions are devoid of 5mC. Notably, 5hmC enriched regions are frequently found at CpG-rich gene promoters, pluripotency transcription factor binding sites and insulator CTCF binding sites,[9-12] whereas 5mC is generally depleted from these gene regulatory elements,[26] consistent with the notion that DNA methylation has a negative effect on most protein-DNA interactions.

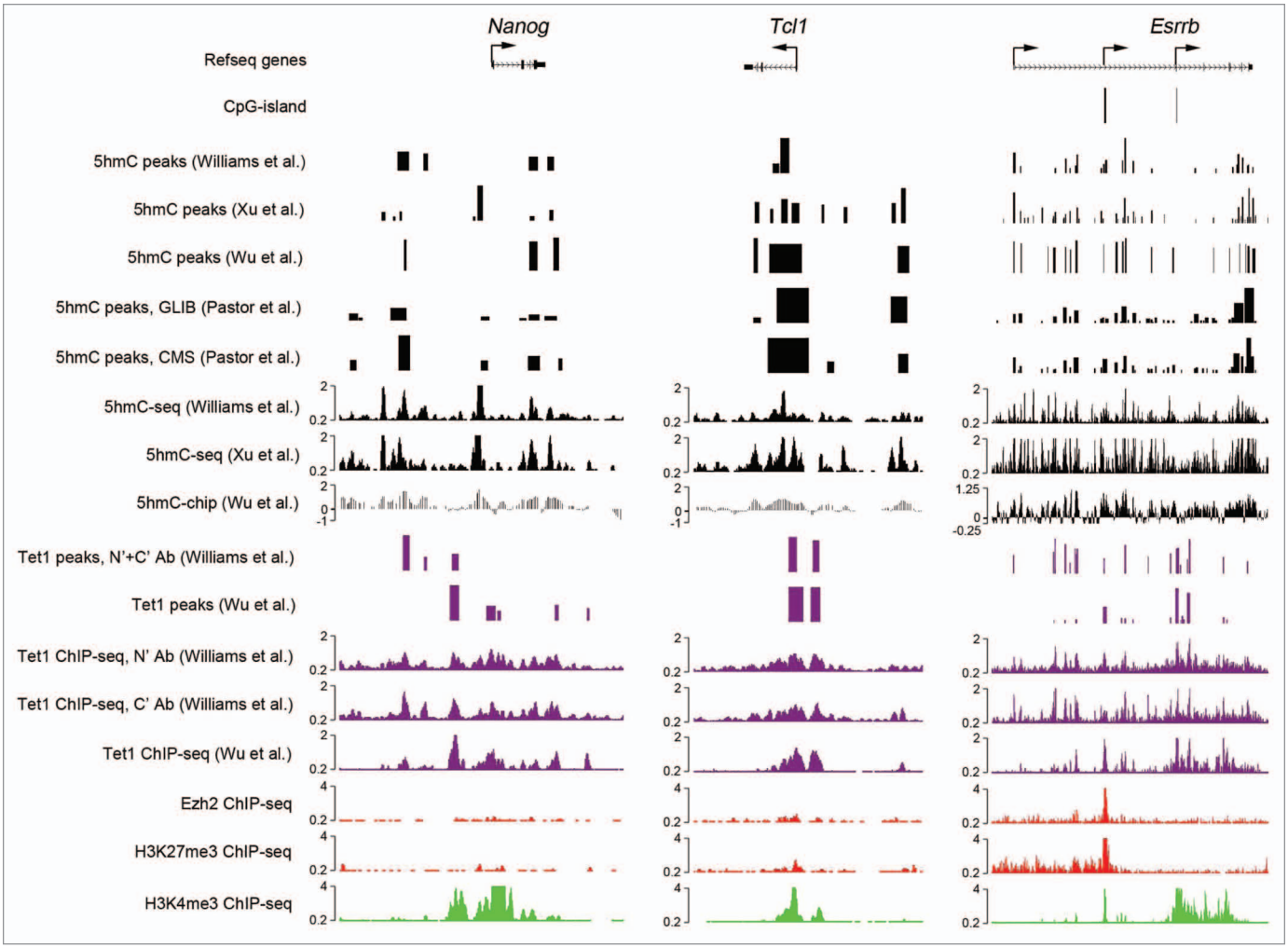

Figure 4. Enrichment of Tet1 and 5hmC at representative loci of pluripotency factors. Shown are Tet1 (purple), 5hmC (black) and other epigenetic regulator/marks (red, Ezh2/H3K27me3; green, H3K4me3) at three loci encoding pluripotency-related factors. Note that only one of three alternative promoters of the *Esrrb* gene is highly expressed in mouse ES cells. ChIP-seq data are shown in reads per million with the y-axis floor set to 0.2 reads per million. 5hmC-chip data are show in log2 ratios (IP/input). Genomic regions with statistically significant enrichment of Tet1 [$p < 10^{-8}$, measured by -log10 (peak p value)] and 5hmC ($p < 10^{-5}$ for 5hmC-seq and KS-score >2.3 for 5hmC-ChIP) are also indicated as vertical bars (Tet1: purple; 5hmC: black).

Further analysis of 5hmC distribution at CGI-containing promoters indicates that 5hmC is highly enriched at promoter regions (immediately upstream of TSSs and 5' end of gene bodies) of Polycomb-repressed genes (**Figs. 3 and 5**). In contrast, 5hmC is preferentially enriched within intragenic regions (particularly at 3' end of gene bodies) of actively transcribed, H3K4me3-only genes. Thus, while both groups of CGI-containing