Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity

Johannes M. Freudenberg1, Swati Ghosh1, Brad L. Lackford2, Sailu Yellaboina1, Xiaofeng Zheng2, Ruifang Li2, Suresh Cuddapah3, Paul A. Wade2, Guang Hu2,* and Raja Jothi1,*

1Systems Biology Section, Biostatistics Branch, 2Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), 111 TW Alexander Drive, Research Triangle Park, NC 27709, USA and 3Department of Environmental Medicine, New York University School of Medicine, 57 Old Forge Road, Tuxedo, NY 10987, USA

Received November 1, 2011; Revised November 29, 2011; Accepted December 1, 2011

ABSTRACT

The TET family of FE(II) and 2-oxoglutarate-dependent enzymes (Tet1/2/3) promote DNA demethylation by converting 5-methylcytosine to 5-hydroxymethylcytosine (5hmC), which they further oxidize into 5-formylcytosine and 5-carboxylcytosine. Tet1 is robustly expressed in mouse embryonic stem cells (mESCs) and has been implicated in mESC maintenance. Here we demonstrate that, unlike genetic deletion, RNAi-mediated depletion of Tet1 in mESCs led to a significant reduction in 5hmC and loss of mESC identity. The differentiation phenotype due to Tet1 depletion positively correlated with the extent of 5hmC loss. Meta-analyses of genomic data sets suggested interaction between Tet1 and leukemia inhibitory factor (LIF) signaling. LIF signaling is known to promote self-renewal and pluripotency in mESCs partly by opposing MAPK/ERK-mediated differentiation. Withdrawal of LIF leads to differentiation of mESCs. We discovered that Tet1 depletion impaired LIF-dependent Stat3-mediated gene activation by affecting Stat3’s ability to bind to its target sites on chromatin. Nanog overexpression or inhibition of MAPK/ERK signaling, both known to maintain mESCs in the absence of LIF, rescued Tet1 depletion, further supporting the dependence of LIF/Stat3 signaling on Tet1. These data support the conclusion that analysis of mESCs in the hours/days immediately following efficient Tet1 depletion reveals Tet1’s normal physiological role in maintaining the pluripotent state that may be subject to homeostatic compensation in genetic models.

INTRODUCTION

Embryonic stem cells (ESCs) appear to have an unique epigenetic state that maintains the pluripotent genome in a stable program of self-renewal, while allowing rapid induction of alternate transcriptional programs to initiate differentiation (1–8). DNA methylation is one of the principle regulators of the epigenetic landscape that shapes and refines gene expression programs during embryogenesis and stem cell differentiation. While mechanisms of establishment and maintenance of DNA methylation by DNA methyltransferases are well characterized (9,10), it was less clear which enzymatic machinery is responsible for DNA demethylation or even which pathways lead to DNA demethylation (11,12). Recently, it was discovered that proteins of the Tet family of FE(II) and 2-oxoglutarate-dependent enzymes (Tet1/2/3) oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (13–17), which they further oxidize into 5-formylcytosine and 5-carboxylcytosine, thereby promoting DNA
demethylation (18–23). Tet proteins and 5hmC have been shown to have roles in cancer and stem cell biology (13,16,24,25). Mouse embryonic stem cells (mESCs) deficient in DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b lack 5mC as well as 5hmC implying that 5hmC marks are probably derived from pre-existing 5mC marks (26,27).

In mESCs, Tet1 is expressed at high levels, comparable to those of the master pluripotency factor Oct4 (16,28). Tet2, though well expressed in mESCs, is 5-fold less abundant than Tet1 (28). Tet3 is variably expressed in many tissues but not in mESCs (16). Though various tissues express one or more Tet proteins, the 5hmC modification is particularly abundant in mESCs and Purkinje neurons (13,14,16,29). In mESCs, Tet1 binding and 5hmC occupancy are correlated with CpG density, are enriched at the promoters and gene bodies of nearly two-thirds of all genes, and have been linked to both gene activation and repression (27,30–33). Tet1 has also been shown to be required for the recruitment of transcriptional repressors Ezh2 (30) and Sin3a (27) at CpG-rich promoters of developmental regulators. Though Tet1-mediated regulation is believed to be due to its catalytic activity, Tet1 might have other functions in addition to converting 5mC to 5hmC (27).

A recent study reported that RNAi-mediated knockdown (KD) of Tet1 in mESCs resulted in downregulation of pluripotency marker Nanog and loss of undifferentiated state implicating Tet1 in mESC maintenance (16). In contrast, subsequent studies found that Tet1 KD cells were morphologically indistinguishable from control mESCs with no changes in Nanog expression, modest reduction in total 5hmC levels, and a minor-to-moderate increase in 5mC levels (26–28,31). Discrepancy between these studies may be due to differences in KD efficiency, off-target effects, or homeostatic compensation masked by antibiotic selection. More recently, it was reported that Tet1−/− mice are viable, fertile and grossly normal, and that Tet1−/− mESCs maintained a normal undifferentiated mESC morphology with only ~35% reduction in total 5hmC levels (34). Since only a modest reduction in 5hmC levels is observed in this genetic model, it is possible that maternal contribution or unaltered Tet2 in Tet1 null cells may be compensating Tet1’s function, thereby obscuring the direct impact of Tet1 loss. The discrepancies observed in the various RNAi-mediated depletion experiments in addition to the puzzling retention of 5hmC in the Tet1 null mESCs suggest that unanswered questions remain regarding the role of Tet1 in the maintenance of the pluripotent state of mESCs.

