nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0060-9

RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells

Diana Guallar[1,2], Xianju Bi[3,10], Jose Angel Pardavila[2,10], Xin Huang[1,10], Carmen Saenz[1], Xianle Shi[1,4], Hongwei Zhou[1], Francesco Faiola[1], Junjun Ding[1], Phensinee Haruehanroengra[5], Fan Yang[1,6], Dan Li[1,7], Carlos Sanchez-Priego[1,7], Arven Saunders[1,7], Feng Pan[8], Victor Julian Valdes[9], Kevin Kelley[1], Miguel G. Blanco[2], Lingyi Chen[4], Huayan Wang[6], Jia Sheng[5], Mingjiang Xu[8], Miguel Fidalgo[2], Xiaohua Shen[3] and Jianlong Wang[1,7]*

Ten-eleven translocation (TET) proteins play key roles in the regulation of DNA-methylation status by oxidizing 5-methylcytosine (5mC) to generate 5-hydroxymethylcytosine (5hmC), which can both serve as a stable epigenetic mark and participate in active demethylation. Unlike the other members of the TET family, TET2 does not contain a DNA-binding domain, and it remains unclear how it is recruited to chromatin. Here we show that TET2 is recruited by the RNA-binding protein Paraspeckle component 1 (PSPC1) through transcriptionally active loci, including endogenous retroviruses (ERVs) whose long terminal repeats (LTRs) have been co-opted by mammalian genomes as stage- and tissue-specific transcriptional regulatory modules. We found that PSPC1 and TET2 contribute to ERVL and ERVL-associated gene regulation by both transcriptional repression via histone deacetylases and post-transcriptional destabilization of RNAs through 5hmC modification. Our findings provide evidence for a functional role of transcriptionally active ERVs as specific docking sites for RNA epigenetic modulation and gene regulation.

TET proteins maintain appropriate patterns of gene expression through epigenetic mechanisms that are relevant in stem cell and cancer biology[1]. Extensive studies on TET functions in mammalian gene regulation and chromatin dynamics have highlighted the contribution of a number of sequence-specific DNA-binding transcription factors such as NANOG, PRDM14, PU.1, and WT1 (reviewed by Wu and Zhang[2]) to 5hmC deposition at the genome, which leads to active demethylation of target genes. Whereas 5mC modification of RNA is firmly established[3], the potential roles of TET proteins in mediating 5mC-to-5hmC oxidation in RNA are just beginning to be appreciated[4–8].

Pluripotent mouse embryonic stem cells (ESCs) are derived from the inner cell mass of the preimplantation blastocyst. ESCs characteristically suppress the transcription of most ERVs[9] but fluctuate with MERVL activity in the two-cell (2C)-like population with an expanded potency[10]. ESCs express all components of the methylation and demethylation pathways, with all oxidized forms of 5mC detected at the DNA level. Despite extensive research into the role of TET proteins in genome regulation, little is known about their functions in controlling ERVs, which make up 8–10% of mouse and human genomes.

Here we defined the TET2 interactome in mouse ESCs and identified the RNA-binding protein PSPC1 as a binding partner of TET2. We show that TET2 can be recruited to chromatin in an RNA-dependent manner through its physical association with PSPC1. By identifying RNA targets of PSPC1, we demonstrated that PSPC1, while binding to *MERVL* transcripts, recruits TET2 function for both transcriptional and post-transcriptional regulation of *MERVL* through HDAC1/2-mediated repression and RNA hydroxymethylation (5hmC)-mediated degradation.

Results

TET2 interaction with PSPC1 is required for recruitment to chromatin. To identify factors that may regulate TET2 chromatin binding, we investigated the TET2 interactome in mouse ESCs. Using a 3×Flag-tagged *Tet2*-knockin ESC line, we carried out affinity purification coupled with mass spectrometry analysis of TET2-containing protein complexes (Supplementary Fig. 1a–c), following our well-established strategies[11,12]. Among the top TET2-interacting partners, we found the nuclear protein PSPC1 (Fig. 1a, Supplementary Fig. 2a and Supplementary Table 1). The interaction between PSPC1 and TET2 was further confirmed by immunoprecipitation and coimmunoprecipitation experiments (Fig. 1c), and was not compromised by the absence of other TET2-interacting partners such as OGT, SIN3A and NONO (Fig. 1a and Supplementary Fig. 2b,c). *Pspc1* showed a gene expression pattern similar to that

[1]The Black Family Stem Cell Institute and Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]CiMUS, Universidade de Santiago de Compostela-Health Research Institute (IDIS), Santiago de Compostela, Coruña, Spain. [3]Tsinghua-Peking Center for Life Sciences, School of Medicine, Tsinghua University, Beijing, China. [4]State Key Laboratory of Medicinal Chemical Biology and College of Life Sciences, Nankai University, Tianjin, China. [5]Department of Chemistry and The RNA Institute, University at Albany, State University of New York, Albany, NY, USA. [6]Department of Animal Biotechnology, College of Veterinary Medicine, Northwest A&F University, Xianyang, China. [7]The Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]Sylvester Comprehensive Cancer Center, Department of Biochemistry and Molecular Biology, University of Miami, Miami, FL, USA. [9]Department of Cell Biology and Development, Instituto de Fisiologia Celular, UNAM, Mexico City, Mexico. [10]These authors contributed equally: Xianju Bi, Jose Angel Pardavila and Xin Huang. *e-mail: jianlong.wang@mssm.edu

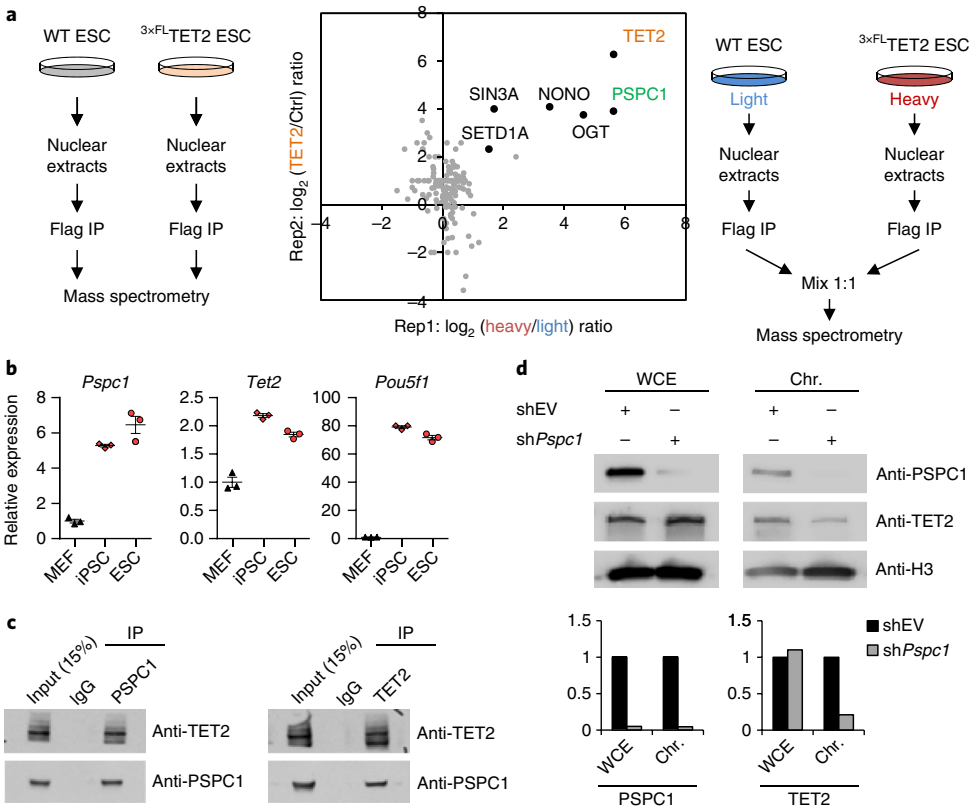

Fig. 1 | TET2 is recruited to chromatin by the RNA-binding protein PSPC1. a, Illustration of the two complementary techniques used to identify TET2-interacting proteins in mouse ESCs. Left, the experimental scheme for Flag immunoprecipitation (IP) followed by mass spectrometry of $^{3\times FL}$*Tet2*-knockin and wild-type (WT) control ESC lines (Rep1). Right, scheme of the SILAC-based in vivo labeling approach used to identify TET2 partners by IP with anti-Flag using the nuclear extracts from $^{3\times FL}$*Tet2*-knockin ESCs and WT ESCs followed by mass spectrometry (Rep2). Center, ratios of TET2-interacting peptides versus nonspecific peptides detected by affinity purification coupled with mass spectrometry in both IP experiments. Ctrl, control. **b**, Relative RNA expression levels of *Pspc1*, *Tet2* and the ESC marker *Pou5f1* (also known as *Oct4*) in differentiated mouse embryonic fibroblast (MEF) versus pluripotent (induced pluripotent stem cell (iPSC) and ESC) cell lines. Data are from one representative experiment ($n=3$ technical replicates) and are presented as mean ± s.e.m. **c**, Validation of the interaction between endogenous PSPC1 and TET2 by coimmunoprecipitation followed by western blotting analysis. IgG was used as a negative control for the IP. The percentage of input (15%) is shown. **d**, Reduced TET2 chromatin occupancy after *Pspc1* depletion. Top, western blotting analysis of total (whole cell extract (WCE)) and chromatin-bound (Chr.) PSPC1 and TET2 in ESCs treated with control empty vector (shEV) or shRNA targeting *Pspc1*. Histone H3 was used as a loading control. Bottom, quantitation of the relative levels of PSPC1 and TET2 in WCE and chromatin from cells treated with empty vector or sh*Pspc1*. Images in **c**,**d** are representative of immunoblots from 2 independent experiments.

of *Tet2* across multiple tissues, including much higher enrichment in pluripotent cells than in somatic mouse embryonic fibroblasts (Fig. 1b and Supplementary Fig. 2d,e).

To test the effect of PSPC1 loss on TET2 chromatin occupancy, we analyzed TET2 levels in pure chromatin extracts after PSPC1 knockdown, and observed reduced occupancy (<30%) of TET2 (Fig. 1d and Supplementary Fig. 3a). Notably, knockdown of PSPC1 did not dramatically affect ESC properties (Supplementary Fig. 3b–d), which is similar to what is observed after the loss of TET2 function[13,14]. These results suggest that the partnership between TET2 and PSPC1 may be necessary for proper function of TET2 at the chromatin level, although other TET2 partners could also participate in its recruitment, given that an appreciable level of TET2 still remained at the chromatin after PSPC1 depletion (Figs. 1d and 2b).

PSPC1 recruits TET2 to chromatin through RNA. Because PSPC1 is an RNA-binding protein, we explored the possibility that its RNA-binding ability could be relevant to TET2 recruitment, similar to what has been observed for other transcription factors and epigenetic regulators[15,16]. To distinguish RNA-dependent from RNA-independent functions of PSPC1 in TET2 recruitment, we used the CRISPR–Cas9 nuclease system to generate a *Pspc1*-knockout (KO) ESC line (Supplementary Fig. 4a,b), and then rescued PSPC1 loss with Piggybac vectors expressing either wild-type PSPC1 or PSPC1 bearing four mutations in its RNA-binding domains[17] (F118A, F120A, K197A and F199A; hereinafter referred to as PSPC1Mut) (Fig. 2a and Supplementary Fig. 4c,d). We found that whereas the physical interaction between PSPC1 and TET2 was independent of PSPC1 RNA-binding capacity (Fig. 2a and Supplementary Fig. 4e), both TET2 (Fig. 2b and Supplementary Fig. 4f) and PSPC1 (Fig. 2c) chromatin occupancies were largely dependent on intact RNA-binding domains of PSPC1.

