

Analysis of epigenetic features characteristic of L1 loci expressed in human cells

Benjamin Freeman^{1,2}, Travis White⁵, Tiffany Kaul², Emily C. Stow^{1,2}, Melody Baddoo⁴, Nathan Ungerleider⁴, Maria Morales², Hanlin Yang², Dawn Deharo^{1,2}, Prescott Deininger^{2,3} and Victoria P. Belancio^{1,2,*}

¹Department of Structural and Cellular Biology, Tulane University School of Medicine, Tulane Cancer Center, Tulane Center for Aging, New Orleans, LA 70112, USA, ²Tulane Cancer Center, Tulane Health Sciences Center, 1700 Tulane Ave, New Orleans, LA 70112, USA, ³Department of Epidemiology, Tulane School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA, ⁴Department of Pathology, Tulane University School of Medicine, Tulane Cancer Center, New Orleans, LA 70112, USA and ⁵Sloan Kettering Institute for Cancer Research, NY, NY 10065, USA

Received June 24, 2021; Revised December 27, 2021; Editorial Decision December 29, 2021; Accepted January 24, 2022

ABSTRACT

Only a select few L1 loci in the human genome are expressed in any given cell line or organ, likely to minimize damage done to the genome. The epigenetic features and requirements of expressed L1 loci are currently unknown. Using human cells and comprehensive epigenetic analysis of individual expressed and unexpressed L1 loci, we determined that endogenous L1 transcription depends on a combination of epigenetic factors, including open chromatin, activating histone modifications, and hypomethylation at the L1 promoter. We demonstrate that the L1 promoter seems to require interaction with enhancer elements for optimal function. We utilize epigenetic context to predict the expression status of L1Hs loci that are poorly mappable with RNA-Seq. Our analysis identified a population of ‘transitional’ L1 loci that likely have greater potential to be activated during the epigenetic dysregulation seen in tumors and during aging because they are the most responsive to targeted CRISPR-mediated delivery of transactivating domains. We demonstrate that an engineered increase in endogenous L1 mRNA expression increases Alu mobilization. Overall, our findings present the first global and comprehensive analysis of epigenetic status of individual L1 loci based on their expression status and demonstrate the importance of epigenetic context for L1 expression heterogeneity.

INTRODUCTION

Retrotransposons are mobile elements that have colonized the human genome for tens of millions of years (1). These elements increase their copy number in the host genome via a copy-and-paste mechanism involving an RNA intermediate. As a result, retrotransposon sequences make up almost half of the human genome (2). Long Interspersed Element-1 (L1) is the only active, autonomous retrotransposon in humans, accounting for ~17% of the genome (2,3). These elements consist of a 5' UTR containing a promoter sequence, two open reading frames (termed ORF1 and ORF2), and a polyA tail defining the end of their 3' UTR (4,5). ORF1 and ORF2 encode two proteins, ORF1p and ORF2p, that are necessary for L1 retrotransposition (6–9).

In order to create a new copy, a full-length L1 element is first transcribed into L1 mRNA which is then exported to the cytoplasm (5). The L1 mRNA is translated to produce ORF1p and ORF2p, which associate with their parental mRNA in order to form the L1 ribonucleoprotein (RNP) (10–14). ORF1p trimerizes and coats the L1 mRNA, which is likely important for stabilization of the L1 RNP and has a putative role in L1 RNP nuclear import (6,7,15). ORF2p is generated at much lower levels than ORF1p (16,17). It contains endonuclease and reverse transcriptase activities (8,9) needed for the genomic insertion process. After formation in the cytoplasm, the L1 RNP localizes to the nucleus and accesses the chromatin to generate a new copy through target-primed reverse transcription (TPRT), the process in which the L1 mRNA is reverse-transcribed into DNA by ORF2p and integrated into the chromatin (18–20). This copy-and-paste mode of amplification has resulted in the accumulation of over 500,000 L1 insertions in the human genome. Approximately 5,000 of these are full-length L1

*To whom correspondence should be addressed. Tel: +1 504 988 4506; Fax: +1 504 988 1687; Email: vperepe@tulane.edu

elements capable of expression from an L1 promoter. The remaining L1 elements are truncated at the 5' end during insertion (21,22). Without the 5' UTR containing the L1 internal promoter sequence, these truncated L1 elements are not considered active for retrotransposition as they cannot form the L1 mRNA.

Multiple mechanisms contribute to the suppression of L1 retrotransposition (23). The first step of suppression occurs at the transcriptional level, as epigenetic silencing and/or accumulation of mutations in the L1 internal promoter can suppress expression of full-length elements (21,22). The existence of deliberate epigenetic silencing of L1 expression by the host is supported by reactivation of L1 expression associated with engineered or natural loss of functions involved in epigenetic maintenance (24,25) and in pathological states associated with epigenetic remodeling such as cancers and induced cell reprogramming (23,26–29). Furthermore, recent efforts to characterize L1 mRNA expression at single locus resolution have determined that, in any given cell line, most L1 transcripts come from a few loci, including from loci of the LIHs sub-family, the youngest and most active L1 sub-family (21,22,30–32). Studies of endogenous L1 mRNA expression in mouse organs determined that this finding is true *in vivo* because only different small sets of L1 loci are expressed in each organ (33). These data strongly support the potential that genomic environment influences which L1 loci can be expressed.

A number of epigenetic factors have been studied in connection with bulk L1 expression (23,34–36). The biggest weakness of these studies is that they analyze epigenetic marks at promoters of all L1 loci or those belonging to a specific sub-family. This was the result of previous technical limitations preventing analysis of L1 mRNA expression from individual loci.

Several typical mechanisms of transcriptional silencing have been identified in mammalian cells. Most thoroughly studied thus far are the effects of promoter methylation on repetitive elements (36–49). Methylation of CpG islands leads to stable transcriptional silencing by recruiting proteins involved in transcriptional repression and physically blocking transcription factor binding (50,51). CpG islands are overrepresented in promoter regions of the human genome (42) and are a key feature of the L1 promoter (38,39). Studies on L1 methylation highlight the importance of methylation in restricting LINE-1 elements, especially younger and more active sub-families (31). CpG demethylation has been shown to enhance L1 expression (31,36), but hypomethylation is not an absolute requirement for L1 expression (44). Recent studies that characterized promoter methylation of specific expressed loci have focused on only a handful of L1 elements. Scott *et al.* identified a hyperactive LIHs element in somatic cells that contains a CpG mutation likely resulting in its hypomethylation (31). Tubio *et al.* demonstrated that a few active L1 elements, as identified by 3' transductions events, show consistent promoter hypomethylation in a variety of tumor types (29). Sanchez-Luque *et al.* found hypomethylation of a L1 donor element in human neurons (36). Together, these studies suggest that L1 promoter methylation status may play a role in distinguishing expressed and unexpressed L1 loci on a broad scale.

