H4K16ac activates the transcription of transposable elements and contributes to their cis regulatory function.

Debosree Pal[1]†, Manthan Patel[1]†, Fanny Boulet[1], Jayakumar Sundarraj[1 2], Olivia A Grant[1,3], Miguel R. Branco[1], Srinjan Basu[4], Nicolae Radu Zabet[1], Paola Scaffidi[5], & Madapura M Pradeepa[1*]

1 - Blizard Institute; Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK
2 - Bhabha Atomic Research Centre, Mumbai, India
3 - School of Life Sciences, University of Essex, UK
4 - Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, UK
5 - Francis Crick Institute; London, UK
† Both the authors contributed equally
[*] Corresponding author Madapura M Pradeepa: p.m.madapura@qmul.ac.uk

Abstract:

Mammalian genomes harbor a large number of transposable elements (TEs) and their remnants. Most TEs are incapable of retrotransposition. Although most TEs are epigenetically repressed, transcriptional silencing is partially released to permit developmental or tissue-specific expression of TEs. Some TEs have also evolved as cis-regulatory elements (CREs), enabling them to recruit host-encoded transcription factors. Understanding the contribution of TEs in the regulation of the mammalian genome is an active area of research. Previously, the noncoding long terminal repeat (LTR) part of the endogenous retrovirus (ERV) families has been shown to function as enhancers. We show that new LTR families and the promoter region of LINE1 (L1) are enriched with H4K16ac and H3K122ac and the chromatin features associated with active enhancers. Depletion of MSL complex and H4K16ac levels leads to a significant reduction in the expression of L1 and ERV/LTRs. We demonstrate that H4K16ac regulates TE transcription by maintaining a permissive chromatin structure. Furthermore, CRISPR-based perturbation of candidate TEs led to the downregulation of genes *in cis*. We conclude that H4K16ac and H3K122ac regulate a significant portion of the mammalian genome by opening local chromatin structure and transcriptional activity at TEs.

Introduction

Dysregulation of TEs and their insertions into gene exons are usually disruptive and are implicated in cancer and neurological disorders (Burns, 2017; Hancks and Kazazian, 2016). When inserted into noncoding DNA, including introns, they can affect the host gene expression in *cis* or *trans*. Most TEs are incapable of transposing due to acquired mutations, epigenetic, and post-transcriptional silencing mechanisms, reviewed in (Almeida et al., 2022; Molaro and Malik, 2016). However, TEs are transiently upregulated during the early development (Jachowicz et al., 2017), in the neuronal lineage (Upton et al., 2015) and cancer (Hancks and Kazazian, 2016). The endogenous retrovirus (ERV) superfamily of long-terminal repeats (LTR) and short interspersed nuclear element (SINE/*Alu*) often exhibit chromatin features associated with active enhancers (cis-acting regulatory elements, CREs) (Fueyo et al., 2022; Sundaram and Wysocka, 2020) and are shown to function either as enhancers to regulate genes in cis or act as alternate promoters (Fueyo et al., 2022). Long interspersed nuclear elements (LINE-1, L1) are also bound by tissue-specific transcription factors (TFs) and can function as nuclear noncoding RNAs (Jachowicz et al., 2017; Percharde et al., 2018), but it is unclear whether they can act as CREs. TEs are suggested to contribute to nearly one-quarter of the regulatory epigenome (Chuong et al., 2017; Hermant and Torres-Padilla, 2021; Jachowicz et al., 2017; Schmidt et al., 2012).

Histone post-translational modifications (PTMs) – H3K4me1 in combination with H3K27ac, bidirectional transcription of enhancer RNAs (eRNAs) and accessible chromatin (e.g., using ATAC-seq) are widely used to predict active enhancers in the genome (Andersson et al., 2014; Buenrostro et al., 2013; Creyghton et al., 2010). However, most histone acetylations, including H3K27ac, appear as a consequence of transcription and play a supportive role rather than a regulatory role in the transcription (Martin et al., 2021; Wang et al., 2022). Moreover, the H3K27ac level does not correlate with or is dispensable for enhancer activity, suggesting other uncharacterized chromatin features could be contributing to transcriptional regulatory activity (Kheradpour et al., 2013; Taylor et al., 2013; Wang et al., 2022). Apart from H3K27ac, other histone acetylations such as H3K122ac, H4K16ac, H3K18ac and H3K9ac are also found across active enhancers (Karmodiya et al., 2012; Pradeepa, 2017; Pradeepa et al., 2016; Taylor et al., 2013; Wolfe et al., 2021). Indeed we previously showed that new repertoires of enhancers that lack detectable H3K27ac are enriched with H3K122ac or H4K16ac in the mouse genome (Pradeepa et al., 2016; Taylor et al., 2013). H4K16ac and H3K122ac are particularly interesting among the many histone acetylations as they directly alter chromatin structure and increase transcription *in vitro* (Shogren-Knaak et al., 2006; Tropberger et al., 2013). However, it is challenging to decipher the causal role of specific histone acetylation marks because many acetylations, including H3K27ac, are catalyzed by multiple lysine acetyltransferases (KATs). KATs also have a broad substrate specificity. H4K16ac is an exception as it is catalyzed explicitly by KAT8 when associated with male-specific lethal (MSL) complex. Still, when KAT8 is associated with non-specific lethal (NSL), it catalyzes H4K5ac, H4K8ac and H4K12ac (Chatterjee et al., 2016; Chelmicki et al., 2014; Radzisheuskaya et al., 2021; Ravens et al., 2014). In mouse embryonic stem cells (mESCs), KAT8 and H4K16ac mark active enhancers and promoters of genes that maintain the identity of mESCs (Li et al. 2012; Taylor et al. 2013). Loss of function mutations in KAT8 or *MSL3* accompanied by reduced H4K16ac were shown to cause neurodevelopmental disorders (Basilicata et al., 2018; Li et al., 2020). However, the mechanism through which MSL/KAT8-mediated H4K16ac contributes to genome regulation during normal development is less clear, especially in the human genome.

Here, by profiling histone modifications associated with regulatory elements, we show that unlike H3K27ac, which marks active genes and enhancers, H4K16ac and H3K122ac are enriched

across the L1 and ERV/LTR superfamily of TEs in human cell lines. Depletion of H4K16ac by knockout or knockdown of MSL proteins is sufficient to downregulate these TEs. We show that H4K16ac contributes to opening local chromatin structure and transcription at TEs. The genome-wide enhancer reporter assay data shows that H4K16ac marked TEs show higher enhancer reporter activity. Furthermore, CRISPR-mediated perturbation of H4K16ac marked TEs leads to the downregulation of genes associated in 3D space.

Results

H4K16ac and H3K122ac are enriched at TEs in human and mouse cell lines.

