HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 August 26.

Published in final edited form as:

Nat Genet. 2018 March ; 50(3): 443–451. doi: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,#], **Jose Angel Pardavila**[2,#], **Xin Huang**[1,#], **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,*]

[1]The Black Family Stem Cell Institute and Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, USA

[2]CiMUS, Universidade de Santiago de Compostela-Health Research Institute (IDIS), Santiago de Compostela, 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, USA

[6]Department of Animal Biotechnology, College of Veterinary Medicine, Northwest A&F University, Yangling, Shaanxi, China

[7]The Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, USA

[8]Sylvester Comprehensive Cancer Center, Department of Biochemistry and Molecular Biology, University of Miami, Miami, USA

[*]Corresponding author: Jianlong Wang, PhD., Icahn School of Medicine at Mount Sinai, Black Family Stem Cell Institute, Dept. of Cell, Developmental and Regenerative Biology, Atran Building, AB7-10D, 1428 Madison Ave, New York, NY10029, jianlong.wang@mssm.edu, Tel: 1-212-241-7425.

[#]These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

D.G. conceived, designed and conducted the studies. D.G. and M.F. wrote the manuscript with contributions from all authors. F.P. and M.X. generated the *Tet2* knock-in ESC line, J.D. and F.F. performed the TET2 interactomes, X.H. conducted computational analysis, J.A.P., C.S., X.S., 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.S. conducted CLIP-seq experiments and provided helpful discussion. J.W. conceived, designed, supervised the project, wrote and approved the final manuscript.

DECLARATION OF COMPETING INTERESTS

The authors declare no competing financial interests.

[9]Department of Cell Biology and Development, Instituto de Fisiologia Celular, UNAM, Mexico City, Mexico

Abstract

Ten-eleven translocation (TET) proteins play key roles in regulating the methylation status of DNA through oxidizing methylcytosines (5mC), generating 5-hydroxymethylcytosines (5hmC) that can both serve as stable epigenetic marks and participate in active demethylation. Unlike the other TET-family members, 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 find that PSPC1 and TET2 contribute to ERVL and ERVL-associated gene regulation by both transcriptional repression via histone deacetylases and posttranscriptional 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.

Ten-eleven translocation (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 revealed the contribution of a number of sequence-specific DNA binding transcription factors including NANOG, PRDM14, PU.1, and WT1 (reviewed by Wu and Zhang[2]) to 5-hydroxymethyl cytosine (5hmC) deposition at the genome, leading to active demethylation of target genes. While 5mC modification of RNA is firmly established (reviewed by Frye and Blanco[3]), the potential roles of TET proteins in mediating 5mC to 5hmC oxidation in RNA are just begun 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 transcription of most members of endogenous retroviruses (ERVs)[9] but fluctuate with MERVL activity in the 2-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 Paraspeckle component 1 (PSPC1) as a binding partner of TET2. We showed 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 posttranscriptional regulation of *MERVL* through HDAC1/2-mediated repression and RNA hydroxymethylation (5hmC)-mediated degradation.

RESULTS

TET2 interaction with PSPC1 is required for its recruitment to chromatin

In search of factors that may regulate TET2 chromatin binding, we investigated the TET2 interactome in ESCs. To this end, we performed affinity purification (AP) of TET2-containing protein complexes from a 3xFLAG-tagged *Tet2* knock-in ESC line (Supplementary Fig. 1, a–c) coupled with mass spectrometry analysis (AP-MS), 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 (IP) and co-immunoprecipitation (coIP) (Fig. 1c), and was not compromised by the absence of other TET2-interacting partners such as OGT, SIN3A or NONO (Fig. 1a and Supplementary Fig. 2, b and c). PSPC1 displays a similar gene expression pattern to TET2 across multiple tissues, including a much higher enrichment in pluripotent cells than in somatic mouse embryonic fibroblasts (Fig. 1b and Supplementary Fig. 2, d and e).

In order to test the impact of PSPC1 loss on TET2 chromatin occupancy, we analyzed TET2 levels in pure chromatin extracts, revealing a reduced occupancy (< 30%) upon PSPC1 knock-down (Fig. 1d and Supplementary Fig. 3a). Notably, knock-down of PSPC1 did not dramatically affect ESC properties (Supplementary Fig. 3, b–d), similar to 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 remains at the chromatin upon PSPC1 depletion (Fig. 1d and 2b).

PSPC1 recruits TET2 to chromatin through RNA

Given that 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 PSPC1 RNA-dependent from independent functions in TET2 recruitment, we employed the CRISPR-Cas9 nuclease system to generate a *Pspc1* knock-out (KO) ESC line (Supplementary Fig. 4, a and b), and then rescued PSPC1 loss with Piggybac vectors expressing either wild-type PSPC1 (PSPC1WT), or PSPC1 bearing four mutations in its RNA binding domains[17] (F118A/F120A/K197A/F199A, hereafter PSPC1Mut) (Fig. 2a and Supplementary Fig. 4, c and d). We found that, while 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 occupancy were largely dependent on intact RNA binding domains of PSPC1.

We next explored whether PSPC1 can mediate TET2 interaction with RNA. 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 is independent of TET2 (Supplementary Fig. 5c).

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

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 performed cross-linking followed by FLAG immunoprecipitation and high-throughput sequencing (CLIP-seq) to identify PSPC1-associated RNA species in ESCs over-expressing 3xFLAG- and biotin-tagged-PSPC1 (^{3xFLBio}PSPC1) (Fig. 3a, left panel and Supplementary Table 2). Supporting a functional connection between PSPC1 and TET2, an *in silico* analysis of the available TET2 ChIP-seq data set in ESCs[19] showed high correlation between TET2 occupancy at DNA regions with PSPC1-bound RNA peaks (Fig. 3a, right panel), thereby pointing to a co-transcriptional recruitment of TET2 to those loci through PSPC1-RNA interactions. We validated this observation by chromatin immunoprecipitation coupled with PCR (ChIP-qPCR) analysis, demonstrating that TET2 chromatin occupancy at representative loci (e.g., *Adss* and *Ywhae*), which transcribe RNAs that are bound by PSPC1, is compromised by the loss of PSPC1 RNA-binding capacity (Supplementary Fig. 6a).

PSPC1 regulates endogenous retrovirus expression

Given that PSPC1 recruitment to chromatin is stabilized by its binding 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 sequencing (RNA-seq). Although PSPC1 loss-of-function led to up-regulation and down-regulation of 362 and 610 coding genes, respectively (Supplementary Fig. 6b and Supplementary Table 3), only approximately 18% of these genes deregulated after PSPC1 depletion were bound at their RNAs by PSPC1 (Supplementary Fig. 6c). Gene ontology (GO) analysis revealed that deregulated genes upon PSPC1 knock-down are involved in developmental processes (Supplementary Fig. 6d). Compared to a published RNA-seq data set of early development of mouse embryos[22], we observed a specific enrichment of deregulated targets in the two-cell stage embryos (Supplementary Fig. 6e), wherein a remarkable co-regulation of host stage-specific genes and retrotransposable elements (REs) is prevalent[23].

REs are vestiges of ancient retroviral infections which comprise 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[25,26]). Given that PSPC1 binds both coding and non-coding RNAs (Fig. 3b), and that only a minority of coding genes with

transcribed RNAs bound by PSPC1 were affected upon PSPC1 knock-down (Supplementary Fig. 6, b and c), we hypothesized that PSPC1 might regulate gene expression through binding to REs that in turn regulate neighboring genes. Consistent with this hypothesis, further examination of CLIP-seq data revealed that, among PSPC1-bound RNAs other than coding ones, a vast majority arose from intergenic regions (85%) (Fig. 3b), including long terminal repeat (LTR)-containing (ERVs) and non-LTR (*SINEs* and *LINEs*) subclasses of REs (Supplementary Fig. 7a). The interaction between PSPC1 and RE-derived RNAs was further confirmed by RNA CLIP-qPCR experiments (Supplementary Fig. 7, b, d and e). It is well known that ESC potency fluctuates with endogenous retrovirus (ERV) activity[10] and that ERV expression is transcriptionally regulated by multiple epigenetic pathways[27,28]. We discovered abundant representation of ERVK, ERVL, and MaLR ERV families among PSPC1-bound RNAs whose binding is independent of their relative abundance (Fig. 3c and Supplementary Fig. 7, c–e), and observed a global deregulation of their expression upon PSPC1 depletion (Fig. 3d and Supplementary Fig. 7, f and g). These results suggest that the interaction of ERV RNA species with PSPC1 is required for the regulation of their cellular levels.

