Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2024 January 04.

Published in final edited form as:

Nat Genet. 2023 August ; 55(8): 1324–1335. doi:10.1038/s41588-023-01452-5.

Crosstalk between RNA m⁶A and DNA methylation regulates transposable element chromatin activation and cell fate in human pluripotent stem cells

Tongyu Sun[1], **Yueyuan Xu**[1,2,3], **Yu Xiang**[1,2], **Jianhong Ou**[2], **Erik J Soderblom**[1,4], **Yarui Diao**[1,2,3,5,6,7,#]

[1]Department of Cell Biology, Duke University Medical Center, Durham, NC 27710, USA

[2]Duke Regeneration Center, Duke University Medical Center, Durham, NC 27710, USA

[3]Center for Advanced Genomic Technologies, Duke University, Durham, NC 27708, USA

[4]Proteomics and Metabolomics Shared Resource, Duke Center for Genomic and Computational Biology, Duke University, Durham, North Carolina 27708, USA

[5]Duke Cancer Institute, Duke University Medical Center, Durham, NC 27710, USA

[6]Department of Orthopaedics Surgery, Duke University Medical Center, Durham, NC 27710, USA

[7]Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA

Abstract

Transposable elements (TEs) are parasitic DNA sequences accounting for over half of the human genome. Tight control of the repression and activation states of TEs is critical for genome integrity, development, immunity, and diseases, including cancer. However, precisely how this regulation is achieved remains unclear. Here, we develop a targeted proteomic proximity labeling approach to capture TE-associated proteins in human embryonic stem cells (hESCs). We find that the RNA N6-methyladenosine(m^6A)-reader, YTHDC2, occupies genomic loci of the primate-specific TE, LTR7/HERV-H, specifically through its interaction with m^6A-modified HERV-H RNAs. Unexpectedly, YTHDC2 recruits the DNA 5-methylcytosine(5mC)-demethylase, TET1, to remove 5mC from LTR7/HERV-H and prevent epigenetic silencing. Functionally, the YTHDC2/LTR7-axis inhibits neural differentiation of hESCs. Our results reveal both an underappreciated crosstalk between RNA m^6A and DNA 5mC, the most abundant regulatory modifications of RNA and DNA in eukaryotes, and the fact that in hESCs this interplay controls TE activity and cell fate.

[#]Correspondence: yarui.diao@duke.edu.

AUTHOR CONTRIBUTIONS

T.S. and Y.D. conceived the idea of this study. T.S. performed all the experiments and data analysis with the help from Y.Xu, Y.Xiang, and J.O. on bioinformatics. T.S. and E.S. performed LC-MS/MS proteomics analysis. T.S. and Y.D. wrote the paper.

Competing interests

All authors declare no competing interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

INTRODUCTION

Transposable elements (TEs) are mobile DNA elements derived from ancient viral infections accounting for more than half of our genomic sequences in humans[1-4]. To maintain the integrity of the host genome, most TEs in the mammalian genome are silenced by repressive heterochromatic marks [5,6]. On the other hand, through evolution, some TE sequences have been co-opted as cis-regulatory elements controlling spatial-temporal gene expression in numerous biological processes[7]. Therefore, a tight balance between the transcriptionally silenced and activated chromatin states is critical for genome integrity and spatial-temporal gene regulation. However, how this regulation is achieved is still not fully elucidated.

Among eukaryotes, the hypermethylation of DNA sequences is a common strategy utilized to silence TEs[5,6]. In mammals, the most common DNA methylation, 5-methylcytosine (5mC) of CpG dinucleotides, is deposited and maintained by DNA methyltransferases including DNMT3A/3B and DNMT1[8,9]. The family of Ten-eleven translocation (TET) methylcytosine dioxygenase, including TET1-3, can remove DNA methylation through the stepwise oxidation of 5mC to 5-hydroxymethycytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[10,11]. Recently, several studies have revealed a critical role for the posttranscriptional modification of RNA by N6-methyladenosine (m^6A) in regulating TE silencing through modulating histone modifications[12-18]. m^6A is installed by the methyltransferase complex, can be removed by “erasers”, and is recognized by the “reader” proteins, including YTHDC1-2 and YTHDF1-3 that contain a highly conserved m^6A-binding, YTH-domain[18,19]. Recent studies show that m^6A-modified “chromatin-associated regulatory RNAs” (carRNAs), including RNAs transcribed from promoters, enhancers, and TE loci, are bound by YTHDC1, which either induces degradation of TE transcripts or recruits histone modifiers to induce epigenetic silencing of TE loci by controlling histone modifications[20-22]. In human cancer cell lines, it has also been shown that the RNA m^6A reader FXR1 recruits TET1 to regulate chromatin accessibility and transcription of enhancers and promoters[23]. To date, however, whether and how the epitranscriptomics modifications interplay with DNA methylation machinery in regulating TE activity in mammals remains an intriguing and unanswered question.

Here, we develop CARGO-BioID, a CRISPR-based TE-centric proteomics approach, to identify the proteins associated with the genomic loci of the primate-specific TE, LTR7/HERV-H, in human embryonic stem cells (hESCs). We find that the RNA m^6A-reader, YTHDC2[24], is recruited to the LTR7 DNA regions through its interaction with m^6A-modified HERV-H transcripts; and engages TET1[11] to prevent the epigenetic silencing of LTR7/HERV-H sequences via DNA demethylation. Functionally, the YTHDC2/LTR7-axis inhibits the neuronal fate of hESCs. These results uncover mechanisms controlling TE activity and hESC fate, as well as crosstalk between RNA m^6A and DNA methylation.

RESULTS

CARGO-BioID identifies TE-associated proteome

The large copy numbers and highly repetitive sequences of TEs, as well as their locus-specific mutations or indels, pose technical challenges for identifying their regulatory

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

complexes (Fig. 1a). To overcome this problem, we combined the multiplexed CRISPR targeting approach "Chimeric Array of gRNA Oligos" (CARGO)[25,26] with biotin proximity labeling (BioID) to develop a CRISPR-based TE-centric proteomics method which we termed CARGO-BioID. The CARGO construct enables the simultaneous expression of up to 15 sgRNAs to target TE sequences in the native chromatin environment[25,26]. We fused the catalytic inactive "dead" Cas9 (dCas9) that can bind to but not cleave DNA with biotin ligase TurboID[27,28], that is predicted to biotinylate proteins proximate to dCas9-TurboID and the DNA sequences to which it is directed[29-32]. The proteins are then purified by streptavidin selection for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis (Fig. 1b). By comparing the biotinylated proteins purified from the dCas9-TurboID-expressing cells with CARGO versus those expressing the control non-targeting sgRNAs, we expect to identify proteins that are enriched specifically at targeted TE loci.

We selected the primate-specific TE, LTR7/HERV-H, and designed 15 sgRNAs to target these loci (Fig. 1a, Extended Data Fig. 1a; Supplementary Table 1) for oligonucleotide sequences). In hESCs, the LTR7s sequences are enriched with active histone marks such as histone H3K4me1, H3K4me3, and H3K27ac (Extended Data Fig. 1b). By contrast, there is less representation of these active chromatin marks in more differentiated lineages like mesendoderm, neural progenitors, and trophoblast[33,34] (Extended Data Fig. 1c). ChIP-seq showed that this CARGO construct directs dCas9-TurboID to 1,815 copies of LTR7 (73.5% of the total) in hESCs (Fig. 1c, 1d). Since the binding of dCas9-TurboID protein to DNA may lead to the risk of disrupting the endogenous interactome, including transcription factors (TFs), we designed sgRNAs to avoid the predicated TF binding motifs in LTR7s. With this approach, dCas9-TurboID/CARGO did not affect the RNA levels of HERV-H and its retroviral genes, Gag, Pol, and Env, or the epigenetic modification of LTR7s compared to the cells expressing dCas9-Turbo-ID/control sgRNAs (Extended Data Fig. 1e, 1f). These data suggest that the native interactome key for LTR7/HERV-H transcriptional and epigenetic activities was not disrupted by CARGO-BioID.

After three hours of exogenous biotin treatment, we observed distinct biotinylation patterns of nuclear proteins in hESCs stably expressing dCas9-TurboID CARGO sgRNAs, versus cells expressing control sgRNAs (Fig. 1e, Extended Data Fig. 1g, 1h). Mass spectrometry identified 144 candidate proteins that were highly enriched at LTR7 loci (Fig. 1f, Supplementary Table 2). LTR7s in hESCs serve as enhancers, promoters, pluripotency factor binding sites, and topologically associating domains (TAD) boundaries [34-37]. Correspondingly, the putative LTR7-associated proteins are involved in histone modification, chromosome organization, transcription initiation, mRNA processing, and pluripotency factor signaling (Fig. 1g, Supplementary Table 2).

YTHDC2 binds to LTR7s and limits their silencing

Intriguingly, several proteins involved in RNA methylation were enriched on LTR7 loci, including YTHDC2, an RNA m^6A-reader that contains the m^6A-binding YTH domain[18,24] (Fig. 1f, 1g, Extended Data Fig. 2a). We conducted ChIP-seq using an antibody against YTHDC2, and identified 1,307 LTR7 elements that overlap with YTHDC2 ChIP-seq peaks. Permutation test revealed that LTR7 is the second most enriched (empirical p-value < 0.001)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

repeat family in the YTHDC2-occupied sequences (Extended Data Fig. 2b, Supplementary Table 3). Compared to the 1,162 LTR7s without YTHDC2 binding, the YTHDC2-occupied LTR7s exhibit higher levels of the active chromatin mark H3K27ac and chromatin accessibility but lower levels of silencing mark H3K9me3 (Fig. 2a, 2b). In addition to the LTR7s, the YTHDC2-occupied sequences are associated with active chromatin marks, binding of the transcription activator, BRD4[38,39], and a minimal representation of the repressive heterochromatin mark (Extended Data Fig. 2c). By taking advantage of the 15 chromatin states of hESCs defined by ChromHMM[40,41], we found that the most enriched chromatin states across all the YTHDC2 ChIP-seq peaks include active enhancers (Enh), bivalent enhancers (EnhBiv), and the sequences flanking bivalent promoters and enhancers (BivFlnk) (Extended Data Fig. 2d, 2e, Supplementary Table 3).

To interrogate YTHDC2 function, we generated two independent knock-out (KO) hESC clones using CRISPR/Cas9 editing (Extended Data Fig. 2f). Interestingly, RNA levels of HERV-H, Gag, Pol, and Env, were significantly reduced in YTHDC2 KO cells (Fig. 2c). We examined the expression of the pluripotency markers POU5F1, SOX2, and NANOG at RNA and protein levels (Extended Data Fig. 2g, 2h), and assessed the RNA decay rate of HERV-H RNAs (Extended Data Fig. 2i). The results suggest that the reduction of HERV-H RNAs in YTHDC2 KO hESCs is caused by reduced transcription, rather than loss of pluripotency or RNA degradation. Indeed, using 5-Ethynyluridine (EU) to label newly synthesized RNAs[42] indicated that there are far fewer (~20%) nascent HERV-H transcripts in YTHDC2 KO hESCs than in control cells (Fig. 2d).

Since the phenotypes of two independent YTHDC2 KO H1 hESC clones were similar (Fig. 2c, 2d; Extended Data Fig. 2f-2i), we selected one for further in-depth analysis. We performed a rescue experiment with re-expression of wild-type YTHDC2 cDNA (Extended Data Fig. 2j). This restored the expression of HERV-H transcripts (Fig. 2e), indicating only limited, if any, effects of potential Cas9 off-target mutations. Recent studies have shown that YTHDC2 can function as an RNA helicase in the context of spermatogenesis[43-45]. We overexpressed in YTHDC2 KO cells two mutant forms of YTHDC2 that lack either helicase activity (YTHDC2^{E317Q})[43-45] or m^{6}A binding capability (YTHDC2$^{\Delta YTH}$). Interestingly, YTHDC2^{E317Q} could partially rescue HERV-H expression, whereas YTHDC2$^{\Delta YTH}$ had no detectable rescue activity, despite the higher protein expression level (Fig 2e, Extended Data Fig. 2j). These data indicate that the m^{6}A-binding capacity of YTHDC2 is indispensable for its regulation of HERV-H transcription, and that its helicase activity also contributes.

In eukaryotic cells, the silenced TEs are associated with the deposition of repressive histone marks such as H3K9me3 and the loss of the active mark H3K27ac [5]. From RNA-seq and ChIP-seq data, we observed significantly decreased HERV-H expression (Extended Data Fig. 2k, 2l), gain of H3K9me3 and loss of H3K27ac on the LTR7/HERV-H genomic loci in the absence of YTHDC2 (Fig. 2f-2i). We identified 1,552 and 1,573 LTR7s exhibiting reduced H3K27ac and increased H3K9me3 signals, respectively (Fig 2j, fold-change > 1.2, p-value < 0.05). These YTHDC2-regulated LTR7s, totaling 1,661 sequences, significantly overlap with the 1,307 YTHDC2-bound LTR7s (Fig 2j, Chi-square test, p-value < 0.0001, Fig. 2k). Taking all these results into consideration, we conclude that YTHDC2 occupies the LTR7/HERV-H genomic loci and prevents their transcriptional and epigenetic silencing.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The YTHDC2/LTR7 axis regulates neuronal fate commitment

Previous studies have indicated roles for HERV-H RNAs in the regulation of pluripotency maintenance [46,47] and neuronal differentiation[48,49]. By contrast, a recent study showed that shRNA-mediated knockdown of pan HERV-H RNAs neither induced differentiation of human pluripotent stem cells, nor altered the expression of pluripotency markers in the self-renewal conditions [50]. In YTHDC2 KO hESCs, despite a significant reduction of HERV-H RNAs, there is no change in the expression of pluripotency markers in self-renewal medium (Extended Data Fig. 2g, 2h). However, Gene Ontology (GO) analysis did reveal differential expression of genes related to nervous system development (Extended Data Fig. 3a). YTHDC2 ChIP-seq also showed enriched binding of YTHDC2 in sequences near genes related to nervous system development (Extended Data Fig. 3b). These results suggest potential roles of YTHDC2 in regulating neuronal fate.