We aimed to investigate the role of lesser studied histone acetylations, such as H4K16ac and H3K122ac, in human genome regulation. We performed Cleavage Under Targets and Tagmentation (CUT&Tag) (Kaya-okur et al., 2019) on the histone modifications that are known to be associated with active regulatory elements (H3K27ac, H3K122ac, H4K12ac, H4K16ac, H3K4me1 and H3K4me3), polycomb repressed domains (H3K27me3) and heterochromatin (H3K9me3) (fig. S1A). While we focused our study on human embryonic stem cells (H9-hESCs), we also validated our findings in mouse embryonic stem cells (E14 mESCs), prostate (PC3, LNCaP), erythroleukemia (K562), ovarian (HeLa) and neuroblastoma (SH-SY5Y) cancer cells together with prostate epithelium (RWPE), embryonic kidney (HEK293T) and transformed dermal fibroblast (TDF) cells. CUT&Tag sequencing reads that mapped more than once were filtered out, and only uniquely mapped reads were retained for analyses. The Hidden Markov model's chromatin-state discovery and genome annotation analysis (12 ChromHMM states) revealed H3K4me1, H3K4me3, H3K27ac and H4K12ac's expected enrichment at genomic features associated with active transcription, including active promoter and enhancers in hESCs. H3K36me3 also showed expected enrichment at transcription elongation and weak transcription features. Intriguingly, H4K16ac and H3K122ac but not H3K27ac and H4K12ac were enriched at heterochromatin, insulator, and repetitive DNA (Fig. 1A). Further enrichment analysis at different classes of TEs from the RepeatMasker database and protein-coding genes revealed specific enrichment of H4K16ac and H3K122ac at the 5' UTR of full-length L1s, ERV/LTRs, and SINE/*Alu* elements (Fig. 1B, C, D, E). H4K16ac enrichment at a large number of TEs is consistent with the mass spectrometry data showing this acetylation is highly abundant, and nearly 30% of the histone H4 proteins are acetylated at lysine 16 (Radzisheuskaya et al., 2021). Apart from hESCs, the other seven tested human somatic cell lines also showed higher enrichment of H4K16ac at TEs than H3K27ac (fig. S1B, C). Consistent with the CUT&Tag data from human cell lines, reanalysis of publicly available H4K16ac ChIP-seq datasets from the human brain tissues (prefrontal lobe) (Nativio et al., 2018) showed specific enrichment of H4K16ac at 5' UTR of L1 subfamilies (fig S2A). Moreover, CUT&Tag data from mESCs showed that H4K16ac is enriched at 5' UTR of L1 and many TE families, suggesting enrichment of H4K16ac at TEs is not unique to human cell lines and tissues but also conserved in mice (fig. S2B, C).

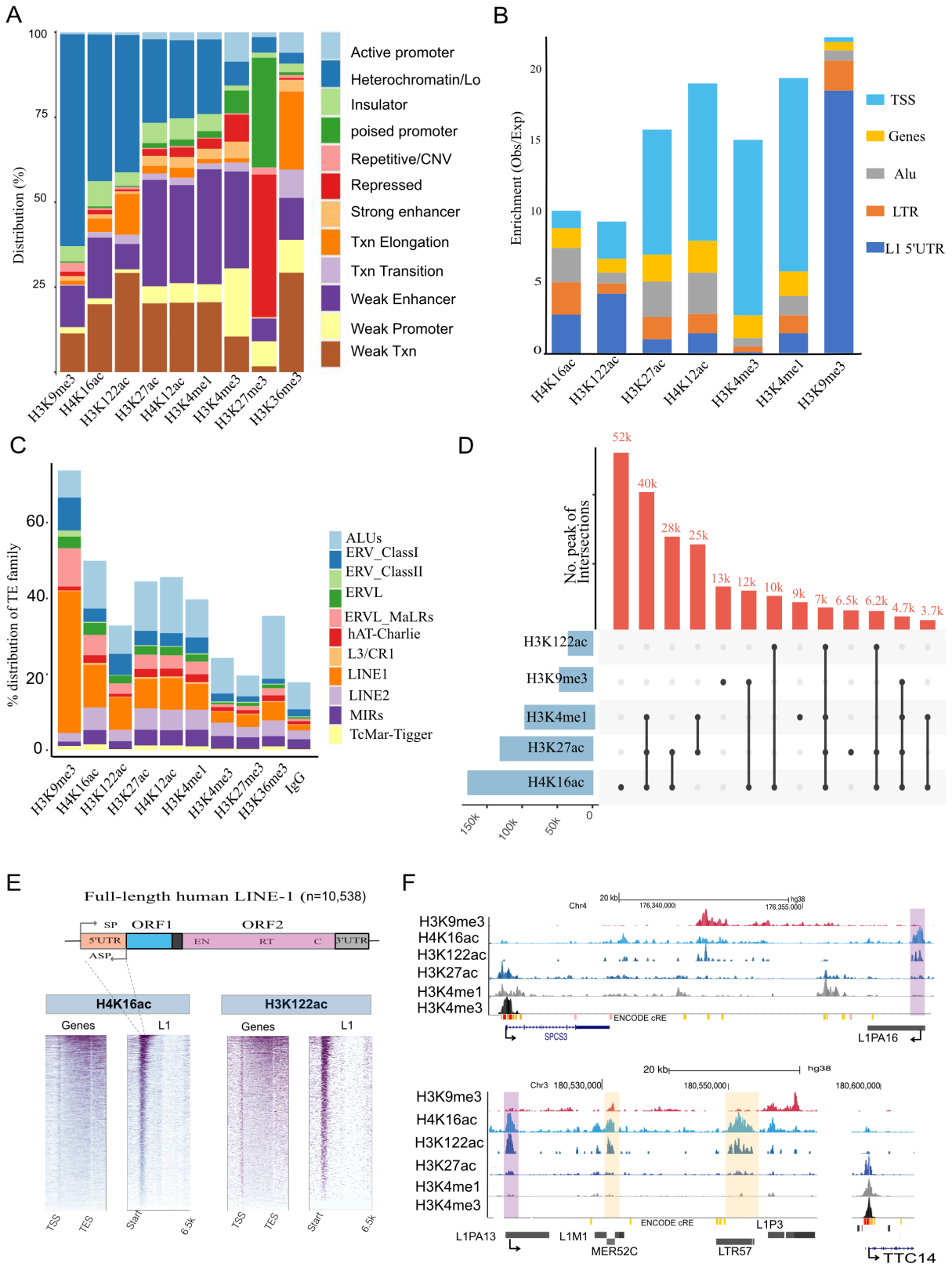
A
Distribution (%)
100
75
50
25
H3K9me3
H4K16ac
H3K122ac
H3K27ac
H4K12ac
H3K4me1
H3K4me3
H3K27me3
H3K36me3
Active promoter
Heterochromatin/Lo
Insulator
poised promoter
Repetitive/CNV
Repressed
Strong enhancer
Txn Elongation
Txn Transition
Weak Enhancer
Weak Promoter
Weak Txn
B
Enrichment (Obs/Exp)
20
15
10
5
0
H4K16ac
H3K122ac
H3K27ac
H4K12ac
H3K4me3
H3K4me1
H3K9me3
TSS
Genes
Alu
LTR
L1 5'UTR
C
% distribution of TE family
60
40
20
0
H3K9me3
H4K16ac
H3K122ac
H3K27ac
H4K12ac
H3K4me1
H3K4me3
H3K27me3
H3K36me3
IgG
ALUs
ERV_ClassI
ERV_ClassII
ERVL
ERVL_MaLRs
hAT-Charlie
L3/CR1
LINE1
LINE2
MIRs
TcMar-Tigger
D
No. peak of Intersections
52k
40k
28k
25k
13k
12k
10k
9k
7k
6.5k
6.2k
4.7k
3.7k
H3K122ac
H3K9me3
H3K4me1
H3K27ac
H4K16ac
150k
100k
50k
0
E
Full-length human LINE-1 (n=10,538)
SP
ORF1
ORF2
5'UTR
EN
RT
C
3'UTR
ASP
H4K16ac
H3K122ac
Genes
L1
Genes
L1
TSS
TES
Start
6.5k
TSS
TES
Start
6.5k
F
Chr4
20 kb
hg38
176,340,000
176,355,000
H3K9me3
H4K16ac
H3K122ac
H3K27ac
H3K4me1
H3K4me3
ENCODE cRE
SPCS3
L1PA16
20 kb
hg38
Chr3 180,530,000
180,550,000
180,600,000
H3K9me3
H4K16ac
H3K122ac
H3K27ac
H3K4me1
H3K4me3
ENCODE cRE
L1PA13
L1M1
MER52C
LTR57
L1P3
TTC14

