LETTER

doi:10.1038/nature25179

Selective silencing of euchromatic L1s revealed by genome–wide screens for L1 regulators

Nian Liu[1]*, Cameron H. Lee[2]*, Tomek Swigut[1], Edward Grow[2]†, Bo Gu[1], Michael C. Bassik[2,3] & Joanna Wysocka[1,4,5,6]

Transposable elements, also known as transposons, are now recognized not only as parasitic DNA, the spread of which in the genome must be controlled by the host, but also as major players in genome evolution and regulation[1–6]. Long interspersed element-1 (LINE-1, also known as L1), the only currently autonomous mobile transposon in humans, occupies 17% of the genome and generates inter- and intra-individual genetic variation, in some cases resulting in disease[1–7]. However, how L1 activity is controlled and the function of L1s in host gene regulation are not completely understood. Here we use CRISPR–Cas9 screening strategies in two distinct human cell lines to provide a genome-wide survey of genes involved in the control of L1 retrotransposition. We identify functionally diverse genes that either promote or restrict L1 retrotransposition. These genes, which are often associated with human diseases, control the L1 life cycle at the transcriptional or the post-transcriptional level in a manner that can depend on the endogenous L1 nucleotide sequence, underscoring the complexity of L1 regulation. We further investigate the restriction of L1 by the protein MORC2 and by the human silencing hub (HUSH) complex subunits MPP8 and TASOR[8]. HUSH and MORC2 can selectively bind evolutionarily young, full-length L1s located within transcriptionally permissive euchromatic environments, and promote deposition of histone H3 Lys9 trimethylation (H3K9me3) for transcriptional silencing. Notably, these silencing events often occur within introns of transcriptionally active genes, and lead to the downregulation of host gene expression in a HUSH-, MORC2-, and L1-dependent manner. Together, these results provide a rich resource for studies of L1 retrotransposition, elucidate a novel L1 restriction pathway and illustrate how epigenetic silencing of transposable elements rewires host gene expression programs.

Most of our knowledge about the control of L1 retrotransposition comes from studies examining individual candidate genes[2–6]. To systematically identify genes regulating L1 retrotransposition, we performed a genome-wide CRISPR–Cas9 screen in human chronic myeloid leukaemia K562 cells using an L1-G418^R retrotransposition reporter[9] (in which G418^R indicates resistance to the antibiotic G418, also known as geneticin) (Fig. 1a, b). The L1-G418^R reporter was modified to be driven by a doxycycline (dox)-responsive promoter, rather than the native L1 5′ untranslated region (5′ UTR), to avoid leaky retrotransposition before the functional screen (Extended Data Fig. 1a–c). The cells become G418^R antibiotic-resistant only when the L1-G418^R reporter undergoes a successful retrotransposition event after dox induction (Fig. 1b). For the screen, we transduced clonal L1-G418^R cells with a lentiviral genome-wide single-guide RNA (sgRNA) library such that each cell expressed a single sgRNA[10]. We then treated the cells with dox to activate the L1-G418^R reporter for retrotransposition, and split the cells into G418-selected conditions and unselected conditions, which served to eliminate cell growth bias in the screen analysis. The frequencies of sgRNAs in the two populations were measured by deep sequencing (Fig. 1a) and analysed using Cas9 high-throughput maximum likelihood estimator (CasTLE)[11]. Consequently, cells transduced with sgRNAs targeting L1 suppressors would have more retrotransposition events than negative control cells and would be enriched through the G418 selection; conversely, cells transduced with sgRNAs targeting L1 activators would be depleted.

Using this strategy, we identified 25 putative L1 regulators at a 10% false discovery rate (FDR) cutoff, and 150 genes at a 30% FDR cutoff (Fig. 1c and Extended Data Fig. 1d; see Supplementary Table 1 for the full list). Despite low statistical confidence, many of the 30%-FDR-cutoff genes overlapped previously characterized L1 regulators (for example, *ALKBH1* and *SETDB1*) and genes functioning in complexes with our top 10% FDR hits (for example, the Fanconi anaemia pathway and the HUSH complex), suggesting that they probably encompassed biologically relevant hits. To increase statistical power in distinguishing bona fide L1 regulators among these, we performed a high-coverage secondary screen targeting the 30% FDR hits (150 genes) and an additional 100 genes that were either functionally related to our top hits or were otherwise previously known to regulate L1 but fell outside of the 30%-FDR-cutoff threshold (see Supplementary Table 2 for the full list). This secondary screen validated 90 genes out of the top 150 genome-wide screen hits, a fraction close to that expected with the 30% FDR cutoff (Fig. 1d and Extended Data Fig. 2a–c).

Our two-tier screening approach identified 142 human genes that either activate or repress L1 retrotransposition in K562 cells, encompassing over 20 previously known L1 regulators (Extended Data Fig. 2d). Novel candidates are involved in functionally diverse pathways, including those of chromatin regulation, DNA damage and repair, and RNA processing (Extended Data Fig. 2e, f). Whereas many DNA damage or repair factors—particularly the Fanconi anaemia factors—suppress the activity of L1, genes implicated in the non-homologous end joining repair pathway promote L1 retrotransposition (Extended Data Fig. 2f). In agreement with this, mutations in some of the factors identified in this pathway were previously found to result in decreased retrotransposition frequencies[12]. Notably, many hits uncovered by our screen (for example, Fanconi anaemia factors, *MORC2* and *SETX*) are associated with human disorders[13–17].

To extend our survey of L1 regulators to another cell type, we performed both a genome-wide and a secondary screen in HeLa cells (Extended Data Fig. 1b, e) with the same sgRNA libraries used in the K562 screens. The top hits identified in the K562 genome-wide screen were also found in the HeLa screen (for example, *MORC2*, *TASOR* (also known as *FAM208A*), *SETX* and *MOV10*) (Extended Data Fig. 3a). Furthermore, secondary screens in both K562 and HeLa cells showed concordant effects for groups of genes; for example, the suppressive

[1]Department of Chemical and Systems Biology, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA. [2]Department of Genetics, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA. [3]Stanford University Chemistry, Engineering, and Medicine for Human Health (ChEM-H), Stanford School of Medicine, Stanford University, Stanford, California 94305, USA. [4]Institute of Stem Cell Biology and Regenerative Medicine, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA. [5]Department of Developmental Biology, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA. [6]Howard Hughes Medical Institute, Stanford School of Medicine, Stanford University, Stanford, California 94305, USA. †Present address: Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112-5550, USA.
*These authors contributed equally to this work.

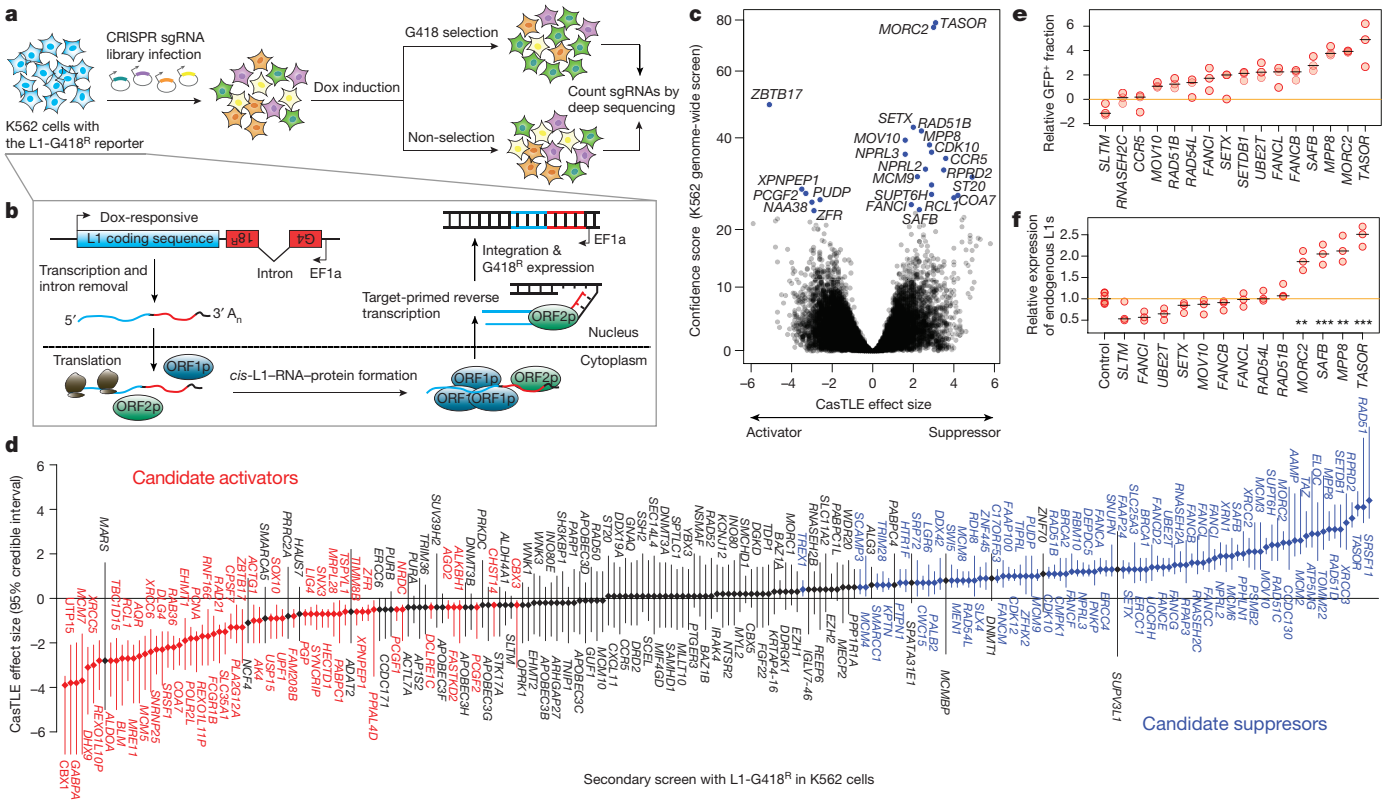


Figure 1 | Genome-wide screen for L1 activators and suppressors in K562 cells. a, Schematic of the screening process. **b**, Schematic of the L1-G418^R retrotransposition. **c**, CasTLE analysis of two independent K562 genome-wide screens. Genes at 10% FDR cutoff coloured in blue, analysed by CasTLE likelihood ratio test[11]. **d**, The maximum effect size (centre value) estimated by CasTLE from two independent K562 secondary screens with ten independent sgRNAs per gene. Bars, 95% credible interval. L1 activators are shown in red; L1 suppressors are shown in blue; and insignificant genes for which the credible interval includes zero are shown in grey. **e**, L1–GFP retrotransposition in control (infected with negative control sgRNAs, hereafter referred to as 'control') and mutant K562 cells as indicated. GFP$^+$ cell fractions are normalized to the control, the centre value is the median. $n = 3$ biological replicates per gene. **f**, RT–qPCR measuring the expression of endogenous L1Hs in mutant K562 cells, normalized to control. The centre value is the median. $n = 3$ technical replicates per gene. $^{**}P < 0.01$; $^{***}P < 0.001$; two-sided Welch t-test.

effects of the Fanconi anaemia complex genes, and the activating effects of the non-homologous end joining pathway genes (Extended Data Fig. 3b–e). Of note, however, is that a subset of genes showed cell-line-selective effects (Extended Data Fig. 3c). At the same time, some of the previously known L1 regulators did not appear as hits in our screen. Several factors could limit our ability to identify all the genes controlling L1 retrotransposition to saturation. For example, a subset of regulators may function in a cell-type specific manner that is not captured by either K562 or HeLa screens; essential genes with strong negative effects on cell growth may have dropped out; or regulators that strictly require native L1 UTR sequences may have been missed owing to our reporter design. However, our combined screens identify many novel candidates for the control of L1 retrotransposition in human cells and provide a rich resource for mechanistic studies of transposable elements.

Select screen hits were further validated in K562 cells using a well-characterized L1–GFP reporter[18] (Extended Data Fig. 1a), confirming 13 suppressors and 1 activator (*SLTM*) out of 16 examined genes (Fig. 1e). Notably, chromatin regulators (TASOR, MORC2, MPP8, SAFB and SETDB1) suppress the retrotransposition of the L1–GFP reporter, but not that of a previously described codon-optimized L1–GFP reporter (hereafter referred to as (opt)-L1–GFP)[19,20], indicating that these factors regulate L1 retrotransposition in a manner dependent upon the open reading frame (ORF) nucleotide sequence of the native L1 (Extended Data Fig. 3f, g). An additional secondary screen against the codon-optimized (opt)-L1-G418^R reporter in K562 cells confirmed the sequence-dependent nature of these L1 regulators, and systematically partitioned our top screen hits into native L1 sequence-dependent and -independent candidates (Extended Data Fig. 3h; see Supplementary Table 2 for the full list).

We next examined whether the identified regulators influence the expression of endogenous L1Hs, the youngest and only retrotransposition-competent L1 subfamily in humans. CRISPR-deletion of some genes (*TASOR*, *MPP8* (also known as *MPHOSPH8*), *SAFB* and *MORC2*) significantly increased the expression of endogenous L1Hs, whereas deletion of other genes—such as *SETX*, *RAD51* or Fanconi anaemia complex components—had little effect (Fig. 1f). Because all of the genes we studied restrict L1–GFP retrotransposition into the genome (Fig. 1e and Extended Data Fig. 4a), our results suggest that the identified suppressors can function at either the transcriptional or the post-transcriptional level.

We further investigated three candidate transcriptional regulators of L1: MORC2, TASOR and MPP8. TASOR and MPP8 (along with PPHLN1) comprise the HUSH complex, and recruit the H3K9me3 methyltransferase SETDB1 to repress genes[8]. Notably, PPHLN1 and SETDB1 were also identified as L1 suppressors in our screen (Fig. 1d and Extended Data Fig. 3b). MORC2, which has recently been shown to biochemically and functionally interact with HUSH[21], is a member of the microrchidia (MORC) protein family that has been implicated in transposon silencing in plants and mice[22,23]. Although MORC2 and HUSH have been previously implicated in heterochromatin formation, most heterochromatin factors had no effect on L1 retrotransposition, suggesting a selective effect (Fig. 2a and Extended Data Fig. 4b).

Several independent experiments in clonal knockout K562 lines confirmed that HUSH and MORC2 suppress the retrotransposition of the L1–GFP reporter by silencing its transcription (Fig. 2b, c and

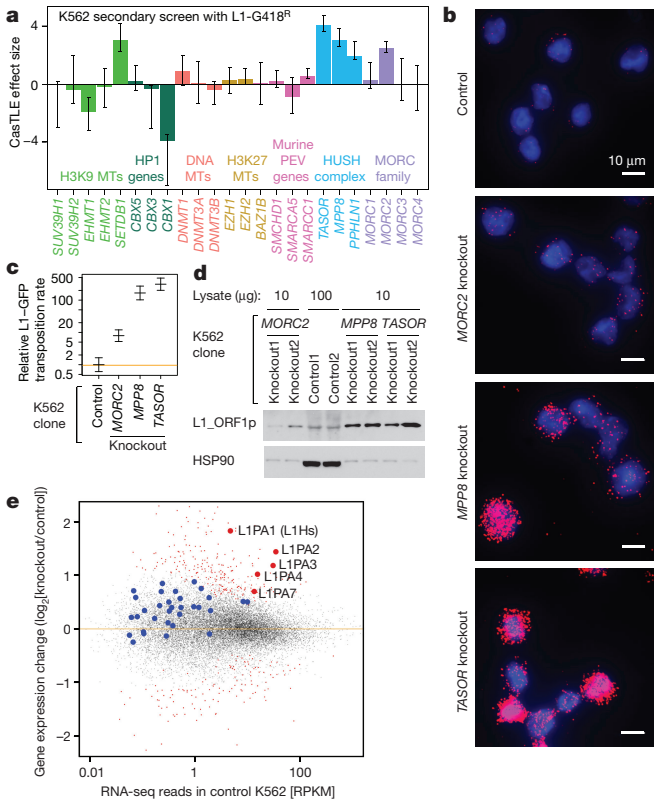


Figure 2 | HUSH and MORC2 silence L1 transcription to inhibit retrotransposition. a, The maximum effect size (centre value) of the indicated heterochromatin regulators, estimated by CasTLE from two independent K562 secondary screens with ten independent sgRNAs per gene. Error bars, 95% credible intervals. MT, methyltransferase. **b**, Micrographs of L1–GFP mRNAs in dox-induced K562 clones, from single single-molecule fluorescent *in situ* hybridization (smFISH) experiments that were independently repeated twice with similar results. See also Extended Data Fig. 4d, e. Blue, stained with 4′,6-diamidino-2-phenylindole (DAPI); pink, L1–GFP RNA. **c**, The L1–GFP retrotransposition rate[18] (centre value) in K562 clones, from logistic regression fit of the GFP$^+$ cell counts at 7 time points (0, 5, 10, 15, 20, 25 and 30 days post-induction) and two independent clones per gene. Over 200 GFP$^+$ cells per cell count; data normalized to control. Bars, 95% credible interval. **d**, Endogenous L1_ORF1p levels in K562 clones shown by western blotting with HSP90 as a loading control, independently repeated three times with similar results. **e**, RNA-seq read counts from *MORC2* knockout, *MPP8* knockout and *TASOR* knockout K562 clones, compared to control RNA-seq reads. $n = 6 + 2$ biologically independent RNA-seq experiments. Dots represent transcripts; large dots represent L1 transcripts. Red indicates significant changes (padj < 0.1, DESeq analysis); blue and grey indicate insignificant changes.