Open chromatin allows for transcriptional machinery, including specific transcriptional factors, to access promoter regions. Several transcription factors, such as YY1 (52,53) and RUNX3 (54,55), have been implicated in L1 regulation either experimentally or via identification of predicted transcription factor binding sites (52–54,56–59). While co-factors involved with chromatin remodeling have been found to bind to L1 elements, a role for chromatin state in L1 transcription has not been studied directly (60,61). Chromatin state is influenced by posttranslational modifications of core histones. Histone tails can be methylated and/or acetylated to regulate activation of nearby promoters by recruiting proteins involved in transcriptional regulation (24). Chromatin Immunoprecipitation (CHIP) experiments have shown that both repressive and activating histone marks associate with L1 promoters, though their direct influence on L1 expression has not been studied because none of the published approaches were guided by locus-specific L1 mRNA expression (62–64). The activating histone marks analyzed in connection to L1 expression facilitate transcription through a variety of functions: H3K27Ac and H3K9Ac have roles in chromatin remodeling and facilitating transcriptional elongation (65–68), while H3K27Ac also associates with active enhancers (69). H3K4Me1-3 histone modifications at promoters are associated with regions of open chromatin (70,71). H4K20Me1 is associated with transcription elongation factors (72). H3K4Me3 has been found to localize nearby promoters of active LIHs-Ta elements in MCF7 cells (32). Bulk analysis of these marks in relation to L1 loci has found that five of these marks, H3K27Ac, H3K9Ac and H3K4Me1-3, are significantly enriched in certain L1 sub-families in the mouse genome (73).

Additional complexity in regulation of expression of L1 loci may arise due to long range interactions between the L1 promoter and enhancers in the general area. Promoter-enhancer associations generally fall within larger DNA loops that are determined by CTCF binding sites (74,75). Alternative DNA loops may occur depending on where CTCF binds, allowing for different loop confirmations and dynamic promoter-enhancer interactions (76). Previous studies have found that CTCF and RNA polymerase II colocalize at the L1 5' promoter (56). It has been postulated for many years that the L1 promoter is relatively weak (77,78), suggesting a possibility that it may be strengthened by an association with nearby enhancers (56). Interactions between enhancers and L1 elements have only been studied indirectly via enhancer-associated histone marks (73). In combination, all of these epigenetic factors have the potential to silence or activate individual L1 loci. However, whether some of them are more or less critical for efficient expression of endogenous L1 loci remains unknown.

Expression of L1 loci containing intact ORFs as well as those without could have biological implications. L1 loci capable of retrotransposition mobilize themselves and other elements. Based on the current knowledge, the biological impact of expression of L1 loci without intact ORFs could play some cellular role either due to expression or incomplete ORFs (79,80). L1 loci containing intact ORF2p have the potential to generate DNA double strand breaks even in the absence of retrotransposition [80,101]. Additionally, L1 sequences contain a bidirectional promoter, splice sites, and

polyadenylation sites, all of which can contribute to generation of chimeric transcripts, interfere with normal gene expression, or drive expression of genomic regions that are otherwise not expressed (81–87).

In this manuscript, we take advantage of our ability to identify individual expressed L1 loci to evaluate epigenetic features distinguishing expressed from transcriptionally silent loci. We focus on L1 loci in MCF7 cells as this cell line has unusually high levels of L1 mRNA expression and the number of expressed L1 loci compared to a much more limited number of expressed L1 loci and low expression levels in other cell lines, such as HeLa and HEK293. We combine a series of studies on epigenetics around expressed vs. unexpressed L1 loci to develop a better understanding of the epigenetic regulation of individual loci expression. Through these efforts, we demonstrate that epigenetic factors can be used for predicting expression of poorly mapped L1Hs loci and that endogenous L1 loci expression can be manipulated in a biologically relevant manner via CRISPR/Cas9 activation and inhibition approaches.

MATERIALS AND METHODS

RNA Sequencing and alignment

Cytoplasmic, polyA selected, strand-specific Illumina RNA-Seq (2 × 100 bp) as previously described (22,88) was used for this study (Beijing Genomics Incorporated). Alignment to the hg19 genome and quantification of reads with BOWTIE, bedtools intersect, and manual curation were performed as previously described (22,88). Briefly, this process involves visual inspection of RNA-Seq BAM files in IGV to confirm that reads mapping to a particular L1 element originate from the L1 promoter and not from an upstream promoter. The annotation for the full-length L1 elements present in the reference genome utilized BLAST to identify loci with the L1 promoter region and that intersected with the REPEATMASKER L1 annotation for L1 elements with at least a 5500 bp concordance (22,88). Key factors in the alignment include a requirement that the reads be concordant with the genome as a pair, as the L1 mRNA involved in retrotransposition is not processed. In addition, the ‘tryhard’ setting is used to force the software to look for all possible matches in the genome prior to determining whether one is uniquely better than all others. MCF7 cell line fastq files were downloaded from: <https://www.encodeproject.org/experiments/ENCSR000CTU/> as fastq files, ENCF000HSC and ENCF000HSC. HEK293 cell line fastq files can be found listed as SRR1275413 from the Gene Expression Omnibus (GEO) traces website. HeLa cell line fastq files are the same files used as the negative control of Figure 10. The unzipped Illumina sequencing fastq file sizes for MCF7, HeLa, and HEK293 are 80.3GB, 40.52GB and 27.54GB, respectively. The total number of mapped reads in MCF7, HeLa and HEK293 BAM files is 30 324 454, 16 581 681 and 7 419 177, respectively. To compensate for these differences, the number of mapped reads for each locus in HeLa and HEK293 were increased by 1.83 and 4.09 times, respectively. We rounded the number of mapped reads for each locus to the nearest whole number and used the adjusted totals for analyses. Expression thresholds were chosen to ensure that

the expressed loci are robustly transcribed, the population of expressed loci in MCF7 cells is large enough for rigorous statistical analyses, and that we could consistently compare our analyses across the three cell lines that were sequenced to various depths. This approach was validated by down-sampling of the MCF7 file which was performed using the Picard DownsampleSam tool. Biological replicates for RNA-Seq data were performed for all three cell lines using either in-house sequencing for MCF7 and HeLa or existing datasets for HEK293-SRR710092.

L1 loci association with ATAC peaks

ATAC library preparation and PE sequencing were done by the Tulane Center for Translational Research in Infection and Inflammation NextGen Sequencing core according to their standard protocols seeding 2 million cells. Sequencing reads were mapped using BOWTIE as described for the RNA-Seq data (22,88) and broad peaks were called using MACS2 (89). L1 loci are considered positive for an ATAC-peak association if an ATAC peak overlaps within 500 bp of the L1 5'UTR start site. Biological replicates for ATAC-Seq data were performed for all three cell lines (MCF7 – SRR9684005; HeLa – SRR8171287; HEK293 – SRR6418071).

L1 loci overlap with genes

Bedtools INTERSECT was used to determine whether each L1 locus overlapped with a gene. Gene coordinates for hg19 were downloaded from the UCSC genome browser (90).