To test the impact of YTHDC2/LTR7 on the neurogenic potential of hESCs, we used two complementary experimental systems: a lineage-specific neuronal cell differentiation assay, based on a neural induction medium that contains SMAD2/3 inhibitors[51], and an embryoid-body (EB) formation assay that allows spontaneous multi-lineage differentiation of hESCs. In the neural-induction medium, YTHDC2 KO cells express ~6-8 fold higher levels of neuronal marker genes than wild-type cells [52,53] (Fig. 3a, 3b, Extended Data Fig. 3c). The YTHDC2 mutant EBs also displayed significantly higher levels (~2-3 fold) of these markers (Fig. 3c, Extended Data Fig. 3c). In contrast, the expression of other lineage markers (SOX17, GATA6, CDX2, HAND1, GSC)[54] were significantly downregulated in the mutant EBs, although the early mesoderm marker TBX1 (also known as T or Brachyury) was upregulated by ~2 fold (Extended Data Fig. 3d). Collectively, these data indicate that YTHDC2 normally inhibits the ability of hESCs to differentiate toward neuronal lineages and maybe required for differentiation beyond the primitive streak stage[54].

To determine whether LTR7 sequences themselves influence the fate of hESCs, we induced epigenetic silencing of LTR7 sequences *en masse* by co-expression of the epigenetic repressor dCas9-KRAB[55] together with LTR7 CARGO in hESCs (CRISPR interference, CRISPRi)[25,26]. We focused on the 1,815 LTR7stargeted by dCas9/CARGO (Fig 1c), and found that these CARGO-targeted LTR7 regions exhibited global loss of the active chromatin mark H3K27ac and gain of the repressive mark H3K9me3(Fig. 3d; Extended Data Fig. 3e, 3f). Concomitantly, this perturbation caused a significant reduction of HERV-H transcripts (Fig. 3e, Wilcoxon $P < 2.2e\text{-}16$; Extended Data Fig. 3g, 3h). Significantly, key observations indicate that LTR7 CRISPRi phenocopies YTHDC2 KO in hESCs. First, LTR7-silenced hESCs were maintained in a pluripotent state in self-renewal medium (Extended Data Fig. 3i - 3k). Second, they expressed a higher level (~3-6 fold) of neuronal markers in both hESC-derived neuronal cells and EBs (Fig. 3f-3h; Extended Data Fig. 3l). Taken together, our results suggest a model in which the YTHDC2/LTR7 axis is dispensable for pluripotency maintenance but negatively regulates the switch from pluripotency to the neuronal fate.

To identify the key effectors that are directly regulated by the YTHDC2/LTR7 axis and influence the fate of hESCs, we examined genes and long non-coding RNAs (lncRNAs) with significantly lower levels in YTHDC2 KO hESCs, LTR7-silenced hESCs, and hESC-derived

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

neuronal cells compared to control pluripotent hESCs (Fig. 3i, DESeq2 $\log_2$ fold change > 1, p-value < 0.05, Supplementary Table 4). This comparison identified 12 genes and lncRNAs, including LINC-ROR, an LTR7/HERV-H-derived lncRNA initially identified as a regulator involved in human iPSC reprogramming[56]. RT-qPCR indicated that LINC-ROR levels are indeed significantly reduced in YTHDC2 KO and in LTR7-silenced hESCs (Extended Data Fig. 3m, 3n). LINC-ROR has been shown to maintain hESC self-renewal by acting as a microRNA sponge to trap miR-145, which would otherwise target multiple core pluripotency factors including POU5F1, NANOG, and SOX2[57,58]. However, results from other studies showed that shRNA-mediated knockdown of LINC-ROR did not alter the expression of NANOG in human pluripotent stem cells[48]. We employed an additional approach to determine the causal role of LINC-ROR in pluripotency maintenance. CasRx is an RNA-guided RNA-targeting type VI CRISPR-Cas that can induce targeted RNA degradation in a gRNA-dependent manner[59,60]. Compared to the 21-nucleotide siRNAs used in previous studies, CasRx requires 30bp sgRNA which increases the specificity of knockdown and reduces the potential risk of off-target effects. Using two independent gRNAs, CasRx reduced the LINC-ROR RNA level to 10-30% but did not alter the expression of pluripotency markers (Extended Data Fig. 3o, 3p). Next, we overexpressed LINC-ROR in the YTHDC2 KO and LTR7 silenced hESCs. This manipulation reduced the expression of neuronal markers in these cells when cultured in neural induction medium (Fig. 3a, 3b, 3f, 3g). Collectively, we conclude that neither the epigenetic activity of LTR7 DNA, nor the expression of HERV-H RNAs including LINC-ROR, is required for pluripotency maintenance in self-renewal culture conditions. Significantly, LINC-ROR is one of the effector molecules that is directly regulated by the YTHDC2/LTR7 axis and restricts the neural differentiation potency of hESCs (Fig. 3j).

The m^{6}A-modified HERV-Hs interact with YTHDC2 and LTR7 DNA

HERV-H transcripts are expressed from their LTR7 promoters, supporting the feasibility that YTHDC2 interacts with m^{6}A-modified HERV-H RNAs in the vicinity of LTR7 DNA (Fig. 4a). To test this idea, we analyzed published m^{6}A RNA immunoprecipitation sequencing (RIP-seq)[61] data and found that HERV-H transcripts are indeed extensively modified by m^{6}A (Fig. 4b). Next, we conducted RIP-seq and Enhanced cross-linking and immunoprecipitation (eCLIP)-seq[62] using the antibody against YTHDC2. Results from both assays showed that HERV-H RNAs are extensively bound by YTHDC2 (Fig. 4c; Extended Data Fig. 4a-4c). Since the results from RIP-seq and eCLIP-seq were concordant, we used YTHDC2 RIP-seq for the rest of this study. We identified 6,172 YTHDC2 RIP-seq peaks in hESCs (p-value < 1e-5), including 6,043 peaks (97.9%) overlapping with the annotated repeat regions of the human genome. Permutation test showed that the HERV-H and LTR7-derived RNAs are the two most enriched repeat RNAs in the YTHDC2 RIP-seq peaks (empirical p-value < 0.001, Fig. 4d, Supplementary Table 3). Together, these data indicate that the HERV-H RNAs undergo m^{6}A modification and are enriched for YTHDC2 binding in hESCs.

Next, we asked whether YTHDC2 simply binds to many highly expressed RNAs, or specifically interacts with HERV-H transcripts in a motif-dependent manner. To probe this question, we identified 1,119 mRNAs associated with enriched YTHDC2 binding ($\log_2$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fold change > 1, p-value < 0.05). Compared to the rest of 12,326 genes with lower levels of YTHDC2 enrichment, the expression of the two groups of genes showed no significant difference (Extended Data Fig. 4d, Wilcoxon test, p-value = 0.92). Further analysis revealed that a CCGGACGG motif, which harbors the canonical m^{6}A consensus motif RRACH[24], is significantly enriched in the YTHDC2-occupied mRNA sequences (Extended Data Fig. 4e, p-value = 1e-6). Particularly, we identified a significantly enriched canonical m^{6}A motif GGACU in the YTHDC2-occupied HERV-H RNAs from RIP-seq (p-value = 1e-131) and eCLIP-seq (p-value = 1e-116) (Fig. 4e). To further assess whether YTHDC2 directly interacts with m^{6}A-modified HERV-H RNAs, we conducted RNA pulldown and electrophoretic mobility shift assay (EMSA) using a pair of biotin-conjugated RNA oligonucleotides that contain a 30-base conserved HERV-H sequence flanking the GGACU motif, with or without m^{6}A-modification in this motif (Supplementary Table 1). Compared to the unmethylated RNA probe, the methylated probe showed a much stronger binding capacity with YTHDC2 protein prepared from hESC nuclear lysates (Fig 4f). For EMSA, the recombinant protein of the YTH-domain of human YTHDC2 (1218-1488 amino acids) only binds to the methylated RNA probe, in a protein dosage-dependent manner, but not to the unmethylated one (Fig 4g, Extended Data Fig. 4f). Taken together, we conclude that YTHDC2 directly binds to the m^{6}A-modified HERV-H RNAs in a motif-dependent manner.

To interrogate the physical interaction between HERV-H RNA and LTR7 DNA, we analyzed published hESC MARGI (MApping RNA-Genome Interactions) data, which quantitatively reports genome-wide RNA-chromatin interactions[63]. This revealed that HERV-H RNAs indeed form interactions with their local LTR7s flanking HERV-H DNA, rather than with solo LTR7s (Fig. 4h, 4i). The interaction frequency between HERV-H RNA and LTR7 DNA significantly correlates with HERV-H transcription levels (Pearson correlation coefficient R = 0.76, p-value < 2.2e-16). Interestingly, HERV-H RNAs exhibit stronger interaction with their 5'- and 3'- flanking LTR7 regions compared to their gene body (Extended Data Fig. 4g). The HERV-H RNAs occupied by YTHDC2 or modified by m^{6}A also showed stronger interaction with their promoter LTR7s compared to those without YTHDC2 binding or m^{6}A modification (Extended Data Fig. 4h, 4i). These results suggest that the interactions between HERV-H transcripts and LTR7 loci are likely to be dynamically regulated and stabilized by YTHDC2 and m^{6}A modification, and do not merely represent a nascent interaction between newly transcribed RNA and template DNA.

The m^{6}A-modified HERV-Hs control YTHDC2/LTR7 interaction

To determine the role of m^{6}A modification of HERV-H in YTHDC2/LTR7 axis, we treated hESCs with STM2457[64], a highly potent and selective catalytic inhibitor of the m^{6}A writer METTL3. STM2457 treatment did not induce the degradation of HERV-H RNAs or METTL3 protein, but reduced the m^{6}A level of all HERV-H RNAs by ~80% after two days (Extended Data Fig. 5a - 5c). Significantly, STM2457 treatment disrupted the binding of YTHDC2 on both HERV-H transcripts and LTR7 DNA genome-wide (Fig. 5a, 5b, Extended Data Fig. 5d, 5e), and led to a global gain of repressive histone mark H3K9me3 on LTR7/HERV-H loci (Fig 5c). These data suggest that the RNA m^{6}A modification regulates epigenetic activity and YTHDC2 occupancy of LTR7/HERV-H sequences.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To address the potential concern that the effect of STM2457 treatment is indirect, we conducted a sequence-specific m^{6}A editing experiment using LINC-ROR as a representative locus (Fig. 5d). To edit m^{6}A modifications of LINC-ROR, we fused the m^{6}A-eraser ALKBH5 with dead CasRx (dCasRx-ALKBH5) [59,60], which can remove m^{6}A from LINC-ROR [65,66] with two gRNAs targeting the m^{6}A-modified sequence (Fig. 5d, blue arrow). We found that dCasRx-ALKBH5, but not its mutant form dCasRx-dALKBH5 (dead ALKBH5) lacking m^{6}A-eraser activity, can effectively reduce the extent of m^{6}A modification of LINC-ROR (Fig. 5e). Significantly, this manipulation resulted in a markedly decrease of YTHDC2 binding on LINC-ROR RNA and its LTR7 DNA loci (Fig. 5f, 5g), as well as a significant reduction of LINC-ROR RNA (Fig. 5h). These changes were not due to the binding of dCasRx-dALKBH5, nor to the destabilization of LINC-ROR transcripts (Fig. 5h, Extended Data Fig. 5f). Notably, reduction of m^{6}A in LINC-ROR RNA also resulted in a significant gain of the repressive marks H3K9me3 on the LINC-ROR locus (Fig. 5i), indicating epigenetic silencing. By contrast, in the absence of YTHDC2, removal of m^{6}A from LINC-ROR did not induce a further reduction of LINC-ROR or gain of H3K9me3 modification on its LTR7 loci (Extended Data Fig. 5g-5i). Collectively, our data support a model in which the m^{6}A modification of HERV-H RNAs plays a causal role in preventing the epigenetic silencing of LTR7/HERV-H genomic loci by recruiting the RNA m^{6}A reader YTHDC2 (Fig. 4a).