We next explored whether PSPC1 can mediate TET2–RNA interactions. We adapted an RNA immunoprecipitation protocol[18] to develop an in vitro RNA immunoprecipitation approach, termed iv-RIP, wherein PSPC1 or TET2 protein complexes were affinity-purified in the absence of endogenous nucleic acids, and their abilities to interact with total RNA were subsequently assayed (Fig. 2d). Our results indicate that TET2 protein complexes interact with RNA species in a PSPC1-dependent manner, both in vitro (Fig. 2d and Supplementary Fig. 5a) and in vivo (Supplementary Fig. 5b). In contrast, PSPC1 binding to RNA was independent of TET2 (Supplementary Fig. 5c).

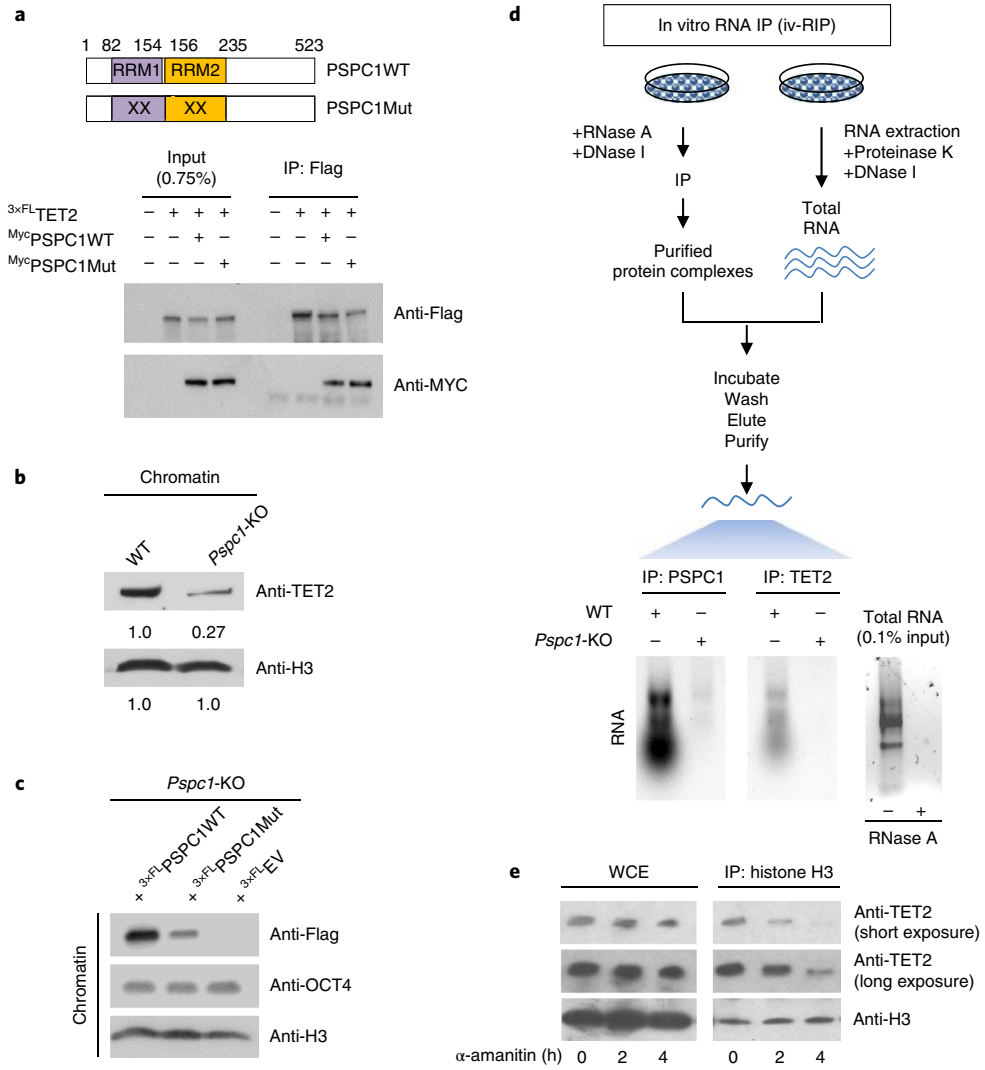

Fig. 2 | PSPC1 mediates TET2 recruitment to chromatin through RNA. a, Top, a depiction of the wild-type (PSPC1WT) and RNA-binding mutant (PSPC1Mut) PSPC1 protein structures. Bottom, the maintained interactions of $^{3\times FL}$TET2 with both ^{Myc}PSPC1WT and ^{Myc}PSPC1Mut as determined by coimmunoprecipitation. HEK293T cells were transiently transfected with constructs expressing $^{3\times FL}$TET2 and ^{Myc}PSPC1 or ^{Myc}PSPC1Mut and then subjected to immunoprecipitation (IP) with anti-Flag and western blotting with the indicated antibodies. **b**, Reduction of chromatin-bound TET2 in *Pspc1*-KO lines relative to that in the wild-type (WT) control. The relative levels of TET2, normalized to histone H3 (H3) and wild-type controls, are indicated. **c**, Chromatin-bound $^{3\times FL}$PSPC1WT and $^{3\times FL}$PSPC1Mut proteins in the *Pspc1*-KO rescued cell lines. OCT4 is shown as a negative control. EV, empty vector. **d**, Top, a depiction of the in vitro RNA-immunoprecipitation assay (iv-RIP) protocol. Bottom, agarose gel analysis of RNA bound by PSPC1 or TET2 protein complexes in wild-type and *Pspc1*-KO ESCs. Total RNA used for IP was incubated with DNase I to ensure the complete absence of DNA contamination (0.1% input is shown on the right). **e**, TET2 chromatin binding was reduced after the inhibition of transcription by α-amanitin treatment. TET2 levels in chromatin were evaluated by nucleosome pulldown (IP: histone H3) compared to total TET2 protein in whole-cell extract (WCE) after transcriptional inhibition with α-amanitin for 2 and 4 h, compared with that in untreated cells (0 h). Histone H3 was used as a loading control in **b**,**c**,**e**. All images are representative of at least 2 independent experiments.

Given that TET2 chromatin recruitment is facilitated by PSPC1 RNA-binding properties, we hypothesized that transcriptional inhibition might affect chromatin targeting of TET2. We treated ESCs with α-amanitin briefly to induce global transcription inhibition while avoiding TET2 protein changes (Fig. 2e, left), and confirmed by nucleosome pulldown that TET2 binding to chromatin was reduced after transcriptional inhibition compared with that in untreated cells (Fig. 2e, right).

Characterization of the RNA interactome involved in PSPC1–TET2 chromatin occupancy. To characterize the RNA interactome of PSPC1 involved in the binding and maintenance of PSPC1 and TET2 at chromatin, we carried out cross-linking followed by Flag immunoprecipitation and high-throughput sequencing (CLIP-seq) to identify PSPC1-associated RNA species in ESCs overexpressing 3×Flag- and biotin-tagged PSPC1 (Fig. 3a, left, and Supplementary Table 2). Supporting a functional connection between PSPC1 and TET2, an in silico analysis of the available TET2 ChIP-seq dataset for ESCs[19] showed high correlation between TET2 occupancy at DNA regions and PSPC1-bound RNA peaks (Fig. 3a, right), thereby pointing to cotranscriptional recruitment of TET2 to those loci through PSPC1–RNA interactions. We validated this observation by chromatin immunoprecipitation coupled with quantitative PCR (ChIP-qPCR), which showed that TET2 chromatin occupancy at representative loci (for example, *Adss* and *Ywhae*) that transcribe RNAs bound by PSPC1