promoters are enriched with Tet1 and associated with low levels of 5mC, Polycomb-repressed (bivalent) and actively transcribed (H3K4me3-only) CpG-rich promoters are marked with high and low levels of 5hmC, respectively. Gene ontology analysis indicates that genes functionally related to development (e.g., lineage-specific transcription factors) are highly enriched in Polycomb-repressed genes, whereas genes involved in housekeeping functions are enriched in actively transcribed H3K4me3-only genes.[13] It is tempting to speculate that the distinct patterns of 5hmC may contribute to the establishment and/or maintenance of different chromatin structures at CpG-rich gene promoters in mouse ES cells.

Consistent with the known enzymatic activity of Tet1, 5hmC is preferentially enriched at Tet1-bound gene promoters and intragenic regions.[12] Tet1 depletion leads to a more pronounced decrease in 5hmC levels at intragenic regions (e.g., exons) than at promoter regions,[9,12] possibly due to different turnover rate of 5hmC at distinct genomic regions and/or partial functional redundancy between Tet1 and Tet2, which may also be present at Tet1 bound gene promoters.

Dual Functions of Tet1 and 5hmC in Transcriptional Regulation

The enrichment of Tet1 and 5hmC at the gene promoters suggests a role for the Tet-mediated hydroxymethylation in transcriptional regulation. Depletion of Tet1/2 leads to a decrease in expression of a cohort of genes, including pluripotency-related factors such as *Nanog*, *Esrrb* and *Tcl1*.[3,9,13] Independent genome-wide mapping datasets have confirmed that

