

Published in final edited form as:

Nature. 2011 May 19; 473(7347): 343–348. doi:10.1038/nature10066.

Tet1 and hydroxymethylcytosine in transcription and DNA methylation fidelity

Kristine Williams^{1,2,*}, Jesper Christensen^{1,2,*}, Marianne Terndrup Pedersen^{1,2,*}, Jens V. Johansen^{1,3}, Paul A. C. Cloos^{1,2}, Juri Rappsilber⁴, and Kristian Helin^{1,2}

¹Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark

²Centre for Epigenetics, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark

³The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark

⁴Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, EH9 3JR, UK

Abstract

Enzymes catalysing the methylation of the 5-position of cytosine (mC) have essential roles in regulating gene expression and maintaining cellular identity. Recently, TET1 was found to hydroxylate the methyl group of mC, converting it to 5-hydroxymethyl cytosine (hmC). Here we show that TET1 binds throughout the genome of embryonic stem cells, with the majority of binding sites located at transcription start sites (TSSs) of CpG-rich promoters and within genes. The hmC modification is found in gene bodies and in contrast to mC is also enriched at CpG-rich TSSs. We provide evidence further that TET1 has a role in transcriptional repression. TET1 binds a significant proportion of Polycomb group target genes. Furthermore, TET1 associates and colocalizes with the SIN3A co-repressor complex. We propose that TET1 fine-tunes transcription, opposes aberrant DNA methylation at CpG-rich sequences and thereby contributes to the regulation of DNA methylation fidelity.

Introduction

The majority of CpGs in mammalian genomes are methylated. An exception to this is CpG islands, which are found in more than 60% of all mammalian gene promoters. These are often unmethylated and can be either transcriptionally active or inactive depending on other factors, including histone modifications and the activity of cell-type-specific transcription factors^{1, 2, 3, 4, 5}. In current models for gene regulation, CpG methylation in promoters leads to stable gene silencing, whereas the function of intragenic methylation might, like

Correspondence: kristian.helin@bric.ku.dk (K.H.).

*These authors contributed equally to this work.

Contributions K.W. performed the major part of experiments in Figs 1, 3, 4a, b, h and Supplementary Figs 1a–c, 2, 3, 5, 7a, 9a–d, 10a, 11 and 12c, d. J.C. developed and characterized the new reagents used in this study, and participated in most experiments. M.T.P. performed the major part of experiments in Figs 2, 4c, g and Supplementary Figs 1d, 6b, 7b, c, 8, 10b and 12a, b. J.V.J. performed bioinformatics analyses. P.A.C.C. assisted in characterizing reagents. J.R. performed the mass spectrometry analysis. J.C. and K.H. supervised the project and all authors contributed to the writing of the manuscript.

Accession codes Primary accessions: Gene Expression Omnibus <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24843>

Competing financial interests K.H., J.C. and P.A.C.C. are cofounders of EpiTherapeutics and have shares and warrants in the company. All other authors declare that they have no competing financial interests.

trimethylation of histone 3 lysine 36 (H3K36me3), repress the initiation of intragenic transcription⁶.

DNA methyltransferases are essential for embryogenesis, and the methylation pattern of the mammalian genome undergoes major changes during development. As an example, global waves of DNA demethylation and remethylation take place after fertilization, and gene-specific de novo methylation occurs during differentiation of embryonic stem (ES) cells^{6, 7}. Importantly, patterns of DNA methylation are perturbed in human diseases such as imprinting disorders and cancer⁸. So far there is very limited knowledge regarding the mechanisms leading to DNA hypermethylation of CpG-island promoters in cancer, and how CpG-islands generally remain unmethylated in somatic cells.

Enzymes contributing to DNA demethylation could potentially provide a fidelity system for DNA methylation, but such enzymes were not known until recently. In a ground-breaking paper, TET1 was shown to catalyse the hydroxylation of mC9, which has led to the proposal of several models for how TET1 and hmC may contribute to DNA demethylation and gene regulation. One possibility is that hydroxylation of mC by TET1 might interfere with DNMT1 activity, leading to a subsequent passive loss of methylation following replication. Alternatively, hmC may be converted to cytosine through hitherto unknown enzymatic mechanisms. In addition, hydroxylation of mC may promote transcriptional de-repression by dissociation of mC-binding proteins and/or recruitment of effector proteins. The demonstration that hmC is highly abundant in ES cells and in neuronal Purkinje cells indicates that this modification is stably present in the mammalian genome and that it might be important for gene regulation^{9, 10}.

TET1 binds CpG-rich transcription start sites

TET1 is highly expressed in mouse ES cells and is rapidly downregulated during their differentiation^{9, 11}. To obtain more information regarding the function of TET1, we inhibited TET1 expression in mouse ES cells using two different shRNA constructs (Fig. 1a and Supplementary Fig. 1a). The efficient knockdown of Tet1 did not lead to any change in proliferation rate or expression of NANOG and OCT4 (Fig. 1a and Supplementary Fig. 1a, b). These data are in agreement with a recently published study¹², but in contrast to results reported by others¹³. We also observed inhibition of growth and decreased levels of NANOG in mouse ES cells when using the Tet1 shRNA sequences published in the latter study (Supplementary Fig. 1c, d). However, as these shRNA sequences do not lead to greater knockdown efficiency than the ones we have used (Supplementary Fig. 1c), it is possible that shRNA off-target effects could cause the observed phenotype.

