TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells

Nipun Verma1,2, Heng Pan3,4, Louis C. Doré5, Abhijit Shukla1, Qing V. Li6, Bobbie Pelham-Webb2, Virginia Teijeiro1,4, Federico González19, Andrei Krivtsov7, Chan-Jung Chang8, Eirini P. Papapetrou8, Chuan He5, Olivier Elemento3* and Danwei Huangfu1* 

TET enzymes oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), which can lead to DNA demethylation. However, direct connections between TET-mediated DNA demethylation and transcriptional output are difficult to establish owing to challenges in distinguishing global versus locus-specific effects. Here we show that TET1, TET2 and TET3 triple-knockout (TKO) human embryonic stem cells (hESCs) exhibit prominent bivalent promoter hypermethylation without an overall corresponding decrease in gene expression in the undifferentiated state. Focusing on the bivalent PAX6 locus, we find that increased DNMT3B binding is associated with promoter hypermethylation, which precipitates a neural differentiation defect and failure of PAX6 induction during differentiation. dCas9-mediated locus-specific demethylation and global inactivation of DNMT3B in TKO hESCs partially reverses the hypermethylation at the PAX6 promoter and improves differentiation to neuroectoderm. Taking these findings together with further genome-wide methylation and TET1 and DNMT3B ChIP–seq analyses, we conclude that TET proteins safeguard bivalent promoters from de novo methylation to ensure robust lineage-specific transcription upon differentiation.

DNA methylation is a key mechanism for transcriptional regulation, and dramatic changes in DNA methylation of regulatory regions occur during normal development and in pathological conditions1–2. Individual deletion of the DNA methyltransferases (Dnmt1, Dnmt3a and Dnmt3b) in mice leads to embryonic or postnatal lethality3–4. Proper development also requires active and passive DNA demethylation5. The three mammalian TET proteins (TET1, TET2 and TET3) catalyze the oxidation of 5mC to 5hmC, which can serve as an intermediate in active DNA demethylation6–8. Loss of TET proteins leads to hypermethylation of the promoters and enhancers of developmental genes in mouse embryos and mouse embryonic stem cells (mESCs)9–11. Although TET gene inactivation clearly impairs embryonic development and cellular differentiation, it is difficult to establish direct causal relationships between TET-mediated DNA demethylation, transcriptional output, and developmental or pathological phenotypes owing to the general challenge of distinguishing global versus locus-specific effects for epigenetic regulators12–14.

To link transcriptional output to TET-mediated demethylation, previous studies have focused on TET actions at enhancers and found examples in which loss of the TET proteins caused hypermethylation and decreased gene expression11,15. However, the consequences of TET activities at promoters remain ambiguous. In particular, bivalent promoters, which have H3K4me3 and H3K27me3 histone marks on the same or adjacent nucleosomes, are hypomethylated like active promoters. Yet, unlike active promoters, which support productive transcription, bivalent promoters are associated with negligible transcription, similar to silent promoters that have high levels of DNA methylation16–19. Thus, at bivalent promoters, the importance of DNA methylation for gene expression regulation is not readily apparent.

hESCs reflect a later developmental stage than mESCs. Here we have generated viable hESC lines with mutations in all three TET genes (TKO hESCs). Although hESCs have greater global CpG methylation than mESCs20, inactivating the TET genes still produces hypermethylation in a locus-specific manner. This hypermethylation is observed among enhancers and other regulatory regions, and is particularly prominent at bivalent promoters. In the absence of the TET proteins, the de novo methyltransferase DNMT3B causes aberrant hypermethylation at bivalent promoters, which leads to impaired gene activation upon differentiation. Thus, the TET proteins are necessary to maintain hypomethylation at bivalent promoters, which is critical for proper cellular differentiation during early human development.

Results

Bivalent promoter hypermethylation in TKO hESCs. Because all three TET genes were expressed in hESCs (Supplementary Fig. 1a)
and none had been genetically deleted previously, we used the iCRISPR platform developed in our laboratory\(^1\) to generate a panel of TET1\(^{-}\), TET2\(^{-}\) and TET3\(^{-}\)-knockout lines in the HUES8 and MEL-1 hESC backgrounds (Fig. 1a and Supplementary Tables 1 and 2). hESCs in which all three TET genes were inactivated (TKO hESCs) had no detectable 5hmC signal by mass spectrometry or 5hmC dot blot (Fig. 1b and Supplementary Fig. 1b,c) but showed no difference in morphology, self-renewal capacity or pluripotency marker expression when compared to wild-type (WT) hESCs (Fig. 1c,d and Supplementary Fig. 1d). However, TKO hESCs showed a complete inability to form teratomas and impaired induction of key early differentiation genes upon spontaneous embryoid body differentiation (Fig. 1e,f), suggesting that the TET proteins may be particularly important for the regulation of cellular differentiation.

Loss of the TET genes resulted in locus-specific hypermethylation rather than a global gain of methylation. Mass spectrometry analysis did not show a difference in 5mC levels between TKO and WT hESCs (Fig. 1b and Supplementary Fig. 1c), similar to previous findings in mESCs\(^1\). Instead, whole-genome bisulfite sequencing (WGBS) of WT and TKO HUES8 hESCs identified 3,523 hypermethylated differentially methylated regions (hyper-DMRs) each with at least five hypermethylated CpGs and a \(\geq 10\%\) methylation

**Fig. 1 | TET TKO hESCs exhibit differentiation defects.** a, TET-knockout mutants were generated using CRISPR gRNAs (arrowheads) that target the sequences corresponding to the beginning of the catalytic domain in TET1, TET2 and TET3. b, Analysis of 5hmC (left) and 5mC (right) in WT and TET-knockout HUES8 hESCs by mass spectrometry. For all mass spectrometry analyses, two mutant lines were used for all knockout genotypes except for the TKO genotype; for TKO, two different passages of the same line were used. Human fibroblasts were used as a negative control for mass spectrometry analysis of 5hmC. Data are presented as means \(\pm\) s.d. Black lines indicate comparisons to WT. Statistical analysis was performed by one-way ANOVA: ****\(P < 0.0001\); ns, not significant. c, The pluripotency markers NANOG, OCT4 and SOX2 were detected by immunofluorescence. Scale bar, 100 \(\mu\)m. d, Growth curves for WT and TKO hESCs; \(n = 3\) independent experiments. Data are presented as means \(\pm\) s.d. e, Hematoxylin and eosin staining of teratoma sections from WT teratomas. Arrows point to representative tissues for the respective germ layers. No teratomas were obtained 6 months after injection of TKO hESCs. One of the mutated TET1 alleles in the TKO hESCs was repaired to the WT sequence to form a rescued line, TKO-r1. f, Expression of markers of mesoderm (GSC, \(T\)), endoderm (SOX17, FOXA2), neuroectoderm (OTX2, PAX6, SOX1, FOXG1) and neural crest (SOX10) at day 12 of spontaneous embryoid body differentiation; \(n = 3\) independent experiments. Data are presented as means \(\pm\) s.d. Statistical analysis was performed by Student’s t test (two-sided): *\(P < 0.05\), **\(P < 0.01\), ***\(P < 0.001\).
difference when comparing TKO to WT hESCs (Supplementary Data 1). Using the same criterion, we also observed 3,832 hypo-methylated differentially methylated regions (hypo-DMRs) (Fig. 2a). These hypo-DMRs primarily occurred at CpGs outside of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased methylation of CpG islands (CGIs) (Fig. 2b). They could be a direct result of TET gene inactivation or a secondary effect, possibly due to redirection of the DNMT proteins to novel sites in TKO hESCs. Notably, CGIs were enriched in regulatory regions and showed increased
methylation in TKO hESCs (Fig. 2b). Furthermore, hyper-DMRs were enriched for regulatory regions such as promoters and enhancers (Fig. 2c). In comparison to other promoter types, bivalent promoters had the greatest magnitude of methylation increase (Fig. 2b), and a relatively high proportion of bivalent promoters gained methylation (Fig. 2d), as exemplified by the HOXA7 and HOX59 loci[11] (Supplementary Fig. 1e and Supplementary Data 2). Indeed, significant gain of methylation in TKO hESCs was found to center around genomic sites with both H3K4me3 and H3K27me3 marks (Fig. 2e). Finally, we performed genome-wide 5hmC profiling through 5hmC-Seal[12] and found that DNA hypermethylation in TKO hESCs was detected most strongly in regions that had 5hmC peaks in WT hESCs (Fig. 2e), supporting our hypothesis that loss of the TET proteins is directly responsible for gain of methylation. Gene Ontology (GO) analysis showed that hypermethylated bivalent promoters in particular, but also hypermethylated poised enhancers (marked by H3K4me1 but not H3K27ac), were associated with developmental categories, suggesting that the methylation aberrations could be responsible for the differentiation defects of TKO hESCs (Supplementary Fig. 2a).