Fig. 1: H4K16ac and H3K122ac are enriched at 5' UTR of L1 and ERV/LTRs in H9 hESCs
A. *Bar chart showing the percentage distribution (Y-axis) of CUT&Tag peaks from various histone PTMs (X-axis) across ChromHMM chromatin states.* ***B.*** *Enrichment (observed/expected) of histone modification peaks across gene transcription start-sites (TSS), TE families (L1, ERV/LTRs, SINE/Alu) and gene body.* ***C.*** *Percentage distribution of repeat element across histone modification CUT&Tag peaks.* ***D***. *Intersection of histone modification CUT&Tag peaks at TE (LTR, Alu and full-length L1) families, X-axis shows the number of peaks, and the Y-axis is the number of peaks intersected.* ***E***. *Above, illustration showing the structure of human L1. 5' and 3' untranslated regions (UTR), open reading frames (ORF1 and ORF2), sense and antisense promoters (SP and ASP). Below are heatmaps showing H4K16ac and H3K122ac CUT&Tag signal covering RefSeq genes and full length (>5kb, n=10538) L1s.* ***F***. *Example UCSC genome browser tracks showing average counts per million (CPM) for two replicates of H3K4me3, H3K4me1, H3K9me3, H4K16ac, H3K122ac and H3K27ac CUT&Tag data from hESCs. RepeatMasker tracks showing 5' UTRs of L1 (purple highlight) and ERV/LTRs (yellow highlight) and ENCODE cis-regulatory elements (CRE) are shown below.*

MSL mediated H4K16ac regulates the transcription of L1 and ERV/LTRs

Next, we aimed to investigate the role of H4K16ac at TEs in altering chromatin structure and transcription. Depletion of KAT8 affects H4K16ac, H4K5ac, H4K8ac and H4K12ac as KAT8 when associated with NSL complex catalyzes H4K5, H4K8 and H4K12 acetylations while it is bound to MSL complex it explicitly catalyzes H4K16ac (Fig. 2A). As knockout of individual MSL complex proteins such as MSL1, MSL2 and MSL3 are sufficient to deplete most of the H4K16ac (Monserrat et al., 2021). We depleted MSL3, in hESCs, using two independent lentiviral shRNAs (Fig. 2B, C). shRNA-mediated depletion of MSL3 and H4K16ac led to reduced L1-ORF1 and HERV at the protein level (Fig. 2C). Furthermore, RT-qPCR and RNA-seq analysis showed significant downregulation of both human-specific (L1HS) and primate-specific (L1PA2 to L1PA16) full-length L1 and LTR sub-family transcripts (Fig. 2B, E). Pluripotency and differentiation markers were unaffected upon MSL3 depletion, suggesting that downregulation of TEs is not due to the altered pluripotency state of hESCs (fig. S3).

Furthermore, to demonstrate that our findings are not restricted to hESCs, we also used doxycycline-inducible Cas9 (iCas9) mediated knockout (KO) of MSL1 and MSL3 in transformed dermal fibroblasts (TDFs) (Fig. 2F, G). MSL1 KO significantly reduced the H4K16ac level in TDFs (Fig. 2F, G). Similarly, MSL1 and MSL3 KO led to bulk depletion of H4K16ac levels and at TEs (Fig. 2F, G, and fig S4A). Furthermore, a reanalysis of RNAseq data from Monserrat et al. (Monserrat et al., 2021) showed significant downregulation of L1 and LTR transcripts in MSL1 and MSL3 KO TDFs (fig. S4). Notably, L1 and LTR regions that intersect with H4K16ac peaks were particularly downregulated in MSL1 KO cells compared to L1 and LTRs that lacked detectable levels of H4K16ac (Fig. 2H). Since H4K16ac decompacts chromatin *in vitro*, and has been shown to increase chromatin accessibility (Samata et al., 2020; Shogren-Knaak et al., 2006), we next asked whether the lack of H4K16ac leads to altered accessibility at TEs. Assay for transposase-accessible chromatin using sequencing (ATAC-seq) showed specific depletion of accessible DNA at the 5' UTR of L1s and LTRs in MSL3-depleted hESCs and MSL KO fibroblasts (Fig. 2D). Reanalysis of publicly available ATAC-seq data from MSL1 depleted monocyte cell line (THP1) (Radzisheuskaya et al., 2021) also showed reduced ATAC-seq signal at TEs. Interestingly, we found cell type specific variation in reduced DNA accessibility at different subfamilies of L1 and LTRs, suggesting the role of H4K16ac in cell type-specific regulation of TE subfamilies (Fig. 2D). Therefore, we conclude that MSL/KAT8-mediated H4K16ac leads to the opening of chromatin structure and increased transcription at TEs.