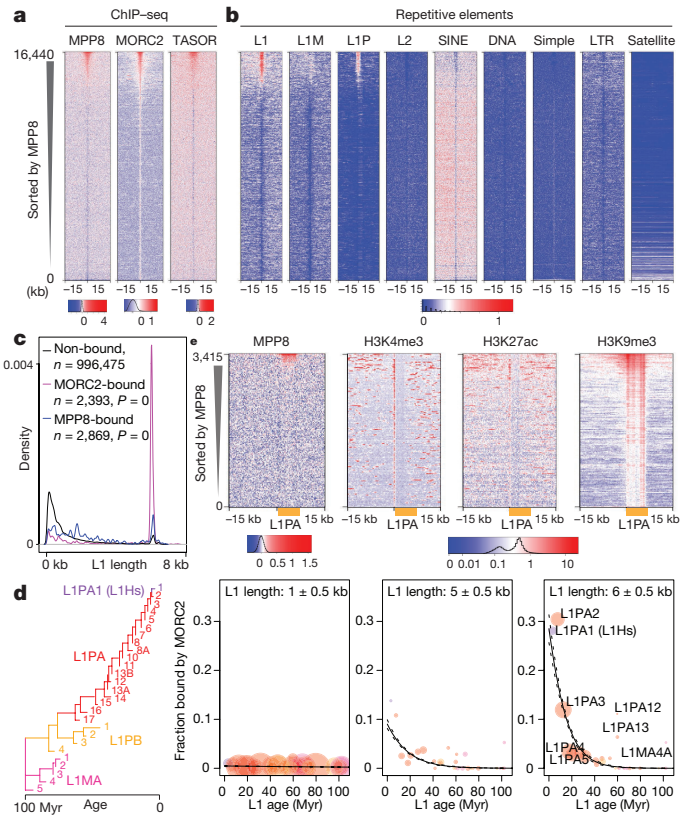


Figure 3 | HUSH and MORC2 target young full-length L1s in euchromatic environments. a, Heat maps showing signal enrichment of ChIPs with indicated antibodies in K562 cells, sorted by MPP8 ChIP signal and centred on MPP8 and MORC2 peaks. Plotted is the normalized ChIP signal (control subtracted with corresponding knockout). **b**, Heat maps showing MPP8 and MORC2 ChIP signal enrichment over repetitive elements, centred and sorted as in **a**. **c**, Size distribution of the L1s bound or unbound by MORC2 or MPP8 in K562 cells. *P* values, two-tailed Kolmogorov–Smirnov test. **d**, Fraction of MORC2-bound L1s (centre values) as a function of L1 length (three size classes are presented) and age (predicted from the phylogenetic analysis[27]) in K562 cells. Coloured circles represent L1 families, with areas proportional to the number of L1 instances with the indicated age and length. $n = 1{,}501$ MORC2-bound L1 + 200,160 unbound L1. $P = 2.2 \times 10^{-90}$ for the age–length interaction term, lower for simple terms (analysis of variance, χ^2 test), plotted are the logistic regression lines with 95% credible interval. Myr, million years. **e**, Heat maps showing signal enrichment of ChIPs with the indicated antibodies in K562 cells, centred on the 5′ end of full-length L1PAs.

Extended Data Fig. 4c–f). Additionally, HUSH and MORC2 repressed endogenous (non-reporter) L1Hs RNA and protein expression in both K562 and human embryonic stem cells[24] (hES cells, H9) (Fig. 2d and Extended Data Fig. 4g–k). PolyA-selected RNA sequencing (RNA-seq) experiments revealed the upregulated expression of evolutionarily younger L1PA families (including L1Hs) upon HUSH or MORC2 knockout in K562 cells (Fig. 2e). Taken together, these data demonstrate that HUSH and MORC2 silence both the reporter transgene as well as endogenous, evolutionarily young L1s.

Chromatin immunoprecipitation followed by sequencing (ChIP–seq) from K562 cells and hES cells demonstrated that MORC2, MPP8 and TASOR co-bind genomic regions characterized by specific L1 instances. Elements from the primate-specific L1P family showed greater enrichment than elements from the older L1M family (Fig. 3a, b and Extended Data Figs 5a, b, 6a, b), consistent with the preferential derepression of the former upon HUSH or MORC2 knockout (Fig. 2e).

Moreover, this enrichment was specific to L1s, as other major repeat classes were not enriched (Fig. 3b and Extended Data Fig. 6b), although all three proteins also targeted the expressed KRAB zinc-finger (KRAB-ZNF) genes (Extended Data Fig. 5c, d). HUSH knockout in K562 cells almost completely abrogated MORC2 binding at L1s (consistent with recently published observations that HUSH recruits MORC2 for transcriptional repression[21]), whereas MORC2 deletion led to a modest but appreciable decrease of HUSH subunit binding (Extended Data Fig. 7). In mouse ES cells, MPP8 was found to bind retrotransposition-competent L1Md-A and L1Md-T, as well as IAP elements, which are a class of murine endogenous retroviruses that remain currently mobile in the mouse genome (Extended Data Fig. 6c, d). This suggests that regulators uncovered by our study in human cells may, in other species, target additional active transposons beyond L1s.

Notably, only a subset of L1Ps is bound by HUSH or MORC2 in either K562 cells or hESCs; this is seen even for the younger human L1Ps. We sought to identify genomic or epigenomic features that could explain this selectivity. We found that HUSH and MORC2 selectively

target young, full-length L1s, particularly the L1PA1-5 in human cells (Fig. 3c, d) and the L1Md-A/T in mice (Extended Data Fig. 6e). Both MPP8 and MORC2 bind broadly across the L1; whereas MORC2 binding is skewed towards the 5′ end, MPP8 shows higher enrichments within the body and at the 3′ end of L1PAs, including the L1Hs (L1PA1) elements (Extended Data Fig. 6f, g).

However, the preference for full-length, evolutionarily younger L1PAs can only partially explain the observed selectivity of HUSH and MORC2, as only a subset of such elements are targeted by the complex (Fig. 3d). We found that the additional layer of selectivity can be explained by the state of the surrounding chromatin, with HUSH- or MORC2-occupied L1s preferentially located within the transcriptionally permissive euchromatic environment marked by modifications such as H3K4me3 (histone H3 Lys4 trimethylation) and H3K27ac (histone H3 Lys27 acetylation) (Fig. 3e). In agreement with this, HUSH- or MORC2-bound L1s are enriched within introns of actively transcribed genes (Extended Data Fig. 8a, b). Furthermore, although most HUSH- or MORC2-bound L1s are concordant between K562 and hES cells, those that are bound in a cell-type-specific manner tend to be associated with genes that are differentially active between the two cell types (Extended Data Fig. 8c). To understand the role of transcription in the HUSH and MORC2 targeting of L1s, we investigated MORC2 and MPP8 occupancy at the inducible L1 transgene. We observed increased binding of these factors upon transcriptional induction (Extended Data Fig. 8d), suggesting that transcription through L1 sequences facilitates HUSH and MORC2 binding. Together, HUSH and MORC2 selectively target young, full-length L1s located within transcriptionally permissive euchromatic regions. These L1s are precisely the elements that pose the highest threat to genome integrity, as a subset of them remains mobile and transcription is the first step of L1 mobilization.

Despite their location within the euchromatic environment, HUSH- and MORC2-bound L1s themselves are heavily modified with the transcriptionally repressive H3K9me3 (Fig. 3e), consistent with the role of HUSH in facilitating H3K9me3 deposition at target sites[8]. HUSH or MORC2 knockout decreased H3K9me3 levels preferentially at L1 rather than at non-L1 HUSH or MORC2 genomic targets, and at bound versus unbound L1s (Fig. 4a and Extended Data Fig. 9a, b). Because HUSH- and MORC2-bound L1s are significantly enriched within introns of transcriptionally active genes (Extended Data Fig. 8a–c), we examined whether HUSH or MORC2 recruitment and its associated H3K9me3 deposition can influence chromatin modification and expression of the host genes. Despite a transcriptionally active status (Extended Data Fig. 8a, b), promoters and especially bodies of genes harbouring MORC2- or HUSH-bound L1s show appreciable levels of H3K9me3. This enrichment is substantially diminished in the knockout lines (Extended Data Fig. 9c), with concomitant upregulation of genes that harbour MORC2- or HUSH-bound L1s, but not of those with unbound intronic L1s (Fig. 4b). Thus, HUSH or MORC2 binding at intronic L1s leads to a modest but significant downregulation of the active genes that harbour them (Fig. 4c and Extended Data Figs 9d–g, 10a).

Inserting L1 sequences in a transcript leads to a decrease in RNA expression due to inadequate transcript elongation[25], and this effect has been attributed to the A/T enrichment of L1s. However, our results argue that the transcriptional attenuation of host gene expression could be a consequence of epigenetic silencing by HUSH or MORC2 (Fig. 4b, c and Extended Data Figs 9d–g, 10a), and this possibility is consistent with the described role of genic H3K9me3 in decreasing the Pol II elongation rate, leading to its accumulation over the H3K9me3 region[26]. If such a mechanism is involved, then HUSH knockout should decrease accumulation of the elongating Pol II over L1 bodies, and this is indeed what we observe in Pol II ChIP–seq experiments (although notably, at the 5′ UTRs of L1s, Pol II levels are relatively increased in the knockouts) (Extended Data Fig. 10b).

Host gene regulation is directly dependent on the presence of the intronic L1, as deletion of select MORC2- or HUSH-bound L1s from the intron leads to the upregulation of host mRNA to a level commensurate with the magnitude of changes caused by HUSH or MORC2 knockout (Fig. 4d, e and Extended Data Fig. 10c, d). Thus, attenuated expression level of an active gene can be a by-product of a retrotransposition event and associated HUSH- and MORC2-mediated L1 silencing (Fig. 4f). Although the observed effects on active host genes are only modulatory, they occur to various extents at hundreds of human genes, illustrating how transposable element activity can rewire host gene expression patterns.

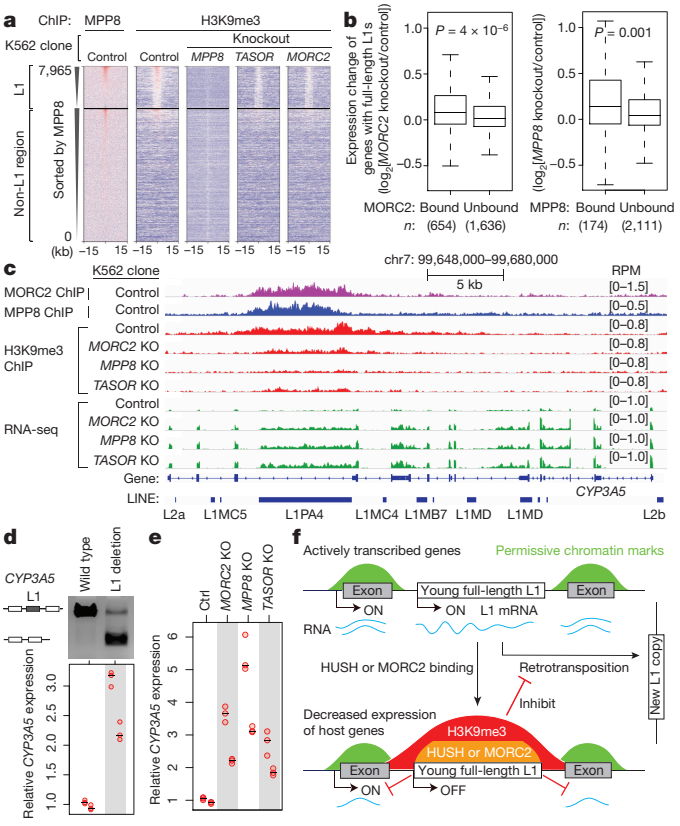


Figure 4 | HUSH or MORC2 binding at L1s decreases active host gene expression. a, Heat maps showing MPP8 and H3K9me3 ChIP signal enrichment, centred on MPP8 and MORC2 summits and separated by the presence or absence of L1. **b**, Gene expression changes with intronic full-length L1s that are bound or unbound by MORC2 or MPP8 (RNA-seq reads from knockout K562 clones compared to control). Box plots show the median and interquartile range (IQR), whiskers are 1.5 × IQR. *P* value, two-sided Mann–Whitney–Wilcoxon test. **c**, Genome browser tracks. Loss of HUSH or MORC2 causes a decrease of H3K9me3 at the target L1 and an increase in expression of both the target L1 and its host gene, independently repeated once with similar results. KO, knockout. **d**, Deleting the target intronic L1 from *CYP3A5* in K562 increases *CYP3A5* expression, as measured by RT–qPCR normalized to a wild-type sample. $n = 2$ biological replicates × 3 technical replicates; centre value is median. The gel confirms L1 deletion; two experiments repeated independently with similar results. **e**, RT–qPCR for *CYP3A5* expression in K562 clones, normalized to the control. $n = 2$ biological replicates × 3 technical replicates; centre value is median. KO, knockout. **f**, HUSH and MORC2 bind young full-length L1s within transcriptionally active genes, and promote H3K9me3 deposition at target L1s to silence L1 transcription. This pathway not only inhibits L1 retrotransposition, but also decreases host gene expression.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 May; accepted 28 November 2017.
Published online 6 December 2017.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

2. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12,** 615–627 (2011).
3. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* **12,** 187–215 (2011).
4. Mita, P. & Boeke, J. D. How retrotransposons shape genome regulation. *Curr. Opin. Genet. Dev.* **37,** 90–100 (2016).
5. Goodier, J. L. Restricting retrotransposons: a review. *Mob. DNA* **7,** 16 (2016).
6. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18,** 71–86 (2017).
7. Philippe, C. *et al.* Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *eLife* **5,** e13926 (2016).
8. Tchasovnikarova, I. A. *et al.* Epigenetic silencing by the HUSH complex mediates position–effect variegation in human cells. *Science* **348,** 1481–1485 (2015).
9. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87,** 917–927 (1996).
10. Morgens, D. W. *et al.* Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8,** 15178 (2017).
11. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34,** 634–636 (2016).
12. Suzuki, J. *et al.* Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.* **5,** e1000461 (2009).
13. Chance, P. F. *et al.* Linkage of the gene for an autosomal dominant form of juvenile amyotrophic lateral sclerosis to chromosome 9q34. *Am. J. Hum. Genet.* **62,** 633–640 (1998).
14. Németh, A. H. *et al.* Autosomal recessive cerebellar ataxia with oculomotor apraxia (ataxia-telangiectasia-like syndrome) is linked to chromosome 9q34. *Am. J. Hum. Genet.* **67,** 1320–1326 (2000).
15. Albulym, O. M. *et al. MORC2* mutations cause axonal Charcot–Marie–Tooth disease with pyramidal signs. *Ann. Neurol.* **79,** 419–427 (2016).
16. Schottmann, G., Wagner, C., Seifert, F., Stenzel, W. & Schuelke, M. *MORC2* mutation causes severe spinal muscular atrophy-phenotype, cerebellar atrophy, and diaphragmatic paralysis. *Brain* **139,** 1–4 (2016).
17. Brégnard, C. *et al.* Upregulated LINE-1 activity in the Fanconi anemia cancer susceptibility syndrome leads to spontaneous pro-inflammatory cytokine production. *EBioMedicine* **8,** 184–194 (2016).
18. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28,** 1418–1423 (2000).
19. Han, J. S. & Boeke, J. D. A highly active synthetic mammalian retrotransposon. *Nature* **429,** 314–318 (2004).
20. Wagstaff, B. J., Barnerssoi, M. & Roy-Engel, A. M. Evolutionary conservation of the functional modularity of primate and murine LINE-1 elements. *PLoS ONE* **6,** e19672 (2011).
21. Tchasovnikarova, I. A. *et al.* Hyperactivation of HUSH complex function by Charcot–Marie–Tooth disease mutation in *MORC2*. *Nat. Genet.* **49,** 1035–1044 (2017).
22. Moissiard, G. *et al.* MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336,** 1448–1451 (2012).
23. Pastor, W. A. *et al.* MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* **5,** 5795 (2014).
24. Garcia-Perez, J. L. *et al.* LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16,** 1569–1577 (2007).
25. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429,** 268–274 (2004).
26. Saint-André, V., Batsché, E., Rachez, C. & Muchardt, C. Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* **18,** 337–344 (2011).
27. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16,** 78–87 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Moran for the LRE-GFP plasmid and A. Engel for the codon-optimized L1 construct; D. Fuentes, A. Spencley, R. Srinivasan, J. Mohammed, V. Bajpai, K. Tsui, G. Hess, D. Morgens and G. Cornelis for assistance and discussions; K. Cimprich, A. Fire and A. Urban for comments on the manuscript; and L. Bruhn, S. Altschuler, B. Borgo, P. Sheffield and C. Carstens (Agilent) for discussions and oligonucleotide synthesis. This work was funded by grants from the Jane Coffin Childs Memorial Fund for Medical Research (N.L.), National Science Foundation DGE-114747 (C.H.L.), National Institutes of Health (NIH) R01HG008150, 1UM1HG009436-01 and NIH 1DP2HD084069-01 (M.C.B.), NIH R01 GM112720, Stinehart Reed Award and Howard Hughes Medical Institute (J.W.).