To understand and clarify the role of Tet1 and 5hmC in epigenetic and transcriptional regulation of mESCs, we used RNAi to acutely deplete Tet1 in mESCs and performed expression and genome-wide 5hmC occupancy studies. We find that acute short-term depletion as opposed to genetic deletion of Tet1 results in a significant decrease in 5hmC levels, downregulation of pluripotency-associated factors, impairment of LIF-dependent Stat3-mediated gene activation, and loss of embryonic stem cell identity.

**MATERIALS AND METHODS**

**Mouse ES cell culture, RNAi and alkaline phosphatase staining**

Oct4GiP mESCs were kindly provided by Dr Austin Smith. E14Tg2a mESCs were obtained from Mutant Mouse Research Resource Centers, and J1 mESCs were obtained from ATCC. The cells were routinely maintained on gelatin-coated plates in the ESGRO complete plus clonal grade medium (Millipore), and were used at passage 20–35 for experiments. For siRNA transfections, mESCs were cultured on gelatin-coated plates in M15 medium: Dulbecco’s Modified Eagle Medium (Invitrogen) supplemented with 15% Fetal Bovine Serum (FBS) (Invitrogen), 10 μM 2-mercaptoethanol, 0.1 mM non-essential amino acids (Invitrogen), 1× EmbryoMax nucleosides (Millipore), 1000 U of ESGRO (Millipore). For cells to be harvested or stained 96 h after transfection, 20–25 × 10^3 mESCs were transfected with siRNAs at 100 nM in M15 medium in one well of a 96-well plate. About 0.7 μl of Lipofectamine 2000 (Invitrogen) was pre-mixed with 10 μl of Opti-MEM (Invitrogen) and then mixed with 1 × 10^−11 mol siRNAs. Dissociated mESCs were plated at 25 × 10^3 per well in 100 μl of M15 medium in gelatin-coated plates. siRNA–lipids complexes were incubated at room temperature for 15–30 min and then added to the cells. About 25–50% of the cells in each well was re-plated into one well of a 12-well plate the next day, and cultured in M15 medium for another 3 days. For cells to be harvested 48 h after transfection, 75 × 10^3 mESCs were transfected with siRNAs at 50 nM in M15 medium in one well of a 24-well plate, and cells were collected 2 days after transfection. RNAi experiments were performed using indicated individual siRNAs: Tet1 siRNA #1 (Invitrogen, MSS284895), Tet1 siRNA #2 (Invitrogen, MSS284897), Tet1 siRNA #3 (Dharmacon, D-062861-01), Tet1 siRNA #4 (Dharmacon, D-062861-02), Tet1 siRNA #5 (Dharmacon, D-062861-03), Tet1 siRNA #6 (Dharmacon, D-062861-04) and control siRNA duplex targeting firefly luciferase (Dharmacon, 5′-CGTACGCGGAATACCTCGA). For lineage marker [quantitative reverse-transcriptase PCR (qRT-PCR)] and microarray analysis, cells were harvested 48 and 96 h after transfection for RNA extraction. For alkaline phosphatase (AP) staining, cells were fixed, permeabilized and stained for AP activity with the Alkaline Phosphatase Detection Kit (Millipore).

**Nanog overexpression cells**

Mouse Nanog was cloned into pcDNA-EF-HA/Flag vector and sequence verified (see Supplementary Material). The resulting plasmid, mNanog-pcDNA-EF-HA/Flag, was transfected into Oct4GiP cells with Lipofectamine 2000 (Invitrogen), and the cells were selected with G418 and plated at colonial density. Single clones were picked and expanded, and clone #2 was chosen for the experiment. The expression of the exogenous HA/Flag-tagged Nanog was estimated to be at the same level of the endogenous Nanog based on western blot.
Oct4GiP reporter assay

Oct4GiP cells were transfected in 96-well plates: 0.35 µl of Lipofectamine 2000 (Invitrogen) was pre-mixed with 10 µl of Opti-MEM (Invitrogen) and then mixed with 5 x 10^-12 mol siRNAs. Dissociated cells were plated at 10 x 10^3 per well in 100 µl of M15 medium in gelatin-coated plates. siRNA–lipid complexes were incubated at room temperature for 15–30 min and then added to the cells. Cells were cultured in M15 medium with daily medium change. Four days after transfection, cells were lifted and dissociated by trypsinization, pipetted into single cell suspension in 10% FBS in Phosphate Buffered Saline (PBS), and fluorescence-activated cell sorting (FACS) analyzed on LSRII FACs analyzer (BD). Percentage of differentiation was determined by measuring the percentage of green fluorescent protein (GFP)-negative cells. For each experimental condition, 3–4 independent transfections were carried out, and data was plotted as mean ± SEM (standard error of the mean).

Western blot

Cells were directly lysed in SDS–PAGE sample buffer. Proteins were resolved by SDS–PAGE, transferred to nitrocellulose membrane using iBlot (Invitrogen), and probed with the indicated antibodies. Loading was normalized based on Ran or Tubulin. Primary antibodies used: Nanog (Millipore, AB9220), Stat3–Y705-phospho-specific (Cell Signaling, 9132), Dnmt3a (Santa Cruz, H-295), Dnmt3b (IMEGENEX, IMG-184A), Ran (BD, 610340) and Tubulin (Santa Cruz, sc-9104).

Quantitative RT-PCR

Total RNAs were prepared from cells using the RNeasy (Qiagen), and cDNAs were generated using the Omniscript RT kit (Qiagen) according to the manufacturer’s instructions. Quantitative PCRs were performed using the Ssofast Evagreen Supermix (Bio-Rad) on the Bio-Rad CFX-384 or CFX-96 Real-Time PCR System. At least three biological repeats were carried out for each experiment. For every biological repeat, triplicate or quadruplicate PCR reactions were performed for each sample, and Actin was used for normalization. Data was plotted as mean ± SEM. Please refer to Supplementary Table S1 for gene specific primers used for RT-PCR analysis.