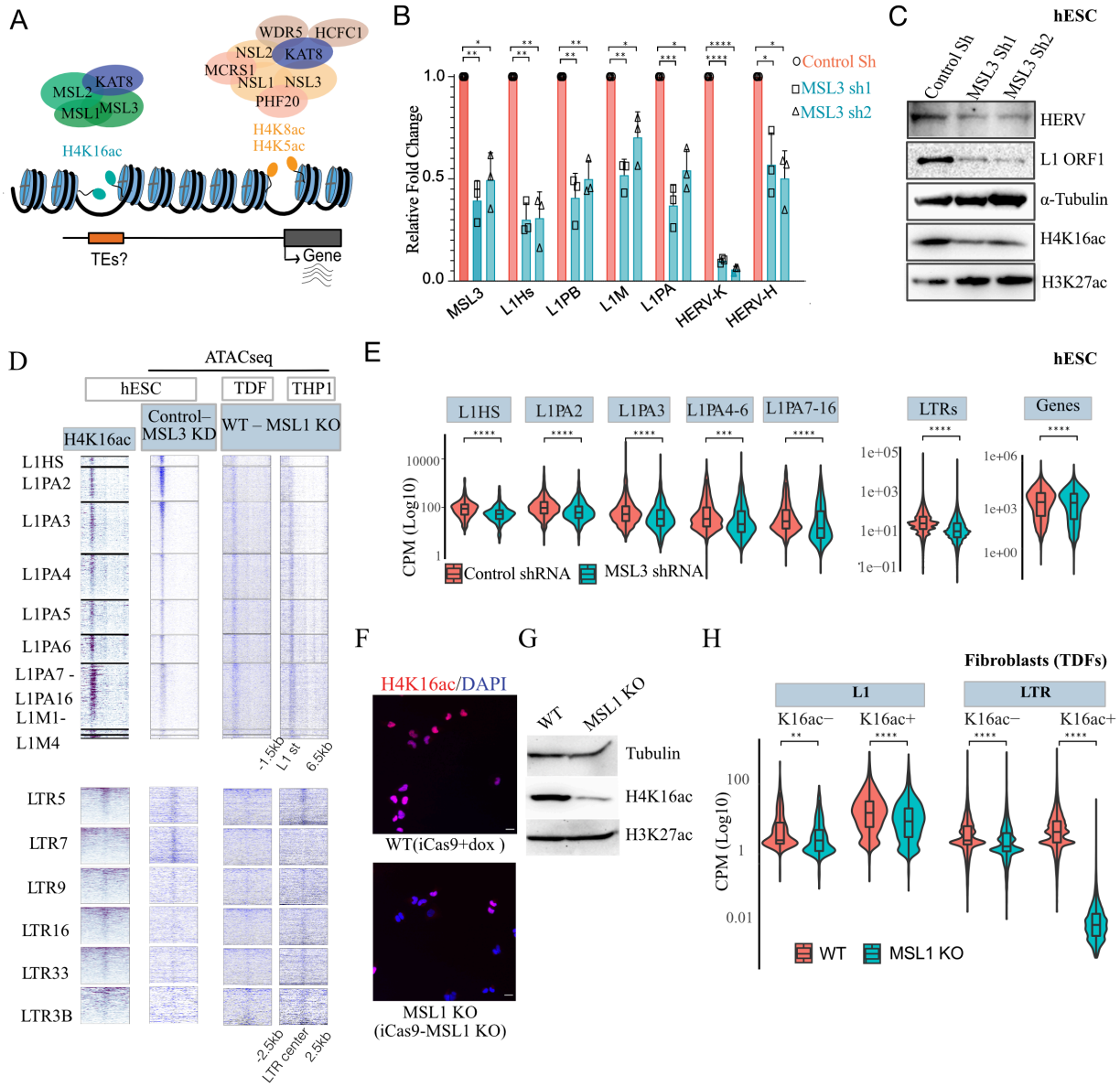

Fig. 2 MSL, which specifically deposits H4K16ac and regulates the transcription of TEs.
A. *Illustration showing the presence of KAT8 in two distinct complexes, KAT8 in the MSL complex catalyzes H4K16ac at TEs, while in the NSL complex, it catalyzes H4K5ac, H4K8ac and H4K12ac at gene TSSs.* ***B.*** *RT-qPCR data showing mean (±SD, n=3 independent experiments) fold change (normalized to β-ACTIN) in transcript level for MSL3, subfamilies of L1s (L1Hs, L1PB, L1M and L1PA) and ERVs (HERV-K gag and HERV-H gag).* ***C.*** *Western blot showing HERV envelope protein, L1-ORF1 and H4K16ac levels upon knockdown of MSL3 using two independent shRNAs (MSL3 sh1 and MSL3 sh2) versus non-target control. α-Tubulin and H3K27ac served as controls.* ***D.*** *Heatmap showing the fold change in ATACseq signal at full-length L1 and LTRs in MSL3 knockdown in hESCs, MSL1 knockout (KO) in TDF and THP1 cell line normalized to respective controls* (Radzisheuskaya et al., 2021) *compared to WT control.* ***E.*** *Violin plots showing the sum of RNAseq reads as counts per million (CPM) normalized with ERCC spike-in control read counts at LTR, Alu and full-length L1 sub-families and genes from MSL3 shRNA and control hESCs. Boxes indicate the median and interquartile range, with whiskers showing the first and the fourth quartile.* ***F.*** *Immunofluorescence images showing H4K16ac levels (Magenta) in TDFs after doxycycline induced MSL1 knock-out (MSL KO) (day 4) and in a parental cell line (iCas9).* ***G.*** *Western blots showing the levels of H4K16ac MSL1 KO and control cells, Tubulin and H3K27ac served as controls.* ***H.*** *Like E, but for RNA*

levels at L1 and LTRs that intersect with H4K16ac peaks (K16ac+) vs. L1 and LTRs that lack H4K16ac peaks (K16ac–). P-values and statistical tests for all RT-qPCR and violin plots are detailed in data S2.

H4K16ac marked L1 and LTRs function as enhancers.

LTR subfamilies, particularly LTR5 and LTR7 function as enhancers in hESCs (Chuong et al., 2017; Fuentes et al., 2018; Pontis et al., 2019). However, it is not known whether L1s can also act as enhancer elements. We have previously shown that H4K16ac and H3K122ac mark active enhancers in mESCs (Pradeepa et al., 2016; Taylor et al., 2013). Here we show that H4K16ac, H3K122ac and H3K27ac were particularly enriched at the 5' UTR of full-length L1s and correlate with chromatin features associated with active enhancers such as H3K27ac, H4K16ac, H3K4me1, BRD4 and ATACseq signal in hESCs (Fig. 1C and Fig. 3A). H4K16ac was notably higher in older (L1PA7 to L1PA16) compared to younger primate-specific (L1PA1 to L1PA6) and human-specific (L1HS) L1s. Analysis of genome-wide enhancer activity data (STARRseq) from neuroblastoma (SH-SY5Y) and erythroleukemia (K562) cell lines generated by ENCODE (Lee et al., 2020) revealed high enhancer activity specifically at the 5' UTR of the L1 family (Fig. 3B). The presence of chromatin features associated with active enhancers (Fig. 3A), together with the ability of L1 5' UTR to drive transcription in an in vitro enhancer activity assay (Fig. 3B), suggests that L1s can function as enhancers to regulate genes in cis.

Apart from LTRs with known enhancer activity (LTR5 and LTR7), our data show LTR9, LTR16, LTR33 and MER21C also serve as enhancers as they are enriched with H4K16ac along with other active enhancer chromatin features (Fig. 3C, E). Active enhancers can be distinguished from inactive enhancers based on the co-occurrence of H3K4me1 with H3K27ac, H3K122ac, H4K16ac and enhancer RNA transcription (Pradeepa, 2017). LTRs with H4K16ac and H3K4me1 peaks showed significantly higher enhancer activity (STARRseq signal) than LTRs that intersected with only H3K4me1 or H4K16ac (Fig. 3D). This is consistent with the specific downregulation of LTR transcripts upon depletion of MSL/H4K16ac (Fig. 2B, C, E and fig S4B, C), supporting the role of H4K16ac in the transcription of LTR transcripts. Our findings show that 5' UTR of full-length L1 and LTRs marked with H4K16ac and H3K4me1 are likely to function as enhancer elements. The rest of the LTR and Alu families are not likely to act as active enhancers in hESCs, as they either lack enhancer chromatin features and/or enhancer activity (fig. S5A). Reduced chromatin accessibility across L1 and LTR subfamilies upon depletion of MSL/H4K16ac (Fig. 2D), further supports the role of H4K16ac in enhancer activity. In hESCs, LTR5 and LTR7 show a more significant reduction in chromatin accessibility, consistent with the known function of these LTR subfamilies as enhancers in this cell type. Interestingly, cell type-specific reduction in chromatin accessibility is observed across L1 and LTR subfamilies (Fig. 3B, C), indicating that these LTR subfamilies can act as enhancers in a cell type-specific manner.

