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.

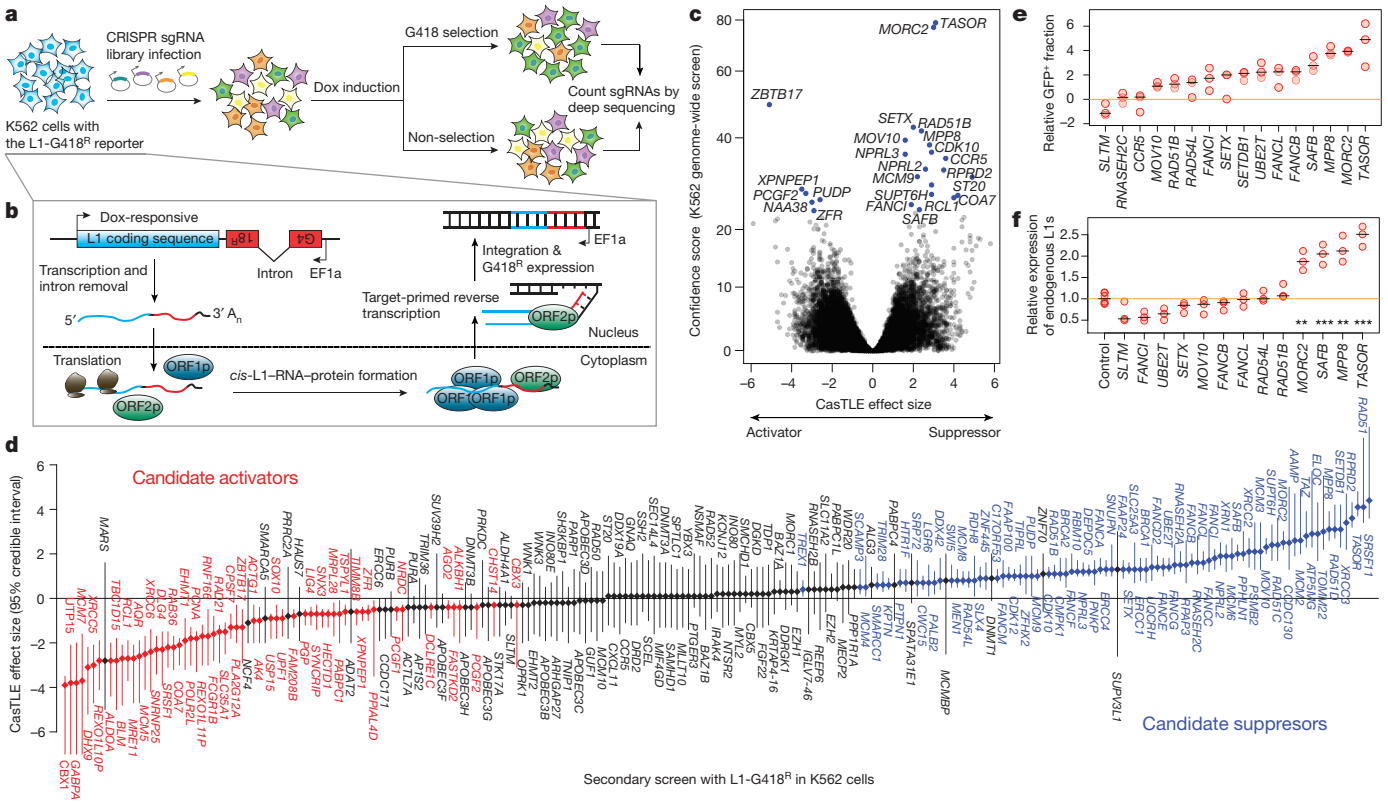

Figure 1 | Genome-wide screen for L1 activators and suppressors in K562 cells. a, Schematic of the screening process. **b**, Schematic of the L1-G418^R retrotransposition. **c**, CasTLE analysis of two independent K562 genome-wide screens. Genes at 10% FDR cutoff coloured in blue, analysed by CasTLE 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