L1 loci association with activating histone marks

Analysis of activating histone marks associating with individual L1 loci was carried out using bed files downloaded from ENCODE (see below) for each of the CHIP-Seq experiments for activating histone marks. Bedtools INTERSECT was used to identify overlaps between the peaks from the CHIP-Seq and the 5' end of the full-length L1 annotations. An individual locus is positive for a certain histone mark if the peak overlaps within 500 bp upstream and 300 bp downstream of the L1 5'UTR start site.

L1 loci association with CpG Methylation

Whole Genome Bisulfite Sequencing (WGBS) using 100 bp paired-end reads was carried out by BGI Global Genomic Services using a DNBseq platform. Bismark-0.20.1 (91) and BOWTIE were used to identify and map methylated/unmethylated CpG islands. Bedtools INTERSECT was used to identify the percentage of methylated CpG islands within the first 500 bp of each L1 locus 5'UTR sequence. Biological replicates for WGBS data were performed for all three cell lines using MCF7 – SRR7707730; HeLa – ENCF0751KHK, ENCF0192ITK; HEK293 – SRR1020524. Total CpGs within the first 500 bp of each L1 5'UTR was assayed with bedtools INTERSECT and a file containing all hg19 CpGs downloaded from: https://figshare.com/articles/All_CpG_sites_for_hg19/1415416/1.

L1 loci promoter motif sites

Promoter motifs were analyzed based on the canonical motif sequences. The YY1 motif 'AGCCAAGATGGC-CGAATAG' and RUNX3 motif 'TGCATTTCCATCT-GAGGTA' used for alignment were derived from the consensus active L1 element (55). All nucleotides in both the YY1 and RUNX3 binding motifs are present in >90% of the L1 loci used to form the consensus. Alignment of loci within the TTC28 gene with the consensus motifs was performed with ClustalW.

L1 association with CTCF loops, Pol II loops, and DHS linkages

All screenshots of CTCF loops, Pol II loops, and DHS linkages are taken from the following website: <http://3dgenome.fsm.northwestern.edu/chiapet.php>. CTCF and Pol II loops were identified with Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) (92). Analysis was performed on 10 L1 loci of each group. Limiting the number to 10 loci allowed us to only look at the expressed loci with the most reads and transitional loci with the strongest ATAC peaks. The remaining loci are at most 40% of the expression of the highest ranking locus in the top 10 expressed L1 loci, and therefore we would expect a gradual decrease in data quality.

Detection of L1Hs loci in MCF7 cells

Paired-end whole genome sequencing of MCF7 cells was obtained from NCBI SRA accession number SRR8652105. The paired alignment files were aligned separately to the human L1 consensus sequence using STAR v2.3.0e and allowing one alignment per read (-outFilterMultimapNmax 1) and a maximum of 25 mismatches (-outFilterMismatchNmax 25). Alignments that occurred in the first 700 bp of the L1 consensus sequence and aligned in the reverse orientation to L1 were extracted. These reads were then used to find their pair based on matching read IDs. The opposite read pair was then aligned to the human genome (hg38) using BOWTIE v0.12.8, requiring unique alignments (-m 1) and allowing three mismatches (-v 3). Alignments in the resulting file were then parsed for read alignments that occurred within the 5' upstream region of annotated L1Hs loci in hg38. This was done using Bedtools v2.22.0. We cross-referenced the loci detected in MCF7 cells using this method with the 4973 L1 elements from the REPEATMASKER annotation (developed by A.F.A. Smit, R. Hubley and P. Green; see <http://www.repeatmasker.org/>) analyzed in this manuscript and identified the 246 L1Hs loci present in MCF7 cells analyzed in Figure 8. Liftover (93) was used to convert genomic coordinates of these L1Hs loci from hg38 to hg19.

Transactivation of L1 mRNA expression in HeLa cells

HeLa cells were cultured in MEM (HyClone) supplemented with 1% sodium pyruvate, L-glutamate, NEAA and 10% FBS as previously described (94). 1.5×10^5 HeLa cells were seeded in each T75 flask 16–18 h prior to transfection. Cells were transfected with 100 ng Cas9m4-VP64 (or

Cas9m4-KRAB) (95,96), 100 ng gRNA (either singly or total of a mix), 100 ng of Renilla luciferase-containing plasmid, and 100 ng of plasmid containing the L1 5' UTR driving Firefly luciferase in the pGL3-basic plasmid (Promega) (83,97). The Renilla luciferase plasmid contains an HSV-tk promoter, a Renilla luciferase reporter, and a polyA signal (pRL-TK, Promega) as a transfection control. Both plasmids and the gRNAs used are illustrated in Figure 9A. Transfection reaction used 12 uL of plus reagent (Invitrogen) and 5 uL of lipofectamine (Invitrogen). All gRNA pools are designed to transfect the same total amount of gRNA plasmid. The exact gRNA sequences used can be found in Additional File 2. The plasmid containing the gRNA targeting AAV (gRNA_AAVS1-T2) was purchased from Addgene (98).

Alu retrotransposition driven by endogenous L1s

For the Alu retrotransposition assay, experiments were carried out as previously described (80). 1×10^6 HeLa cells were seeded in T75 flasks 16–18 h prior to transfection. Briefly, 1 ug of the Alu reporter plasmid expressing reverse-complementary neo resistance tag (99) was co-transfected with 1 ug plasmid expressing Cas9-VP64 (or Cas9-KRAB) and 1 ug plasmid expressing various gRNAs. After two weeks of G418-selection, colonies were stained and counted. As a control to assess any potential toxicity of our transfection protocols, a plasmid expressing the neo resistance tag (1 ug) was co-transfected with plasmid expressing Cas9-VP64 (or Cas9-KRAB) (1 ug) and a plasmid expressing the gRNA (1 ug). After two weeks of G418-selection, colonies were stained and counted. Transfection reaction used 12 uL of plus reagent (Invitrogen) and 5 uL of lipofectamine (Invitrogen). Workflow for both assays is shown in Supplemental Figure S8B.

RNA-Seq following stimulation of endogenous L1 mRNA expression

HeLa cells were cultured as previously described in MEM (HyClone) supplemented with 1% sodium pyruvate, L-glutamate, NEAA and 10% FBS as previously described (94). 1×10^5 HeLa cells were seeded in T75 flasks 16–18 h prior to transfection. Cells were co-transfected with 1 ug Cas9-VP64 and 1 ug of pool 2 gRNA expression plasmids using 12 uL of plus reagent (Invitrogen) and 5 uL of lipofectamine (Invitrogen). Cytoplasmic RNA was isolated, polyA-selected and used for PE sequencing as described previously (82). RNA-Seq reads were mapped to Hg19 using BOWTIE and manually curated for authentic expression as described (22,88). The total number of mapped reads from RNA-Seq BAM files in the pool 2 and control HeLa cells is 17 927 082 and 16 581 681, respectively. Biological replicates for pool 2 and control RNA-Seq data were performed.