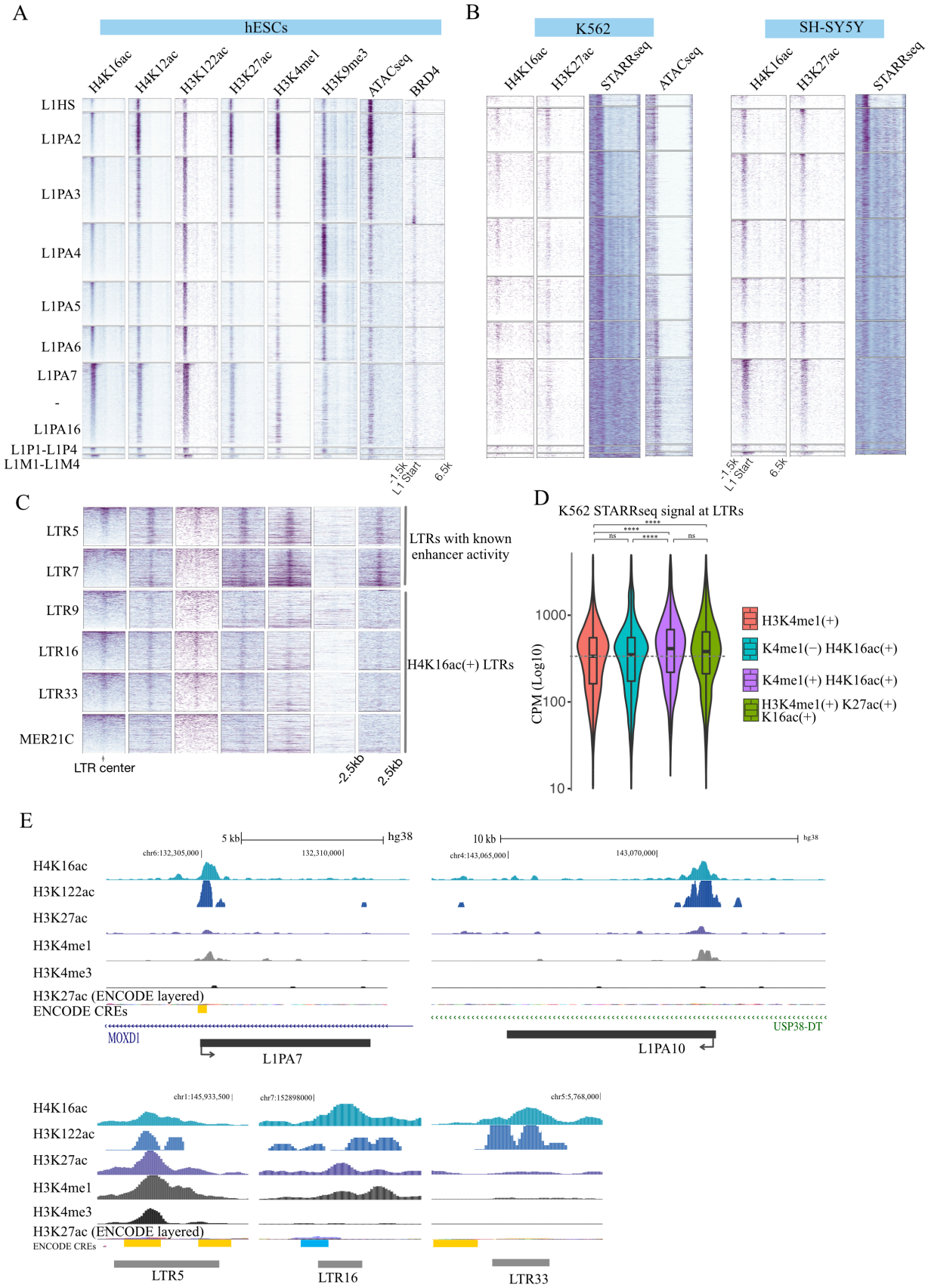
A
hESCs
H4K16ac
H4K12ac
H3K122ac
H3K27ac
H3K4me1
H3K9me3
ATACseq
BRD4
L1HS
L1PA2
L1PA3
L1PA4
L1PA5
L1PA6
L1PA7
-
L1PA16
L1P1-L1P4
L1M1-L1M4
-1.5k
L1 Start
6.5k
B
K562
SH-SY5Y
H4K16ac
H3K27ac
STARRseq
ATACseq
C
LTR5
LTR7
LTR9
LTR16
LTR33
MER21C
LTRs with known enhancer activity
H4K16ac(+) LTRs
LTR center
-2.5kb
2.5kb
D
K562 STARRseq signal at LTRs
CPM (Log10)
1000
100
10
ns
****
H3K4me1(+)
K4me1(−) H4K16ac(+)
K4me1(+) H4K16ac(+)
H3K4me1(+) K27ac(+) K16ac(+)
E
5 kb
hg38
chr6:132,305,000
132,310,000
10 kb
chr4:143,065,000
143,070,000
H4K16ac
H3K122ac
H3K27ac
H3K4me1
H3K4me3
H3K27ac (ENCODE layered)
ENCODE CREs
MOXD1
L1PA7
USP38-DT
L1PA10
chr1:145,933,500
chr7:152898000
chr5:5,768,000
LTR5
LTR16
LTR33

Fig. 3 H4K16ac marked TEs function as enhancers to regulate genes *in cis*.
A. *Heatmap of CUT&Tag signals for histone modifications and BRD4 normalized to IgG control, ATACseq signal at TE sub-families; –1.5kb to +6.5kb from L1 (L1HS, L1PA2, L1PA3, L1PA4, L1PA5, L1PA6, L1PA7-L1PA16, L1P1-L1P4 and L1M1-L1M4).* ***B.*** *B) Heatmap shows H4K16ac and H3K27ac CUT&Tag signals and STARRseq signals in K562 and SH-SY5Y cells* ***C.*** *Like A, but for ±2.5kb around ERV/LTR center (LTR5, 7, 9, 16, 33 MER21C, HERV and ERV3, data for rest of the LTR sub-families in fig S5).* ***D.*** *Violin plots showing K562 STARRseq signal at LTRs that intersect with H3K4me1 peaks; H3K4me1, H3K27ac and H4K16ac peaks; H4K16ac peaks but not H3K4me1 peaks; and H3K4me1, H4K16ac peaks. Statistical tests are detailed in table S2.* ***E.*** *Example UCSC genome browser tracks showing average counts per million (CPM) for two replicates of H4K16ac, H3K122ac, H3K27ac, H3K4me1 and H3K4me3 CUT&Tag data from hESCs. RepeatMasker tracks showing L1, LTR5, LTR16 and LTR33 and ENCODE layered H3K27ac and cis-regulatory elements (CRE) are shown below.*

H4K16ac marked TEs are bound by Cohesin and CTCF and contribute to chromatin topology.

We carried out a TF enrichment analysis to determine whether H4K16ac-specific TEs act as enhancers by recruiting specific TFs. This revealed expected enrichment for EP300 at H3K27ac marked but not H4K16ac and H3K122ac marked TEs (Fig. 4A). YY1 was observed to be enriched at L1s harboring all three acetylations, supporting the previously known role of YY1 in the regulation of L1 transcription (Athanikar et al 2004). MSL complex recruits YY1 to the promoter region of *Tsix* to activate its expression in mESCs (Chelmicki et al., 2014), suggesting a possible interplay between these two factors in regulating L1 transcription. Several TFs were uniquely enriched at H4K16ac marked TEs (Fig. 4A, fig. S7A), suggesting these TFs could contribute to maintaining H4K16ac and H3K122ac levels along with transcriptional activity at these TEs. Interestingly, architectural proteins such as CTCF, RAD21, and their cofactor ZNF143 were enriched at H4K16ac- rather than H3K27ac- marked TEs (Fig. 4A). We, therefore, asked whether H4K16ac marked TEs together with the CTCF, RAD21, and ZNF143 could contribute to 3D genome folding at these loci. Analysis of Hi-C data revealed that TEs that overlap with histone acetylation peaks are closer to topologically associated domain (TAD) borders than TEs that lack acetylations (Fig. 4B). We also found that H4K16ac levels are relatively higher at the TEs overlapping with the TAD borders than at the TEs that don't overlap with TAD borders (fig. S6). Furthermore, since transcriptionally active HERVs contribute to TAD boundary formation (Zhang et al., 2019) and MSL/H4K16ac axis drives the transcription of TEs, including HERVs (Fig. 2B, C and E, fig S4B), it is likely that H4K16ac-mediated transcription at TEs contributes to TAD border formation and 3D chromatin organization. This, in turn, could prevent the spread of heterochromatinization at TEs by repressor complexes.