We determined the genome-wide location of TET1 by using two different antibodies to TET1 (Tet1-N and Tet1-C) for chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq). These experiments were performed in control or TET1-depleted mouse ES cells. The two TET1 antibodies were highly specific as shown in the examples provided in Fig. 1b and by the fact that 97–99% of the identified TET1 binding sites were not found in the TET1-depleted cells (Supplementary Fig. 2a). The majority of TET1 binding sites were found in gene bodies, with the highest density around TSSs (Fig. 1c). Gene annotation of TET1 binding sites, using a false discovery rate (FDR) < 0.01, showed that TET1 binds in the vicinity of the TSS of 6,573 genes (Fig. 1d and Supplementary Table 1), of which all tested so far have been independently validated by ChIP followed by real-time quantitative PCR (ChIP-qPCR, Supplementary Fig. 2b and data not shown). Peak detection analysis using FDR < 0.1 indicates that TET1 could have up to 9,241 target genes (Supplementary Fig. 3a). Gene Ontology analysis showed that TET1 target genes are involved in a variety of basic cellular processes, and in more specific processes such as

development and differentiation (Supplementary Fig. 3b). The majority of the TET1 target genes are associated with high and intermediate density CpG promoters (HCPs and ICPs, Fig. 1e), which are positive for H3K4me3 (Fig. 1f). The correlation between TET1 binding and high CpG density is also found outside of TSSs (Supplementary Fig. 4). Interestingly, TET1 binding does not predict whether a promoter is active, poised for activation (non-productive) or inactive (Fig. 1g). In agreement with this, we found that a significant fraction of TET1 was associated with promoters containing the H3K27me3 Polycomb repressive mark (Fig. 1f). Indeed, independent analysis showed a highly significant overlap of genes bound by TET1 and the Polycomb group (PcG) protein, SUZ12, in ES cells (Supplementary Fig. 5a, b).

hmC is enriched at TSSs and gene bodies

To gain information regarding a possible function of hmC, we generated an affinity-purified polyclonal antibody to hmC that binds with high specificity and sensitivity to this mark, as shown by enzyme-linked immunosorbent (ELISA) and DNA immunoprecipitation (DIP) assays (Supplementary Fig. 6). Genome-wide DIP-seq assays were performed using anti-hmC, anti-mC and IgG on genomic DNA purified from control or TET1-depleted ES cells as well as from Dnmt triple knockout (TKO) mouse ES cells, lacking Dnmt1, Dnmt3a and Dnmt3b14. We confirmed by ChIP-qPCR that TET1 localizes to its target genes in the Dnmt TKO cells (Supplementary Fig. 7a). The analyses showed that hmC is located as discrete peaks throughout the genome (Fig. 2a). Furthermore, the majority of signals obtained with the hmC antibody were absent in Dnmt TKO mouse ES cells, confirming that generation of hmC requires the pre-existence of mC (Fig. 2a). The hmC modification in mouse ES cells is particularly enriched within gene bodies as also observed for the mC mark¹⁵ and recently reported for hmC in mouse cerebellum¹⁶ (Fig. 2b, c). Strikingly, in contrast to the localization of mC, hmC is also significantly enriched at the TSS coinciding with TET1 (Fig. 2c), indicating that a significant fraction of mC is converted to hmC at the TSS. Also, the hmC modification is generally not detectable at repetitive elements such as intracisternal A particle (IAP) elements and minor satellite repeats by DIP-qPCR (Supplementary Fig. 7b), further demonstrating that hmC and mC show distinct genomic distributions.

Gene annotation of hmC positive regions around the TSS (−0.7 kilobases to +0.3 kb) showed that 2,424 regions are hmC-positive in wild-type ES cells compared to Dnmt TKO ES cells. Approximately 28% of these regions showed a more than twofold reduction in hmC signal in the DIP-seq analyses upon downregulation of TET1 (Fig. 2d) and in validation experiments the knockdown of Tet1 led to a significant decrease in hmC levels on tested genes (Fig. 2e and data not shown). Depending on the used false discovery rate cut-off for TET1, between 35% (FDR < 0.01) and 50% (FDR < 0.1) of hmC-positive genes are bound by TET1 (Fig. 2f). These results are in agreement with reports showing that Tet1 knockdown only causes a partial decrease in global hmC levels in mouse ES cells^{9, 12}, and imply that, although TET1 is important for the generation of hmC, other enzymes such as TET2 are also likely to contribute to hmC levels in mouse ES cells.

As for TET1, Gene Ontology analysis of the hmC-positive genes showed enrichment for genes involved in basic cellular processes, but also in the regulation of development and differentiation (Supplementary Fig. 7c). Moreover, hmC positivity does not correlate with transcriptional activation and surprisingly, most hmC-positive genes seem not to be expressed in mouse ES cells (Fig. 2g).