Author Contributions N.L., C. H.L., T.S., J.W. and M.C.B. designed and performed experiments, analysed data and wrote the manuscript. E.G., C.H.L., J.W. and M.C.B. initiated the K562 genome-wide screen. B.G. analysed smFISH data. J.W. and M.C.B. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.W. (wysocka@stanford.edu) or M.C.B. (bassik@stanford.edu).

Reviewer Information *Nature* thanks D. Bourc'his and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture and antibodies. K562 cells (American Type Culture Collection, ATCC) were grown in Roswell Park Memorial Institute (RPMI) 1640 Medium (11875093, Life Technologies) supplemented with 10% fetal bovine serum (Fisher, SH30910), 2 mM L-glutamine (Fisher, SH3003401) and 1% penicillin–streptomycin (Fisher, SV30010), and cultured at 37 °C with 5% CO_2. HeLa cells (ATCC) were grown in Dulbecco's modified Eagle's medium (Life Technologies, 11995073) supplemented with 10% FBS, 2 mM L-glutamine, and 1% penicillin–streptomycin, and cultured at 37 °C with 5% CO_2. H9 human ES cells (WiCell) were expanded in feeder-free, serum-free medium mTeSR-1 from StemCell technologies, passaged 1:6 every 5–6 days using accutase (Invitrogen) and replated on tissue culture dishes coated overnight with growth-factor-reduced Matrigel (BD Biosciences). Male mouse embryonic stem cells (ES-E14TG2a, ATCC) were grown as described[28]. Cell lines were authenticated by the vendor. Cell cultures were routinely tested and found to be negative for mycoplasma infection (MycoAlert, Lonza).

Rabbit MORC2 antibody (A300-149A, Bethyl Laboratories), Rabbit MPP8 antibody (16796-1-AP, Protein Technologies Inc.), Rabbit TASOR antibody (HPA006735, Atlas Antibodies) were used in western blots (1:1,000 dilution) and ChIP assays. Mouse anti-LINE-1 ORF1p antibody (MABC1152, Millipore)[29], Rabbit HSP90 (C45G5, Cell Signalling, 4877) and β-actin antibody (ab49900, Abcam) were used in western blots. Histone H3 (tri-methyl K9) antibody (ab8898, Abcam) and RNA Pol II (Santa Cruz Biotechnology, N-20 sc-899) were used in ChIP assays.

L1 reporters. The L1–ORF1–ORF2 sequence is derived from the LRE–GFP[30], a gift from J. Moran. To make the L1–GFP reporter, we used Gibson assembly to clone the L1_ORF1/2 fragment and a GFP–β-globin-intron cassette driven by the mammalian promoter EF1a into the pB transgene using a dox-inducible promoter (modified from PBQM812A-1, System Biosciences) to drive the L1 sequence and a UBC–RTTA3–ires Blast as a selectable marker for reporter integration. To make the L1-G418^R reporter, we replaced the GFP–β-globin-intron fragment in the L1–GFP reporter with a NEO-intron–NEO cassette driven by the mammalian promoter EF1a. The codon-optimized L1–ORF1–ORF2 sequence in our (opt)-L1 reporter is derived from the SynL1_optORF1_neo, a gift from A. Engel[31]. We replaced the self-splicing Tetrahymena NEO-intron–NEO cassette with the neo–β-globin-intron–neo cassette driven by the EF1a promoter or the GFP–β-globin-intron–GFP cassette driven by the EF1a promoter. This L1–syn-ORF1–ORF2–indicator cassette was inserted into the pB transgene using a dox-inducible promoter and a UBC–RTTA3–ires Blast, as described previously.

Genome-wide screen in K562 cells. The K562 cell line (with a BFP–Cas9 lentiviral transgene) was nucleofected with the pB–tetO–L1-G418^R/Blast construct and the piggyBac transposase (PB210PA-1, System Biosciences) following the manufacturer's instructions (Lonza 2b nucleofector, T-016 program). The nucleofected cells were sorted using limiting dilution in 96-well plates, and positive clones were screened first for sensitivity to Blast, and then the ability to generate G418-resistant cells after dox induction. The Cas9/L1-G418^R cells were lentivirally infected with a genome-wide sgRNA library as described previously[10], containing approximately 200,000 sgRNAs targeting 20,549 protein-coding genes and 13,500 negative-control sgRNAs at a multiplicity of infection of 0.3–0.4 (as measured by the mCherry fluorescence from the lentiviral vector), and selected for lentiviral integration using puromycin (1 μg ml^{-1}) for 3 days as the cultures were expanded for the screens. In duplicate, 200×10^6 library-infected cells were dox-induced (1 μg ml^{-1}) for 10 consecutive days, with a logarithmic growth (500,000 cells per ml) maintained each day of the dox induction. After dox induction, the cells were recovered in normal RPMI complete medium for 24 h, and then split into the G418-selection condition (300 μg ml^{-1} G418, Life Technologies, 11811031) and non-selection conditions. After 7 days of maintaining cells at 500,000 per ml, 200 million cells under each condition were recovered in normal RPMI medium for 24 h, before they were pelleted by centrifugation for genomic DNA extraction using Qiagen DNA Blood Maxi kit (51194) as described[32]. The sgRNA-encoding constructs were PCR-amplified using Agilent Herculase II Fusion DNA Polymerase (600675) (see Supplementary Table 4 for the primer sequences used). These libraries were then sequenced across two Illumina NextSeq flow cells (approximately 40 million reads per condition; around 200× coverage per library element). Computational analysis of genome-wide screen was performed as previously described[10,11] using CasTLE, which is a maximum likelihood estimator that uses a background of negative control sgRNAs as a null model to estimate gene effect sizes. See Supplementary Table 1 for the K562 genome-wide screen results.

Secondary screen in K562 cells. The secondary screen library included the following, non-comprehensive sets of genes (253 genes in total, around 10 sgRNAs per gene, plus 2,500 negative control sgRNAs): all genes falling within approximately 30% FDR from the K562 genome-wide screen (around 150 genes); genes known to be functionally related to the 30% FDR genes, genes previously implicated in L1 biology; and genes involved in epigenetic regulation or position effect variegation (see Supplementary Table 2 for a complete list). The library oligonucleotides were synthesized by Agilent Technologies and cloned into pMCB320 using BstXI/BlpI overhangs after PCR amplification. The Cas9/L1-G418^R (or Cas9/(opt)-L1-G418^R) K562 cell line was lentivirally infected with the secondary library (around 4,500 elements) at an multiplicity of infection of 0.3–0.4 as described previously[33]. After puromycin selection (1 μg ml^{-1} for 3 days) and expansion, 40 million (approximately 9,000 coverage per library element) cells were dox-induced for 10 days in replicate, recovered for 1 day, and split for 7-day G418-selection and non-selection conditions, with a logarithmic growth (500,000 cells per ml) maintained as in the K562 genome-wide screen. 10 million cells under each condition were used for genomic extractions, sequenced (approximately 6–10 million reads per condition; around 1,000–2,000× coverage per library element) and analysed using CasTLE as described above[10,11]. Supplementary Table 2 lists the K562 secondary screen results with L1-G418^R and (opt)-L1-G418^R.

Genome-wide screen and secondary screen in HeLa cells. The pB–tetO–L1-G418^R/Blast construct was integrated into Cas9 expressing HeLa cells with piggyBac transposase via nucleofection (Lonza 2b nucleofector, I-013 program) following the manufacturer's instructions. The Cas9/L1-G418^R HeLa cells were blasticidin (10 μg ml^{-1})-selected, screened for sensitivity to G418 and the ability to generate G418 resistance cells after dox induction, and lentivirally infected with the genome-wide sgRNA library or with the secondary sgRNA library. Infected cells were then puromycin selected (1 μg ml^{-1}) for 5 days and expanded for the screens.

For the genome-wide screen, approximately 200×10^6 Cas9/L1-G418^R HeLa cells (around 1,000× coverage of the sgRNA library) were dox-induced for 10 days in replicate, recovered for 1 day, and split for 8-day G418-selection and non-selection conditions, with cells being split every other day to maintain the sgRNA library at a minimum of around 350× coverage. Approximately 200 million (1,000× coverage) cells per condition were used for genomic extractions and sequencing as described above for the K562 screens. Supplementary Table 1 presents the HeLa genome-wide screen results.

For the secondary screen, approximately 1×10^7 Cas9/L1-G418^R HeLa cells (around 2,000× coverage of the sgRNA library) were dox-induced for 10 days in replicate, recovered for 1 day, and split for 8-day G418-selection and non-selection conditions, with cells being split every other day to maintain around 400× coverage. Approximately 5 million (1,000× coverage) cells per condition were used for genomic extractions and sequencing as described above. Supplementary Table 2 presents the HeLa secondary screen results.

Validation of individual candidates using the L1–GFP retrotransposition assay. To validate the genome-wide screen hits, we infected clonal Cas9/L1–GFP K562 cells with individual sgRNAs as previously described[32], three independent mutant cell lines per gene, each with a different sgRNA (cloned into pMCB320 using BstXI/BlpI overhangs; mU6:sgRNA; EF1a:puromycin–t2a–mCherry). Supplementary Table 3 lists the sgRNA sequences used. The infected cells were selected against puromycin (1 μg ml^{-1}) for 3 days, recovered in fresh RPMI medium for 1 day, and dox-induced for 10 days. Next, the percentage of GFP$^+$ cells was measured on a BD Accuri C6 Flow Cytometer (GFP fluorescence detected in FL1 using 488-nm laser) after gating for live mCherry$^+$ cells.

CRISPR-mediated deletion of individual genes and intronic L1s. To delete genes in H9 ES cells, we cloned target sgRNAs in pSpCas9(BB)-2A–GFP (PX458) as described[34]. The sgRNA plasmids were prepared with the Nucleospin plasmid kit (Macherey Nagel) and transfected into H9 ES cells using Fugene following the manufacturer's instructions. After 48–72 h, GFP-positive transfected cells were sorted and expanded. Gene depletion effects were validated by western blots.

To delete the L1 from the host gene intron, we designed sgRNAs targeting both upstream and downstream sides of the L1 within the intron; one was cloned into pSpCas9(BB)-2A–BFP, whereas the other was cloned into pSpCas9(BB)-2A–GFP. The two sgRNA plasmids were mixed at a 1:1 ratio and nucleofected into K562 cells via electroporation following the manufacturer's instructions. After 48–72 h, BFP or GFP-positive transfected cells were single-cell sorted and expanded. The genetic deletion effects were validated by PCR assay.

Western blotting. Live cells were lysed for 30 min at 4 °C in protein extraction buffer (300 mM NaCl, 100 mM Tris pH 8, 0.2 mM EDTA, 0.1% NP40, 10% glycerol) with protease inhibitors, and centrifuged to collect the supernatant lysate. The cell lysate was measured with Bradford reagent (Biorad), separated on SDS–PAGE gels and transferred to nitrocellulose membranes. The L1-reporter containing K562 cells had not been dox-induced when used for western blot assays characterizing endogenous L1_ORF1p levels (Fig. 2d and Extended Data Fig. 4k).

PCR and gel electrophoresis. PCR experiments characterizing the L1-G418^R retrotransposition and the deletion of intronic L1s were performed with Phusion High-Fidelity DNA Polymerase (M0530S, New England Biolabs), following the

manufacturer's instructions. In general, 30 cycles of PCR reactions were performed at an annealing temperature 5 °C below the melting temperature of the primer. No 'spliced' PCR products were detected without dox induction, even with 40 PCR cycles. PCR reaction products were separated on 1% agarose gels with ethidium bromide. Primer sequences are listed in Supplementary Table 4.

qRT–PCR and PspGI-assisted qPCR. Total RNA was isolated from live cells using the RNeasy kit (74104, Qiagen) and treated with RNase-free DNase Set (79254, Qiagen) to remove genomic DNA, according to the manufacturer's instructions. Total RNA (500 ng) was reverse-transcribed with SuperScriptA III First-Strand Synthesis System (18080051, Life Technologies) following the manufacturer's instructions. β-actin mRNA was used as an internal control within each RNA sample (Figs 1f and 4d, e). The sequences of PCR primers, including the one targeting the 5′ UTR of L1Hs[35–37], are summarized in Supplementary Table 4.

Genomic DNA was isolated using PureLink Genomic DNA Mini Kit (K182001, Life Technologies) with RNase A digestion to remove contaminant RNA, according to the manufacturer's instructions. Genomic DNA (300 ng per sample) was digested with 50 units PspGI (R0611S, New England Biolabs) in 1× Smart buffer (New England Biolabs) at 75 °C for 1 h, to cut uniquely at the intron of the GFP cassette. The reaction mixture was then used in qPCR experiments with primers flanking the intron in the GFP cassette (Supplementary Table 4). Because of the PspGI digestion, the original unspliced L1–GFP reporter will not be amplified by PCR. Only newly integrated GFP cassettes, in which the intron was removed during the retrotransposition process, can be PCR-amplified. qPCR runs and analysis were performed on the Light Cycler 480II machine (Roche).

Northern blotting. Northern blotting was conducted as previously described[38]. In brief, 15 μg of total RNA from K562 cells or H9 ES cells was separated on the 0.7% formaldehyde agarose gel, capillary transferred overnight in 20× saline sodium citrate to the Hybond N membrane (GE Healthcare), crosslinked with a Stratalinker (Stratagene), and hybridized with ^{32}P-labelled single-stranded DNA probes (10^6 counts per min per ml) in ULTRAhyb-Oligo Hybridization Buffer (AM8663, Life Technologies) following the manufacturer's instructions. Blots were washed twice with wash buffer (2× saline sodium citrate, 0.5% SDS), and then exposed to film overnight for several days at −80 °C with an intensifying screen. The sequence of oligonucleotide probes is listed in Supplementary Table 3.