To further investigate the methylation changes at bivalent promoters, we performed enhanced reduced-representation bisulfite sequencing (ERRBS), which focuses on CGIs and thus increases sequencing coverage of promoter regions[13]. By ERRBS, we observed similar methylation changes in the MEL-1 TKE line as in the HUES8 TKO line. ERRBS analysis showed ~12,000 hyper-DMRs and an insignificant number of hypo-DMRs in TKO lines as compared to WT hESCs (Fig. 2a and Supplementary Data 3 and 4). We further analyzed the methylation changes according to promoter types: bivalent promoters (marked by H3K4me3 and H3K27me3), active promoters (marked by H3K4me3 and H3K79me2), initiated promoters (marked by H3K4me3 only) and silent promoters (based on the absence of H3K4me3)[14]. Overall, about half of bivalent promoters showed hypermethylation by ERRBS and WGBS analyses (Fig. 2d and Supplementary Fig. 2b). Although WGBS also showed that a fraction of promoters became hypomethylated after TET gene inactivation, this was not observed by ERRBS. The hypomethylation detected by WGBS is likely due to hypomethylation of non-CGI CpGs that are frequently found within or near promoter regions.

In comparison to other promoter types, bivalent and silent promoters showed a greater magnitude of methylation increase in HUES8 and MEL-1 TKE lines (Fig. 2f and Supplementary Fig. 2c). Notably, we observed increasing fractions of bivalent promoters among promoters with greater methylation changes (Fig. 2g and Supplementary Fig. 2d). Although bivalent promoters composed 27% of all promoters evaluated by ERRBS, they constituted 60–74% of all promoters with a >60% increase in DNA methylation. We observed similar results in HUES8 and MEL-1 TKE hESCs, with 1,326 and 1,579 bivalent promoters gaining methylation, respectively (Supplementary Data 5 and 6). Individual bivalent promoters showed similar methylation changes in these two lines (Supplementary Fig. 2e), and 87% of the 1,326 hypermethylated bivalent promoters in HUES8 TKO hESCs also gained methylation in MEL-1 TKO hESCs (Fig. 2h). Overall, these results indicate that the TET proteins are critical to preserve hypomethylation at bivalent promoters and, in their absence, a reproducible subset of bivalent promoters becomes aberrantly hypermethylated.

We also compared our data with previous TKO mESC data[11]. Of the 732 bivalent promoters that became hypermethylated (>20% increase) in TKO mESCs, 517 were also bivalent in hESCs, of which 289 (~56%) were associated with hyper-DMRs in TKO hESCs. The remaining 215 promoters were not bivalent in hESCs, and only 55 (~25.6%) were associated with hyper-DMRs in TKO hESCs (Supplementary Fig. 2f). Thus, bivalent promoters found in both mESCs and hESCs tend to show similar methylation changes after TET gene inactivation.

Hypermethylation of the PAX6 promoter. TKO hESCs showed relatively few transcriptional changes in comparison to WT cells (Supplementary Fig. 3a). We found that hypermethylation of active, initiated and silent promoters was associated with a decrease in gene expression. In contrast, genes associated with hypermethylated bivalent promoters and poised enhancers did not show an overall change in expression in TKO hESCs (Supplementary Fig. 3b). 5mC MassARRAY EpiTYPER analysis confirmed that the bivalent promoters of selected developmental genes (FOX2A, GATA2, PAX6, SOX10 and SOX17) showed significant hypermethylation in TKO hESCs, whereas the active promoters for housekeeping and pluripotency genes did not (Supplementary Fig. 3c). Hypermethylation of the LEFTY2 enhancer (an active enhancer marked by H3K4me1 and H3K27ac) in TKO hESCs was associated with a significant decrease in gene expression, as has previously been described in TKO mouse embryos[16] (Supplementary Fig. 3d,c), whereas no expression change was detected for housekeeping and pluripotency genes. Notably, hypermethylation of the bivalent promoters was not associated with a consistent decrease in gene expression. For instance, no change in PAX6 expression was observed. A few genes associated with bivalent promoters showed up- or downregulation, but expression levels were generally low (Supplementary Fig. 3d and Supplementary Data 7).

We were intrigued that, despite promoter hypermethylation, genes with bivalent promoters such as PAX6 did not show a change in expression at the hESC stage (Supplementary Fig. 3d). Noticing that PAX6 (as well as FOX2A and SOX10) failed to be upregulated upon spontaneous embryoid body differentiation (Fig. 1f), we speculated that the hypermethylation at bivalent promoters could affect activation of gene expression following exposure to differentiation signals. We chose to focus on PAX6 because it is a critical gene for human neural development and is highly expressed during in vitro differentiation of hESCs into the neural lineage[17–20], thus allowing us to use neural differentiation to track the effects of TET gene deletion on PAX6 expression as well as cellular differentiation. PAX6 has well-annotated promoters and enhancers (Fig. 3a and Supplementary Fig. 4a)[21,22]. 5mC MassARRAY analysis of these regions revealed that only the bivalent P0 promoter of PAX6 showed hypermethylation in TKO HUES8 hESCs (Fig. 3b and Supplementary Fig. 4b), which was recapitated in the MEL-1 TKE line (Supplementary Fig. 4c). We further generated two ‘TKO-repaired’ lines, TKO-r1 and TKO-r2, through CRISPR–Cas9-mediated homology-directed repair (HDR) for comparison with isogenic WT and TKO hESC lines to rule out the possibility of CRISPR off-target effects (Fig. 3c). TKO hESCs that underwent the HDR targeting but retained the original biallelic TET1 mutations, TKO-r, were used as passage-matched controls. Repair of one TET1 allele to the WT sequence in TKO hESCs was sufficient to restore 5mcM to near WT levels as assessed by mass spectrometry (Fig. 3d), and it also reversed PAX6 P0 promoter hypermethylation (Fig. 3b). Notably, the TKO-repaired line was able to form teratomas, indicating a rescue in differentiation capacity (Fig. 1e).

