Article

DNA methylation shapes the Polycomb landscape during the exit from naive pluripotency

Check for updates

A list of authors and their affiliations appears at the end of the paper

In mammals, 5-methylcytosine (5mC) and Polycomb repressive complex 2 (PRC2)-deposited histone 3 lysine 27 trimethylation (H3K27me3) are generally mutually exclusive at CpG-rich regions. As mouse embryonic stem cells exit the naive pluripotent state, there is massive gain of 5mC concomitantly with restriction of broad H3K27me3 to 5mC-free, CpG-rich regions. To formally assess how 5mC shapes the H3K27me3 landscape, we profiled the epigenome of naive and differentiated cells in the presence and absence of the DNA methylation machinery. Surprisingly, we found that 5mC accumulation is not required to restrict most H3K27me3 domains. Instead, this 5mC-independent H3K27me3 restriction is mediated by aberrant expression of the PRC2 antagonist *Ezhip* (encoding EZH inhibitory protein). At the subset of regions where 5mC appears to genuinely supplant H3K27me3, we identified 163 candidate genes that appeared to require 5mC deposition and/or H3K27me3 depletion for their activation in differentiated cells. Using site-directed epigenome editing to directly modulate 5mC levels, we demonstrated that 5mC deposition is sufficient to antagonize H3K27me3 deposition and confer gene activation at individual candidates. Altogether, we systematically measured the antagonistic interplay between 5mC and H3K27me3 in a system that recapitulates early embryonic dynamics. Our results suggest that H3K27me3 restraint depends on 5mC, both directly and indirectly. Our study also implies a noncanonical role of 5mC in gene activation, which may be important not only for normal development but also for cancer progression, as oncogenic cells frequently exhibit dynamic replacement of 5mC for H3K27me3 and vice versa.

5-Methylcytosine (5mC) represents a highly conserved chemical modification that occurs mainly in the CpG dinucleotide context in vertebrates. 5mC is part of a hierarchical system that impacts gene regulation and can impart epigenetic memory of transcriptional states across cell division. For example, methylated CpG-rich gene promoters are associated with stable transcriptional silencing[1,2]. In this regard, 5mC exhibits properties similar to those of histone 3 lysine 27 trimethylation (H3K27me3), which is deposited by Polycomb repressive complex

2 (PRC2) and is also associated with stable silencing when enriched at CpG-rich regulatory regions[3–6]. Given the predominance of 5mC occurring in the CpG dinucleotide context, we use '5mC' as a signifier of CpG methylation throughout.

In mammals, the fundamental importance of 5mC and H3K27me3 is highlighted by embryonic lethality in mouse knockout (KO) studies[7–11] and grossly perturbed 5mC and/or H3K27me3 landscapes in a number of cancers[12–14]. Notably, 5mC and H3K27me3 are largely

e-mail: maxim.greenberg@ijm.fr

mutually exclusive at CpG-rich regions (CpG islands, CGIs)—regulatory sequences that frequently overlap gene promoters[15–17]. While the vast majority of CGI promoters are DNA hypomethylated[18], most of the mammalian genome contains diminished CpG density and 5mC is the default state in this context. In regions with reduced CpG content, 5mC appears less anticorrelated with H3K27me3 (refs. [16,17]). Thus, the mutual antagonism of 5mC and H3K27me3 appears contingent on a threshold of local CpG content.

H3K27me3–5mC reciprocal dynamics is a hallmark of mouse embryogenesis; genome-wide deposition of 5mC during postimplantation development is coincident with restriction of H3K27me3 to hypomethylated CGIs by the epiblast stage[19–21]. The chromatin established in the epiblast arises just before lineage specification; however, testing the relevance of H3K27me3 restriction at this stage is complicated by the lethality of embryos lacking 5mC.

Pluripotent mouse embryonic stem cells (mES cells) are an attractive model to dissect the interaction between the two forms of regulation. Past studies have largely relied on metastable mES cells cultured in FBS, which exhibit ~70% global CpG methylation[16,22–24] and are resilient to genetic ablation of both DNA methylation[25] and PRC2 machinery[17,26]. However, metastable mES cells represent a heterogenous population of cells consisting of the spectrum of pluripotency states, including 'naive', 'primed' and the intermediate 'formative' state[27,28]. We, therefore, sought to measure the interplay between 5mC and H3K27me3 in a differentiation system that recapitulates the exit from naive pluripotency while mitigating cellular heterogeneity in a time-resolved manner[29–31]. To do so, we used a differentiation system of naive ES cells to epiblast-like cells (EpiLCs) in which cells pass from a DNA hypomethylated to hypermethylated state within 4 days and exhibit a restricted H3K27me3 landscape within 7 days[32,33]. Importantly, we previously demonstrated that cells lacking 5mC (that is, triple KO (TKO) of DNA methyltransferases (DNMTs)) successfully exit the naive state and exhibit hallmarks of primed pluripotent cells, despite substantial transcriptional misregulation[34]. Thus, we used this model to formally assess the effect of genome-wide 5mC deposition on H3K27me3 patterning and the transcriptional impact on underlying genes.

Unexpectedly, we observed extensive H3K27me3 restriction in TKO EpiLCs despite the complete lack of global 5mC deposition. We showed that, in the absence of 5mC, ectopic expression of *Ezhip* (encoding EZH inhibitory protein), a PRC2 antagonist that is normally silenced in a 5mC-dependent manner, restrains H3K27me3 spreading. A distinct set of loci harboring intermediate CpG density aberrantly maintained H3K27me3 in TKO EpiLCs, suggesting that 5mC normally supplants H3K27me3 at thousands of regions. Interestingly, a subset of H3K27me3-to-5mC switch regions (SWRs) is associated with transcriptional activation in wild-type (WT) EpiLCs. Single-locus modulation of 5mC levels in naive and differentiated cells using epigenome editing demonstrated that 5mC can directly antagonize H3K27me3-mediated repression. Previously, we showed in the case of the *Zdbf2* gene not only that embryonic H3K27me3-to-5mC switching is required for activation of this gene but that failure of this switch results in a lifelong growth defect. Here, we show that this form of embryonic epigenetic regulation is of a broader scope and may potentially impact myriad downstream developmental processes.

Results

H3K27me3-to-5mC dynamics during the exit of the naive state

To investigate the interplay between 5mC and PRC2, we performed whole-genome bisulfite sequencing (WGBS) and H3K27me3 cleavage under targets and tagmentation (CUT&Tag)[35] in ES cells cultured with double inhibition of extracellular signal-regulated kinase (ERK) and glycogen synthase kinase (GSK) 3β supplemented with vitamin C (2i + vitC)[36] and EpiLCs[29]—a cellular differentiation system that recapitulates the dramatic transition from a globally cytosine hypomethylated

to hypermethylated genome that occurs during in vivo embryonic development (Fig. 1a and Extended Data Fig. 1a,b). Chromatin profiling was complemented with RNA sequencing (RNAseq) to correlate DNA and histone methylation dynamics with changes in gene expression. We performed a protracted EpiLC differentiation protocol (7 days) to ensure complete de novo deposition of DNA methylation and H3K27me3 turnover.

During differentiation, 5mC and H3K27me3 levels over CGIs were largely static (Extended Data Fig. 1c). That is to say, the CGI promoters of PRC2 target genes remained hypomethylated and maintained high levels of H3K27me3 in EpiLCs in association with transcriptional silencing (Extended Data Fig. 1c–e). Only a minority of CGI promoters exhibited dynamic H3K27me3 patterns that generally anticorrelated with the transcriptional state. For example, genes active in EpiLCs showed depletion of H3K27me3 over the course of differentiation (Extended Data Fig. 1c,f,g). Conversely, naive pluripotency marker genes gain H3K27me3 in both EpiLCs and the epiblast (Extended Data Fig. 1h,i). Lastly, a small subset of CGIs, including the promoters of germline-expressed genes that are known to be regulated by DNA methylation, gain 5mC (>75%) at the expense of H3K27me3 in EpiLCs (Supplementary Fig. 1a). This switch from H3K27me3 to dense 5mC levels was associated with transcriptional silencing in both cell types, with the notable feature that at least some germline genes exhibited modest and transient transcriptional activity during the H3K27me3–5mC transition both in vitro and in vivo (Supplementary Fig. 1b). Consistent with our differentiation model, 66% of CGI promoters that significantly (linear modeling using Limma, fold change > 2, *t*-test adjusted *P* < 0.05) lose H3K27me3 coincident with the gain of 5mC in WT EpiLCs also gain 5mC in the developing epiblast (Supplementary Fig. 1c). Nevertheless, CGIs are generally devoid of DNA methylation (<10% 5mC) in our EpiLC differentiation system (83% of CGIs) and during normal development (86% of CGIs) (Supplementary Fig. 1d).

Unlike CGIs, the remainder of the genome showed widespread H3K27me3 loss during EpiLC differentiation (Extended Data Fig. 1a). Notably, 93% of regions that lost H3K27me3 gained >75% 5mC levels in EpiLCs (Supplementary Fig. 2a), suggesting that 5mC may broadly antagonize H3K27me3 deposition. To test whether 5mC drives H3K27me3 loss during EpiLC differentiation, we performed parallel experiments in TKO cells that are completely devoid of 5mC (Fig. 1b)[25,34,37]. Analysis of genome-wide H3K27me3 levels showed a strong correlation between globally hypomethylated WT and TKO ES cells (Supplementary Fig. 2b). Notably, relative to WT EpiLCs, TKO EpiLCs exhibited a clearly elevated H3K27me3 distribution (Fig. 1c). Consistent with our observations that 5mC restricts H3K27me3 over CGIs, 94% of regions that aberrantly maintain H3K27me3 in TKO EpiLCs were normally hypermethylated (>75% 5mC) in WT EpiLCs (Supplementary Fig. 2c). It should be noted that elevated levels of H3K27me3 in TKO EpiLCs could potentially be confounded by differentiation defects; however, aberrantly H3K27me3-marked regions in TKO EpiLCs included less than half of those that have H3K27me3 in ES cells, as well as thousands of additional regions, indicating that TKO EpiLCs do not simply exhibit a naive ES cell profile (Supplementary Fig. 2d). Further supporting this notion, analysis of marker gene expression indicated that differentiation occurred with similar dynamics even in the absence of 5mC, which is consistent with previous reports (Extended Data Fig. 2a)[34]. Moreover, principal component analysis (PCA) of WT and TKO RNAseq data with in vivo data showed clustering of ES cells with blastocyst-stage embryos and EpiLCs with epiblast-stage embryos, regardless of genetic background (Extended Data Fig. 2b). These results suggest that 5mC restricts H3K27me3 levels across the genome during normal EpiLC differentiation and that, in the absence of 5mC, H3K27me3 can remain relatively broadly distributed.

To quantify H3K27me3-to-5mC dynamics, we performed agnostic *k*-means clustering of H3K27me3 domains (Fig. 1d,e). Three primary
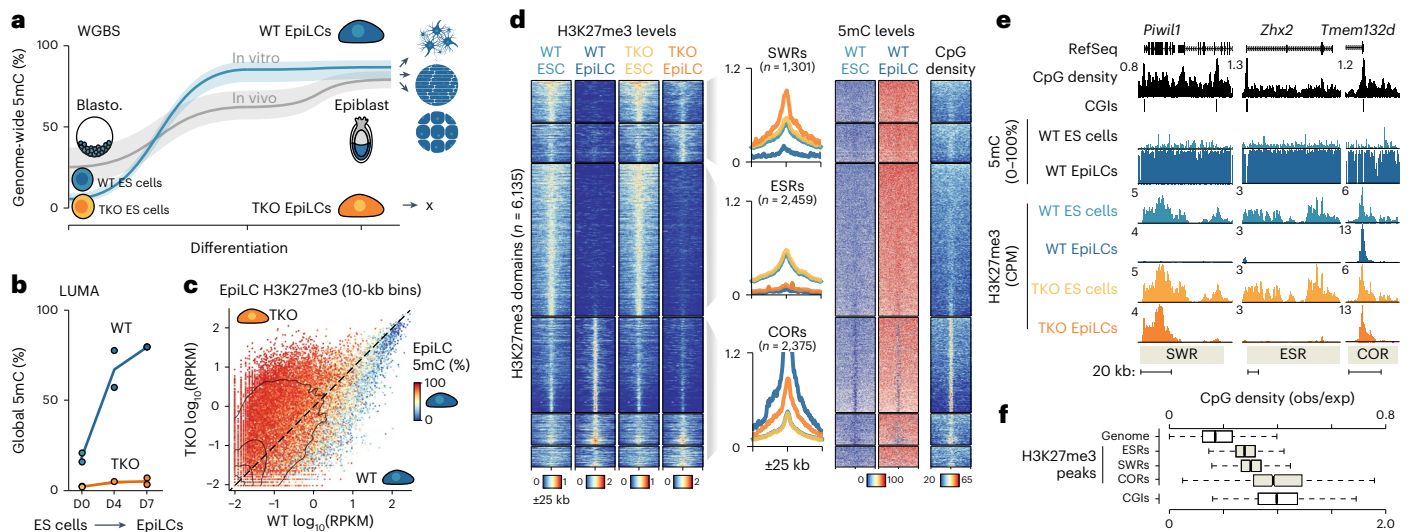
Fig. 1 | H3K27me3 persists in a large fraction of the genome in absence of 5mC. a, WGBS data showing the median (line) and 25th and 75th quartiles (shaded area) of 5mC levels over 10-kb bins ($n = 273,121$) during EpiLC differentiation. In vitro (blue): ES cells grown in 2i + vitC (left) and EpiLCs for 4 days (middle) and 7 days (right) EpiLCs. In vivo (gray): ICM of the blastocyst on embryonic day 3.5 (E3.5) (left) and of the epiblast on E6.5 (middle) and E7.5 (right). In vivo data from Wang et al.[19]. **b**, LUMA showing global 5mC levels in WT and TKO ES cells (day 0) and EpiLCs (days 4 and 7). **c**, Scatter plot showing global H3K27me3 levels over 10-kb bins in WT and TKO EpiLCs (day 7). Data points are colored by average 5mC levels in WT EpiLCs. Here, 10-kb bins with ≥10 CpGs covered by ≥5 reads are shown ($n = 252,559$); contours are drawn at isoproportions of the density of data points. **d**, Heat map of *k*-means clustered H3K27me3 domains, showing H3K27me3, 5mC and