Single-molecule fluorescent *in situ* hybridization. SmFISH assays were performed following the affymetrix Quantigene ViewRNA ISH Cell Assay user manual. Between 2.5 and 5 million live K562 cells were fixed within 4% formaldehyde in 1× PBS for 60 min at room temperature, resuspended in 1× PBS, pipetted onto poly-L-lysine coated glass cover slips (around 20,000 total cells per spot; spread out with a pipette tip), and kept in a dry oven at 50 ± 1 °C for 30 min to fix the cells onto the glass slip, followed by digestion with Protease QS (1:4,000) in 1× PBS for 10 min at room temperature. Cells were hybridized with smFISH probes, designed to target β-actin mRNA (fluorescein isothiocyanate channel) and the L1–GFP reporter mRNA (Cy3 channel), stained for 5 min with 4′,6-diamidino-2-phenylindole (DAPI), and mounted with Prolong Gold Antifade Reagent (10 ml per sample). Images were taken using a spinning disk confocal microscope equipped with 60× 1.27-numerical aperture water immersion objective with an effective pixel size of 108 × 108 nm. Specifically, for each field of view, a *z*-series of 8 μm is taken with 0.5 μm per *z*-step for all three channels. For quantification, maximum-projected images from the *z*-series is used and analysed by a custom-written MATLAB script. In brief, all images are first subtracted with the background determined with the Otsu method[39] from the log-transformed image after pillbox blurring with a radius of 3 pixels. mRNA puncta are segmented by top-hat filter using the background subtracted images and only the ones above the 25th percentile intensity of all segmented puncta are taken for downstream analysis. Each punctum is then assigned to the nuclear mask identified by image areas above the previously determined background. For each single cell, the assigned pixel area of L1–GFP mRNA is then normalized to the assigned pixel area of β-actin mRNA per cell.

RNA-seq. Two independent biological replicates of K562 cells in culture were extracted to isolate a DNA-free total RNA sample, using the RNeasy kit (74104, Qiagen) combined with the RNase-free DNase Set (79254, Qiagen). PolyA-selected RNA were isolated using 'Dynabeads mRNA Purification Kit for mRNA Purification from Total RNA preps' (610-06, Life Technologies) following the manuals. PolyA-selected RNA (100 ng) was fragmented with NEBNext Magnesium RNA Fragmentation Module (E6150S, New England Biolabs), and used for first strand cDNA synthesis with SuperScriptII (18064-014, Invitrogen) and random hexamers, followed by second strand cDNA synthesis with RNaseH (18021-014, Invitrogen) and DNA Pol I (18010-025, Invitrogen). The cDNA was purified, quantified, multiplexed and sequenced with 2× 75-bp pair-end reads on an Illumina NEXT-seq (Stanford Functional Genomics Facility).

RNA-seq reads were aligned to the hg38 reference genome with hisat2 (–no-mixed,–no-discordant) without constraining to known transcriptome. Known (gencode 25) and *de novo* transcript coverages were quantified with featureCount. Repeat Masker coverage was quantified with bedtools coverage. Reads mapping to the same repeat family were then tabulated together, because individual read coverage was too low to obtain meaningful results. Differential expression analysis of join gene-repeat data was performed with DESeq2[40].

ChIP–seq. Two replicates of ChIP experiments per sample were performed as previously described[41,42]. Approximately 0.5–1 × 10^7 cells in culture per sample were crosslinked with 1% paraformaldehyde for 10 min at room temperature, and quenched by 0.125 M glycine for 10 min at room temperature. Chromatin was sonicated to an average size of 0.2–0.7 kb using a Covaris (E220 evolution). Sonicated chromatin was incubated with 5–10 μg antibody bound to 100 μl protein G Dynabeads (Invitrogen) and incubated overnight at 4 °C, with 5% kept as input DNA. Chromatin was eluted from Dynabeads after washing five times (50 mM HEPES, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% sodium deoxycholate), and incubated in a 65 °C water bath overnight (12–16 h) to reverse crosslinks. ChIP DNA were subject to end repair, A-tailing, adaptor ligation and cleavage with USER enzyme, followed by size selection to 250–500 bp and amplification with NEBNext sequencing primers. Libraries were purified, quantified, multiplexed (with NEBNext Multiplex Oligos for Illumina kit, E7335S) and sequenced with 2× 75-bp pair-end reads on an Illumina NEXT-seq (Stanford Functional Genomics Facility).

ChIP–seq reads were trimmed with cutadapt (−m 50 −q 10) and aligned with bowtie2 (version 2.2.9, −no-mixed –no-discordant −end-to-end −maxins 500) to the hg38 reference genome. ChIP peaks were called with macs2 (version 2.1.1.20160309) callpeak function with broad peak option and human genome effective size using reads form corresponding to loss of gene lines as background model. Visualization tracks were generated with bedtools genomecov (−bg −scale) with scaling factor being 10^6 per number aligned reads and converted to bigWig with bedGraphToBigWig (Kent tools). BigWigs were plotted with IGV browser. Individual alignments were inspected with the IGB browser.

Heat maps were generated by intersecting bam alignment files with intervals of interest (bedtools v2.25.0), followed by tabulation of the distances of the reads relative to the centre of the interval and scaling to account for total aligned read numbers (10^6 per number aligned). Heat maps were plotted using a custom R function. Aggregate plots were generated by averaging rows of the heat map matrix. For ChIPs in control and knockout K562 clones, ChIP–seq signals in the corresponding knockout cells were used as the null reference.

For ChIP–seq repetitive sequence relationship analysis, repeat masker was intersected with ChIP–seq peak calls to classify each masker entry as MPP8-bound, MORC2-bound or unbound. Enriched families of repeats were identified with R fisher.test() followed by FDR correction with qvalue(). The distribution of sizes of occupied versus non-occupied L1 was plotted using R density() with sizes being taken from repeat masker. ks.test() was used to reject the null hypothesis that the distribution of sizes for bound and unbound L1s is the same. To investigate the relationship between L1 age, length and occupancy, logistic regression was performed with R glm() engine.

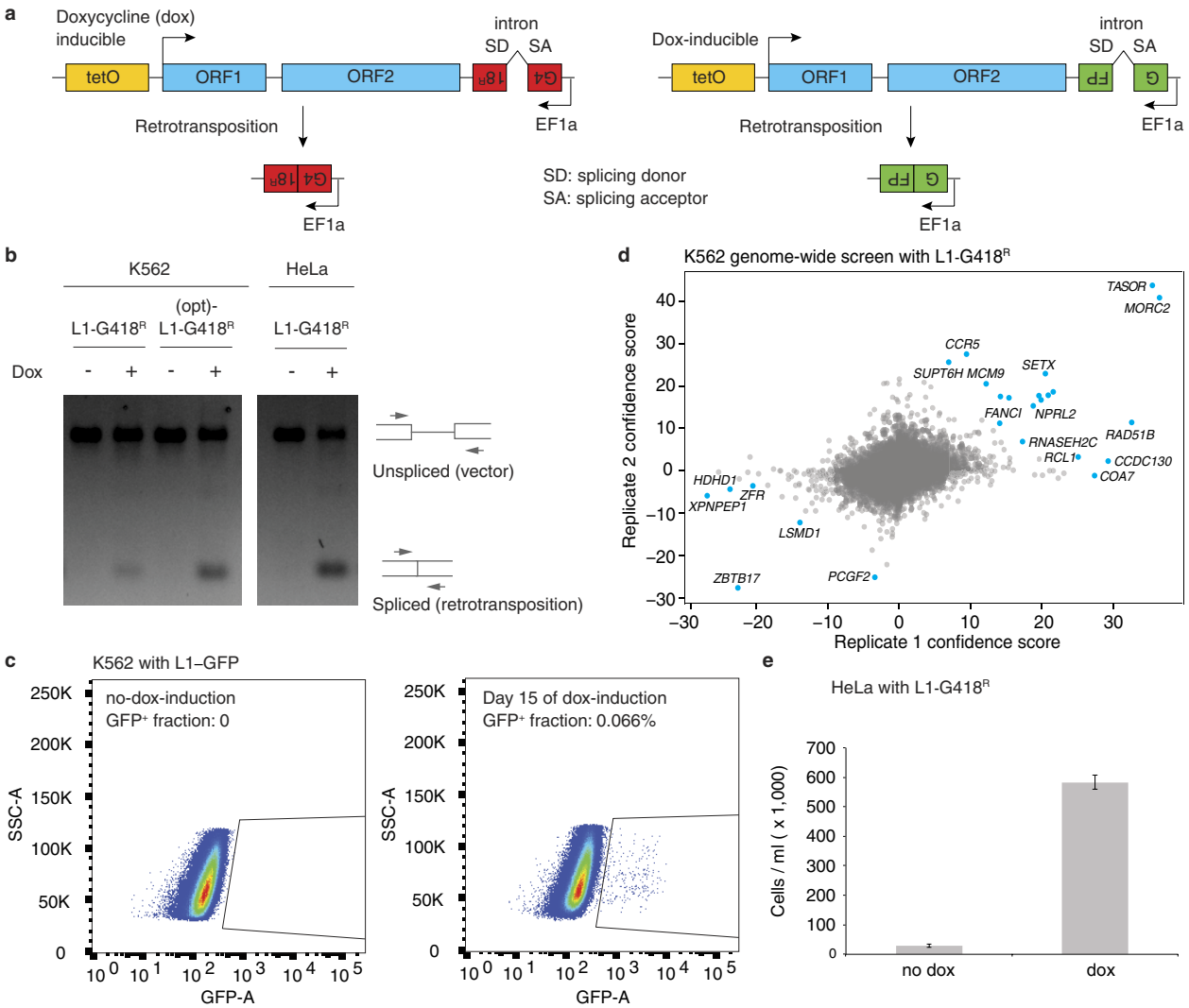
Quantitative analysis of H3K9me3 changes was performed by first identifying regions of significant enrichment in each sample relative to corresponding input sample (macs2 callpeak), merging the intervals into a common superset. This superset was joined with a decoy randomized set of intervals, twice the size of the actual experimental interval set, with the same size distribution (bedtools shuffle). Next the read coverage was determined for each sample (bedtools coverage) and regions with significant change together with fold changes were identified using DESeq2[40]. H3K9me3 regions were classified into bound or unbound by performing intersect with MORC2 and MPP8 ChIP peak calls.

Data availability. All sequencing data generated in this work has been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE95374. H3K4me3 and H3K27ac K562 ChIP–seq datasets in Fig. 3e are from BioProject (accession number PRJEB8620). hES-cell RNA-seq datasets in Extended Data Fig. 8c are from Sequence Read Archive (SRA) run entries SRR2043329 and SRR2043330. The complete results of genome-wide screens in K562 and HeLa cells are in Supplementary Table 1; the complete results of secondary screens in K562 and HeLa cells are in Supplementary Table 2. The sequences of guide RNAs and oligonucleotides used in this work are listed in Supplementary Table 3 and Supplementary Table 4. The uncropped scans with size marker indications are shown in the Supplementary Figure. All data are available from the corresponding author upon reasonable request.

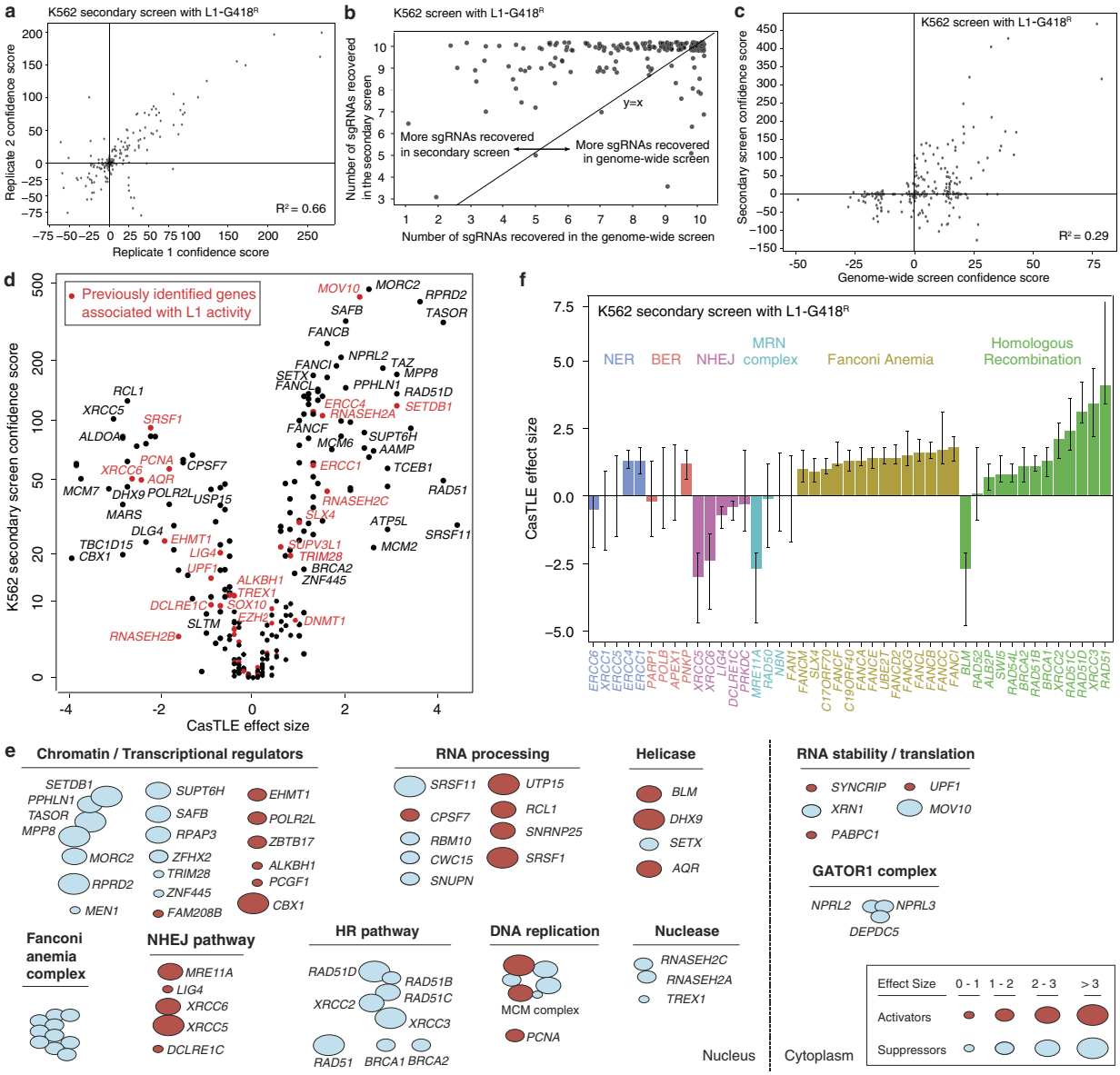
Code availability. Detailed data and further code information are available on request from the authors.

28. Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14,** 838–853 (2014).
29. Taylor, M. S. *et al.* Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155,** 1034–1048 (2013).