Identification L1 elements with intact ORF2

L1 loci with intact ORF2 were determined based on database information of 'Human Full-Length, Intact LINE-1 Elements' and 'Human ORF2 Intact LINE-1 Elements' from L1Base 2 (100)

RESULTS

Locus-specific L1 mRNA expression in MCF7, HeLa and HEK293 cells

Utilizing our RNA-Seq method that involves ‘unique’ alignment of reads combined with manual curation to validate the nature of the transcript (22,88) we find that MCF7 cells have much higher levels of expressed L1 loci than other cells, including HeLa and HEK293 (Figure 1A). Using 20 mapped reads/L1 locus in MCF7 as a cutoff to identify the most robustly expressed L1 loci, we detected 162 of the 4973 full-length L1 loci as being expressed (Additional File, page 1). There were an additional 1,783 loci that either were rejected as not being generated from the L1 promoter during manual validation or whose expression was ambiguous because it was below the 20 read threshold (Figure 1B). Over 3,000 full-length L1 loci had no reads mapping in MCF7 and will be referred to as unexpressed. Downsampling of the MCF7 BAM files to match the total number of mapped reads in HeLa BAM files identified 159 of the 162 loci that are expressed in MCF7 and meeting the expression threshold used for HeLa. Downsampling of the MCF7 BAM files to match the total number of mapped reads in HEK293 BAM files identified 138 loci that are expressed in MCF7 cells meeting the expression threshold used for HEK293. Comparing normalized L1 mRNA expression (see Methods) in HeLa and HEK293 data to MCF7 showed significantly fewer expressed L1 loci in HeLa and HEK293 compared to MCF7 (HeLa: 27 loci, $P < 0.0001$; HEK293: 42 loci, $P < 0.0001$; Chi Square analysis) (Figure 1A). The average levels of reads mapped to a locus expressed in MCF7 (71.2 reads, $n = 162$ loci) was significantly higher than in HeLa (29.8 reads, $P = 0.005$, Student’s *t*-test) and near-significant in HEK293 (41.4 reads, $P = 0.06$, Student’s *t*-test), although the HeLa results are partially a reflection of the lower threshold of mapped reads for expression (see Methods). We then determined the expression status of the 162 MCF7-expressed loci in HeLa and HEK293 cells, finding only 16 and 19 of these loci to be expressed in HeLa and HEK293, respectively (Figure 1C). Because of the high number of uniquely expressed loci in MCF7 cells relative to HeLa and HEK293, we proceeded to focus our analysis of epigenetic features of expressed L1 loci on the 162 loci expressed in MCF7 and use L1 loci expressed in both HeLa and HEK293 cell lines for comparative analysis. We also included analysis of total L1 loci in each cell line in our studies because it represents results obtained for bulk analysis of L1 promoters. Evaluation of the location of the 162 L1 loci relative to known genes showed that there was a modest but significant enrichment for expressed compared to unexpressed L1 loci to be within genes ($P < 0.05$, Chi square analysis) (Supplemental Figure S1). This may indicate that genes provide a potentially advantageous chromatin architecture for L1 mRNA expression.

To assess the known L1Hs elements with mutagenic potential, we determined that of the 162 loci identified to be expressed in MCF7 cells, 2 loci contain intact ORF2, both of which are L1Hs elements and includes L1-5219 (Additional File, page 1). L1 loci containing intact ORF2p have the potential to generate DNA double strand breaks even in the absence of retrotransposition (80,101). We also an-

alyzed 38 L1 elements previously assessed for their retrotransposition potential (as identified by Rodriguez-Martin *et al.* (102)) and found three loci had 14–26 mapped reads and 4 more had only 1–6 reads and are therefore potentially background level of mapping in MCF7 (Additional File, page 4). There were also similar number of loci from these 38 loci expressed in HEK293 and HeLa cells.

Promoters of most expressed L1 loci overlap with an ATAC peak

Chromatin accessibility measured by ATAC-Seq (103) identified peaks at 89,503, 67,953 and 19,216 locations throughout the genomes of MCF7, HeLa and HEK293, respectively. The majority of L1 loci expressed in MCF7 had an ATAC peak overlapping with their promoter (134/162 L1 loci, 83%), while only 280 out of 3,028 unexpressed L1 loci (9%) had ATAC peaks ($P < 0.0001$, Chi square analysis) (Figure 2). Some of these 280 loci with ATAC peaks but no L1 mRNA expression may represent a transitional state that has either lost or not yet gained a key component required for detectable expression. We will refer to these L1 loci as ‘transitional’ loci. The L1 loci with mapped reads that primarily originated from a non-L1 promoter showed only 14 of 50 loci (28%) having an ATAC peak, a significant decrease in ATAC peak association as compared to the 162 expressed loci ($P < 0.0001$, Chi square analysis). It is possible that some of these loci express at a low level that is indistinguishable from the surrounding passive transcription. Analyses of ATAC peaks in HeLa and HEK293 show a significant increase in ATAC peaks detected at expressed versus unexpressed L1 loci in both HeLa (37.04% versus 13.71% of loci, $P < 0.01$, Chi square analysis) and HEK293 (28.57% versus 1.94%, $P < 0.0001$, Chi square analysis) (Figure 2). Thus, in all three cell lines, MCF7, HeLa, and HEK293, ATAC peaks are significantly more likely to overlap with expressed as opposed to unexpressed loci. When we evaluate the presence of ATAC peaks at all L1 elements regardless of their expression status, we see many shared loci, as well as cell-line specific loci (Supplemental Figure S2). Only 49 out of the 1,325 loci having an ATAC peak in at least one cell line were found in common between all three cell lines. These findings are generally consistent with the cell-line specificity of L1 expression (Figure 1).

Activating histone marks are enriched at promoters of expressed L1 loci

The open chromatin that is associated with transcription generally contains specific histone modifications (104,105). We analyzed expressed, unexpressed, and total L1 loci for the presence of activating histone marks H3K27Ac, H3K9Ac, H3K4Me3, H3K4Me2, H3K4Me1 and H4K20Me1 using ENCODE CHIP-Seq data sets performed in duplicate using DNA/protein extracts from MCF7, HeLa and HEK293 cells (see Materials and Methods). Only data concerning H3K27Ac, H3K4Me3 and H3K4Me1 status were available for HEK293 cells. This analysis determined that five of the six histone marks, H3K27Ac, H3K9Ac, H3K4Me3,