CpG density levels 25 kb upstream and downstream of the domain centers ($n = 6,135$). Three groups of H3K27me3 peaks are indicated: H3K27me3-to-5mC SWRs, ESRs and CORs. Metaplots showing H3K27me3 enrichment over three H3K27me3 peak categories are included. **e**, UCSC Genome Browser screenshot of example SWR, ESR and COR. Data from WT and TKO ES cells and EpiLCs (day 7) are shown. CPM, counts per million aligned reads. Coordinates: chr5:128,725,246–128,800,866 (left), chr15:57,686,919–57,853,110 (middle) and chr5:128,422,318–128,454,166 (right). **f**, Box plot representing the distribution of CpG density levels across the genome, over CGIs and over H3K27me3 domains (SWRs, ESRs and CORs). Note: CGIs are placed on a different scale. The median (line inside box), first and third quartiles (box) and 1.5 times the interquartile range (whiskers) are shown.

groups emerged: (1) constitutive regions (CORs), which largely overlapped with CGIs and showed H3K27me3 enrichment and hypomethylation regardless of cell type; (2) ES cell-specific regions (ESRs), which lost H3K27me3 levels in both WT and TKO EpiLCs; and (3) H3K27me3-to-5mC SWRs in WT EpiLCs, which aberrantly maintained H3K27me3 levels in TKO EpiLCs. In other words, the H3K27me3 landscape is shaped in a 5mC-independent (ESR) and 5mC-dependent (SWR) manner.

CGIs, which are a general feature of CORs (Extended Data Fig. 2c), are known to be protected from DNA methylation through several mechanisms[38]; thus, we were more intrigued to investigate the features that delineate the other two classes. Both ESRs ($n = 2,459$) and SWRs ($n = 1,301$) exhibited similar H3K27me3 levels in ES cells, indicating that H3K27me3 enrichment before differentiation was not a predictor of H3K27me3 or 5mC levels in EpiLCs (Fig. 1d). We next wondered whether specific transcription factor-binding motifs might explain the differential H3K27me3 dynamics. Given that the H3K27me3 domains are quite large, we focused our analysis on 1-kb regions in ESRs and SWRs that overlapped with publicly available Polycomb complex subunit binding data[22]. While SWRs were enriched for G-C and C-G dinucleotides (Extended Data Fig. 2d), no clear transcription factor-binding motifs were found that would indicate specific regulation for either ESRs or SWRs.

PRC2 in mammals has a well-described affinity for CpG-rich regions, such as CGIs[39]. Consistent with this, all three classes of H3K27me3-enriched regions exhibited CpG density higher than the genome average (Fig. 1f). However, ESRs contained the lowest CpG density, with SWRs intermediate and CORs substantially higher. This indicates that, in WT EpiLCs, PRC2 affinity is shifted toward regions with higher CpG content (CORs) and that the gain of DNA methylation at relatively CpG-rich regions (SWRs) appears refractory to PRC2 activity. Nevertheless, an intriguing question arose: what accounted

for the decrease of H3K27me3 deposition not only in WT but also in TKO EpiLCs in a large fraction of the genome (ESRs)?

H3K36me3 is unlinked to H3K27me3 misregulation in TKO EpiLCs

Given the H3K27me3 loss concomitant with 5mC gain during the exit from the naive state, it was somewhat surprising that such a large proportion of regions lost the histone mark in the complete absence of DNA methylation machinery. In fact, ESRs comprised the greatest proportion of H3K27me3-marked regions (Fig. 1d). Concordantly, genome-wide loss of H3K27me3 levels was observed in differentiating TKO cells (Fig. 2a). We hypothesized that perhaps H3K36 methylation, which has been shown to antagonize H3K27me3 in vitro[40,41], may have a role in limiting broad H3K27me3 deposition; indeed, both in vivo and cell culture models of H3K36 misregulation have reported extensive H3K27me3 remodeling[24,42]. Given the link between H3K36me2/3 and 5mC[43–45], we tested whether the H3K36 methylation landscape is altered in TKO EpiLCs. As expected, H3K36me3 CUT&Tag analysis showed a clear anticorrelation between H3K36me3 and H3K27me3 in ES cells and EpiLCs (Extended Data Fig. 3a,b) and the vast majority of H3K36me3-marked sites were DNA methylated in WT EpiLCs (Extended Data Fig. 3b). However, differential analysis in WT versus TKO showed highly correlative H3K36me3 patterns (Extended Data Fig. 3c,d). Indeed, regions that showed increased H3K27me3 levels in TKO EpiLCs relative to WT EpiLCs did not exhibit a concomitant depletion of H3K36me3 (Extended Data Fig. 3e,f). While we did not formally exclude a role for H3K36me2 misregulation in TKO EpiLCs, previous reports demonstrated that H3K36me2 is broadly deposited in mouse pluripotent cells and not perturbed in DNA methylation mutants[24,46]. As such, we proceeded to search for alternative explanations for the observed restriction of H3K27me3 in TKO EpiLCs.
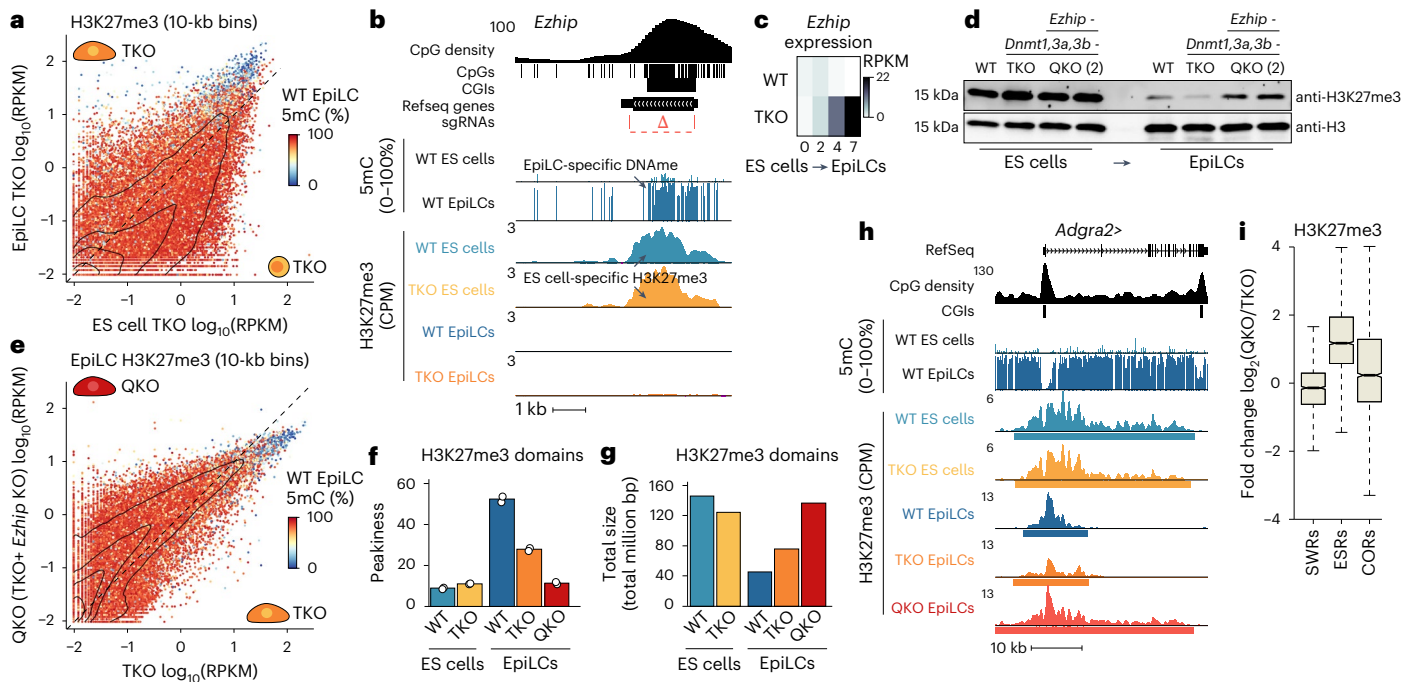
Fig. 2 | 5mC indirectly regulates H3K27me3 spreading through *Ezhip* silencing. a, A two-dimensional (2D) scatter plot showing H3K27me3 enrichment over genome-wide 10-kb bins in differentiating TKO cells. Data points are colored by 5mC levels in WT EpiLCs. Here, 10-kb bins with ≥10 CpGs covered by ≥5 reads are shown (*n* = 252,559) and the density of data points is included. **b**, UCSC Genome Browser screenshot of the *Ezhip* locus displaying H3K27me3 enrichment over the CGI promoter in WT and TKO ES cells and 5mC enrichment in WT EpiLCs (day 4). The location of CRISPR–Cas9 gRNAs and deleted region is shown. Coordinates: chrX:6,078,117–6,084,409. **c**, Heat map showing *Ezhip* expression levels in WT and TKO ES cells (day 0) and EpiLCs (days 2, 4 and 7). The mean of two replicates is shown. **d**, Western blot showing global H3K27me3 levels in WT, TKO and two QKO subclones in ES cells and EpiLCs (day 7). **e**, A 2D

scatter plot showing H3K27me3 enrichment over genome-wide 10-kb bins in TKO and QKO EpiLCs. Data points are colored by 5mC levels in WT EpiLCs as in **a**. **f**, Bar plot of H3K27me3 domain peakiness scores. Peakiness is defined as the average RPKM level of the top 10% most enriched H3K27me3 domains. **g**, Bar plot showing the sum of all H3K27me3 domain sizes. Note the anticorrelation with peakiness scores. **h**, UCSC Genome Browser screenshot of the *Adgra2* gene showing 5mC-independent, *Ezhip*-dependent H3K27me3 restriction during EpiLC differentiation. H3K27me3 peak calls are shown as rectangles. Coordinates: chr8:27,074,539–27,123,049. **i**, Box plot showing the relative change in H3K27me3 level enrichment over H3K27me3 domains between TKO and QKO EpiLCs. The median (line inside box), first and third quartiles (box) and 1.5 times the interquartile range (whiskers) are shown.

Ezhip expression mitigates H3K27me3 spreading in TKO EpiLCs

Recently, a number of papers described EZHIP as a PRC2 antagonist in mouse germline development and in some human cancers[47–51]. *Ezhip* was strongly enriched for H3K27me3 in ES cells (Fig. 2b) and published expression data in ES cells lacking PRC2 components exhibit *Ezhip* upregulation (Extended Data Fig. 4a)[52,53]. In WT EpiLCs, the *Ezhip* promoter exhibited dense 5mC and was transcriptionally repressed (Fig. 2b,c). However, consistent with previous reports[49], unmethylated *Ezhip* was highly transcribed in TKO EpiLCs (>300-fold upregulated compared to WT) (Fig. 2c).

To determine whether *Ezhip* expression restricts H3K27me3 in EpiLCs lacking 5mC, we used clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 to genetically ablate *Ezhip* in TKO ES cells and performed EpiLC differentiation. TKO + *Ezhip* KO (quadruple KO (QKO)) EpiLCs resembled both WT and TKO EpiLCs morphologically and expressed marker genes at comparable levels (Extended Data Fig. 4b,c). Western blot analysis revealed similar global H3K27me3 levels in WT and TKO ES cells and EpiLCs, with a reduction in H3K27me3 signal in the EpiLC state (Fig. 2d). Strikingly, QKO EpiLCs showed a global increase in H3K27me3 levels compared to TKO EpiLCs (Fig. 2d). It should be noted that QKO H3K27me3 levels did not appear as high as in ES cells, which may be because of downregulation of core PRC2 machinery at the protein and RNA levels in EpiLCs (Extended Data Fig. 4d,e)[54]. Nevertheless, genome-wide analysis of H3K27me3 distribution by CUT&Tag in QKO versus TKO EpiLCs uncovered increased enrichment of H3K27me3 in cells lacking *Ezhip*, indicative of spreading (Fig. 2e).