YTHDC2 recruits TET1 to demethylate LTR7 DNA

From CARGO BioID data, both TET1 and TET3 were identified as candidate proteins enriched on LTR7 loci (Fig. 1f, 1g). A previous study showed that in hESCs genetic depletion of TET1, but not TET2/3, leads to a marked decrease of global 5hmC by ~70%[52], suggesting that TET1 plays a primary role in controlling 5hmC-mediated DNA demethylation. Thus, we asked whether TET1 proteins play a role in the YTDHC2-mediated regulation of LTR7/HERV-H via DNA demethylation. To enable specific immunoprecipitation of TET1 protein, in the parental H1 hESC clones with intact or mutant YTHDC2 alleles, a V5 epitope was tagged to the N-terminus of the endogenous TET1 protein (TET1-V5) using a CRISPR/Cas9-mediated knock-in strategy[67] (Extended Data Fig. 6a-6c). The endogenous YTHDC2 and TET1 proteins interact with each other in hESCs, and such interaction relies on the TYH-domain of YTHDC2, the presence of RNA molecules, and the m^{6}A-modification of RNAs (Fig. 6a, Extended Data Fig. 6d-6f). Furthermore, we found that TET1 was significantly enriched on the LTR7s loci occupied by YTHDC2, and vice versa (Fig. 6b, Extended Data Fig. 6g). The TET1 and YTHDC2 ChIP-seq signals were highly correlated with each other across all the LTR7 loci (Extended Data Fig. 6h, Pearson correlation coefficient = 0.61, $P < 0.0001$). Importantly, the binding of TET1 on LTR7 sequences was markedly decreased in YTHDC2 KO cells (Fig. 6c), accompanied by a significant gain of DNA 5mC modification and a marked decrease of 5hmC modification surrounding the LTR7/HERV-H loci (Fig. 6d - 6f). These data strongly suggest that TET1 is recruited by YTHDC2 to the LTR7/HERV-H loci where it induces DNA CpG demethylation of TE sequences.

Next, we analyzed the published RNA-seq and methyl-capture sequencing data generated from wild-type and TET1 KO H1 hESCs[67]. We found that TET1 KO indeed led to a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

significant reduction of HERV-H expression (Fig. 6g, Wilcoxon p < 2.2e-16; Extended Data Fig. 6i), accompanied by a gain of DNA CpG methylation in LTR/HERV-H sequences (Fig. 6h). Notably, the reduced HERV-H expression in the YTHDC2 KO hESCs can be rescued by overexpression of LTR7 CARGO together with TET1CD, a fusion protein containing dCas9 and the catalytic domain (CD) of TET1 that can convert DNA 5mC to 5hmC on its targeted chromatin regions[68] (Fig. 6i). Collectively, these data establish the causal role of TET1 in controlling LTR7/HERV-H activity in hESCs and show that the silencing of LTR7/HERV-H in the absence of YTHDC2 is due at least in part to the loss of TET1 binding on LTR7/HERV-H DNA.

HERV-Hs are required for TET1 binding to LTR7 DNA

Lastly, we asked whether m^6A modification of HERV-H RNA is required for TET1 engagement on LTR7/HERV-H DNA and conversion of 5mC to 5hmC. We found that upon targeted removal of m^6A from LINC-ROR RNA (Fig. 5d, 5e), the binding of TET1 on the two LTR7 loci of LINC-ROR was markedly decreased (Fig. 6j) in the cells with intact YTHDC2 expression. By contrast, erasing m^6A from LINC-ROR in the absence of YTHDC2 did not lead to further reduction of TET1 binding on the LINC-ROR locus (Extended Data Fig. 6j). These data strongly suggest that the recruitment of TET1 to the LTR7/HERV-H DNA requires both m^6A modification of HERV-H RNA and YTHDC2. Notably, the loss of TET1 binding on the LINC-ROR locus is accompanied by a concordant gain of DNA 5mC and loss of DNA 5hmC across LINC-ROR sequences on both LTR7 loci and the gene body of LINC-ROR (Fig. 6k, Extended Data Fig. 6k, 6l). These data support a model in which m^6A modification of HERV-H transcripts recruits TET1 to prevent epigenetic silencing of LTR7/HERV-H loci, via a DNA demethylation mechanism.

DISCUSSION

In eukaryotes, RNA m^6A modification is the most abundant epitranscriptome modification[18], and the DNA cytosine 5mC and 5hmC marks are widely distributed throughout the genome[69,70]. A recent study has reported the crosstalk between m^6A and DNA methylation, mediated by m^6A reader FXR1 and TET1, in the regulation of chromatin accessibility and transcription in cancer cell lines[23]. However, the biological significance of the phenomenon remains largely unknown in physiological and pathological conditions[71]. Here, we demonstrate that the crosstalk between YTHDC2 and TET1 is critical for TE activity and the hESC fate. Our findings significantly expand our knowledge of the regulatory crosstalk between the epitranscriptome and the epigenome.

Despite the critical role of TEs in development and diseases, it has been technically challenging to identify the effectors controlling the activity of endogenous TE sequences. Our study demonstrates that CARGO-BioID is a powerful technology to identify key regulators of TE activity. Considering the critical roles of TEs in regulating development, immunity, aging, and cancer[1-4], we expect that CARGO-BioID will become a powerful technology to further our understanding of TE function and regulation in many biological processes and human diseases. Extended discussion of this study can be found in the section of Supplementary Information.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

METHODS

Cell culture and differentiation

H1 hESC line (WiCell, NIHhESC-10-0043) was maintained in mTseR plus medium (Stemcell Tech, 100-0276) and detached using 5 mM EDTA/PBS or Accutase (Sigma, A6964). The TET1-V5 H1 hESC line with intact YTHDC2 expression was a kind gift from Dr. Danwei Huangfu (Sloan Kettering Institute, New York, NY, USA)[67]. The HEK293T cell line was obtained from ATCC. HEK293T cells were cultured in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10% FBS and 1% penicillin-streptomycin at 37 °C with 5% CO_2. All cells were tested to confirm that there was no mycoplasma contamination. To induce embryoid body formation, H1 cells were dissociated and aggregated in EB medium in 6-well ultra-low attachment plates (Corning). Medium was changed daily. Neural cells were induced using STEMdiff™ SMADi Neural Induction Kit (Stemcell Technologies, 08581) following monolayer cell culture protocol.

Antibodies

All antibodies used in this study can be found in the supplementary file (Supplementary Table 1).

Plasmid construction

To construct the plasmids, we employed three different methods: restriction enzyme-based cloning, Gibson Assembly, and Golden Gate Cloning. To generate the dCas9-TurboID construct, the KRAB domain on Lenti-dCas9-KRAB-blast (Addgene, 89567) was replaced with TurboID sequence (Addgene, 107169) using restriction enzyme-based cloning. For sgRNA design targeting LTR7s, we identified transcription factor binding motifs at LTR7s, used the Flashfly tool[72] to obtain candidate sgRNAs, manually filtered sgRNAs to exclude those with high off-target ratios, and ranked the sgRNAs based on their on-target numbers. The CARGO construct was created using Golden Gate Cloning by ligating together each individual gRNA expression fragment with BsmB1 cutting sites at both ends. The dCasRX-ALKBH5 and dCasRX-dALKBH5 constructs were generated using Gibson Assembly, where dCasRX fragments (Addgene, 109050), ALKBH5 fragments (Addgene, 134783) or dALKBH5 fragments (Addgene, 134784), and backbones (Addgene, 89567) were PCR amplified and assembled using Gibson Assembly Master Mix (NEB, E2611L). To produce the lenti-YTHDC2 construct, the Cas9 sequence (Addgene, 52962) was replaced with full-length YTHDC2 (Horizon, MHS6278-213245943) using Gibson Assembly, with point mutations introduced by PCR amplification and validated by Sanger sequencing.

Proximity labeling assay and Quantitative LC–MS/MS analysis

In proximity labeling assays, cells expressing dCas9-TurboID and gRNAs were treated with 500 μM biotin (Sigma, B4501) for varying time points (0-4 h) as previously described[27]. The nuclear biotinylated proteins were extracted with RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM sodium chloride,1.0% Igepal CA-630, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate) and subjected to immunoprecipitation and Western blot analyses.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For quantitative LC-MS/MS analysis, 10 million hESCs were used for each replicate and biotinylated protein complexes were eluted with the elution buffer (2% SDS, 25 mM Tris pH 7.4, 5 mM DTT, and 5 mM free biotin). All enriched proteins (FC > 1.5, p-value < 0.05) were used for DAVID GO analysis[73,74] and ingenuity pathway analysis (IPA) with default parameters[75].

Generation of knockout cell lines

Guide RNAs (sgRNAs) targeting YTHDC2 were selected from Human CRISPR Knockout Pooled Library (GeCKO v2)[76]. YTHDC2 targeting gRNAs were cloned into lentiCRISPRv2 vectors to make viruses. hESCs were infected with viruses or transiently transfected with plasmids using FuGENE HD Transfection Reagent (Promega, E2311). Knockout cell lines were sequenced to validate genotypes.

Generation of knock-in cell lines

To create knock-in cell lines, V5-P2A-puromycin fragments flanked by homology arms were ligated to TOPO vectors using the Zero Blunt™ TOPO™ PCR Cloning Kit (Thermo Fisher Scientific, 450031) to generate donor constructs. Cells were then transfected with 750 ng of the donor constructs and 250 ng of gRNA-expressing vectors in 24-well plates. Cells were selected using 0.5 μg/μl Puromycin (Thermo Fisher Scientific, A1113802) for three days. Single clones were genotyped and sequenced, and TET1 proteins were immunoprecipitated from the transgenic clones to validate V5 epitope expression.

m^{6}A editing by dCasRx-ALKBH5

To target specific m^{6}A sites on RNA, cells were co-transfected with gRNAs against m^{6}A peak regions and dCasR-ALKBH5 using the FuGENE HD Transfection Reagent. (Promega, E2311) Negative controls included cells transfected with scramble gRNA or dCasRx-dALKBH5. Following selection with Puromycin or Blasticidin (Thermo Fisher Scientific, A1113903) for 5-7 days, total RNA was extracted and subjected to qRT-PCR and m^{6}A RIP-qPCR assays to evaluate changes in RNA levels and m^{6}A modification. The m^{6}A peak-negative region of the edited RNA was used as an internal control for m^{6}A level normalization.

Co-immunoprecipitation and western blot

Co-IP assay was performed as previously described[77]. We used 100 million hESCs to detect the interaction between YTHDC2 and TET1. hESCs were crosslinked in a solution of 200 μg/mL DSP (Thermo Fisher Scientific, 22585) at room temperature for 20 minutes. Cell lysates were prepared using the co-IP lysis buffer (50 mM Tris-HCl pH 7.5, 120 mM NaCl, 0.5% NP-40, 0.1% SDS, and 1× EDTA-free protease inhibitor cocktail). The supernatants were incubated with either 5 μg of antibody overnight. 200 μL of washed magnetic beads (Thermo Fisher Scientific, 11203D) were added to each sample, followed by 3 hours of incubation. The beads were then collected and washed 5 times with the wash buffer (2 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, and 0.5% NP-40) at 4°C. The enriched proteins were used for Western blot analysis.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA extraction and Real-time Quantitative Analysis

Total RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo Research, 11-331), following the vendor's protocol. Genomic DNA contamination was removed by treating the extracted RNA with DNase. Reverse transcription was carried out using the SMARTScribe Reverse Transcriptase kit (TaKaRa, 639538). Real-time Quantitative Analysis (RT-qPCR) was performed using iTaq Universal SYBR Green Supermix (Bio-Rad, 1725122). The Primers used for RT-qPCR are listed in the Extended data 1.

RNA immunoprecipitation

RNA immunoprecipitation (RIP) was performed as previously described[78]. For each replicate, 40 million hESCs were washed with ice-cold PBS. Cells were lysed with 10 ml lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM DTT, 0.5% Igepal CA-630, supplemented with 40 U/ml of RNase inhibitor (Invitrogen, 10777019) and 1 × EDTA free protease inhibitor). The supernatants were incubated with 5 μg of antibody overnight at 4 °C. 200 μl prewashed magnetic beads were added and incubated for 3 hours. Bead-RNA complexes were washed with RIP lysis buffer. The RNA complexes were eluted using 500 μl TRIzol reagent (Ambion, 15596108) for RNA isolation.

m^6A-RIP

m^6A-RIP was performed as previously described[66,79]. For each replicate, 10 μg of total RNA was fragmented into approximately 200-nt-long fragments using 1 × RNA fragmentation buffer (10 mM Tris-HCl, 10 mM $ZnCl_2$ in nuclease-free H_2O). After purification, the RNA was diluted with 1 × IP buffer (150 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.1% Igepal CA-630 in nuclease-free H_2O supplemented with RNase inhibitor). 15 μl anti-m^6A antibodies (Cell Signaling Technology, D9D9W) were added to the samples, which were then incubated overnight at 4°C. Protein-A magnetic beads (Thermo Fisher Scientific, 10002D) and protein-G magnetic beads (Thermo Fisher Scientific, 10004D) were washed three times with 1 × IP buffer and added to the samples for another 4 hours of incubation. The RNA-bead complexes were washed 5 times with 1 mL of 1 × IP buffer. The RNA complexes were eluted using 500 μl TRIzol reagent (Ambion, 15596108), and isolated using Direct-zol RNA MiniPrep Kit (Zymo Research, 11-331). m^6A levels were measured using RT-qPCR.

eCLIP-seq library preparation

eCLIP libraries were generated as reported[62]. 10 million cells were treated with UV crosslinking (254 nm, 400 mJ/cm^2) and sonicated with Bioruptor (30 s ON/30 s OFF, 10 cycles) in the lysis buffer. The lysate was digested with RNase I and incubated with YTHDC2 antibody-coated beads, followed by an overnight incubation at 4 °C. The enriched RNAs were dephosphorylated and ligated with a barcoded RNA adapter before being run on PAGE gels and transferred to nitrocellulose membranes. A region from 160 kDa - 235 kDa was collected and treated with proteinase K to isolate RNAs. Free RNA was reverse transcribed. The 5’ ends of cDNA were ligated with DNA adaptors, and after cleaning up the reaction, the cDNA was PCR amplified to generate libraries.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA electrophoretic mobility shift

EMSA assays were performed with the LightShift Chemiluminescent RNA EMSA Kit (Thermo Scientific, 20158). Recombinant YTH domain proteins were expressed and purified from E. coli BL21 (DE3). Biotinylated RNA probes, with or without m^6A, had the sequence of 5′-CCACUGGAAAUCGGXCUGUCCAACUCACCU-3′ (where X is A or m6A). To determine the binding affinity of YTHDC2 for the RNA probes, 1 μl of RNA probes (4 nM) was incubated with recombinant YTHDC2 proteins (0, 20 nM, 100 nM, 500 nM, 2 μM and 5 μM) at room temperature for 30 min. The protein-RNA complexes were separated on a native gel. The protein-RNA complexes were then transferred onto nylon membranes and crosslinked by UV irradiation (254 nm, 120mJ/cm^2) to be detected using chemiluminescence.