PSPC1 and TET2 participate in the repression of *MERVL* and adjacent genes during development

In line with our hypothesis that ERVs bound by PSPC1 likely influence adjacent genes that are deregulated by PSPC1 depletion, we found that PSPC1-interacting ERVs are more frequently located near the transcription start sites (TSSs) of PSPC1-deregulated genes than those RNAs encoding LTRs not bound by PSPC1 (Supplementary Fig. 8a). Notably, transcripts corresponding to *MERVL* elements, which are expressed in the two-cell (2C) stage embryo[23,29], were amongst the most strongly induced ERVs in the absence of PSPC1 (Fig. 3d and Supplementary Fig. 7, f and g). In agreement with this, we observed that 2C-embryo genes, deregulated upon PSPC1 knock-down (Supplementary Fig. 6b), had Class III LTR ERVL elements in close proximity to their TSSs (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* using a CRISPR SAM system[31] could recapitulate PSPC1 depletion in derepressing those PSPC1 bound and unbound two-cell specific genes (Supplementary Fig. 8, d and 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]. More importantly, we identified an RNA-binding dependent control of ERV expression, which was exemplified by the repressive effect of PSPC1 on *MERVL* and evident by the *MERVL* repression in PSPC1WT, but not PSPC1Mut, rescued *Pspc1* KO ESCs (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 importantly, that such interactions were also dependent on PSPC1 (Fig. 3f and Supplementary Fig. 9, a and b). Interestingly, depletion of TET2 (Fig. 3g and Supplementary Fig. 9c) in ESCs caused similar

expression changes of these REs as those of PSPC1 depletion (Fig. 3d and Supplementary Fig. 7f–g), and comparable to, if not more significant than, transcriptional deregulation of these REs in ESCs lacking other well-known ERV regulators such as *Kap1* or *G9a*[28,30,33] (Supplementary Fig. 10, a–d). These results suggest that the PSPC1-TET2 partnership may be critically involved in the regulation of REs whose expressions are dynamically regulated during early embryonic development[23].

MERVL expression peaks at the 2C stage of embryonic development and is greatly reduced by the blastocyst stage[23]. In order to validate our ESC data of PSPC1/TET2-mediated *MERVL* regulation in an *in vivo* developmental setting, we injected siRNA against *Pspc1* (si*Pspc1*), *Tet2* (si*Tet2*) or a non-targeting control (siNT) into mouse zygotes and cultured those embryos *in vitro* until the blastocyst stage (Fig. 3h and Supplementary Fig. 11, a and b). In line with previous reports on the dispensability of PSPC1[34] or TET2[35] for early development, we did not observe any significant delay in embryonic development upon PSPC1 or TET2 loss. However, we observed derepression of *MERVL* elements and *MERVL*-associated genes during early development upon depletion of PSPC1 or TET2 (Fig. 3h and Supplementary Fig. 11, c–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

In order to understand how PSPC1 might cooperate with TET2 to regulate ERV expression, we examined the correlation between our PSPC1 CLIP-seq peaks and 5-hydroxymethylcytosine (5hmC) as well as 5-methylcytosine (5mC) enrichment at the DNA level in ESCs[36]. In contrast to the high enrichment of 5mC/5hmC on non-repetitive 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 (i.e., *IAP* and *MusD*) whose proper activation was dependent on PSPC1 RNA-binding ability and TET2 presence (Supplementary Figs. 7h, 9c, 12d), consistent with 5hmC-mediated DNA demethylation and transcriptional activation. These results indicate distinct regulatory mechanisms of Class II (*IAP, MusD*) and Class III (*MERVL*) ERVs by PSCP1/TET2, and also suggested that catalytic activity dependent and independent functions of TET2 may be involved in the transcriptional activation of Class II ERVs *IAP*/*MusD* and repression of Class III ERVs *MERVL* elements, respectively.

TET2 mediates 5-hydroxymethylation (5hmC) of *MERVL* RNAs

To further understand these distinct regulatory mechanisms, we rescued *Tet1/2/3* triple knock-out (*Tet* TKO) ESCs with a wild-type (TET2WT) or a catalytic mutant (TET2Mut) TET2 (Supplementary Fig. 12c). As expected, we found that TET2Mut failed to 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 TET2WT-rescued cells (Fig. 4a), suggesting that the catalytic activity of TET2 has likely contributed to *MERVL* repression, possibly through a DNA-independent mechanism.

Recent findings uncover 5hmC as an epigenetic mark on RNA species in mammalian[5,6,8,37] and non-vertebrate organisms[7], but the functional aspects of this novel RNA epigenetic modification remain poorly defined. We therefore decided to determine whether PSPC1 and TET2 could mediate 5hmC-modification of *MERVL* and how such RNA modification might control *MERVL* abundance. We immunoprecipitated DNA-free RNA using an antibody against 5mC or 5hmC, and identified *MERVL* and *MERVL*-chimeric transcripts, but not *IAP* or *MusD*, among 5hmC-modified RNAs in ESCs (Supplementary Fig. 13, a–d). We also confirmed the presence of 5mC-modified *MERVL* transcripts (data not shown), consistent with a recent report on global 5mC profiling of poly(A) RNA in mouse ESCs[38]. Importantly, we found that the PSPC1-TET2 complex can bind both 5mC and 5hmC-modified *MERVL* 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- than 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 causing the subsequent release of the protein complex from RNA (Supplementary Fig. 13f). The functional contribution of TET2 to *MERVL* 5hmC modification was confirmed by a rescue experiment indicating that only TET2WT, but not catalytic mutant TET2Mut or either TET1 or TET3, can rescue 5hmC levels on *MERVL* transcripts in *Tet* TKO cells (Fig. 4b and Supplementary Fig. 14a). Importantly, consistent with PSPC1-dependent RNA association of TET2 (Fig. 2d, and Supplementary Figs. 5a and 9a), only the rescue of *Pspc1* KO cells with an RNA-binding competent PSPC1 (PSPC1WT) could restore 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, a.k.a. 5mC, can have stabilizing roles on mRNA[39,40], and that *MERVL* abundance is increased upon TET2 and PSPC1 loss (Fig. 3, e and g, Supplementary Figs. 7f and 9c), we evaluated the impact of 5hmC deposition in the stability of *MERVL* transcripts. To this end, we monitored *MERVL* levels after transcription inhibition with α-Amanitin or after a 5-ethylnyl uridine (EU) incorporation pulse in WT, *Pspc1* KO, and *Tet2* KO ESCs. Our analyses revealed a significant increase on the stability of *MERVL* transcripts in cells depleted of PSPC1 or TET2 (Fig. 4d and Supplementary Fig. 14, b and c), which strongly correlates with the absence of 5hmC in these transcripts (Fig. 4, b and c). Moreover, similar results were obtained by treatment with transcriptional inhibitor Triptolide (data not shown). To further confirm the hypothesis that 5hmC deposition on *MERVL* transcripts facilitates their destabilization, we identified PSPC1 consensus RNA-binding motifs, and performed a minigene reporter (d2EGFP) assay to assess the impact of PSPC1 binding on EGFP stability (Supplementary Fig. 14, d and e). We found that some of the PSPC1 binding motifs (motifs #1, #4 and #5), when fused to the d2EGFP minigene, mediated a significant decrease in EGFP protein expression, evident by the comparison of WT and mutant (Mut) motifs in generating relative mean fluorescence intensity (MFI) upon 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 with TET2Mut, although to a lesser extent than TET2WT, could also lead to a modest and yet appreciable decrease in *MERVL* expression (Fig. 4a), suggesting that the catalytic activity of TET2 and 5hmC-mediated RNA degradation alone 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 complexes in leukemia cells[41]. Indeed, histone deacetylases are among TET2 partners in our interactome (Supplementary Table 1), and the interactions of both HDAC1 and HDAC2 (HDAC1/2) with PSPC1 were confirmed (Supplementary Fig. 15, a and b). More importantly, we detected a reduction in HDAC1/2 occupancy in *MERVL loci* in the absence of PSPC1 (Supplementary Fig. 15c). Such a PSPC1-dependent HDAC1/2 binding to *MERVL loci* is also reliant on PSPC1 RNA-binding capacity, as was supported by the failure of HDAC1/2 binding in *Pspc1* KO cells rescued with PSPC1Mut, compared with that in PSPC1WT rescued cells (Supplementary Fig. 15c). These data suggest that HDAC1/2 and/or their histone deacetylase activity could mediate the transcriptional repression of *MERVL*. Consistent with this, both chemical inhibition of HDAC activity using valproic acid (VPA) and *Hdac1/2* knock-down (sh*Hdac1/2*) led to an increased expression of *MERVL* and *MERVL*-associated genes (e.g., *Zfp352*) (Supplementary Fig. 15, d and e), in line with what we observed for PSPC1 and TET2 depletion. Interestingly, the repressive activity of HDAC1/2 on *MERVL* expression was dependent on the presence of an RNA-binding competent PSPC1, and 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. By applying the interactome study in mouse ESCs, we not only rediscovered the well-known TET2 partner protein OGT[19,42,43], 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 Erk signaling and mouse ESC pluripotency, lack of which stabilizes ESCs at 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 posttranscriptional repression of MERVL in pluripotent stem cells. Specifically, PSPC1 recruits TET2 for the deposition of 5hmC onto Class III LTR ERVL RNAs, leading to their destabilization, while facilitating concomitant recruitment of HDAC activity to repress their transcription in mouse ESCs (Fig. 4e). While high-resolution mapping of RNA-binding regions in the nuclear proteome of ESCs revealed