Indeed, H3K27me3 was most restricted in WT EpiLCs, relatively intermediately spread in TKO EpiLCs and broadly distributed in QKO EpiLCs (Fig. 2f,g). The broad H3K27me3 profiles of QKO EpiLCs were most reminiscent of ES cells (Fig. 2f–h), suggesting that, in the absence of 5mC, differentiating cells can nevertheless restrict H3K27me3 through *Ezhip* expression. Nevertheless, QKO EpiLCs showed distinct H3K27me3 patterns compared to ES cells (Supplementary Fig. 3a,b). EZHIP was a more effective PRC2 antagonist at ESRs, which exhibit less CpG density and, thus, potentially less PRC2 affinity than at CORs and SWRs (Fig. 2i and Supplementary Fig. 3b). We conclude from our genetic analyses that H3K27me3 can be directly restricted by global deposition of DNA methylation and by ectopic *Ezhip* expression in the absence of 5mC.

We wondered whether the more naive H3K27me3 patterning in the QKO EpiLCs impacted the transcriptional landscape. Curiously, RNAseq data suggested only minor differences between the TKO and QKO backgrounds (Supplementary Fig. 3c). This raises the question of the biological function of H3K27me3 restriction at ESRs. Hopefully, future studies will interrogate the repercussions of broad H3K27me3 domains in more differentiated cell types, where the effects on gene regulation may be more acute.

5mC accumulates at a subset of upregulated genes

Despite ESRs comprising the largest class of H3K27me3-marked regions, a sizeable fraction (SWRs, ~21%) maintained H3K27me3 in TKO EpiLCs, even in the presence of EZHIP (Fig. 1d). Given the relatively higher affinity for PRC2 activity, we were curious to know whether SWRs spanned regions wherein H3K27me3 exerts a gene-regulatory effect
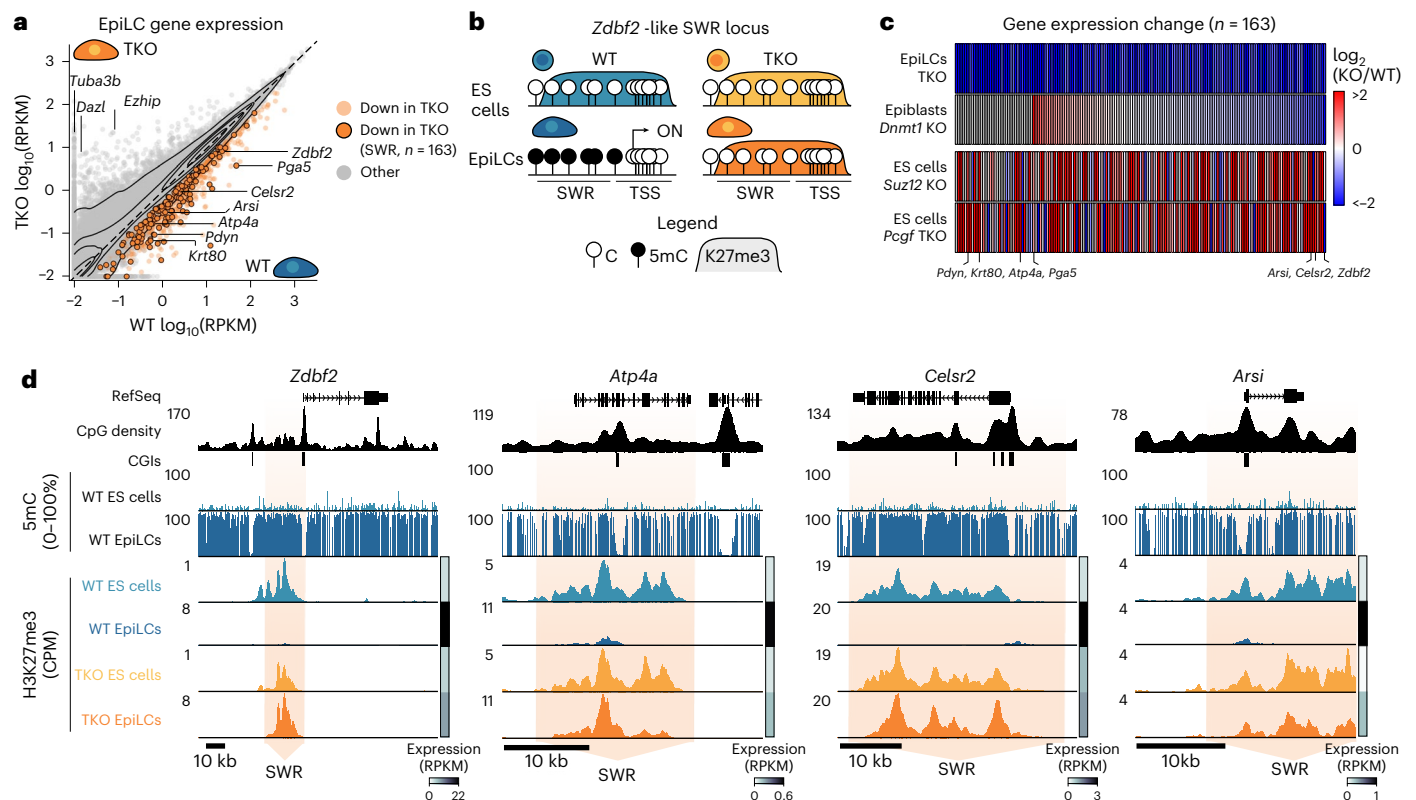
Fig. 3 | 5mC may facilitate the activation of a subset of genes during differentiation. a, A 2D scatter plot showing gene expression levels in WT and TKO EpiLCs (day 7). Significantly downregulated genes in TKO EpiLCs ($n = 1,048$) are colored in orange. Significantly downregulated genes whose TSS overlaps an SWR are indicated ($n = 163$) and the density of data points is included. **b**, Schema showing H3K27me3 and gene expression patterns used to identify SWR regions that may be sensitive to 5mC-mediated gene activation. **c**, Heat map showing relative expression values of candidate 5mC-activated genes near SWRs ($n = 163$). Each column displays the fold change between KO (TKO, *Dnmt1* KO, *Suz12* KO and *Pcgf3*, *Pcgf5* and *Pcgf6* TKO) and control WT cells in $\log_2$ scale. Lowly expressed genes in both conditions (average CPM < 1) are shown in gray. The full list of data mined in this study is provided in Supplementary Table 1. **d**, UCSC Genome Browser screenshots of candidate 5mC-activated genes *Zdbf2*, *Atp4a*, *Celsr2* and *Arsi* adjacent to SWRs. *Zdbf2* chr1:63,222,008–63,338,671; *Atp4a* chr7:30,703,888–30,733,844; *Celsr2* chr3:108,388,342–108,426,060; *Arsi* chr18:60,899,912–60,924,326.

and, by extension, whether DNA methylation deposition impacted PRC2-mediated gene control.

Canonically, 5mC is a mark of transcriptional repression in CpG-rich regulatory regions, such as CGIs. At CGIs, an SWR would presumably not lead to a change in transcriptional activity because H3K27me3 is supplanted by 5mC (inactivated SWR; Extended Data Fig. 5a). For example, the germline gene *Piwil1* CGI promoter normally acquires 5mC during EpiLC differentiation; however, H3K27me3 was aberrantly maintained in TKO EpiLCs (Extended Data Fig. 5b), suggesting that both pathways converge to silence this gene. Indeed, *Piwil1* is repressed in WT and TKO ES cells and EpiLCs (Extended Data Fig. 5c).

While 2,426 genes were upregulated in TKO EpiLCs, intriguingly, a substantial set of 1,048 genes were significantly downregulated in TKO EpiLCs (Fig. 3a). This could either reflect indirect effects of DNA methylation loss or examples where 5mC has a noncanonical activating role (activating SWR). Giving credence to the latter possibility, our previous work demonstrated that the *Zdbf2* gene requires 5mC to displace H3K27me3 (ref. 32). However, to our knowledge, *Zdbf2* is the only gene described as regulated in such a manner during this window of differentiation. To determine whether 5mC deposition contributes to gene activation, we intersected a list of statistically significant SWRs (Methods; $n = 2,679$) with genes significantly downregulated in TKO EpiLCs (Fig. 3b). SWRs overlapped the promoters of 163 downregulated genes in TKO EpiLCs, including *Zdbf2* (Fig. 3a–c). Importantly, 87 were downregulated (fold change < 1) in *Dnmt1* KO epiblasts, including *Zdbf2* and *Celsr2*, which substantiates our in cellula model (Fig. 3c)[55]. H3K27me3 enrichment is likely responsible for repression, as the majority (101/140) of candidate SWR-activated genes expressed in ES cells were upregulated (fold change > 1) upon *Suz12* KO (Fig. 3c)[52]. Additionally, 112 candidate SWR genes were upregulated in ES cells harboring a TKO for PRC1 subunits *Pcgf3*, *Pcgf5* and *Pcgf6*, indicating that both Polycomb complexes cooperate to repress expression of SWR genes in hypomethylated cells (Fig. 3c)[56].

Integrated analysis of 5mC, H3K27me3 and expression levels revealed that broad H3K27me3 upstream and overlapping the *Zdbf2* CGI promoter is normally restricted in WT EpiLCs in association with 5mC deposition and gene activation (Fig. 3d). In contrast, TKO EpiLCs exhibited aberrantly maintained H3K27me3, which extends into the CGI promoter and is linked with maintained transcriptional repression (Fig. 3d). Although H3K27me3 appears depleted at the *Zdbf2* promoter, this is likely because of the low background of the CUT&Tag technique, as we previously observed enrichment using chromatin immunoprecipitation. A similar pattern of 5mC-mediated H3K27me3 restriction and gene activation was observed at the other 162 candidate genes including *Apt4a*, *Celsr2*, *Pdyn*, *Krt80*, *Pga5* and *Arsi* (Fig. 3a–d and Extended Data Fig. 5d). These data suggest that 5mC may have an activating role in gene expression through Polycomb antagonism at a much larger set of genes than was previously realized as cells differentiate from the naive state.

Precision 5mC deposition at *Zdbf2* antagonizes H3K27me3

While our genomic analyses in EpiLCs enabled us to characterize and identify candidate 5mC-activated genes near SWRs, we could not exclude potentially misleading confounding effects. To directly assess

the effect of 5mC deposition on H3K27me3 restriction and nearby gene expression, we established and used two complementary site-directed epigenome-editing approaches.

First, we generated an inducible cell line to selectively target 5mC to candidate SWRs in hypomethylated ES cells (Fig. 4a). Briefly, two doxycycline (Dox)-inducible constructs were stably integrated in the two alleles of the *Rosa26* locus: a catalytically dead *Cas9* fused to ten GCN4 epitopes (dCas9–SunTag)[57] and a DNMT3A catalytic domain (3ACD) fused to green fluorescent protein (GFP) and a single-chain antibody that recognizes the GCN4 epitope (Extended Data Fig. 6a). Both inducible constructs contain a FK506-binding protein 12-derived destabilizing domain, minimizing leaky expression in uninduced cells and enabling robust expression in the presence of Dox and the Shield-1 ligand (Extended Data Fig. 6b)[58,59]. Cells expressing d3ACD were used as a control. Lastly, we stably integrated single guide RNAs (sgRNAs) designed to target the predicted Polycomb nucleation site (Methods) of candidate SWRs (Extended Data Fig. 6a,c).

To assess the efficiency of our epigenome-editing system, we targeted the *Zdbf2* SWR as a proof of principle. We individually tested the extent of 5mC induction using four individual sgRNAs (*Zdbf2* gRNA1–gRNA4) followed by targeted deep methylation sequencing, with gRNA1 showing an 11.8% increase in 5mC levels spanning ~2 kb adjacent to the sgRNA target site (Supplementary Fig. 4). Subsequently, dCas9–SunTag/3ACD induction experiments were performed in hypomethylated ES cells cultured in titrated 2i with the hope that this modified medium composition would help stabilize the maintenance of newly deposited 5mC[36,60,61]. Using sodium bisulfite treatment followed by pyrosequencing (BS-pyro), we obtained robust and highly reproducible 5mC deposition at the *Zdbf2* SWR (22–29%) compared to cells expressing d3ACD or uninduced controls (Fig. 4b and Extended Data Fig. 7a).

While the extent of 5mC induction was modest, we nevertheless assessed the effect on H3K27me3 enrichment using cleavage under targets and release under nuclease (CUT&RUN)[62] followed by qPCR (Fig. 4c and Extended Data Fig. 7b,c). Notably, H3K27me3 levels significantly decreased across a ~16-kb region adjacent to the sgRNA target site in ES cells expressing dCas9–SunTag/3ACD compared to d3ACD. Despite loss of H3K27me3, *Zdbf2* expression levels were not significantly altered in epigenome-edited cells (Fig. 4d). This may be because of incomplete depletion of H3K27me3; it also may be because of persistent isolation from EpiLC-specific enhancers that are known to regulate *Zdbf2* (ref. 63). Nevertheless, we were very encouraged by our ability to demonstrate direct antagonism of 5mC against PRC2 using a targeted approach.