A significant proportion of the TSSs classified as positive for hmC has intermediate or high CpG content (Fig. 2h and Supplementary Fig. 4). Genome-wide analyses of the hmC distribution relative to CpG content showed that the hmC mark is enriched in regions with

relatively high CpG content compared to mC (Fig. 2i). Whereas only 15% of hmC-positive TSSs also contain a high mC signal, we find that several hmC-positive regions have low levels of mC, implying that the two marks often co-exist. Upon Tet1 knockdown only a minor global increase in mC was observed as evaluated by genome-wide anti-mC DIP (Me-DIP) (Supplementary Fig. 8a). However, a few hundred genes show modest TSS specific increases in mC levels after Tet1 knockdown (Supplementary Fig. 8b). Gene Ontology analyses for these genes showed enrichments for specialized developmental processes (Supplementary Fig. 8c). Interestingly, we found that approximately a third of the genes reported to acquire DNA methylation during ES cell differentiation^{2, 3} are marked by hmC in the ES cell state (Supplementary Table 2). Taken together, these results show that hmC colocalizes with mC in gene-bodies, and that hmC, in contrast to mC, is enriched at TSSs with intermediate to high CpG density, where it may contribute to the regulation of DNA methylation patterns.

TET1 contributes to transcriptional repression

To understand how TET1 contributes to the regulation of target genes, we performed genome-wide expression analyses of mouse ES cells expressing two different Tet1 shRNAs or a scrambled shRNA (Supplementary Fig. 9a, b and Supplementary Table 3). As shown in Fig. 3a and Supplementary Fig. 9c, we observed a significant decrease in expression of 556 genes and a significant increase in expression of 851 genes common to both shRNAs. Of these approximately 700 were direct target genes of TET1, and therefore only around 10% of all TET1 target genes change expression following Tet1 knockdown. Whereas we expected to observe a significant fraction of the downregulated genes to be direct targets for TET1, we were surprised to find that an even higher fraction of the upregulated genes were associated with TET1 (Fig. 3a). To validate these results, we performed qPCR analysis of a number of downregulated and upregulated genes (Fig. 3b) that were also directly bound by TET1 (Supplementary Fig. 2b). Moreover, several of the identified targets show similar expression change upon differentiation of mouse ES cells by retinoic acid, which leads to decreased levels of TET1 (Supplementary Fig. 9d).

To investigate whether the transcriptional effects of TET1 are mediated by modulating hmC and mC levels, we performed knockdown of Tet1 in Dnmt TKO cells (Supplementary Fig. 10a). We found that all the tested transcriptional effects by knockdown of Tet1 were similar in Dnmt TKO and normal ES cells (Fig. 3c and Supplementary Fig. 10b), indicating that the effects are independent of catalytic activity. However, we cannot rule out that TET1-dependent modulation of hmC and mC might contribute to transcriptional fine-tuning at some target genes. Taken together, these results indicate that TET1 can contribute to transcriptional repression, and to a minor extent also transcriptional activation, and that the majority of TET1-mediated transcriptional effects are independent of conversion of mC to hmC.

TET1 associates with the SIN3A complex

The mechanism by which TET1 contributes to transcriptional repression is unknown. Although we find an extensive overlap between TET1 and PcG target genes, we have not been able to detect a physical interaction of TET1 with PcG proteins. Therefore, we purified proteins associated with double-epitope Flag-haemagglutinin (Flag-HA)-tagged TET1 expressed in HEK293 cells. This purification led to the identification of SIN3A and several other core components of the SIN3A co-repressor complex, which we did not find associated with the TET2 hydroxylase (Fig. 4a). The SIN3A co-repressor complex is thought to contribute to transcriptional repression by mediating histone deacetylation¹⁷. We validated the interaction between SIN3A and TET1 *in vivo* by co-immunoprecipitation of

endogenous proteins with and without the DNA intercalating agent ethidium bromide (Fig. 4b and Supplementary Fig. 11a). Furthermore, TET1 expressed as a fusion protein with the GAL4 DNA binding domain was sufficient to recruit SIN3A to the GAL4 DNA binding sites in vivo (Supplementary Fig. 11b–e).

To understand if SIN3A also colocalizes with TET1 on target genes, we performed ChIP-seq analysis using two different commercial antibodies to SIN3A (Fig. 4c, Supplementary Table 1). This analysis showed that SIN3A has a similar binding profile as TET1 (Fig. 4d, e and Supplementary Fig. 4), and that TET1 and SIN3A display a significant overlap of target genes (Fig. 4f and Supplementary Fig. 12a). Moreover, ChIP experiments showed that TET1 contributes significantly to the recruitment of SIN3A (Fig. 4g), whereas depletion of SIN3A had no or modest effect on TET1 binding to tested target genes (Supplementary Fig. 12b). To understand if SIN3A is required for the silencing of TET1 repressed genes, we performed gene expression analysis of Sin3A knockdown cells (Supplementary Fig. 12c and Supplementary Table 4). Here we found an extensive overlap between genes with increased expression after Tet1 and Sin3A knockdown that are also directly bound by both TET1 and SIN3A (Supplementary Fig. 12d). This implies that SIN3A is required for the repression of a subset of TET1 target genes that show increased expression upon TET1 downregulation (Fig. 4h).