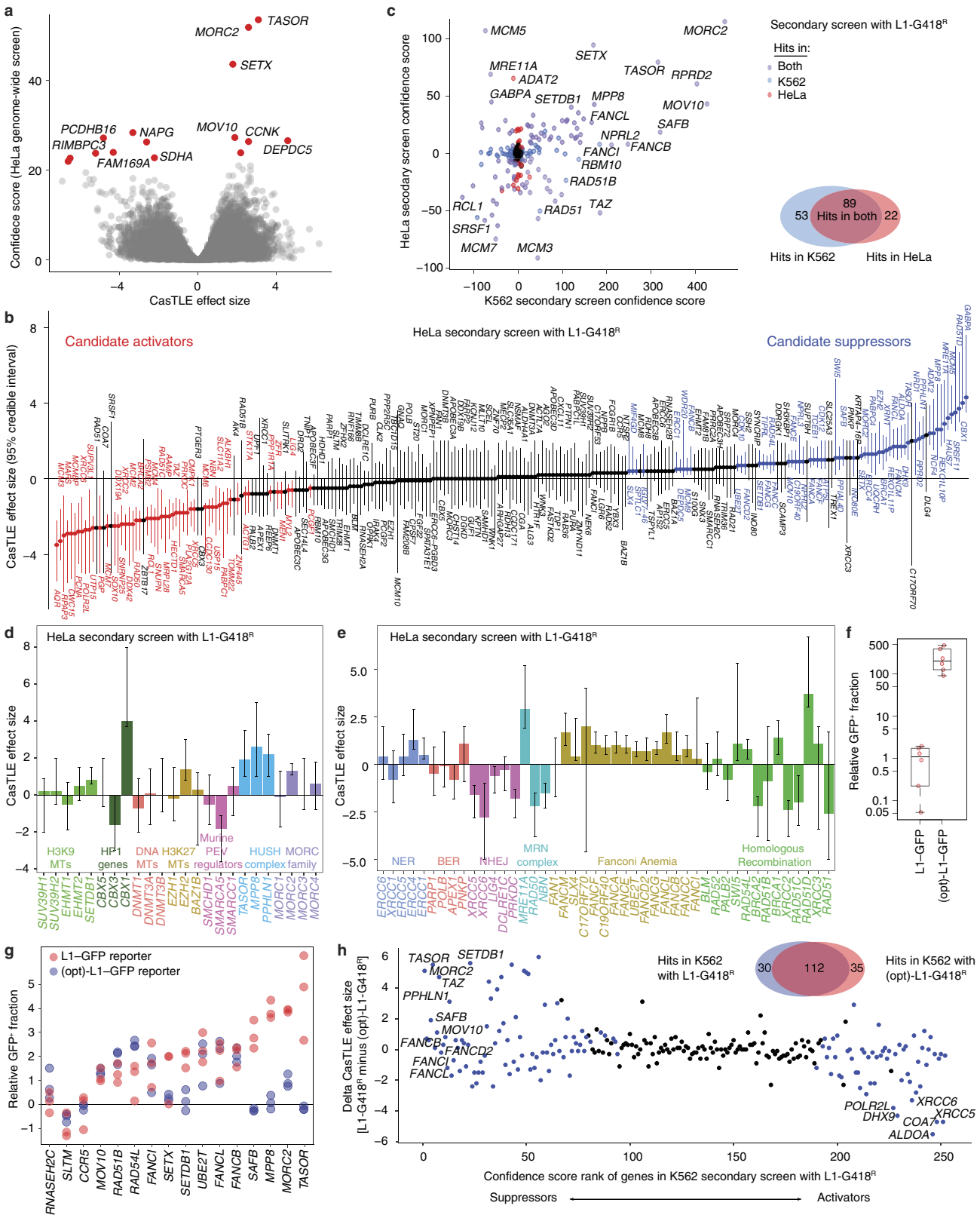
30. Brouha, B. *et al.* Evidence consistent with human L1 retrotransposition in maternal meiosis I. *Am. J. Hum. Genet.* **71,** 327–336 (2002).
31. Gasior, S. L., Roy-Engel, A. M. & Deininger, P. L. ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst.)* **7,** 983–989 (2008).
32. Deans, R. M. *et al.* Parallel shRNA and CRISPR–Cas9 screens enable antiviral drug target identification. *Nat. Chem. Biol.* **12,** 361–366 (2016).
33. Bassik, M. C. *et al.* A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152,** 909–922 (2013).
34. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
35. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460,** 1127–1131 (2009).
36. Shukla, R. *et al.* Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153,** 101–111 (2013).
37. Carreira, P. E. *et al.* Evidence for L1-associated DNA rearrangements and negligible L1 retrotransposition in glioblastoma multiforme. *Mob. DNA* **7,** 21–34 (2016).
38. Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y. & Moran, J. V. A. A 3′ poly(A) tract is required for LINE-1 retrotransposition. *Mol. Cell* **60,** 728–741 (2015).
39. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9,** 62–66 (1979).
40. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
41. Bajpai, R. *et al.* CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463,** 958–962 (2010).
42. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).



Extended Data Figure 1 | Genome-wide CRISPR–Cas9 screen for L1 regulators in K562 cells. a, Schematic of L1-G418^R and L1–GFP reporters used in this work. **b**, PCR assay on genomic DNA using primers that flank the engineered intron within the G418^R cassette. Two experiments repeated independently with similar results. The spliced PCR bands were not observed before dox induction in either K562 or HeLa cells, suggesting that the L1-G418^R reporter was not activated before the screening. However, there may exist an extremely low level of reporter leakiness that is below the detection limits of the PCR assay. **c**. Fluorescence-activated cell sorting experiments show that the L1–GFP cells have no GFP signals without dox induction (0 out of approximately 300,000 cells), and begin to produce GFP after dox induction. Therefore, the level of reporter leakiness without dox induction is insignificant. Two experiments repeated independently with similar results. **d**. CasTLE analysis of genome-wide screens in K562 cells, with 20,488 genes represented as individual points. Genes falling under 10% FDR coloured in blue, analysed by CasTLE likelihood ratio test[11]. $n = 2$ biologically independent screens. **e**. HeLa cells with L1-G418^R are resistant to G418 after dox induction (seven days of dox induction followed by ten days of G418 selection). Live cells in equal volumes were counted in a single ($n = 1$) fluorescence-activated cell sorting experiment. The centre value indicates the total number of live cells. Error bar, square root of total events assuming Poisson distribution of counts.

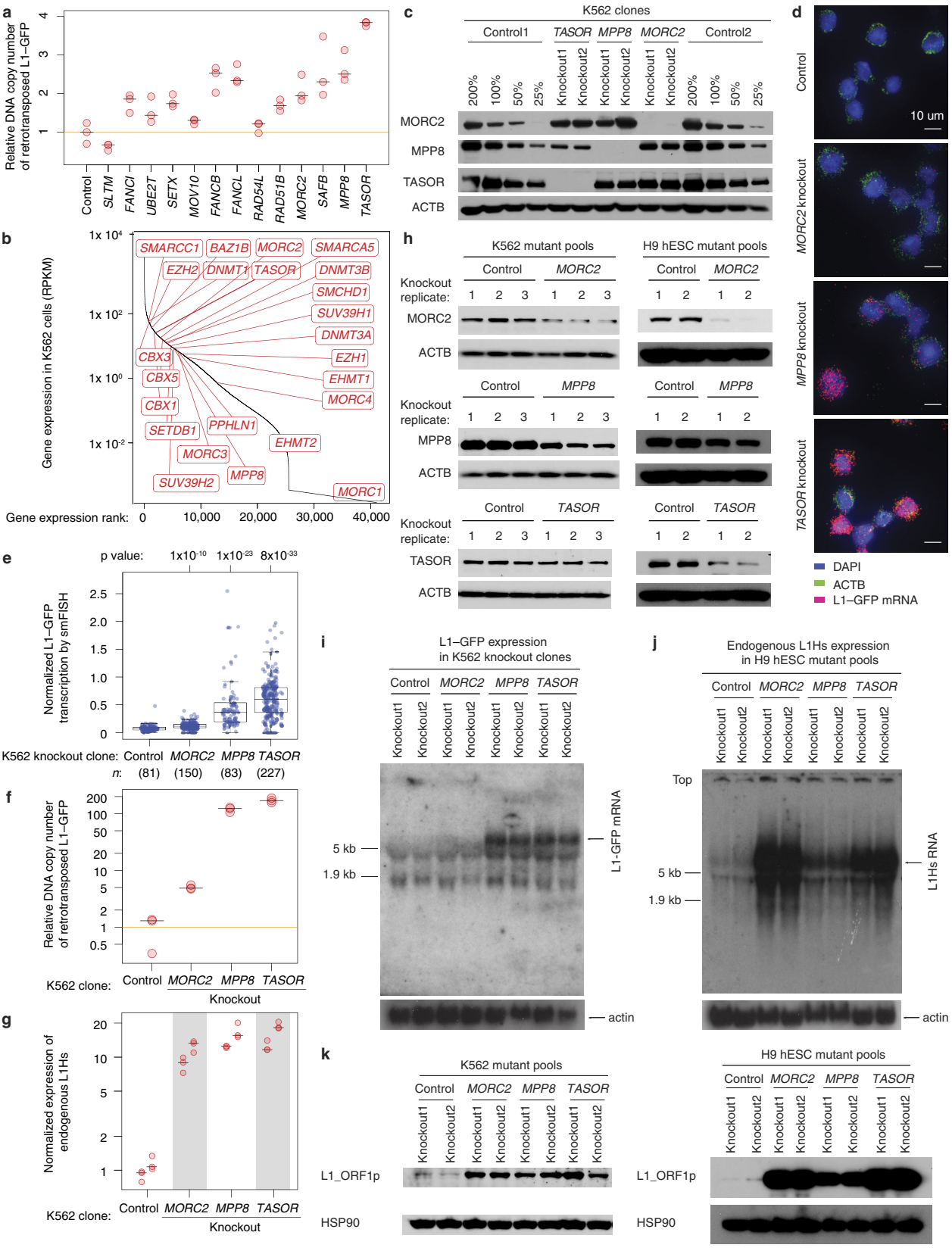


Extended Data Figure 2 | A secondary screen identifies functionally diverse L1 regulators in K562 cells. a, The reproducibility between two independent secondary screens ($n = 2$) in K562 cells. *R*-squared value is from a linear regression model. **b**. The K562 secondary screen recovers more sgRNAs than the K562 genome-wide screen, suggesting a higher detection sensitivity in the secondary screen. **c**, Comparison of the secondary screen data (252 genes from $n = 2$ independent screens) with the genome-wide screen data ($n = 2$ independent screens) in K562 cells. The *R*-squared value is from a linear regression model. **d**, Volcano plot showing K562 secondary screen results (252 genes from two independent screens), with genes previously implicated in L1 biology coloured in red. **e**, Classification-diverse L1 activators and suppressors identified in K562 cells by their known biological processes. **f**, The maximum effect size (centre value) of indicated DNA repair genes, estimated by CasTLE from two independent K562 secondary screens with ten different sgRNAs per gene. Error bars, 95% credible intervals of the estimated effect size.



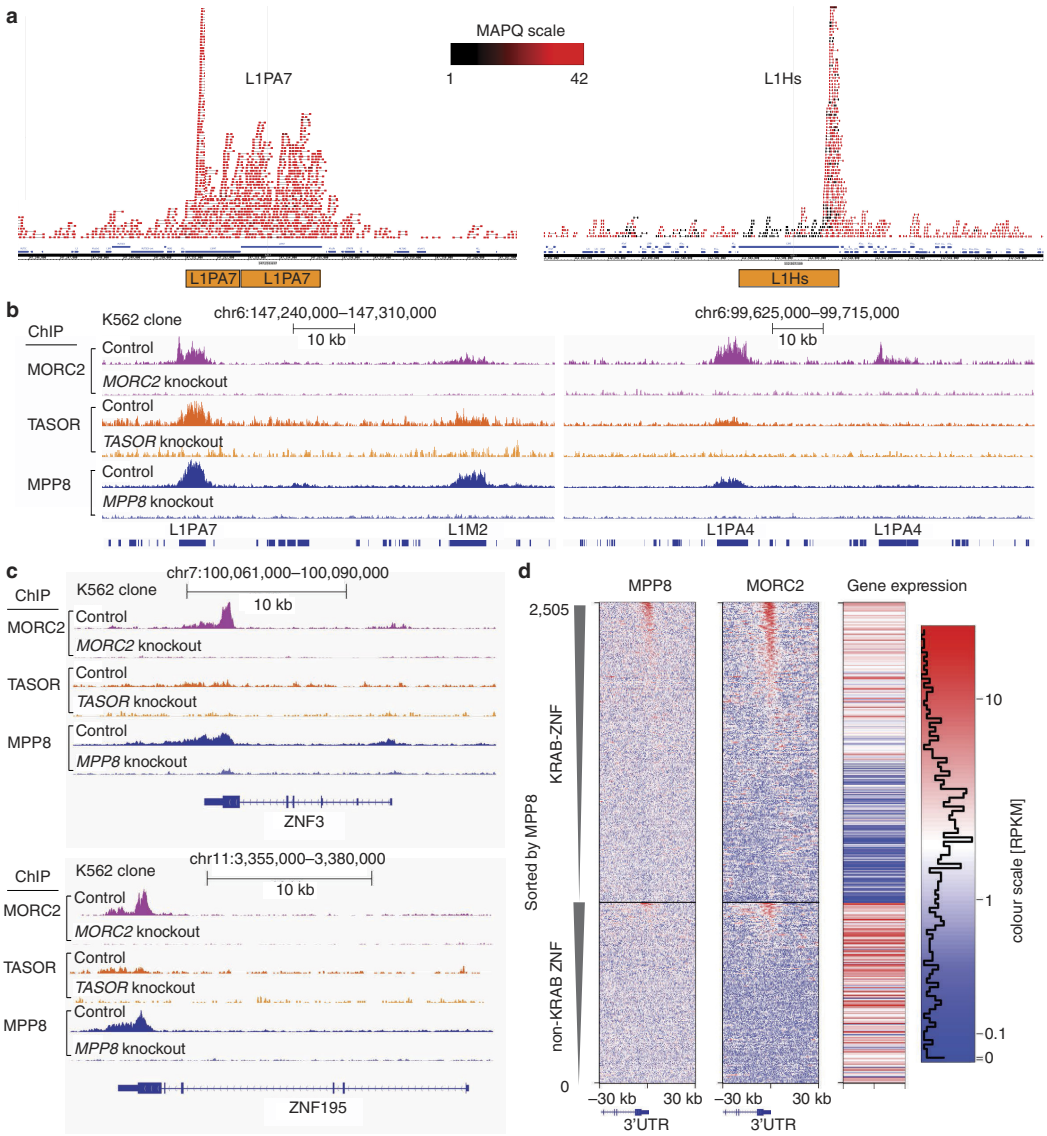
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Screen for L1 regulators in HeLa cells and and L1- sequence-dependent L1 regulators. a, CasTLE analysis of two independent genome-wide screens in HeLa cells, with 20,514 genes represented as individual points. Genes at 10% FDR cutoff coloured in red, analysed by CasTLE likelihood ratio test[11]. **b**, The maximum effect size (centre value) estimated by CasTLE from two independent HeLa secondary screens with ten different sgRNAs per gene. Bars, 95% credible interval. L1 activators are shown in red; L1 suppressors are shown in blue. Genes for which the credible interval includes zero are coloured in grey and are considered non-effective against L1. **c**, Scatter plots showing the secondary screen hits identified in K562 cells and HeLa cells (252 genes from two independent screens in each cell line), with a Venn diagram comparing hits in the two cell lines shown on the right. **d**, The maximum effect size (centre value) of indicated heterochromatin regulators, estimated by CasTLE from two independent HeLa secondary screens with ten different sgRNAs per gene. Error bars, 95% credible intervals of the estimated effect size. **e**, The maximum effect size (centre value) of indicated DNA repair genes, estimated by CasTLE from two independent HeLa secondary screens with ten different sgRNAs per gene. Error bars, 95% credible intervals of the estimated effect size. **f**, The (opt)-L1–GFP reporter retrotransposed more frequently than did L1–GFP in K562. The GFP^+ fraction of cells with the indicated L1 reporter after 15 days of dox induction was normalized to the L1–GFP sample. Box plots show median and IQR, whiskers are 1.5 × IQR. $n = 6$ biologically independent replicates. **g**, The GFP^+ fraction of dox-induced control and mutant cell pools with the L1–GFP reporter or (opt)-L1–GFP reporter. Experiments were performed as in Fig. 1e. Chromatin regulators (for example, TASOR, MORC2, MPP8 and SAFB) did not suppress the (opt)-L1–GFP reporter, in which 24% of the L1 ORF nucleotide sequence is altered, without changes in the encoded amino acid sequence[19,20], indicating that their L1 regulation depends on the native nucleotide sequence of L1Hs. **h**, K562 secondary screen with the (opt)-L1-$G418^R$ reporter (252 genes from $n = 2$ independent screens) revealed genes that regulate retrotransposition dependent or non-dependent on the native L1 nucleotide sequence. The K562 secondary screen candidates identified with L1-$G418^R$ (252 genes from $n = 2$ independent screens) are labelled in blue. A Venn diagram comparing hits identified from the two L1 reporters is shown on the top right.