To confirm that hypermethylation of the PAX6 P0 promoter is a direct consequence of losing the TET proteins, we performed 5mC analysis using the Epimark 5mC Analysis kit[23] and genome-wide hMe-Seal[24]. Both methods showed 5mC enrichment at the PAX6 P0 promoter in WT hESCs (Fig. 3e,f). We also detected TET1 binding at the PAX6 P0 promoter in WT hESCs by ChiP-seq and ChiP-qPCR (Fig. 3f,g). Approximately 51% of 5mC peaks at promoters and 75% of 5mC peaks globally overlapped with TET1 peaks (Supplementary Fig. 4d,e). Incomplete overlap between 5mC and TET1 peaks could be due to 5mC production by TET2 or TET3. Globally, TET1 bound to bivalent, active and initiated promoters at the transcription start site (TSS). Binding of TET1 overlapped with 5mC signals, which extended into the promoter and gene body (Fig. 3h and Supplementary Fig. 4f).
**Fig. 3 | TET1 TKO hESCs show hypermethylation at the PAX6 P0 bivalent promoter.**

a, Schematic for analysis of 5mC, 5hmC and TET1 binding at the PAX6 locus. Arrows represent the P0 and P1 promoters, the gray box represents the PAX6 mRNA transcript and the black box represents the PAX6 protein. The region analyzed for 5mC using MassARRAY and for 5hmC using hMe-Seal profiling is indicated by a green box. 
b, Heat map of MassARRAY analysis of 5mC at the PAX6 P0 promoter. The location (in base pairs) of each row of CpGs with respect to the P0 TSS is shown to the left of the heat map. The color key for percent methylation is shown to the right of the heat map. For each cell line, three independent experiments are shown as three columns. NE, neuroectoderm differentiation. Statistical analysis was performed by Student’s t test (two-sided): ****P < 0.0001.
c, Diagram of HDR of the TET1 locus in TKO hESCs. Red letters correspond to mutations of the WT sequence. The sequences of the two repaired lines (TKO-r1, TKO-r2) are shown below. 
d, Mass spectrometry analysis of 5hmC levels in WT hESCs, TET1-knockout lines, TET1-knockout lines in which one allele of TET1 was repaired (TKO-R) and TET1-knockout lines that underwent HDR targeting but retained the mutations in the TET1 locus (TKO-nr). For all mass spectrometry analysis, two mutant lines were used for all genotypes except for the TKO genotype; for TKO, two different passages of the same line were used. Fib, fibroblasts. Data are presented as means ± s.d. Statistical analysis was performed by one-way ANOVA: ***P < 0.001. 
e, Analysis of the percentage of 5hmC at the PAX6 P0 promoter by Epimark; n = 3 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by Student’s t test (two-sided): ****P < 0.0001.
f, Top track, analysis of the 5hmC peak at the PAX6 P0 promoter by hMe-Seal in WT hESCs. The shaded areas indicate the regions of the PAX6 P0 promoter assayed for 5hmC by Epimark in e and for TET1 binding by ChIP–qPCR in g. ChIP–qPCR for TET1 in WT and TKO hESCs; n = 3 independent experiments. Primers that bind to the P0 promoter (PO_1, PO_2, PO_3), the P1 promoter (P1_1, P1_2), and the upstream (E6k) and downstream (E156k) enhancers were used. Data are presented as means ± s.d. Statistical analysis was performed by Student’s t test (two-sided): *P < 0.05, **P < 0.01, ***P < 0.001.
g, Analysis of 5hmC and TET1 peaks at bivalent promoters in WT hESCs. The height above the x axis reflects the normalized tag count. 
h, Percentage of DNA methylation change (TKO – WT) at bivalent promoters with 5hmC peaks in WT hESCs (red line) as compared to bivalent promoters without 5hmC peaks in WT hESCs (dashed blue line).
**Fig. 4 | TET TKO hESCs show a defect in neuroectoderm differentiation.**

(a) Schematic of neuroectoderm differentiation. SB431542, TGF-β receptor inhibitor; LDN, BMP receptor inhibitor. 

(b) Left, representative FACS plots of PAX6 staining at days 4, 6, 8 and 10 of neuroectoderm differentiation for WT and TKO cells. Right, quantification of PAX6+ cells at days 4, 6, 8 and 10 of neuroectoderm differentiation; n = 3 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by Student’s t test (two-sided): ***P < 0.001.

(c) Immunofluorescence of PAX6, SOX1 and OCT4 at the endpoint of differentiation (day 10) for WT, TKO and TKO-r1 cells. Scale bar, 100 μm.

(d) Representative FACS plots (top) and quantification (bottom) of PAX6 staining at day 10 of neuroectoderm differentiation; n = 3 independent experiments. FSC, forward scatter. Data are presented as means ± s.d. Statistical analysis was performed by one-way ANOVA: ****P < 0.0001.

(e) Representative FACS plots (top) and quantification (bottom) of OCT4 staining at day 10 of neuroectoderm differentiation; n = 3 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by one-way ANOVA: ****P < 0.0001.

(f) qPCR analysis for epiblast (OTX2), neuroectoderm (PAX6, SOX1 and OTX2) and neural crest (SOX10) markers during neuroectoderm differentiation; n = 3 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by one-way ANOVA: *P < 0.05, **P < 0.01, ***P < 0.001.
The functional relevance of TET binding and 5hmC signal is supported by a greater methylation increase in TKO hESCs at bivalent promoters that had 5hmC peaks than at bivalent promoters that did not have 5hmC peaks (Fig. 3i and Supplementary Fig. 4g). Overall, ~92% and ~50% of hyper-DMRs at bivalent promoters overlapped with 5hmC and TET1 peaks, respectively (Supplementary Fig. 4h). The greater overlap between hyper-DMRs and 5hmC peaks is likely due to production of 5hmC by TET2 and TET3. Altogether, these findings support the conclusion that TET1 binding at bivalent promoters (such as the PAX6 P0 promoter) leads to 5hmC production while TET inactivation causes 5hmC depletion and aberrant promoter hypermethylation.

**Impaired neural differentiation of TKO hESCs.** Because PAX6 is expressed in early neuroectoderm derived from hESCs and has been shown to be both necessary and sufficient for neuroectoderm formation from hESCs, we speculate that hypermethylation of the PAX6 P0 promoter in TKO hESCs may impede hES differentiation into the neural lineage, as suggested by the embryoid body differentiation results (Fig. 1f).

We performed neuroectoderm differentiation using the dual-SMAD-inhibition protocol (Fig. 4a and Supplementary Fig. 5a). In contrast to HUES8 WT and TKO-repaired hESCs, HUES8 TKO hESCs formed significantly fewer PAX6+ neuroectoderm cells at each point during differentiation (Fig. 4b–d), suggesting a defect in the acquisition rather than the maintenance of the neuroectoderm fate. Notably, in differentiation of HUES8 TKO hESCs, ~40% of cells remained positive for OCT4 (also known as POU5F1) after 10 d of hypermethylation of the P0 promoter is responsible for the loss of PAX6 expression accounts for the neuroectoderm differentiation defect of TKO hESCs (Fig. 4f and Supplementary Fig. 5b–d). Notably, comparison of the HUES8 TKO line to HUES8 hESCs with individual and double knockout of TET genes showed that the severity of the neuroectoderm differentiation defect depended on TET gene dosage. Loss of TET1 had the largest effect on bulk 5hmC levels (Fig. 1b) as well as neuroectoderm differentiation, as determined by FACS, immunostaining, and RT–qPCR analysis for PAX6 and SOX1 expression (Supplementary Fig. 6a–c).