5hmC slot blot, and genome-wide mapping and analysis of 5hmC

Mouse E14Tg2a cells were transfected with control siRNA or an siRNA targeting Tet1 in 6-well plates. Cells were lysed 96 h after transfection in Trizol reagent, and both total RNA and genomic DNA were prepared according to manufacturer’s instructions. Total RNA was used to determine the Tet1 KD efficiency. Genomic DNA was quantified with NanoDrop, and 100 or 500 ng was loaded on GE Hybond+ nylon membrane using the GE slot blot device according to manufacturer’s instructions and blotted for 5hmC. Primary antibody used: 5hmC (Active Motif, 39769). Relative signal strengths of 5hmC bands were quantified using the Image-J software (http://imagej.nih.gov/ij/).

For genome-wide mapping of 5hmC, the genomic DNA obtained from control KD and Tet1 KD cells were processed using the QUEST 5hmC detection kit (Zymo Research) according to manufacturer’s instructions. Briefly, the genomic DNA was treated with 5hmC glucosyltransferase enzyme, which specifically tags the 5hmC in DNA with a glucose moiety yielding a modified base, glucosyl-5hmC. Next the DNA was digested using glucosyl-5hmC sensitive restriction endonuclease MsPI. MsPI efficiently cleaves the DNA when cytosine, 5mC, or 5hmC but not glucosyl-5hmC is located at the inner C position within its recognition sequence (i.e. CCGG). In other words, the 5hmC sites in genomic DNA treated with 5-glucosyltransferase enzyme are protected from MsPI digestion. By comparing the qPCR or next-generation sequencing results for treated versus untreated genomic DNA, 5hmC sites are detected (see below for details). Computational analysis indicated that digesting mouse genomic DNA with MsPI, selecting 40–220bp fragments, and performing 36-bp sequencing would cover ~1 million distinct CpG dinucleotides (~5% of all CpGs), with roughly half located within ‘CpG islands’ (including sequences from 90% of all CpG islands) and the rest distributed between other relatively CpG-poor sequence features (35). Every Mspl cleaved sequence read will thus include at least one informative CpG position. Based on this information, Mspl-digested DNA from treated and untreated samples derived from control and Tet1 KD mESCs were size-selected (40–220 bp) and sequenced using Illumina GAII.

Sequenced 36-bp reads were aligned to the reference genome (mouse NCBI36/mm8 assembly) using Bowtie 0.12.2 (36). Those reads that mapped to unique genomic locations with at most two mismatches were retained. Roughly ~50–65% of all reads were mapped to unique genomic locations. To increase the mapping efficiency, we trimmed one base from the 3′-end of the unmapped reads, aligned the residual 36-bp reads to the reference genome and retained those that mapped to unique genomic locations. This procedure was repeated until the residual reads were of length 18, and, in each iteration, reads that mapped to unique genomic locations were retained for further analysis. The reads that mapped to unique genomic locations but whose corresponding 5′- to 3′-sequence in the reference genome neither started with CCG nor ended with GGC were considered noise and discarded. The rationale for this is because the Mspl cleaves the DNA containing its recognition site CCGG at the inner C position (C/CGG) (37), the DNA fragments that is sequenced must all start with CCG or end with GGC (complementary strand). Reads that start or end with anything other than CCG or GGC, respectively, are likely to be byproducts of DNA degradation or sequencing errors. The resulting mapped reads for the four samples (control KD treated, control KD untreated, Tet1 KD treated and Tet1 KD untreated) were normalized by total reads in each sample so that the number of reads mapping to a specific site is directly comparable across all
four samples. Sites with less than two reads in all four samples were excluded from the analysis. A site was defined as a 5hmC site only if the normalized read count at this site in the control untreated sample is at least 1.5-fold higher than that in the control treated sample. A 5hmC site was defined to have reduced 5hmC levels in Tet1 KD mESCs only if

$$\Delta \text{5hmC} = \frac{N_{\text{treated}}}{N_{\text{control}}} / \frac{N_{\text{untreated}}}{N_{\text{control}}} \geq 1.5$$

where $\Delta \text{5hmC}$ measures the reduction in 5hmC levels in Tet1 KD cells and $N_i$ is the normalized read count in sample $s$ with treatment $t$. Genes were considered to have 5hmC occupancy if they have one or more detected 5hmC sites within 5 kb upstream of the transcription start site (TSS) to the transcription end site (TES). Please refer to Supplementary Table S2 for PCR primers used for quantitative 5hmC analysis.

Chromatin immunoprecipitation

E14Tg2a mESCs were crosslinked with 1% formaldehyde in PBS for 10 min, and the resulting chromatin was sonicated using Diagenode Bioruptor for 15–18 cycles, 30 s per cycle with a 30 s rest between each cycle, to obtain ~200–500 bp fragments. Chromatin immunoprecipitation (ChIP) was performed by incubating indicated antibodies with Dynabeads protein A magnetic beads (Invitrogen, 100-02D) for 4–6 h at 4°C, followed by overnight incubation with the chromatin at 4°C. Subsequent washing and reverse crosslinking steps were performed as described earlier (38). Each ChIP was performed using chromatin from ~2 × 10⁶ cells. Primary antibodies used: Stat3 (Santa Cruz, C-20), H3K9me3 (Abcam, ab8898) and IgG (Santa Cruz, H-270). Please refer to Supplementary Table S3 for ChIP primers used to analyze Stat3-target/binding sites.