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, 3f, Supplementary Fig. 5a).

ERVs are well-recognized for their roles in contributing to host genome evolution and gene regulatory networks, and their aberrant regulatory activities also link to pathological and oncological conditions[1,2,47]. The pluripotent embryonic cells serve as the “battle ground” for an evolutionary arms race between transposable elements and 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[10,48]. While transcriptional and epigenetic control of ERVs via DNA (de)methylation and histone modifications are well established, our study for the first time suggests that ERV transcripts can also be posttranscriptionally 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[49]. In particular, our study provides novel 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 posttranscriptional level via an RNA hydroxymethylation-mediated mechanism (Fig. 4e). However, our findings on RNA-dependent chromatin targeting of TET2 for ERV control are not universal, but varied depending on ERV classes. This is evident by the fact that the PSPC1-TET2 partnership has a positive role on other ERV family expression (i.e., Class II LTR ERVK family members *IAP* and *MusD*; Fig. 3c and d) via PSPC1-TET2 mediated transcriptional activation (Supplementary Figs. 7h, 12b and 12d). Future studies will be needed to dissect such transcriptional activation mechanism for Class II ERVs and 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[50] or contribute to 5hmC *MERVL* RNA hydroxymethylation (Supplementary Fig. 14a).

In sum, our study provides a new paradigm for posttranscriptional silencing of Class III ERVs (i.e., *MERVL*) RNAs, via 5hmC modification by an RNA-binding protein (i.e., PSPC1) mediated TET2 recruitment. Since ERV reactivation has been widely related to aging, cancer, and autoimmune diseases[51,52], our findings should also open new avenues for exploring posttranscriptional ERV control by RNA hydroxymethylation in health and disease.

ONLINE METHODS

Murine embryonic stem cell (ESC) 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% fetal bovine serum (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 of 2-mercaptoethanol (Sigma M6250)

and homemade recombinant leukemia inhibitory factor (LIF) 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/3xFLAG/V5* sequence. A 2.2 kb 5′ and a 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 blot.

Two independent affinity purification approaches were employed 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[53]. Briefly, five large square dishes (245 × 245 mm) of each cell line were washed with PBS, scrapped, and cytoplasmic fraction was removed by incubating cells with Buffer A (10 mM HEPES pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl) supplemented with proteinase inhibitors. Afterwards, 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, salt concentration was decreased to 100 mM by dialyzing with Buffer D (20mM HEPES pH 7.6, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol) 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 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 hours at 4°C. After five washes in Buffer D supplemented with 0.02% NP-40, the 3xFLAG-tagged TET2 protein complexes were eluted four times for 1 h each at 4°C, with 0.3 mg/ml 3xFLAG peptide in Buffer D supplemented with 0.02% NP-40. After concentration, protein complexes were boiled 5 min with Laemmli sample buffer and separated in a 10% SDS-PAGE gel.

For the second approach, we used SILAC IP-MS where WT and *Tet2* knock-in ESC lines were cultured in medium labeled by light L-Arginine-HCL, L-Lysine-2HCL (Thermo Scientific 88427) or heavy L-Arginine-HCL (U-13C6, 99%; U-15N4, 99%) (Cambridge Isotope Laboratories CNLM-539-H-0.25), L-Lysine-2HCL (U-13C6, 99%; U-15N2, 99%) (Cambridge Isotope Laboratories CNLM-291-H-0.25) amino acids, respectively. 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 3xFLAG 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 mass spectrometry analysis.

Mass spectrometry data analysis

MS data were processed by Thermo Proteome Discoverer software with SEQUEST engine against International Protein Index (IPI) mouse protein sequence database (v.3.68). Protein lists were filtered by minimal number of identified peptides > 2. Common contamination proteins (trypsin, keratins, actin, tubulins) were removed. Duplicated records were removed by unique protein symbol. Then protein enrichment ratio was calculated by Heavy/Light ratio (SILAC) or spectrum counts (label-free) of TET2 pulldown versus control pulldown.

Co-immunoprecipitation and Western Blot

Nuclear extracts from CCE ESCs prepared as described before 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 using the following primary antibodies: 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), 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 Image J software.

Lentiviral infection for shRNA knockdown

Small hairpin RNAs (shRNAs) for *Pspc1* knock-down 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[54]. All shRNA sequences are provided in Supplementary Table 4.

RNA extraction and analysis by quantitative PCR

Total RNA from ESCs, iPSCs and MEFs was extracted with the RNeasy kit (Qiagen #74136) and converted to cDNA using qSCRIPT (Quanta #95048). Gene expression was analyzed using the Lightcycler 480 SYBR Green Master Mix (Roche #4729749001) on the LightCycler480 Real-Time PCR System (Roche).

One adult male and one adult female mice were dissected to isolate organs of interest. Thirty milligrams of each tissue were disaggregated with QIAshredder columns (Qiagen #79656). Total RNA was extracted with RNeasy kit and subjected to reverse transcription and qPCR quantitation as described per manufacturer's instructions.

RNA from 1.5, 2.5 and 4.5 dpc embryos injected with siRNAs was extracted with TRIzol (Thermo Fisher #15596026) according to manufacturer's instructions. Total purified RNA was subjected to reverse transcription as explained 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 a housekeeping gene as indicated (*Hprt*, *U6* or *β-Actin*). Error bars indicate standard error of the mean (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

Whole cell protein extracts were obtained by lysing cell pellets in RIPA buffer (60 mM Tris-HCl pH 6.8, 2% SDS, 10% Glycerol and 10 mM DTT) supplemented with PMSF and proteinase inhibitors. Chromatin bound fraction of proteins was prepared using the chromatin extraction kit (Thermo Scientific #PI-78840) according to manufacturer's instructions.

Immunofluorescence

Cells were grown on 24-well plates coated with 0.1% of gelatin. After fixation with 4% paraformaldehyde for 15 min at room temperature (RT), cells were permeabilized with 0.25% Triton X-100 in PBS for 5 min at RT and blocked with 10% bovine serum albumin (BSA, AMRESCO) for 30 min at 37°C. Immunostaining was performed by incubating the cells overnight at 4°C with primary antibodies anti-SOX2 (Santa Cruz #sc-17320), anti-PSPC1, anti-OCT4 in PBS with 3% BSA. 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 with a 20x magnification.