**Promoter hypermethylation hinders PAX6 expression upon differentiation.** PAX6 is expressed at a very low level in hESCs, and hypermethylation of the P0 promoter in TKO hESCs had no effect on PAX6 gene expression (Supplementary Fig. 3d). We hypothesized that hypermethylation of the PAX6 P0 promoter prevents activation of PAX6 expression upon differentiation and leads to the neuroectoderm differentiation defect in TKO hESCs. Supporting this hypothesis, 5mC MassARRAY analysis showed aberrant hypermethylation at the PAX6 P0 promoter in TKO hESCs both before and during neuroectoderm differentiation (Fig. 3b). To establish direct causality, we needed to determine whether loss of PAX6 expression accounts for the neuroectoderm differentiation defect of TKO hESCs and then investigate whether hypermethylation of the P0 promoter is responsible for the loss of PAX6 expression. We first performed a rescue experiment in which we expressed the PAX6 transgene under the control of a doxycycline-inducible promoter in TKO cells during neuroectoderm differentiation (Fig. 5a). TKO cells exposed to doxycycline were able to upregulate the neuroectoderm markers SOX1 and FOXG1 and downregulate the pluripotency markers POU5F1 and NANOG (Fig. 5b,c and Supplementary Fig. 7a). However, expression of SOX10 and endogenous PAX6 was not restored (Supplementary Fig. 7a). In comparison to the promoters of SOX1 and FOXG1, the SOX10 and PAX6 promoters showed a much greater methylation increase in TKO hESCs (Supplementary Data 5), which may prevent their expression even when the PAX6 transgene is overexpressed.

De novo methylation causes PAX6 promoter hypermethylation. The DNA methyltransferases—DNMT1, DNMT3A and DNMT3B—are responsible for cytosine methylation. We speculated that hypermethylation of the PAX6 P0 promoter and the resulting neuroectoderm differentiation defect could be due to
Fig. 5 | Hypermethylation of the PAX6 P0 bivalent promoter in TET TKO hESCs leads to a failure of PAX6 induction upon neuroectoderm differentiation.

a, Schema of rescue of the neuroectoderm differentiation defect in TKO hESCs using PAX6 overexpression or targeted demethylation of the PAX6 P0 promoter. b, qPCR analysis of neuroectoderm (SOX1) and pluripotency (POUSF1 (OCT4)) markers at day 10 of neuroectoderm differentiation in WT and TKO cells without (TKO) and with (TKO + PAX6) doxycycline treatment; n = 3 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by one-way ANOVA: *P < 0.05, **P < 0.01, ***P < 0.001. c, Immunofluorescence of PAX6, SOX1 and OCT4 at the endpoint of differentiation (day 10) for TKO cells without (TKO) and with (TKO + PAX6) doxycycline treatment. Scale bar, 100 μm. d, Heat map of MassARRAY analysis of 5mC at the PAX6 P0 promoter for TKO hESCs expressing PAX6-targeting gRNAs with either the dCas9-TET1CD/Mut (left) or dCas9-TET1CD (right) fusion protein. The location (in base pairs) of each row of CpGs with respect to the P0 TSS is shown to the left of the heat map. The color key for percent methylation is shown to the right of the heat map. Methylation analysis at the PAX6 P0 promoter was performed for these cell lines with and without doxycycline (Dox) treatment; n = 3 independent experiments. Statistical analysis was performed by Student’s t test (two-sided): ****P < 0.0001. e, qPCR of PAX6 expression on day 10 of neuroectoderm differentiation for TKO hESCs expressing PAX6-targeting gRNAs with either the dCas9-TET1CD/Mut or dCas9-TET1CD fusion protein. PAX6 expression was analyzed for these cell lines with and without doxycycline treatment before differentiation; n = 3 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by Student’s t test (two-sided): ***P < 0.001. f, Immunofluorescence of PAX6 on day 10 of neuroectoderm differentiation for TKO hESCs expressing PAX6-targeting gRNAs with either the dCas9-TET1CD/Mut or dCas9-TET1CD fusion protein. TKO hESCs expressing the dCas9-TET1CD fusion and a nontargeting gRNA were also used as a control. PAX6 immunofluorescence was analyzed for these cell lines with and without doxycycline treatment before differentiation. Scale bar, 100 μm.
increased expression or activity of the DNMTs in TKO hESCs. There were no differences in expression of the DNMT genes between WT and TKO hESCs (Supplementary Fig. 8a); however, ChIP-qPCR analysis showed increased binding of DNMT3B, but not DNMT1 or DNMT3A, at the PAX6 P0 promoter in TKO hESCs as compared to WT hESCs (Fig. 6a and Supplementary Fig. 8b). Thus, we further investigated whether DNMT3B is responsible for the hypermethylation at the PAX6 P0 promoter in TKO hESCs through genetic deletion. By applying a CRISPR gRNA that targets the sequence corresponding to the cytosine C5 methyltransferase domain of DNMT3B in TKO hESCs, we generated a TET1, TET2, TET3 and DNMT3B quadruple-knockout (QKO) line. QKO hESCs had a ~35% reduction in methylation at the PAX6 P0 promoter as compared to passage-matched TKO hESCs (Fig. 6b). Furthermore, there was partial rescue of the neuroectoderm differentiation phenotype as compared to TKO hESCs: QKO cells formed PAX6 and SOX1 double-positive cells (Fig. 6c), and few cells remained OCT4+ after 10 d of neuroectoderm differentiation. RT-qPCR analysis also showed rescue of expression for neuroectoderm markers PAX6 and SOX1 and proper downregulation of pluripotency markers POU5F1 and NANOG (Fig. 6d). Our results suggest that DNMT3B activity at the PAX6 P0 promoter is responsible for the hypermethylation and neuroectoderm differentiation phenotypes observed in TKO hESCs. In WT hESCs, TET proteins or resulting 5hmC marks may...
We performed ERRBS for passage-matched WT, TKO and QKO lines to investigate whether our findings at the PAX6 P0 promoter apply to other bivalent promoters. We found that QKO hESCs showed a global decrease in methylation, both within and outside of CGIs, and in all promoter types. Bivalent promoters showed a larger decrease in methylation between QKO and TKO hESCs than active and initiated promoter types, but less of a decrease than silent promoters (Fig. 8a, b, Supplementary Fig. 8c, d and Supplementary Data 8). Furthermore, bivalent promoters, in comparison to non-bivalent promoters, had greater overlap with hyper-DMRs that were present after TET gene inactivation as well as higher overlap with hypo-DMRs that were present after DNMT3B inactivation in the TKO background (Fig. 8c and Supplementary Fig. 8e). This suggests that the TET and DNMT proteins dynamically regulate the methylation state of bivalent promoters. Approximately 90% of the bivalent promoters that lost methylation in QKO hESCs had previously gained methylation in TKO hESCs (in comparison to WT hESCs). Conversely, ~57% of the bivalent promoters that gained methylation after TET gene inactivation lost methylation after DNMT3B was mutated (Fig. 8d). Thus, at many bivalent promoters, continuous DNMT3B activity is needed for the hypermethylation phenotype in TKO hESCs.