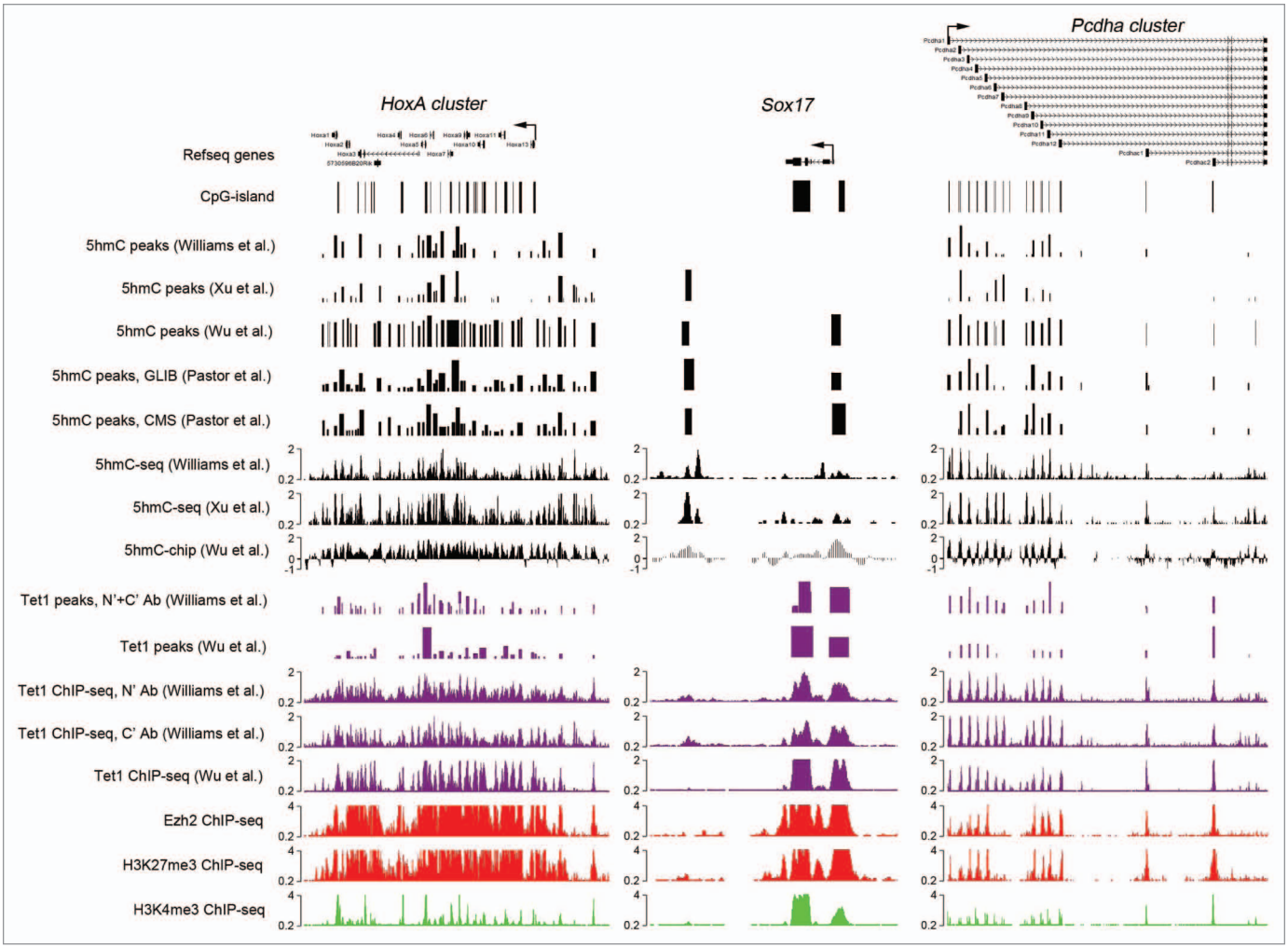

Figure 5. Enrichment of Tet1 and 5hmC at representative PRC2-repressed loci. Shown are Tet1 (purple), 5hmC (black) and other epigenetic regulator/marks (red, Ezh2/H3K27me3; green, H3K4me3) at three loci repressed by PRC2 in mouse ES cells. ChIP-seq data are shown in reads per million with the y-axis floor set to 0.2 reads per million. 5hmC-chip data are show in log2 ratios (IP/input). Genomic regions with statistically significant enrichment of Tet1 $p < 10^{-8}$, measured by -log10 (peak p values) and 5hmC ($p < 10^{-5}$ for 5hmC-seq and KS-score >2.3 for 5hmC-ChIP) are also indicated as vertical bars (Tet1, purple; 5hmC, black).

Tet1 and 5hmC are enriched at 5' gene regulatory regions of these pluripotency factors (**Fig. 3**), supporting a direct role for Tet1/2 and 5hmC in promoting transcription of a subset of pluripotency genes. In agreement with this notion, depletion of Tet1 in mouse ES cells leads to an increase in 5mC levels concomitant with decreased expression of certain pluripotency genes.[9,13] Thus, in undifferentiated mouse ES cells, Tet1, possibly in conjunction with Tet2, are required for promoting transcription of a cohort of pluripotency factors by maintaining a hypomethylated state at their promoters.

Surprisingly, gene expression microarray or RNA-seq analysis of Tet1-depleted mouse ES cells revealed that Tet1 predominantly has repressive, rather than activating, roles on its direct target genes.[10,11,13,14] Many Tet1-repressed target genes are also bound by PRC2. Although a direct interaction between Tet1 and PRC2 is not detected,[11,13] Tet1 can directly or indirectly facilitate the recruitment of PRC2 to many Tet1 target genes.[13] Recent studies indicate that DNA methylation and PRC2 are generally localized at distinct gene promoters in ES cells or cancer cells,[27,28] and high levels of 5mC may inhibit recruitment of PRC2 to chromatin.[29,30] Moreover, at PRC2-repressed target genes, high level of non-proximal promoter DNA methylation seems to be associated with increased transcription.[26,30] Thus, Tet1 may positively regulate PRC2 recruitment to chromatin, at least in part, by reducing 5mC levels at PRC2 binding sites. In addition, Tet1 may exert its repressive role on a subset of target genes by recruiting the Sin3A co-repressor complex.[11] Tet1 seems to repress these genes independent of its enzymatic activity as Tet1-depletion in TKO ES cells, which are devoid of both 5mC and 5hmC, exhibited similar changes in gene expression on these genes.[11] However, we note that this conclusion is in direct contrast with a previous study demonstrating a role of Tet1 enzymatic activity in transcriptional regulation.[16] Further studies are needed to determine whether the capacity of Tet1 in regulating gene expression requires its enzymatic activity.