Cell cycle analysis and flow cytometry

For cell cycle analysis, an equal number of cells were washed with DPBS (Dulbecco's phosphate buffered saline), permeabilized with 0.1% Triton X-100 in DPBS, stained with 10 μM dye 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 using the Dean-Jett-Fox cell cycle model.

CRISPR-Cas9 generation of *Pspc1* knock-out ESC line

Pspc1 knock-out mouse ESCs were generated using the CRISPR-editing tool as described[55]. Briefly, an sgRNA (Supplementary Table 4) was designed to target the transcription start site of *Pspc1* gene using the guidelines described in http://crispr.mit.edu/ 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)[23] using the 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* targeted (+317F4 from gag ATG) sgRNAs were cloned using BsmB1 restriction sites. Lentiviruses containing each of the plasmids were generated and

ESCs were infected with a 1:1:1 mix of the viruses in presence of 8 μg/ml Polybrene. Infected cells were selected for 48 h with 10 μg/ml of Blasticidin, 250 μg/ml of Hygromycin, and 250 μg/ml Zeocin. Selected cells were collected for total RNA isolation and qPCR analysis.

Chromatin immunoprecipitation (ChIP) coupled with qPCR

ChIP assays were performed as previously described[56]. Briefly, cells were crosslinked with 1% (w/v) formaldehyde for 10 min at RT, followed by the addition of 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) antibodies. The immunoprecipitated DNA was purified with ChIP DNA Clean & Concentrator columns (Zymo Research) and analyzed by qPCR using the 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 (iv-RIP)

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 antibodies 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 incubating 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 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 using Qubit® High Sensitive Assay kit (Life Technologies #Q32852). Immunoprecipitated RNA was visualized in an ethidium bromide agarose gel and retrotranscribed with qSCRIPT (Quanta #95048) to be analyzed by qPCR.

RNA immunoprecipitation (RIP) of crosslinked cells

Crosslinked 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) which were 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 rotating at 4°C. After washing, immuno-complexes were eluted with 100 μl of elution buffer (10 mM Tris-HCl pH 8.0 with 1mM EDTA) and RNA was extracted with TRIzol. Immunoprecipitated RNA was treated with DNase I and with or without RNase A followed by phenol:chloroform extraction. The resulting RNA was retrotranscribed with qScript and cDNA was visualized in a polyacrylamide gel by silver staining (Thermo Scientific #24600) following manufacturer's instructions.

UV-crosslinking and RNA immunoprecipitation coupled with qPCR (RIP-qPCR)

Cells were trypsinized and UV-crosslinked following previously published protocols[57]. Briefly, cells were irradiated with 400 mJ/cm2 in a CL-1000 UVP UV-crosslinker 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) in 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) in 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 VPA treatments

Cells in culture were treated with 10 μg/ml of the RNA Pol II and Pol III inhibitor α-Amanitin (Santa Cruz #sc-202440) or milli-Q water control (vehicle) for 0 h, 1 h, 2 h, or 4 h. For HDAC inhibition, cells were cultured in presence of 0.5 μM Class I HDAC inhibitor Valproic Acid (VPA) (Stemgent #04-0007) or DMSO control (vehicle) for 24 h. For both experiments, cells were harvested and processed for Western blotting or qPCR analysis.

RNA stability assay by nascent RNA capture

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

Crosslinking immunoprecipitation and massive parallel sequencing (CLIP-seq)

UV-crosslinking and immunoprecipitation were performed as previously described[58] with some modifications. Briefly, J1 mouse ESCs expressing a 3xFLAG-biotin tagged PSPC1 construct were crosslinked in PBS with UV type C (254 nm) at 600 mJ per cm^2 on ice. Cells were harvested, pelleted and lysed in PXL lysis buffer (1× PBS, 0.1% SDS, 0.5% NP-40 and 0.5% sodium deoxycholate) supplemented with proteinase and RNase inhibitors and RQ1 DNase (Promega #M6101). After 30 min 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 antibody conjugated with 30 μl Protein G Dynabeads overnight at 4°C. After immunoprecipitation, the sample was washed with PXL lysis buffer twice and then high salt buffer twice (5xPBS, 0.1% SDS, 0.5 NP-40, 0.5% sodium deoxycholate). Then the protein-RNA complex was subjected to MNase digestion (New England BioLabs #M0247). To dephosphorylate RNA, each immunoprecipitated sample was incubated in 80 μl of 1x 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) and twice with PNK buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, and 0.5% NP-40). To ligate the 3′ linker, each of the washed immunoprecipitates was incubated in 40 μl of 1x

ligation reaction mixture with 80 pmol 3′ linker (TAKARA, Supplementary Table 4), and 3 μl truncated T4 RNA ligase 2 (NEB #M0242) for overnight at 16°C. After the reaction, immunoprecipitates were washed twice with PXL buffer, twice with PNK buffer. Then the RNAs were phosphorylated by incubating the immunoprecipitates in 1x reaction mixture containing 2 μl of T4 PNK (NEB #M0201) and 1 μl of hot ATP (Perkin Elmer) for 5 min at 37°C to label the RNA-protein complexes. After the labeling, 5 μl of 2 mM ATP was added and incubated for another 5 min at 37°C. After PNK treatment, immunoprecipitates were washed 4 times with PNK buffer and mixed with 2x 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 an NuPage SDS gel and transferred into a nitrocellulose membrane.

For RNA isolation, nitrocellulose membranes were fragmented with a clean scalpel blade and treated with 4 mg/ml of proteinase K (TAKARA #9034) in 200 μl of PK buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10m M 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. 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 were carried out 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 using the Piranha tools (http://smithlabresearch.org/software/piranha/) with the parameters -s -b 200 and annotated with the annotatePeaks module in HOMER program (v4.6) against the mm9 mouse genome.

Motif finding in CLIP-seq

CLIP-seq peaks were used as input for de novo motif discovery using HOMER with the parameters “-rna -len 6”, and the mm9 reference genome.

RNA stability reporter assay

To generate the mRNA stability minigene construct, the top five consensus motifs of PSPC1 CLIP-seq were cloned upstream of the EGFP CDS in the pd2EGFP-N1 vector (BD Biosciences). This vector expresses a destabilized version of EGFP protein with a half-life of 2h, 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 (see Supplementary Table 4). The corresponding mutated motifs were designed by replacing Cytosines with Adenines for each motif. All the motifs were cloned with XhoI and BamHI sites into the reporter vector. Each vector was transfected with JetPrime polyplus reagent into ESCs in 48 wells. Next day, transcription was inhibited by adding 10 μg/ml of α-Amanitin or vehicle (milli-Q) to the medium. 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, indicator of RNA stability, was gated on the GFP^+ singlet population.

ChIP-seq, MeDIP-seq and hMeDIP-seq analysis

External data for ChIP-seq, MeDIP-seq and 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 the bowtie (v1.0.0) program, 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 previously described[59]. Briefly, RepeatMasker track (RMSK) from UCSC Genome Browser was used and ChIP-seq intensity at each LTR region was counted by HTseq software (v0.6.1) with parameters -a 10 -m intersection-nonempty. ChIP-seq intensity of TET2 and 5hmC at each LTR region was normalized by total mapped reads as reads per million (RPM).

hMeDIP-qPCR

Genomic DNA from wild type ESCs was isolated following manufacturer's instructions (Qiagen #158388). Four micrograms of gDNA were denatured by boiling 10 min in water, and immunoprecipitated with 2.5 μl of anti-5hmC antibody in immunoprecipitation buffer (IP buffer: 100 mM Sodium Phosphate pH 7.0, 1.4 M NaCl and 0.5% Triton X-100). After 6 hours of incubation at 4°C, the immunobound DNA was recovered by adding 20 μl of preequilibrated Dynabeads G and rotated at 4°C overnight. 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 for *Pspc1* or control shRNAs. Biological duplicates were prepared for RNA-seq analysis, following our previously described protocol[60]. Briefly, total RNA from each sample was extracted by RNeasy kit and paired-end sequencing was performed with the Illumina HiSeq-2500, following a RiboZero selection protocol according to 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 performed using Cufflinks (v2.1.1). Assembling of novel transcripts was allowed (-g), other parameters followed the default setting. The summed FPKM (fragments per kilobase per million mapped reads) of transcripts sharing each gene_id was calculated and significance of differential expression test was estimated via a genome-wide false discovery rate (FDR) after Benjamin-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 and sh*Pspc1* samples were compared and *p*-values were calculated based on student *t*-test.