Our results thus far indicate that DNMT3B is responsible for the majority of the bivalent promoter hypermethylation that occurs after TET gene inactivation. We performed DNMT3B ChIP–seq to investigate whether bivalent promoters are directly targeted by DNMT3B. DNMT3B showed relatively insignificant overall binding to promoter regions (Supplementary Fig. 8e) in WT and TKO hESCs, similar to previous results of DNMT3B ChIP–seq in WT mESCs40. This may be due to weak DNMT3B binding or technical difficulties in DNMT3B ChIP. Nevertheless, among the promoters with DNMT3B peaks in TKO hESCs, 74% of the 293 bivalent promoters and 21% of the 1,017 non-bivalent promoters gained methylation after TET gene inactivation. Furthermore, bivalent promoters with DNMT3B peaks in TKO hESCs showed greater methylation increase (TKO versus WT) and greater methylation decrease (QKO versus TKO) than bivalent promoters that lacked DNMT3B peaks in TKO hESCs and non-bivalent promoters with DNMT3B peaks in TKO hESCs (Fig. 8e). Thus, DNMT3B binding correlates with more dynamic changes in DNA methylation at bivalent promoters, suggesting that at these promoters the TET and DNMT3B proteins function in a competitive manner.

**Discussion**

In mouse and human ESCs25,41, the promoter regions of differentiation-associated genes are enriched for bivalent marks. Previous studies have focused on the establishment and maintenance of bivalent histone marks. However, it was unclear whether DNA methylation is also actively regulated at bivalent promoters and, if so, whether it has functional relevance for cell differentiation. Here we show that the TET proteins are critical for maintaining a hypomethylated state at bivalent promoters in hESCs. Although we focused on lineage regulators such as PAX6 and SOX10, bivalent promoters may also regulate signaling pathways during development that, along with regulation of enhancers and other regulatory regions, contribute to proper embryonic development and cellular differentiation. Notably, alteration of DNA methylation at bivalent promoters does not cause immediate changes in transcription in hESCs but nonetheless impairs hESC differentiation. We therefore propose to revise the predominant approach for studying epigenetic regulators, which focuses on epigenetic changes (including changes in DNA methylation) that have an immediate impact on gene expression, by expanding the analysis to additional epigenetic changes that do not immediately affect gene expression but may influence future cell behavior during embryonic development or adult stem/progenitor cell differentiation.

Previous studies have indicated that genomic regions marked by the H3K4me3 modification are refractory to de novo DNA methylation42,43. TET gene deletion causes bivalent promoter hypermethylation without causing significant changes in H3K4me3 occupancy (Supplementary Fig. 8f). One possibility is that some of the effects associated with the H3K4me3 mark could be due to TET proteins, TET-dependent 5hmC, or TET-associated proteins, which may often co-occur with H3K4me344. A second, non-mutually exclusive possibility is that de novo methyltransferases may be actively recruited to bivalent promoters in the absence of TET proteins, overcoming the repulsion by H3K4me3. In fact, it was recently shown that the Polycomb repressive complex 2 (PRC2) recruits DNMT3L, DNMT3A and DNMT3B to the bivalent promoters of genes involved in germ cell differentiation45.

It was also interesting that this aberrant hypermethylation at bivalent promoters, once established in TKO hESCs, is not adequately maintained by DNMT1. On the basis of the >99.7% overall fidelity of DNMT1 in preserving methylation46, DNMT1 would be expected to largely preserve the hypermethylation seen in TKO hESCs during the approximately six passages it takes to generate and expand QKO cells for analysis. Yet, we observed a global reduction in DNA methylation at bivalent promoters, including the PAX6 P0 promoter, in QKO cells, indicating that the hypermethylation phenotype requires continuous DNMT3B activity. This result also suggests that additional mechanisms, such as transient transcription47,
Fig. 8 | DNMT3B regulates the methylation level at bivalent promoters. 

**a**, Average methylation at different genomic regions and bivalent promoters for WT, TKO and QKO hESCs by ERRBS; n = 2 independent experiments. Data are presented as means ± s.d. Statistical analysis was performed by one-way ANOVA: *P < 0.05, **P < 0.01. 

**b**, Percentage methylation in WT, TKO and QKO hESCs for active, initiated, bivalent and silent promoters. Error bars show the 10–90% confidence interval. The lower and upper limits of the box represent the first and third quartile, respectively, and the bar at the center of the box indicates the median. n = 2 independent experiments. Statistical analysis was performed by one-way ANOVA: **P < 0.01, ***P < 0.001, ****P < 0.0001.

**c**, Top, overlap of bivalent promoters with hyper-DMRs (TKO versus WT) and hypo-DMRs (QKO versus TKO) at promoter regions. Bottom, overlap of non-bivalent promoters with hyper-DMRs (TKO versus WT) and hypo-DMRs (QKO versus TKO) at promoter regions. The odds ratio (OR) and P value for the comparison between bivalent and non-bivalent promoters are provided (Fisher’s exact test).

**d**, Overlap between bivalent promoters associated with hyper-DMRs (TKO versus WT) and hypo-DMRs (QKO versus TKO).

**e**, Left, methylation changes (TKO – WT) for bivalent and non-bivalent promoters that either had (+ DNMT3B) or did not have (– DNMT3B) DNMT3B peaks in TKO hESCs. Right, methylation changes (QKO – TKO) for bivalent and non-bivalent promoters that either had (+ DNMT3B) or did not have (– DNMT3B) DNMT3B peaks in TKO hESCs. Error bars show the 10–90% confidence interval. The lower and upper limits of the box represent the first and third quartile, respectively, and the bar at the center of the box indicates the median. n = 2 independent experiments. Statistical analysis was performed by one-way ANOVA: ****P < 0.0001.
may inhibit DNM1 activity at these loci, which may be an additional protective measure against hypermethylation of bivalent promoters even in TKO hESCs. Alternatively, although not necessarily in a mutually exclusive manner, DNM3B may function as both a de novo and maintenance methyltransferase, as previously suggested. It must be noted, however, that the maintenance function of DNM3A and DNM3B has only been observed for limited genomic regions in mESCs, such as repetitive sequences, and inactivation of both DNM3A and DNM3B in hESCs only causes gradual DNA demethylation.

Previous studies have shown that DNM3A and DNM3B have largely overlapping targets, yet deletion of DNM3B alone in TKO hESCs was sufficient to partially reverse bivalent promoter hypermethylation along with the associated neuroectodermal differentiation defect. DNM3B may have stronger preference or activity, in comparison to DNM3A, at bivalent promoters such as the PAX6 P0 promoter, and such loci have been identified previously in hESCs. Perhaps more likely, the rescue by DNM3B deletion alone could be due to the relatively low DNM3A expression as compared to that of DNM3B (Supplementary Fig. 8a) and the lack of compensatory increase in DNM3A expression upon DNM3B deletion (Supplementary Fig. 8g). Similarly, a recent paper found that inactivating either DNM3A or DNM3B in TET TKO mouse embryos was able to rescue the gastrulation phenotype. It is possible that in other cell types DNM3A, along with or instead of DNM3B, may counteract the TET proteins to regulate DNA methylation.

Our work highlights the utility of locus-specific epigenome editing tools to directly probe the functional consequences of epigenetic changes and to distinguish direct, locus-specific effects from indirect effects. The competitive balance between the TET proteins and de novo methyltransferases at bivalent promoters and other genomic loci could facilitate rapid changes in methylation state to either activate or silence transcription in a cell-lineage- and locus-specific manner. Further work could investigate the factors that influence whether the methylation state of a genomic region is dynamically regulated and, ultimately, how methylation states predict (in the context of cell differentiation) cell-type-specific transcriptional programs.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0002-y.