We next selected four uncharacterized SWR regions for DNA methylation editing that looked promising from their genomic profiles: *Celsr2*, *Pdyn*, *Pga5* and *Krt80* (Fig. 3d and Extended Data Fig. 5d). Unfortunately, applying the same strategy did not yield a significant increase in 5mC levels (Extended Data Fig. 7d). In total, using ES cells cultured in serum-free conditions, only one of 20 sgRNAs successfully induced 5mC gain at a target SWR. In a last attempt, we generated a TKO cell line that constitutively expresses *Dnmt1* (Extended Data Fig. 8a). Our rationale was that this cell line would exhibit no de novo DNA methylation beyond epigenome editing of target loci and, in serum-grown conditions, 5mC would be more stably maintained. Indeed, a luminometric methylation assay (LUMA) confirmed no detectable 5mC in this cell line (Extended Data Fig. 8b). We then introduced constitutively expressed dCas9–SunTag/3ACD or d3ACD machinery and scrambled or target sgRNAs (Extended Data Fig. 8c). Encouragingly, the *Zdbf2* and *Celsr2* SWRs showed robust 5mC gain (46–71%) and concomitant modest upregulation (Extended Data Fig. 8d,e). However, global 5mC levels also increased by ~30% presumably because of pervasive 3ACD activity that was maintained by DNMT1 (Extended Data Fig. 8b). This was reflected by an increase in 5mC and a modest increase in expression of our genes of interest in the scrambled gRNA control cell lines (Extended Data Fig. 8d,e). Such off-target effects compromised the direct assessment of the role of 5mC at SWRs on nearby gene expression.
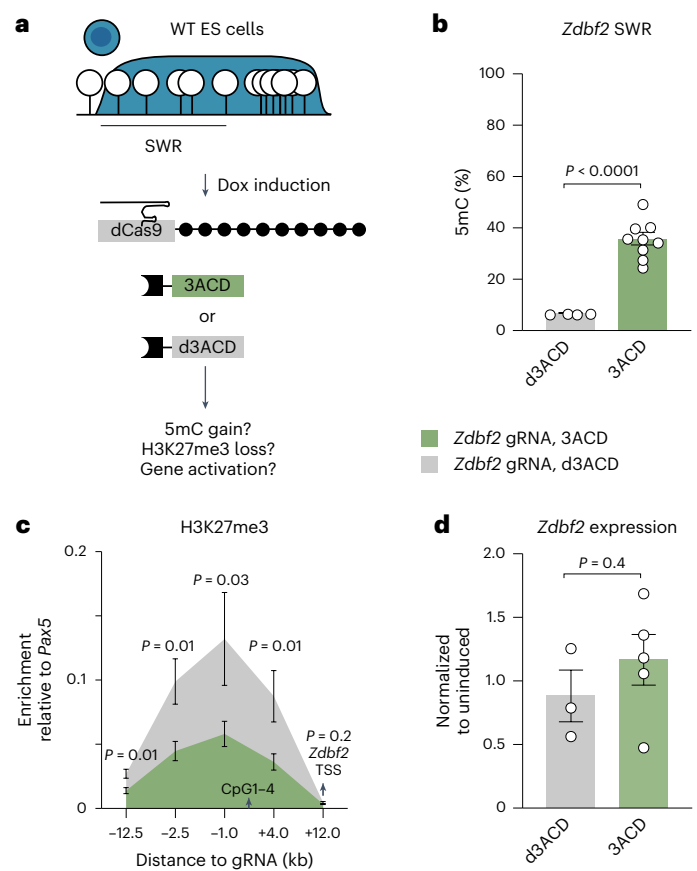
Fig. 4 | Site-directed 5mC deposition in naive ES cells is sufficient to antagonize H3K27me3 at the *Zdbf2* SWR. a, Schema of induced targeted 5mC at candidate hypomethylated (white lollipops), H3K27me3-enriched (blue) SWRs. A construct composed of dCas9 and ten GCN4 peptides (black circles) recruits multiple copies of the mouse 3ACD (green box) through single-chain variable fragment antibodies (black squares). d3ACD (gray box) was used as a control. The construct was induced in naive, globally hypomethylated ES cells and targeted to the *Zdbf2* SWR by sgRNA (gRNA1). 5mC, H3K27me3 and expression levels were assessed 7 days after induction. Not shown: GFP fused to 3ACD and d3ACD. **b,** BS-pyro results at the *Zdbf2* SWR. Each data point represents the average 5mC level over the pyrosequencing amplicon (composed of four CpGs) for each replicate ($n = 4$ for d3ACD and $n = 9$ for 3ACD). **c,** H3K27me3 CUT&RUN–qPCR levels normalized to the positive control *Pax5* CGI promoter (ΔC_t method). Data represent eight and five replicates for 3ACD and d3ACD induced samples, respectively. **d,** Bar plot showing *Zdbf2* expression levels measured by RT–qPCR. Levels were normalized to the average C_t of two housekeeping genes (*Rplp0* and *Rrm2*) and subsequently normalized to uninduced sample expression levels ($\Delta\Delta C_t$ method). Data represent five and three replicates for 3ACD and d3ACD, respectively, denoted by dots. In **b**–**d**, data are shown as the mean ± s.e.m. of replicates from independent experiments. Replicates and their mean are shown. P values were calculated using a two-tailed unpaired t-test assuming equal variance. Source data are provided.

Given these technical challenges, we took a complementary approach to DNA methylation editing.

Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) editing leads to maintenance of H3K27me3 and silencing

For our second approach, our goal was to perform locus-specific 5mC erasure in EpiLCs. We initially used the same inducible system described above, replacing the 3ACD with the TET1 catalytic domain (TET1CD), which progressively oxidizes 5mC leading to 5mC loss, thus counteracting de novo 5mC deposition at target sites (Supplementary Fig. 5a,b)[64]. However, once again, we were met with a technical challenge, as our induction strategy was compromised by transgene
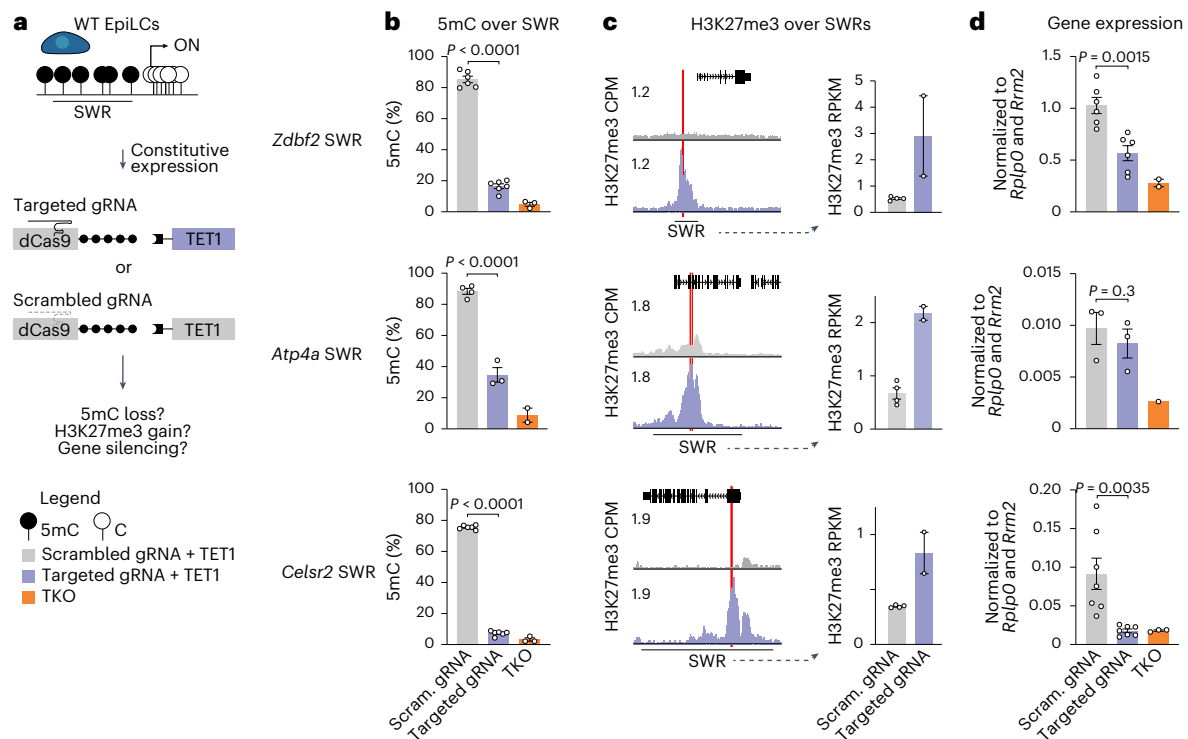
Fig. 5 | Targeted demethylation of candidate SWRs in differentiating ES cells leads to H3K27me3 maintenance and failure to activate *Zdbf2*, *Atp4a* and *Celsr2*. a, Schema of site-directed 5mC erasure at candidate SWRs. Demethylation is carried out by the human TET1CD (purple). Site-specific demethylation is achieved by constitutive expression of dCas9–SunTag/scFv–TET1CD and gRNAs targeted to candidate SWRs. Cells constitutively expressing scrambled gRNAs were used as a control (gray). 5mC, H3K27me3 and expression levels were assessed in EpiLCs (day 4). **b**, BS-pyro showing 5mC levels at the *Zdbf2* (top), *Apt4a* (middle) and *Celsr2* (bottom) candidate SWRs. Purple, targeted gRNAs; gray, scrambled gRNA; orange, TKO. Each data point represents the average 5mC level over the pyrosequencing amplicon (CpGs per amplicon: four, *Zdbf2*; seven, *Atp4a*; five, *Celsr2*) for each independent experiment (*Zdbf2*: scrambled, $n = 6$, targeted gRNA $n = 6$, TKO, $n = 3$; *Atp4a*: scrambled, $n = 4$,

targeted gRNA, $n = 3$, TKO, $n = 2$; *Celsr2*: scrambled, $n = 6$, targeted gRNA, $n = 6$, TKO, $n = 3$. **c**, UCSC genome browser screenshots of candidate SWRs showing H3K27me3 CUT&RUN followed by sequencing in epigenome-edited EpiLCs. gRNA target sites are indicated in red. Right, using the same data, an RPKM enrichment score was calculated over the SWR, represented by a bar chart (scrambled, $n = 4$; targeted, $n = 2$; no statistical tests applied). **d**, Bar plot showing expression levels by RT–qPCR. Levels were normalized to the average C_t of two housekeeping genes (*Rplp0* and *Rrm2*) (ΔC_t method). *Zdbf2*: scrambled, $n = 6$, targeted gRNA, $n = 6$, TKO, $n = 2$; *Atp4a*: scrambled, $n = 3$, targeted gRNA, $n = 3$, TKO, $n = 1$; *Celsr2*: scrambled, $n = 7$, targeted gRNA, $n = 7$, TKO, $n = 3$. In **b**–**d**, data are shown as the mean ± s.e.m. of replicates from independent experiments. Replicates and their mean are shown. *P* values were calculated using a two-tailed unpaired *t*-test assuming equal variance. Source data are provided.

silencing in differentiated cells (Supplementary Fig. 5c). To overcome this limitation, we integrated a constitutively expressed dCas9–SunTag/TET1CD construct using *piggyBac* transposition (Fig. 5a and Supplementary Fig. 5d)[65]. In this manner, robust expression of the construct was observed in the majority of cells after 4 days of differentiation (Supplementary Fig. 5c); GFP-positive cells were then enriched by fluorescence-activated cell sorting (FACS). Cells expressing a scrambled gRNA were used as controls.

Using the optimized approach, we obtained a remarkably robust and highly significant decrease in 5mC levels (average 54%) for all targets in EpiLCs (Fig. 5b and Extended Data Fig. 9a). Next, we set out to determine the impact on H3K27me3 using CUT&RUN followed by sequencing. For *Zdbf2*, *Atp4a* and *Celsr2*, we observed a significant enrichment of H3K27me3 exclusively over the target SWR in epigenome-edited cells compared to scrambled controls (Fig. 5c and Extended Data Fig. 9b–e). The candidate SWR near *Krt80* showed a modest decrease in H3K27me3 levels in edited cells and the candidate SWRs near *Pdyn* and *Pga5* show depletion of H3K27me3 regardless of 5mC editing, suggesting that H3K27me3 depletion is uncoupled from 5mC deposition at the respective loci (Supplementary Fig. 6a). The *Arsi* locus, which showed the lowest change in 5mC levels (22%) of all candidates, did not show differences in H3K27me3 levels in epigenome-edited cells (Supplementary Fig. 6b). Consistent with a lack of an H3K27me3 effect, targeted 5mC demethylation of *Pdyn*, *Pga5*

and *Krt80* did not lead to relative repression at two different time points (Supplementary Fig. 6c and Supplementary Fig. 7).