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 regions was counted by HTseq and normalized as RPM values. PSPC1-bound (intensity > 0.5 RPM) and PSPC1-unbound (intensity=0) 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 distribution of distances between PSPC1-upregulated genes by RNA-seq and nearest PSPC1-bound or -unbound sites were calculated and plotted. Significance of the number of nearest PSPC1-bound versus -unbound sites (< 50 kb) were calculated by binomial test.

RNA-seq analysis of previously published datasets

External RNA-seq data for *Kap1* cKO (GSE41903) and *G9a* cKO (GSE33923) 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[22] and 16 ng 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 kit DualGlo® Luciferase Assay (Promega #E2920) following manufacturer's instructions in a Perkin Elmer EnSpire Alpha Luminometer. Luciferase/Renilla ratio was calculated for all the samples. Measurements were performed in triplicate biological samples.

Mouse embryo collection and microinjection

C57Bl/6JxDBA/2J (B6D2) female mice were superovulated with intraperitoneal injection of 5 IU pregnant mare's serum gonadotropin (PMSG; National Hormone and Peptide Program) followed by intraperitoneal injection of 5 IU human chorionic gonadotropin (hCG; National Hormone and Peptide Program) 48 h later. After overnight mating with males, one-cell embryos were collected in HEPES-buffered FHM media (Millipore #MR-024-D) and 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 five to ten picoliter of 20 μM non-targeting 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. Microinjections were performed using a Nikon Diaphot inverted microscope equipped with a Narashige micromanipulator system. After injection, embryos were inspected every day to determine developmental progress. Total RNA was extracted for qPCR analysis and processed as described before. 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. Zygote injections were performed in the Mouse Genetic Shared Research Facility (SRF) at Mount Sinai.

Methylated and hydroxymethylated RNA Immunoprecipitation (MeRIP and hMeRIP)

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) antibodies or an IgG (Millipore #PP64) control based on a previously described protocol[61]. Briefly, 6 to 10 μg of sonicated RNA were incubated with 2 μg of antibody at 4°C overnight. Next day, 5hmC-modified RNAs were purified by incubating with 20 μl of Dynabead Protein G beads. After being washed to reduce non-specific background, bound RNA was eluted with TRIzol (Invitrogen) and extracted following manufacturer's instructions. Finally, immunoprecipitated RNAs were subjected to reverse transcription and qPCR quantitation.

RNA electrophoretic mobility shift assay (REMSA)

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 the ion-exchange HPLC purification using Dionex PA-200 column. HEK293T cells were transfected with plasmids coding for ^{3xFL}TET2 and ^{Myc}PSPC1 and total protein extracts were incubated with anti-FLAG antibody to immunopurify PSPC1-TET2 complexes. The complex was competitively eluted by 3xFLAG peptide. Increasing amount of PSPC1-TET2 protein complex (0, 0.5, 1, 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, and 40 U/ml RNasin, 1 mM DTT, 50% glycerol, 5 ng/μl BSA) for 30 min at RT. Then, 1 μl glutaraldehyde (0.2% final concentration) was added into the mixture, which was incubated at room temperature for 15 min. The total protein-RNA mixture was loaded into a 6% TBE acrylamide gel and run 30 min at 80 V on ice. The gel was transferred onto hybond-N^+ membrane (GE Healthcare #95038-362) in 1x TBE buffer and nucleic acids detected by the chemiluminescent nucleic acid detection module (Thermo Fisher #89880) following the manufacturer's instructions. Quantification of each band was carried out with ImageJ software and the percentage of bound RNA probe were calculated as [intensity of bound probe]/([intensity of bound probe] + [intensity of free probe]) *100.

Data availability

RNA-seq (GSE103267) and CLIP-seq (GSE103268) datasets generated in this study have been deposited in the Gene Expression Omnibus (GEO) database. A Life Sciences Reporting Summary for this paper is available.

URLs

CRISPR sgRNA design: http://crispr.mit.edu/; DAVID gene Functional Classification Tool: https://david.ncifcrf.gov/; Image J: 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: Analysing high-throughput sequencing data with Python: https://htseq.readthedocs.io/en/release_0.9.1/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Y. Kurihara for Pspc1 constructs, T. Macfarlan for the *Zfp352*-luciferase reporter construct, R. Jaenisch for *Tet* TKO ESC line, D. Trono for Kap1 cKO line, and D.M. Gilbert and Y. Shinkai for G9a 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 grants from the National Institutes of Health (NIH) to J.W. (1R01-GM095942 and R21HD087722), the Empire State Stem Cell Fund through New York State Department of Health (NYSTEM) to J.W. (C028103, C028121). J.W. is a recipient of 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. M.X. is supported by NIH R01HL112294. The research from the Shen laboratory is supported by the National Basic Research Program of China (2012CB966703), the National Natural Science Foundation of China (31471219, 8141101062, 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) (AEI/FEDER, UE), and the Consellería de Cultura, Educación e Ordenación Universitaria (ED431F 2016/016) to M.F.

References

1. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 2016; 30:733–50. [PubMed: 27036965]
2. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet. 2017; 18:517–534. [PubMed: 28555658]
3. Frye M, Blanco S. Post-transcriptional modifications in development and stem cells. Development. 2016; 143:3871–3881. [PubMed: 27803056]
4. Masiello I, Biggiogera M. Ultrastructural localization of 5-methylcytosine on DNA and RNA. Cell Mol Life Sci. 2017
5. Zhang HY, Xiong J, Qi BL, Feng YQ, Yuan BF. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chem Commun (Camb). 2016; 52:737–40. [PubMed: 26562407]
6. Miao Z, et al. 5-hydroxymethylcytosine is detected in RNA from mouse brain tissues. Brain Res. 2016; 1642:546–52. [PubMed: 27117867]
7. Delatte B, et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016; 351:282–5. [PubMed: 26816380]
8. Fu L, et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc. 2014; 136:11582–5. [PubMed: 25073028]
9. Schlesinger S, Goff SP. Retroviral transcriptional regulation and embryonic stem cells: war and peace. Mol Cell Biol. 2015; 35:770–7. [PubMed: 25547290]
10. Macfarlan TS, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012; 487:57–63. [PubMed: 22722858]
11. Kim J, Cantor AB, Orkin SH, Wang J. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nat Protoc. 2009; 4:506–17. [PubMed: 19325547]
12. Wang J, et al. A protein interaction network for pluripotency of embryonic stem cells. Nature. 2006; 444:364–8. [PubMed: 17093407]
13. Huang Y, et al. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. Proc Natl Acad Sci U S A. 2014; 111:1361–6. [PubMed: 24474761]

14. Hon GC, et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell. 2014; 56:286–97. [PubMed: 25263596]
15. Sigova AA, et al. Transcription factor trapping by RNA in gene regulatory elements. Science. 2015
16. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010; 329:689–93. [PubMed: 20616235]
17. Fox AH, Bond CS, Lamond AI. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. Mol Biol Cell. 2005; 16:5304–15. [PubMed: 16148043]
18. Gilbert SL, Pehrson JR, Sharp PA. XIST RNA associates with specific regions of the inactive X chromatin. J Biol Chem. 2000; 275:36491–4. [PubMed: 11006266]
19. Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013; 493:561–4. [PubMed: 23222540]
20. Knott GJ, Bond CS, Fox AH. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. Nucleic Acids Res. 2016; 44:3989–4004. [PubMed: 27084935]
21. Ma C, et al. Nono, a Bivalent Domain Factor, Regulates Erk Signaling and Mouse Embryonic Stem Cell Pluripotency. Cell Rep. 2016; 17:997–1007. [PubMed: 27760330]
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. 2014; 15:4. [PubMed: 24383959]
23. Peaston AE, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell. 2004; 7:597–606. [PubMed: 15469847]
24. Mouse Genome Sequencing C et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002; 420:520–62. [PubMed: 12466850]
25. Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Biol. 2015; 31:429–51. [PubMed: 26393776]
26. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016; 351:1083–7. [PubMed: 26941318]
27. Leeb M, et al. Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev. 2010; 24:265–76. [PubMed: 20123906]
28. Maksakova IA, 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. 2013; 6:15. [PubMed: 23735015]
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. 2003; 68:651–4. [PubMed: 12533431]
30. Macfarlan TS, et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. 2011
31. Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015; 517:583–8. [PubMed: 25494202]
32. Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virology. 2011; 411:273–87. [PubMed: 21251689]
33. Rowe HM, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature. 2010; 463:237–40. [PubMed: 20075919]
34. Kowalska E, et al. Distinct roles of DBHS family members in the circadian transcriptional feedback loop. Mol Cell Biol. 2012; 32:4585–94. [PubMed: 22966205]
35. Li Z, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood. 2011
36. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–8. [PubMed: 21490601]
37. Huber SM, et al. Formation and abundance of 5-hydroxymethylcytosine in RNA. Chembiochem. 2015; 16:752–5. [PubMed: 25676849]
38. Amort T, et al. Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. Genome Biol. 2017; 18:1. [PubMed: 28077169]
39. Zhang X, et al. The tRNA methyltransferase NSun2 stabilizes p16INK(4) mRNA by methylating the 3′-untranslated region of p16. Nat Commun. 2012; 3:712. [PubMed: 22395603]