Received: 31 October 2016; Accepted: 25 October 2017; Published online: 4 December 2017

References

1. Ziller, M. J. et al. Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477–481 (2013).
2. Robertson, K. D. DNA methylation and human disease. Nat. Rev. Genet. 6, 597–610 (2005).
3. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915–926 (1992).
4. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257 (1999).
5. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14, 204–220 (2013).
6. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935 (2009).
7. Itô, S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466, 1129–1133 (2010).
8. Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxilation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell 145, 423–434 (2011).
9. Dawlaty, M. M. et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev. Cell 29, 102–111 (2014).
10. Dai, H. Q. et al. TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. Nature 538, 528–532 (2016).
11. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. I. & Zhang, V. Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. 28, 2103–2119 (2014).
12. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 33 (Suppl.), 245–254 (2003).
13. Kouzarides, T. Chromatin modifications and their function. Cell 128, 645–653 (2007).
14. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 30, 733–750 (2016).
15. Bogoianov, O. et al. Active DNA demethylation at enhancers during the vertebrate phyloptotic period. Nat. Genet. 48, 417–426 (2016).
16. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560 (2007).
17. Fouse, S. D. 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 2, 160–169 (2008).
18. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315–322 (2009).
19. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766–770 (2008).
20. von Meyenn, F. et al. Comparative principles of DNA methylation reprogramming during human and mouse in vitro primordial germ cell specification. Dev. Cell 39, 104–116 (2015).
21. González, F. et al. An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell 15, 215–226 (2014).
22. Song, C. X. et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. 29, 68–72 (2011).
23. Gu, H. et al. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat. Protoc. 6, 468–481 (2011).
24. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. 12, 7–18 (2011).
25. Zhang, X. et al. Pax6 is a human neuroectodermal cell fate determinant. Cell Stem Cell 7, 90–100 (2010).
26. Li, X. J. et al. Specification of motoneurons from human embryonic stem cells. Nat. Biotechnol. 23, 215–221 (2005).
27. Pankratz, M. T. et al. Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. Stem Cells 25, 1511–1520 (2007).
28. Anderson, T. R., Hedlund, E. & Carpenter, E. M. Differential Pax6 promoter activity and transcript expression during forebrain development. Mech. Dev. 114, 171–175 (2002).
29. Kamandari, B. et al. Distinct cis-essential modules direct the time–space pattern of the Foxa2 gene activity. Dev. Biol. 205, 79–97 (1998).
30. Plaza, S., Saule, S. & Dozier, C. High conservation of cis-regulatory elements between quail and human for the Pax-6 gene. Dev. Genes Evol. 209, 165–173 (1999).
31. Xu, Z. P. & Saunders, G. F. Pax6 intronic sequence targets expression to the spinal cord. Dev. Genet. 23, 259–263 (1998).
32. Zheng, F. B., Zhou, Y. H., Masty, T., Liao, W. S. & Saunders, G. F. Activation of the human Pax6 gene through the exon 1 enhancer by transcription factors SEF and Sp1. Nucleic Acids Res. 29, 4070–4078 (2001).
33. Stroud, H. et al. Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 12, R34 (2011).
34. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iPSC cells by dual inhibition of SMAD signaling. Nat. Biotechnol. 27, 275–280 (2009).
35. Xu, X. et al. A CRISPR-based approach for targeted DNA demethylation. Cell Discov. 2, 16009 (2016).
36. Lui, X. S. et al. Editing DNA methylation in the mammalian genome. Cell 167, 233–247 (2016).
37. Choudhary, S. R., Cui, Y., Lubecka, K., Stefanaka, R. & Irudayaraj, J. CRISPR–dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget 7, 46545–46556 (2016).
38. Morita, S. et al. Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. Nat. Biotechnol. 34, 1511–1520 (2016).
39. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNM3B in genic methylation. Nature 520, 243–247 (2015).
40. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326 (2006).
41. Liu, J. et al. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell 1, 299–312 (2007).
42. Guo, H. et al. The DNA methylation landscape of human early embryos. Nature 511, 606–610 (2014).
43. Ooi, S. K. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
44. Xu, Y. et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**, 451–464 (2011).
45. Neri, F. et al. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell* **155**, 121–134 (2013).
46. Liao, J. et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* **47**, 469–478 (2015).
47. Di Ruscio, A. et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
48. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
49. Liang, G. et al. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell. Biol.* **22**, 480–491 (2002).
50. Zhang, X. et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat. Genet.* **48**, 1014–1023 (2016).
51. Rinaldi, L. et al. Dnmt3a and Dnmt3b associate with enhancers to regulate human epidermal stem cell homeostasis. *Cell Stem Cell* **19**, 491–501 (2016).
52. Pulecio, J., Verma, N., Meija-Ramirez, E., Huangfu, D. & Raya, A. CRISPR/Cas9-based engineering of the epigenome. *Cell Stem Cell* **21**, 431–447 (2017).

Acknowledgements

We thank the WCMC Epigenomics core for ERRBS and 5mC MassARRAY analysis and the MSKCC Integrated Genomics core for performing RNA-seq and WGBS. We also thank L. Studer and J. Tchieu for advice on neuroectoderm differentiation, E. Apostolou and M. Donohoe for comments on the manuscript, and M. Goll and members of the Huangfu laboratory for insightful discussions and critical reading of the manuscript. This study was funded in part by the Tri-Institutional Stem Cell Initiative (2016-032), New York State Stem Cell Science (NYSTEM C029156) and an MSKCC Cancer Center Support grant (P30 CA008748). N.V. is supported by the Weill Graduate School of Medical Sciences at the Cornell University/The Rockefeller University/Sloan Kettering Institute Tri-Institutional MD–PhD Program.

Author contributions

N.V. and D.H. devised experiments and interpreted results. N.V. performed most experiments and collected data. H.P. and O.E. performed computational analysis on WGBS, ERRBS, RNA-seq, ChIP–seq and 5hmC profiling. L.C.D. and C.H. performed 5hmC profiling, sequencing and analysis, and mass spectrometry. A.K. generated libraries for WGBS. A.S., Q.V.L., B.P.-W., V.T., F.G., E.P.P. and C.-J.C. assisted with additional experiments. N.V. and D.H. wrote the manuscript; all other authors provided editorial advice.

Competing interests

5hmC-Seal has been licensed to Active Motif and Epican by the University of Chicago. This statement is relevant to C.H.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-017-0002-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to O.E. or D.H.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
**Methods**

**hESC culture.** HUES8 (NHIIhESC-09-0021) and MEI-1 (NHIIhESC-11-0139) hESCs were cultured on irradiated mouse embryonic fibroblast (iMEF) feeder layers in DMEM:F12 medium (Life Technologies, 12500-106) supplemented with 20% KnockOut Serum Replacement (Life Technologies, 10829028), 1 μM Non-Essential Amino Acids (Life Technologies, 11104050), 1× GlutaMax (Life Technologies, 35050079), 100 U/ml penicillin and 100 µg/ml streptomycin (Gemini, 15070063), 0.055 mM 2-mercaptoethanol (Life Technologies, 21985023) and 10 ng/ml recombinant human bFGF (EMD Millipore, GP9058F). Cells were incubated at 37°C with 5% CO2, and medium was changed daily. Cultures were passaged at a 1:6 to 1:12 split ratio every 4–6 d using TrypLE (Life Technologies, 12563-029). 5 μM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals, S1049) was added into the culture medium when passaging or thawing cells. Cells are regularly confirmed to be mycoplasma free by the MSKCC Antibody & Bioresource Core Facility.