Discussion

One of the major findings presented in this paper is that TET1 localizes to gene bodies and TSSs of a large number of genes and is particularly enriched on genes with high CpG-content. In contrast to the global pattern of mC, which is found predominantly in low CpG density regions, we found that hmC colocalizes with TET1 at high and intermediate CpG-content sequences. This finding indicates that TET1 could have an important role in the metabolism of mC at CpG-rich sequences by converting it to hmC. Statistically significant hmC levels were not detected around the TSS at the majority of TET1 target genes. It is possible that these genes are not methylated and therefore cannot be subsequently hydroxymethylated. Alternatively, it is tempting to speculate that low and stochastically placed methylations on these CpG-rich genes are passively eliminated through replication in rapidly dividing ES cells, following TET1-mediated hydroxylation. If so, the generated hmC will most likely not be detected by DIP-analyses because it will only occur in few cells in the total cell population. In this way the role of TET1 would be to remove aberrant stochastic DNA methylation and contribute to regulating DNA methylation fidelity in ES cells. However, we also found a large number of hmC-positive genes and, interestingly, many of these become hypermethylated in differentiated cells, for example, *Dazl*, *Hormad1*, *Sycp1* and *Sycp2* (ref. 2; Supplementary Table 2 and data not shown). This suggests a dual biological role of TET1, one in which it removes aberrant DNA methylation and another that ensures the timely DNA methylation and silencing of target genes during differentiation.

We also provide evidence that TET1 has a role in transcriptional repression. Interestingly, downregulation of TET1 in Dnmt TKO ES cells leads to upregulation of the same genes as observed in wild-type ES cells, indicating that the repressive function of TET1 is independent of its catalytic activity. We found that TET1 interacts with the SIN3A complex and the extensive colocalization of TET1 and the SIN3A co-repressor complex at target genes suggests that SIN3A has an important function in TET1-mediated gene repression.

In summary, our results indicate that TET1 is required for the timely expression of genes during development. We propose that TET1 by converting mC to hmC serves an important function in the regulation of DNA methylation fidelity. In turn this conversion may lead to a

reduction of DNA methylation at CpG-rich gene regulatory sequences. Thus, loss of function of the TET proteins would promote the stochastic hypermethylation of promoters leading to deregulation of transcription and differentiation. Interestingly, the related TET2 oxygenase is frequently mutated in a variety of haematopoietic neoplasms supporting an important role of conversion of mC to hmC in cellular homeostasis 18, 19.

Methods

Cell culture

Low passage (p17) E14TG2a.4 feeder independent ES cells were grown on 0.1% gelatin-coated plates in Glasgow medium (Sigma) supplemented with glutamine (Gibco), nonessential amino acids (Gibco), sodium pyruvate (Gibco), 50 μ M β -mercaptoethanol, and 15% fetal bovine serum (HyClone) in the presence of leukaemia inhibitory factor (LIF). Recombinant lentiviruses encoding Tet1 and Sin3A shRNA were produced by standard methods employing co-transfection of pLKO.1 shRNA and packaging vectors in 293FT cells. shRNA-transduced ES cells were selected 36 h post transduction with 2 μ g per ml of puromycin for 72 h. For Sin3A knockdown, cells were harvested after 48 h to minimize differentiation. Tet1 shRNAs had the following sequences, shTet1#3: 5'-tgtagaccatcactgttcgac-3', shTet1#4: 5'-tcactactctcacctagtg-3', shTet1#5: 5'-agagaacctggtgcacagat-3', shTet1#A: 5'-gcagatggccgtgacacaaat-3' and shTet1#B: 5'-gctcatggagactaggttgg-3'. Sin3A shRNA had the following sequence, shSin3A#73: 5'-gctgtccgattgccttaaa-3'.

Cloning procedures

The open-reading frames (ORF) of mouse Tet1 and Tet2 were amplified by PCR using cDNA from mouse ES cells or LPS-stimulated RAW264.7 mouse macrophages as template, respectively. The amplified fragments were cloned into the pCR8/GW gateway entry vector (Invitrogen), and the DNA sequence was verified by sequencing. Coding errors according to the GenBank reference sequences of mouse Tet1 and Tet2 were corrected by site-directed mutagenesis. To generate expression vectors, the appropriate entry clones were transferred into gateway-compatible pCDNA5 TO Flag-HA. shRNA constructs targeting Tet1 were constructed in pLKO.1. shRNAs targeting murine Sin3A were obtained from Sigma-Aldrich.

Generation of antibodies to mouse TET1 and hydroxymethylcytosine

Polyclonal antibodies were generated by immunizing rabbits with affinity-purified bacterially expressed GST-Tet1-N (amino acids 1–308) and GST-Tet1-C (amino acids 1739–2039). The antibodies were absorbed on GST-coupled cyanogen bromide-activated Sepharose (GE Healthcare) and subsequently affinity purified using Sepharose coupled with GST-Tet1-N or GST-Tet1-C. Antibody specificity was confirmed by immunoblotting and immunoprecipitation. To generate antibodies against hydroxymethylcytosine, 5-hydroxymethylcytidine (Berry & Associates), was covalently coupled to BSA essentially as described²⁸ and used for immunization of rabbits. Affinity-purified anti-hydroxymethylcytosine (hmC) antibodies were produced by column absorption of the rabbit antisera on methylcytidine-ovalbumin coupled to cyanogen bromide-activated Sepharose followed by column-affinity purification on hydroxymethylcytidine-ovalbumin coupled to Sepharose. The antibodies were eluted with 0.1 M glycine-HCl, neutralised, dialysed against PBS and stored at -80°C . The specificity of the purified anti-hmC antibodies were analysed by ELISA and in hme-DIP assays. For the hme-DIP assays, synthetic 300-base-pair probes incorporating 5, 20 and 100% hmC or mC, respectively, were amplified by PCR using pCR8/GW (nucleotides 701–1000) as template. The probes (0.001 ng) were spiked into the

hmeC/meC reactions containing 1 μg of sonicated ES DNA. Antibody reactivity with the probes was detected by qPCR.