40. Warren L, et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell. 2010; 7:618–30. [PubMed: 20888316]
41. Zhang Q, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature. 2015; 525:389–93. [PubMed: 26287468]
42. Vella P, et al. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell. 2013; 49:645–56. [PubMed: 23352454]
43. Deplus R, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. Embo J. 2013
44. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell. 2009; 35:467–78. [PubMed: 19716791]
45. Fox AH, Lamond AI. Paraspeckles. Cold Spring Harb Perspect Biol. 2010; 2:a000687. [PubMed: 20573717]
46. He C, et al. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Mol Cell. 2016; 64:416–430. [PubMed: 27768875]
47. Ko M, An J, Rao A. DNA methylation and hydroxymethylation in hematologic differentiation and transformation. Curr Opin Cell Biol. 2015; 37:91–101. [PubMed: 26595486]
48. Choi YJ, et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. Science. 2017; 355
49. Gerdes P, Richardson SR, Faulkner GJ. TET enzymes: double agents in the transposable element-host genome conflict. Genome Biol. 2016; 17:259. [PubMed: 27993162]
50. de la Rica L, et al. TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. Genome Biol. 2016; 17:234. [PubMed: 27863519]
51. Lee E, et al. Landscape of somatic retrotransposition in human cancers. Science. 2012; 337:967–71. [PubMed: 22745252]
52. Li W, et al. Human endogenous retrovirus-K contributes to motor neuron disease. Sci Transl Med. 2015; 7:307ra153.
53. Ding J, Xu H, Faiola F, Ma'ayan A, Wang J. Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res. 2012; 22:155–67. [PubMed: 22083510]
54. Ivanova N, et al. Dissecting self-renewal in stem cells with RNA interference. Nature. 2006; 442:533–8. [PubMed: 16767105]
55. Ran FA, et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013; 8:2281–308. [PubMed: 24157548]
56. Lee TI, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006; 125:301–13. [PubMed: 16630818]
57. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods. 2005; 37:376–86. [PubMed: 16314267]
58. 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. 2009; 36:996–1006. [PubMed: 20064465]
59. Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature. 2015; 522:240–4. [PubMed: 25938714]
60. Ding J, et al. Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming. Cell Stem Cell. 2015; 16:653–68. [PubMed: 25936917]
61. Dominissini D, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012; 485:201–6. [PubMed: 22575960]
62. Yokochi T, et al. G9a selectively represses a class of late-replicating genes at the nuclear periphery. Proc Natl Acad Sci USA. 2009; 106(46):19363–8. [PubMed: 19889976]

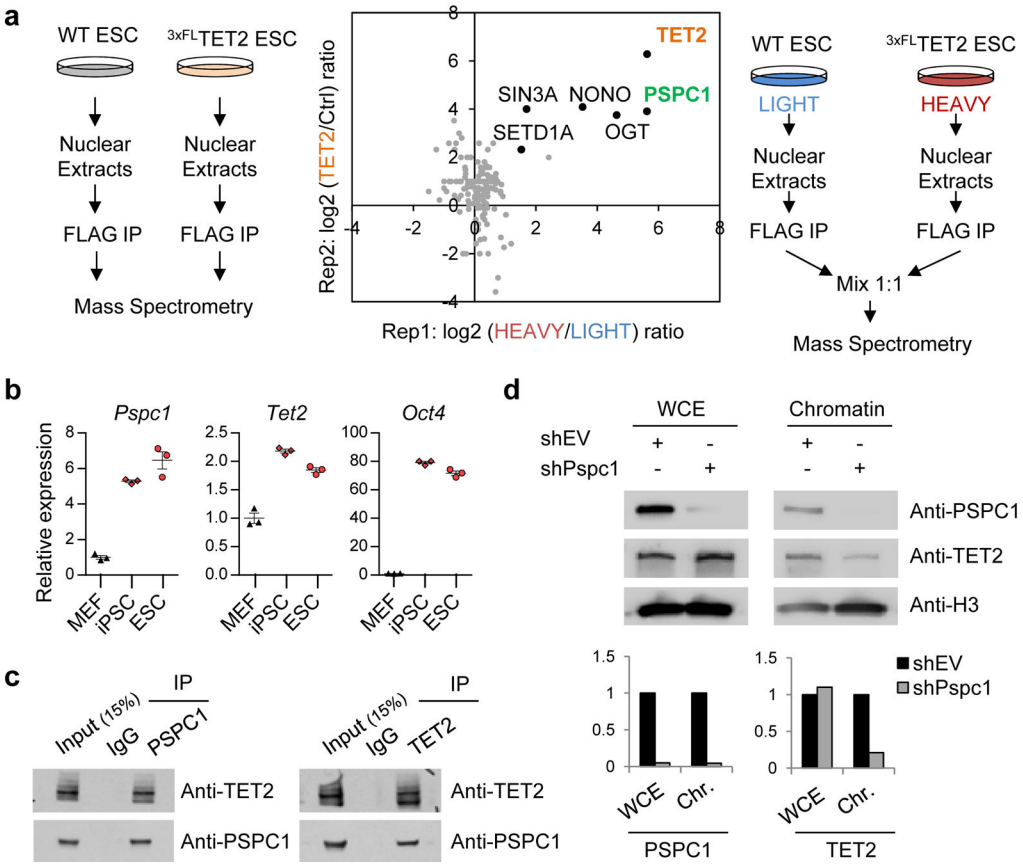

Figure 1. TET2 is recruited to chromatin by the RNA-binding protein PSPC1

a, Illustration of the two complementary techniques (Rep1 and Rep2) employed to identify TET2- interacting proteins in mouse ESCs. (Left) The experimental scheme for FLAG immunoprecipitation (IP) followed by mass spectrometry (MS) of *3xFL Tet2* knock-in and wild-type (WT) control ESC lines. (Right) Scheme of the SILAC-based *in vivo* labeling approach used to determine TET2 partners by IP with an anti-FLAG antibody using the nuclear extracts from *3xFL Tet2* knock-in ESCs and wild-type (WT) ESCs followed by MS analysis. (Center) Ratios of TET2-interacting peptides versus non-specific peptides detected by AP-MS in both IP-MS experiments. **b**, Relative RNA expression levels of *Pspc1*, *Tet2* and the ESC marker *Oct4* in differentiated (MEF) versus pluripotent (iPSC, ESC) cell lines. Data are from one representative experiment (n=3 technical replicates) and 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 upon *Pspc1* depletion. (Top) Western blot analysis of total (whole cell extract, WCE) and chromatin-bound (Chromatin) PSPC1 and TET2 in control (shEV) and PSPC1 knock-down (shPspc1) ESCs. Histone H3 was used as a loading control. (Bottom) Quantitation of the relative levels of PSPC1 and TET2 in WCE and chromatin (Chr.) compared to H3 and shEV. In c and d, images are representative of immunoblots from 2 independent experiments.