ChIP-Seq factor binding data analysis

Published ChIP-Seq data sets for various transcription factors and chromatin remodelers (listed in Supplementary Table S4) were downloaded from the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) or the BROAD Institute website. In cases where the data was mapped to mouse reference genome assembly mm9, the data was realigned to mm8 assembly using the liftOver tool (39). Where available for download, binding sites (peaks) as defined in the original publication were used. Otherwise SISSSRs (40) was used for peak calling. Direct target genes of a particular factor (for example, Tet1) were defined as genes that have one or more detected factor binding sites within 5-kb upstream of the TSS to the TES.

Microarray gene expression data analysis

Raw CEL files (Affymetrix array type Mouse 430 2.0) generated from this study were processed using relevant R/Bioconductor packages (41). Specifically, RMA (42) was used for background correction and data normalization. Probe sets were defined using Entrez gene based custom chip definition files (CDFs), version 13 (http://brainarray.mbnii.med.umich.edu/Brainarray/Data base/CustomCDF/genomic_curated_CDF.asp) (43). This resulted in a single expression intensity measure per Entrez gene ID. All subsequent analyses were carried out on the log2 scale. To determine differentially expressed genes, a family-wise moderated t-test (44) (KD versus control) was performed followed by a multiple testing correction procedure to control the false discovery rate (FDR) (45). Unless otherwise noted, genes were considered differentially expressed if their FDR ≤ 0.1 and fold change ≥ 1.5. Since global gene expression changes due to Tet1 KDs using Tet1 siRNAs #1 and #2 were highly similar, we treated them as biological replicates in the differential expression analysis. Unsupervised hierarchical clustering was performed using the CLEAN software package (46). Functional enrichment analysis of differentially regulated lists of genes was performed using R packages CLEAN (46) and GO.db (41).

Published microarray data sets (listed in Supplementary Table S5) were downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo/). Where available, raw CEL files were re-processed using the above described microarray processing pipeline. Otherwise, the respective Series Matrix File was used to obtain preprocessed expression values and corresponding probe annotations. All data were transformed and analyzed on the log2 scale. Each probe identifier was mapped to an Entrez gene ID using relevant Bioconductor packages by continually matching annotations in the following order until an Entrez gene ID was found: Entrez gene ID, RefSeq ID, gene symbol or alias, gene name, unigene ID, Ensembl transcript ID. Probes mapping to the same Entrez gene ID were averaged (after the log-transformation). Fold changes and adjusted P-values (KD versus control) using moderated t-statistic described above were then computed for each data set independently.

Data availability

All microarray and 5hmC sequencing data generated for this study were deposited in NCBI GEO repository under the accession number GSE34267.

RESULTS

Acute depletion of Tet1 results in loss of embryonic stem cell identity

To elucidate the role of Tet1 and 5hmC in epigenetic and transcriptional regulation of mESC self-renewal and pluripotency, we used RNAi to acutely deplete Tet1 in Oct4GiP mESCs, which express the GFP under the control of the Oct4 promoter (47). Since Oct4 is exclusively expressed in ESCs, the GFP expression faithfully correlates with the ESC identity. Two small-interfering RNAs (siRNAs #1 and #2) targeting Tet1 were used to ensure that the effects were specific. Cells transfected with siRNAs were maintained in normal mESC culture conditions for 4 days, and subjected to FACS to determine the fraction of differentiated (GFP-negative) cells. Acute deple- tion of Tet1 resulted in small but significant increase in the percentage of differentiated cells (Figure 1A).
Tet1 knockdown (KD) efficiency was confirmed using qRT-PCR and western blot (Figure 1B and Supplementary Figure S1A). Examination of Tet1 KD mESCs for colony morphology and AP staining revealed morphological changes and loss of AP staining consistent with differentiation (Figure 1C). The outcome remained similar even when E14Tg2a or J1 cell lines were used (Supplementary Figure S1B). To rule out the possibility that the observed changes were due to siRNA off-target effects, we used four additional siRNAs (#3–6), some of which were used in other studies that observed no obvious phenotype (28,31), and confirmed that acute depletion of Tet1 indeed induces morphological changes and loss of AP staining consistent with differentiation (Supplementary Figure S1B).

ESC markers are downregulated and differentiation markers are upregulated upon Tet1 depletion

To determine whether Tet1 depletion has an effect on gene expression programs influencing mESC self-renewal, pluripotency and differentiation, we used microarrays to profile global gene expression in mESCs transfected with control and Tet1 siRNAs at 96 h after transfection. Pluripotency-associated genes such as Nanog, Esrrb, Tcl1, Tbx3, Klf2, Klf4, Lefty1, Lefty2, Tcfcp2l1 and Prdm14 were downregulated, and differentiation-associated genes including ectoderm and neuronal markers Fgf5, Pitx2, Nestin, Nefm, CD133 (Prom1), CD44, Lef1 and Zic1 and trophodermectodermarkers Eomes and Krt8/18/19 were upregulated in response to Tet1 KD (Figure 1D). Gene ontology analysis of transcripts downregulated at 96 h revealed enrichment of many terms related to stem cell maintenance and proliferation, whereas upregulated transcripts were enriched for terms related to development and differentiation (Figure 1E). To understand the kinetics of downstream effects of Tet1 KD, we also carried out microarray analysis at 48 h after siRNA transfection when Tet1-depleted cells (Figure 1B and Supplementary Figure S1A) were morphologically indistinguishable from control cells (Supplementary Figure S1D). Interestingly, our 48 h Tet1 KD data clustered tightly with those from shRNA-mediated Tet1 KD studies that did not observe morphological changes even past 96 h (Figure 1D and Supplementary Figure S1C), and away from our 96 h Tet1 KD data that accompanies morphological changes, suggesting a kinetic effect. The lack of phenotype in shRNA-mediated depletion studies could be due to the (i) differences in reaction rates in shRNA-versus siRNA-mediated knockdowns as a result of additional processing time necessary for the shRNA hairpin structure to be cleaved into siRNA or (ii) homeostatic compensation of Tet1’s function masked by antibiotic selection.