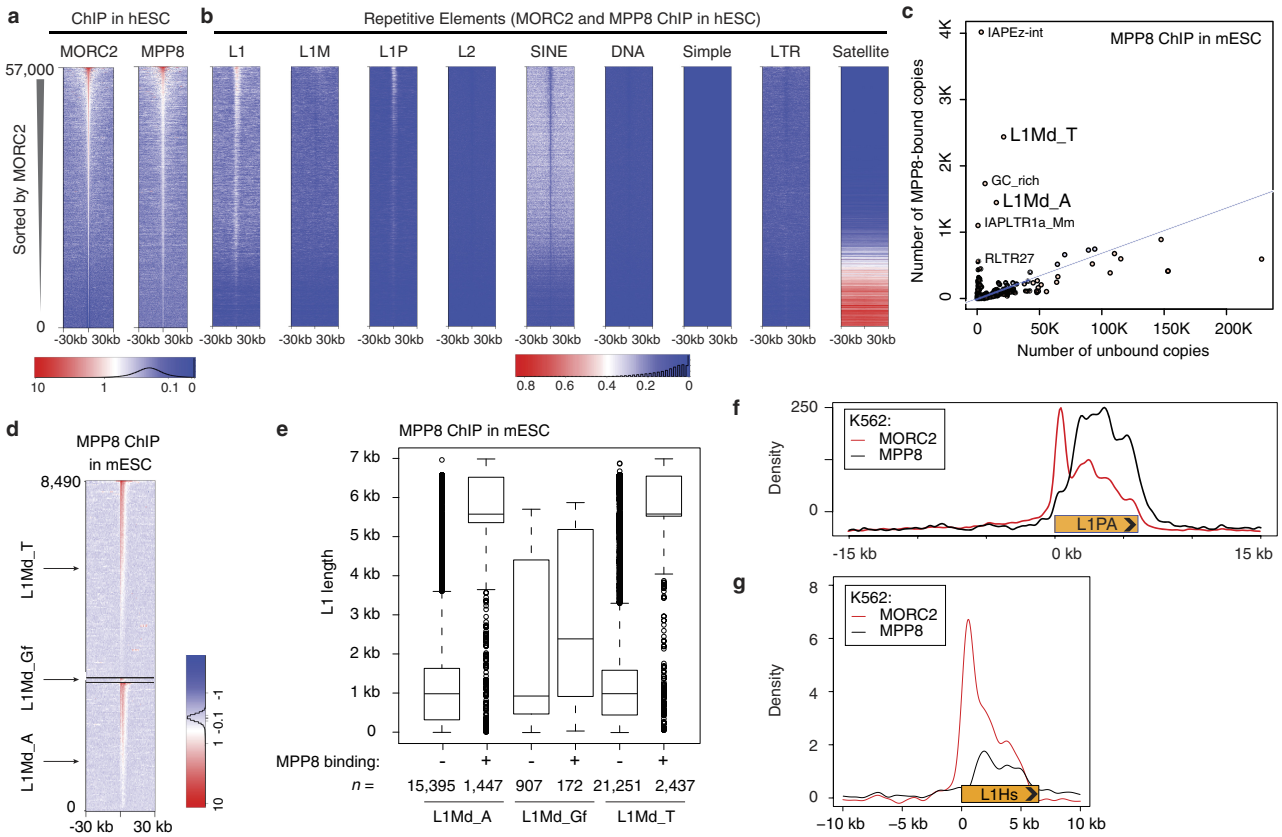


Extended Data Figure 4 | See next page for caption.

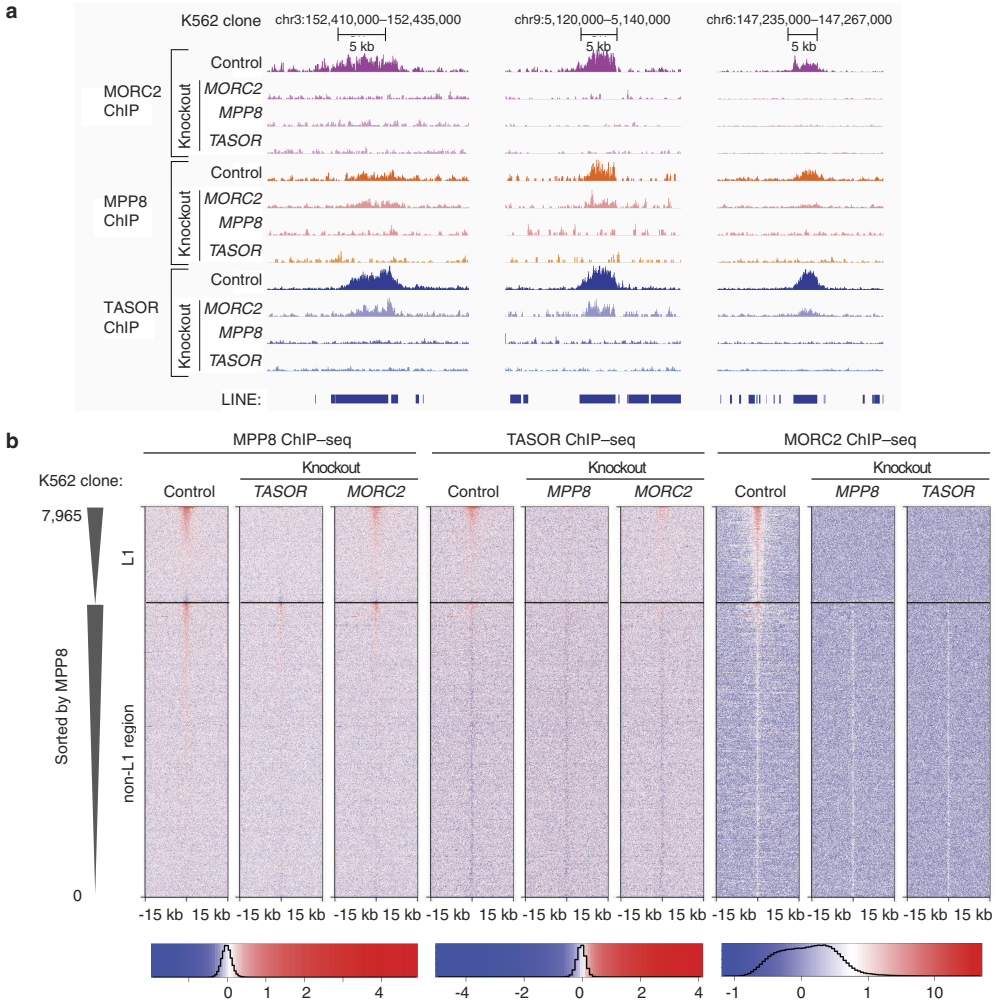
Extended Data Figure 4 | MORC2, MPP8 and TASOR silence L1 transcription. a, Relative genomic copy number of newly integrated L1–GFP reporters in the indicated mutant K562 pools after dox induction. The PspGI-assisted qPCR assay used here was designed to selectively detect spliced GFP rather than the unspliced version (see Methods). The L1–GFP copies were normalized to β-actin DNAs; data were then normalized to the control. As a putative L1 activator, SLTM shows an opposite effect on the DNA copy number compared with L1 suppressors. Centre value is the median; $n = 3$ technical replicates per gene. **b**, RNA-seq data in control K562 cells show that most heterochromatin regulators in Fig. 2a are expressed, supporting the selective effect of HUSH and MORC2 in L1 regulation. **c**, Western blots validating the effects of knockout in independent knockout K562 cell clones. Control samples were loaded at four different amounts (200%, 100%, 50% and 25% of knockout clones). Three experiments, repeated independently with similar results. To obtain knockout clones, we sorted pools of mutant K562 cells (those used in Fig. 1e, f) into 96-well plates, expanded the cells and screened for knockout clones through western blotting. Of note is that all K562 knockout clones were derived from the same starting L1–GFP reporter line, and therefore do not differ in reporter transgene integrations among the clones. **d**, Representative images of smFISH assays targeting ACTB mRNAs and RNA transcripts from L1–GFP reporters in control and knockout K562 clones after five days of dox induction. No signal was observed from L1–GFP reporters without dox induction (data not shown). Two experiments repeated independently with similar results. See also **e** and Fig. 2b (showing L1–GFP mRNA only). **e**, Quantification of the L1–GFP transcription level from the indicated number of K562 cells, determined by smFISH assays (**d** and Fig. 2b). The number of L1–GFP mRNA transcripts is normalized to the number of β-actin mRNAs within each K562 cell. Box plots show median and IQR, whiskers are 1.5 × IQR. *P* value, two-sided Wilcoxon test. 95% credible interval for median from 1,000× bootstrap: control: 0.059–0.082; MORC2: 0.106–0.123; MPP8: 0.264–0.410; TASOR: 0.514–0.671. **f**, MORC2, MPP8 and TASOR knockouts increase the genomic copy number of newly integrated L1–GFP reporters. PspGI-assisted qPCR assays were performed as in **a**, but using clonal knockout K562 clones instead of mutant cell pools. Data are normalized to the control. $n = 3$ technical replicates; centre value is median. **g**, MORC2 knockout, MPP8 knockout, and TASOR knockout increase the expression of endogenous L1s. RT–qPCR experiments were performed as in Fig. 1f, but using clonal knockout K562 clones instead of mutant cell pools. $n = 2$ biological replicates × 3 technical replicates; centre value as median. The primers do not target the L1–GFP reporter and the cell lines were not dox-induced, so these RT–qPCR assays will not detect L1–GFP transcripts. **h**, Western blots showing the effects of depletion of MORC2, MPP8 and TASOR in the mutant pools of K562 cells (left) and in the mutant pools of H9 hES cells without transgenic L1 reporters (right), independently repeated twice with similar results. **i**, Northern blots showing increased transcription from the L1–GFP reporter in knockout K562 clones (same cell lines as in **c**) after five days of dox induction, independently repeated twice with similar results. As observed in Fig. 2b, whereas HUSH knockout significantly increases L1–GFP transcription, MORC2 knockout leads to only a modest increase. This is probably because the L1–GFP reporter does not contain the native L1 5′ UTR sequence, to which MORC2 binds strongly (see Extended Data Fig. 6f, g). The 5-kb and 1.9-kb marks on the membrane refer to the 28S rRNA and 18S rRNA bands, respectively. **j**, Northern blots showing that disruption of MORC2, MPP8 and TASOR increases the expression level of endogenous L1Hs in hES cells, using the same cell lines as in **h**. Size marker indicated as in **i**. Independently repeated twice with similar results. **k**, Western blots showing protein abundance of L1_ORF1p and HSP90 in the mutant pools of K562 cells and hES cells (same cell line as shown in **h**), independently repeated twice with similar results. Experiments were performed without dox induction of the transgenic L1 reporter. Owing to the strong signal of bands from the knockout samples, the blots were exposed for a very short time and the band signals in the control samples were relatively very weak compared to the knockout samples; this is was also the case in **i**, **j**.



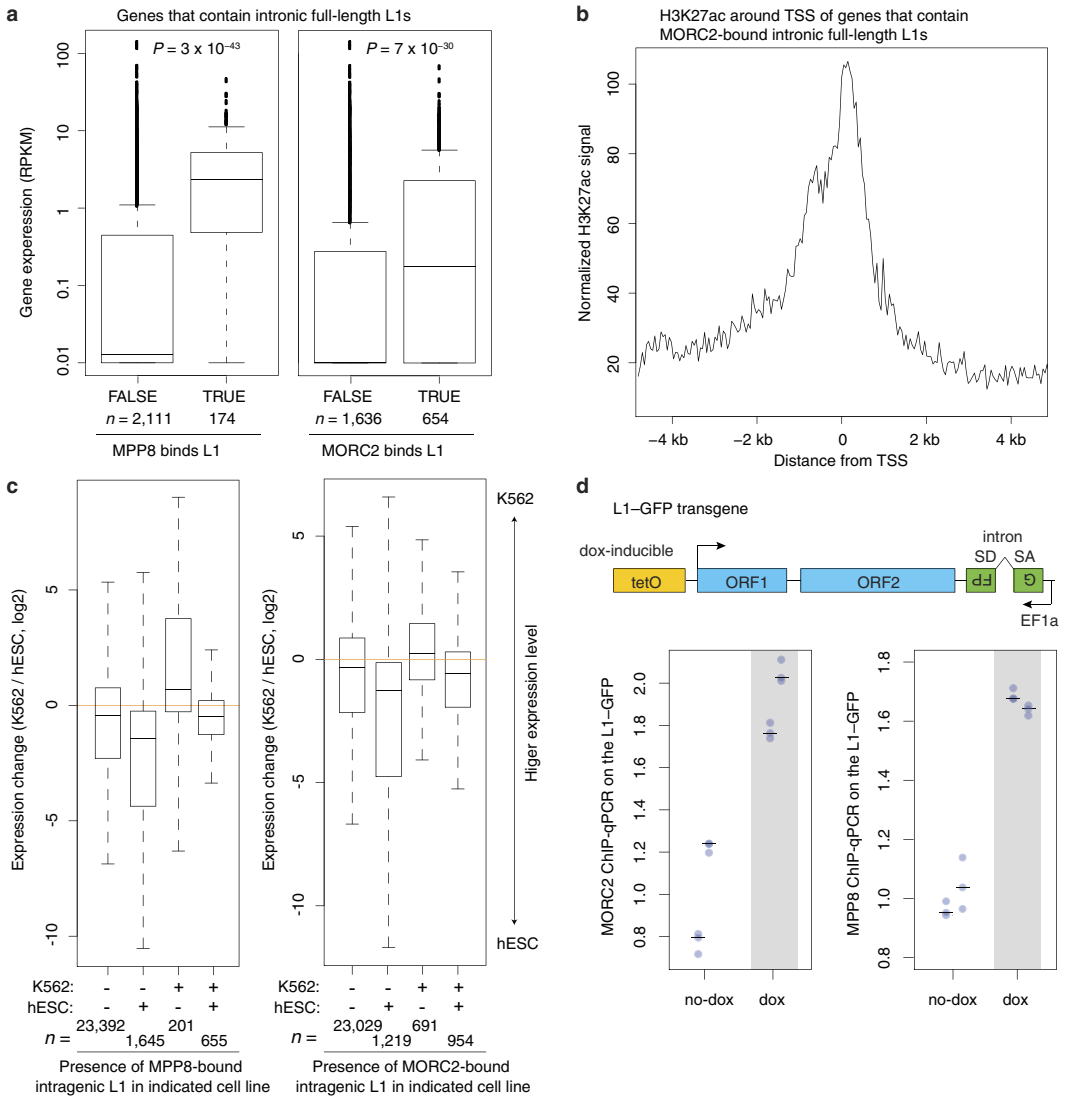
Extended Data Figure 5 | The binding profiles of MORC2, MPP8 and TASOR revealed by ChIP–seq in K562 cells. a, Using a paired-end sequencing strategy for ChIP–seq, together with the sequence divergence within native L1 elements, we could map ChIP–seq reads to individual L1 instances in the genome. Genome browser snapshots of MORC2 ChIP–seq reads alignment over L1PA7 (left) and L1Hs (right). The experiment was repeated once with similar results. The colour scale indicates the mapping quality score MAPQ for each read pair. MAPQ = 10log10p, where p is the probability that true alignment belongs elsewhere. With the exception of L1Hs, which is the youngest and least sequence-divergent family, the bodies of L1 repeats are uniquely mappable. In case of L1Hs, the 5′ UTR is still mappable to determine the level of L1Hs in control and knockout clones. **b**, Genome browser snapshots for MPP8 (blue), TASOR (orange) and MORC2 (purple) ChIP–seq read densities from control and corresponding knockout K562 clones at two representative genomic loci. The experiment was repeated once with similar results. LINE element occurrences are indicated by blue rectangles at the bottom of the plot. Four instances of long L1 elements are named, indicating the L1 families to which they belong. Note the complete absence of ChIP–seq signals from knockout lines, and selectivity towards some but not other L1 instances. Of note is that, whereas MPP8 and MORC2 ChIP signals were robust, TASOR ChIPs showed relatively weak enrichments (either owing to poor antibody quality or genuine biological properties); for this reason, a subset of our downstream analyses is focused on MORC2 and MPP8. **c**, In addition to full-length L1, the HUSH complex and MORC2 bind 3′ UTR of KRAB zinc-finger (KRAB-ZNF) genes. Genome browser snapshots of ChIP–seq read densities over representative examples, from both control and corresponding knockout K562 clones. The experiment was repeated once with similar results. **d**, The HUSH complex and MORC2 preferentially bind expressed KRAB-ZNF genes over other ZNF genes. Heat maps of MPP8 (left) and MORC2 (centre) signals over 2,600 ZNF genes, centred at the 3′ end of the genes and sorted first by the presence of the KRAB domain and then by the MPP8 ChIP signal. The upper 1,600 genes are KRAB-ZNF genes, the lower 1,000 genes are non-KRAB ZNF genes. The heat maps on the right code the absolute expression level of each gene in RPKM scale from the K562 RNA-seq data (far right).