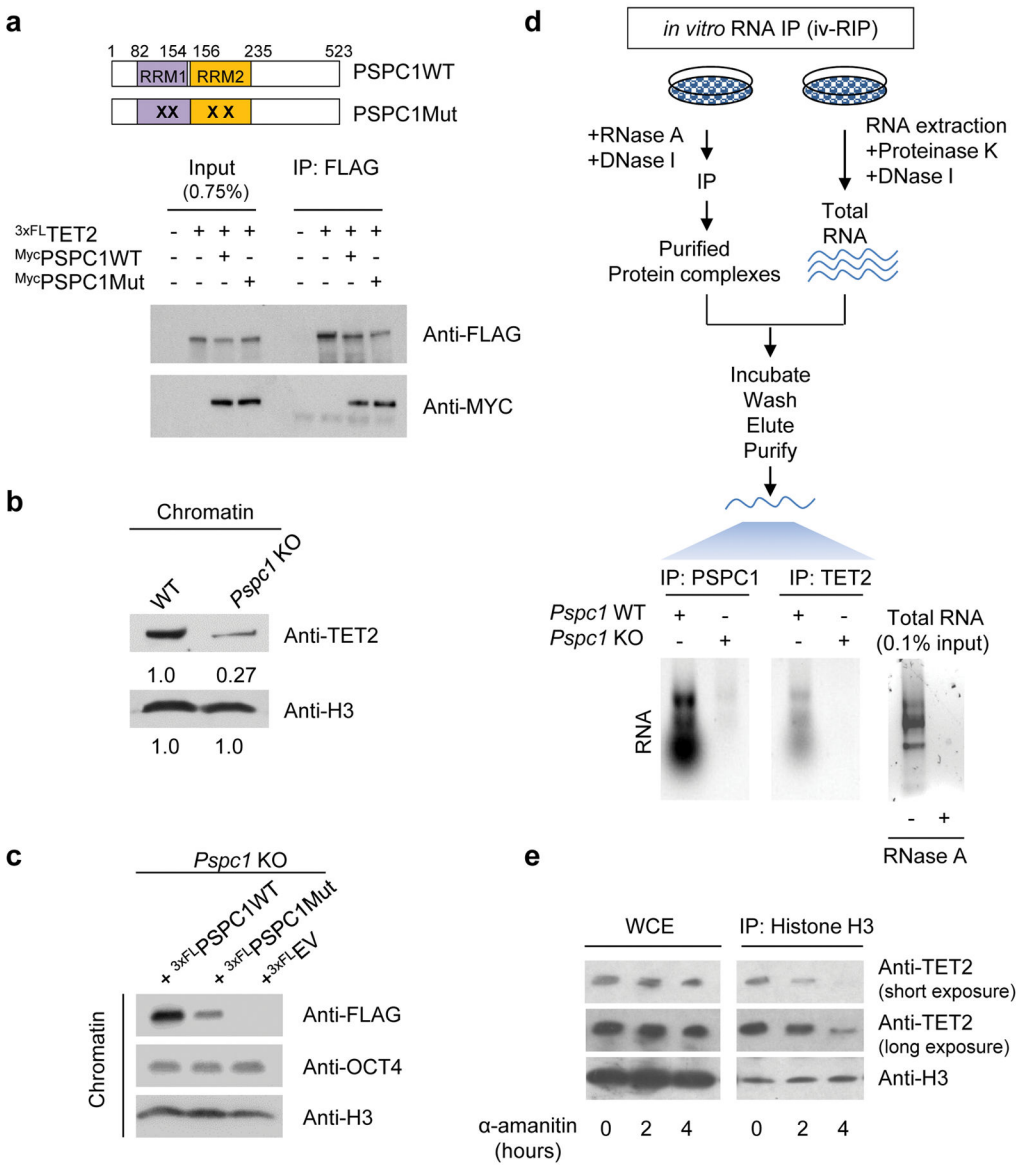

Figure 2. PSPC1 mediates TET2 recruitment to chromatin through RNA

a, (Top) Depiction of the wild-type (PSPC1WT) and RNA binding mutant (PSPC1Mut) PSPC1 protein structures. (Bottom) The maintained interactions of ^{3xFL}TET2 with both ^{Myc}PSPC1WT and ^{Myc}PSPC1Mut by coimmunoprecipitation (CoIP). HEK293T cells were transiently transfected with constructs expressing ^{3xFL}TET2 and ^{Myc}PSPC1 or ^{Myc}PSPC1Mut followed by IP with anti-FLAG and western blotting with the indicated antibodies. **b**, Reduction of chromatin-bound TET2 in *Pspc1* KO lines relative to WT control. The relative levels of TET2, normalized to Histone 3 (H3) and WT controls, are indicated. **c**, Chromatin-bound ^{3xFL}PSPC1WT and ^{3xFL}PSPC1Mut proteins in the *Pspc1* KO rescued cell lines. OCT4 is shown as a negative control. **d**, (Top) Depiction of the *in vitro* RNA-immunoprecipitation assay (iv-RIP) protocol. (Bottom) Agarose gel analysis of RNA bound by PSPC1 (IP: PSPC1) or TET2 (IP: TET2) protein complexes in *Pspc1* WT and KO ESCs. Total RNA used for the IP was incubated with DNase I to ensure the complete

absence of DNA contamination (0.1% input is shown on the right panel). **e**, TET2 chromatin binding is reduced upon 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 hours, compared to untreated cells (0 hours). Histone H3 was used as a loading control in b, c, and e. All images are representative of at least 2 independent experiments.