Comparing independently derived lists of genes differentially expressed between control and Tet1 knockdown (KD) mouse ES cells (both studies used the E14Tg2A ES cell line) indicate a small, but statistically significant overlap (**Fig. 6A and Table S3**). Since Williams and colleagues did not observe a downregulation of Nanog in Tet1-depleted cells using distinct shRNA sequences,[11] they suggest that two Tet1 shRNA sequences used in our previous studies[3] may represent an off-target effect. To further examine this issue, we compared multiple recently published gene expression datasets of Tet1 or Tet1/2-depleted mouse ES cells.[9,11,13,14] Using the list of differentially expressed genes identified by Wu et al. (left parts in **Fig. 6B**) or Williams et al. (right parts in **Fig. 6B**) as a reference, unbiased hierarchical clustering analysis of log2 ratios of (Tet KD/Mock KD) indicates that the expression datasets of Williams et al. and other studies cluster separately (**Fig. 6B**). Since different ES cell lines and knockdown methods were employed in each study (**Table S4**), it is unlikely that the observed discrepancy between Williams et al. and other studies are due to technical issues. In addition, the demonstration by multiple groups that Tet1 and 5hmC are enriched at the Nanog promoter (**Fig. 4**), and that 90% of dysregulated genes identified in our study are Tet1 direct targets,[13] as well as the fact that Tet1 knockdown phenotypes can be partially rescued by exogeneous Nanog[3] all argue against the suggested off-target effect. Thus, it remains an open question as to which Tet1 KD shRNA is more likely to have an off-target effect. To further clarify this issue, knockdown followed by Tet1 rescue is required to definitively address the effect of Tet1/2 deficiency on gene expression networks in mouse ES cells.