**Generation of mutant lines.** In vitro transcription of gRNAs. CRISPR gRNAs were designed to target the genomic sequences corresponding to the beginning of the catalytic domains for TET1, TET2 and TET3. For TET1 and TET3, the two gRNAs most efficient at producing indel mutations were used for the targeting experiments. For TET2, only one gRNA was found to be efficient at producing frameshift indel mutation. A CRISPR gRNA from a previous study was used to target the sequence encoding the C5 methyltransferase domain of DNMT3B46. The results were confirmed to be efficient at producing only one gRNA was found to be efficient at producing frameshift indel mutation. The reactions were cleaned up using a Zymo DNA Clean & Concentrator kit (Zymo, D4003), and a biont moiety was then added to the aldehyde-labeled DNA via a copper-free click chemistry reaction with BDO-S-S-PEG3-biotin in water at 37°C for 1 h. Reactions were then centrifuged to separate the Zymo kit and then bound to the MSKCC Integrated Genomics Core for library preparation and sequencing.

**RNA-seq.** For RNA-seq, total RNA was isolated with the RNeasy Mini kit (Qiagen, 74136) from WT and TKO HUES8 hESCs (n = 2 cultures each). RNA samples were submitted to the MSKCC Integrated Genomics Core for library preparation and sequencing.

RNA-seq data were aligned to the hg19 reference genome using Bowtie1. Read counts were derived from the HTSeq scripts.count module in HTSeq-0.6.0 with default parameters in a non-directional model. Lists of differentially expressed genes were generated by DESeq2-1.4.5 in R. Unregulated genes were defined by log(fold change) > 2 and P < 0.1 (n = 134). Downregulated genes were defined by log(fold change) < -2 and P < 0.1 (n = 233).

**5hmC profiling.** 5hmC-Seal was performed as previously described1. 40 µg of genomic DNA was sonicated into 200–400 bp fragments using a Diagenode Bioruptor Sonicator. Sonicated DNA was then labeled with azide glucose in a 1:3 reaction at 37°C catalyzed by recombinant β-GT using UDP-6-N3-glucose as the sugar donor. The reactions were cleaned up using a Zymo DNA Clean & Concentrator kit (Zymo, D4003), and a biont moiety was then added to the aldehyde-labeled DNA via a copper-free click chemistry reaction with BDO-S-S-PEG3-biotin in water at 37°C for 1 h. Reactions were then centrifuged to separate the Zymo kit and then bound to the MSKCC Integrated Genomics Core for library preparation and sequencing.
and silent. Active promoters were associated with H3K4me3 (in the 1.5-kb region flanking the TSS) and H3K79me2 (in the 5-kb region downstream of the TSS). Initiated promoters were associated with H3K4me3 (in the 1.5-kb region flanking the TSS), bivalent promoters were associated with H3K4me3 and H3K27me3 (in the 1.5-kb region flanking the TSS). Silent promoters were not associated with H3K4me3 (in the 1.5-kb region flanking the TSS). With WGBS (TKO versus wild-type HUES8 cells), 6,695 active promoters, 989 initiated promoters, 3,327 bivalent promoters and 4,707 silent promoters were analyzed. With ERRBS (TKO versus wild-type HUES8 cells), 6,450 active promoters, 846 initiated promoters, 3,234 bivalent promoters and 1,476 silent promoters were analyzed. With ERRBS (TKO versus wild-type MEL-1 cells), 6,420 active promoters, 3,222 bivalent promoters, 835 initiated promoters and 1,413 silent promoters were analyzed.

Enhancers were defined as regions outside of promoters and exons (RefSeq). Poised enhancers were defined as regions overlapping H3K4me1 peaks only. Active enhancers were identified as regions overlapping H3K4me1 and H3K27ac peaks.

Overlap of 5mc changes with histone marks was determined using previous ChIP-seq datasets for histone marks (H3K4me1, GSM733782; H3K4me2, GSM733780; H3K4me3, GSM733786; H3K27ac, GSM733784; H3K27me3, GSM733785; H3K9me3, GSM733788; H3K9me2, GSM733789; and H3K36me3, GSM733790).

ChIP-seq. ChIP-seq was performed for TET1 (WT hESCs) and DNMT3B (WT and TKO hESCs). WT and TKO hESCs were cultured in standard hESC medium, as described above. Approximately 5 × 10^6 cells were fixed, washed and snap frozen according to the Cell Fixation protocol from Active Motif (see URLs). ChIP and DNA sequencing were performed by Active Motif.

ChIP-seq data were aligned to the hg19 reference genome using bowtie 2-0.12.9 with default parameters except --n 2 -best. Peak calling and analysis of read density in peak regions were performed by macs14 1.4.2 with default parameters11. The technical success of ChIP-seq analysis was confirmed using standard quality control measures and determined the overlap between TET1 binding and the presence of 5hmC. We observed that 52.6% of 5hmC peaks associated with gene promoters overlapped TET1 peaks (Supplementary Fig. 5f), similar to previous results obtained from TET1 ChIP-seq in mESCs44,62. Peaks associated with gene promoters overlapped TET1 peaks (Supplementary Fig. 5f), similar to previous results obtained from TET1 ChIP-seq in mESCs44,62. Approximately 75% of total TET1 peaks overlapped with total 5hmC peaks present in WT HUES8 hESCs; in contrast, only 25% of TET1 peaks overlapped with randomly generated 5hmC peaks (Supplementary Fig. 4e). Random peaks were generated with ChIPseekerCreateRandomRegions in ChIPseeker-2.1 with default parameters. The lack of a complete overlap between 5hmC and TET1 peaks could be due to a number of factors, including the following: 5hmC production by TET2 and TET3, rapid turnover of 5hmC and reduced binding of TET1 to 5hmC. Peaks from ChIP-seq were annotated using ChIPseekerAnnotate from the ChIPseeker package.

Cloning of the dCas9–TET1 catalytic domain fusion vector (described above) was modified at two sites (p.His1671Ala and p.Asp1673Ala) to produce a catalytic-null control (dCas9–TET1CD/Mut). Mutagenesis was performed using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies, 205251) according to the manufacturer's protocol. The primers used for mutagenesis are listed in Supplementary Table 6.

Expressing dCas9–TET1CD in TKO hESCs. The dCas9–TET1CD construct was cloned into the pINDUCER lentiviral backbone (Addgene, plasmid 46948) under the control of a doxycycline-inducible promoter. This lentiviral backbone also constitutively expresses GFp, which allowed us to isolate infected cells. Prior to infection of TKO hESCs with pINDUCER-dCas9–TET1CD, the already AAVS1-integrated Cas9 was inactivated through CRISPR targeting of the sequence at the beginning of the Cas9 protein. The gRNA used to inactivate Cas9 is listed in Supplementary Table 1. After knocking out Cas9, clones were isolated and infected into the pCRg Entry vector to be used as a nontargeting control. For the PA6 P0 promoter, we identified three gRNAs (C6, Cr7 and Cr9) that were able to produce demethylation. We performed Golden Gate cloning to assemble these three gRNAs in tandem. An LR reaction (Thermo Fisher Scientific, 11791100) was then used to transfer the U6 promoter and either the tandem array of Cr6, Cr7 and Cr9 gRNAs for the PA6 P0 promoter or the individual gRNAs for SOX10, LEFTY2 and HBB (nontargeting control) into a lentiviral backbone containing a hygromycin selection cassette. TKO hESCs containing the dCas9–TET1CD construct were then infected by the lentiviruses. Infected cells were isolated by 4 d of hygromycin selection (40 μg/ml) and then amplified and frozen down. As described in the text, targeted clones were treated with doxycycline for 10 d before methylation and RNA expression analyses at the hESC stage. For targeted demethylation of PA6 and SOX10, cells were also differentiated using the standard neuroectoderm differentiation protocol and RNA expression was then analyzed by qPCR.