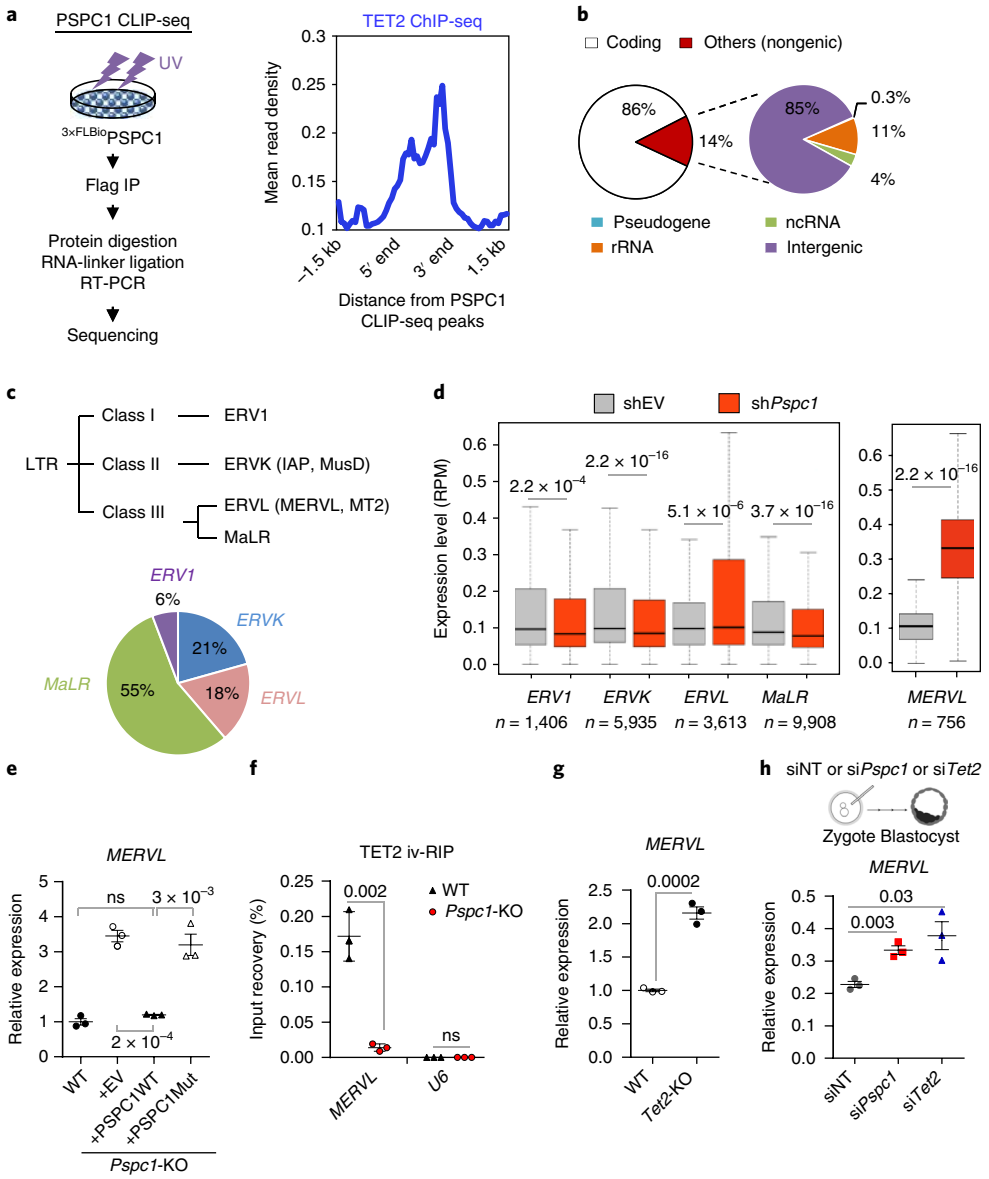

Fig. 3 | PSPC1 binds to *MERVL* RNAs with TET2 to repress their expression. a, Left, a depiction of $^{3\times FLBio}$PSPC1 RNA immunoprecipitation (IP) followed by high-throughput sequencing (CLIP-seq) in ESCs. Right, the mean read density of TET2 ChIP-seq[19] at the chromatin loci corresponding to PSPC1 CLIP-seq peaks. Data were generated from one experiment. **b**, Distribution of PSPC1 CLIP-seq peaks in coding and nongenic RNAs. **c**, Top, a schematic tree and classification of LTR-containing retrotransposons. Bottom, quantitation of LTR-containing RNAs of each LTR family bound by PSPC1. **d**, Expression (in reads per million mapped reads (RPM)) of class I, II and III LTR ERVs, including *MERVL*, after PSPC1 depletion (sh*Pspc1*) compared with that in the control (shEV). In box plots, center lines indicate the median, box edges denote the interquartile range (IQR), and whiskers extend to ±1.5× the IQR. Statistical analysis was done by Wilcoxon rank sum test with continuity correction; *P* values are shown in the figure. **e**, Quantitative PCR analysis of *MERVL* expression in wild-type (WT) and *Pspc1*-KO ESCs rescued with $^{3\times FL}$PSPC1WT or $^{3\times FL}$PSPC1Mut. The data are relative to expression in wild-type ESCs. **f**, Quantitative PCR analysis of *MERVL* abundance among TET2-interacting RNAs analyzed by iv-RIP in wild-type and *Pspc1*-KO cells. *U6* was used as a negative control. **g**, *MERVL* expression in wild-type and *Tet2*-KO ESCs. **h**, Top, a cartoon depicting zygote injection with siRNAs to deplete PSPC1 (si*Pspc1*) or TET2 (si*Tet2*) or with a nontargeting control siRNA (siNT) followed by in vitro culture until the blastocyst stage. Bottom, quantitative PCR of *MERVL* expression in blastocysts obtained in the three indicated conditions. Data are relative to β-actin (*Actb*) and siNT controls. Data in **e**–**h** are shown as the mean ± s.e.m. ($n = 3$ independent experiments). Two-tailed Student's *t*-test; *P* values are shown in the figure. ns, not significant.

was compromised by the loss of PSPC1 RNA-binding capacity (Supplementary Fig. 6a).

PSPC1 regulates endogenous retrovirus expression. Given that recruitment of PSPC1 to chromatin is stabilized by binding of the protein to RNA, and that paraspeckle proteins have been previously implicated in transcriptional regulation[20,21], we speculated that PSPC1 might participate in transcriptional regulation of its RNA targets. To test this hypothesis, we knocked down PSPC1 and examined global expression changes by RNA-seq. Although loss of PSPC1 function led to upregulation and downregulation of 362 and 610 coding genes, respectively (Supplementary Fig. 6b and Supplementary Table 3), only approximately 18% of the genes deregulated after PSPC1 depletion were bound at their RNAs by PSPC1 (Supplementary Fig. 6c). Gene Ontology analysis showed that genes observed to be deregulated after PSPC1 knockdown are involved

in developmental processes (Supplementary Fig. 6d). Compared with published RNA-seq data for mouse embryos in early development[22], we observed specific enrichment of deregulated targets in the 2C-stage embryos (Supplementary Fig. 6e), in which remarkable coregulation of host stage-specific genes and retrotransposable elements (REs) is prevalent[23].

REs are vestiges of ancient retroviral infections that constitute nearly 40% of the mammalian genome[24] and have been proposed to act as *cis*-regulatory sequences for transcriptional control of neighboring genes (reviewed in refs [25,26]). Given that PSPC1 binds both coding and noncoding RNAs (Fig. 3b) and that only a minority of coding genes with transcribed RNAs bound by PSPC1 were affected by PSPC1 knockdown (Supplementary Fig. 6b,c), we hypothesized that PSPC1 might regulate gene expression by binding to REs that in turn regulate neighboring genes. In agreement with this hypothesis, further examination of CLIP-seq data showed that among PSPC1-bound RNAs other than coding ones, a vast majority arose from intergenic regions (85%) (Fig. 3b), including LTR-containing (ERVs) and non-LTR (long and short interspersed nuclear elements) subclasses of REs (Supplementary Fig. 7a). The interaction between PSPC1 and RE-derived RNAs was further confirmed by RNA CLIP-qPCR experiments (Supplementary Fig. 7b,d,e). It is well known that ESC potency fluctuates with ERV activity[10] and that ERV expression is transcriptionally regulated by multiple epigenetic pathways[27,28]. We discovered abundant representation of the ERVK, ERVL and MaLR ERV families among PSPC1-bound RNAs whose binding was independent of their relative abundance (Fig. 3c and Supplementary Fig. 7c–e), and observed global deregulation of their expression after PSPC1 depletion (Fig. 3d and Supplementary Fig. 7f,g). These results suggest that the interaction of ERV RNA species with PSPC1 is required for the regulation of cellular levels of these species.

PSPC1 and TET2 participate in the repression of *MERVL* and adjacent genes during development. In line with our hypothesis that ERVs bound by PSPC1 are likely to influence adjacent genes that are deregulated by PSPC1 depletion, we found that PSPC1-interacting ERVs were more frequently located near the transcription start sites of PSPC1-deregulated genes than those ERV LTRs not bound by PSPC1 (Supplementary Fig. 8a). Notably, transcripts corresponding to *MERVL* elements, which are expressed in the 2C-stage embryo[23,29], were among the most strongly induced ERVs in the absence of PSPC1 (Fig. 3d and Supplementary Fig. 7f,g). In agreement with this, we observed that 2C-embryo genes that were deregulated after PSPC1 knockdown (Supplementary Fig. 6b) had class III ERVL elements in close proximity to their transcription start sites (Supplementary Fig. 8b). Using a luciferase reporter system containing the *MERVL* element regulating the 2C-stage-specific gene *Zfp352*[30], we further confirmed PSPC1-mediated repression (Supplementary Fig. 8c). Moreover, we also demonstrated that artificial activation of endogenous *MERVL* by a CRISPR SAM (synergistic activation mediator) system[31] could recapitulate PSPC1 depletion in derepressing those PSPC1 bound and unbound 2C-stage-specific genes (Supplementary Fig. 8d,e). Thus, our results are consistent with the critical roles of ERVs in shaping the evolution of gene regulatory networks that underlie early embryonic development[32]. In addition, we identified RNA-binding-dependent control of ERV expression, which was exemplified by the repressive effect of PSPC1 on *MERVL* and evident in the *MERVL* repression in *Pspc1*-KO ESCs rescued by wild-type PSPC1, but not by PSPC1Mut (Fig. 3e).

To gain insight into the molecular mechanisms by which PSPC1 represses *MERVL* elements, we investigated the contribution of its interacting partner TET2 to the observed repression. We found that TET2 was also able to bind LTR-containing elements, such as *MERVL*, *IAP* and *MusD*, as well as non-LTR elements, and, more important, that such interactions were also dependent on PSPC1 (Fig. 3f and Supplementary Fig. 9a,b). Depletion of TET2 (Fig. 3g and Supplementary Fig. 9c) in ESCs caused expression changes in these REs similar to those observed after PSPC1 depletion (Fig. 3d and Supplementary Fig. 7f,g), and comparable to, if not more significant than, the transcriptional deregulation of these REs in ESCs lacking other well-known ERV regulators such as *Trim28* and *Ehmt2*[28,30,33] (Supplementary Fig. 10a–d). These results suggest that the PSPC1–TET2 partnership may be critically involved in the regulation of REs whose expression is dynamically controlled during early embryonic development[23].

MERVL expression peaks at the 2C stage of embryonic development and is greatly reduced by the blastocyst stage[23]. To validate our ESC data for PSPC1/TET2-mediated *MERVL* regulation in an in vivo developmental setting, we injected small interfering RNAs (siRNAs) targeting *Pspc1* or *Tet2*, or a nontargeting control siRNA, into mouse zygotes and cultured those embryos in vitro until the blastocyst stage (Fig. 3h and Supplementary Fig. 11a,b). In line with previous reports on the dispensability of PSPC1[34] and TET2[35] for early development, we did not observe any significant delay in embryonic development after knockdown of PSPC1 or TET2. However, we observed derepression of *MERVL* elements and *MERVL*-associated genes during early development after depletion of either PSPC1 or TET2 (Fig. 3h and Supplementary Fig. 11c–f). These results show that both PSPC1 and TET2 participate in the regulation of *MERVL* elements and their adjacent genes during development.

Differential regulation of PSPC1-bound ERV families by TET2. To understand how PSPC1 might cooperate with TET2 to regulate ERV expression, we examined the correlation between our PSPC1 CLIP-seq peaks and enrichment of 5hmC and 5mC at the DNA level in ESCs[36]. In contrast to the high enrichment of 5mC/5hmC on nonrepetitive coding sequences bound by PSPC1, 5mC and 5hmC were barely detectable at LTR-containing genomic loci (Supplementary Fig. 12a). Whereas 5hmC was absent in *MERVL* genomic loci, this epigenetic mark was readily detectable at the DNA of those class II ERV elements whose proper activation was dependent on PSPC1 RNA-binding ability and TET2 presence (i.e., *IAP* and *MusD*; Supplementary Figs. 7h, 9c and 12d), consistent with 5hmC-mediated DNA demethylation and transcriptional activation. These results indicate distinct mechanisms for the regulation of class II (*IAP*, *MusD*) and class III (*MERVL*) ERVs by PSPC1 and/or TET2, and also suggest that catalytic-activity-dependent and -independent functions of TET2 may be involved in the transcriptional activation of class II ERVs *IAP* and *MusD* and repression of class III ERV *MERVL* elements, respectively.

TET2 mediates 5-hydroxymethylation of *MERVL* RNAs. To further understand these distinct regulatory mechanisms, we rescued *Tet1/2/3* triple-knockout (*Tet*-TKO) ESCs with wild-type TET2 or a catalytic mutant form of TET2 (TET2Mut) (Supplementary Fig. 12c). As expected, we found that TET2Mut did not rescue *IAP* and *MusD* expression in *Tet*-TKO cells (Supplementary Fig. 12d). To our surprise, we also found that TET2Mut could not efficiently rescue *MERVL* repression observed in cells rescued by wild-type TET2 (Fig. 4a), which suggests that the catalytic activity of TET2 probably contributes to *MERVL* repression, possibly through a DNA-independent mechanism.