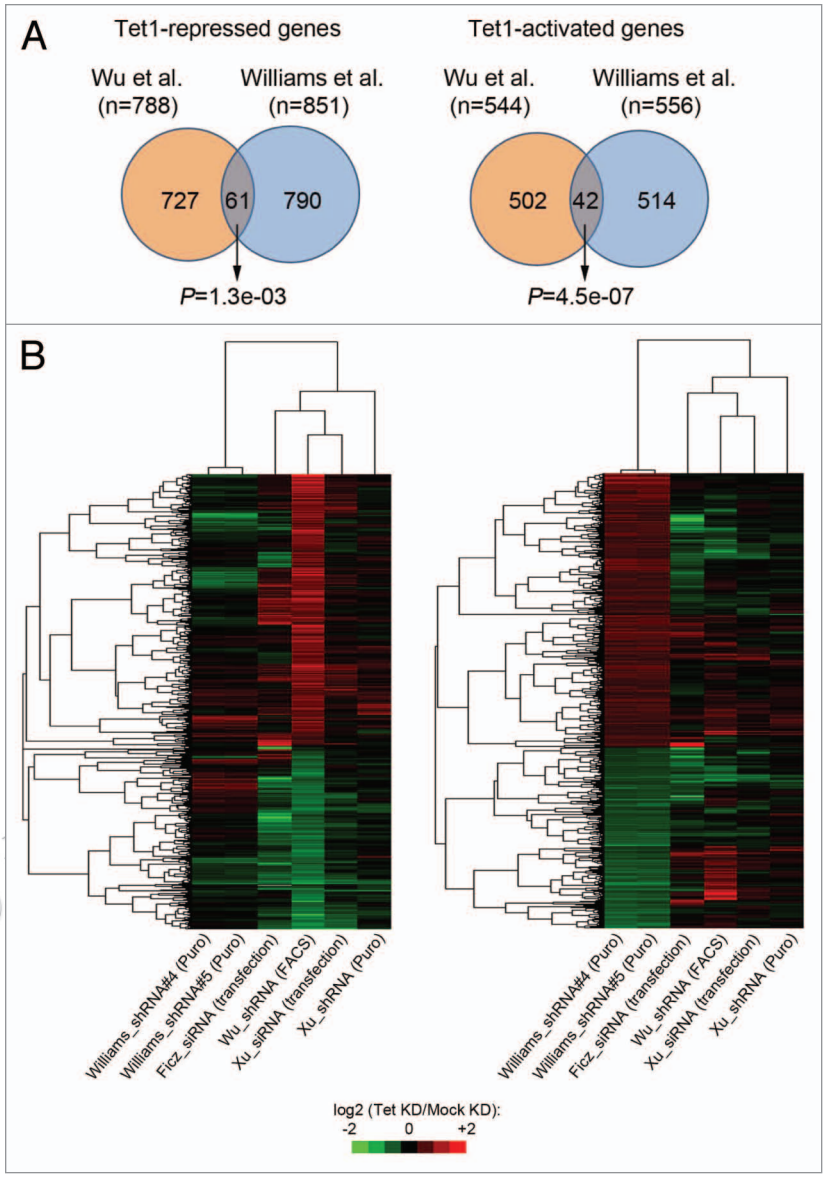

Figure 6. Comparison of differentially expressed genes in response to Tet1-depletion identified by different groups. (A) Venn diagram illustrating the overlap of independently derived lists of differentially expressed genes between wild-type and Tet1-depleted mouse ES cells. (B) Unsupervised hierarchical clustering of changes in gene expression in response to Tet-depletion determined by multiple studies. Shown on the left part is the heatmap representation based on the list of differentially expressed genes identified by Wu et al.; Shown on the right part is the heatmap representation based on the list of differentially expressed genes identified by Williams et al.

Concluding Remarks

In summary, genome-wide analysis of Tet1 and 5hmC distribution in mouse ES cells sheds new light on the molecular functions of Tet proteins in mediating dynamic changes in 5mC and ES cell differentiation. Tet1 deposits 5hmC marks at actively transcribed gene bodies as well as Polycomb-repressed CpG-rich gene promoters. Thus, depending on the context, Tet1 can associate with either active or repressed transcriptional states. Several lines of evidence suggest that this new epigenetic factor plays an important role in the establishment and/or maintenance of a pluripotent epigenetic state during early embryonic development. It is notable that dysregulation of DNA methylation at CpG-rich gene promoters is linked to tumorigenesis, e.g., DNA hypermethylation mediated inactivation of tumor suppressor genes, and that human TET2 gene is frequently mutated in a variety of haematopoietic neoplasms.[31-34] We anticipate that future functional studies will further reveal the roles of Tet proteins and 5hmC in early development and cellular homeostatsis.

Materials and Methods

Analysis of Tet1 ChIP-seq datasets. To compare Tet1 binding sites derived from different studies, BED files containing genomic coordinates of mapped sequencing reads (NCBI build 37/mm9) were downloaded from the GEO database (Williams et al.: GSE24843; Xu et al.: GSE28500). Genomic coordinates were re-mapped to the NCBI build 36 (mm8) via the LiftOver tool and re-mapped sequencing reads were analyzed by the MACS program (v1.3.7.1) using identical parameters (peak p value $<10^{-8}$ and fold enrichment over IgG >10) as previously described in reference 13. To identify Tet1 enriched gene promoters, Tet1 peaks were mapped to 2-kb genomic regions flanking non-redundant transcriptional start sites compiled from a complete set of Refseq genes that was downloaded from the UCSC Table browser (May, 2010). ChIP-seq datasets of Ezh2 (GSE13084), RNA Pol II (GSE12680), H3K4me3 and H3K27me3 (GSE11074) were obtained from previous publications and were re-analyzed in MACS (peak p value $<10^{-5}$ and fold enrichment over input >10). For visualizing ChIP-seq datasets at representative loci (**Fig. 4 and 5**), sequencing reads were binned into 400-bp windows at 100-bp steps along the genome (mm8) as previously described in reference 13. For heatmap representations in the **Figure 3**, Tet1 enriched regions (measured by $-\log_{10}$ peak p-value) were binned into 200-bp intervals at 100-bp steps within 5-kp up and downstream of TSSs of annotated RefSeq genes. The heatmap is generated and visualized using Cluster3 and Java TreeView, respectively.