We performed qRT-PCR and confirmed gene expression changes observed in the microarray analysis. Tet1 depletion decreased the expression of key pluripotency factors: Nanog by nearly 3-fold, and Oct4 and Sox2 by small, but reproducible and statistically significant, ~1.25-fold. Other pluripotency-associated genes such as Klf4, Esrrb, Tcl1, Tbx3, Prdm14 and Nodal antagonists Lefty1 and Lefty2, which are among the earliest to be downregulated upon ESC differentiation (48), were downregulated by ~2- to 4-fold (Figure 1F). Tet1 depletion also increased the expression of several differentiation genes including Pitx2, Nestin, Nefm, Lef1 and Zic1 that are largely associated with Ectoderm lineage (Figure 1F). Furthermore, these changes were consistent across Oct4GiP, E14Tg2a and J1 mESCs (Supplementary Figure S2A), and remained consistent even when using an siRNA that was used in two previous studies that observed no obvious phenotype (28,31) (Supplementary Figure S2B). Taken together, we concluded that acute RNAi-mediated depletion of Tet1 results in the loss of mESC identity.

**Tet1 depletion results in significant reduction in total 5hmC levels**

Examination of total 5hmC levels in Tet1-depleted mESCs using slot blot revealed a significant reduction in 5hmC compared to cells transfected with control siRNA (Supplementary Figure S2C). The severity of observed morphological changes positively correlated with the extent of 5hmC loss (Figure 1G). We suspected that the downregulation of Tet2 in Tet1-depleted cells (Figure 1D), also observed in one other report with findings consistent with ours (30), might also be a contributor to reduced 5hmC levels. Quantitative RT-PCR analysis of Tet2 in Tet1-depleted mESCs revealed a reproducible and statistically significant ~1.5-fold reduction in Tet2 levels (Supplementary Figure S2D). This along with Tet1 occupancy at Tet2 promoter (Supplementary Figure S2E) suggested regulation of Tet2 by Tet1. Together, these data revealed that the acute reduction in 5hmC primarily due to Tet1 KD and also possibly due to indirect reduction in Tet2 levels is probably the cause for the loss of mESC identity.

**Tet1-mediated gene regulation is context dependent**

To investigate the role of Tet1 and 5hmC in the maintenance of mESC identity on the molecular level, we examined the genes that were differentially expressed upon Tet1 depletion in our microarray analysis. Using a reasonably stringent criterion (FDR < 0.1; fold ≥ 1.5), we identified 530 upregulated and 389 downregulated genes (Supplementary Table S7). Examination of published Tet1 occupancy data (27,30) revealed that Tet1 binds to 72% of upregulated and 63% of downregulated genes with no statistically significant preference for one group over the other, which is similar to the pattern observed with other factors known to play diverse regulatory roles in gene regulation such as the insulator binding transcription factor CTCF (Figure 2A and Supplementary Figure S3A). No correlation between Tet1 binding and gene expression changes was observed even when analyzing all genes with expression above detectable levels (Supplementary Figure S3B). These data indicate that Tet1 binding alone is not predictive of Tet1’s function as an activator or a repressor.

Tet1 has been reported to be required for the recruitment of transcriptional repressors Ezh2 (30) and Sin3a (27) at CpG-rich promoters of developmental regulators.
Figure 1. Depletion of Tet1 and 5hmC levels results in loss of mouse embryonic stem identity. (A) Oct4GiP mESCs were transfected with indicated siRNAs in normal ESC medium and cultured for 96 h. The percentage of differentiated cells was determined by measuring the percentage of GFP-negative cells using FACS at 96 h after transfection (**P < 0.001; *P < 0.01). Error bars represent SEM of three experiments. (B) Relative Tet1 mRNA level in control and Tet1 KD Oct4GiP mESCs 48 h after transfection. Data are normalized to Actin. Error bars represent SEM of three experiments (**P < 0.001). Error bars represent SEM of three experiments. (C) AP staining of Oct4GiP mESCs transfected with control siRNA, and Tet1 siRNAs #1 and #2. Cells were cultured in normal ESC medium, and AP staining was performed 96 h after transfection.

(continued)
We find that genes upregulated upon Tet1 KD are enriched for binding of polycomb members Suz12, Ezh2 and Ring1b (Figure 2A and Supplementary Figure S3A), whereas downregulated genes are enriched for binding of Stat3, Klf4, Oct4, Esrrb and Sin3a among others (Figure 2A). Since Tet1 binds to nearly two-thirds of all genes, its co-occupancy with other transcription factors is rather high, which could be misinterpreted for its functional interaction with these factors. When normalized by expected frequencies, Tet1’s co-occupancy with other factors appears insignificant compared to that between factors that are known to function together (Supplementary Figure S3C). Like esBAF, a mESC specific chromatin remodeling complex (49) that facilitates both activation and repression (50,51), we propose that Tet1 probably plays a similar role as a general facilitator/recruiter of activating and repressing factors in a context dependent manner.