CRISPR interference (CRISPRi) based perturbation of H4K16ac/H3K122ac marked TEs revealed their role expression of genes *in cis*

To demonstrate that TEs were regulating transcription of nearby genes, we visualized Micro-C data and identified putative target genes for H4K16ac marked L1/HERVs. Micro-C data from hESCs revealed that L1 and HERVs marked by H4K16ac and H3K122ac overlap with TAD borders or interact with the promoter region of putative target genes (Fig. 4C, fig. S7). To further validate H4K16ac and H3K122ac marked L1 and HERVs regulate genes *in vivo* in the endogenous context, we performed CRISPRi by transiently transfecting plasmids encoding guide RNAs targeting L1 5' UTR and HERV/LTR elements to hESCs stably transduced with dCAS9-KRAB. RT-qPCR analysis showed downregulation of putative target genes *USP38* and *TANC2*, located at ~110kb and ~270kb respectively, from H4K16ac and H3K122ac marked L1 5' UTRs. CRISPRi for LTR18/HERVL located ~10kb from the *ODF2L* promoter led to the downregulation of *ODFL*. Similarly, CRISPRi for LTR7/HERVH located between *GATAD1* and *PEX1* locus led to downregulation of both GATAD1 and PEX1 (Fig 4D).

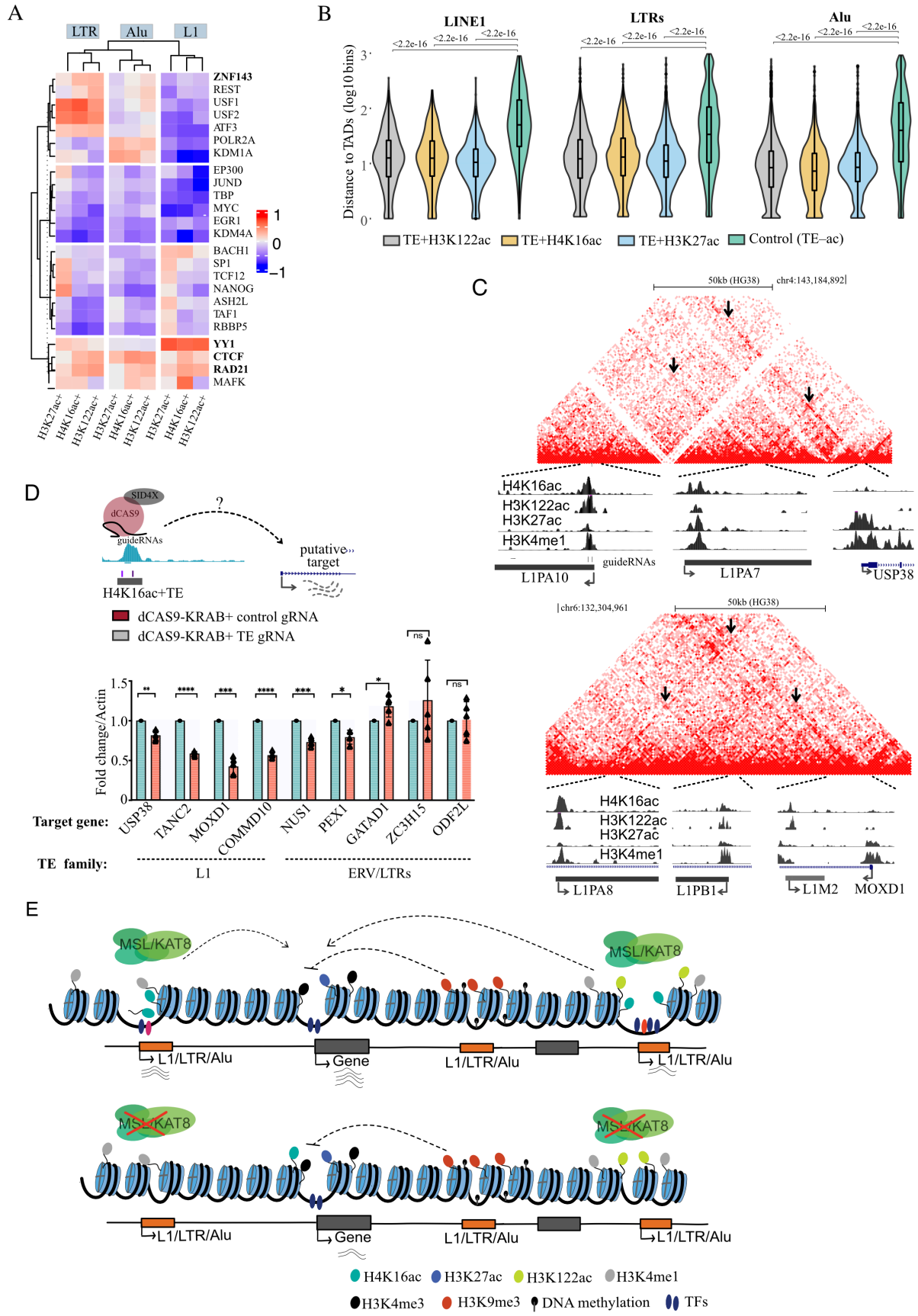
A
LTR
Alu
L1
ZNF143
REST
USF1
USF2
ATF3
POLR2A
KDM1A
EP300
JUND
TBP
MYC
EGR1
KDM4A
BACH1
SP1
TCF12
NANOG
ASH2L
TAF1
RBBP5
YY1
CTCF
RAD21
MAFK
H3K27ac+
H4K16ac+
H3K122ac+
B
LINE1
LTRs
Alu
Distance to TADs (log10 bins)
TE+H3K122ac
TE+H4K16ac
TE+H3K27ac
Control (TE–ac)
C
50kb (HG38)
chr4:143,184,892
H4K16ac
H3K122ac
H3K27ac
H3K4me1
guideRNAs
L1PA10
L1PA7
USP38
chr6:132,304,961
L1PA8
L1PB1
L1M2
MOXD1
D
SID4X
dCAS9
guideRNAs
putative target
H4K16ac+TE
dCAS9-KRAB+ control gRNA
dCAS9-KRAB+ TE gRNA
Fold change/Actin
Target gene:
USP38
TANC2
MOXD1
COMMD10
NUS1
PEX1
GATAD1
ZC3H15
ODF2L
TE family:
L1
ERV/LTRs
E
MSL/KAT8
L1/LTR/Alu
Gene
H4K16ac
H3K27ac
H3K122ac
H3K4me1
H3K4me3
H3K9me3
DNA methylation
TFs