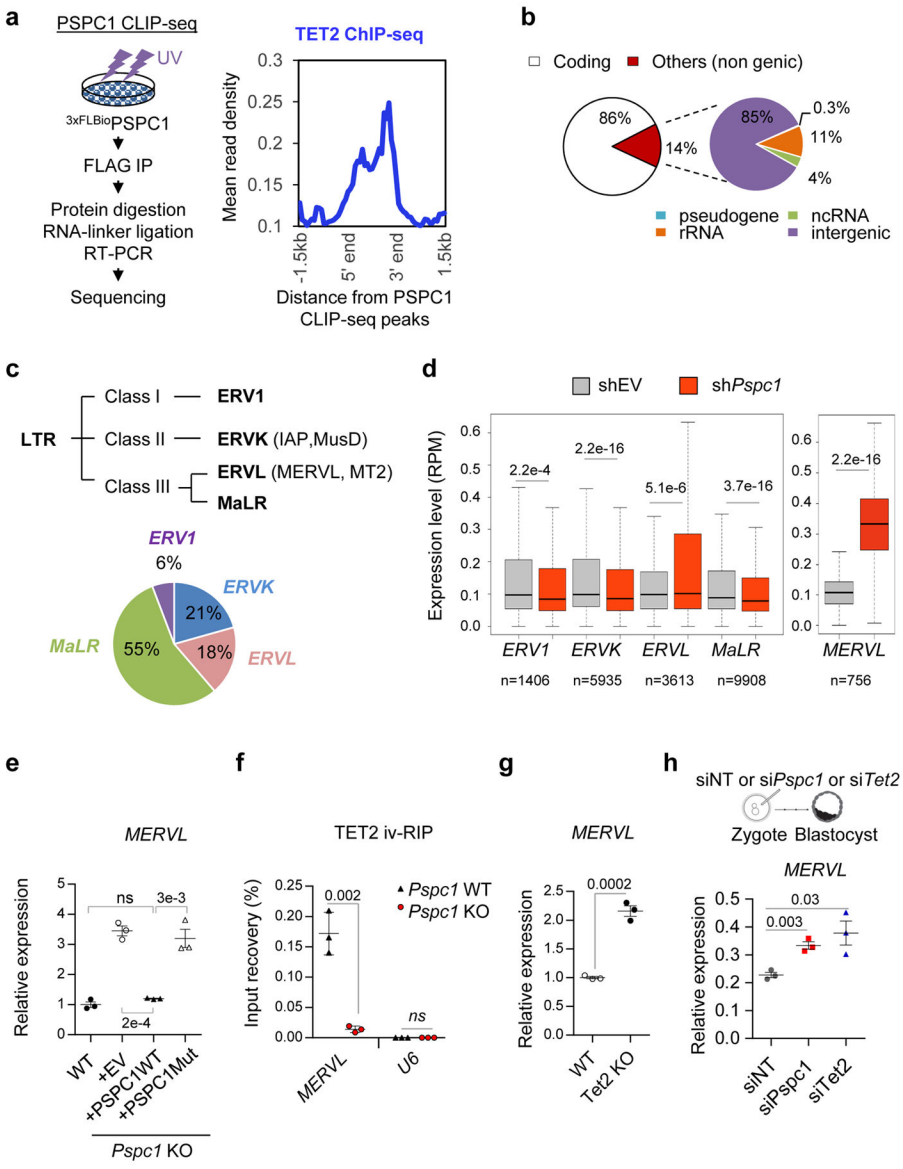

Figure 3. PSPC1 binds to *MERVL* RNAs with TET2 to repress their expression

a, (Left) Depiction of ^{3xFLBio}PSPC1 RNA-immunoprecipitation followed by high-throughput sequencing (CLIP-seq) in ESCs. Data were generated from one experiment. (Right) Mean read density of TET2 ChIP-seq[19] at the chromatin *loci* corresponding to PSPC1 CLIP-seq peaks. **b**, Distribution of PSPC1 CLIP-seq peaks in coding and non-genic RNAs (in percentage). **c**, (Top) Schematic tree and classification of LTR-containing retrotransposons. (Bottom) Quantitation of LTR-containing RNAs bound by PSPC1 of each LTR family. **d**, Expression (in reads per million, RPM) of Class I, II, and III LTR ERVs including *MERVL* after PSPC1 depletion (sh*Pspc1*) compared to a knockdown control (shEV). Box plot (n is indicated in the figure) with whiskers extending to ±1.5 × IQR (the interquartile range); line across the box indicates the median. Statistical analysis: Wilcoxon rank sum test with continuity correction. **e**, Quantitative PCR analysis of *MERVL* expression in WT and *Pspc1* KO ESCs rescued with ^{3xFL}PSPC1WT or ^{3xFL}PSPC1Mut. The data are

relative to WT ESCs. **f**, Quantitative PCR analysis of *MERVL* abundance among TET2-interacting RNAs analyzed by *iv*-RIP in *Pspc1* WT and *Pspc1* KO cells. *U6* is used as a negative control. **g**, *MERVL* expression in WT and *Tet2* knock-out ESCs (*Tet2* KO), relative to WT. **h**, (Top) Cartoon depicting zygote injection with siRNAs to deplete PSPC1 (si*Pspc1*) or TET2 (si*Tet2*) compared to a non-targeting control (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 represented as mean ± s.e.m. (n=3 independent experiments) by two-tailed Student's *t*-test. ns, not significant.

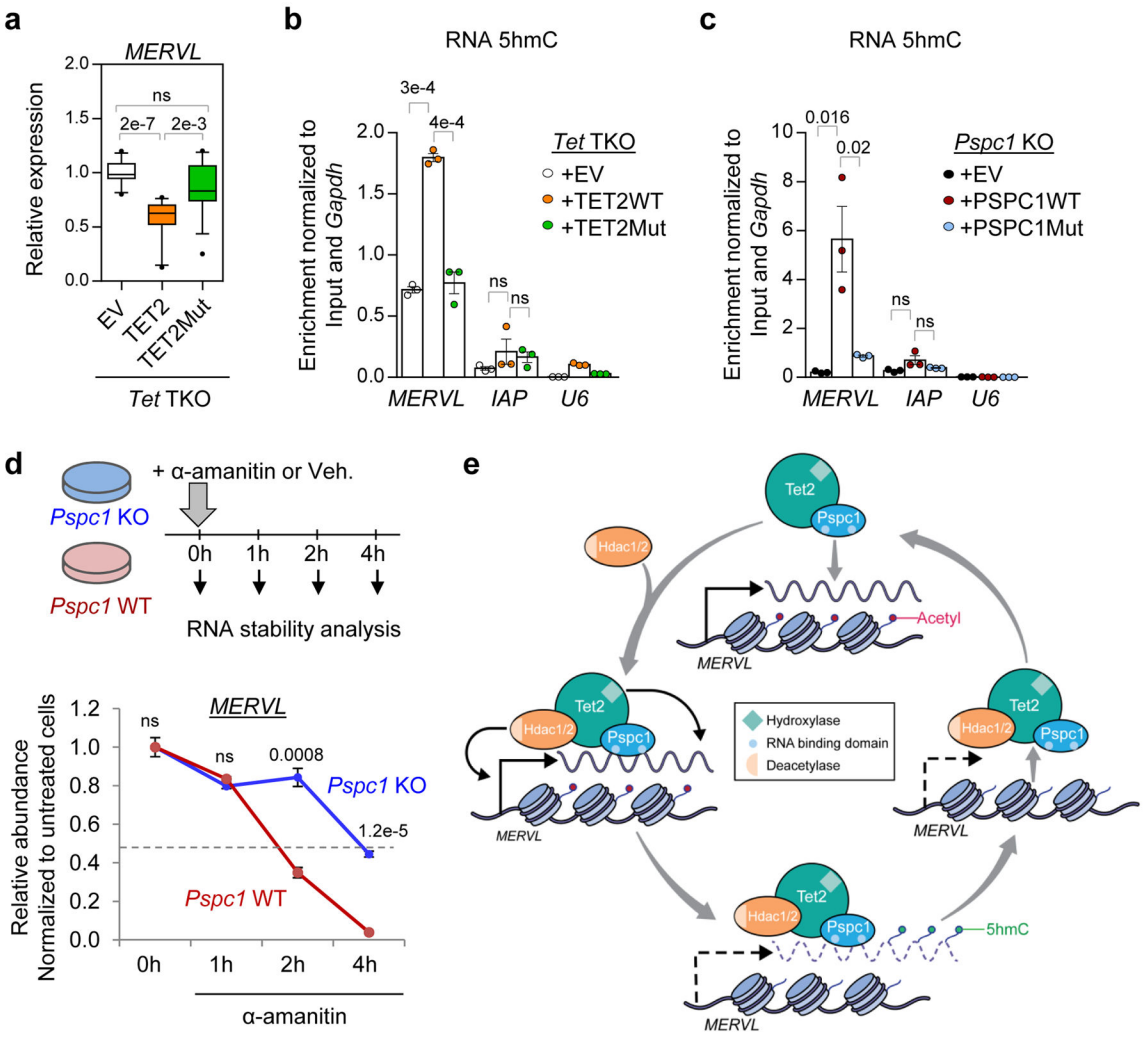

Figure 4. PSPC1 and TET2 silence *MERVL* transcriptionally and post-transcriptionally

a, *MERVL* expression in *Tet1/2/3* triple knock-out (*Tet* TKO) ESCs rescued with an empty vector (+EV), a wild-type (+TET2WT), or a catalytic mutant (+TET2Mut) TET2. Center line, median; box and whisker plots: ± 10th–90th percentile range. Data are from 5 independent experiments (n=14 total technical replicates for each rescue). Two-tailed Student's *t*-test was applied. ns, not significant. **b–c**, *MERVL* and *IAP* enrichment, compared to 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, or a mutant TET2/PSPC1. Data are presented as mean ± s.e.m. (n=3 independent experiments). Two-tailed Student's *t*-test was applied. ns, not significant. **d**, (Top) Schematic of the protocol used for inhibition of transcription with α-Amanitin for RNA stability assay. (Bottom) Relative abundance of *MERVL* RNA in *Pspc1* WT and *d* KO ESCs after transcriptional inhibition for 1, 2, or 4 hours with α-Amanitin. Data are normalized to untreated cells at time 0 h (Vehicle without treatment). Error bars indicate s.e.m. (n=3). Two-tailed Student's *t*-test was applied. ns, not significant. **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 HDAC1/2 deacetylate histones at the chromatin level leading to transcriptional repression of the *MERVL* loci. Transcriptional and posttranscriptional 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 PSPC1-TET2-HDAC1/2 for transcriptional and posttranscriptional control of *MERVL* and coordinated gene expression. Illustration by Jill Gregory. Printed with permission of ©Mount Sinai Health System.