Purification of TET1 and TET2 complexes

To isolate TET1 and TET2-containing complexes, two-step affinity purification was performed followed by mass spectrometry analysis. Nuclear extracts (250–500 mg, 3×10^9 cells) from Flp-In-T-REx-293 cell lines expressing Flag–HA-tagged murine TET1 or TET2 were precleared and incubated with a 700 μl packed volume of anti-Flag beads (anti-Flag M2-agarose, Sigma) overnight at 4 °C with rotation. The beads were collected by centrifugation at 700g for 5 min and washed six times with 40 \times resin bed volume of buffer A (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 10% glycerol, 0.2 mM PMSF, 1 mM DTT, 1 $\mu\text{g ml}^{-1}$ aproVnin and 1 $\mu\text{g ml}^{-1}$ leupepVn). The beads were transferred into a 10-ml poly-prep chromatography column (Bio-Rad) and complexes were then eluted five times after 10 min of incubation using one resin bed volume of buffer A supplemented with 0.5 $\mu\text{g ml}^{-1}$ Flag pepVde. The eluate was subjected to a second round of purification using an antibody against the HA-tag. The Flag-IP elute was incubated with 200 μl of a 50% slurry of HA-beads overnight. The beads were washed four times with buffer A and eluted with 100 μl buffer A supplemented with 1 $\mu\text{g ml}^{-1}$ HA pepVde for 2 h. The samples were boiled in SDS loading buffer and run shortly into a SDS–PAGE gel in order to remove the Flag and HA peptide and other contaminations. A gel slice containing the purified proteins was isolated for mass spectrometry analysis.

ChIP/DIP assays and ChIP/DIP-seq

Chromatin immunoprecipitation assays (ChIP) were performed and analysed as previously described²¹. The antibodies used were anti-mSin3A (Abcam AB3479, Santa Cruz sc-994X) and the antibodies to TET1 described above. ES cell DNA was sonicated to an average size between 300 and 600 bp. Adaptor-ligated libraries for hmC or mC DNA immunoprecipitations assays (hm-DIP/me-DIP) were constructed using the NEBNext DNA Sample Prep Master Mix, NEB combined with Illumina adaptors. hme/me-DIP assays were performed as described²² using 1 μg of denatured sonicated or adaptor-ligated DNA in 100 μl of binding buffer and 0.1–4 μg of affinity-purified rabbit hmC antibody or monoclonal mC antibody (Eurogentec BI-MECY-0500). The samples were incubated for four hours at 4 °C before addition of 10 μl of anti-rabbit/mouse Dynabeads (Invitrogen). After 2 h of incubation, the samples were washed four times and bound DNA was eluted by incubation for one hour at 55 °C in 100 μl of 50 mM Tris-HCl, 10 mM EDTA, 0.5% SDS and 20 μg proteinase K. The DNA was purified using a QIAquick PCR purification kit (Qiagen) and amplified by 16 cycles of PCR. For the MeCAP (methylated DNA capture by affinity purification) experiments, the MethylCap kit (Diagenode) was used according to manufacturer's instructions. For ChIP-seq analysis, the DNA obtained from the ChIP assays were adaptor-ligated and amplified using a kit from Illumina (IP-102-1001). The amplified DNA from hme/me-DIP or ChIP-seq experiments was analysed by Solexa/Illumina high-throughput sequencing. After prefiltering the raw data by removing sequenced adapters and low quality reads, the tags were mapped to the mouse genome (assembly mm9) with the Bowtie alignment tool. To avoid any PCR bias we allowed only one read per chromosomal position (unless otherwise specified) thus eliminating spurious spikes. Peak detection were performed in the CisGenome program²³ at an FDR cut-off value <0.1 or <0.01 as indicated in the text. IgG was used as control for normalization. Venn diagram analysis was performed with Galaxy browser (www.galaxy.psu.edu). Most standard peak detection programs are typically optimized for transcription binding site data and anticipate a defined narrow bell-shaped density profile. However, for epigenetics data, such as mC and hmC, the peaks tends to be broad and low-intensity, thus requiring a different peak detection program. We used the MEDIPS tool²⁴ (bin size = 50, fragment length = 250, frame size = 500, step =