Recent findings identified 5hmC as an epigenetic mark on RNA species in mammalian[5,6,8,37] and nonvertebrate organisms[7], but the functional aspects of this RNA epigenetic modification remain poorly defined. We therefore decided to determine whether PSPC1 and TET2 could mediate 5hmC modification of *MERVL* transcripts, and how such RNA modification might control *MERVL* abundance. We immunoprecipitated DNA-free RNA by using an antibody

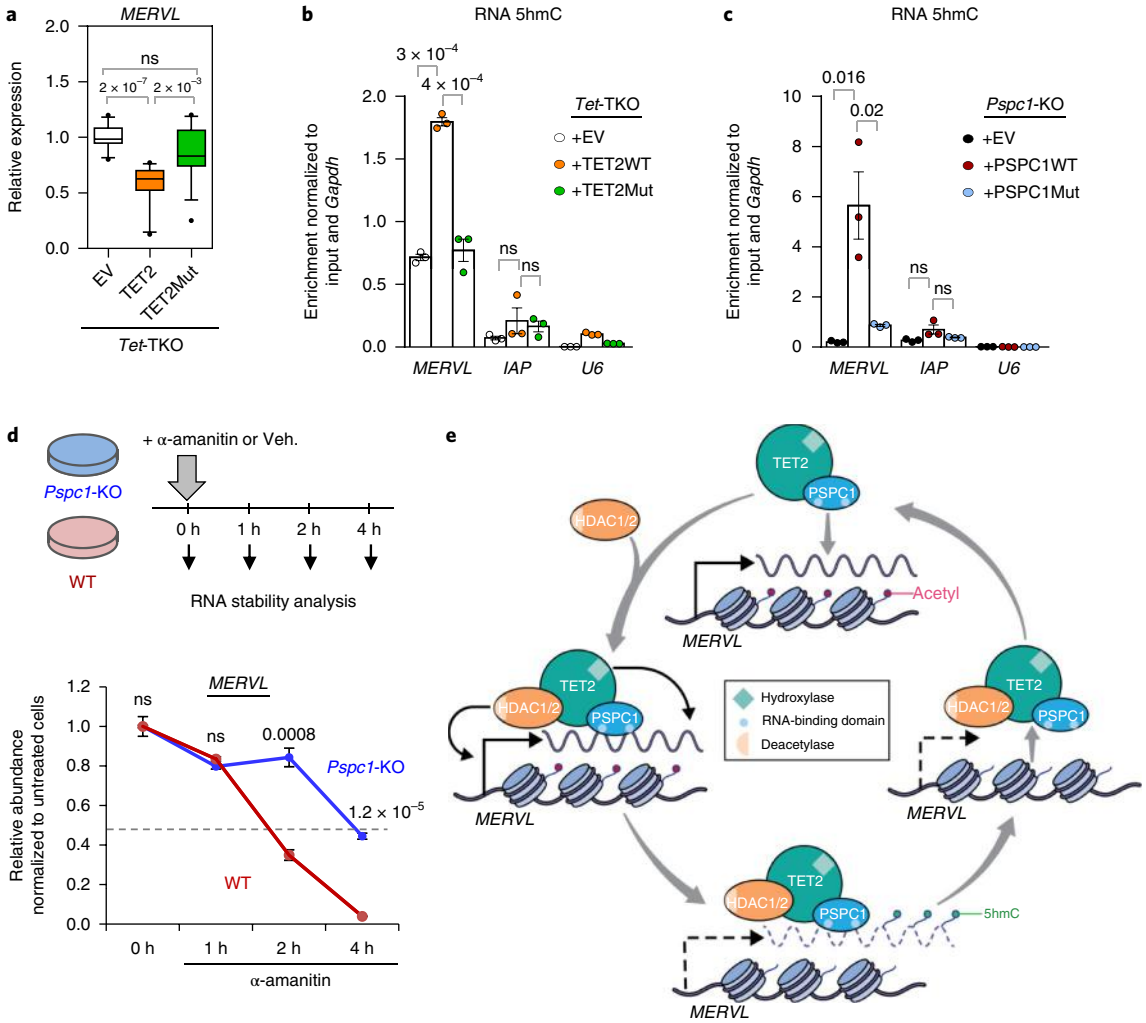

Fig. 4 | PSPC1 and TET2 silence *MERVL* transcriptionally and post-transcriptionally. a, Box plots showing *MERVL* expression in *Tet*-TKO ESCs rescued with an empty vector (EV), a wild-type protein (TET2WT) or a catalytic mutant protein (TET2Mut). Center lines, medians; whiskers extend to the 10th/90th percentile. Data are from 5 independent experiments ($n = 14$ total technical replicates for each rescue). **b**,**c**, *MERVL* and *IAP* enrichment, compared with a *U6* negative control, among anti-5hmC-immunoprecipitated RNAs in *Tet*-TKO (**b**) and *Pspc1*-KO (**c**) ESCs rescued with an empty vector (+EV), a wild-type protein or a mutant protein. Data are presented as mean ± s.e.m. ($n = 3$ independent experiments). **d**, Top, schematic of the protocol used for inhibition of transcription with α-amanitin for the RNA stability assay. Bottom, relative abundance of *MERVL* RNA in wild-type and *Pspc1*-KO ESCs after transcriptional inhibition for 1, 2 or 4 h with α-amanitin. Data are normalized to abundance in untreated cells at time 0 (vehicle (Veh.) without treatment). Error bars indicate ± s.e.m. ($n = 3$). **e**, A model of *MERVL* regulation by PSPC1/TET2 and HDAC1/2 in ESCs. PSPC1 binding to actively transcribed *MERVL* RNAs recruits TET2 and HDAC1/2 to chromatin. TET2 catalyzes 5hmC modification of *MERVL* RNAs, resulting in their destabilization, and HDAC1 and HDAC2 deacetylate histones at the chromatin level, leading to transcriptional repression of the *MERVL* loci. Transcriptional and post-transcriptional repression of *MERVL* leads to the release of the PSPC1–TET2–HDAC1/2 complex from chromatin. Sporadic reactivation of *MERVL* expression, well-recognized in conventionally cultured ESCs[10], via a yet-to-be defined mechanism, leads to the recruitment of PSPC1–TET2–HDAC1/2 for transcriptional and post-transcriptional control of *MERVL* and coordinated gene expression. Illustration by Jill Gregory. Printed with permission from ©Mount Sinai Health System. **a**–**d**, Two-tailed Student's *t*-test; *P* values are shown in the figure. ns, not significant.

raised against 5mC or 5hmC, and identified *MERVL* and *MERVL*-chimeric transcripts, but not *IAP* or *MusD*, among 5hmC-modified RNAs in ESCs (Supplementary Fig. 13a–d). We also confirmed the presence of 5mC-modified *MERVL* transcripts (data not shown), in agreement with a recent report on global 5mC profiling of poly(A) RNA in mouse ESCs[38]. We found that the PSPC1–TET2 complex could bind both 5mC- and 5hmC-modified RNAs in vitro, although with a much weaker affinity for the latter (Supplementary Fig. 13e). The observation that PSPC1–TET2 had a much higher affinity for 5mC-modified than for 5hmC-modified RNAs suggests a possibly predominant 5mC 'reader' function of TET2 that is followed by its 'writing' function in the oxidation of 5mC to 5hmC, thereby causing the subsequent release of the protein complex from RNA (Supplementary Fig. 13f). To confirm the functional contribution of TET2 to *MERVL* 5hmC modification, we carried out a rescue experiment, and observed that only wild-type TET2, and not TET2Mut or either TET1 or TET3, rescued 5hmC levels on *MERVL* transcripts in *Tet*-TKO cells (Fig. 4b and Supplementary Fig. 14a). Consistent with PSPC1-dependent RNA association of TET2 (Fig. 2d and Supplementary Figs. 5a and 9a), only the rescue of *Pspc1*-KO cells with RNA-binding-competent PSPC1 (wild-type PSPC1) restored the 5hmC modification on *MERVL* transcripts (Fig. 4c). These results establish the requirement of PSPC1 and TET2 for 5hmC modification of *MERVL* RNAs.

5hmC modification of *MERVL* RNAs leads to their destabilization. Given that previous studies have shown that the 5hmC precursor, 5mC, can have stabilizing effects on mRNA[39,40], and that *MERVL* abundance is increased after the loss of TET2 and PSPC1 (Fig. 3e,g and Supplementary Figs. 7f and 9c), we evaluated the effects of 5hmC deposition on the stability of *MERVL* transcripts. To this end, we monitored *MERVL* levels after transcription inhibition with α-amanitin or after pulsed incorporation of 5-ethynyl uridine in wild-type, *Pspc1*-KO and *Tet2*-KO ESCs. Our analyses showed a significant increase in the stability of *MERVL* transcripts in cells depleted of PSPC1 or TET2 compared with that in wild-type cells (Fig. 4d and Supplementary Fig. 14b,c), which correlated strongly with the absence of 5hmC in these transcripts (Fig. 4b,c). Moreover, we obtained similar results when we treated cells with the transcriptional inhibitor triptolide (data not shown). To further investigate the hypothesis that 5hmC deposition on *MERVL* transcripts facilitates their destabilization, we identified PSPC1 consensus RNA-binding motifs and carried out a minigene reporter (d2EGFP) assay to assess the effect of PSPC1 binding on EGFP stability (Supplementary Fig. 14d,e). We found that some of the PSPC1 binding motifs (motifs 1, 4 and 5), when fused to the d2EGFP minigene, mediated a substantial decrease in EGFP protein expression, evident in the reduced mean fluorescence intensities of wild-type versus mutant motifs after α-amanitin treatment (Supplementary Fig. 14f). These results indicate that 5hmC deposition on *MERVL* transcripts facilitates their degradation, and that the PSPC1–TET2 partnership contributes to *MERVL* destabilization in ESCs (Supplementary Fig. 13f).

The PSPC1–TET2 complex recruits HDAC1/2 for *MERVL* transcriptional repression. We noted that the rescue of *Tet*-TKO cells with TET2Mut could lead to a modest yet appreciable decrease in *MERVL* expression, although to a lesser extent than observed in cells rescued by wild-type TET2 (Fig. 4a), which suggested that the catalytic activity of TET2 and 5hmC-mediated RNA degradation might not be the only mechanism for *MERVL* repression. In line with this, a recent study showed that TET2 can mediate transcriptional repression in chromatin in a catalytic-activity-independent fashion through histone deacetylase (HDAC) complexes in leukemia cells[41]. Indeed, HDACs were among the TET2 partners in our interactome (Supplementary Table 1), and we confirmed the interaction of both HDAC1 and HDAC2 (HDAC1/2) with PSPC1 (Supplementary Fig. 15a,b). In addition, we detected a reduction in HDAC1/2 occupancy in *MERVL* loci in the absence of PSPC1 (Supplementary Fig. 15c). Such PSPC1-dependent HDAC1/2 binding to *MERVL* loci is also reliant on PSPC1 RNA-binding capacity, as demonstrated by the lack of HDAC1/2 binding in *Pspc1*-KO cells rescued with PSPC1Mut compared with that in cells rescued by wild-type PSPC1 (Supplementary Fig. 15c). These data suggest that HDAC1/2 and/or other HDAC activity could mediate the transcriptional repression of *MERVL*. Consistent with this, both chemical inhibition of HDAC activity by valproic acid and shRNA-mediated knockdown of *Hdac1/2* led to increased expression of *MERVL* and *MERVL*-associated genes (e.g., *Zfp352*) (Supplementary Fig. 15d,e), in line with what we observed after PSPC1 and TET2 depletion. The repressive activity of HDAC1/2 on *MERVL* expression was dependent on the presence of RNA-binding-competent PSPC1, and was independent of TET2 catalytic activity (Supplementary Fig. 15f). These observations suggest that PSPC1 and TET2 may also act together with HDAC1/2 for transcriptional silencing of *MERVL* and its associated gene regulation, in a manner independent of TET2 catalytic activity and distinct from the well-recognized epigenetic mechanism via histone methylation[28,30] (Supplementary Fig. 15g).