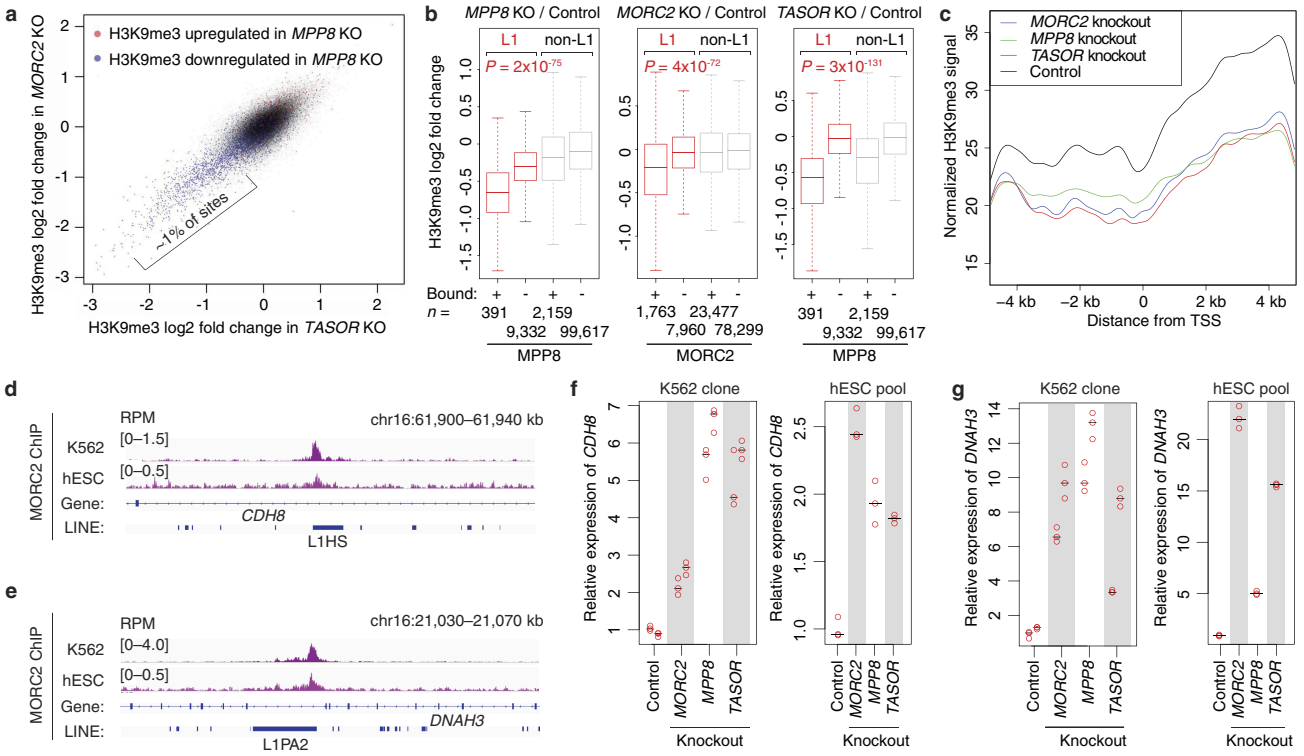
Extended Data Figure 6 | HUSH and MORC2 preferentially bind full-length L1 instances in human ES cells, mouse ES cells and K562 cells. **a**, Widespread genomic co-binding of MPP8 and MORC2 in hES cells. Heat map representation of ChIP–seq results at 57,000 genomic loci, centred on MPP8 and MORC2 summits and sorted by MORC2 ChIP–seq signal. Plotted is the normalized ChIP read density from hES cells. **b**, Heat maps of MORC2/MPP8 ChIP–seq density over the indicated repeat classes, centred and sorted as in **a**. The HUSH complex and MORC2 bind predominantly to L1 elements in hES cells, in particular to the primate-specific L1P families, suggesting that HUSH- and MORC2-dependent silencing is relevant in many embryonic and somatic cell types. **c**, L1 families that encompass active L1 copies, such as L1Md-T and L1Md-A, are significantly enriched among MPP8 binding sites in mouse ES cells. L1Md_Gf is also enriched, but not shown owing to the low number of instances. Thus, HUSH-mediated L1 regulation appears to be conserved among species. Of note is that MPP8 is also strongly enriched at IAP elements, a class of murine endogenous retroviruses that remain currently mobile in the mouse genome. **d**, MPP8 ChIP–seq heat maps in mouse embryonic stem cells featuring retrotransposition-competent L1Md-T, L1Md-A and L1Md-Gf. **e**, MPP8 preferentially binds full-length L1Md-A and L1Md-T in mouse ES cells. Plotted is the size distribution of the indicated L1 instances that overlap with MPP8 ChIP–seq peaks, or remaining L1s that do not overlap with such ChIP–seq signals. Box plots show median and IQR, whiskers are 1.5 × IQR. **f**, Aggregate plots of MORC2 (red) and MPP8 (black) ChIP–seq signals over 500 full-length, MPP8-bound L1PAs, centred on the L1 5′ end. **g**, Aggregate plots of MORC2 (red) and MPP8 (black) ChIP–seq signals on L1Hs (L1PA1). Similar to the binding profile on L1PA (**f**), MPP8 and MORC2 occupy the whole body of L1Hs, with MORC2 additionally binding L1Hs 5′ UTR. It is to be noted that ChIP–seq fragments are much less likely to be uniquely mapped—and thus removed by the alignment criteria—within the L1Hs non-5′ UTR region, owing to their minimal sequence divergence (Extended Data Fig. 5a).



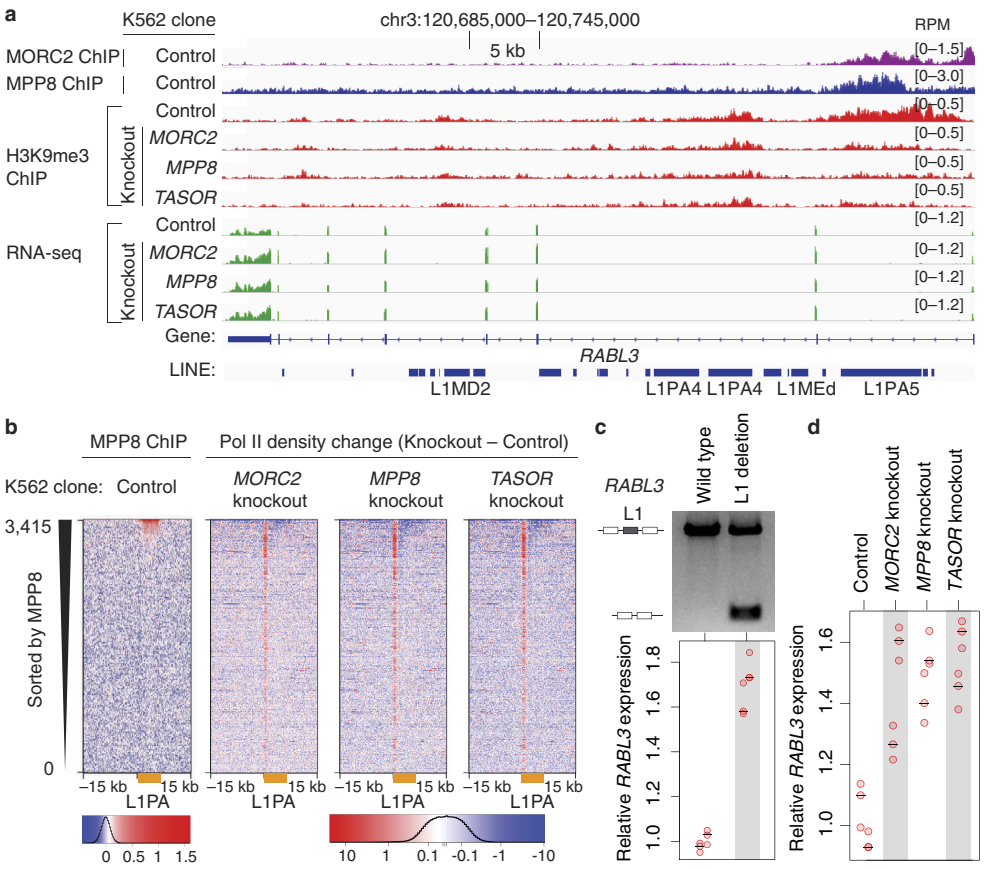
Extended Data Figure 7 | HUSH and MORC2 collaborate in binding target L1s. a, A representative genome browser view of normalized ChIP–seq read densities over L1 elements. The experiment was repeated once with similar results. The loss of MPP8 and TASOR results in no detectable binding by MORC2, MPP8 and TASOR, whereas the loss of MORC2 results in partially diminished recruitment of HUSH complex subunits. **b**, Heat maps of MPP8 (left), TASOR (centre) and MORC2 (right) ChIP–seq signals subtracted for the ChIP signal from corresponding knockout lines. Heat maps are centred on MPP8 and MORC2 peaks, separated by the presence or absence of underlying L1 and then sorted by MPP8 ChIP signal strength. The loss of MORC2 has only a partial effect on the recruitment of MPP8 and TASOR to the L1 elements, whereas the loss of either MPP8 or TASOR abrogates MORC2 recruitment.



Extended Data Figure 8 | HUSH and MORC2 preferentially bind intronic L1s within actively transcribed genes. a. Genes that contain MPP8- or MORC2-bound intronic L1s are expressed at significantly higher levels in control K562 cells, compared to genes that contain intronic full-length L1s unbound by MPP8 or MORC2. *P* value, two-sided Mann–Whitney-Wilcoxon test. Box plots show median and IQR, whiskers are 1.5 × IQR. **b**. The promoters of genes that contain MPP8- or MORC2-bound intronic full-length L1s are marked by transcriptionally permissive H3K27ac in wild-type K562 cells. H3K27ac ChIP–seq data are taken from a K562 epigenome pilot study, accession number PRJEB8620. TSS, transcription start site. **c**, Genes selectively occupied by MORC2 or MPP8 either in K562 or in hES cells exhibit higher gene expression in the corresponding cell line ($P = 4.3 \times 10^{-107}$ for MPP8 binding; 5.0×10^{-92} for MORC2 binding, Kruskal–Wallis test). Boxplots defined as in **a**. RNA-seq datasets for hES cells are from SRA entries SRR2043329 and SRR2043330. **d**, ChIP–qPCR assays quantifying HUSH and MORC2 binding to an inducible L1 transgene in K562 cells before or after its transcriptional induction by dox. Transcriptional induction increases binding of MORC2 and MPP8 to the L1 transgene. $n = 2$ biological replicates × 3 technical replicates; centre value is median.



Extended Data Figure 9 | HUSH and MORC2 facilitate H3K9me3 at their L1 targets for transcription repression. a, A concordant subset (approximately 1%) of 111,499 H3K9me3 sites in the genome lose the H3K9me3 signal in MORC2 knockout, MPP8 knockout and TASOR knockout K562 clones. Two independent lines each for wild-type, MORC2 knockout, TASOR knockout or MPP8 knockout clone. Plotted is the $\log_2$ fold change in the H3K9me3 ChIP signal in the TASOR knockout relative to the control (*x* axis) and the $\log_2$ fold change in the H3K9me3 ChIP signal in the MORC2 knockout relative to the control (*y* axis). Points are colour-coded with blue sites having significant H3K9me3 loss in MPP8 knockout, red sites significantly gaining a signal in MPP8 knockout, and grey sites having no detectable change. Sites that significantly lose the H3K9me3 signal in a knockout line are more likely to show a corresponding loss in other knockout lines. Odds ratios: 26.23 with 95% confidence intervals (23,66, 29.10) for MORC2 versus MPP8; 21.70 with 95% confidence intervals (19.75, 23.83) for TASOR versus MPP8; 122.53 with 95% confidence intervals (109.21, 137.43) for TASOR versus MORC2. $P = 0$ each case, two-sided Fisher's exact test. **b**, Genomic sites that exhibit the strongest loss of H3K9me3 in MORC2, MPP8 or TASOR knockouts are preferentially L1-occupied by MORC2, MPP8 or TASOR respectively. Boxplots of $\log_2$ fold change in H3K9me3 relative to control for MPP8 knockout (left), MORC2 knockout (centre) and TASOR knockout (right). Box plots show median and IQR, whiskers are 1.5 × IQR. MPP8- and MORC2-bound L1s show a significant loss of H3K9me3 (*P* values, two-sided Mann–Whitney–Wilcoxon test). **c**, Averaged distribution of H3K9me3 ChIP–seq signals in control and knockout K562 clones over the host genes that contain the MORC2-targeted intronic full-length L1s, centred on the TSS of the host genes. **d**, Genome browser track showing MORC2 binding at the intronic full-length L1Hs within *CDH8* in both K562 and hES cells. The experiment was repeated once with similar results. **e**, Genome browser track showing MORC2 binding at the intronic full-length L1PA2 within *DNAH3* in both K562 and hES cells. The experiment was repeated once with similar results. **f**, Depletion of MORC2 or HUSH increases the expression of *CDH8* in both K562 ($n = 2$ biological replicates × 3 technical replicates) and hES cells ($n = 3$ technical replicates), as measured by RT–qPCR assay. The *CDH8* expression level was normalized to β-actin mRNA. All samples were then normalized to the control sample. Centre value is median. **g**, Depletion of MORC2 or HUSH increases the expression of *DNAH3* in both K562 ($n = 2$ biological replicates × 3 technical replicates) and hES cells ($n = 3$ technical replicates), as measured by RT–qPCR assay. The *DNAH3* expression level was normalized to β-actin mRNA. All samples were then normalized to the control sample. Centre value is median.



Extended Data Figure 10 | HUSH and MORC2 binding at intronic L1s results in the decreased expression of active host genes. a, Genome browser tracks illustrating that the loss of HUSH and MORC2 causes decreased H3K9me3 over the intronic L1PA5 element and concomitant increase in the expression of host gene *RABL3*. The experiment was repeated once with similar results. **b**, The loss of HUSH or MORC2 leads to increased Pol II signals at the 5′ UTR and decreased Pol II signals within L1 bodies at HUSH-bound L1PA elements (orange bars). Heat maps show Pol II density change in knockout K562 clones compared to control, centred on the L1 5′ end and sorted by MPP8 ChIP signal. **c**, Deletion of the intronic L1 within *RABL3* causes increased *RABL3* expression. Top, an agarose gel analysis of the PCR assay with primers flanking the HUSH- or MORC2-bound intronic L1; two experiments repeated independently with similar results. Bottom, RT–qPCR analysis of *RABL3* expression. The *RABL3* expression level was normalized to β-actin mRNA. All samples were then normalized to the wild-type sample. $n = 2$ biological replicates × 3 technical replicates (centre value is median). **d**, Depletion of MORC2, MPP8 or TASOR increases *RABL3* expression. RT–qPCR data normalized as in **c**. $n = 2$ biological replicates × 3 technical replicates (centre value is median).

nature research

Corresponding author(s): Joanna Wysocka and Michael Bassik

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

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

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

Experimental design

1. Sample size

Describe how sample size was determined.

For the screen, two independent replicate screens were performed, which are sufficient for screening technologies. See Methods section, 'Genome-wide screen in K562 cells', 'Secondary screen in K562 cells' and 'Genome-wide screen and secondary screen in HeLa cells'.

2. Data exclusions

Describe any data exclusions.

No data was excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Once experiments and procedures were fully optimized, all attempts at replication were successful.

4. Randomization

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

K562 KO clones were allocated into experimental groups based on their genotype (Extended Data Figure 4c). For smFISH experiments, individual cells were allocated into experimental groups based on their genotype (Figure 2b and Extended Data Figure 4d).

5. Blinding

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

Blinding was performed during the smFISH data analyses (Extended Data Figure 4e), where B.G. who analyzed smFISH images was blinded to all conditions.