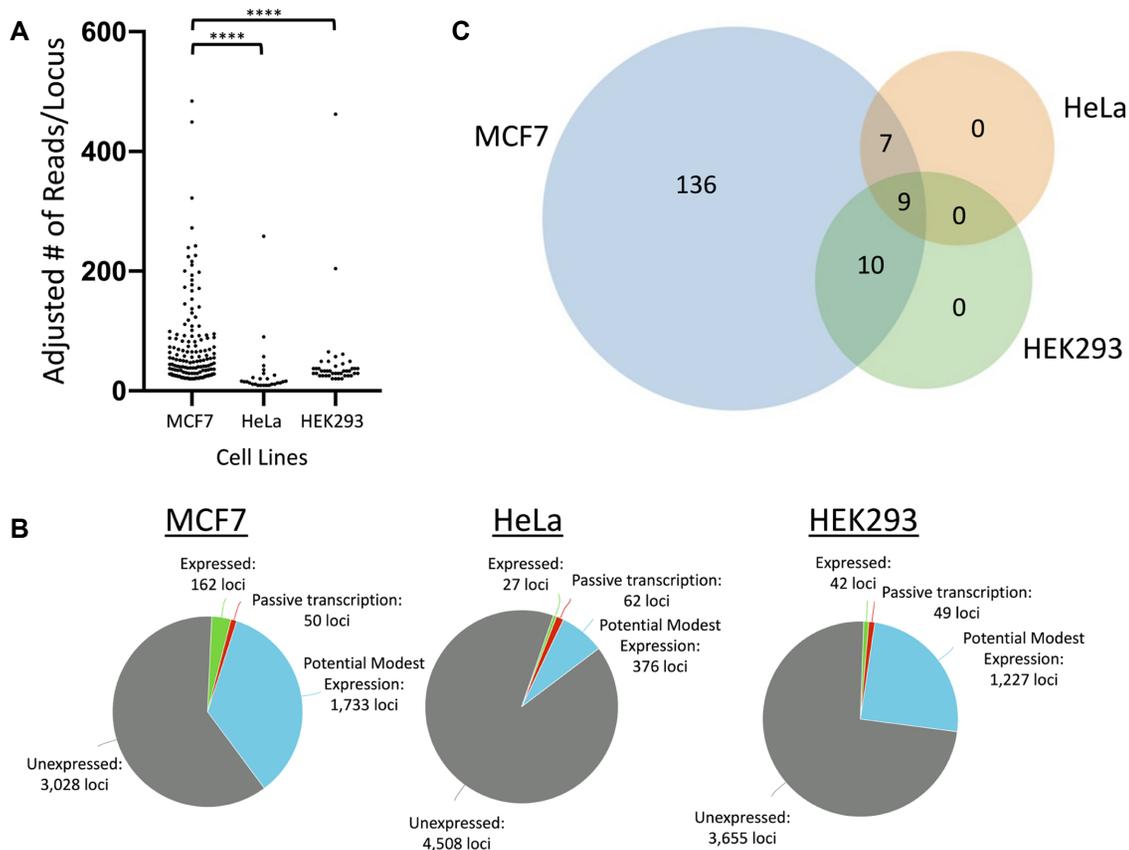


Figure 1. L1 mRNA expression in MCF7, HeLa, and HEK293 cells. (A) Reads corresponding to expressed L1 loci as determined by manual curation of PE stranded RNA-Seq are plotted for three different cell lines. The number of mapped reads for each locus in HeLa and HEK293 is adjusted to normalize the differences in total number of mapped reads from each RNA-Seq file (see Materials and Methods). L1 loci with 20 or greater mapped reads (20+) are shown for MCF7 ($n = 162$ loci) and HEK293 ($n = 42$ loci) cells. Too few loci met this threshold in HeLa cells ($n = 10$), so a threshold of 9+ mapped reads was used instead ($n = 27$ loci) to establish a larger cohort of expressed loci for further analysis. Chi square analysis with Yates' correction was used to assess the significance of the number of expressed loci with 20+ (9+ in HeLa) mapped reads as a proportion of all L1 loci between cell lines (**** $P < 0.0001$). (B) Expression status of L1 loci in MCF7, HeLa, and HEK293 cells as identified by stranded PE RNA-Seq followed by manual validation of resulting mapped reads. Categories of L1 loci for each cell line are: 'Expressed' are loci with 20+ mapped reads (9+ in HeLa) and passed manual curation, as described in (A), 'Passive transcription' are loci that failed manual curation due to mapped reads that may not have come from the L1 promoter, 'Potential modest expression' are loci with 1–19 mapped reads (1–8 in HeLa) that did not undergo manual curation, and 'Unexpressed' are loci with 0 mapped reads. (C) Expression status of loci expressed in MCF7 ($n = 162$ loci) in MCF7 (blue), HeLa (orange) and HEK293 (green) cells.

H3K4Me2 and H3K4Me1, are significantly enriched ($P < 0.0001$, Chi square analysis) at promoters of L1 loci expressed in MCF7 cells as compared to unexpressed loci (Figure 3A). Of the six activating histone marks analyzed, the marks most commonly present at expressed L1 loci are H3K4Me3 (78.09% of loci) and H3K4Me2 (80.25%). We also analyzed activating histone marks at promoters of expressed, unexpressed, and total L1 loci in HeLa and HEK293 cells. Similar to findings in MCF7, there was a significant difference in the presence of the H3K27Ac, H3K9Ac, H3K4Me3, H3K4Me2 and H3K4Me1 histone marks at the promoters of expressed versus unexpressed loci in HeLa cells ($P < 0.0001$, Chi Square analysis) (Supplemental Figure S3A). In HEK293 cells, a significant difference was seen for all three histone marks analyzed (H3K27Ac, H3K4Me3 and H3K4Me1) between expressed versus unexpressed loci (Supplemental Figure S3B). A significant difference was also seen when comparing the number of activating histone marks per locus of expressed and unexpressed loci in MCF7 (5.91 marks versus 0.21

marks, $P < 0.0001$, Student's t -test) (Figure 3B), HeLa (2.85 marks versus 0.19 marks, $P < 0.0001$, Student's t -test), and HEK293 cells (0.9 marks versus 0.05 marks, $P < 0.0001$, Student's t -test) (Supplemental Figure S3C). In addition, HeLa cells have fewer overall of these activating marks per locus regardless of its expression status than do MCF7 cells (Supplemental Figure S3D, left). When analyzing only the H3K27Ac, H3K4Me3 and H3K4Me1 histone marks, both HeLa and HEK293 cells contained fewer overall of these activating marks per L1 locus regardless of its expression status than did MCF7 (Supplemental Figure S3D, right). These findings are consistent with the relatively high level of L1 mRNA expression detected in MCF7 as opposed to HeLa and HEK293 cells (Figure 1A). This is supported by the finding that in MCF7 cells there was a correlation between L1 mRNA expression levels and the number of activating histone marks that were detected at individual expressed L1 loci ($P = 0.017$, $R^2 = 0.035$, Pearson correlation test) (Supplemental Figure S3E).