Discussion

Here we sought to understand how TET2 is recruited to chromatin for epigenetic control in pluripotent stem cells. Our interactome study in mouse ESCs not only identified the well-known TET2 partner protein OGT[19,42,43], thus validating the approach, but also uncovered novel interacting partner proteins, in particular the two RNA-binding proteins NONO and PSPC1 (Fig. 1). Both NONO and PSPC1 are components of paraspeckles with functions in RNA processing, nuclear retention of mRNA, and stress response[44,45]. However, ESCs do not form paraspeckles[21,44], and NONO was recently found to be a bivalent chromatin domain factor that regulates mouse ESC pluripotency through Erk signaling, the lack of which stabilizes ESCs in a naive pluripotent state[21]. In contrast, the potential roles of PSPC1 in stem cells are not known. Our study establishes PSPC1 as an important recruiter of the epigenetic regulators TET2 and HDAC1/2 to actively transcribed *MERVL* loci for a dual transcriptional and post-transcriptional repression of MERVL in pluripotent stem cells. Specifically, PSPC1 recruits TET2 for the deposition of 5hmC onto class III ERVL RNAs, thereby leading to their destabilization, while facilitating concomitant recruitment of HDAC activity to repress their transcription in mouse ESCs (Fig. 4e). Although high-resolution mapping of RNA-binding regions in the nuclear proteome of ESCs identified the potential RNA-binding capacity of TET2[46], our study indicates that the majority of TET2 RNA binding is dependent on PSPC1 and its RNA-binding domains (Figs. 2d and 3f, and Supplementary Fig. 5a).

ERVs are well recognized for their contribution to host genome evolution and gene regulatory networks, and their aberrant regulatory activities also are linked to pathological and oncological conditions[9,25,26,47,48]. Pluripotent embryonic cells serve as the 'battleground' for an evolutionary arms race between transposable elements and the host genome during development[9], and reactivation of *MERVL* and its co-opted 2C genes has been correlated with totipotent features in 2C embryos[10,32] and 2C-like cells within pluripotent ESC cultures[10,49]. While transcriptional and epigenetic control of ERVs via DNA (de)methylation and histone modifications are well established, our study suggests that ERV transcripts can also be post-transcriptionally regulated via TET-mediated RNA hydroxymethylation. This is in line with the recognition that multiple silencing mechanisms involving TET enzymes act in concert to control retrotransposon activity in pluripotent cells[50]. In particular, our study provides potential insights into the unique dynamic cyclic fluctuation between totipotent 2C-like cells and pluripotent ESCs in culture that is regulated by PSPC1 and TET2 at the post-transcriptional level via an RNA-hydroxymethylation-mediated mechanism (Fig. 4e). However, our findings on RNA-dependent chromatin targeting of TET2 for ERV control were not universal, but varied depending on ERV classes. This is evident in the fact that the PSPC1–TET2 partnership had a positive effect on the expression of class IIERVK family members *IAP* and *MusD* (Fig. 3c,d) via PSPC1–TET2-mediated transcriptional activation (Supplementary Figs. 7h and 12b,d). Future studies will be needed to dissect such transcriptional activation mechanisms for class II ERVs, as well as to investigate how other TET members (e.g., TET1) may participate in ERV control. In this regard, it is interesting to note that TET1 does not bind *MERVL* genomic loci for transcriptional regulation of MERVL[51] or contribute to 5hmC *MERVL* RNA hydroxymethylation (Supplementary Fig. 14a).

In sum, our study provides a paradigm for post-transcriptional silencing of class III ERV (i.e., *MERVL*) RNAs via 5hmC modification, facilitated by an RNA-binding protein (i.e., PSPC1), which mediates TET2 recruitment. As ERV reactivation has been widely related to aging, cancer and autoimmune diseases[47,48], our findings should also open new avenues for exploring post-transcriptional ERV control by RNA hydroxymethylation in health and disease.

URLs. CRISPR sgRNA design, http://crispr.mit.edu/; DAVID gene functional classification tool, https://david.ncifcrf.gov/; ImageJ, https://imagej.nih.gov/ij/; Piranha CLIP-seq annotation tool,

http://smithlabresearch.org/software/piranha/; Homer Motif Discovery Software, http://homer.ucsd.edu/homer/; HTSeq, https://htseq.readthedocs.io/en/release_0.9.1/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0060-9.

Received: 17 March 2017; Accepted: 16 January 2018;
Published online: 26 February 2018

References

1. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* **30**, 733–750 (2016).
2. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534 (2017).
3. Frye, M. & Blanco, S. Post-transcriptional modifications in development and stem cells. *Development* **143**, 3871–3881 (2016).
4. Masiello, I. & Biggiogera, M. Ultrastructural localization of 5-methylcytosine on DNA and RNA. *Cell. Mol. Life Sci.* **74**, 3057–3064 (2017).
5. Zhang, H. Y., Xiong, J., Qi, B. L., Feng, Y. Q. & Yuan, B. F. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. *Chem. Commun. (Camb.)* **52**, 737–740 (2016).
6. Miao, Z. et al. 5-hydroxymethylcytosine is detected in RNA from mouse brain tissues. *Brain Res.* **1642**, 546–552 (2016).
7. Delatte, B. et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282–285 (2016).
8. Fu, L. et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136**, 11582–11585 (2014).
9. Schlesinger, S. & Goff, S. P. Retroviral transcriptional regulation and embryonic stem cells: war and peace. *Mol. Cell. Biol.* **35**, 770–777 (2015).
10. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
11. Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat. Protoc.* **4**, 506–517 (2009).
12. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).
13. Huang, Y. et al. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **111**, 1361–1366 (2014).
14. Hon, G. C. et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* **56**, 286–297 (2014).
15. Sigova, A. A. et al. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981 (2015).
16. Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
17. Fox, A. H., Bond, C. S. & Lamond, A. I. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol. Biol. Cell* **16**, 5304–5315 (2005).
18. Gilbert, S. L., Pehrson, J. R. & Sharp, P. A. *XIST* RNA associates with specific regions of the inactive X chromatin. *J. Biol. Chem.* **275**, 36491–36494 (2000).
19. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone *O*-GlcNAcylation during gene transcription. *Nature* **493**, 561–564 (2013).
20. Knott, G. J., Bond, C. S. & Fox, A. H. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res.* **44**, 3989–4004 (2016).
21. Ma, C. et al. Nono, a bivalent domain factor, regulates Erk signaling and mouse embryonic stem cell pluripotency. *Cell Rep.* **17**, 997–1007 (2016).
22. Cao, S. et al. Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. *BMC Genomics* **15**, 4 (2014).
23. Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
24. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
25. Friedli, M. & Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* **31**, 429–451 (2015).
26. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 (2016).
27. Leeb, M. et al. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24**, 265–276 (2010).
28. Maksakova, I. A. et al. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6**, 15 (2013).
29. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654 (2003).
30. Macfarlan, T. S. et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
31. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
32. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).
33. Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
34. Kowalska, E. et al. Distinct roles of DBHS family members in the circadian transcriptional feedback loop. *Mol. Cell. Biol.* **32**, 4585–4594 (2012).
35. Li, Z. et al. Deletion of *Tet2* in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509–4518 (2011).
36. Williams, K. et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
37. Huber, S. M. et al. Formation and abundance of 5-hydroxymethylcytosine in RNA. *ChemBioChem* **16**, 752–755 (2015).
38. Amort, T. et al. Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* **18**, 1 (2017).
39. Zhang, X. et al. The tRNA methyltransferase NSun2 stabilizes p16INK^4 mRNA by methylating the 3′-untranslated region of p16. *Nat. Commun.* **3**, 712 (2012).
40. Warren, L. et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).
41. Zhang, Q. et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* **525**, 389–393 (2015).
42. Vella, P. et al. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* **49**, 645–656 (2013).
43. Deplus, R. et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32**, 645–655 (2013).
44. Chen, L. L. & Carmichael, G. G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* **35**, 467–478 (2009).
45. Fox, A. H. & Lamond, A. I. Paraspeckles. *Cold Spring Harb. Perspect. Biol.* **2**, a000687 (2010).
46. He, C. et al. High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* **64**, 416–430 (2016).
47. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967–971 (2012).
48. Li, W. et al. Human endogenous retrovirus-K contributes to motor neuron disease. *Sci. Transl. Med.* **7**, 307ra153 (2015).
49. Choi, Y. J. et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**, eaag1927 (2017).
50. Gerdes, P., Richardson, S. R. & Faulkner, G. J. TET enzymes: double agents in the transposable element-host genome conflict. *Genome Biol.* **17**, 259 (2016).
51. de la Rica, L. et al. TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* **17**, 234 (2016).

Acknowledgements

We thank Y. Kurihara for PSPC1 constructs, T. Macfarlan for the *Zfp352*-luciferase reporter construct, R. Jaenisch for the *Tet*-TKO ESC line, D. Trono for the *Trim28*-cKO line, and D.M. Gilbert and Y. Shinkai for the *Ehmt2*-cKO line. We also thank the medical illustrator J. Gregory from Icahn School of Medicine at Mount Sinai for the model drawing. This research was funded by the US National Institutes of Health (NIH) (grants 1R01-GM095942 and R21HD087722 to J.W.; grant R01HL112294 to M.X.) and the Empire State Stem Cell Fund through the New York State Department of Health (NYSTEM) (grants C028103 and C028121 to J.W.). J.W. is a recipient of an Irma T. Hirschl and Weill-Caulier Trusts Career Scientist Award, and M.F. is a recipient of a Ramón y Cajal contract (RYC-2014-16779) from the Ministerio de Economía y Competitividad of Spain. The research from the Shen laboratory is supported by the

National Natural Science Foundation of China (31471219 and 31428010) and the Center for Life Sciences (CLS) at Tsinghua University. Additional support was provided by the Agencia Estatal de Investigación (BFU2016-80899-P to M.F.) (AEI/FEDER, UE) and the Consellería de Cultura, Educación e Ordenación Universitaria (ED431F 2016/016 to M.F.).

Author contributions

D.G. conceived, designed and conducted the studies. D.G. and M.F. wrote the manuscript with contributions from all other authors. F.P. and M.X. generated the *Tet2*-knockin ESC line. J.D. and F.F. performed the TET2 interactome studies. X.H. conducted computational analysis. J.A.P., C.S., X. Shi, H.Z., P.H., F.Y., D.L., C.S.-P., A.S., M.G.B., L.C., H.W., J.S., and M.F. provided reagents and performed experiments. K.K. conducted embryo microinjections. V.J.V. provided technical advice and helpful discussion. X.B. and X. Shen conducted CLIP-seq experiments and provided helpful discussion. J.W. conceived, designed and supervised the project, and wrote and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0060-9.

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

Correspondence and requests for materials should be addressed to J.W.

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

Methods

Murine embryonic stem cell culture. ESCs were grown under standard culture conditions. Briefly, cells were cultured on 0.1% gelatinized (Gibco, 214340) tissue culture plates in medium containing high-glucose DMEM (Gibco, 11965-092), 15% FBS (Corning, 35-010-CV), 100 µM nonessential amino acids (Gibco, 11140-050), 2 mM L-glutamine (Gibco, 25030-081), 1% nucleoside mix (Sigma; U3003, A4036, C4654, T1895, G6264), 100 U/ml penicillin, 100 µg/ml streptomycin (Gibco, 15140-122), 8 nl/ml 2-mercaptoethanol (Sigma, M6250) and homemade recombinant leukemia inhibitory factor tested for efficient self-renewal maintenance.