Fig 4. H4K16ac is enriched at LTRs and Alu with TAD borders, and loss of H4K16ac leads to reduced DNA accessibility at TEs

A. *Heatmap showing differentially enriched TF Motif (columns) in H4K16ac, H3K27ac, and H3K122ac peaks at L1 5' UTR, ERV/LTR and SINE/Alu versus random background, a complete list of TF motif data in fig. S6.* ***B.*** *Violin plot showing the distance to TAD borders (Y-axis, log 10 bins) for LTR, Alu and L1 with H3K122ac, H4K16ac and H3K122ac, TEs that do not intersect with these three histone acetylation peaks served as control.* ***C.*** *UCSC genome browser tracks showing H4K16ac, H3K122ac and H3K27ac signals from H9 hESC cells at candidate L1s used for functional validation by CRISPRi. Arrows in the Micro-C map indicate the interaction of L1s with gene promoters; the arrowhead indicates the TAD border. Details of the TE coordinates and gRNA positions are given in table S4. The statistical tests are detailed in table S2.* ***D****. Illustration shows the CRISPRi validation design of candidate L1s & ERVs in hESCs (above). RT-qPCR shows a fold change in RNA level normalized to β-ACTIN for putative target genes of L1, and ERV/LTR elements (selected based on the distance from TE and Micro-C data) upon guide RNA mediated targeting of dCAS9-KRAB to TEs; nontargeting guide RNA vector served as control (below).* ***E.*** *The working model shows MSL/KAT8 mediated H4K16ac activates transcription of TEs and contributes to enhancer activity of TEs to regulate genes in cis (above). Depleting MSL-mediated H4K16ac leads to a reduced level of accessible chromatin, TE transcription and expression of genes in cis (below).*

Our previous work showed that recruitment of dCas9 fused with four copies of repressive Sin3 interacting domain (dCAS9-SID4X) leads to histone deacetylation at enhancers and reduced expression of their target genes (Pradeepa *et al.*, 2016). The use of the dCAS9-Sid4x based CRISPRi approach in HEK293T also showed downregulation of putative target genes (fig. S7F), further confirming the dCAS9-KRAB based CRISPRi data from hESCs. Therefore, we conclude that H4K16ac and H3K122ac marked L1 5' UTRs and LTRs function as enhancers. As H4K16ac and H3K122ac marked TEs constitute a substantial number of TE elements, they contribute significantly to regulatory activity in the mammalian genome.

Neuronal cell types have higher L1 expression and high efficiency in the retrotransposition (Macia et al., 2017); retrotransposon dysregulation is also linked with neurological disorders (Hancks and Kazazian, 2016). Consistent with our findings, analysis of publicly available H4K16ac ChIPseq data from human brain tissues revealed specific enrichment of H4K16ac at 5' UTRs of L1s (Nativio et al., 2018)(fig. S2). Loss of function mutations in genes encoding KAT8 containing protein complexes such as *KANSL1, MSL3 and KAT8* lead to neurodevelopmental disorders (Basilicata et al., 2018; Koolen et al., 2012; Sharp et al., 2006; Shaw-Smith et al., 2006). Further investigation is needed on the effect of MSL/KAT8 disruption on TE-mediated genome regulation in neurodevelopmental disease models. Depleting MSL complex proteins and subsequent loss of H4K16ac exhausts the ability of cells to proliferate and increases the chromosomal instability (Monserrat et al., 2021). Highly proliferative stem and cancer cells have high transcriptional activity at TEs. Reanalysis of H4K16ac ChIP-seq data also shows a dramatic loss of H4K16ac across 5' UTR of L1s and LTRs in senescent cells compared to proliferating cells (fig. S8), suggesting that proliferating cells have adapted to permissive chromatin structure at TEs.

Conclusions

TEs have contributed significantly to the evolution of mammalian genomes by helping to shape both the coding and non-coding regulatory landscape. Here we demonstrate that H4K16ac and H3K122ac contribute to the cis-regulatory activity of human and primate-specific L1 and HERV families. The role of MSL complex to co-opt TEs to rewire cis-regulatory elements also appears to be conserved during the evolution of dosage compensation in *Drosophila miranda,* where a mutant helitron TE is shown to recruit MSL complex to evolutionarily young X chromosome to increase transcription (Christopher and Bachtrog, 2013). Although most TEs are epigenetically repressed, silencing means must be partially released to permit developmental or tissue-specific regulation of gene expression programs, discussed in (Bourque et al., 2018). Several layers of TE silencing mechanisms are known; here, for the first time, we demonstrate that transcriptional activator complex (MSL/KAT8) mediated H4K16ac activates the transcription of TEs. Increased DNA accessibility due to H4K16ac and H3K122ac at TEs leads to increased transcription, which could also counteract the epigenetic repressive environment at TEs (Fig. 4E) (Liu et al., 2018). As H4K16ac increases transcription of TEs, we propose that the act of transcription of TEs could contribute to chromatin topology and enhancer mediated regulation of host gene expression in cis, as many TEs are located near genes, within introns and enriched at TAD borders.

References

Almeida, M.V., Vernaz, G., Putman, A.L.K., and Miska, E.A. (2022). Trends in Genetics Taming transposable elements in vertebrates : from epigenetic silencing to domestication. Trends Genet. 1–25.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Athanikar, J.N., Badge, R.M., and Moran, J. V. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res. *32*, 3846–3855.

Basilicata, M.F., Bruel, A.-L., Semplicio, G., Valsecchi, C.I.K., Aktaş, T., Duffourd, Y., Rumpf, T., Morton, J., Bache, I., Szymanski, W.G., et al. (2018). De novo mutations in MSL3 cause an X-linked syndrome marked by impaired histone H4 lysine 16 acetylation. Nat. Genet. *50*, 1.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. Genome Biol. *19*, 1–12.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Burns, K.H. (2017). Transposable elements in cancer. Nat. Rev. Cancer *17*, 415–424.

Chatterjee, A., Seyfferth, J., Lucci, J., Pfanner, N., Becker, T., Akhtar, A., Panhale, A., Stehle, T., Kretz, O., Sahyoun, A.H., et al. (2016). Acetyl Transferase Regulates Transcription and Respiration in Mitochondria. Cell *167*, 722–738.

Chelmicki, T., Dündar, F., Turley, M.J., Khanam, T., Aktas, T., Ramírez, F., Gendrel, A.-V., Wright, P.R., Videm, P., Backofen, R., et al. (2014). MOF-associated complexes ensure stem cell identity and Xist repression. Elife *3*, e02024.

Christopher, E.E., and Bachtrog, D. (2013). Dosage compensation via transposable element

mediated rewiring of a regulatory network. Science (80-.). *342*, 846–850.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Fuentes, D.R., Swigut, T., and Wysocka, J. (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife *7*, 1–29.

Fueyo, R., Judd, J., and Feschotte, C. (2022). Roles of transposable elements in the. Nat. Rev. Mol. Cell Biol. *24*, 19–24.

Hancks, D.C., and Kazazian, H.H. (2016). Roles for retrotransposon insertions in human disease. Mob. DNA *7*.

Hermant, C., and Torres-Padilla, M.E. (2021). TFs for TEs: The transcription factor repertoire of mammalian transposable elements. Genes Dev. *35*, 22–39.

Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat. Genet. *49*, 1502–1510.

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics *13*, 424.

Kaya-okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for ef fi cient epigenomic pro fi ling of small samples and single cells. Nat. Commun. *10*, 1–10.