Notably, *Zdbf2*, *Atp4a* and *Celsr2*, the three genes that showed a clear maintenance of H3K27me3 levels when 5mC was actively removed, were downregulated in epigenome-edited cells (Fig. 5d and Supplementary Fig. 7). While *Atp4a* did not exhibit significant repression, the decrease in *Zdbf2* and *Celsr2* expression was similar to that observed in TKO EpiLCs. *Arsi* expression was also downregulated in epigenome-edited cells, which could be explained by the dCas9 complex interfering with transcriptional machinery, given that the target region was downstream of the transcription start site (TSS) (Extended Data Fig. 6c and Supplementary Fig. 6c)[66].

To distinguish between the genuine effect of 5mC demethylation from any confounding effects of dCas9 machinery bound to chromatin, we performed parallel experiments using dTET1CD. This approach yielded intermediate effects on 5mC decrease and H3K27me3 maintenance, which resulted in relative gene downregulation (Extended Data Fig. 10a–d). Indeed, it was reported that dCas9 targeting alone can induce 5mC loss[67], likely by sterically preventing DNMT action. Nevertheless, the intermediate effects are informative, as dCas9–SunTag/dTET1CD did not lead to as strong a change in H3K27me3 levels as the catalytically active version. This indicates that 5mC, per se, as opposed to dCas9 binding is required to fully antagonize H3K27me3 deposition and PRC2-mediated gene repression.
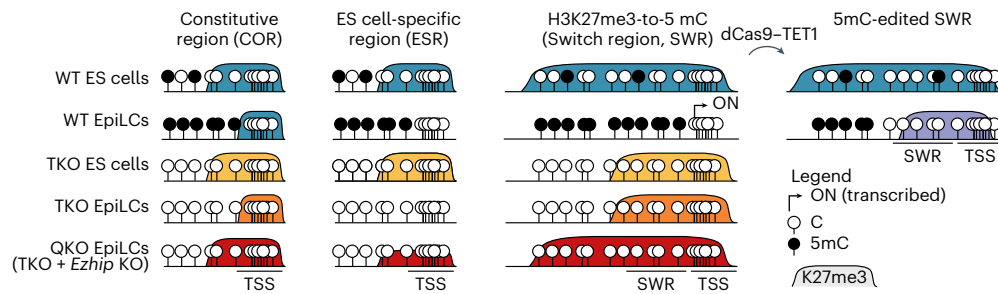
Fig. 6 | Summary of H3K27me3 dynamics during the exit from naive pluripotency. The different categories of H3K27me3 domains as defined in Fig. 1 are illustrated. CORs are constitutively DNA hypomethylated and marked by H3K27me3. ESRs are H3K27me3-enriched in ES cells but lose the histone mark in EpiLCs irrespective of the underlying 5mC state. SWRs lose H3K27me3 in a 5mC-dependent manner. SWRs can be challenged by TET-mediated 5mC editing, resulting in decreased 5mC levels, aberrant maintenance of H3K27me3 and reduced gene expression. QKO EpiLCs show ES cell-like H3K27me3 patterns in all three classes, with the most dramatic effect seen at ESRs.

In sum, in addition to validating our epigenome-editing approach at the proof-of-principle locus, *Zdbf2*, we were able to identify *Atp4a* and *Celsr2* as genes that are activated in a DNA methylation-dependent manner. Importantly, using a targeted approach also allowed us to distinguish bona fide SWRs from loci that exhibited SWR-like behavior in TKO EpiLCs that was likely simply coincidental. Our TET editing strategy emerged as a promising means to demonstrate antagonism between two chromatin pathways typically associated with gene silencing, which in certain cases can lead to a change in gene-regulatory output. It also underscores the power of epigenome editing to discern locus-specific chromatin regulation from confounding effects that frequently occur in constitutive KOs.

Discussion

In this study, we assessed the impact of 5mC deposition on shaping the H3K27me3 landscape in a cell-based model of embryonic progression, leveraging our prior finding that naive mouse ES cells can differentiate to EpiLCs in the absence of DNA methylation machinery (Fig. 6).

Naive ES cells exhibit elevated PRC2 protein abundance and pervasive H3K27me3 distribution throughout the genome, reinforced by PRC2-mediated repression of *Ezhip*. Consistent with this, in vivo postimplantation epiblasts do not express *Ezhip*. However, our differentiation experiments and transcriptional profiling of QKO cells did not suggest an obvious impairment in transitioning from the naive state (aside from a more naive-like H3K27me3 profile), raising questions about the importance of H3K27me3 restriction as the cells prepare for lineage specification. It should be noted that there were perhaps more subtle differentiation or expression defects that we did not detect.

Primordial germ cells (PGCs), a cell type with low levels of DNA methylation, serve as an informative counterpoint to naive ES cells. Mouse PGCs exhibit even lower 5mC levels than inner cell mass cells[19,68] and, while not pluripotent, they share features with ES cells including expression of core pluripotency transcription factors[69]. A key distinction between the cell types is that, unlike in ES cells, *Ezhip* is expressed in PGCs. Moreover, mice mutant for *Ezhip* exhibit increased H3K27me3 in both the male and the female germline, which is associated with decreased fertility[50]. These observations lead us to propose that the CGI promoter of *Ezhip* acts as a global hypomethylation sensor that can help restrain potentially problematic H3K27me3 spreading. Curiously, this adaptation is not exploited in naive ES cells. Does this imply that broad H3K27me3 domains are an important feature of the naive state? If so, this would need to be reconciled with minimal impact on the naive state in the complete absence of PRC2 and H3K27me3 (refs. [26,70,71]), as well as our findings here that EpiLCs appear resilient to broad H3K27me3 deposition.

Even with derepressed *Ezhip* in TKO EpiLCs, we still observed a substantial number of H3K27me3-enriched regions in the genome that appear to be sensitive to deposition of DNA methylation in WT SWRs. A probable important feature of SWRs is their intermediate CpG content. They seemingly occupy middle ground wherein they have exceeded a CpG density threshold that makes them attractive for PRC2 activity but have not exceeded a threshold that would subject them to CGI regulation. With some notable exceptions, promoter CGIs exhibit several layers of protection, such as transcription factor binding[72,73], H3K4me3 (refs. [74–77]) and TET protein enrichment[78,79], that keep them DNA methylation-free. What then could explain the sensitivity of PRC2 to 5mC? A potential interface could be the Polycomb-like (PCL) proteins PHF1, MTF2 and PHF19, which are partially redundant components of the PRC2.1 subcomplex and were shown to be sensitive to DNA methylation[80,81]. However, formal demonstration of PCL protein enrichment or a lack thereof at SWRs in hypomethylated ES cells awaits.

While the primary focus of this study was on PRC2-deposited H3K27me3, as previously mentioned, PRC1 and PRC2 activities are strongly interlinked[39]. The PRC1-associated RING1A and RING1B proteins catalyze H2AK119ub and published data in mouse ES cells indicate that the vast majority of SWRs are enriched for this mark (Supplementary Fig. 8a)[82]. Moreover, after 2 days of EpiLC differentiation, there already appears to be a shift toward H2AK119ub-marked SWRs in conditional *Dnmt* TKO relative to WT cells (Supplementary Fig. 8b,c). Canonically, H3K27me3 deposition serves as a platform for PRC1 binding; however, in mammals, a number of noncanonical PRC1 (ncPRC1) complexes exist, which can bind to chromatin independently of PRC2 activity[83]. Notably, ncPRC1.1 and ncPRC1.6 contain 5mC-sensitive DNA binding factors KDM2B and MAX/MGA, respectively[84,85]. Publicly available ES cell data indicate that PCGF2 (cPRC1.2 component) and PCGF6 (ncPRC1.6 component) exhibit the most overlap with SWRs (Supplementary Fig. 8d)[56]. Most intriguingly, ncPRC1.6 is required for H3K27me3-to-5mC switching at a number of germline-specific gene promoters, although, in these cases, the switch leads to maintained silencing[82,86,87]. It will be worthwhile to pursue a potential link between ncPRC1.6 and SWRs and the possibility that ncPRC1.6 could facilitate gene activation through the same epigenetic switch mechanism that occurs at germline genes.

The development of site-directed epigenome editing has enabled testing of the functional relevance of chromatin marks on local gene expression while bypassing traditional loss-of-function genetics of epigenetic regulators that can have confounding effects[88]. The resistance to targeted 5mC at these loci could be explained by a combination of factors acting against or at least not favoring the maintenance of the mark at the candidate regions in our cell system (for example, impaired DNA methylation maintenance machinery and high TET protein activity)[89]. Moreover, the nature of the sequences we targeted most likely do not contain features that bolster DNA methylation maintenance in the naive state, such as those found at imprint control regions or endogenous retroviruses[90,91]. 5mC deposition and maintenance was improved in cells expressing DNMT1 and grown in serum

conditions, yet this also resulted in a confounding genome-wide 5mC level increase. Lastly, it should be noted that at least one gRNA was able to reliably establish DNA methylation in serum-free conditions at the *Zdbf2* SWR; hence, it is not an impossible task given a suitable guide and target sequence.

Conversely, the dCas9–SunTag/TET1CD system worked robustly and reliably at all candidate SWRs, enabling us to test whether there was a regulatory role for DNA methylation at the underlying genes. Three of the candidates (*Pdyn*, *Pga5* and *Krt80*) did not exhibit relative repression in the absence of DNA methylation and, at four candidates (*Arsi*, *Pdyn*, *Pga5* and *Krt80*), H3K27me3 levels were depleted in WT differentiated cells in the presence or absence of 5mC. Our data underscore the hazard of making conclusions about 5mC-mediated regulation at specific loci when using a constitutive KO.

However, we were encouraged by our demonstration that *Atp4a* and *Celsr2* genuinely require DNA methylation to activate, validating our approach. Although *Atp4a*, *Celsr2* and *Zdbf2* appear to be exceptional cases, it is worth noting that we only analyzed seven of 163 candidate genes using epigenome editing; thus, there are certainly more opportunities to discover similar examples of this mode of gene control. Importantly, we only focused on genes where we could observe differential gene expression in WT and TKO EpiLCs but there could be a much larger set of genes that we omitted because of their activation at later time points in development. This is a limitation of using TKO cells, which do not exhibit the same differentiation potential as WT.

The *Celsr2* gene encodes an atypical cadherin protein that is widely expressed in the developing and adult mouse and human central nervous systems[92–96]. KO studies indicate pleiotropic phenotypes, including hydrocephaly[97], a decrease in cortical plasticity and motor learning in mice[94] and an increase of motor neuron regeneration after injury in mouse and human explants[98]. Consistent with this, publicly available data[99] indicate that *Celsr2* maintains its high 5mC levels in tissues where it is highly expressed (Supplementary Fig. 9a). *Atp4a* encodes an adenosine triphosphatase subunit and KO mice show defects in gastric acid secretion[100]. *Atp4a* also maintains high 5mC in the stomach, where it is almost exclusively expressed (Supplementary Fig. 9b). Neither *Celsr2* nor *Atp4a* has reported embryonic phenotypes, perhaps hinting at a form or epigenetic regulation important for later developmental stages. With emerging in vivo TET-mediated epigenome-editing techniques[65,101], it will be exciting to interrogate the relevance of 5mC-mediated Polycomb antagonism in a physiologically relevant system.

The combined power of genome-wide analyses with single-locus epigenome modulation opens doors to discover and characterize atypical modes of chromatin regulation, not only in the context of development as described here but also in disease. Epigenome misregulation is characteristic of virtually all cancers and SWR regions are prevalent[12,102,103], although it is difficult to discern cause from consequence. With epigenome editing, it is now possible to test whether SWRs are associated with a change in gene expression and possibly disease progression.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-024-01405-4.

References

1. Boyes, J. & Bird, A. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* **11**, 327–333 (1992).
2. Weber, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
3. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
4. Hawkins, R. D. et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
5. Reddington, J. P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).
6. Holoch, D. et al. A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. *Nat. Genet.* **53**, 1686–1697 (2021).
7. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
8. Faust, C., Schumacher, A., Holdener, B. & Magnuson, T. The *eed* mutation disrupts anterior mesoderm production in mice. *Development* **121**, 273–285 (1995).
9. 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).
10. O'Carroll, D. et al. The Polycomb-group gene *Ezh2* is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
11. Grosswendt, S. et al. Epigenetic regulator function through mouse gastrulation. *Nature* **584**, 102–108 (2020).
12. Reddington, J. P., Sproul, D. & Meehan, R. R. DNA methylation reprogramming in cancer: does it act by re-configuring the binding landscape of Polycomb repressive complexes? *Bioessays* **36**, 134–140 (2013).
13. Wassef, M. & Margueron, R. The multiple facets of PRC2 alterations in cancers. *J. Mol. Biol.* **429**, 1978–1993 (2017).
14. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
15. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl Acad. Sci. USA* **104**, 5521–5526 (2007).
16. Brinkman, A. B. et al. Sequential ChIP–bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).
17. Statham, A. L. et al. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* **22**, 1120–1127 (2012).
18. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91–99 (1985).
19. Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **15**, 979–991 (2014).
20. Zheng, H. et al. Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Mol. Cell* **63**, 1066–1079 (2016).
21. Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558–562 (2016).
22. Cooper, S. et al. Targeting Polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* **7**, 1456–1470 (2014).
23. Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).
24. Chen, H. et al. H3K36 dimethylation shapes the epigenetic interaction landscape by directing repressive chromatin modifications in embryonic stem cells. *Genome Res.* **32**, 825–837 (2022).