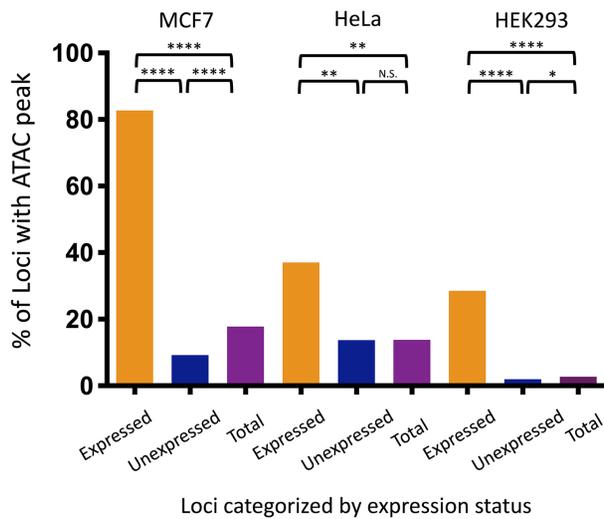


Figure 2. Analysis of ATAC peaks at expressed and unexpressed L1 loci. 82.72% of the 162 L1 loci expressed in MCF7 cells have ATAC peaks around their promoters as compared to 9.25% of the 3,028 unexpressed loci and 17.76% of the 4,973 total loci. 37.04% of the 27 L1 loci expressed in HeLa cells have ATAC peaks as compared to 13.71% of the 4,508 unexpressed loci and 13.75% of the 4,973 total loci. 28.57% of the 42 L1 loci expressed in HEK2993 cells have ATAC peaks as compared to 1.94% of the 3,655 unexpressed loci and 2.67% of the 4973 total loci. Chi square analysis with Yates' correction was used to determine significance (* $P < 0.05$; ** $P < 0.01$; **** $P < 0.0001$).

In addition to the significantly higher levels of activating histone marks associated with expressed versus unexpressed L1 loci in MCF7 cells (Figure 3A, B), we observe a significant difference between the number of activating marks per expressed L1 locus that have an ATAC peak versus the number of activating marks per expressed L1 locus without an ATAC peak (Figure 3B). There is also a subset of unexpressed L1 loci that have an ATAC peak and significantly higher number of activating histone marks than the unexpressed L1 loci with no ATAC. These are the loci that we previously designated as 'transitional' (Figure 2). The lower number of activating histone marks at these loci would be consistent with an incompletely assembled transcriptional apparatus.

CpG methylation of L1 promoters is associated with lower L1 mRNA expression

We performed bisulfite sequencing using genomic DNA extracted from MCF7, HeLa, and HEK2993 cells to use NGS sequencing to determine the methylation status of all CpGs in these three cell lines and specifically for L1 loci. L1 loci expressed in MCF7 cells have a lower percentage of methylated CpGs mapping to the L1 promoter region (36.26% of methylated CpGs) compared to either unexpressed loci (60.82%) or total loci (60.01%) ($P < 0.0001$, Student's t -test) (Figure 4). However, we found no significant difference between methylation at expressed versus unexpressed loci in HeLa (42.85% versus 45.99%, $P = 0.55$, Student's t -test) and HEK2993 cells (41.3% versus 45.14%, $P = 0.37$, Student's t -test) (Supplemental Figure S4A). The significant hypomethylation status of expressed L1 loci in MCF7 cells,

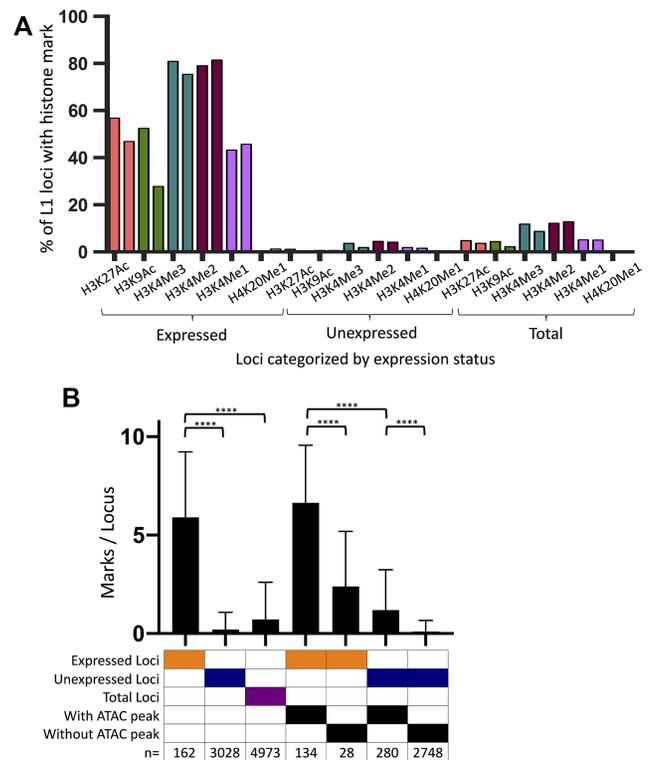


Figure 3. Activating histone marks present in the L1 promoter distinguish expressed from unexpressed L1 loci. (A) Percent of L1 loci containing peaks for each activating histone mark found in promoters of expressed ($n = 162$), unexpressed ($n = 3,028$), and total loci ($n = 4,973$) in MCF7 cells. Two experiments are shown for each histone mark. The percent of H3K27Ac, H3K9Ac, H3K4Me3, H3K4Me2 and H3K4Me1 histone marks present at expressed L1 loci is significantly higher compared to their percent at unexpressed and total L1 loci (Chi square analysis with Yates' correction, $P < 0.0001$). (B) L1 loci with ATAC peaks have more activating histone marks than L1 loci without ATAC peaks in MCF7 cells. The average number of activating histone marks is calculated out of 12 marks (the 6 marks analyzed in two experiments) for expressed loci ($n = 162$), unexpressed loci ($n = 3,028$), total loci ($n = 4,973$), expressed loci with an ATAC peak ($n = 134$), expressed loci without an ATAC peak ($n = 28$), unexpressed loci with an ATAC peak ($n = 280$), unexpressed loci without an ATAC peak ($n = 2,748$). Significance between different groups of L1 loci was determined by Student's t -test (**** $p < 0.0001$).

compared to HeLa and HEK2993, contradicts the higher levels of overall methylation within the MCF7 genome compared to the other cells (Supplemental Figure S4B) as well as higher levels of L1 loci methylation in MCF7 cells when all L1 loci are considered as they are in bulk analysis of L1 methylation (Supplemental Figure S4C). Additionally, in MCF7 cells, loci with an ATAC peak have significantly lower levels of CpG methylation than loci of the same expression status without an ATAC peak (Figure 4). Of particular note is the observation that the 'transitional' loci (unexpressed loci with an ATAC peak) showed significantly lower methylation levels than unexpressed L1 loci without an ATAC peak (Figure 4). Thus, these transitional loci have both lower promoter methylation (Figure 4) and a higher number of activating marks than unexpressed L1 loci without ATAC peaks (Figure 3).