Analysis of 5hmC genome-wide mapping datasets. To determine 5hmC enriched regions derived from high-throughput sequencing of genomic DNA immunoprecipitated by 5hmC antibodies, BED files containing genomic coordinates of mapped sequencing reads (NCBI build 37/mm9) were downloaded from the GEO database (Williams et al.: GSE24843; Xu et al.: GSE28500). Genomic coordinates were re-mapped to NCBI build 36 (mm8) via the LiftOver tool and re-mapped sequencing reads were analyzed by the MACS program (v1.3.7.1) using following parameters (peak p value $<10^{-5}$ and fold enrichment over IgG >10). 5hmC enriched regions reported in Wu et al. and Pastor et al. (GLIB and CMS datasets were re-mapped to mm8 first) were directly obtained from **Supplemental Tables** associated with published papers. Visualization of 5hmC distribution at representative loci and in heatmap representations was performed as described for Tet1 ChIP-seq datasets.

Analysis of gene expression profiling datasets. To compare different gene expression profiling datasets, lists of differentially expressed genes (Tet1 KD vs. Mock KD) reported in Wu et al. and Williams et al. were directly obtained from **Supplemental Tables** associated with published papers. To calculate log2 ratios of expression changes between Tet KD and Mock KD (Affymetrix gene expression data), files containing normalized probe intensity were downloaded from the GEO database (Williams et al.: GSE24842; Xu et al.: GSE28530). For calculation of log2 ratios from RNA-seq datasets (Ficz et al.: ERP000570 in EBI Sequence Read Archive), raw data (FASTQ files) were downloaded and further processed/mapped to annotated RefSeq transcripts (May, 2011; NCBI build 36/mm8) using TopHat (v1.2.0) with default parameters for single-end sequencing experiments. Normalized expression levels (measured by Fragments Per Killobase of exon model per Million mapped fragments, FPKM) for each RefSeq transcript was calculated using Cufflinks (v1.0.1). FPKM values associated with distinct transcripts from the same gene locus were combined to estimate the gene-level expression. Log2 ratios of (Tet1/2 KD vs. Mock KD) were calculated with averaged FPKM values for each non-redundant gene locus in Mock and Tet1/2 KD cells. The complete linkage hierarchical clustering (in **Fig. 6B**) was performed in Cluster 3 and visualized in Java TreeView.

Acknowledgments

We thank Susan Wu for critical reading of the manuscript. This work was supported by NIH grants GM68804 (to Y.Z.). H.W. is supported by a Jane Coffin Childs postdoctoral fellowship. Y.Z. is an Investigator of the Howard Hughes Medical Institute.

Note

Supplemental materials can be found at: www.landesbioscience.com/journals/cc/article/16930

References

1. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet 2008; 9:129-40.
2. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol 2010; 11:607-20.
3. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 2010; 466:1129-33.
4. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009; 324:930-5.
5. Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle 2009; 8:1698-710.
6. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 2009; 324:929-30.
7. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 2010; 38:181.
8. Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, et al. Tet1 and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 2011; 8:200-13.
9. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 2011; 473:398-402.
10. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 2011; 473:394-7.
11. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 2011; 473:343-8.
12. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev 2011; 25:679-84.
13. Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 2011; 473:389-93.
14. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide Regulation of 5hmC, 5mC and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Mol Cell 2011; 42:451-64.
15. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev 2011; 25:1010-22.
16. Zhang H, Zhang X, Clark E, Mulcahey M, Huang S, Shi YG. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Res 2010; 20:1390-3.
17. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 2010; 464:1082-6.

18. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ. CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell 2010; 38:179-90.
19. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006; 125:315-26.
20. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 2002; 298:1039-43.
21. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One 2010; 5:8888.
22. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 2010; 38:125.
23. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 2010; 7:461-5.
24. Robertson AB, Dahl JA, Vagbo CB, Tripathi P, Krokan HE, Klungland A. A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 2010; 39:55.
25. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 2011; 29:68-72.
26. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462:315-22.
27. Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, et al. 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 2008; 2:160-9.
28. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc Natl Acad Sci USA 2008; 105:12979-84.
29. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell 2010; 143:470-84.
30. Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science 2010; 329:444-8.
31. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, et al. Mutation in TET2 in myeloid cancers. N Engl J Med 2009; 360:2289-301.
32. Tefferi A, Lim KH, Levine R. Mutation in TET2 in myeloid cancers. N Engl J Med 2009; 361:1117.
33. Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 2009; 41:838-42.
34. Mullighan CG. TET2 mutations in myelodysplasia and myeloid malignancies. Nat Genet 2009; 41:766-7.