In vitro probe pull-down

In vitro probe pull-down assay performed as previously described[24]. 20 million cells were lysed at 4 °C for 30 min in 5 ml lysis buffer (20mM Tris pH 7.5, 250 mM NaCl, 0.5% NP- 40, 10% glycerol, 1 × Roche Complete Protease Inhibitor, 40 U/ml of RNase inhibitor). Supernatants were diluted 1:1 in 20 mM Tris-HCl pH7.6 and incubated with biotinylated RNA probes with or without m^6A overnight. 100 μl streptavidin beads were washed twice with the lysis buffer and incubated with samples for 2 h. The proteins were eluted and separated with SDS-PAGE gel.

METTL3 inhibition and RNA stability assay

For METTL3 inhibition, STM2457 (MCE, HY-134836) was added into the medium at a final concentration of 5 μM for 1, 2, 3 days[64]. Total RNA was extracted for analysis of the m^6A level of HERV-H. The cells were also used for HEVRH stability analysis. For RNA stability assay, cells were plated 16-24 h before treatment. 80% confluent hESCs were treated with 5 μg ml^{-1} actinomycin D (Sigma, A1410) for 0, 1 h, 3 h, 7 h to inhibit transcription. Total RNA was extracted and used for RT-qPCR analysis.

Nascent RNA synthesis measured by RT-qPCR

We performed nascent RNA capture using Click-iT Nascent RNA Capture Kit (Thermo Fisher Scientific, C10365) according to the manufacturer's instructions. 5-Ethynyl Uridine (EU) was added to a final concentration of 0.5 mM for 60 min and 10 min in the incubator. EU labeled RNA was biotinylated and purified using streptavidin beads. The enriched RNA was reverse transcribed and used for RT-qPCR analysis as described in the RNA extraction and Real-time Quantitative Analysis part.

RNA sequencing and data analysis

Libraries for RNA sequencing were generated using Smart-seq2[80]. Briefly, RNA was reverse transcribed and cDNA was used to generate double stranded DNA (dsDNA). dsDNA was tagmented at 55 °C for 15 min and PCR amplified using Nextera primers. The libraries were purified with magnetic beads and sequenced using the MGISEQ-2000 in pair-end mode with 100 bp per read. For RNA-seq data analysis, raw reads were trimmed using Fastp (0.20.0)[81] and aligned to the hg38 genome using HISAT2 (2.1.0)[82] with default settings. Reads on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

each gene were counted using featureCounts (2.0.1)[83]. Differentially expressed genes were determined using DESeq2 (1.36.0)[84]. Gene ontology analysis was performed using the DAVID webtool with default parameters using all significant differentially expressed genes (p-value < 0.05)[73,74]. Overlapped genes were obtained using the Calculate and draw custom Venn diagrams webtool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

ChIP-seq and data analysis

ChIP libraries were generated as previously reported[27]. 10 million cells were crosslinked with 1% formaldehyde at room temperature for 10 min, quenched in 0.125 M glycine buffer and lysed in 50 μl lysis buffer (1% SDS, 50 mM Tris-HCl, pH 8.0, 20 mM EDTA supplemented with 10x protease inhibitor) on ice. The cell lysates were diluted to 500 μl with the cold TE buffer and sonicated using Covaris M220. The supernatants were transferred into new tubes and DNA concentrations were measured on a NanoDrop. For each replicate, 100 μl supernatants were diluted with 100 μl Binding Buffer (1% Triton-X, 01% Sodium Deoxycholate supplemented with protease inhibitor in TE buffer) The mixture was incubated with antibodies in the cold room overnight on a rotator. The following day, 20 μl lgG pre-washed Dynabeads were added into each reaction and incubated for another 2 h. After incubation, the beads were washed five times with wash buffer (50 mM Hepes, pH 8.0, 1% NP-40, 0.70% Sodium Deoxycholate, 0.5 M LiCl, 1 mM EDTA supplemented with protease inhibitor) and twice with cold 1 × TE. The beads were transferred to new tubes and incubated at 65°C for 20 min at 1300 rpm in the elution buffer (10 mM, Tris-HCl pH 8.0, 1 mM EDTA, 1% SDS). The supernatants containing DNA were transferred into new tubes and incubated at 65°C overnight to reverse crosslink. The following day, 250 μl 1 × TE and 8 μl RNase A were added to each sample and incubated at 37°C for 1 h to remove RNA. Proteinase K was added and incubated at 55°C for 1 h. Then DNA was purified and measured on a Qubit. To generate ChIP libraries, DNA ends were repaired. ployA tails were added to the ends and then adaptors were ligated. Last, the DNA was amplified with TruSeq primers and purified with magnetic beads. Libraries were sequenced on the MGISEQ-2000 platform in pair-end mode with 100 base pair per read.

For histone ChIP-seq analyses, raw reads were trimmed using Fastp (0.20.0)[81] and aligned to the hg38 genome using Bowtie2 (2.4.4)[85] with the default parameter. For TET1 and YTHDC2 ChIP-seq analysis, reads were trimmed using Trim Galore (0.6.4) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and aligned to the hg38 genome using BWA (0.7.17)[86]. Mapped reads were filtered by MAPQ > 10 using SAMtools (1.12)[87]. PCR duplicates were found using Sambamba (0.8.0)[88]. Peak calling was performed using MACS2 (2.2.7.1)[89]. HOMER (4.10)[90] annotatePeaks.pl was used to annotate peaks and for gene ontology analysis. Genome coverage bigWig files for visualization were generated with bamCoverage using deeptools (3.5.0)[91]. Heatmaps and profiles plots were generated using deeptools computerMatrix and plotProfile. The overlapping peaks were calculated using bedtools intersect or HOMER mergePeaks. The Jaccard statistic ratio of the intersection of two sets of peaks was calculated using bedtools (2.30.0)[92].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CUT&RUN and data analysis

CUT&RUN was performed using the CUTANA™ ChIC/CUT&RUN Kit (EpiCypher, 14-1048) according to manufacturer's instructions. For each replicate, 11 µl ConA Beads were activated in the Bead Activation Buffer and kept on ice until cells were ready. 200 000 cells were used and slightly crosslinked in 0.1% formaldehyde for 2 min at room temperature. Cells were washed and resuspended in 100 µl Wash Buffer. 10 µl pre-washed ConA Beads were added. After incubation for 10 min at room temperature, cells were resuspended in the 50 µl ice-cold Antibody Buffer. 1 µl antibodies were added and incubated in the cold room overnight on a rotator. Then beads were washed twice in 200 µl cold Cell Permeabilization Buffer and completely resuspended in 50 µl cold Cell Permeabilization Buffer. 2.5 µl pAG-MNase was added and incubated on the rotator for 10 min at room temperature. Then beads were washed twice in 200 µl cold Cell Permeabilization Buffer and completely resuspended in 50 µl cold Cell Permeabilization Buffer. 1 µl 100 mM Calcium Chloride was added to each tube and incubated for 2 h in the cold room. At the end of incubation, 33 µl Stop Buffer Master Mix was added to each reaction and incubated for 30 min at 37 °C to release DNA fragments from the insoluble nuclear chromatin. DNA was purified and used for library preparation. Libraries were sequenced on the MGISEQ-2000 platform in pair-end mode with 100 bp per read.

For CUT&RUN data analysis, raw reads were trimmed using Fastp (0.20.0)[81] and aligned to the hg38 genome using Bowtie2 (2.4.4)[85]. Mapped reads were filtered by MAPQ > 10 using SAMtools (1.12)[87].. Peak calling was performed using MACS2 (2.2.7.1)[89].

MeDIP sequencing and data analysis

The genomic DNA was first diluted in the TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and fragmented using Covaris sonication. The sonicated DNA (10 µg) was then repaired, A bases were added to the 3' end, and PCR adapters were ligated. After that, DNA was dissolved in the 450 µl TE buffer and denatured by boiling at 95°C for 10 minutes, followed by snap chilling on ice. To this DNA solution, 50 µl of 10x IP buffer (1.4 M NaCl, 100 mM Na-phosphate, pH 7.0, 0.5% Triton X-100) was added and incubated it with 1 µl of anti-5hmC or anti-5mC antibodies overnight. On the following day, 100 µl of magnetic beads were added, and the mixture was incubated at 4°C for another 4 hours. Afterward, the beads were washed five times with 1x IP buffer, and the DNA was eluted from the beads by boiling at 65°C for 15 minutes in the elution buffer (10 mM Tris pH 7.6, 1 mM EDTA, 1% SDS). The DNA was purified and PCR amplified. Libraries were size selected and sequenced on the MGISEQ-2000 platform in pair-end mode with 100 bp per read.

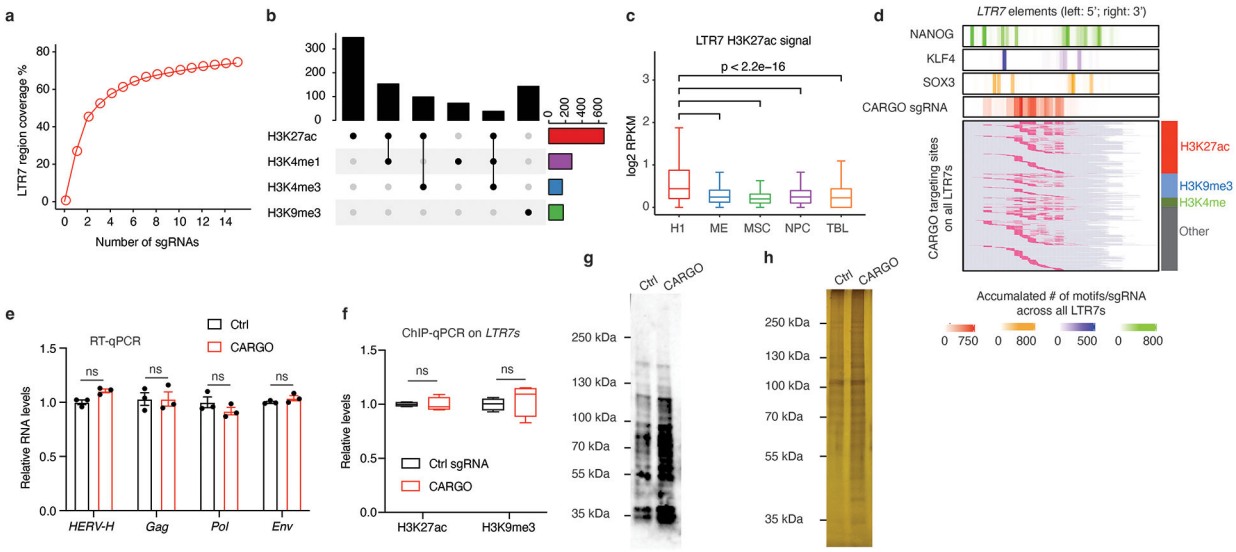
For MeDIP-seq data analysis, raw reads were trimmed using Fastp (0.20.0)[81] and aligned to the hg38 genome using Bowtie2 (2.4.4)[85]. Mapped reads were filtered by MAPQ > 30 using SAMtools (1.12)[87]. PCR duplicates were found using Sambamba (0.8.0)[88]. Peak calling was performed using MACS2 (2.2.7.1)[89].