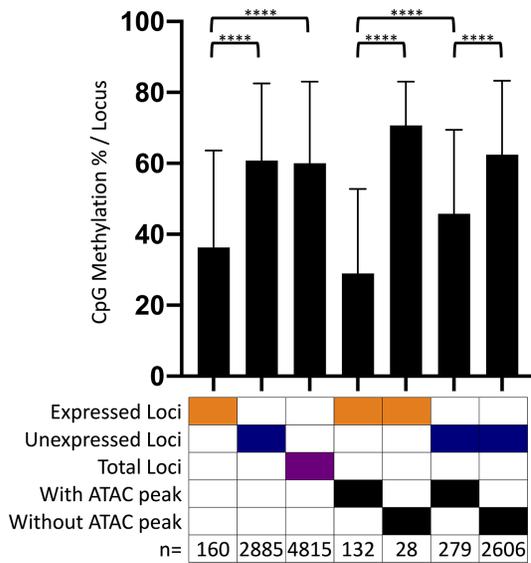


Figure 4. Promoters of L1 loci expressed in MCF7 cells are hypomethylated. Analysis of CpG methylation at promoters of expressed loci ($n = 160$, orange), unexpressed loci ($n = 2,885$, blue), total loci ($n = 4,815$, purple), expressed loci with an ATAC peak ($n = 132$), expressed loci without an ATAC peak ($n = 28$), unexpressed loci with an ATAC peak ($n = 279$), and unexpressed loci without an ATAC peak ($n = 2,606$). A total of 158 loci were unmapped following bisulfite sequencing and thus excluded from this analysis. Significance between different groups of L1 loci was determined by Student's *t*-test (**** $P < 0.0001$).

Analysis of the epigenetic status of L1 loci expressed in MCF7 cells in HeLa and HEK293 cells

Due to the significantly different expression patterns between MCF7, HeLa and HEK293 cells (Figure 1), we hypothesized that there should be associated epigenetic changes at loci exhibiting cell-line specific expression. We used the 162 loci robustly expressed in MCF7 cells as the standard and compared whether there were any reads mapped to the same loci in HeLa or HEK293 cells (Figure 5, heat map, grey). Of these 162 loci expressed in MCF7 cells, only 41 and 68 loci had any reads mapped to them in HeLa and HEK293, respectively, with a total of 885 and 1,580 reads mapped to these 162 loci in each cell line compared to 11,534 mapped reads in MCF7 (Figure 5). We found that five of the six activating histone marks analyzed in HeLa cells (H3K27Ac, H3K9Ac, H3K4Me3, H3K4Me2 and H3K4Me1) and all three marks analyzed in HEK293 (H3K27Ac, H3K4Me3 and H3K4Me1) were able to distinguish between the loci expressed in MCF7 relative to those unexpressed in MCF7 when analyzed in HeLa and HEK293, respectively (Supplemental Figure S5A, B). The percent of loci containing these marks in HeLa and HEK293 cells is significantly lower than in MCF7 cells, consistent with the finding that only some of the 162 L1 loci expressed in MCF7 cells are also expressed in HeLa and HEK293 cells (Figure 5). In HEK293 cells, methylation is significantly higher in expressed loci compared to unexpressed loci (Supplemental Figure S5C), the opposite of what was observed in MCF7 cells (Figure 4), while no significant difference is seen in HeLa cells (Supplemental Figure S5C). Similar to the consistency seen with activating histone marks, as many

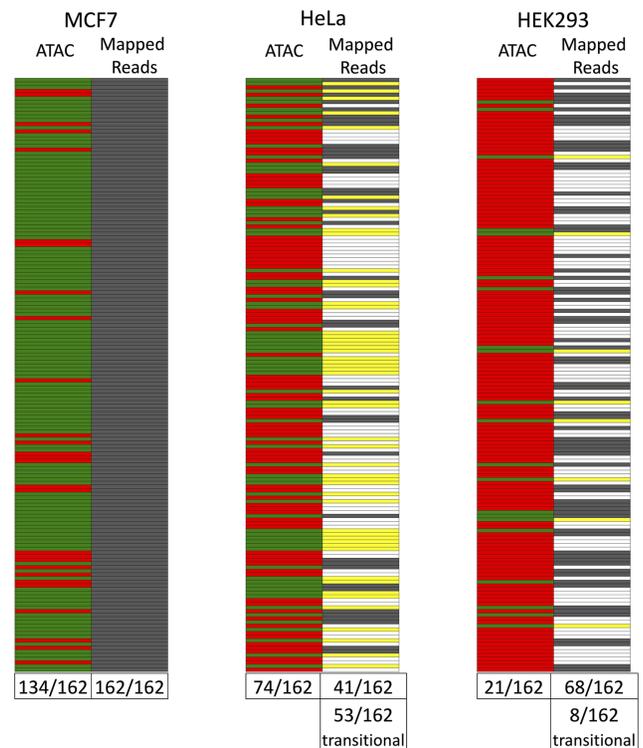


Figure 5. ATAC and expression status of the 162 L1 loci that are expressed in MCF7 cells within HeLa and HEK293 cells. Each locus expressed in MCF7 cells ($n = 162$) is evaluated for their ATAC status in MCF7 cells, and for ATAC and presence of mapped reads (1+ mapped read) status in HeLa and HEK293 cells. Presence of an ATAC peak (green) or with an absence of an ATAC peak (red) and presence (grey) or absence (white) of mapped reads at each locus is shown for each cell line. 'Transitional' loci (unexpressed loci with an ATAC peak) are marked in yellow. Numbers below each respective column indicate the number of loci out of the 162 loci evaluated that have an ATAC peak or have mapped. The number of loci categorized as 'transitional' is listed at the bottom of the 'Mapped Reads' column for HeLa and HEK293. Loci are shown in the same order in each cell line and ordered from highest (top) to lowest (bottom) number of mapped RNA-Seq reads as determined in MCF7 cells.

as 69 out of the 162 loci have shared ATAC peaks between MCF7 and HeLa, with 15 loci in common to all three cell lines (Figure 5, green; Supplemental Figure S5D). Many of the shared L1 loci with ATAC peaks fit into our definition of 'transitional' elements that have ATAC peaks showing some open chromatin but no mapped reads (Figure 5, yellow).

Long-distance genomic interactions contribute to expression of L1 elements

Three dimensional (3D) chromatin interactions can strongly influence transcriptional state and regulation (106). We utilized ChIA-Pet conformation studies in MCF7 to determine whether 3D interactions may explain why some L1 elements are expressed and others are not. The CHIA-PET data (107) analyzed 3D interactions involving RNA polymerase II (pol II) binding that was associated presumably with both the promoter and an enhancer, while also analyzing CTCF binding to determine whether cohesin/CTCF loops might establish Topologi-