**Tet1 depletion results in genome-wide reduction in 5hmC levels**

To determine the extent of 5hmC loss in Tet1 KD cells on a genomic scale, we used a modified version of methylation sensitive restriction enzyme sequencing (52) and generated nucleotide resolution 5hmC maps covering ~90% of sequences located within CpGs islands in control and Tet1 KD mESCs (see ‘Materials and Methods’ section for details). We identified 57895 5hmC sites in control mESCs, of which 32652 had at least 1.5-fold reduction in 5hmC levels in Tet1 KD cells (Supplementary Table S8). Quantitative PCR was performed to validate 5hmC sites and reduction in 5hmC levels (Supplementary Figure S3D). We find that genes that were upregulated or downregulated upon Tet1 depletion are equally likely to contain 5hmC sites and experience reduced 5hmC levels (Figure 2B), which is consistent with the established dual roles of Tet1 and 5hmC in both gene activation and repression (27,30,33). These data indicate that an acute depletion of Tet1 causes significant global reduction in 5hmC levels that were not specific to gene classes, and that the directionality of gene expression change (up/down) is probably determined by the gene context and local transcription factor occupancy (Figure 2A).

**Meta-analysis of genomic data suggests interaction between Tet1 and LIF/Stat3 signaling**

We next examined the fate of Tet1 regulated genes upon KD or knockout of other pluripotency factors. A meta-analysis of microarray data from published studies revealed that genes that are differentially expressed upon Tet1 KD underwent similar expression changes upon leukemia inhibitory factor (LIF) withdrawal or after depletion of key pluripotency-associated factors such as Esrrb, Tet1, Tbx3 and Klf2/4/5, known to be associated with LIF/Stat3 signaling (8,53–55) (Figure 2C), suggesting Tet1’s possible interaction with or co-regulation of transcriptional circuit regulated by LIF/Stat3 signaling (Figure 2A). In contrast, Tet1-regulated genes underwent largely divergent changes upon deletion/depletion of transcriptional repressors Tcf3 and Sall4 (Figure 2C).

**LIF-dependent Stat3-mediated gene activation is dependent on Tet1**

To examine Tet1’s possible mechanistic interaction with LIF/Stat3 signaling (Figure 2A and C), we analyzed global gene expression changes upon Tet1 depletion against those after LIF withdrawal. Tet1-depleted and LIF-starved mESCs undergo similar changes in the global transcriptional profile (Figure 3A). The set of genes co-activated by Tet1 and LIF includes a number of factors essential for pluripotency, including Esrrb, Tcl1 and Tbx3 (56,57). This suggested that Tet1 or Tet1-dependent 5hmC levels might be required for LIF-dependent Stat3-mediated gene activation. To further explore this connection, we decided to study Tet1 KD mESCs 48 h after transfection, when Tet1 is depleted but the cells appeared normal compared to control transfected cells (Figure 1B, Supplementary Figure S1A and D). This time point was chosen so that we could clearly distinguish cause from effect. First, we asked whether LIF/Stat3 signal transduction was impaired in Tet1 KD mESCs. No significant changes in mRNA levels for Stat3 or LIF chimeric receptor consisting of gp130 and LIF receptor (LIFR) were observed at 48 h (Figure 3B and C). Tet1-depleted mESCs at 48 h after transfection had no significant changes in total Stat3 or Nanog levels but a modest reduction in phosphorylated/activated Stat3 (Figure 3D).

Next, we examined the ability of Stat3 to bind to its target sites on chromatin in the absence of Tet1. ChIP experiment
using Stat3 antibody was carried out at 48 h after Tet1 KD. Stat3 binding was affected at many gene targets co-bound and co-activated by Tet1 and Stat3 (Figure 3E) indicating that Tet1 is required to facilitate LIF/Stat3 responsiveness in mESCs by enabling Stat3 to bind to its target sites on chromatin. Since the expression changes due to loss of Stat3 binding at Stat3-activated genes is not evident at 48 h but only at 96 h after Tet1 KD (Figure 3E), we concluded that the loss of Stat3 binding precedes loss of expression. We do not rule out the possibility that a modest reduction in phosphorylated Stat3 levels (Figure 3D) might also have played a role in impaired Stat3 binding. Regardless, the LIF-dependent Stat3-mediated gene activation in mESCs is dependent on Tet1.

**Nanog overexpression or inhibition of MAPK/ERK signaling rescues Tet1 KD phenotype**

Since mESCs can be propagated without activation of Stat3 by imposing constitutive expression of core pluripotency factor Nanog (58), we checked if Nanog overexpression (Figure 4A) could rescue the differentiation phenotype exhibited by Tet1 KD cells. As expected, exogenous Nanog largely rescued the morphological changes and AP activity in Tet1 KD cells (Figure 4B). Also, since mESCs can be propagated without Stat3 activation by suppressing prodifferentiative MAPK/ERK signaling (59), we tested if mESCs transfected with Tet1 siRNA in 2i medium (which inhibits MAPK/ERK and Gsk-3b signaling) blocks differentiation. We found that while Tet1 depletion slightly decreased cell viability in 2i medium (data not shown), the expression of pluripotency-associated factors Nanog, Tcl1, Klf4 and Lefty2 were largely rescued (Figure 4C). In addition, Tet1 KD in 2i medium also suppressed upregulation of differentiation genes such as Pitx2, CD133, Nefm and Lef1. Together, these data provide evidence to further support the dependence of LIF/Stat3 signaling on Tet1.