Affinity purification of TET2 protein complexes in ESCs. The *Tet2:FLAG* knock-in ESC line was generated with a targeting vector containing a *Neo* cassette flanked by two *FRT* sites, followed by a 0.5-kb genomic fragment upstream of the *Tet2* start codon and an *ATG/3×FLAG/V5* sequence. 2.2-kb 5′ and 4.8-kb 3′ arm genomic fragments were subcloned into the vector for gene targeting. The targeting vector was linearized and electroporated into 129/sv mouse ESCs, and positive clones were screened by Southern blotting.

We used two independent affinity-purification approaches to isolate TET2 protein complexes for mass spectrometry (MS) identification. In the first approach, nuclear extracts from both wild-type (WT) and *Tet2* knock-in ESC lines were prepared as previously described[52]. Briefly, five large square dishes (245 × 245 mm) of each cell line were washed with PBS and scraped, and the cytoplasmic fraction was removed by incubation with buffer A (10 mM HEPES, pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl) supplemented with proteinase inhibitors. Afterward, nuclear pellets were incubated with buffer C (20 mM HEPES, pH 7.6, 25% glycerol (v/v), 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA) supplemented with proteinase inhibitors. Finally, the salt concentration was decreased to 100 mM by dialysis with buffer D (20 mM HEPES, pH 7.6, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol (v/v)) at 4 °C for 3 h, and unwanted precipitated proteins were removed by centrifugation. Freshly made nuclear extracts were pre-cleared with 0.5 ml of Protein G agarose beads (Roche, 11243233001) for 1 h at 4 °C in the presence of 750 units of Benzonase (Fisher Scientific, 502308706) to remove DNA and RNA followed by incubation with 0.5 ml of anti-Flag M2 agarose beads (Sigma, F2426) for 3 h at 4 °C. After five washes in buffer D supplemented with 0.02% NP-40 (v/v), the 3×Flag-tagged TET2 protein complexes were eluted four times for 1 h each time at 4 °C, with 0.3 mg/ml 3×Flag peptide in buffer D supplemented with 0.02% NP-40 (v/v). After concentration, protein complexes were boiled for 5 min with Laemmli sample buffer and separated in a 10% SDS–PAGE gel.

For the second approach, we used SILAC IP-MS for which WT and *Tet2* knock-in ESC lines were cultured in medium labeled by L-arginine-HCL and L-lysine-2HCL (Thermo Scientific, 88427) (light) or by L-arginine-HCL (U-13C6, 99%; U-15N4, 99%) (Cambridge Isotope Laboratories, CNLM-539-H-0.25) and L-lysine-2HCL (U-13C6, 99%; U-15N2, 99%) (Cambridge Isotope Laboratories, CNLM-291-H-0.25) (heavy) amino acids. Nuclear extracts were pre-cleared and immunoprecipitated with anti-Flag M2 agarose beads as described above. Immuno-bound complexes from each cell line were combined in a 1:1 ratio before the last wash after immunoprecipitation and elution with 3×Flag peptide (Sigma, F4799) as described above. Protein complexes were concentrated and boiled with Laemmli sample buffer and separated by SDS–PAGE. In both instances, SDS–PAGE gels were stained with GelCode Blue Safe Protein Stain buffer (Thermo, PI-24594) and subjected to whole-lane LC-MS/MS analysis.

Mass spectrometry data analysis. MS data were processed by Thermo Proteome Discoverer software with SEQUEST engine against the International Protein Index mouse protein sequence database (v.3.68). Protein lists were filtered according to the minimal number of identified peptides (>2). Common contamination proteins (e.g., trypsin, keratins, actin, tubulins) were removed. Duplicated records were removed on the basis of unique protein symbols. Then the protein enrichment ratio was calculated on the basis of the heavy/light ratio (SILAC) or spectrum counts (label-free) of TET2 pulldown versus control pulldown.

Coimmunoprecipitation and western blotting. Nuclear extracts from ESCs (mouse ESC line CCE) prepared as described above were incubated with the corresponding antibodies overnight at 4 °C. A fraction of lysate was kept as input. On the second day, equilibrated Dynabeads G (Life Technologies, 10004D) or anti-MYC agarose affinity gel (Sigma, A7470) was added to each reaction and rotated for 4 h at 4 °C. Bound beads were then washed with immunoprecipitation buffer. Immunoprecipitated proteins were visualized by western blotting with the following primary antibodies: anti-PSPC1 (Santa Cruz, sc-84577), anti-TET2 (Abcam, ab124297), anti-Flag (Sigma, F1804), anti-β-actin (Sigma, A5441) anti-GAPDH (Protein Technologies, 10494-1-AP), anti-OCT4 (Santa Cruz, sc-5279) and anti-histone H3 (Abcam, ab1791). True-blot secondary antibodies were used to reduce the detection of IgG used for immunoprecipitation. Western blot bands were quantified with ImageJ software.

Lentiviral infection for shRNA knockdown. Small hairpin RNAs (shRNAs) for *Pspc1* knockdown were designed, synthesized and subcloned into pLKO.1 vectors (Addgene) expressing a puromycin-resistance gene and an mCherry reporter. Lentivirus production and infection were performed as described[53]. All shRNA sequences are provided in Supplementary Table 4.

RNA extraction and analysis by quantitative PCR. Total RNA from ESCs, induced pluripotent stem cells and mouse embryonic fibroblasts was extracted with the RNeasy kit (Qiagen, 74136) and converted to cDNA with qSCRIPT (Quanta, 95048). Gene expression was analyzed with Lightcycler 480 SYBR Green master mix (Roche, 4729749001) on the LightCycler480 real-time PCR system (Roche).

We dissected one adult male and one adult female mouse to isolate organs of interest. Thirty milligrams of each tissue were disaggregated with QIAshredder columns (Qiagen, 79656). Total RNA was extracted with an RNeasy kit and subjected to reverse transcription and qPCR quantitation according to the manufacturer's instructions.

RNA from 1.5-, 2.5- and 4.5-dpc (days post-coitum) embryos injected with siRNAs was extracted with TRIzol (Thermo Fisher, 15596026) according to the manufacturer's instructions. Total purified RNA was subjected to reverse transcription as described above, and expression was quantified by qPCR.

Primers used in this study are shown in Supplementary Table 4. In all cases, average threshold cycles were determined from triplicate reactions, and the levels of gene expression were normalized to those of a housekeeping gene as indicated (*Hprt*, *Rnu6* or *Actb*). Error bars in figures indicate s.e.m. or ranges of fold change relative to the reference sample, as indicated in the legends.

Preparation of whole-cell extracts and chromatin-bound protein fractions. We obtained whole-cell protein extracts by lysing cell pellets in RIPA buffer (60 mM Tris-HCl, pH 6.8, 2% SDS (v/v), 10% glycerol (v/v) and 10 mM DTT) supplemented with PMSF and proteinase inhibitors. Chromatin-bound protein fractions were prepared with a chromatin extraction kit (Thermo Scientific, PI-78840) according to the manufacturer's instructions.

Immunofluorescence. Cells were grown on 24-well plates coated with 0.1% gelatin (w/v). After fixation with 4% paraformaldehyde (w/v) for 15 min at room temperature (RT), cells were permeabilized with 0.25% Triton X-100 (v/v) in PBS for 5 min at RT and blocked with 10% BSA (AMRESCO) for 30 min at 37 °C. For immunostaining, cells were incubated overnight at 4 °C with primary antibodies anti-SOX2 (Santa Cruz, sc-17320), anti-PSPC1 and anti-OCT4 in PBS with 3% BSA (w/v). The next day cells were incubated with fluorophore-labeled secondary antibodies for 1 h at RT. Cells were imaged with a Leica DMI 6000 inverted microscope at 20× magnification.

Cell cycle analysis and flow cytometry. For cell cycle analysis, equal numbers of cells were washed with DPBS (Dulbecco's phosphate-buffered saline), permeabilized with 0.1% Triton X-100 (v/v) in DPBS, stained with 10 µM 4′-6-diamidino-2-phenylindole (DAPI) at RT for 10 min, and analyzed by flow cytometry on an LSRII flow cytometer system (BD Biosciences). Analysis was performed in FlowJo software with the Dean–Jett–Fox cell cycle model.

CRISPR–Cas9 generation of *Pspc1*-knockout ESC line. *Pspc1*-KO mouse ESCs were generated with the CRISPR editing tool as described[54]. Briefly, an sgRNA (Supplementary Table 4) was designed to target the transcription start site of *Pspc1*, using the guidelines described in MIT's online tool ("URLs"), and cloned into the pX330 vector (Addgene #42230) modified to have a GFP reporter gene. ESCs were transfected with the plasmid containing the sgRNA and GFP reporter, and GFP^+ ESCs were sorted 48 h after transfection and seeded at clonal density. One week later, clones were picked and analyzed for PSPC1 expression by western blotting, and the *Pspc1* CRISPR–Cas9-targeted genomic region was PCR-amplified and sequenced in *Pspc1*-KO clones.

CRISPR activation of *MERVL*. CRISPR activation of *MERVL* expression was achieved by the Synergistic Activation Mediator (SAM)[31] using a three-vector system: dCAS-VP64-Blast (Addgene #61425), MS2-P65-HSF1-Hygro (Addgene #61426) and MS2-sgRNA-Zeo (Addgene #61427), where the non-targeting or *MERVL*-targeting (+317F4 from gag ATG) sgRNAs were cloned using *Bsm*BI restriction sites. Lentiviruses containing each of the plasmids were generated, and ESCs were infected with a 1:1:1 mix of the viruses in the presence of 8 µg/ml Polybrene. Infected cells were selected for 48 h with 10 µg/ml blasticidin, 250 µg/ml hygromycin and 250 µg/ml Zeocin. Selected cells were collected for total RNA isolation and qPCR analysis.

Chromatin immunoprecipitation coupled with qPCR. ChIP assays were performed as previously described[55]. Briefly, we cross-linked cells with 1% (w/v) formaldehyde for 10 min at RT, and then added 125 mM glycine to stop the reaction. Chromatin extracts were sonicated into 200–500 bp and immunoprecipitated with anti-TET2, anti-Flag, anti-PSPC1, anti-HDAC1 (Bethyl, A300-713A), anti-HDAC2 (Bethyl, A300-705A), anti-H3K9me2 (Abcam, ab1220) or IgG (Millipore, PP64). The immunoprecipitated DNA was purified with ChIP DNA Clean & Concentrator columns (Zymo Research) and analyzed by qPCR with Roche SYBR Green reagents and a LightCycler480 machine. Primer sequences are listed in Supplementary Table 4. Percentages of input recovery were calculated.

In vitro RNA-binding assay. Whole-cell extracts from WT, *Pspc1*-KO or rescued cells were prepared as previously described. Two milligrams of protein extracts were incubated with 2 µg of anti-TET2 or anti-PSPC1 in the presence of RNase A and DNase I nucleases, to avoid contamination of endogenous RNA or DNA. TET2 or PSPC1 protein complexes were recovered by incubation with 20 µl of Protein G magnetic Dynabeads (Invitrogen) for 4 h at 4 °C. After washing, purified protein complexes were incubated with total RNA for 30 min at RT. Total RNA was obtained from mouse ESCs by TRIzol (Invitrogen) extraction and purification according to the manufacturer's instructions, followed by Proteinase K and DNase I treatment. Beads containing protein–RNA complexes were then washed and eluted in TRIzol, and immunoprecipitated RNA was purified. RNA was treated with or without RNase A (Thermo Fisher, EN0531) and quantified with the Qubit High Sensitive assay kit (Life Technologies, Q32852). Immunoprecipitated RNA was visualized in an ethidium bromide agarose gel and retrotranscribed with qSCRIPT (Quanta, 95048) for analysis by qPCR.