Statistics and Reproducibility

Statistical analyses of the data were conducted as described in the respective figure legends. No statistical method was used to predetermine sample sizes. No data were excluded from

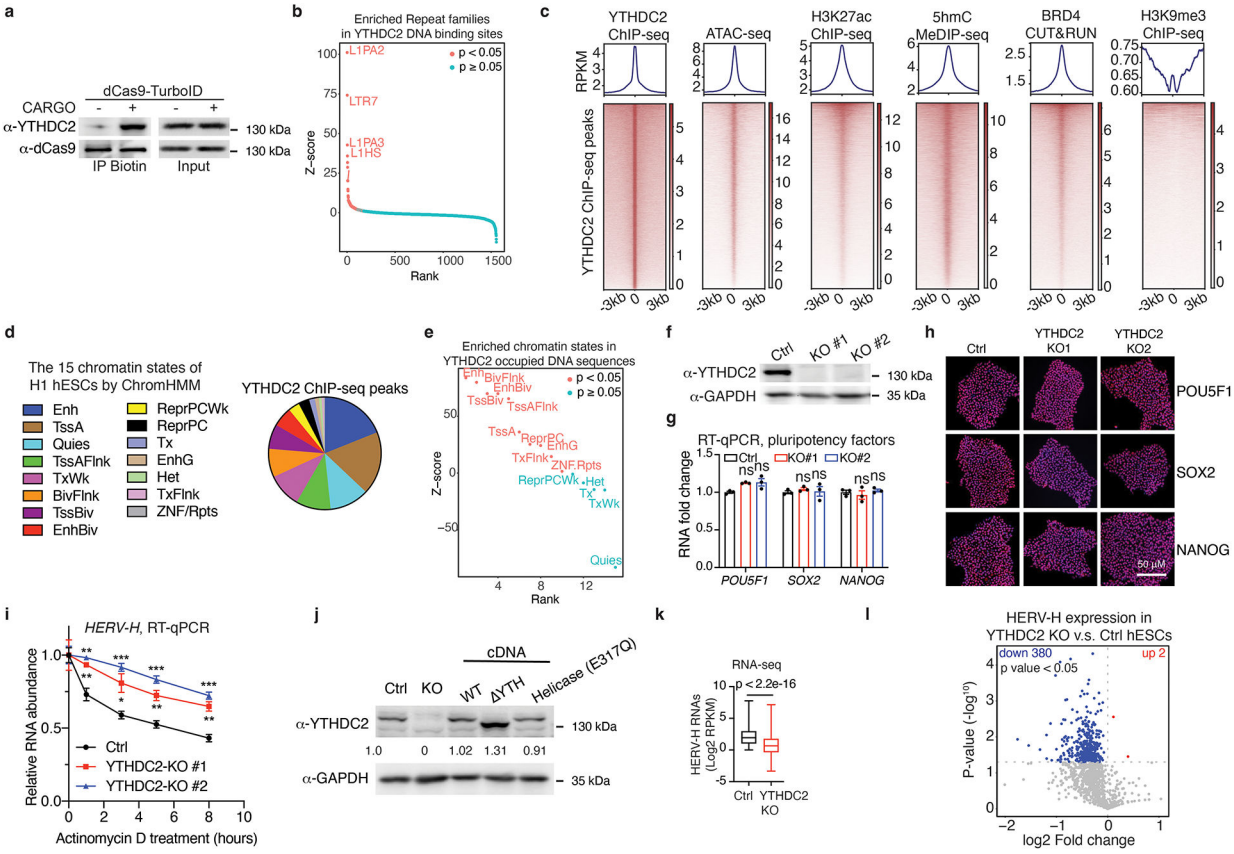
the analyses of all studies. Next-generation sequencing data generated in this study are based on two biological replicates. All results from representative experiments were collected at least three times independently. The experiments in this study were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Extended Data



Extended Data Fig. 1.
CARGO targets LTR7 sequences in human embryonic stem cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

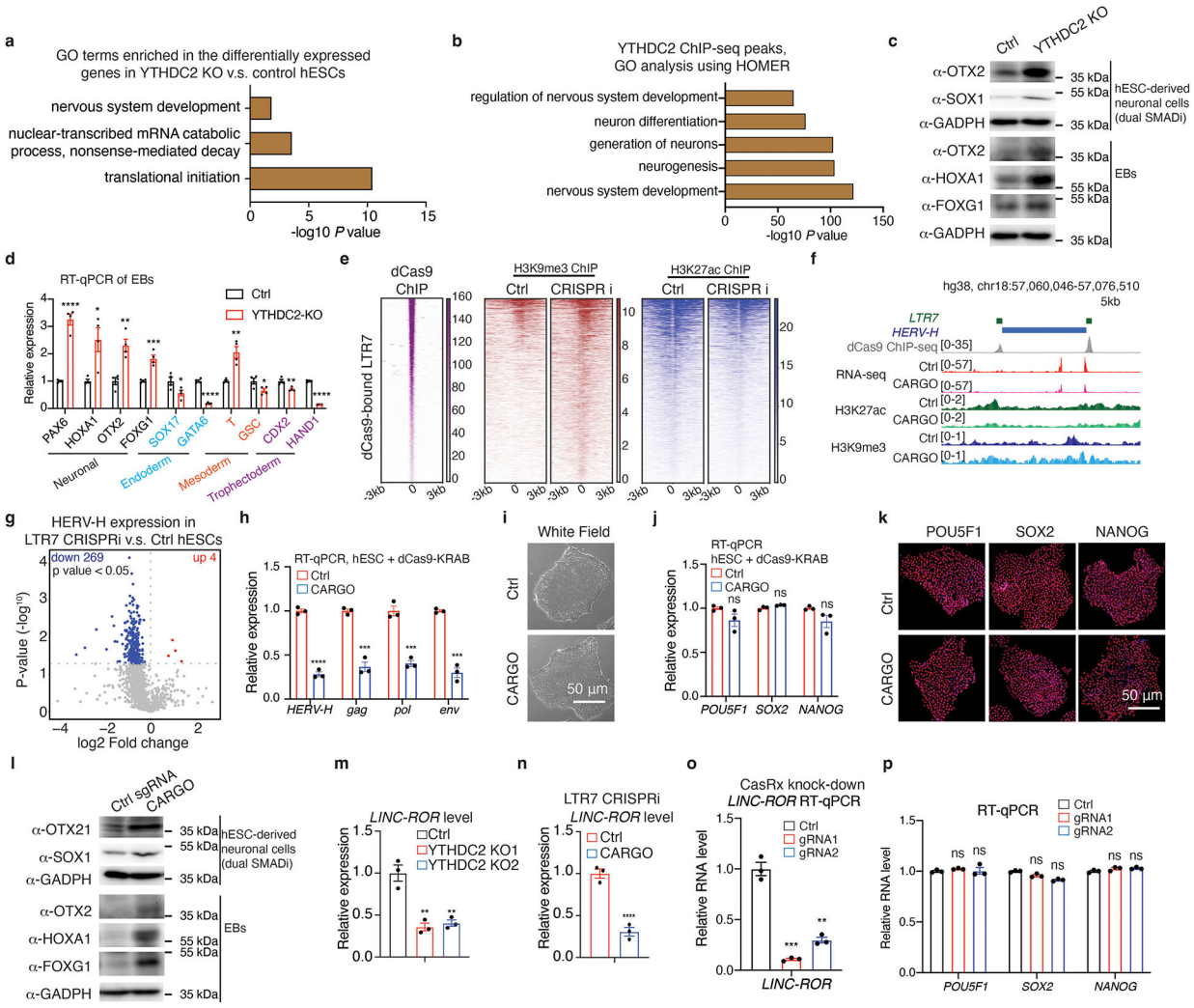


Extended Data Fig. 2.

YTHDC2 occupies active genomic loci and limits transcriptional silencing of LTR7/HERV-H loci in hESCs.

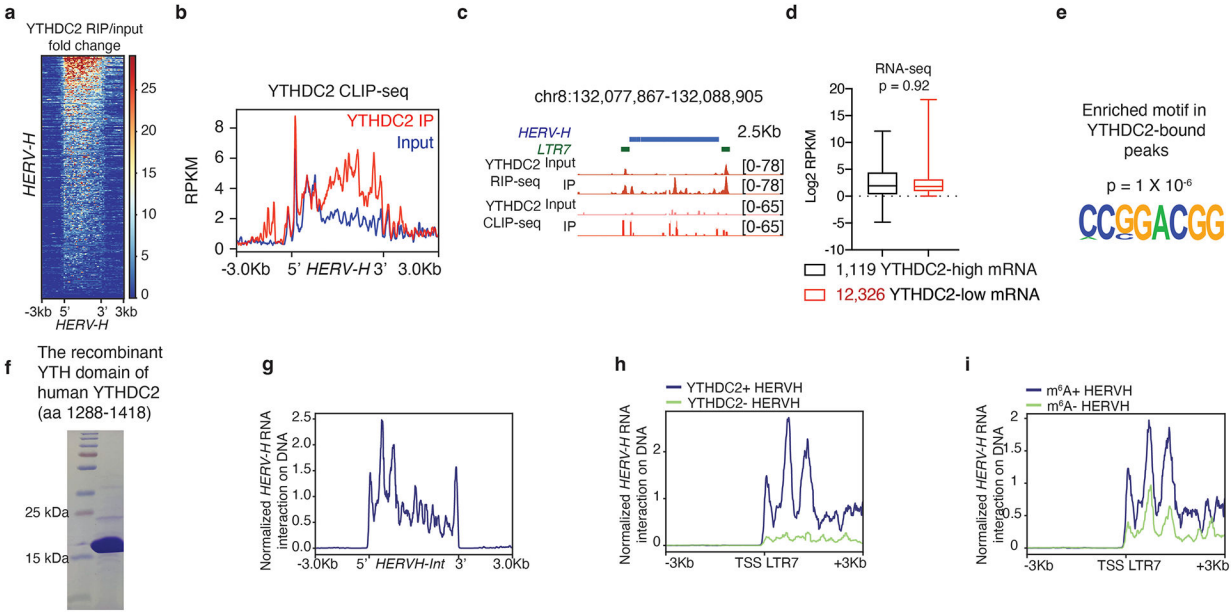
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

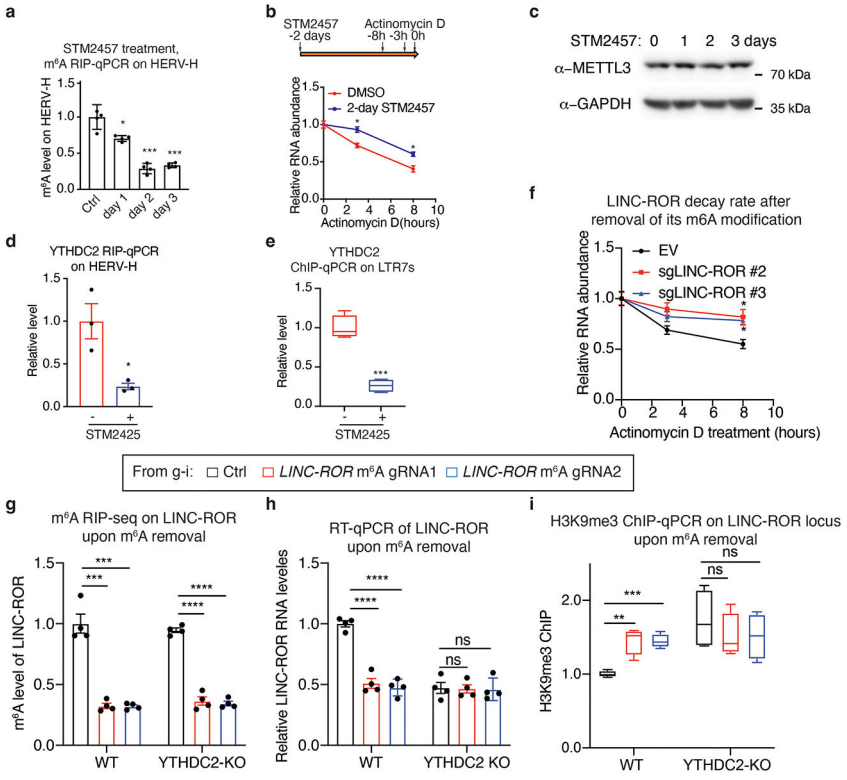


Extended Data Fig. 3.

The YTHDC2/LTR7 axis prevents neuronal fate of hESCs via LINC-ROR.

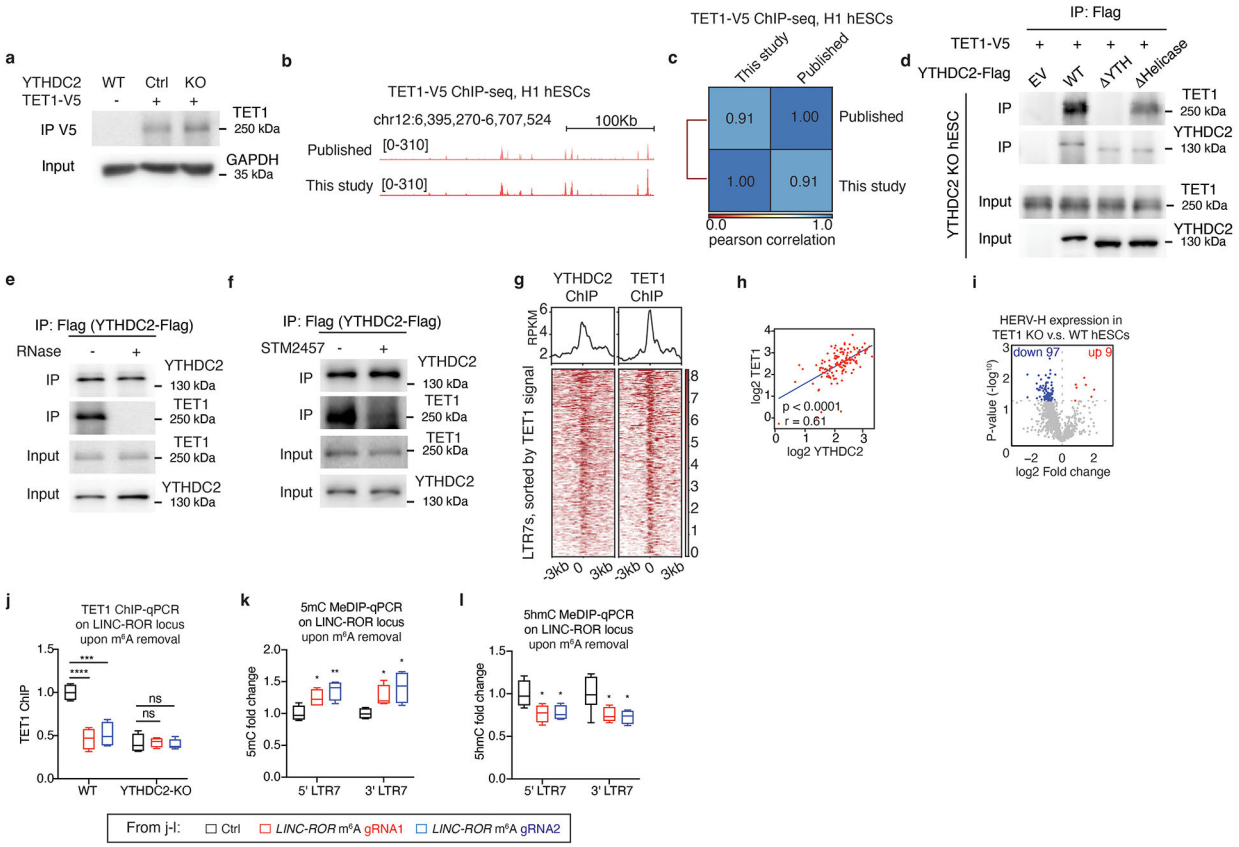


Extended Data Fig. 4.
Characterization of the physical interactions of YTHDC2, m6A-modified HERV-H RNAs, and LTR7 DNA.