**Figure 2.** Meta-analysis of genomic data sets suggests interaction between Tet1 and LIF/Stat3 signaling. (A) Top: Transcription factor occupancy at 919 genes differentially expressed upon Tet1 KD (389 downregulated and 530 upregulated). Genes are represented along the y-axis and factor occupancy is denoted by blue bar. Target gene occupancy is defined as factor occupancy within 5-Kb upstream of the gene’s TSS and/or within its gene body. Bottom: Log-fold enrichment of factor occupancy at up/downregulated genes. Red and green histograms denote enrichment in up- and down-regulated genes, respectively. *P < 0.01 after Bonferroni adjustment for multiple testing. (B) Top: Presence/absence of 5hmC sites (5hmC) in control mESCs, and 5hmC sites with ≥1.5-fold reduced hydroxymethylation levels (Δ5hmC) in Tet1 KD mESCs (96 h). Bottom: Log-fold enrichment at down- versus up-regulated genes, respectively. (C) Expression fold changes of 919 genes differentially expressed upon Tet1 KD (96 h; first column) presented alongside fold changes observed after KD or knockout (KO) of select other pluripotency factors in published reports. Genes are represented along the y-axis. Data sets along the x-axis have been ordered based on unsupervised hierarchical clustering of corresponding gene expression fold changes. Blue and yellow rectangles highlight Tet1-LIF and Polycomb clusters, respectively. Cells with no/missing data are colored in gray (DKO: double KO).
Figure 3. Dependence of LIF/Stat3 signaling on Tet1. (A) Two-dimensional gene density heatmap depicting global gene expression changes 96 h after Tet1 KD against changes observed 48 h after LIF withdrawal (51). The intensity of each square represents the number of genes that fall within that square. Axes indicate degree of fold change, from nil (middle of axis) to >1.5-fold (outermost squares). (B) Gene expression fold changes observed in microarray analysis 48 and 96 h after Tet1 KD in mESCs, and 48 h after LIF withdrawal (51). (C) Relative Stat3 mRNA level in control and Tet1 KD mESCs. The mRNA level in control KD cells is set as one. Data are normalized to Actin. Error bars represent SEM of three experiments. (D) Western blot analysis showing protein levels in control and Tet1 KD mESCs. Tubulin is used as a loading control. (E) Left: ChIP assay of select Stat3 target regions using an antibody against Stat3 in control KD and Tet1 KD mESCs (48 h). The y-axis represents enrichment over input normalized to a negative control region (Yipf2). Refer to Supplementary Figure S4 for data from control ChIP using a non-specific antibody against IgG. Error bars represent SEM of three experiments. Right: Gene expression fold changes observed in microarray analysis 48 and 96 h after Tet1 KD in mESCs, and 48 h after LIF withdrawal (51).

Figure 4. Nanog overexpression or suppression of MAPK/ERK signaling rescues Tet1 KD phenotype. (A) Western blot analysis showing Nanog overexpression with HA-tag. (B) AP staining of E14Tg2a mESCs, with and without Nanog overexpression, transfected with control siRNA and Tet1 siRNA #1. Cells were cultured in normal ESC medium, and AP staining was performed 96 h after transfection. (C) Relative mRNA levels of selected mESC pluripotency genes and differentiation marker genes in control and Tet1 KD mESCs in 2i medium. Oct4GiP cells were transfected with control siRNA or Tet1 siRNA #1 at 50 nM in 24-well plates in 2i medium (which inhibits MAPK/ERK and Gsk-3β signaling) and cells were cultured for 96 h after transfection. The mRNA levels in control cells are set as one. Data are normalized to Actin. Error bars represent SEM of three experiments.
Figure 5. Tet1 negatively regulates de novo DNA methyltransferase Dnmt3b. (A) ChIP assay of select Stat3 target regions using an antibody against H3K9me3 in control and Tet1 KD mESCs (48 h). The y-axis represents enrichment over input normalized to a positive control region (Myod1) for H3K9me3. Error bars represent SEM of three experiments. (B) Relative mRNA levels of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b in control and Tet1 KD mESCs 96 h after transfection. The mRNA levels in control cells are set as 1. Data are normalized to Actin. Error bars represent SEM of three experiments. (C) Western blot analysis showing protein levels of Nanog, Dnmt3a and Dnmt3b in control and Tet1 KD mESCs 96 h after transfection. Ran is used as a loading control. (D) Genome browser shot showing a region containing the Dnmt3b gene and results from Tet1 and Sin3a ChIP-Seq experiments by various groups (GSM numbers denote GEO accession). The red open rectangle highlights Tet1 occupancy at the promoter region of Dnmt3b, where a CpG island is present (green-filled rectangle). (E) Relative 5hmC levels at Dnmt3b locus in control and Tet1 KD (96 h) mESCs. Error bars represent SEM of three experiments.
Tet1 negatively regulates de novo DNA methyltransferase Dnmt3b

Given that genes downregulated in Tet1 KD cells undergo a minor-to-moderate increase in 5mC levels (26–28,31), and that repressive histone mark H3K9me3 has been known to recruit DNA methyltransferases and be associated with DNA methylation (35,60), we hypothesized that the loss of Stat3 binding and Stat3-mediated activation of its targets accompanies an increase in H3K9me3. ChIP using an antibody against H3K9me3 revealed that this was indeed the case (Figure 5A). It is not clear, however, whether Tet proteins alleviate/prevent gene silencing at Stat3 targets by preventing deposition of H3K9me3, or by maintaining a DNA-hypomethylated state. Interestingly, we noticed a 2-fold increase in Dnmt3b levels in Tet1 KD cells in microarray analysis (Figure 1D). Quantitative RT-PCR and western blot confirmed the increase in Dnmt3b at the mRNA and protein levels (Figure 5B and C). Since Tet1 targets Dnmt3b (Figure 5D), we examined the 5mC levels at the Dnmt3b locus and found >3-fold reduction in 5mC in Tet1 KD cells (Figure 5E). Based on this evidence, it is possible that Tet1’s regulation of DNA methylation might also involve its negative regulation, direct or indirect, of de novo DNA methyltransferase Dnmt3b.