Statistical analysis. Data are presented as means ± s.d. (unless otherwise noted) and were derived from at least three independent experiments. Data on replicates (n) are given in figure legends. Statistical analysis was performed using the two-sided Student’s t test (comparing two groups) or one-way-comparison ANOVA (comparing multiple groups against one group). Variance was similar between the groups that were being compared. The distribution of the raw data approximated a normal distribution (Kolmogorov–Smirnov test, p > 0.05) and sufficient number of replicates to test for normality. No method of randomization was performed, and investigators were not blinded to the genotypes of cell lines. Statistical analysis for WGBS, ERRBS and RNA-seq data are described in the corresponding sections. Additional methodological details are provided in the Supplementary Note.

URLs. R language, http://www.r-project.org/; DAVID Bioinformatics Resources 6.8, https://david.ncifcrf.gov/; Cell Fixation protocol from Active Motif, http://www.activemotif.com/documents/1848.pdf; ENCODE: Encyclopedia of DNA Elements, https://www.encodeproject.org/.

Life sciences reporting summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. All sequencing datasets are available from the Gene Expression Omnibus (GEO) under accession GSE89728.

References
53. Zhu, Z., González, F. & Huangfu, D. The iCRISPR platform for rapid genome editing in human pluripotent stem cells. Methods Enzymol. 546, 215–250 (2014).
54. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
55. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
58. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
59. Akalin, A. et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.* **8**, e1002781 (2012).
60. Pan, H. et al. Epigenomic evolution in diffuse large B-cell lymphomas. *Nat. Commun.* **6**, 6921 (2015).
61. Giannopoulou, E. G. & Elemento, O. An integrated ChIP-seq analysis platform with customizable workflows. *BMC Bioinformatics* **12**, 277 (2011).
62. Williams, K. et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
63. Ehrich, M. et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **102**, 15785–15790 (2005).
**Experimental design**

1. **Sample size**
   - Describe how sample size was determined.
   - The molecular pathways in fertilized eggs from normal and obese mice were investigated in this study, so no sample size calculation was performed.

2. **Data exclusions**
   - Describe any data exclusions.
   - No data were excluded from the analyses.

3. **Replication**
   - Describe whether the experimental findings were reliably reproduced.
   - All attempts at replication are successful.

4. **Randomization**
   - Describe how samples/organisms/participants were allocated into experimental groups.
   - In this study, mice were randomly divided into two diet groups, one group receiving a HFD and the other group received a ND, and then their zygotes were examined.

5. **Blinding**
   - Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   - No blinding was done in most experiments (except DNA damage analysis) in this study. Fertilized eggs were examined, not a typical animal study.

6. **Statistical parameters**
   - For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

   | n/a | Confirmed |
   |-----|-----------|
   | ☑   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
   | ☑   | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly. |
   | ☑   | A statement indicating how many times each experiment was replicated |
   | ☑   | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
   | ☑   | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
   | ☑   | The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted |
   | ☑   | A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
   | ☑   | Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Statistical analysis was performed using GraphPad Prism (Version 5.0) for Windows;
DAVID (Version 6.8) was used for analysis of GO enrichment;
IGV (Version 3.3.94) was used for the analysis of methylation level;
Metilene (Version 0.2-7) was used for defining differential methylated regions;
Trim Galore (Version 0.4.2) was used for quality control of sequencing data;
ImageJ (Version 1.46r) was used for fluorescence quantification.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used in this study.
9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

**Primary antibodies:**

Rabbit anti-Myc antibody (1:2000; Abcam, ab9106); Citation: Webster CP et al. The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J. 35:1656-76 (2016).

Mouse anti-γH2AX antibody (1:300; Abcam, ab22551); Citation: Ladstätter S & Tachibana-Konwalski K. A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming. Cell. 167:1774-1787.e13 (2016).

Rabbit anti-Stella antibody (1:1000; Abcam, ab19878); Citation: White YA et al. Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. Nat Med. 18:413-21 (2012).

Rat anti-Tet3 antibody (IF 1:500, WB 1:1000; Active Motif, 61743); Citation: Bauer C et al. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosaminetransferase (OGT). J Biol Chem. 290:4801-12.(2015)

Anti-5mC/5hmC antibody (1:500; Active Motif, CA, Cat#: 39769; 1:500; Calbiochem, Cat#: NA81); Citation: Kurotaki YK et al. Impaired active DNA demethylation in zygotes generated by round spermatid injection.Hum Reprod. 30:1178-87(2015)

Mouse anti-β-Actin antibody (1:2000; Sigma, A5441); Citation: Ji Zhou et al. Activation of the unfolded protein response occurs at all stages of atherosclerotic lesion development in apolipoprotein E-deficient mice. Circulation. 111(111) (2005)

**Secondary antibodies:**

Alexa Fluor 488 Goat anti-Rabbit IgG (H+L) Secondary Antibody (1:150; Molecular Probes, A-11034); Citation: Bai Het al. Yes-associated protein impacts adherens junction assembly through regulating actin cytoskeleton organization.Am J Physiol Gastrointest Liver Physiol. 311:G396-411(2016)

Alexa Fluor 555 Goat anti-mouseIgG (H+L) Secondary Antibody (1:150; Molecular Probes, A-21422); Citation: Ishiuchi Tet al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat Struct Mol Biol. 22:662-71. (2015)

Alexa Fluor 555 Goat anti-ratIgG (H+L) Secondary Antibody (1:150; Molecular Probes, A-21434); Citation: Zhang L et al. Up-regulation of the active form of small GTPase Rab13 promotes macroautophagy in vascular endothelial cells. Biochim Biophys Acta. 1864:613-624.(2017)

HRP Goat anti-rabbit IgG (H+L) Secondary Antibody (1:5000; Thermo fisher Scientific, 31460); Citation: Xie N et al. PRKAA/AMPK restricts HBV replication through promotion of autophagic degradation. Autophagy. 12:1507-20. (2016)

HRP Goat anti-rat IgG (H+L) Secondary Antibody (1:5000; Thermo fisher Scientific, 31470); Citation: Moutzis SS et al. Gene expression profile associated with oncogenic ras-induced senescence, cell death, and transforming properties in human cells. Cancer Invest. 28:563-87. (2010)
10. Eukaryotic cell lines
   a. State the source of each eukaryotic cell line used.
      No eukaryotic cell lines were used in this study.
   b. Describe the method of cell line authentication used.
      No eukaryotic cell lines were used in this study.
   c. Report whether the cell lines were tested for mycoplasma contamination.
      No eukaryotic cell lines were used in this study.
   d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.
      No eukaryotic cell lines were used in this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.
   Female ICR mice three-week of age were purchased from Charles River Laboratories China Inc.
   Four-week-old female C57BL/6J and 10-20 weeks male ICR mice were purchased from Charles RiverLaboratories China Inc.
   C57BL/6J-lep(ob/ob) mice were purchased from The Jackson Laboratories, Bar Harbor, US.

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.
   This study did not involve human research participants.