Extended Data Fig. 5.
YTHDC2 binds to LTR7/HERV-H loci and prevents their silencing in a m6A dependent manner.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 6.
YTHDC2 recruits TET1 to demethylate LTR7/HERV-H DNA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Drs. Brigid Hogan and Kenneth Poss for discussion, Drs. Scott Soderling and Christopher Nicchitta for advising the BioID experiments, Dr. Sharyn Endow for recombinant protein purification. Y.D. acknowledges support from the Duke Regeneration Center, Duke Whitehead Scholarship, Glenn Foundation for Medical Research and AFAR Grants for Junior Faculty, NIH U01HL156064, and R35HG011328. Y.Xiang is supported by a Duke Regeneration Center postdoctoral fellowship. Y.Xu is supported by the Center for Advanced Genomic Technologies and the Duke Regeneration Center Fellowships.

DATA AVAILABILITY

Sequencing data have been deposited to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE210867. Publicly available RNA-seq of WT and TET1 knockout H1 cells (GSM5183601, GSM5183602, GSM5183607, GSM5183608), methyl capture sequencing (GSM5183585, GSM5183586, GSM5183589, GSM5183590), H3K4me1 (GSM409307), H3K4me3 (GSM409308) and MeRIP-seq (GSM1272365, GSM1272366, GSM1272367, GSM1272368) were obtained from the GEO. H3K27ac (ENCFF103PND) and H3K9me3 (ENCFF385ZBQ) were obtained from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the ENCODE project (https://www.encodeproject.org/). Proteomics data are available in the MassIVE database (Massive.ucsd.edu) under accession MSV000091407. Additional experimental materials used in this study, including plasmids and engineered cell lines, are available upon request.

DATA AVAILABILITY

All the data were analyzed using published pipelines with parameters described in the Methods section. No custom codes were developed in this study.

REFERENCES

1. Bourque G. et al. Ten things you should know about transposable elements. Genome Biol. 19, 199 (2018). [PubMed: 30454069]
2. Chuong EB, Elde NC & Feschotte C Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet 18, 71–86 (2017). [PubMed: 27867194]
3. Senft AD & Macfarlan TS Transposable elements shape the evolution of mammalian development. Nat. Rev. Genet (2021) doi:10.1038/s41576-021-00385-1.
4. Fueyo R, Judd J, Feschotte C & Wysocka J Roles of transposable elements in the regulation of mammalian transcription. Nat. Rev. Mol. Cell Biol 23, 481–497 (2022). [PubMed: 35228718]
5. Deniz Ö, Frost JM & Branco MR Regulation of transposable elements by DNA modifications. Nat. Rev. Genet 20, 417–431 (2019). [PubMed: 30867571]
6. Padeken J, Methot SP & Gasser SM Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. Nat. Rev. Mol. Cell Biol (2022) doi:10.1038/s41580-022-00483-w.
7. Hermant C & Torres-Padilla M-E TFs for TEs: the transcription factor repertoire of mammalian transposable elements. Genes Dev. 35, 22–39 (2021). [PubMed: 33397727]
8. Hermann A, Goyal R & Jeltsch A The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J. Biol. Chem 279, 48350–48359 (2004). [PubMed: 15339928]
9. Okano M, Bell DW, Haber DA & Li E DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257 (1999). [PubMed: 10555141]
10. Ito S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466, 1129–1133 (2010). [PubMed: 20639862]
11. Tahiliani M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935 (2009). [PubMed: 19372391]
12. Liu J. et al. N6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. Science 367, 580–586 (2020). [PubMed: 31949099]
13. Xu W. et al. METTL3 regulates heterochromatin in mouse embryonic stem cells. Nature (2021) doi:10.1038/s41586-021-03210-1.
14. Liu J. et al. The RNA m6A reader YTHDC1 silences retrotransposons and guards ES cell identity. Nature 591, 322–326 (2021). [PubMed: 33658714]
15. Wei J. et al. FTO mediates LINE1 m6A demethylation and chromatin regulation in mESCs and mouse development. Science eabe9582 (2022).
16. Li Y. et al. N6-Methyladenosine co-transcriptionally directs the demethylation of histone H3K9me2. Nat. Genet 52, 870–877 (2020). [PubMed: 32778823]
17. Lee J-H et al. Enhancer RNA m6A methylation facilitates transcriptional condensate formation and gene activation. Mol. Cell 81, 3368–3385.e9 (2021). [PubMed: 34375583]
18. Shi H, Wei J & He C Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. Mol. Cell 74, 640–650 (2019). [PubMed: 31100245]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

19. Yang Y, Hsu PJ, Chen Y-S & Yang Y-G Dynamic transcriptomic m6A decoration: writers, erasers, readers and functions in RNA metabolism. Cell Res. 28, 616–624 (2018). [PubMed: 29789545]
20. Michalak EM, Burr ML, Bannister AJ & Dawson MA The roles of DNA, RNA and histone methylation in ageing and cancer. Nat. Rev. Mol. Cell Biol 20, 573–589 (2019). [PubMed: 31270442]
21. Zhao S, Allis CD & Wang GG The language of chromatin modification in human cancers. Nat. Rev. Cancer 21, 413–430 (2021). [PubMed: 34002060]
22. Jambhekar A, Dhall A & Shi Y Roles and regulation of histone methylation in animal development. Nat. Rev. Mol. Cell Biol 20, 625–641 (2019). [PubMed: 31267065]
23. Deng S. et al. RNA m6A regulates transcription via DNA demethylation and chromatin accessibility. Nat. Genet 54, 1427–1437 (2022). [PubMed: 36071173]
24. Hsu PJ et al. Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res. 27, 1115–1127 (2017). [PubMed: 28809393]
25. Fuentes DR, Swigut T & Wysocka J Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife 7, (2018).
26. Gu B. et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science 359, 1050–1055 (2018). [PubMed: 29371426]
27. Branon TC et al. Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol 36, 880–887 (2018). [PubMed: 30125270]
28. Kim DI et al. Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. U. S. A 111, E2453–61 (2014). [PubMed: 24927568]
29. Schmidtmann E, Anton T, Rombaut P, Herzog F & Leonhardt H Determination of local chromatin composition by CasID. Nucleus 7, 476–484 (2016). [PubMed: 27676121]
30. Gao XD, Rodríguez TC & Sontheimer EJ Adapting dCas9-APEX2 for subnuclear proteomic profiling. Methods Enzymol. 616, 365–383 (2019). [PubMed: 30691651]
31. Liu X. et al. Multiplexed capture of spatial configuration and temporal dynamics of locus-specific 3D chromatin by biotinylated dCas9. Genome Biol. 21, 59 (2020). [PubMed: 32138752]
32. Ugur E, Bartoschek MD & Leonhardt H Locus-Specific Chromatin Proteome Revealed by Mass Spectrometry-Based CasID. Methods Mol. Biol 2175, 109–121 (2020). [PubMed: 32681487]
33. Xie W. et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell 148, 816–831 (2012). [PubMed: 22341451]
34. Leung D. et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature 518, 350–354 (2015). [PubMed: 25693566]
35. Zhang Y. et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet 51, 1380–1388 (2019). [PubMed: 31427791]
36. Glinsky GV Transposable Elements and DNA Methylation Create in Embryonic Stem Cells Human-Specific Regulatory Sequences Associated with Distal Enhancers and Noncoding RNAs. Genome Biol. Evol 7, 1432–1454 (2015). [PubMed: 25956794]
37. Gorkin DU, Leung D & Ren B The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell 14, 762–775 (2014). [PubMed: 24905166]
38. Shi J & Vakoc CR The mechanisms behind the therapeutic activity of BET bromodomain inhibition. Mol. Cell 54, 728–736 (2014). [PubMed: 24905006]
39. Dhalluin C. et al. Structure and ligand of a histone acetyltransferase bromodomain. Nature 399, 491–496 (1999). [PubMed: 10365964]
40. Ernst J & Kellis M Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol 28, 817–825 (2010). [PubMed: 20657582]
41. Ernst J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43–49 (2011). [PubMed: 21441907]
42. Jao CY & Salic A Exploring RNA transcription and turnover *in vivo* by using click chemistry. Proc. Natl. Acad. Sci. U. S. A 105, 15779–15784 (2008). [PubMed: 18840688]
43. Saito Y. et al. YTHDC2 control of gametogenesis requires helicase activity but not m6A binding. Genes Dev. 36, 180–194 (2022). [PubMed: 35058317]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

44. Li L. et al. The XRN1-regulated RNA helicase activity of YTHDC2 ensures mouse fertility independently of m6A recognition. Mol. Cell 82, 1678–1690.e12 (2022). [PubMed: 35305312]
45. Larivera S & Meister G Domain confusion 2: m6A-independent role of YTHDC2. Molecular cell vol. 82 1608–1609 (2022). [PubMed: 35523125]
46. Lu X. et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol 21, 423–425 (2014). [PubMed: 24681886]
47. Wang J. et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature 516, 405–409 (2014). [PubMed: 25317556]
48. Ohnuki M. et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proc. Natl. Acad. Sci. U. S. A 111, 12426–12431 (2014). [PubMed: 25097266]
49. Koyanagi-Aoi M. et al. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. Proceedings of the National Academy of Sciences vol. 110 20569–20574 Preprint at 10.1073/pnas.1319061110 (2013).
50. Takahashi K. et al. The pluripotent stem cell-specific transcript ESRG is dispensable for human pluripotency. Cold Spring Harbor Laboratory 2020.11.25.397935 (2020) doi:10.1101/2020.11.25.397935.
51. Chambers SM et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol 27, 275–280 (2009). [PubMed: 19252484]
52. Verma N. et al. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat. Genet 50, 83–95 (2018). [PubMed: 29203910]
53. Ahmad R. et al. Functional neuronal cells generated by human parthenogenetic stem cells. PLoS One 7, e42800 (2012). [PubMed: 22880113]
54. Zheng Y. et al. Single-cell analysis of embryoids reveals lineage diversification roadmaps of early human development. Cell Stem Cell 29, 1402–1419.e8 (2022). [PubMed: 36055194]
55. Thakore PI et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods 12, 1143–1149 (2015). [PubMed: 26501517]
56. Loewer S. et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet 42, 1113–1117 (2010). [PubMed: 21057500]
57. Cheng E-C & Lin H Repressing the repressor: a lincRNA as a MicroRNA sponge in embryonic stem cell self-renewal. Developmental cell vol. 25 1–2 (2013). [PubMed: 23597480]
58. Wang Y. et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev. Cell 25, 69–80 (2013). [PubMed: 23541921]
59. Konermann S. et al. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell 173, 665–676.e14 (2018). [PubMed: 29551272]
60. Yan WX et al. Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. Mol. Cell 70, 327–339.e5 (2018). [PubMed: 29551514]
61. Batista PJ et al. m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. Cell Stem Cell 15, 707–719 (2014). [PubMed: 25456834]
62. Van Nostrand EL et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 13, 508–514 (2016). [PubMed: 27018577]
63. Sridhar B. et al. Systematic Mapping of RNA-Chromatin Interactions In Vivo. Curr. Biol 27, 602–609 (2017). [PubMed: 28132817]
64. Yankova E. et al. Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. Nature 593, 597–601 (2021). [PubMed: 33902106]
65. Zheng G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell 49, 18–29 (2013). [PubMed: 23177736]
66. Xia Z. et al. Epitranscriptomic editing of the RNA N6-methyladenosine modification by dCasRx conjugated methyltransferase and demethylase. Nucleic Acids Res. 49, 7361–7374 (2021). [PubMed: 34181729]
67. Dixon G. et al. QSER1 protects DNA methylation valleys from de novo methylation. Science 372, (2021).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

68. Liu XS et al. Editing DNA Methylation in the Mammalian Genome. Cell 167, 233–247.e17 (2016). [PubMed: 27662091]
69. Ziller MJ et al. Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477–481 (2013). [PubMed: 23925113]
70. Robertson KD DNA methylation and human disease. Nat. Rev. Genet 6, 597–610 (2005). [PubMed: 16136652]
71. Xu W & Shen H When RNA methylation meets DNA methylation. Nature genetics vol. 54 1261–1262 (2022). [PubMed: 36071174]