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

6. Statistical parameters

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

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

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

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	For Images analyses we used MATLAB 2016b (MathWorks). Flowjo 9.9 was used for flow cytometry analyses. For statistical analysis we used R 3.3.2. For ChIP-seq genomic alignments we used bowtie2 v.2.2.9, peak calls with MACS2 v2.1.1.20160309, IGV_2.3.92 and IGB 9.0.0 for visualization, bedtools v2.17.0 and GNU awk 4.1.3 for overlap statistics and genome interval manipulation. For humans hg38 reference genome was used, for mouse mm10. RNAseq alignments were performed with hisat2 v2..0.5, followed by stringtie v 1.3.3b and featureCounts v1.4.6-p2, further analysis was performed with Bioconductor 3.4 and DESeq2 1.14.1, human genocde 25 transcript models were used.

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

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	1) Rabbit MORC2 antibody (A300-149A, Bethyl Laboratories), validated by vendor, and used in previous literature. 2) Rabbit MPP8 antibody (16796-1-AP, Protein Technologies Inc), validated by vendor, and used in previous literature. 3) Rabbit TASOR antibody (HPA006735, Atlas Antibodies), validated by vendor, and used in previous literature. 3) Mouse anti-LINE-1 ORF1p antibody (MABC1152, Millipore), validated by vendor, and used in previous literature. 4) Rabbit HSP90 (C45G5, Cell Signalling, #4877), Extensively used in the literature. 5) Beta actin antibody (ab49900, Abcam), Extensively used in the literature. 6) Histone H3 (tri-methyl K9) antibody (ab8898, Abcam), validated by vendor, and used in previous literature. 7) RNA Pol II (Santa Cruz Biotechnology, N-20 sc-899), validated by vendor, and used in previous literature.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Cell lines are from commercial sources. HeLa and K562: ATCC Human Embryonic Stem Cells, H9: WiCell Mouse Embryonic Stem Cells, ES-E14TG2a: ATCC
b. Describe the method of cell line authentication used.	Cell lines were authenticated by the vendor. All cells were obtained from comerical sources. HeLa, K562 and mESC (ATCC). Human Embryonic Stem Cells H9 (WiCell).
c. Report whether the cell lines were tested for mycoplasma contamination.	Cell cultures were routinely tested and found negative for mycoplasma infection (MycoAlert, Lonza).
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used in this study are in the database of commonly misidentified cell lines.

Animals and human research participants

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

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	This study did not involve human research participants.

nature research

Corresponding author(s): Joanna Wysocka, Michael Bassik

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

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

Data deposition

1. For all ChIP-seq data:

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

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

2. Provide all necessary reviewer access links.
The entry may remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=ojerwuukpzsbjsb&acc=GSE95374

3. Provide a list of all files available in the database submission.

GSM2509455
ChIP:MORC2_Cell:WT_rep1
GSM2509456
ChIP:MORC2_Cell:WT_rep2
GSM2509457
ChIP:MORC2_Cell:MORC2-KO_rep1
GSM2509458
ChIP:MORC2_Cell:MORC2-KO_rep2
GSM2509459
ChIP:MORC2_Cell:MPP8-KO_rep1
GSM2509460
ChIP:MORC2_Cell:MPP8-KO_rep2
GSM2509461
ChIP:MORC2_Cell:TASOR-KO_rep1
GSM2509462
ChIP:MORC2_Cell:TASOR-KO_rep2
GSM2509463
ChIP:MPP8_Cell:WT_rep1
GSM2509464
ChIP:MPP8_Cell:WT_rep2
GSM2509465
ChIP:MPP8_Cell:MORC2-KO_rep1
GSM2509466
ChIP:MPP8_Cell:MORC2-KO_rep2
GSM2509467
ChIP:MPP8_Cell:MPP8-KO_rep1
GSM2509468
ChIP:MPP8_Cell:MPP8-KO_rep2
GSM2509469
ChIP:MPP8_Cell:TASOR-KO_rep1
GSM2509470
ChIP:MPP8_Cell:TASOR-KO_rep2
GSM2509471
ChIP:TASOR_Cell:WT_rep1
GSM2509472
ChIP:TASOR_Cell:WT_rep2
GSM2509473
ChIP:TASOR_Cell:MORC2-KO_rep1
GSM2509474
ChIP:TASOR_Cell:MORC2-KO_rep2

GSM2509475
ChIP:TASOR_Cell:MPP8-KO_rep1
GSM2509476
ChIP:TASOR_Cell:MPP8-KO_rep2
GSM2509477
ChIP:TASOR_Cell:TASOR-KO_rep1
GSM2509478
ChIP:TASOR_Cell:TASOR-KO_rep2
GSM2509479
ChIP:Input(MORC2, MPP8, TASOR)_Cell:WT_rep1
GSM2509480
ChIP:Input(MORC2, MPP8, TASOR)_Cell:WT_rep2
GSM2509481
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MORC2-KO_rep1
GSM2509482
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MORC2-KO_rep2
GSM2509483
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MPP8-KO_rep1
GSM2509484
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MPP8-KO_rep2
GSM2509485
ChIP:Input(MORC2, MPP8, TASOR)_Cell:TASOR-KO_rep1
GSM2509486
ChIP:Input(MORC2, MPP8, TASOR)_Cell:TASOR-KO_rep2
GSM2509487
ChIP:H3K9me3_Cell:WT_rep1
GSM2509488
ChIP:H3K9me3_Cell:WT_rep2
GSM2509489
ChIP:H3K9me3_Cell:MORC2-KO_rep1
GSM2509490
ChIP:H3K9me3_Cell:MORC2-KO_rep2
GSM2509491
ChIP:H3K9me3_Cell:TASOR-KO_rep1
GSM2509492
ChIP:H3K9me3_Cell:TASOR-KO_rep2
GSM2509493
ChIP:Input(H3K9me3)_Cell:WT_rep1
GSM2509494
ChIP:Input(H3K9me3)_Cell:WT_rep2
GSM2509495
ChIP:Input(H3K9me3)_Cell:MORC2-KO_rep1
GSM2509496
ChIP:Input(H3K9me3)_Cell:MORC2-KO_rep2
GSM2509497
ChIP:Input(H3K9me3)_Cell:TASOR-KO_rep1
GSM2509498
ChIP:Input(H3K9me3)_Cell:TASOR-KO_rep2
GSM2509503
ChIP:H3K9me3_Cell:WT_rep3
GSM2509504
ChIP:H3K9me3_Cell:WT_rep4
GSM2509505
ChIP:H3K9me3_Cell:MPP8-KO_rep1
GSM2509506
ChIP:H3K9me3_Cell:MPP8-KO_rep2
GSM2509507
ChIP:Input(H3K9me3)_Cell:WT_rep1
GSM2509508
ChIP:Input(H3K9me3)_Cell:WT_rep2

GSM2509509
ChIP:Input(H3K9me3)_Cell:MPP8-KO_rep1
GSM2509510
ChIP:Input(H3K9me3)_Cell:MPP8-KO_rep2
GSM2789802
ChIP:PolII_Cell_K562:WT_rep1
GSM2789803
ChIP:PolII_Cell_K562:WT_rep2
GSM2789804
ChIP:PolII_Cell_K562:MORC2-KO_rep1
GSM2789805
ChIP:PolII_Cell_K562:MORC2-KO_rep2
GSM2789806
ChIP:PolII_Cell_K562:MPP8-KO_rep1
GSM2789807
ChIP:PolII_Cell_K562:MPP8-KO_rep2
GSM2789808
ChIP:PolII_Cell_K562:TASOR-KO_rep1
GSM2789809
ChIP:PolII_Cell_K562:TASOR-KO_rep2
GSM2789810
ChIP:Input_Cell_K562:WT_rep1
GSM2789811
ChIP:Input_Cell_K562:WT_rep2
GSM2789812
ChIP:MPP8_Cell_hESC:WT_rep1
GSM2789813
ChIP:MPP8_Cell_hESC:WT_rep2
GSM2789814
ChIP:MORC2_Cell_hESC:WT_rep1
GSM2789815
ChIP:MORC2_Cell_hESC:WT_rep2
GSM2789816
ChIP:Input_Cell_hESC:WT_rep1
GSM2789817
ChIP:Input_Cell_hESC:WT_rep2
GSM2789818
ChIP:MPP8_Cell_mESC:WT_rep1
GSM2789819
ChIP:MPP8_Cell_mESC:WT_rep2
GSM2789820
ChIP:Input_Cell_mESC:WT_rep1
GSM2789821
ChIP:Input_Cell_mESC:WT_rep2
MORC2KO1_CHIP_INPUT_1.bw
MORC2KO1_CHIP_INPUT.bw
MORC2KO1_CHIP_K9me3.bw
MORC2KO1_CHIP_MORC2_1.bw
MORC2KO1_CHIP_MPP8_1.bw
MORC2KO1_CHIP_TASOR_1.bw
MORC2KO2_CHIP_INPUT_1.bw
MORC2KO2_CHIP_INPUT.bw
MORC2KO2_CHIP_K9me3.bw
MORC2KO2_CHIP_MORC2_1.bw
MORC2KO2_CHIP_MPP8_1.bw
MORC2KO2_CHIP_TASOR_1.bw
MPP8_1.bw
MPP8_2.bw
MPP8KO1_CHIP_INPUT_1.bw
MPP8KO1_CHIP_INPUT_K9_1.bw

MPP8KO1_CHIP_K9me3.bw
MPP8KO1_CHIP_MORC2_1.bw
MPP8KO1_CHIP_MPP8_1.bw
MPP8KO1_CHIP_TASOR_1.bw
MPP8KO2_CHIP_INPUT_1.bw
MPP8KO2_CHIP_INPUT_K9.bw
MPP8KO2_CHIP_K9me3.bw
MPP8KO2_CHIP_MORC2_1.bw
MPP8KO2_CHIP_MPP8_1.bw
MPP8KO2_CHIP_TASOR_1.bw
SAFE_1.bw
SAFE_2.bw
TASOR_1.bw
TASOR_2.bw
TASORKO1_CHIP_INPUT_1.bw
TASORKO1_CHIP_INPUT.bw
TASORKO1_CHIP_K9me3.bw
TASORKO1_CHIP_MORC2_1.bw
TASORKO1_CHIP_MPP8_1.bw
TASORKO1_CHIP_TASOR_1.bw
TASORKO2_CHIP_INPUT_1.bw
TASORKO2_CHIP_INPUT.bw
TASORKO2_CHIP_K9me3.bw
TASORKO2_CHIP_MORC2_1.bw
TASORKO2_CHIP_MPP8_1.bw
TASORKO2_CHIP_TASOR_1.bw
WT1_CHIP_INPUT_1.bw
WT1_CHIP_INPUT.bw
WT1_CHIP_INPUT_K9.bw
WT1_CHIP_K9me3.bw
WT1_CHIP_MORC2_1.bw
WT1_CHIP_MPP8_1.bw
WT1_CHIP_TASOR_1.bw
WT2_CHIP_INPUT_1.bw
WT2_CHIP_INPUT.bw
WT2_CHIP_INPUT_K9.bw
WT2_CHIP_K9me3.bw
WT2_CHIP_MORC2_1.bw
WT2_CHIP_MPP8_1.bw
WT2_CHIP_TASOR_1.bw
hESC_Input1.bw
hESC_Input2.bw
hESC_MORC2_ChIP1.BW
hESC_MORC2_ChIP2.BW
hESC_MPP8_ChIP1.BW
hESC_MPP8_ChIP2.BW
INPUT_1.bw
INPUT_2.bw
INPUT_mESC1.bw
INPUT_mESC2.bw
MPP8_mESC1.bw
MPP8_mESC2.bw
PolII_MORC2_KO1.bw
PolII_MORC2_KO2.bw
PolII_MPP8_KO1.bw
PolII_MPP8_KO2.bw
PolII_TASOR_KO1.bw
PolII_TASOR_KO2.bw
PolII_WT1.bw

	PolII_WT2.bw
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).	n/a

Methodological details

5. Describe the experimental replicates.	ChIP experiments (MORC2, MPP8, TASOR, H3K9me3 and RNA PolII) were performed in two biological replicates each, with indicated antibodies. Peaks were extensively validate using ChIP-qPCR.
6. Describe the sequencing depth for each experiment.	To amplify each library we used quantitative PCR (qPCR) to endure that all libraries were amplified similarly and avoid bottlenecking of the libraries. ChIP-Seqs are pair ended, 75 bp was the read length. On average, each ChIP-seq sample contain ~40 million reads, with above 70-80% alignment.
7. Describe the antibodies used for the ChIP-seq experiments.	Rabbit MORC2 antibody (A300-149A, Bethyl Laboratories), Rabbit MPP8 antibody (16796-1-AP, Protein Technologies Inc), Rabbit TASOR antibody (HPA006735, Atlas Antibodies), Histone H3 (tri-methyl K9) antibody (ab8898, Abcam) and RNA Pol II (Santa Cruz Biotechnology, N-20 sc-899) were used in ChIP experiments.
8. Describe the peak calling parameters.	Pair-end reads were trimmed with cutadapt (-m 50 -q 10) and aligned with bowtie2 (version 2.2.9, --no-mixed --no-discordant --end-to-end -maxins 500) to the hg38 reference genome. ChIP peak calls were performed with macs2 callpeak using default settings, except for --broad flag. Background files were either ChIP input sequencing or ChIPseq from knockout cell lines for factor ChIPped. For final list of sites MACS2 peak calls were merged, combined with 2x amount of shuffled decoy sites and read coverage for each sequencing file was obtained using bedtool coverage. Combined coverage matrix was subjected to DESeq2 procedure to reject false positives from MACS2
9. Describe the methods used to ensure data quality.	Visualization tracks were generated with bedtools genomecov (-bg -scale) with scaling factor being 10^6/number aligned reads and converted to bigWig with bedGraphToBigWig (Kent tools). BigWigs were plotted with IGV browser. Individual alignments were inspected with IGB browser. Heatmaps were generated by intersecting bam alignment files with intervals of interest (bedtools v2.25.0), followed by tabulation of the distances of the reads relative to the center of the interval and scaling to account for total aligned read numbers (10^6/number aligned). Heatmaps were plotted using a custom R function. Aggregate plots were generated by averaging rows of the heatmap matrix. ChIP-seq repetitive sequence relationship analysis. Repeat masker was intersected with ChIP-seq peak calls to classify each masker entry as MPP8 bound, MORC2-bound or unbound. Enriched families of repeats were identified with R fisher.test() followed by FDR correction with qvalue(). Distribution of sizes of occupied vs non-occupied L1 was plotted using R density() with sizes being taken from repeat masker. ks.test() was used to reject null hypothesis that distribution of sizes for bound and unbound L1s is the same. To investigate relationship between L1 age, length and occupancy, logistic regression was performed with R glm() engine. Quantitative analysis of H3K9me3 changes was performed by first identifying regions of significant enrichment in each sample relative to corresponding input sample (macs2 callpeak), merging the intervals into a common superset. This superset was joined with a decoy randomized set of intervals, twice the size of actual experimental interval set, with the same size distribution (bedtools shuffle). Next the read coverage was

	determined for each sample (bedtools coverage) and regions with significant change together with fold changes were identified using DESeq2 analysis paradigm. H3K9me3 regions were classified into bound vs unbound by performing intersect with MORC2 and MPP8 ChIP peak calls.
10. Describe the software used to collect and analyze the ChIP-seq data.	For ChIP-seq, we used standard and available softwares: Bowtie, MACS2, Bedtools. Details are provided in the Methods sections. All sequencing samples reported have been deposited at GEO under the accession number: GSE95374. Detailed Data and further code information are available on request from the authors.

nature research

Corresponding author(s): Joanna Wysocka and Michael Bassik

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

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

Data presentation

For all flow cytometry data, confirm that:

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

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

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

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

Methodological details

5. Describe the sample preparation.	Live cells were sorted. No staining involved.
6. Identify the instrument used for data collection.	BD LSR Fortessa
7. Describe the software used to collect and analyze the flow cytometry data.	BD Diva for collection and FlowJo for analysis
8. Describe the abundance of the relevant cell populations within post-sort fractions.	The abundance of transposition positive cells is generally low. ~300,000 gated events were collected for each sample to determine GFP(+) fractions, with target of at least 200 positive cells collected.
9. Describe the gating strategy used.	Cells were gated for live/dead and doublet exclusion using FSC and SSC channels, then cells were gated for presence of mCherry signal (reporting on presence of gRNA). Events passing above gating strategy were classified as positive or negative based on SSC and GFP channel signals.

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