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. Genome Res. *23*, 800–811.

Koolen, D.A., Kramer, J.M., Neveling, K., Nillesen, W.M., Moore-barton, H.L., Elmslie, F. V, Toutain, A., Amiel, J., Malan, V., Tsai, A.C., et al. (2012). Mutations in the chromatin modifier gene KANSL1 cause the. Nat. Genet. *44*, 2011–2013.

Lee, D., Shi, M., Moran, J., Wall, M., Zhang, J., Liu, J., Fitzgerald, D., Kyono, Y., Ma, L., White, K.P., et al. (2020). STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. Genome Biol. *21*, 1–24.

Li, L., Ghorbani, M., Weisz-Hubshman, M., Rousseau, J., Thiffault, I., Schnur, R.E., Breen, C., Oegema, R., Weiss, M.M.M., Waisfisz, Q., et al. (2020). Lysine acetyltransferase 8 is involved in cerebral development and syndromic intellectual disability. J. Clin. Invest. *130*, 1431–1445.

Li, X., Li, L., Pandey, R., Byun, J.S., Gardner, K., Qin, Z., and Dou, Y. (2012). The Histone Acetyltransferase MOF Is a Key Regulator of the Embryonic Stem Cell Core Transcriptional Network. Cell Stem Cell *11*, 163–178.

Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C., and Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators.

Nature *553*, 228–232.

Macia, A., Widmann, T.J., Heras, S.R., Ayllon, V., Sanchez, L., Benkaddour-boumzaouad, M., Muñoz-lopez, M., Rubio, A., Amador-cubero, S., Blanco-jimenez, E., et al. (2017). Engineered LINE-1 retrotransposition in nondividing human neurons. Genome Res. *27*, 335–348.

Martin, B.J.E., Brind'Amour, J., Kuzmin, A., Jensen, K.N., Liu, Z.C., Lorincz, M., and Howe, L.A.J. (2021). Transcription shapes genome-wide histone acetylation patterns. Nat. Commun. *12*, 1–9.

Molaro, A., and Malik, H.S. (2016). Hide and seek: How chromatin-based pathways silence retroelements in the mammalian germline. Curr. Opin. Genet. Dev. *37*, 51–58.

Monserrat, J., Morales Torres, C., Richardson, L., Wilson, T.S., Patel, H., Domart, M.C., Horswell, S., Song, O.R., Jiang, M., Crawford, M., et al. (2021). Disruption of the MSL complex inhibits tumour maintenance by exacerbating chromosomal instability. Nat. Cell Biol. *23*, 401–412.

Nativio, R., Donahue, G., Berson, A., Lan, Y., Amlie-Wolf, A., Tuzer, F., Toledo, J.B., Gosai, S.J., Gregory, B.D., Torres, C., et al. (2018). Publisher Correction: Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease (Nature Neuroscience, (2018), 21, 4, (497-505), 10.1038/s41593-018-0101-9). Nat. Neurosci. *21*, 1018.

Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell *174*, 391-405.e19.

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T.W., Jaenisch, R., and Trono, D. (2019). Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell *24*, 724-735.e5.

Pradeepa, M.M. (2017). Causal role of histone acetylations in enhancer function. Transcription *8*, 40–47.

Pradeepa, M.M., Grimes, G.R., Kumar, Y., Olley, G., Taylor, G.C.A., Schneider, R., and Bickmore, W.A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. Nat. Genet. *48*, 681–686.

Radzisheuskaya, A., Shliaha, P. V., Grinev, V. V., Shlyueva, D., Damhofer, H., Koche, R., Gorshkov, V., Kovalchuk, S., Zhan, Y., Rodriguez, K.L., et al. (2021). Complex-dependent histone acetyltransferase activity of KAT8 determines its role in transcription and cellular homeostasis. Mol. Cell *81*, 1749-1765.e8.

Ravens, S., Fournier, M., Ye, T., Stierle, M., Dembele, D., Chavant, V., and Tora, L. (2014). Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. Elife *2014*, 1–23.

Samata, M., Alexiadis, A., Richard, G., Georgiev, P., Nuebler, J., Kulkarni, T., Renschler, G., Basilicata, M.F., Zenk, F.L., Shvedunova, M., et al. (2020). Intergenerationally Maintained Histone H4 Lysine 16 Acetylation Is Instructive for Future Gene Activation. Cell *182*, 127-144.e23.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonalves, Â., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel

genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348.

Sharp, A.J., Hansen, S., Selzer, R.R., Cheng, Z., Regan, R., Hurst, J.A., Stewart, H., Price, S.M., Blair, E., Hennekam, R.C., et al. (2006). Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat. Genet. *38*, 1038–1042.

Shaw-Smith, C., Pittman, A.M., Willatt, L., Martin, H., Rickman, L., Gribble, S., Curley, R., Cumming, S., Dunn, C., Kalaitzopoulos, D., et al. (2006). Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat. Genet. *38*, 1032–1037.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844–847.

Sundaram, V., and Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philos. Trans. R. Soc. B Biol. Sci. *375*.

Taylor, G., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M., and Bickmore, W.A. (2013). H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. *23*, 2053–2065.

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Bühler, M., et al. (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell *152*, 859–872.

Upton, K.R., Gerhardt, D.J., Jesuadian, J.S., Richardson, S.R., Sánchez-Luque, F.J., Bodea, G.O., Ewing, A.D., Salvador-Palomeque, C., Van Der Knaap, M.S., Brennan, P.M., et al. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. Cell *161*, 228–239.

Wang, Z., Chivu, A.G., Choate, L.A., Rice, E.J., Miller, D.C., Chu, T., Chou, S., Kingsley, N.B., Petersen, J.L., Finno, C.J., et al. (2022). Prediction of histone post-translational modification patterns based on nascent transcription data. Nat. Genet. *54*, 295–305.

Wolfe, J.C., Mikheeva, L.A., Hagras, H., and Zabet, N.R. (2021). An explainable artificial intelligence approach for decoding the enhancer histone modifications code and identification of novel enhancers in Drosophila. Genome Biol. *22*, 1–23.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

Acknowledgments: We thank QMUL epigenetics hub members for discussions and reading the manuscript. We thank Ludovic Vallier (Cambridge UK, with MTA from WiCell) for sharing the H9 cell line. pATn5 expression plasmid was a kind gift from Steve Henikoff (Fred Hutchinson Cancer Research Center) lab. We thank Edda Schulz (Max Planck Institute for Molecular Genetics) and Chema Martin (Queen Mary University of London) labs for sharing purified pATn5. Pankaj Dubey, Ivan Alic and Aoife Murrey help in hESC cell culture. This research utilized Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT.

Funding:

Medical Research Council UKRI/MRC grant (MR/T000783/1) (MMP, DP, MP, FB)
Barts charity small grant (MGU0475) (MMP). Marie Skłodowska-Curie grant 896079 (JS)

Author contributions:

Conceptualization: MMP, DP, MP

Methodology: MMP, DP, MP, FB, JS, MRB, PS, OAG, NRZ

Investigation: MMP, DP, MP, FB, JS, SB

Visualization: MMP, DP, MP, FB, MRB

Funding acquisition: MMP

Project administration: MMP

Supervision: MMP

Writing – original draft: MMP, DP, MP

Writing – review & editing: MMP, DP, MP, FB, JS, MRB

Competing interests: Authors declare that they have no competing interests.