RNA immunoprecipitation of cross-linked cells. Cross-linked nuclear extracts were prepared in the same way as ChIP extracts, described above. After sonication, nuclear extracts were incubated with 1.5 µg of the corresponding antibody (IgG, anti-PSPC1 or anti-TET2) pre-bound to 25 µl of Dynabeads Protein G (Thermo Fisher, 10004D) in the presence of proteinase and RNase inhibitors (Thermo Scientific, AM2694 and 10777019) overnight with rotation at 4 °C. After washing, immuno-complexes were eluted with 100 µl of elution buffer (10 mM Tris-HCl, pH 8.0, with 1 mM EDTA) and RNA was extracted with TRIzol. Immunoprecipitated RNA was treated with DNase I and with or without RNase A, after which it was subjected to phenol:chloroform extraction. The resulting RNA was retrotranscribed with qScript, and cDNA was visualized in a polyacrylamide gel by silver staining (Thermo Scientific, 24600) according to the manufacturer's instructions.

UV-cross-linking and RNA immunoprecipitation coupled with qPCR. Cells were trypsinized and UV-cross-linked according to previously published protocols[56]. Briefly, cells were irradiated at 400 mJ/cm^2 in a CL-1000 UVP UV-cross-linker and then subjected to cell lysis by incubation with nuclear suspension buffer (248 mM sucrose, 8 mM Tris-HCl, pH 7.5, 4 mM $MgCl_2$, 0.1 mM DTT, 0.8% Triton X-100 (v/v)) in the presence of protease and RNase inhibitors. Nuclear pellets were obtained by centrifugation, and nuclear content was released by incubation in RIP buffer (150 mM KCl, 25 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40 (v/v)) in the presence of protease and RNase inhibitors, and subjected to brief sonication to help break the nuclear envelope. Immunoprecipitation of PSPC1 or TET2 and purification of their target RNAs were performed as described above.

α-Amanitin and valproic acid treatments. Cells in culture were treated with 10 µg/ml α-amanitin (Santa Cruz, sc-202440), an inhibitor of RNA Pol II and Pol III, or with Milli-Q water control (vehicle) for 0, 1, 2 or 4 h. For HDAC inhibition, cells were cultured in the presence of 0.5 µM class I HDAC inhibitor valproic acid (Stemgent, 04-0007) or DMSO control (vehicle) for 24 h. For both experiments, cells were collected and processed for western blotting or qPCR analysis.

RNA stability assay by nascent RNA capture. To monitor RNA degradation, we treated ESCs with 0.2 mM 5-ethynyl uridine (EU) in growth medium for 16 h. Total RNA was extracted at 0 h and 8 h after the removal of EU from culture medium. EU-labeled RNAs were biotinylated and captured with the Click-iT nascent RNA capture kit (Invitrogen, C10365) according to the manufacturer's instructions. RNA was reverse-transcribed with SUPERSCRIPT IV VILO (Invitrogen, 11756050) and quantified by qPCR.

Cross-linking immunoprecipitation and massive parallel sequencing. UV cross-linking and immunoprecipitation were performed as described[57], with some modifications. Briefly, J1 mouse ESCs expressing a 3×Flag–biotin-tagged PSPC1 construct were cross-linked in PBS with UV type C (254 nm) at 600 mJ/cm^2 on ice. Cells were harvested, pelleted, and lysed in PXL lysis buffer (1× PBS, 0.1% SDS (w/v), 0.5% NP-40 (v/v) and 0.5% sodium deoxycholate (w/v)) supplemented with proteinase and RNase inhibitors and RQ1 DNase (Promega, M6101). After 30 min of incubation on ice, cells were centrifuged, and the supernatant was carefully collected. For immunoprecipitation, the supernatant was incubated with beads prebound with 4 µg of anti-Flag conjugated with 30 µl of Protein G Dynabeads overnight at 4 °C. After immunoprecipitation, the sample was washed with PXL lysis buffer twice and then with high-salt buffer twice (5× PBS, 0.1% SDS (w/v), 0.5% NP-40 (v/v), 0.5% sodium deoxycholate (w/v)). Then the protein–RNA complex was subjected to MNase digestion (New England BioLabs (NEB), M0247). To dephosphorylate RNA, we incubated each immunoprecipitated sample in 80 µl of 1× reaction mixture including 3 µl of CIP (NEB, M0290S) for 10 min at 37 °C. After the CIP treatment, the immunoprecipitates were washed twice with PNK + EGTA buffer (50 mM Tris-HCl, pH 7.4, 20 mM EGTA, and 0.5% NP-40 (v/v)) and twice with PNK buffer (50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, and 0.5% NP-40 (v/v)). To ligate the 3′ linker, we incubated each of the washed immunoprecipitates in 40 µl of 1× ligation reaction mixture with 80 pmol of 3′ linker (Takara; Supplementary Table 4) and 3 µl of truncated T4 RNA ligase 2 (NEB, M0242) overnight at 16 °C. After the reaction, immunoprecipitates were washed twice with PXL buffer and twice with PNK buffer. Then we phosphorylated the RNAs by incubating the immunoprecipitates in 1× reaction mixture containing 2 µl of T4 PNK (NEB, M0201) and 1 µl of hot ATP (PerkinElmer) for 5 min at 37 °C to label the RNA–protein complexes. After the labeling, 5 µl of 2 mM ATP was added and the mixture was incubated for another 5 min at 37 °C. After PNK treatment, immunoprecipitates were washed four times with PNK buffer and mixed with 2× LDS loading dye (Invitrogen, NP0007). The samples were incubated at 70 °C for 10 min to elute RNA–protein complexes from the beads. Immunoprecipitated RNA was loaded onto a NuPage SDS gel and transferred onto a nitrocellulose membrane.

For RNA isolation, nitrocellulose membranes were fragmented with a clean scalpel blade and treated with 4 mg/ml proteinase K (Takara, 9034) in 200 µl of PK buffer (100 mM Tris-HCl, pH 7.4, 50 mM NaCl, 10 mM EDTA) for 20 min at 37 °C, and incubated in 200 µl of PK + urea buffer (100 mM Tris-HCl, pH 7.4, 50 mM NaCl, 10 mM EDTA, and 7 M urea) for another 20 min at 37 °C. RNA was extracted with TRIzol and ligated with 5′ RNA linkers (Supplementary Table 4) for 16 h at 16 °C. We carried out reverse transcription with RT primer followed by PCR with forward SR and reverse index primers (all are listed in Supplementary Table 4) for 25 cycles to generate and amplify the cDNA library, respectively. High-throughput sequencing of the resulting cDNA was performed on a HiSeq-2000 sequencer.

CLIP-seq analysis. CLIP-seq reads were aligned to the mouse genome (mm9) using TopHat (v2.0.10) and Bowtie2 (v2.1.0) with the default parameter settings. CLIP-seq peaks were determined with the Piranha tools ("URLs") with the parameters -s -b 200 and annotated with the "annotatePeaks" module in HOMER (v4.6) against the mm9 mouse genome.

Motif finding in CLIP-seq. CLIP-seq peaks were used as input for de novo motif discovery in HOMER with the parameters "-rna -len 6" and the mm9 reference genome.

RNA stability reporter assay. To generate the mRNA stability minigene construct, we cloned the top five consensus motifs of PSPC1 CLIP-seq upstream of the EGFP coding sequence in the pd2EGFP-N1 vector (BD Biosciences). This vector expresses a destabilized version of EGFP protein with a half-life of 2 h, which is useful for studies that require rapid reporter turnover. To design the sequence surrounding the motifs, we specifically chose sequences enriched in the *MERVL* RNA sequence with CG regions (Supplementary Table 4). We designed the corresponding mutated motifs by replacing cytosines with adenines in each motif. All the motifs were cloned with *Xho*I and *Bam*HI sites into the reporter vector. Each vector was transfected with JetPrime polyplus reagent into ESCs in 48 wells. The next day, we added 10 µg/ml of α-amanitin (or vehicle (Milli-Q)) to the medium to inhibit transcription. Cells were analyzed by flow cytometry with an Accuri C6 instrument (BD Biosciences), and data were analyzed with FlowJo software (Treestar). The mean fluorescence intensity of EGFP, an indicator of RNA stability, was gated on the GFP$^+$ singlet population.

ChIP-seq, methylated DNA immunoprecipitation sequencing and hydroxymethylated DNA immunoprecipitation sequencing analysis. External data for ChIP-seq, methylated DNA immunoprecipitation sequencing (MeDIP-seq) and hydroxymethylated DNA immunoprecipitation sequencing (hMeDIP-seq) analysis were downloaded from GEO (5mC, GSM611203; 5hmC, GSM611199; TET2, GSM1023124). Reads were aligned to the mouse genome (NCBI build 37, mm9) using Bowtie (v1.0.0), with parameters -M 1 --best --chunkmbs 200. The duplicated reads of the aligned data were removed, then filtered reads were sorted with SAMtools (v0.1.19). For LTR sequence annotation, analysis was performed as described[58]. Briefly, RepeatMasker track (RMSK) from UCSC Genome Browser was used, and ChIP-seq intensity at each LTR region was measured by HTseq software (v0.6.1) with parameters -a 10 -m intersection-nonempty. The ChIP-seq intensity of TET2 and 5hmC at each LTR region was normalized by total mapped reads as reads per million mapped reads (RPM).

hMeDIP-qPCR. Genomic DNA from wild-type ESCs was isolated according to the manufacturer's instructions (Qiagen, 158388). Four micrograms of gDNA were denatured by boiling for 10 min in water and immunoprecipitated with 2.5 µl of anti-5hmC in immunoprecipitation buffer (100 mM sodium phosphate, pH 7.0, 1.4 M NaCl and 0.5% Triton X-100 (v/v)). After 6 h of incubation at 4 °C, the immunobound DNA was recovered by the addition of 20 µl of pre-equilibrated Dynabeads G and rotated at 4 °C overnight. The next day, beads were recovered and washed with IP buffer, antibody was removed by digestion with Proteinase K, and DNA was extracted with phenol:chlorophorm:isoamylalcohol and ethanol precipitation. Immunoprecipitated DNA was analyzed by qPCR.

RNA-seq of PSPC1-depleted ESCs. Mouse ESCs were infected with pLKO.Puro-IRES-mCherry constructs carrying shRNAs targeting *Pspc1* or control shRNAs. Biological duplicates were prepared for RNA-seq analysis, according to our previously described protocol[59]. Briefly, total RNA from each sample was extracted

with an RNeasy kit, and paired-end sequencing was carried out with the Illumina HiSeq-2500 according to a RiboZero selection protocol per the manufacturer's instructions. Reads were aligned to the mouse genome (NCBI build 37, mm9) using TopHat (v2.0.10) and Bowtie2 (v2.1.0) with the default parameter settings. Transcript assembly and differential expression analysis were done with Cufflinks (v2.1.1). Assembly of novel transcripts was allowed (-g); other parameters followed the default settings. The summed FPKM (fragments per kilobase per million mapped reads) of transcripts sharing each gene_id was calculated, and the significance of differential expression tests was estimated via a genome-wide false discovery rate after Benjamini–Hochberg correction for multiple testing.

For the LTR regions, a reference genome with all LTRs was created based on the RMSK database. RNA-seq intensity at each LTR region was counted by HTseq software (v0.6.1) with parameters -a 10 -m intersection-nonempty, and normalized to total mapped reads (RPM). LTR expression in shEV (empty vector) and sh*Pspc1* samples was compared, and *P* values were calculated by Student's *t*-test.