METHODS-ONLY REFERENCES

72. McKenna A & Shendure J FlashFry: a fast and flexible tool for large-scale CRISPR target design. BMC Biol. 16, 74 (2018). [PubMed: 29976198]
73. Huang DW, Sherman BT & Lempicki RA Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc 4, 44–57 (2009). [PubMed: 19131956]
74. Sherman BT et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. (2022) doi:10.1093/nar/gkac194.
75. Kramer A, Green J, Pollard J Jr., & Tugendreich S Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30, 523–530 (2014). [PubMed: 24336805]
76. Sanjana NE, Shalem O & Zhang F Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784 (2014). [PubMed: 25075903]
77. Diao Y, Wang X & Wu Z SOCS1, SOCS3, and PIAS1 promote myogenic differentiation by inhibiting the leukemia inhibitory factor-induced JAK1/STAT1/STAT3 pathway. Mol. Cell. Biol 29, 5084–5093 (2009). [PubMed: 19620279]
78. Chelmicki T. et al. m6A RNA methylation regulates the fate of endogenous retroviruses. Nature Preprint at 10.1038/s41586-020-03135-1 (2021).
79. Zeng Y. et al. Refined RIP-seq protocol for epitranscriptome analysis with low input materials. PLoS Biol. 16, e2006092 (2018). [PubMed: 30212448]
80. Picelli S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc 9, 171–181 (2014). [PubMed: 24385147]
81. Chen S, Zhou Y, Chen Y & Gu J fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics vol. 34 i884–i890 Preprint at 10.1093/bioinformatics/bty560 (2018). [PubMed: 30423086]
82. Kim D, Paggi JM, Park C, Bennett C & Salzberg SL Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol 37, 907–915 (2019). [PubMed: 31375807]
83. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014). [PubMed: 24227677]
84. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014). [PubMed: 25516281]
85. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]
86. Li H & Durbin R Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]
87. Li Handsaker, Wysoker Fennell & Ruan. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics.
88. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ & Prins P Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034 (2015). [PubMed: 25697820]
89. Feng J, Liu T, Qin B, Zhang Y & Liu XS Identifying ChIP-seq enrichment using MACS. Nat. Protoc 7, 1728–1740 (2012). [PubMed: 22936215]
90. Heinz S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010). [PubMed: 20513432]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

91. Ramirez F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–5 (2016). [PubMed: 27079975]

92. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

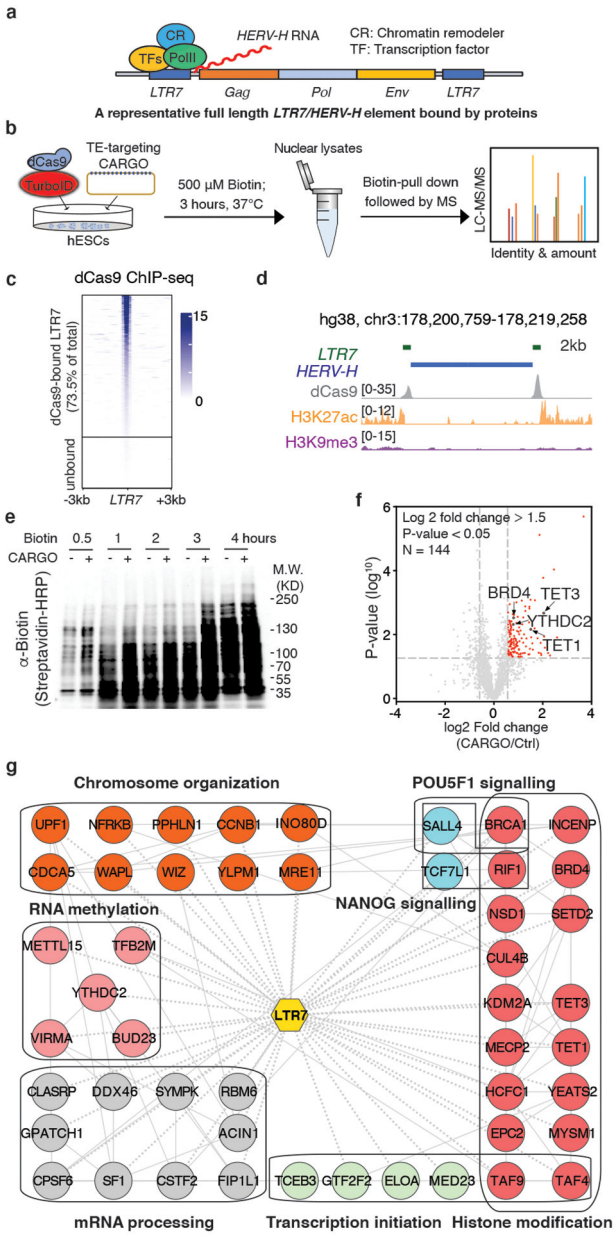


Figure 1. Development of CARGO-BioID to identify LTR7-associated proteome in hESCs.

a, Schematic of a representative full-length LTR7/HERV-H element. The promoter LTR7 drives the expression of HERV-H RNA, and is bound by transcription factors (TFs), RNA polymerase II (PolII), and chromatin remodelers (CR).

b, Illustration of the CARGO-BioID experimental design. CARGO: Chimeric Array of gRNA Oligos; BioID: proximity-dependent biotin identification.

c, d, Heatmap (**c**) and a genome browser snapshot (**d**) illustrating dCas9 binding and histone H3K27ac and H3K9me3 modifications on LTR7/HERV-H sequences.

e, A representative Western blotting (WB) result showing the patterns of nuclear proteins that are biotinylated by CARGO-BioID at indicated experimental conditions. Three hours

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of biotin labeling is selected for proteomic analysis. n = 3 biologically independent experiments.

f, Volcano plot showing 144 significantly enriched proteins that are spatially proximal to the LTR7 loci. Exact p values are in Source Data 1.

g, Representative protein interaction network, significantly enriched Gene Ontology (GO) terms ($p < 0.05$) and signaling pathways revealed by Ingenuity Pathways analysis (IPA) ($p < 0.05$) of these LTR7-bound proteins in H1 hESC cells. The protein-protein interaction information is retrieved from the STRING protein interaction database with default confidence cutoff using Cytoscape. Proteins are grouped by their known molecular function in different colors. Node titles show the corresponding gene symbols. Dashed edges show the interactions with LTR7s. Solid edges show the known protein-protein interactions. Exact p values are in Supplementary Table 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

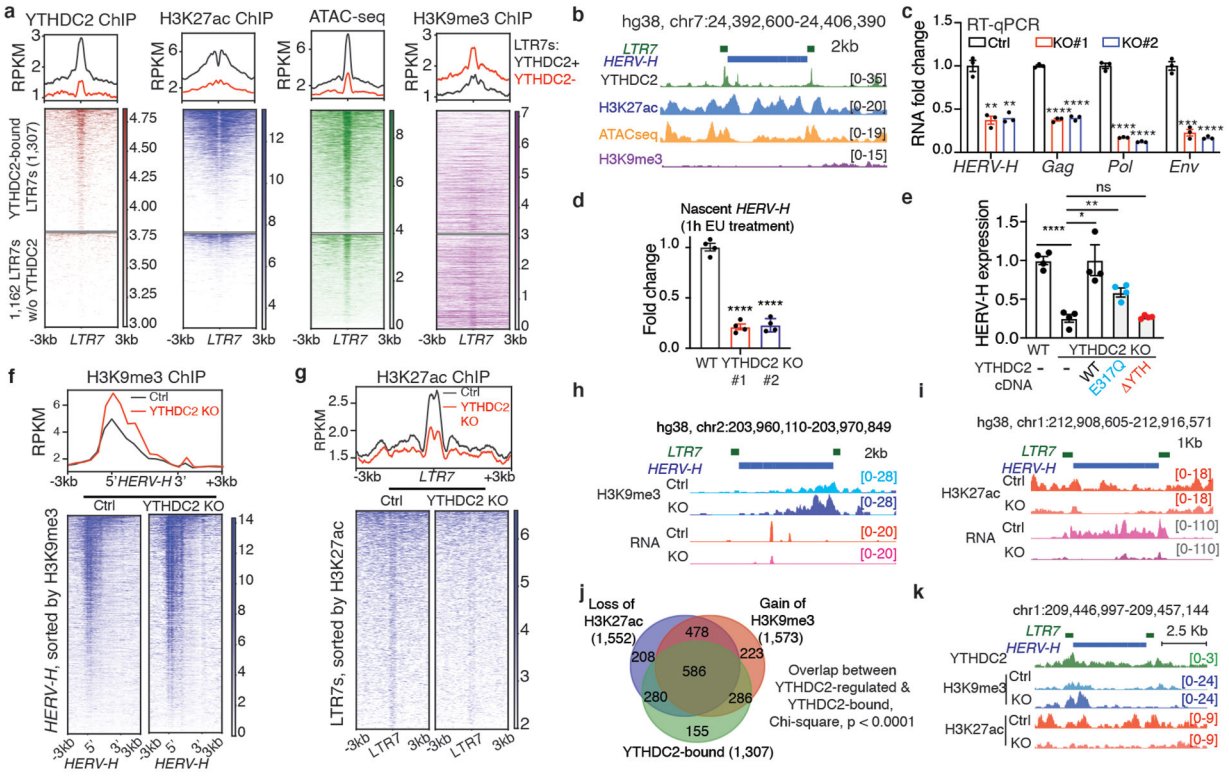


Figure 2. YTHDC2 occupies LTR7 genomic loci and limits their transcriptional silencing.

a, b, Heatmaps (**a**) and a genome browser snapshot (**b**) showing indicated signals on LTR7 sequences. Each row of the heatmap represents one LTR7 element. All LTR7 loci are ranked in the same order and sorted based on their YTHDC2 ChIP-seq signal across all heatmap plots.

c, RT-qPCR analysis showing the expression level of HERV-H and its retroviral genes in YTHDC2 Knock-out (KO) compared to wild-type (WT) cells.

d, RT-qPCR analysis of newly transcribed HERV-H.

e, The WT and two mutant forms of YTHDC2 cDNA were expressed in YTHDC2 KO cells, followed by RT-qPCR analysis to test their effect in rescuing the expression of HERV-H transcripts. E317Q denotes a point mutation disrupting ATP binding of helicase. ΔYTH indicates deletion of the entire YTH domain from YTHDC2.

f-i, Aggregated reads density (the top panels in **f** and **g**), heatmaps (the bottom panel in **f** and **g**), and a genome browser snapshots (**h, i**), showing altered histone H3K9me3 and H3K27ac modifications on LTR7 loci in wild-type control (Ctrl) and YTHDC2 KO hESCs.

j, Venn diagram showing that the YTHDC2-bound (green) LTR7s significantly overlap with the YTHDC2-regulated (red and blue) LTR7s (Chi-square test: p-value < 0.0001). Red: LTR7s gain H3K9me3 in YTHDC2 KO (p-value < 0.05, fold change > 2); blue: the LTR7s lose H3K27ac in YTHDC2 KO (p-value < 0.05, fold change >1.2).

k, Genome browser snapshot showing one representative LTR7/HERV-H element that is bound and regulated by YTHDC2, and that shows increased H3K9me3 and reduced H3K27ac ChIP-seq signal in YTHDC2 KO compared to the wild-type control hESCs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data represent mean ± s.e.m from three (**c**) or four (**d, e**) independent experiments. ns, not significant; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. P values were calculated by two-tailed Student's test in (**c-e**). Exact p values are in Source Data 1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

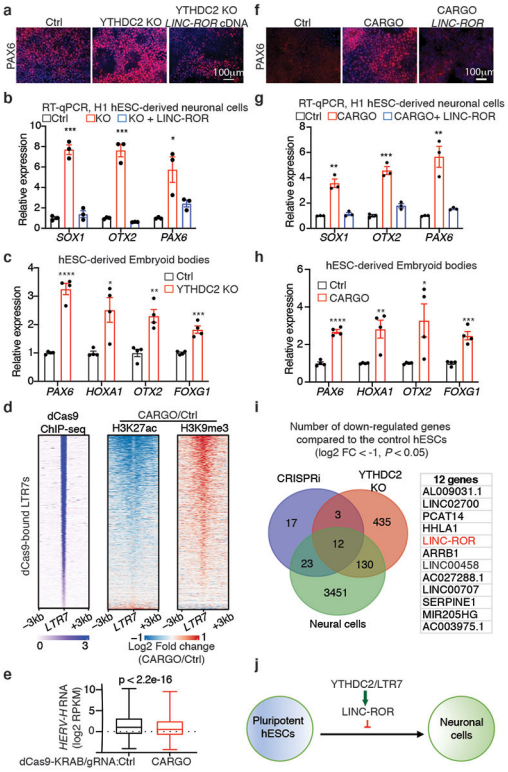


Figure 3. The YTHDC2/LTR7 axis inhibits neuronal fate of hESCs via LINC-ROR.

a, b, Immunofluorescence staining (**a**) and RT-qPCR analysis (**b**) of neuronal markers in control, YTHDC2 KO and LINC-ROR overexpression cells cultured in the neural induction media.

c, RT-qPCR analysis of neuronal markers in embryoid bodies derived from control, YTHDC2 KO and LINC-ROR overexpression hESCs.