25. Tsumura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).

26. Montgomery, N. D. et al. The murine Polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr. Biol.* **15**, 942–947 (2005).

27. Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365–373 (2017).

28. Kolodziejczyk, A. A. et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* **17**, 471–485 (2015).

29. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).

30. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).

31. Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).

32. Greenberg, M. V. C. et al. Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat. Genet.* **49**, 110–118 (2017).

33. Greenberg, M., Teissandier, A., Walter, M., Noordermeer, D. & Bourc'his, D. Dynamic enhancer partitioning instructs activation of a growth-related gene during exit from naïve pluripotency. *eLife* **8**, e44057 (2019).

34. Schulz, M. et al. DNA methylation restricts coordinated germline and neural fates in embryonic stem cell differentiation. *Nat. Struct. Mol. Biol.* **31**, 102–114 (2024).

35. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283 (2020).

36. Blaschke, K. et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).

37. Dubois, A. et al. H3K9 tri-methylation at Nanog times differentiation commitment and enables the acquisition of primitive endoderm fate. *Development* **149**, dev201074 (2022).

38. Hughes, A. L., Kelley, J. R. & Klose, R. J. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194567 (2020).

39. Holoch, D. & Margueron, R. Mechanisms regulating PRC2 recruitment and enzymatic activity. *Trends Biochem. Sci.* **42**, 531–542 (2017).

40. Schmitges, F. W. et al. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* **42**, 330–341 (2011).

41. Finogenova, K. et al. Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *eLife* **9**, e61964 (2020).

42. Xu, Q. et al. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* **51**, 844–856 (2019).

43. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).

44. Weinberg, D. N. et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).

45. Yano, S. et al. Histone H3K36me2 and H3K36me3 form a chromatin platform essential for DNMT3A-dependent DNA methylation in mouse oocytes. *Nat. Commun.* **13**, 4440 (2022).

46. Streubel, G. et al. The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in embryonic stem cells. *Mol. Cell* **70**, 371–379 (2018).

47. Pajtler, K. W. et al. Molecular heterogeneity and *CXorf67* alterations in posterior fossa group A (PFA) ependymomas. *Acta Neuropathol.* **136**, 211–226 (2018).

48. Jain, S. U. et al. PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nat. Commun.* **10**, 2146 (2019).

49. Piunti, A. et al. CATACOMB: an endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci. Adv.* **5**, eaax2887 (2019).

50. Ragazzini, R. et al. EZHIP constrains Polycomb repressive complex 2 activity in germ cells. *Nat. Commun.* **10**, 3858 (2019).

51. Jessa, S. et al. K27M in canonical and noncanonical H3 variants occurs in distinct oligodendroglial cell lineages in brain midline gliomas. *Nat. Genet.* **54**, 1865–1880 (2022).

52. Højfeldt, J. W. et al. Non-core subunits of the PRC2 complex are collectively required for its target-site specificity. *Mol. Cell* **76**, 423–436 (2019).

53. van Mierlo, G. et al. Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-state pluripotency. *Cell Stem Cell* **24**, 123–137 (2019).

54. Ugur, E. et al. Comprehensive chromatin proteomics resolves functional phases of pluripotency and identifies changes in regulatory components. *Nucleic Acids Res.* **51**, 2671–2690 (2023).

55. Dahlet, T. et al. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat. Commun.* **11**, 1–14 (2020).

56. Scelfo, A. et al. Functional landscape of PCGF proteins reveals both RING1A/B-dependent- and RING1A/B-independent-specific activities. *Mol. Cell* **74**, 1037–1052 (2019).

57. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).

58. Banaszynski, L. A., Chen, L.-c, Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995–1004 (2006).

59. Senturk, S. et al. Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat. Commun.* **8**, 14370 (2017).

60. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**, e11418 (2016).

61. Carlini, V., Policarpi, C. & Hackett, J. A. Epigenetic inheritance is gated by naïve pluripotency and *Dppa2*. *EMBO J.* **41**, e108677 (2022).

62. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).

63. Monteagudo-Sánchez, A., Noordermeer, D. & Greenberg, M. V. C. The impact of DNA methylation on CTCF-mediated 3D genome organization. *Nat. Struct. Mol. Biol.* **31**, 404–412 (2024).

64. Morita, S. et al. Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**, 1060–1065 (2016).

65. Horii, T., Morita, S., Kimura, M. & Hatada, I. Efficient generation of epigenetic disease model mice by epigenome editing using the *piggyBac* transposon system. *Epigenetics Chromatin* **15**, 40 (2022).

66. Zukher, I., Dujardin, G., Sousa-Luís, R. & Proudfoot, N. J. Elongation roadblocks mediated by dCas9 across human genes modulate transcription and nascent RNA processing. *Nat. Struct. Mol. Biol.* **30**, 1536–1548 (2023).

67. Sapozhnikov, D. M. & Szyf, M. Unraveling the functional role of DNA demethylation at specific promoters by targeted steric blockage of DNA methyltransferase with CRISPR/dCas9. *Nat. Commun.* **12**, 5711 (2021).

68. Seisenberger, S. et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).

69. Hajkova, P. et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).

70. Shen, X. et al. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).

71. Shan, Y. et al. PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. *Nat. Commun.* **8**, 672 (2017).

72. Lienert, F. et al. Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* **43**, 1091–1097 (2011).

73. Krebs, A. R., Dessus-Babus, S., Burger, L. & Schübeler, D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* **3**, e04094 (2014).

74. Ooi, S. K. T. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).

75. Otani, J. et al. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep.* **10**, 1235–1241 (2009).

76. Zhang, Y. et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* **38**, 4246–4253 (2010).

77. Guo, X. et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640–644 (2015).

78. Manzo, M. et al. Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J.* **36**, 3421–3434 (2017).

79. Gu, T. et al. DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol.* **19**, 88 (2018).

80. Li, H. et al. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287–291 (2017).

81. Perino, M. et al. MTF2 recruits Polycomb repressive complex 2 by helical-shape-selective DNA binding. *Nat. Genet.* **50**, 1002–1010 (2018).

82. Mochizuki, K. et al. Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. *Nat. Commun.* **12**, 7020 (2021).

83. Blackledge, N. P. & Klose, R. J.The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).

84. Farcas, A. M. et al. KDM2B links the Polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**, 6233–6226 (2012).

85. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).

86. Dahlet, T. et al. E2F6 initiates stable epigenetic silencing of germline genes during embryonic development. *Nat. Commun.* **12**, 3582 (2021).

87. Al Adhami, H. et al. Systematic identification of factors involved in the silencing of germline genes in mouse embryonic stem cells. *Nucleic Acids Res.* **51**, 3130–3149 (2023).

88. Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V. & Hackett, J. A. Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. *Nat. Genet.* **56**, 1168–1180 (2024).

89. von Meyenn, F. et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* **62**, 848–861 (2016).

90. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).

91. Kojima, S. et al. Epigenome editing reveals core DNA methylation for imprinting control in the Dlk1–Dio3 imprinted domain. *Nucleic Acids Res.* **50**, 5080–5094 (2022).

92. Shima, Y. et al. Differential expression of the seven-pass transmembrane cadherin genes *Celsr1–3* and distribution of the Celsr2 protein during mouse development. *Dev. Dyn.* **223**, 321–332 (2002).

93. Qu, Y. et al. Atypical cadherins Celsr1–3 differentially regulate migration of facial branchiomotor neurons in mice. *J. Neurosci.* **30**, 9392–9401 (2010).

94. Li, C. et al. Planar cell polarity protein Celsr2 maintains structural and functional integrity of adult cortical synapses. *Prog. Neurobiol.* **219**, 102352 (2022).

95. GTEx, C. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

96. Lindsay, S. J. et al. HDBR expression: a unique resource for global and individual gene expression studies during early human brain development. *Front. Neuroanat.* **10**, 86 (2016).

97. Tissir, F. et al. Lack of cadherins Celsr2 and Celsr3 impairs ependymal ciliogenesis, leading to fatal hydrocephalus. *Nat. Neurosci.* **13**, 700–707 (2010).

98. Wen, Q. et al. Inactivating Celsr2 promotes motor axon fasciculation and regeneration in mouse and human. *Brain* **145**, 670–683 (2022).

99. He, Y. et al. Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature* **583**, 752–759 (2020).

100. Spicer, Z. et al. Stomachs of mice lacking the gastric H,K-ATPase α-subunit have achlorhydria, abnormal parietal cells, and ciliated metaplasia. *J. Biol. Chem.* **275**, 21555–21565 (2000).

101. Horii, T. et al. Successful generation of epigenetic disease model mice by targeted demethylation of the epigenome. *Genome Biol.* **21**, 77 (2020).

102. Reddington, J. P., Pennings, S. & Meehan, R. R. Non-canonical functions of the DNA methylome in gene regulation. *Biochem. J.* **451**, 13–23 (2013).

103. Guo, H. et al. DNA hypomethylation silences anti-tumor immune genes in early prostate cancer and CTCs. *Cell* **186**, 2765–2782 (2023).

Julien Richard Albert ⓘ [1,7], **Teresa Urli** ⓘ [1,7], **Ana Monteagudo-Sánchez** ⓘ [1,2,7], **Anna Le Breton** ⓘ [1,3], **Amina Sultanova**[1,4], **Angélique David** ⓘ [1], **Margherita Scarpa**[1], **Mathieu Schulz** ⓘ [5,6] **& Maxim V. C. Greenberg** ⓘ [1] ✉

[1]Université Paris Cité, CNRS, Institut Jacques Monod, Paris, France. [2]Carlos Simon Foundation, INCLIVA Health Research Institute, Valencia, Spain. [3]Gulbenkian Institute for Molecular Medicine, Lisbon, Portugal. [4]Development and Disease Research Group, Max Planck Institute for Molecular Genetics, Berlin, Germany. [5]Institut Curie, PSL Research University, INSERM U934, CNRS, UMR3215, Paris, France. [6]Department of Pathology and Cell Biology, Faculty of Medicine, University of Montreal Hospital Research Centre, Montréal, Québec, Canada. [7]These authors contributed equally: Julien Richard Albert, Teresa Urli, Ana Monteagudo-Sánchez. ✉e-mail: maxim.greenberg@ijm.fr

Methods

Mouse ES cell culture

E14Tg2a (E14) mouse ES cells represented the parental line used for all experiments in this study, in addition to serving as the background for all transgenic lines. The E14 *Dnmt* TKO line was previously generated in-house[37].

For the cells grown in serum culture conditions, we used Glasgow medium (Gibco) supplemented with 15% FBS, 0.1 mM MEM nonessential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine, penicillin–streptomycin, 0.1 mM β-mercaptoethanol and 1,000 U per ml leukemia inhibitory factor (LIF). To pass, the cells were washed with 1× PBS; then, trypsin was added to detach and disaggregate the cells for 5 min at 37 °C. The desired number of cells were then transferred to the new flask.

For the 2i + vitC culture conditions (naive ES cells), we used N2B27 medium (50% neurobasal medium and 50% DMEM) supplemented with N2 (Gibco), B27 (Gibco), 2 mM L-glutamine, 0.1 mM β-mercaptoethanol, penicillin–streptomycin, LIF and 2i (3 μM GSK3β inhibitor CT-99021 and 1 μM ERK inhibitor PD0325901) and vitamin C (Sigma) at a final concentration of 100 μg ml^{-1}. Titrated 2i (t2i) medium was prepared as 2i medium with two modifications: the ERK inhibitor PD0325901 was used at 0.2 μM and no vitamin C was supplemented. To pass the cells, the medium was removed; then, Accutase (Gibco) was added to detach and disaggregate the cells before incubating for 5 min at room temperature. The desired number of cells were then transferred to the new plate. The ES cells in all three conditions were grown on 0.1% gelatin-coated flask in an incubator at 37 °C and 5% CO_2.

To induce EpiLC differentiation, cells were gently washed with PBS, dissociated and replated at a density of 2×10^5 cells per cm^2 on fibronectin (10 μg ml^{-1}, Sigma)-coated plates in N2B27 medium supplemented with 20 ng ml^{-1} activin A (R&D) and 12 ng ml^{-1} FGF2 (R&D). EpiLCs were passed with Accutase on day 2 or 3 of differentiation depending on the end time point of differentiation (day 4 or 7).

Mycoplasma contamination checks were performed using an ultrasensitive qPCR assay (Eurofins).