The PSPC1 CLIP-seq dataset was used to calculate the distance between PSPC1-bound sites and PSPC1-regulated genes. Briefly, CLIP-seq intensity at each LTR region was determined by HTseq and normalized as an RPM value. PSPC1-bound (intensity > 0.5 RPM) and PSPC1-unbound (intensity = 0) sites were collected. A subset of PSPC1-unbound sites was randomly selected from the PSPC1-unbound pool, with the same number of PSPC1-bound sites ($n = 14{,}220$). Then the distribution of distances between PSPC1-upregulated genes by RNA-seq and nearest PSPC1-bound or -unbound sites was calculated and plotted. The significance of the number of nearest PSPC1-bound versus -unbound sites (<50 kb) was calculated by binomial test.

RNA-seq analysis of previously published datasets. External RNA-seq data for *Trim28*–conditional knockout (cKO) (GSE41903) and *Ehmt2*-cKO (GSE33923) cells were analyzed for ERV expression as described above.

Luciferase reporter assay. Twenty thousand ESCs were transfected with 0.32 µg of luciferase reporter plasmids containing genomic promoter fragments from the *Zfp352* gene including the PSPC1-regulated LTR element[30] and 16 ng of *Renilla* control plasmid. The same promoter without the LTR was used as a negative control. Twenty-four hours after transfection, cells were lysed and luciferase and *Renilla* activity were assayed with the DualGlo luciferase assay kit (Promega, E2920) according to the manufacturer's instructions in a PerkinElmer EnSpire Alpha Luminometer. The luciferase/*Renilla* ratio was calculated for all the samples. Measurements were carried out in triplicate biological samples.

Mouse embryo collection and microinjection. C57BL/6JxDBA/2J (B6D2) female mice were superovulated by intraperitoneal injection of 5 IU pregnant mare's serum gonadotropin (National Hormone and Peptide Program) followed by intraperitoneal injection of 5 IU human chorionic gonadotropin (National Hormone and Peptide Program) 48 h later. The females were mated with males overnight, and one-cell embryos were collected from pregnant females in HEPES-buffered FHM media (Millipore, MR-024-D). Cumulus cells were removed by brief treatment with hyaluronidase (Millipore, MR-056-F) in FHM media. Isolated fertilized eggs (as judged by the presence of two pronuclei) were microinjected in the cytoplasm with 5–10 picoliters of 20 µM nontargeting siRNA (GE-Dharmacon #D-001910-01-05), *Pspc1* siRNA (GE-Dharmacon #E-049216-00-0005) or *Tet2* siRNA (GE-Dharmacon #E-058965-00-0005) and cultured in bicarbonate-buffered KSOM media (Millipore, MR-121-D) at 37 °C with 5% CO_2. We used a Nikon Diaphot inverted microscope equipped with a Narashige micromanipulator system for the microinjections. After injection, embryos were inspected every day for assessment of developmental progress. Total RNA was extracted for qPCR analysis and processed as described before. For each knockdown we processed around 100 injected embryos, with a 50% rate of survival. All mouse procedures were done in accordance with Mount Sinai IACUC policy. Zygote injections were done in the Mouse Genetic Shared Research Facility at Mount Sinai.

Methylated and hydroxymethylated RNA immunoprecipitation. Total RNA was sonicated to an average size of 500 bp and then subjected to immunoprecipitation with either anti-5mC (Sigma, 60612) or anti-5hmC (Active Motif, 39769) or with an IgG control antibody (Millipore, PP64) based on a previously described protocol[60]. Briefly, 6–10 µg of sonicated RNA was incubated with 2 µg of antibody at 4 °C overnight. The next day, 5hmC-modified RNAs were purified by incubation with 20 µl of Dynabead Protein G beads. After being washed to reduce nonspecific background, bound RNA was eluted with TRIzol (Invitrogen) and extracted according to the manufacturer's instructions. Finally, immunoprecipitated RNAs were subjected to reverse transcription and qPCR quantitation.

RNA electrophoretic mobility shift assay. The RNA probe was synthesized by standard solid-phase synthesis using Oligo-800 DNA synthesizer with the sequence 5′-biotin-CCUCUGCCUXCCGAAUCCAA-3′ (where X is 5mC or 5hmC). Both 5mC and 5hmC phosphoramidite building blocks were purchased from ChemGenes. The RNA oligonucleotides were deprotected by AMA (1:1 mixture of ammonium hydroxide and methyl amine solution) and Et3N•3HF treatment, followed by ion-exchange HPLC purification on a Dionex PA-200 column. HEK293T cells were transfected with plasmids encoding $^{3\times FL}$TET2 and ^{Myc}PSPC1, and total protein extracts were incubated with anti-Flag to immunopurify PSPC1–TET2 complexes. The complex was competitively eluted by 3×Flag peptide. Increasing amounts of PSPC1–TET2 protein complex (0, 0.5, 1 and 2 µg of total protein) were incubated with 250 ng of the corresponding RNA probe in binding buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.4 mM EDTA, 0.1% NP-40 (v/v), 40 U/ml RNasin, 1 mM DTT, 50% glycerol (v/v), 5 ng/µl BSA) for 30 min at RT. Then, 1 µl of glutaraldehyde (0.2% final concentration) was added into the mixture, which was incubated at RT for 15 min. The total protein–RNA mixture was loaded onto a 6% TBE acrylamide gel and run for 30 min at 80 V on ice. The gel was transferred onto hybond-N^+ membrane (GE Healthcare, 95038-362) in 1× TBE buffer and nucleic acids were detected by the chemiluminescent nucleic acid detection module (Thermo Fisher, 89880) according to the manufacturer's instructions. Quantification of each band was carried out with ImageJ software, and the percentage of bound RNA probe was calculated as (Intensity of bound probe)/((Intensity of bound probe) + (Intensity of free probe)) × 100.

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

Data availability. RNA-seq (GSE103267) and CLIP-seq (GSE103268) datasets generated in this study have been deposited in the Gene Expression Omnibus (GEO).

References

52. Ding, J., Xu, H., Faiola, F., Ma'ayan, A. & Wang, J. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Res.* **22**, 155–167 (2012).
53. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).
54. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
55. Lee, T. I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
56. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
57. Xue, Y. et al. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* **36**, 996–1006 (2009).
58. Elsässer, S. J., Noh, K. M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**, 240–244 (2015).
59. Ding, J. et al. Tex10 coordinates epigenetic control of super-enhancer activity in pluripotency and reprogramming. *Cell Stem Cell* **16**, 653–668 (2015).
60. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).

nature research

Corresponding author(s): Jianlong Wang

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. *The entry may remain private before publication.*	N/A
3. Provide a list of all files available in the database submission.	GSE103268: GSM2759483 CLIPseq_Pspc1
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).	N/A

Methodological details

5. Describe the experimental replicates.	One experimental replicate was analyzed.
6. Describe the sequencing depth for each experiment.	Total number of reads after quality filtering: 40,584,236; unique mapping; ratio: 44.71%; multiple alignment ratio (>=2): 39.15%; lengthof read: 50bp; single/paired-end reads: single-end
7. Describe the antibodies used for the ChIP-seq experiments.	Anti-FLAG antibody, detailed in Methods.
8. Describe the peak calling parameters.	Piranha -s -b 200
9. Describe the methods used to ensure data quality.	There are 15,291 clip-seq clusters were identified by Piranha, with the setting of background threshold 0.99, means 99% of the lowest scores reads are considered the background.
10. Describe the software used to collect and analyze the ChIP-seq data.	CLIP-seq reads were trimmed by the 3' adaptor sequence: AGATCGGAAGAGCACACGTCT and cleaned by Trim Galore. CLIP-seq reads were aligned to the mouse genome (mm9) using TopHat (v2.0.10) and Bowtie2 (v2.1.0) with the default parameter settings.

nature research

Corresponding author(s): Jianlong Wang

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	ESCs were trypsinized and suspended in DPBS buffer. For Fig. S3b, cells were permeabilized and stained with DAPI.
6. Identify the instrument used for data collection.	Fig. S3b: LSRII Flow Cytometer System (BD Biosciences). Fig. S14f: Accuri C6 (BD Biosciences).
7. Describe the software used to collect and analyze the flow cytometry data.	We run our experiments with the corresponding BD software and analyze the data with FlowJo.
8. Describe the abundance of the relevant cell populations within post-sort fractions.	N/A
9. Describe the gating strategy used.	Alive GFP+ singlet events were analyzed.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☐

nature research

Corresponding author(s): Jianlong Wang

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   All experiments were performed at least in duplicate and analyzed in at least technical triplicate. No statistical methods were used to predetermine sample size.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded from analysis

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   All attempts of replication were successful

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Samples were not randomized for the experiments.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Investigators were not blinded during the experiments.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

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

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

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

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Excel and GraphPrism (v7) softwares were used to analyze the statistics. Image J (v1.50i) was used to quantify the blot bands in this study. FlowJo (v7.6) was used to analyze the cytometry data. Thermo Proteome Discoverer was employed for identifying and quantifying proteins in our MS studies, Piranha (v1.2.0) was used for the peak calling of CLIP-seq reads; Homer (v4.6) was used for motif discovery. HTseq software (v0.6.1) was used for mapping RNA/ChIP-seq intensities at LTR regions.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All materials used here are commercial, except for the cell lines that we generated or that were kind gifts from other researchers as acknowledged.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Primary antibodies used in this study: anti-PSPC1 (Santa Cruz sc-84577), anti-TET2 (Abcam ab124297), anti-FLAG tag (Sigma F1804), anti-β-ACTIN (Sigma A5441) anti-GAPDH (Protein Technologies 10494-1-AP), anti-OCT4 (Santa Cruz sc-5279), anti-SOX2 (Santa Cruz #sc-17320), anti-HDAC1 (Bethyl #A300-713A), anti-HDAC2 (Bethyl #A300-705A), anti-H3K9me2 (Abcam #ab1220), IgG (Millipore #PP64), anti-5mC (Sigma # 60612), anti-5hmC (Active Motif #39769), and anti-Histone H3 (Abcam ab1791). Species validation of all primary antibodies used in this study can be found in the corresponding manufacturer´s websites.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Commercially available mouse CCE ESC line (129/Sv background) , mouse J1 ESC (129S4/SvJae) and human HEK293T cells were used in this study. Tet-TKO (129/C57BL/6), Tet2-KO (129/sv/C57BL/6), Kap1 cKO and G9a cKO lines were kind gifts from other researchers (see Acknowledgements section).
b. Describe the method of cell line authentication used.	Tet2KI ES cells were analyzed by Southern blot and RT-qPCR. PSPC1 KO ES cells were confirmed/authenticated by RT-qPCR, Western blot and sequencing of the genomic sequence targeted by CRISPR-Cas9. Tet-TKO, Tet2-KO, Kap1 cKO and G9a cKO ES cells were confirmed and authenicated by RT-qPCR.
c. Report whether the cell lines were tested for mycoplasma contamination.	All cell lines were tested for mycoplasm contamination (free).
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used are listed in the ICLAC database.

Animals and human research participants

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

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	C57Bl/6JxDBA/2J (B6D2) female mice were superovulated and mated overnight with males. One-cell embryos were collected for in vitro culture until blastocyst stage. The total number of embryos processed for each knock-down consisted of around 100 injected embryos with a 50% rate of survival. All mouse procedures were performed in accordance with Mount Sinai IACUC policy.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	N/A