250) to detect significant enrichment of signal (reads per million, rpm) relative to a control (Dnmt TKO DIP) and an input (IgG DIP) sample at an FDR cut-off value <0.1 and a minimal enrichment of ratio >5 . For the MEDIPS analysis the reads were not limited to one read per chromosomal position, and the total length of the mapped reads were extended in the direction of the 3'-end to a total length of 250 bases, which was our estimate of the mean fragment length. Chromosomal positions (peaks) were annotated to the RefSeq database (mm9) using the UCSC "refFlat" table29. Genes not uniquely mapped to the genome were excluded. Signal vs CpG plot: for the signal vs CpG plots the MEDIPS calculated rpm and CpG (CpG values from transformed "coupling" factors) values were used. To avoid redundancy only the longest transcript variant of each gene was used to define chromosomal locations of promoter, TSS, exons and introns. For each bin (non-overlapping) MEDIPS determines the number of overlapping reads and the CpG content. For a specified region of interest (ROI), for example, an exon, the mean rpm and CpG content of the bins within the range was calculated. The distribution of CpG content within the different genomic categories are distinctly different for example, with the TSS region showing the known bimodal distribution. To depict the rpm as a function of CpG-content, the mean rpm values were stratified according to CpG-content (1% resolution) and the mean of the mean rpms within each stratus calculated. Due to variability in the size of ROIs (except for the genome-wide analysis), the plots for the different genomic categories are not directly comparable. Wiggle-based plots: to avoid redundancy, the longest transcript variant of each gene in the RefSeq database was used as reference. In total the chromosomal mappings of 21,513 unique genes were used. The filtered alignment files were converted to bigWig files from which the tag count information was extracted using unix tools from the UCSC website. Gene Body plots: 40 non-overlapping windows with average tag number per base were calculated for each gene. 10 kb upstream of TSS and 10 kb downstream of transcription end site (TES) was divided into of windows of size 1 kb. Between TSS and TES each gene were divided into 20 windows of equal (gene-specific) size and the average counts was calculated. All statistics and plotting were done using the statistical program R.

mRNA expression analysis

For expression analysis, total RNA was purified from murine embryonic stem cells using RNeasy (Qiagen). The RNA was reverse transcribed using TaqMan reverse transcription reagents from ABI, according to the manufacturer's instructions. For RNA quantification, reversed-transcribed total RNA was analysed by real-time PCR using SYBR Green PCR Master Mix (Fermentas) and an ABI prism 7700 Sequence Detection system. All reactions were analysed in triplicates. Primer sequences are listed in Supplementary Figure 13 and Supplementary Fig. 14. For microarray analysis, RNA was extracted with the RNeasy Plus RNA extraction kit (Qiagen). RNA was hybridized on mouse Gene 1.0 ST arrays by the RH Microarray Center at Rigshospitalet, Copenhagen, following Affymetrix procedures and analysis. Gene expression analyses of RNA from shScr, shTet1#4, shTet1#5 and shSin3A#73 cells were performed in triplicates and in the subsequent data analysis FDR values <0.05 was used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank U. Toftegaard for excellent technical help, M. Okano for the donation of TKO ES cells, and members of the Helin lab for discussions. M.T.P. was supported by a fellowship from the Danish Cancer Society. J.R. is a senior research fellow of the Wellcome Trust. The work in the Helin lab was supported by grants from the Excellence Program of the University of Copenhagen, the Danish National Research Foundation, the Danish

Cancer Society, the Lundbeck foundation, the Novo Nordisk Foundation, and the Danish Medical Research Council.

References

1. Fouse SD, 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; 2:160–169. [PubMed: 18371437]
2. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008; 454:766–770. [PubMed: 18600261]
3. Mohn F, et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell*. 2008; 30:755–766. [PubMed: 18514006]
4. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA*. 2006; 103:1412–1417. [PubMed: 16432200]
5. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA*. 2002; 99:3740–3745. [PubMed: 11891299]
6. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet*. 2008; 9:465–476. [PubMed: 18463664]
7. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Rev. Genet*. 2009; 10:295–304. [PubMed: 19308066]
8. Gal-Yam EN, Saito Y, Egger G, Jones PA. Cancer epigenetics: modifications, screening, and therapy. *Annu. Rev. Med*. 2008; 59:267–280. [PubMed: 17937590]
9. Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324:930–935. [PubMed: 19372391]
10. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009; 324:929–930. [PubMed: 19372393]
11. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res*. 2010; 38:e181. [PubMed: 20685817]
12. Koh KP, et al. Tet1 and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*. 2011; 8:200–213. [PubMed: 21295276]
13. Ito S, et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. 2010; 466:1129–1133. [PubMed: 20639862]
14. 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*. 2006; 11:805–814. [PubMed: 16824199]
15. Maunakea AK, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010; 466:253–257. [PubMed: 20613842]
16. Song CX, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnol*. 2011; 29:68–72. [PubMed: 21151123]
17. Grzenda A, Lomberk G, Zhang JS, Urrutia R. Sin3: master scaffold and transcriptional corepressor. *Biochim. Biophys. Acta*. 2009; 1789:443–450. [PubMed: 19505602]
18. Mohamedali AM, et al. Novel *TET2* mutations associated with UPD4q24 in myelodysplastic syndrome. *J. Clin. Oncol*. 2009; 27:4002–4006. [PubMed: 19528370]
19. Delhommeau F, et al. Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med*. 2009; 360:2289–2301. [PubMed: 19474426]
20. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell Biol*. 2007; 27:3769–3779. [PubMed: 17339329]
21. Pasini D, et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature*. 2010; 464:306–310. [PubMed: 20075857]