Plasmid generation

The Dox-inducible dCas9–SunTag is a modified version of the pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP vector (Addgene, 60903). An N-terminal DD domain and the construct minus the P2A-BFP were PCR-amplified and Gibson-cloned into a modified version of the pEN111 vector that has ~700-bp homology regions to the *Rosa26* locus (a gift from E. Nora) containing a neomycin resistance gene.

The Dox-inducible scFv-GFP-3ACD, d3ACD or TET1CD was generated as follows. pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (Addgene, 60904) was digested with RsrII and SpeI restriction enzymes to remove the VP64 sequence, as previously performed[104]. Then, 3ACD, d3ACD containing a cysteine-to-alanine substitution in the catalytic loop or TET1CD was inserted by Gibson cloning. Subsequently, an N-terminal DD domain and the construct were PCR-amplified and inserted into the original pEN111 vector containing a puromycin resistance gene by Gibson cloning.

The plasmids for constitutive expression of dCas9–SunTag/dTET1, d3ACD and 3ACD were generated from the *piggyBac*-compatible plasmid containing a constitutive dCas9–SunTag/TET1CD construct (a gift from I. Hatada[65]) as follows. The original plasmid was digested using AgeI and BbsI and the human dTET1 sequence containing H1672Y and D1674A substitutions (gBlock, Integrated DNA Technologies) was inserted by Gibson cloning. Alternatively, the plasmid was digested with AgeI and NotI restriction enzymes to remove the TET1 cassette; then, d3ACD or 3ACD, PCR-amplified from the Dox-inducible plasmids containing these constructs, was inserted by Gibson cloning.

All CRISPR–Cas9 guide sequences were cloned into a *piggyBac* donor plasmid containing a U6 promoter, enhanced gRNA scaffold sequence and downstream hygromycin resistance gene[105]. Single guides were cloned by BbsI digest and ligation of double-stranded guide sequences with compatible overhangs. Dual guides were cloned by amplifying a PCR product from the pLKO.1-blast-U6-sgRNA-BfuA1-stuffer plasmid[6] containing one gRNA sequence, the enhanced gRNA scaffold, a modified mouse U6 promoter[106] and a second gRNA sequence and inserting it into the BbsI-digested *piggyBac* donor vector by Gibson cloning. All guide sequences were designed using the CRISPOR web tool (http://crispor.tefor.net/) and are listed in Supplementary Table 2. Epigenome-editing gRNA target sites were determined on the basis of previously published data on RING1B and SUZ12 occupancy[22].

Ezhip deletion strategy

The *Ezhip* deletion was generated by transfecting SpCas9 and two CRISPR sgRNAs specific to target sequences flanking the entire *Ezhip* coding sequence. Guide sequences were designed using the online CRISPOR online program (http://crispor.tefor.net/) and cloned into the pX459 plasmid harboring the *Cas9* gene. All guide sequences are listed in Supplementary Table 2. Transfections were performed in *Dnmt* TKO ES cells cultured in serum using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Following a 2-day treatment of puromycin (1 μg ml^{-1}) to select for transfectants, 96 individual clones were picked and screened by PCR for deletion. Mutated alleles were confirmed by Sanger sequencing of cloned PCR amplicons and two *Ezhip* KO subclones were kept for analysis. All primer sequences are listed in Supplementary Table 2.

Generation of stable epigenome-editing lines

The Dox-inducible epigenome-editing lines were generated as follows. The pEN111-DD-dCas9–SunTag vector (described above) was cotransfected into serum-grown ES cells along with a pX330-EN479 (Addgene, 86234), which contains a Cas9 nickase and a guide targeting the *Rosa26* locus, using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. After 1 day, Genetecin (Gibco) was added to the medium at a final concentration of 250 μg ml^{-1}. Approximately 1 week later, 96 individual neomycin-resistant clones were picked and screened for hemizygous insertion at *Rosa26* by PCR genotyping. The unmodified alleles of positive clones were analyzed by Sanger sequencing to ensure that there were no CRISPR–Cas9-induced mutations at the guide target site. Finally, inducible expression was confirmed by western blot. The scFv-GFP-3ACD/d3ACD/TET1CD constructs were transfected using the same method into a dCas9–SunTag/+ line, using puromycin (Gibco) at a final concentration of 1 μg ml^{-1}. Induction was confirmed by western blot and fluorescence microscopy.

All *piggyBac* donor plasmids were cotransfected with pBroad3_PBase_IRES_RFP (a gift from L. Giorgetti), which contains the *piggyBac* transposase, using Lipofectamine 2000. For guide constructs, stable integration was confirmed by hygromycin (Invitrogen) selection at a final concentration of 200 μg ml^{-1}. Stable integration of the constitutive dCas9–SunTag/TET1CD, dTET1, 3ACD or d3ACD constructs was confirmed by FACS of GFP-positive cells with a Becton Dickinson FACSAria fusion after 10 days. Data were collected using the FACSDiva software and analyzed using FlowJo version 10.8 Software (BD Life Sciences).

To restore DNMT1 expression, TKO cells were cotransfected with a pCAG-*Dnmt1* plasmid with 900-bp homology arms and a single plasmid containing spCas9 and an sgRNA targeting the *Tigre* locus. Transfectants were selected with puromycin (1 μg ml^{-1}) and 36 puromycin-resistant clones were screened by PCR genotyping. Six positive clones were then screened for DNMT1 expression by western blotting (as described below). The puromycin resistance cassette was removed by Flp-FRT-mediated recombination and clone B5 was selected for subsequent experiments.

Epigenome editing

For Dox-inducible 3ACD and d3ACD experiments, cells were grown in 2i + vitC medium for at least 7 days before switching to t2i conditions.

To induce the expression of the epigenome-editing machinery, Dox (Sigma) and Shield-1 (Aobious) were added to the medium at a final concentration of 0.5 µg ml^{-1} and 100 nM, respectively, for 7 days before the collection of the cells.

For Dox-inducible TET1CD experiments, cells were grown in 2i + vitC medium for at least 7 days before starting the differentiation to EpiLC. The induction was performed as above, but only for 4 days and in parallel with EpiLC differentiation.

For constitutive TET1CD/dTET1 experiments, after the lines were generated and confirmed, serum-grown ES cells were converted to 2i + vitC for 7–10 days, and then differentiated to EpiLCs. For both the constitutive and the inducible TET1CD lines, cells were FACS sorted for GFP expression at day 4 or 7 of EpiLC differentiation.

For constitutive 3ACD and d3ACD experiments, cells were cultured in serum conditions.

Targeted deep methylation sequencing

DNA was converted using NEBNext Enzymatic Methyl-seq conversion module (New England Biolabs) according to the manufacturer's instructions. We then followed the protocol described in Leitão et al.[107]. We designed primers for six amplicons of ~300 bp covering a ~2-kb region encompassing the potential Polycomb nucleation site upstream of *Zdbf2* TSS.

After amplification, the DNA was quantified by Qubit and the six amplicons were pooled in an equimolar manner for each sample for a second round of PCR to include Illumina-compatible sequences. The 20 PCRs corresponding to the 20 different samples were pooled in an equimolar manner and purified twice with the solid-phase reversible immobilization-based size selection following the kit instructions for selection of the fragment between 400 and 1,000 bp. Finally, the sample was prepared at a concentration of 4 nM containing 10% of PhiX DNA and 10% of another library provided by colleagues to add sequence complexity. Paired-end, 250-bp sequencing was performed on an Illumina miSeq machine. All primers are listed in Supplementary Table 2.

BS-Pyro

Genomic DNA was isolated from cells using the NucleoSpin tissue kit (Macherey-Nagel). Then, 500 ng of genomic DNA was subject to bisulfite conversion using the EZ DNA Methylation-Gold kit (Zymo). Target genomic regions were amplified by PCR using 1–2 µl of converted DNA with specific primers pairs recognizing BS-converted DNA, one of which was biotin conjugated (Supplementary Table 2), using Immolase (Ozyme) or GoTaq G2 (Promega) polymerases. Pyrosequencing assays were designed using the PyroMark Q24 Advanced 3.0.1 software and the sequencing reaction was performed with PyroMark Q24 Advanced reagents (Qiagen, 970922) according to the manufacturer's instructions. Briefly, 20 µl of the PCR reaction was mixed with streptavidin beads (Cytiva, 17511301) and binding buffer by shaking for 10 min at room temperature. Samples were then denatured with denaturation buffer using a PyroMark workstation (Qiagen) and released in a Pyro-Mark Q24 plate (Qiagen) preloaded with 0.375 µM sequencing primer. The single-strand PCR template was annealed to the sequencing primer by heating at 80 °C for 5 min. The plate was then processed in the PyroMark Q24 pyrosequencer (Qiagen). Results were analyzed with PyroMark Q24 Advanced 3.0.1 software. Graphical representation and statistical analysis were performed with GraphPad Prism version 9.3.1.

Protein extraction and western blot

For protein extraction, cells were incubated with a BC250 lysis solution (25 mM Tris pH 7.9, 0.2 mM EDTA, 20% glycerol and 0.25 M KCl) supplemented with protease inhibitors (Roche) and sonicated three times for 5 s each using a Bioruptor (Diagenode). The lysate was centrifuged at 13,000g for 10 min and the supernatant was transferred to a new tube. Protein concentrations were quantified using the Pierce BCA protein assay kit (Thermo Fisher Scientific) on an Infinite M200 plate reader

(Tecan). For western blotting, 3 µg of the protein samples were prepared with 4× Laemmli loading buffer (Bio-Rad) and denatured at 95 °C for 5 min. The samples were then loaded in Mini-PROTEAN TGX precast gels (Bio-Rad) and run at 100 V for 15 min and and 130 V for 1 h. Proteins were then transferred onto a PVDF membrane and rinsed with PBS–Tween (1%) (PBST). The membrane was blocked in 5% milk in PBST for 30 min. Membranes were then cut and incubated with different primary antibodies diluted in 5% milk at 4 °C overnight; anti-H3 (Abcam, 1791) and anti-H3K27me3 (Cell Signaling, C36B11) were used at 1:5,000 dilution; while anti-Cas9 (Active Motif, 61957), anti-GFP (Millipore Sigma, 11814460001), anti-DNMT1 (Ozyme, 5032T) and anti-lamin B1 (Abcam, ab16048) were used at 1:2,000. Blots were washed three times with PBST for 10 min at room temperature and incubated for 1 h with fluorescent marker-tagged secondary antibody diluted in PBST (goat anti-Rabbit IgG StarBright blue 700 or goat anti-Mouse IgG DyLight 800, Bio-Rad; 1:2,000 dilution). Lastly, we performed three washes with PBST for 10 min before imaging the membrane on a ChemiDoc system (Bio-Rad). The original scans of the blots are provided in the Source Data.

LUMA

First, 500 ng of genomic DNA was digested with MspI + EcoRI or HpaII + EcoRI (New England BioLabs). EcoRI was included for internal normalization. The extent of the enzymatic digestions was quantified by pyrosequencing (PyroMark Q24) and global CpG methylation levels were then calculated from the normalized peak height ratio of HpaII to MspI.

Reverse transcription (RT)–qPCR

Total RNA was extracted from cell pellets using the NucleoSpin RNA kit (Macherey-Nagel) or the MagMAX mirVana Total RNA isolation kit. First-strand complementary DNA (cDNA) synthesis was performed using the SuperScript III reverse transcriptase kit (Invitrogen). First, 500–1,000 ng of total RNA per reaction was mixed with 1 µl of 50 ng µl^{-1} random primers, 1 µl of 10 mM dNTPs and sterile H$_2$O up to 13 µl. The obtained cDNA was diluted 1:5 and, for each RT–qPCR reaction, 1 µl of cDNA was mixed with 5 µl of LightCycler 480 SYBR Green I master mix and 4 µl of forward and reverse primer mix (final primer concentration, 200 nM). RT–qPCR was run on a LightCycler 480 II (Roche Applied Science) using 384-well plates. The samples first underwent an initial incubation at 95 °C for 10 min and then 45 cycles of denaturation at 95 °C for 10 s, annealing at 61 °C for 20 s and extension at 72 °C for 20 s. Samples were amplified in triplicate with appropriate nontemplate controls. Relative gene expression was normalized to the geometric mean of the C_t levels of the two housekeeping genes, *Rrm2* and *Rplp0*. Graphical representation and statistical analysis were performed with GraphPad Prism version 9.3.1. Primer sequences can be found in Supplementary Table 2.

CUT&RUN (qPCR and sequencing)

To measure H3K27me3 enrichment after epigenome editing, we used the original CUT&RUN protocol[62] with minor modifications. Briefly, 2.5–5 × 10^5 cells per sample were pelleted and washed twice with wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM spermidine containing protease inhibitor). The cells were then incubated for binding with concavalin A magnetic beads (Polysciences) by rotating for 10 min at room temperature. The antibody incubation was carried out overnight on a rotor at 4 °C in 100 µl of antibody solution (wash buffer with 0.1% digitonin and 2 mM EDTA) containing 1 µl of anti-H3K27me3 (Cell Signaling, C36B11). Samples were then washed in wash buffer with 0.1% digitonin and incubated with pAG-Mnase on a rotor for 10 min followed by two more washes. The MNase reaction was then activated by adding 2 mM CaCl$_2$. After incubating at 0 °C for 30 min, the reaction was stopped with 1× final concentration of stop solution (2× 340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% digitonin and 1:200 RNase A). Chromatin was released by incubating samples at 37 °C for 10 min.