DISCUSSION

Our studies indicate that acute short-term depletion as opposed to long-term genetic deletion of Tet1 in mESCs results in the loss of the mESC identity. Tet1 depletion, in our hands, reduced total 5mC levels by roughly 2- to 3-fold, Nanog expression by ~3-fold and Tet2 expression by ~1.5-fold. The extent of 5mC reduction positively correlated with morphological changes and extent of loss of AP staining (Figure 1) suggesting that it is the loss of 5mC and not necessarily Tet1 that caused differentiation. Our findings are consistent with those by another group that used shRNAs to deplete Tet1 and observed similar changes in morphology, global gene expression, Nanog, Tet2 and total 5hmC levels (16,30). Our findings, however, are in contrast to Tet1 depletion studies by others (26–28,31), who observed only ~1.5-fold reduction in total 5hmC levels, no reduction in Nanog or Tet2 levels and no distinguishable morphological changes. Most remarkably, our results were reproducible even when using a siRNA that was employed in two previous studies that observed no obvious phenotype (28,31). Interestingly, the molecular changes we observe after Tet1 depletion are consistent with those from siRNA-mediated KD but not shRNA-mediated KD with puromycin selection from a previous study (31) (data from Xu et al in Figure 1D). It should be noted that the only other study that reported a Tet1 KD phenotype used shRNA-mediated depletion of Tet1 with no antibiotic selection. Since the siRNAs we used do not lead to greater KD efficiency than the shRNAs used by studies with no phenotype, it is possible that the lack of phenotype observed by others could be the result of homeostatic compensation of Tet1’s function masked by antibiotic selection.

At present, although we have no ready explanation for the discrepancy between our results and those from Tet1 null mice (34), a modest reduction in 5hmC levels in the genetic model raises an intriguing possibility of maternal contribution or unaltered Tet2 expression compensating for Tet1’s function after prolonged loss of Tet1, therefore masking the direct effect of Tet1 loss that we observe hours/days after RNAi-mediated acute Tet1 KD. Our findings suggest that an acute as opposed to long-term chronic reduction in Tet1 and thus 5mC levels is probably difficult for mESCs to overcome or compensate, thereby resulting in the loss of the pluripotent state. This is consistent with Tet2’s critical role in regulating self-renewal, proliferation, and differentiation in hematopoietic stem cells (61–63), and Tet3’s requirement for paternal genome conversion of 5mC into 5hmC (64–66).

The sharp reduction in Tet1 and 5hmC impaired LIF-dependent Stat3-mediated gene activation by affecting Stat3’s ability to bind to its target sites on chromatin, indicating dependence of LIF/Stat3 signaling on Tet1 and/or 5hmC. Suppression of prodifferentiative
MAPK/ERK signaling or Nanog overexpression, known to maintain mESCs in the absence of LIF or Stat3 activation, rescued Tet1 KD phenotype further supporting the dependence of LIF/Stat3 signaling on Tet1. Collectively, our results support a model in which the Tet1-mediated regulation of 5hmC levels is critical to the maintenance of the pluripotent state in mESCs (Figure 6). Tet1 aids LIF-mediated Stat3 gene activation by facilitating Stat3 binding to chromatin by a yet to be determined mechanism. Tet1 regulates key pluripotency-associated factors including Nanog, Klf4/2, Tcf1, Tbx3, Esrrb, Lefty1/2, Prdm14 and Tcfcp21L. Regulation of DNA methylation by Tet1-mediated conversion of 5mC to 5hmC also involves Tet1’s regulation of Tet2 and Dnmt3b. Our findings suggest that studying mESCs in the hours/days following depletion of Tet1, before any compensatory alterations could take place, reveals Tet1’s normal physiological role in maintaining the pluripotent state, which may be subject to homeostatic compensation in genetic models. Tet1/2 double knockout may shed light into whether Tet2 is in fact compensating for Tet1’s loss in Tet1 null cells. Further investigations are needed to better understand the mechanistic role of Tet1 and 5hmC in facilitating pluripotency, and LIF/Stat3 signaling.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online: Supplementary Figures 1–4, Supplementary Tables 1–8, Supplementary Methods and Supplementary References [67–79].

ACKNOWLEDGEMENTS
We thank Lena Ho (A-Star) for helpful discussions on ChIP experiments and critical comments on the article, NIEHS/NISC sequencing facility for their advice and help in mapping 5hmC sites, and the NIEHS Microarray Core for microarray data generation. We also thank Karen Adelman, Georgette Charles, Senthilkumar Cinghu, Anne Lai, Xiaoling Li, Thomas Kunkel, Larry Lazarus, Daniel Menendez and Sailesh Surapureddi for critical comments on the article. R.J. conceived the study, and wrote the article. G.H. and R.J. designed the experiments with contributions from S.C. and P.A.W. G.H. directed the experiments. S.G., B.L., X.Z., R.L. and G.H. executed the experiments. J.M.F. and R.J. analyzed the data with contributions from S.Y. J.M.F., G.H. and R.J. interpreted the data. All the authors read and approved the article.

FUNDING
Funding for open access charge: The Intramural Research Program of the NIH, National Institute of Environmental Health Sciences [1ZIAES102625-03 (to R.J.), 1ZIAES102745-02 (to G.H.) and 1ZIAES101965-07 (to P.A.W.)].

Conflict of interest statement. None declared.

REFERENCES