22. Weber M, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* 2005; 37:853–862. [PubMed: 16007088]
23. Ji H, et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nature Biotechnol.* 2008; 26:1293–1300. [PubMed: 18978777]
24. Chavez L, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res.* 2010; 20:1441–1450. [PubMed: 20802089]
25. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 2007; 448:553–560. [PubMed: 17603471]
26. Rahl PB, et al. c-Myc regulates transcriptional pause release. *Cell.* 2010; 141:432–445. [PubMed: 20434984]
27. Marson A, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell.* 2008; 134:521–533. [PubMed: 18692474]
28. Erlanger BF, Beiser SM. Antibodies specific for ribonucleosides and ribonucleotides and their reaction with DNA. *Proc. Natl Acad. Sci. USA.* 1964; 52:68–74. [PubMed: 14192660]
29. Rhead B, et al. The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* 2010; 38:D613–D619. [PubMed: 19906737]

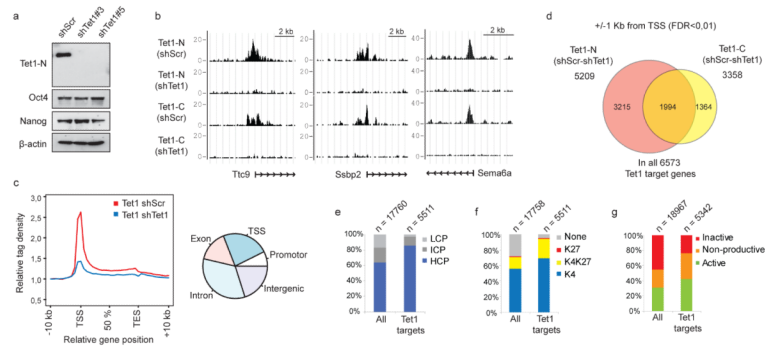


Figure 1.
Identification of TET1 target genes.

a, Western blot showing TET1, OCT4 and NANOG levels for control-transfected (shScr) and TET1-depleted (shTet1#3 and shTet1#5) mouse ES cells. b, Examples of TET1 ChIP-seq results in control or Tet1 knockdown ES cells. ChIP-seq was performed using both an anti-N- and anti-C-terminal TET1 antibody (Tet1-N and Tet1-C). y-axis of binding profiles denotes number of sequence tag reads. c, Left panel, mean distribution of tags across gene bodies for TET1 ChIP-seq in control and TET1 knockdown cells. Right panel, diagram illustrating the overall distribution of TET1 binding sites into TSS (± 1 kb), promoter (-1 to -5 kb), exon, intron and intergenic regions. d, Venn diagram illustrating the overlap of TET1 target genes using anti-TET1-N and -C antibodies. e, f, Histograms showing promoter CpG density, divided into high-, intermediate- or low-density CpG promoters (HCP, ICP or LCP) as defined in ref. 25 (e) or distribution of H3K4me3 (K4) and H3K27me3 (K27)25 (f) for all genes or for TET1 target genes. g, Overlay of TET1 target genes with active genes (RNA polymerase II binding and H3K79me2), non-productive (RNA polymerase II binding, no H3K79me2) and inactive (no RNA polymerase II binding or H3K79me2)26, 27.

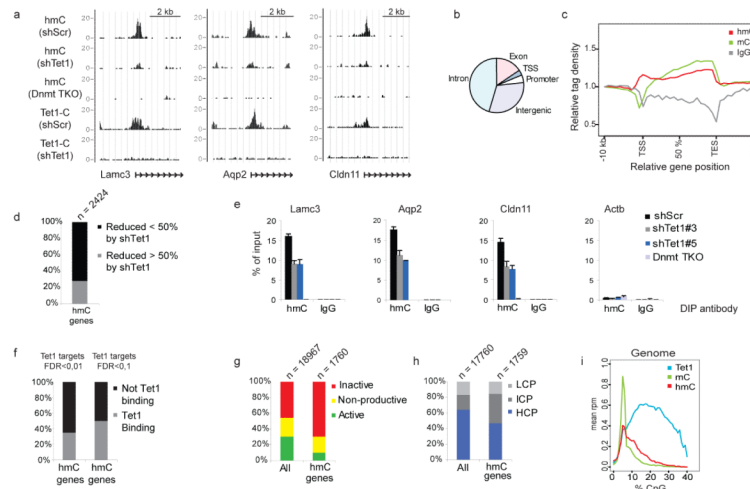
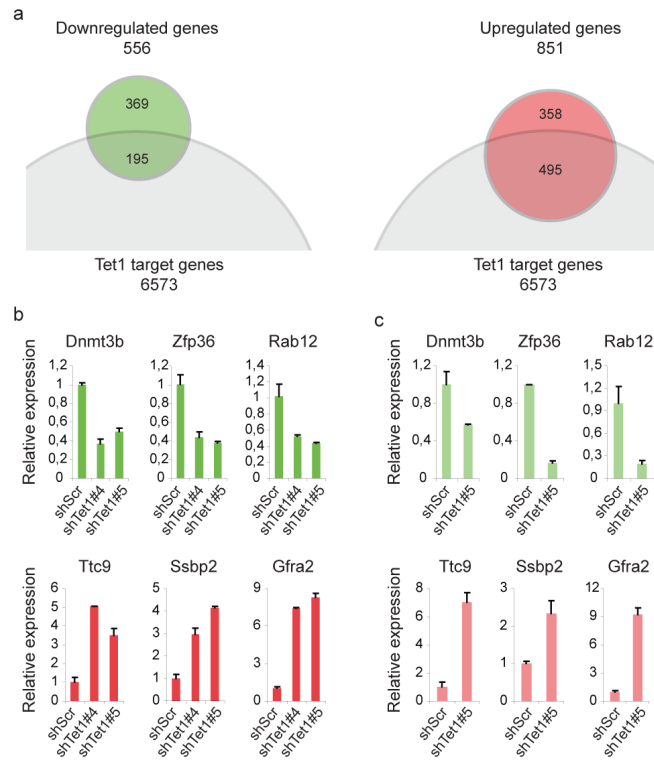