After centrifuging at full speed for 5 min, the supernatant was collected after incubation on magnet stand. Purified DNA was obtained by phenol–chloroform extraction and precipitated with 100% ethanol by centrifugation. The DNA pellet was washed in 70% ethanol, spun down and air-dried before being resuspended in 35 µl of 5 mM Tris-HCl pH 8.5. The pAG-MNase plasmid was obtained from Addgene (123461) and the protein was purified by the Curiecoretech Recombinant Protein Platform.

CUT&RUN DNA fragments were then subjected to qPCR to amplify selected target regions. First, 1 µl of DNA was mixed with 5 µl of LightCycler 480 SYBR Green I master mix and 4 µl of forward and reverse primer mix (final primer concentration 200 nM). Primers specific for target and control regions (positive control, *Pax5*; negative control, *Actb*) were used. RT–qPCR was run on a LightCycler 480 II (Roche Applied Science) using 384-well plates. Relative enrichment of H3K27me3 was calculated by comparing C_t values of target regions to the positive control locus, *Pax5* (ΔC_t method). Graphical representation and statistical analysis were performed with GraphPad Prism version 9.3.1. Primer sequences can be found in Supplementary Table 2.

Select samples were prepared for high-throughput sequencing (HTS). Sequencing libraries were constructed from purified CUT&RUN DNA fragments using the xGen low-input DNA library preparation kit from Integrated DNA Technologies (10009859) following the supplier's instructions. Library quality was assessed using 4200 TapeStation system (Agilent) and sequenced on an Illumina NovaSeq X (paired-end, 150-nt-long reads) by Novogene.

CUT&Tag

We performed CUT&Tag according with the protocol described previously[35], closely following the established procedure[108]. A total of 500,000 cells were used for each sample and duplicates were performed. The primaries antibodies, anti-H3K27me3 (Cell Signaling, C36B11; 1:100) and anti-H3K36me3 (Invitrogen, RM491; 1:100), were incubated overnight at 4 °C on a rotator. Samples were then incubated with a secondary anti-rabbit IgG (ABIN6923140) at room temperature on a rotator. After an incubation of 1 h at room temperature with a pA-Tn5 adaptor complex (Addgene, 124601), the Tn5 enzyme tagmentation reaction was performed by adding a buffer containing MgCl$_2$ to the samples and incubating for 1 h at 37 °C. To stop tagmentation and solubilize DNA fragments, we added 10 µl of 0.5 M EDTA, 3 µl of 10% SDS and 2.5 µl of 20 mg ml^{-1} proteinase K to each sample and samples were incubated at 50 °C for 1 h. Purified DNA was obtained by phenol–chloroform extraction and precipitated with 100% ethanol and centrifugation. The DNA pellet was washed in 80% ethanol, spun down and air-dried before being resuspended in 25 µl of 1 mM Tris-HCl pH 8.0. A universal i5 primer and a barcoded i7 primer were used to perform PCR[109]. Library quality was assessed using 4200 TapeStation system (Agilent) and paired-end 150-bp sequencing on an Illumina NovaSeq X was performed by Novogene.

WGBS

From the genomic DNA we provided, WGBS libraries were generated using the Accel-NGS Methyl-Seq DNA library kit (Swift Biosciences) and sequenced on an Illumina NovaSeq X (paired-end, 150-nt-long reads) by Novogene.

RNAseq

RNA extraction from cell pellets was performed using the KingFisher Duo Prime magnetic particle processor and the MagMAX mirVana total RNA kit, according to the manufacturer's instructions. Total RNA was sent to Novogene for poly(A) RNA library constructions and sequencing on an Illumina NovaSeq X (paired-end, 150-nt-long reads).

HTS read filtering and trimming

Before alignment, HTS library sequencing quality was assessed using FastQC (version 0.11.9)[110]. PCR duplicate reads were removed using BBMap clumpify (version 38.18)[111] and the following parameters: reorder dedupe = t, k = 19, passes = 6 and subs = 0.01 × read length, where the number of substitutions allowed is equal to the length of the read in nucleotides multiplied by the conservative sequencing error rate of Illumina machines. Lastly, adaptor and low-quality sequences were trimmed using Trimmomatic (version 0.39)[112] and the following parameters: 'ILLUMINACLIP:adaptors.fa:2:30:10SLIDINGWINDOW:4:20MINLEN:24', where 'adaptors.fa' corresponds to a FASTA file containing commonly used Illumina sequencing adaptors. H3K36me3 CUT&Tag data were processed with the additional parameter 'CROP:36'.

WGBS analysis

Filtered and trimmed (as described above) HTS reads were aligned to the mm10 genome using Bismark (version 0.23.1)[113] and default parameters. Reads with mates that did not survive read trimming or that could not be aligned in paired-end mode were concatenated and realigned in single-end mode. PCR duplicate reads were removed using deduplicate_bismark and CpG methylation information was extracted using bismark_methylation_extractor. Average CpG methylation levels over regions of interest was calculated using BEDOPS (version 2.4.40)[114]. For each dataset, 10-kb bins and CGIs with at least ten CpGs covered by at least five reads were kept. H3K27me3 peaks with at least five CpGs covered by at least five reads were kept. Spaghetti and violin plots were made using VisR (version 0.9.42)[115].

RNAseq analysis

Filtered and trimmed (as described above) RNAseq reads were aligned to the mm10 genome using STAR (version 2.7.9a)[116] and default parameters. Expression values at the transcript (University of California, Santa Cruz (UCSC) RefGene) and gene (National Center for Biotechnology Information (NCBI) RefSeq) levels were calculated using VisR using uniquely aligned reads (MAPQ = 255). De novo transcriptome assembly was performed for each sample using Stringtie (version 2.1.7)[117] with default parameters. De novo assemblies were subsequently combined and annotated using the UCSC RefGene reference file. The TSS of each isoform was extracted using a custom AWK script and those overlapping SWRs were identified using closestBed (version 2.30.0). Transcript level expression values (RPKM) were normalized for batch effects using pyCombat (version 0.3.2)[118] and input into sklearn PCA (version 1.1.1)[119] using pandas (version 1.4.3)[120,121] and numpy (version 1.23.1)[122] to generate a PCA plot, which was visualized using matplotlib (version 3.5.2)[123,124] and seaborn (version 0.11.2)[125]. Alignment files were converted to BigWig using deepTools bamCoverage (version 3.5.1)[126] and the following parameters: '--binSize 1 --smoothLength 0 --minMappingQuality 255 --normalizeUsing CPM --outFileFormat BigWig --blackListFileName ENCFF547MET.bed --ignoreForNormalization chrX chrM chrY', where 'ENCFF547MET.bed' corresponds to the coordinates of blacklisted regions as defined by the Kundaje lab as part of the ENCODE consortium. Scatter plots and heat maps were generated in Python using matplotlib, numpy, pandas and seaborn. Correlograms were generated using Morpheus and Spearman rank correlation of gene expression (RPKM) values. Linear modeling using Limma (version 3.54.1)[127] in R was used to identify statistically significant differences in gene expression levels between samples. Voom normalization and a stringent threshold of $\log_2$(fold change) > 1 and adjusted P value (t-test) < 0.05 was applied to define statistical enrichment.

CUT&Tag and CUT&RUN data analysis

Filtered and trimmed (as described above) CUT&TAG data were aligned to the mm10 reference using Bowtie2 (version 2.4.5)[128] and the following parameters: '--local-very-sensitive -no-mixed-dovetail -no-discordant -phred33 -I 10 -X 700'. Replicate alignment files were merged before peak calling using SEACR (version 1.3)[129] and the parameter '0.01 non-stringent'. Peaks within 10 kb were merged into domains using BEDTools merge (2.30.0)[130] and H3K27me3 enrichment levels (RPKM) were

calculated using VisR from alignments with MAPQ ≥ 10. H3K27me3 domains were agnostically clustered into six groups using k-means clustering (Morpheus, Broad Institute) of RPKM values. Statistical enrichment was performed as described above for RNAseq using the same filtering criteria. Sunset plots and meta plots of H3K27me3 enrichment over groups were generated using deepTools functions computeMatrix, plotHeatmap and plotGroup. The number of H3K27me3 domains overlapping CGIs, the promoters of NCBI RefSeq genes (TSS ± 1 kb) or bodies of genes was assessed using BEDTools intersect. Peakiness scores were calculated by averaging the enrichment level (RPKM) of the top 10% enriched peaks for each dataset, as previously described[24]. Total domain size was calculated by summing the lengths of all peaks for each dataset. Alignment files were converted to BigWig using deepTools bamCoverage and the same parameters as for RNAseq analysis (as described above) with the exception of '--minMappingQuality 10'. Transcription factor-binding motif analysis of SWRs and ESRs was performed using the MEME[131] and Rsat[132,133] suites.

Measuring CpG density
A CpG density track was generated following a previously described method[2]. Briefly, 100-bp-overlapping 1,000-bp bins were generated using BEDTools makewindows and the numbers of Cs, Gs and CpGs were counted using BEDTools getfasta and faCount. CpG density was then calculated for each bin using the formula CpG observed/expected ratio = no. of CpG/(no. of C + no. of G) and converted into a genome browser compatible track using bedgraphtobigwig[134].

Data visualization
UCSC Genome Browser[135] track data hubs[136] were used to display data over loci of interest.

Statistics and reproducibility
P values reported directly in figures are the result of two-sided t-tests assuming equal variance conducted without multiple-comparison correction. Western blot analysis of H3K27me3 levels in *Ezhip* KO cell lines was conducted twice independently with similar results. Western blot analysis of epigenome-editing constructs was conducted once and transgene expression was confirmed by flow cytometry for each experiment. *Ezhip* KO differentiation assays were conducted twice and representative morphological pictures were taken once.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Datasets generated in this study were uploaded to the NCBI Gene Expression Omnibus under accession number GSE242201. Additional data can be found in the Supplementary Information. Source data are provided with this paper.

Code availability
Scripts used to generate the figures presented are available under an GNU General Public License v3.0 on GitHub (https://github.com/julienrichardalbert/5mC-H3K27me3/releases/tag/v0.2).

References

104. Huang, Y. H. et al. DNA epigenome editing using CRISPR–Cas SunTag-directed DNMT3A. *Genome Biol.* **18**, 176 (2017).
105. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
106. Vidigal, J. A. & Ventura, A. Rapid and efficient one-step generation of paired gRNA CRISPR–Cas9 libraries. *Nat. Commun.* **6**, 8083 (2015).
107. Leitão, E. et al. Locus-specific DNA methylation analysis by targeted deep bisulfite sequencing. *Methods Mol. Biol.* **1767**, 351–366 (2018).
108. Kaya-Okur, H. S. & Henikoff, S. Bench top CUT&Tag v2. *protocols.io* https://doi.org/10.17504/protocols.io.z6hf9b6 (2019).
109. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
110. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
111. Bushnell, B. BBMap: a fast, accurate, splice-aware aligner. *Joint Genome Institute* (2014).
112. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
113. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
114. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
115. Younesy, H., Möller, T., Lorincz, M. C., Karimi, M. M. & Jones, S. J. M. VisRseq: R-based visual framework for analysis of sequencing data. *BMC Bioinformatics* **16**, S2 (2015).
116. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
117. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
118. Behdenna, A. et al. pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinformatics* **24**, 459 (2023).
119. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
120. McKinney, W. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) (SciPy, 2012).
121. Reback, J. et al. pandas-dev/pandas: Pandas 1.0.5. *Zenodo* https://doi.org/10.5281/zenodo.3509134 (2021).
122. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
123. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
124. Ari, N. & Ustazhanov, M. Matplotlib in Python. In *Proceedings of the 11th International Conference on Electronics, Computer and Computation* (eds Mahamat, M., Adeshina, S. A. & Arreytambe, T.) (IEEE, 2014).
125. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
126. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
127. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
128. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
129. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by sparse enrichment analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
130. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
131. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

132. Thomas-Chollier, M. et al. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40**, e31 (2012).
133. Thomas-Chollier, M. et al. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nat. Protoc.* **7**, 1551–1568 (2012).
134. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
135. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
136. Raney, B. J. et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003–1005 (2014).

Author contributions

J.R.A., conceptualization, methodology, formal analysis, data curation, visualization, writing—original draft and writing—review and editing; T.U., investigation, visualization, writing—original draft and writing—review and editing; A.M.-S., investigation; A.L.B., investigation; A.S., investigation; A.D., investigation; M. Scarpa, investigation; M. Schulz, investigation; M.V.C.G., conceptualization, methodology, investigation, writing—review and editing, supervision, project administration and funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-024-01405-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-024-01405-4.

Correspondence and requests for materials should be addressed to Maxim V. C. Greenberg.

Peer review information *Nature Structural & Molecular Biology* thanks Gabriella Ficz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

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