d, ChIP-seq analysis of dCas9, H3K27ac and H3K9me3 ChIP-seq signals on LTR7s bound by dCas9.

e, Boxplot of RNA-seq data showing reduced HERV-H RNA levels upon LTR7 inhibition. P value determined by two-sided Wilcoxon signed-rank test. n = 2 biologically independent experiments. In the plots, center lines represent the median value, box limits the 25th and 75th percentiles, and whiskers denote minima and maxima (1.5× the interquartile range).

f, g, Immunofluorescence staining (**f**) and RT-qPCR analysis (**g**) of neuronal markers in control, CARGO and LINC-ROR overexpression cells cultured in the neural induction media.

h, RT-qPCR analysis of neuronal markers in embryoid bodies derived from control, CARGO and LINC-ROR overexpression hESCs.

i, Venn diagrams highlighting the shared downregulated genes in LTR7 CRISPRi, YTHDC2 KO, and hESC-derived neuronal cells. The 12 genes downregulated in three conditions are listed on the right.

j. Schematic illustration of a model in which LINC-ROR is directly regulated by YTHDC2/LTR7-axis and prevents neural fate of hESCs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data represent mean ± s.e.m from three (**b, g**) or four (**c, h**) independent experiments. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. P values were calculated by two-tailed Student's test (**b, c, g, h**). Exact p values are in Source Data 1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

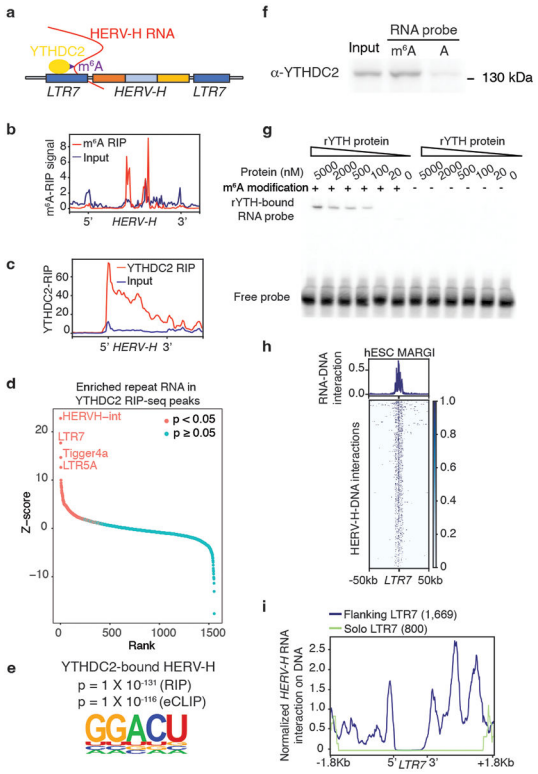


Figure 4. The m^{6}A-modified HERV-H RNAs interact with YTHDC2 and LTR7 DNA.

a, Schematic illustration of the proposed model: YTHDC2 is recruited to the LTR7 DNA sequences via its interaction with m^{6}A-modified HERV-H RNAs associated with LTR7 DNA.

b, c, Aggregated RIP-seq reads counts of m^{6}A MeRIP (**b**) and YTHDC2 RIP-seq (**c**) across the HERV-H RNA sequences.

d, A permutation test was conducted to identify the transcripts of repeat families enriched in the YTHDC2 RIP-seq peaks. The significantly enriched repeat RNAs in the YTHDC2 occupied RNA sequences were highlighted in red (p-value < 0.05). Exact p values are in Supplementary Table 3.

e, The canonical m^{6}A motif GGACU is significantly enriched in the YTHDC2-bound HERV-H RNA sequences identified by RIP-seq or eCLIP-seq.

f, g, RNA pull-down (**f**) and EMSA analysis (**g**) showing YTHDC2 binding to RNA with or without m^{6}A modification. n = 3 biologically independent experiments.

h, i, MARGI analyses showing aggregated signal (top panel of **h**; **i**) and heatmap (bottom panel of **h**) centered on all LTR7s (**h**), LTR7s flanking HERV-H (**i**, blue line), and solo LTR7s (**i**, green line).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

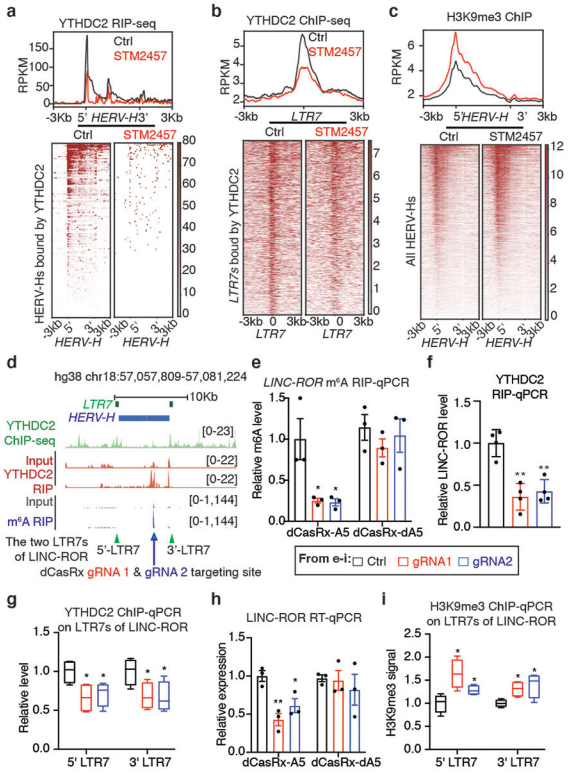


Figure 5. The m^{6}A modification of HERV-H RNAs regulates the interaction and activation of the YTHDC2/LTR7 axis.

a-c, Aggregated read counts (top) and heatmaps (bottom) showing the signal of YTHDC2 RIP-seq (**a**), YTHDC2 ChIP-seq (**b**), and H3K9me3 ChIP-seq (**c**) in hESCs treatment with DMSO or STM2457. Each row of the heatmap represents one HERV-H RNA transcript (**a**) or DNA element (**b, c**) as indicated.

d, Genome browser snapshot on LINC-ROR locus, displaying YTHDC2 ChIP-seq, YTHDC2 RIP-seq, and m^{6}A MeRIP-seq data tracks. Two CasRx gRNAs were designed to target the m^{6}A peak (the dark blue arrow) of LINC-ROR. The two green arrowheads point to the two LTR7 loci of LINC-ROR.

e, MeRIP-qPCR analysis using m^{6}A antibody showing that dCasRx-ALKBH5, but not dCasRx-dALKBH5, specifically reduced m^{6}A modification of LINC-ROR RNA.

f-i, The m^{6}A modification of LINC-ROR was erased by dCasRx-ALKBH5. This manipulation leads to reduced YTHDC2 binding on LINC-ROR RNA, determined by YTHDC2 RIP-qPCR (**f**); decreased YTHDC2 occupancy on the two LTR7 DNA loci flanking LINC-ROR, determined by YTHDC2 ChIP-qPCR (**g**); downregulation of LINC-ROR RNA determined by RT-qPCR (**h**); and gain of H3K9me3 modification on the two LTR7 loci flanking LINC-ROR, determined by H3K9me3 ChIP-qPCR (**i**). dCasRx-A5: dCasRx-ALKBH5; dCasRx-dA5: dCasRx-dALKBH5. Black bars: non-targeting control gRNA; red and blue: the two guide RNAs (#1 and #2) targeting the m^{6}A-modified sequence of LINC-ROR (**e-i**). In the box plots (**g, i**), center lines represent the median value, box limits the 25th and 75th percentiles, and whiskers denote minima and maxima (1.5× the interquartile range) (n = 3 biologically independent experiments). Data represent mean ± s.e.m from three (**e, h**) or four (**f**) independent experiments. *p < 0.05, **p

< 0.01. P values were calculated by two-tailed Student's test in (**e-i**). Exact p values are in Source Data 1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

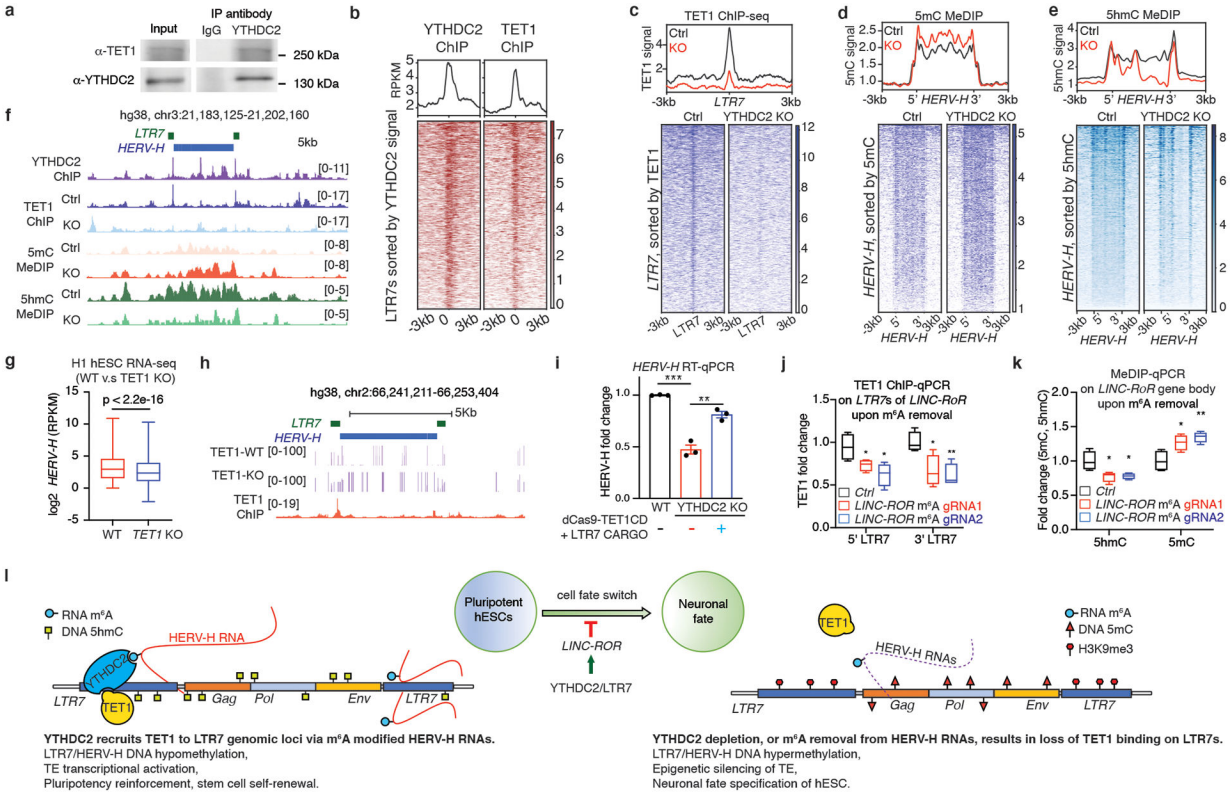


Figure 6. YTHDC2 recruits TET1 to prevent epigenetic silencing of LTR7/HERV-H loci via DNA demethylation.

a, YTHDC2 protein interacts with TET1 protein in H1 hESCs determined by endogenous protein co-immunoprecipitation experiments. n = 3 biologically independent experiments.

b, YTHDC2 and TET1-V5 ChIP-seq signal on LTR7 regions occupied by YTHDC2 in hESCs. The LTR7 loci in the two heatmap plots were ranked in the same order and sorted based on YTHDC2 ChIP-seq signal.

c-e, The TET1-V5 ChIP-seq (**c**), 5mC MeDIP (**d**) and 5hmC MeDIP (**e**) in control and YTHDC2 KO cells were shown as aggregated reads counts coverage (top) and heatmaps (bottom) centered on LTR7 (**c**) and HERV-H (**d, e**) regions.

f, Genome browser snapshot showing a representative LTR7/HERV-H locus bound by YTHDC2/TET1 and exhibiting reduced TET1 binding, increased 5mC, and decreased 5hm signals in YTHDC2 KO cells.

g, Boxplot showing reduction of HERV-H RNAs in TET1 KO H1 hESCs compared to WT cells.

h, Genome browser snapshot showing a representative LTR7/HERV-H locus exhibiting increased 5mC signal upon TET1 knockout.

i, RT-qPCR analysis of HERV-H expression in the indicated hESCs.

j, TET1 ChIP-qPCR analysis in the indicated conditions.

k, 5hmC and 5mC MeDIP-qPCR analysis in the indicated conditions.

l, Cartoon illustration of the model proposed in this study: YTHDC2 recruits TET1 to the genomic loci of LTR7/HERV-H, through its interaction with m^{6}A modified HERV-H RNAs. The engagement of TET1 on LTR7 sequences prevents the epigenetic silencing of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

LTR7/HERV-H and inhibits the neuronal fate of hESCs. LINC-ROR is one of the molecules directly regulated by the YTHDC2/LTR7 axis and inhibits neuronal fate.

In box plots (**g, j, k**), center lines represent the median value, box limits the 25th and 75th percentiles, and whiskers denote minima and maxima (1.5× the interquartile range). P value in (**g**) determined by two-sided Wilcoxon signed-rank test (n = 2 biologically independent experiments). Data in (**i-k**) represent mean ± s.e.m from three independent experiments. *p < 0.05, **p < 0.01, ***p < 0.001. P values were calculated by two-tailed Student's test (**i-k**). Exact p values are in Source Data 1.