Figure 2.

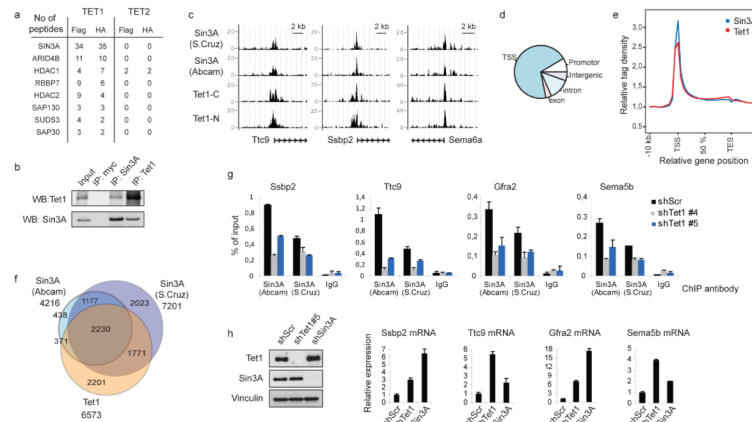
Hydroxymethylcytosine localizes to TSS and gene body.

a, Examples of hmcC DIP-seq results in mouse ES cells. ChIP-seq profiles of TET1 are included for comparison. b, Diagram illustrating the overall distribution of hmcC into TSS (± 1 kb), promoter (-1 to -5 kb), exon, intron and intergenic regions. c, The mean distribution of tags across gene bodies for hmcC, mC and IgG. d, Almost a third (28%) of hmcC positive TSSs showed a more than twofold reduction in hmcC signal in mouse ES cells depleted of TET1. e, DIP-qPCR was performed in control mouse ES cells, Tet1 knockdown cells (shTet1#3 and shTet1#5), and Dnmt TKO cells as indicated. f, Overlay of genes positive for hmcC at the TSS with TET1 target genes using FDR cut-off values of 0.01 or 0.1 in the ChIP-seq analysis. g, Overlay of hmcC positive genes with active genes (RNA polymerase II binding and H3K79me2), non-productive (RNA polymerase II binding, no H3K79me2) and inactive (no RNA polymerase II binding or H3K79me2)26, 27. h, Distribution of high-, intermediate- or low density CpG promoters (HCP, ICP or LCP)25 for all genes or hmcC-positive genes. i, Plot illustrating the genome-wide correlation of TET1, hmcC and mC signal intensity (rpm, reads per million) with CpG density. All error bars denote s.d., $n = 3$.

**Figure 3.**

Knockdown of Tet1 in ES cells affects transcription.

a, Microarray analyses were performed in control (shScr) and Tet1 knockdown cells (shTet1#4 and shTet1#5) in triplicates. Venn diagram showing overlap between TET1-bound genes, and genes up- or downregulated by both shRNAs using a cut-off of $FDR < 0.05$. b, qRT-PCR validation of selected genes. c, Genes that were found upregulated or downregulated by Tet1 knockdown show similar regulation in Dnmt TKO ES cells. All error bars denote s.d., $n = 3$.

**Figure 4.**

TET1 interacts with SIN3A.

a, Peptides identified by mass spectrometry from anti-Flag and tandem anti-Flag–HA purification of Flag–HA–TET1 and Flag–HA–TET2 stably expressed in 293 cells. The presented proteins are all part of the SIN3A complex¹⁷. b, Antibodies specific for TET1, SIN3A and c-Myc (negative control) were used for immunoprecipitation (IP) and western blot (WB) using nuclear extracts from mouse ES cells. Input represents 8%. c, Examples of SIN3A and TET1 ChIP-seq results in mouse ES cells. d, Diagram illustrating the overall distribution of SIN3A binding sites into TSS (± 1 kb), promoter (-1 to -5 kb), exon, intron and intergenic regions. e, Mean distribution of tags across gene bodies for SIN3A and TET1. f, Venn diagram illustrating a significant ($P < 10^{-8}$) overlap between TET1 and SIN3A target genes ($FDR < 0.01$). g, ChIP-qPCR in control or Tet1 knockdown cells (shTet1#4 and shTet1#5). h, Left panel, western blot illustrating knockdown efficiencies of TET1 and SIN3A. Right panel, genes that are upregulated by Tet1 knockdown are also de-repressed by Sin3a knockdown. All error bars denote s.d., $n = 3$.