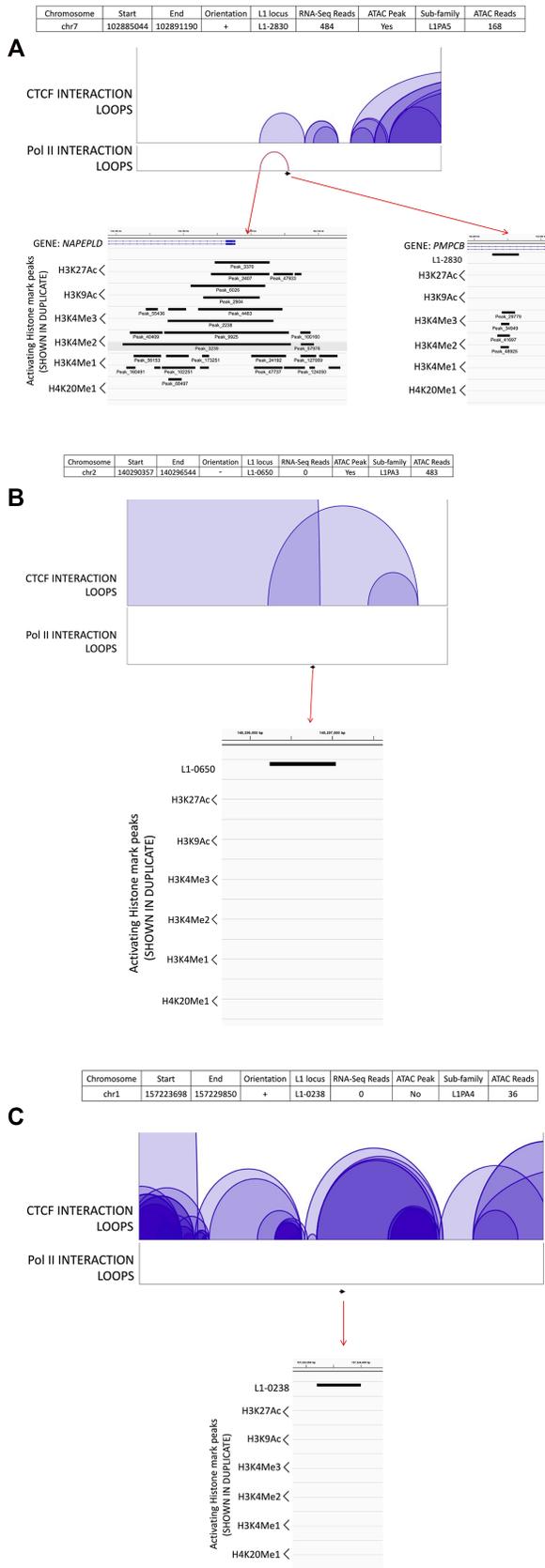


Figure 6. Long-distance interactions of specific L1-loci in MCF7 cells. Screenshots of Integrated Genomics Viewer of L1-2830 (A), L1-0650

cally Associated Domains (TADs) (108,109) that defined which elements are able to interact with an enhancer. We compared genomic interactions of the ten highest expressed loci in MCF7 (Table 1 and Figure 6A, Supplemental Figure S6A–I), 10 transitional loci (unexpressed loci overlapping with an ATAC peak) (Table 1 and Figure 6B, Supplemental Figure S6J–R), and ten randomly chosen unexpressed loci (Figure 6C, Supplemental Figure S6S–AA). We show that nine out of ten expressed loci have their promoter closely associated with a distant sequence via RNA pol II binding (Table 1 and Figure 6A, Supplemental Figure S6A–H). This would be consistent with the L1 promoter interacting with an enhancer (Table 1). We also see that seven of these promoter/enhancer interactions are present within a CTCF loop (Figure 6A, Supplemental Figure S6A–C, E, G, H). Analysis of the ‘transitional’ L1 loci shows that only four of these loci associate with an RNA pol II loop and that all four of these loci are within a CTCF loop (Table 1, Supplemental Figure S6J–L, Q). In contrast, of the ten unexpressed loci (Figure 6C, Supplemental Figure S6S–AA), only one locus (Supplemental Figure S6Y) shows an interaction loop with the RNA pol II.

Analysis of active histone marks shows that seven of the ten expressed L1 loci discussed above are associated with enhancer-like genomic regions that contain high levels of activating histone marks (Table 1, Figure 6A, Supplemental Figure S6B–D,F–H). All but one of these loci have significant levels of activating histone marks on their promoter as well (Table 1, Figure 6A, Supplemental Figure S6B,C,F–H). Thus, all of the expressed loci have significant levels of activating histone marks associated either through their promoter or their enhancer and most have both (Figure 6A, Supplemental Figure S6A–I). Of the ten transitional loci analyzed, only three associate with a potential enhancer containing histone marks through RNA pol II (Table 1, Supplemental Figure S6J–L) and generally have much lower levels of histone activation on both their promoter and enhancer regions (Expressed: 7.1 and 7.6 marks, respectively; Transitional: 3.3 and 2.5 marks, respectively) (Table 1). These loci contrast with the ten random, unexpressed loci where only one has a putative RNA pol II-mediated loop (Supplemental Figure S6Y), but that enhancer and none of the promoters are associated with activating histone marks (Table 1, Figure 6C, Supplemental Figure S6S–AA).

(B) and L1-0238 (C) in MCF7 cells. L1-2830 is an expressed locus with the most mapped RNA-Seq reads in MCF7. L1-0650 is a ‘transitional’ locus, i.e. unexpressed locus with the second most ATAC-sequencing reads mapped. L1-0238 is a randomly-selected unexpressed locus. Information regarding the locus location (chromosome, start site, end site and orientation), RNA-Seq reads, presence of an ATAC peak, L1 sub-family, and number of reads from ATAC sequencing is shown in a table format at the top. Loops indicate CTCF binding sites (purple, shaded) within 500 kb of the L1 start site (black arrow). RNA polymerase II (Pol II) loops (red) are only shown if the pol II binding site overlaps within 500 bp of the L1 start site. Red arrows indicate magnification of the indicated genomic region that includes analysis of the activating histone marks from previously described CHIP-Seq data (Figure 3A). The status of each histone mark is examined in two experiments.

Selective expression of the L1-5219 locus located within the *TTC28* gene

Analysis of L1 loci within the *TTC28* gene allowed exploration of reasons underlying differential expression of L1 loci that are present near one another in the human genome. This gene is located on chromosome 22 and harbors seven intronic, full-length L1 elements (Figure 7A). One locus, L1-5219, has been previously characterized as an unusually active locus that is expressed in multiple cancers (21,22,29,30,32,64,110). L1-5219 shows consistent mRNA expression across multiple cancer cell lines with 26 mapped reads in MCF7 cells, 4 mapped reads in HeLa cells, and 49 mapped reads in HEK293 cells, despite the very low mappability of this element. The higher expression of L1-5219 as compared to the other 6 loci within *TTC28*, which all show little or no authentic expression, makes them a suitable model for identification of factors affecting their expression.

L1-5219 is the only member of the L1Hs-Ta sub-family, the youngest and most retrotranspositionally active L1 subfamily in this gene (Figure 7A). Analysis of the promoter sequences of all seven loci showed that one locus, L1-5215, has a deleted YY1 motif and only two loci, L1-5212 and L1-5219, contains the consensus relative to the motif from the consensus active L1 sequence (55). Analysis of the RUNX3 binding motif determined that two L1 loci, L1-5219 and L1-5217, have the consensus sequence, while the other five contain one mutation in this site (Figure 7A) (55). The impact that these mutations may have on a specific L1 locus expression is unknown.

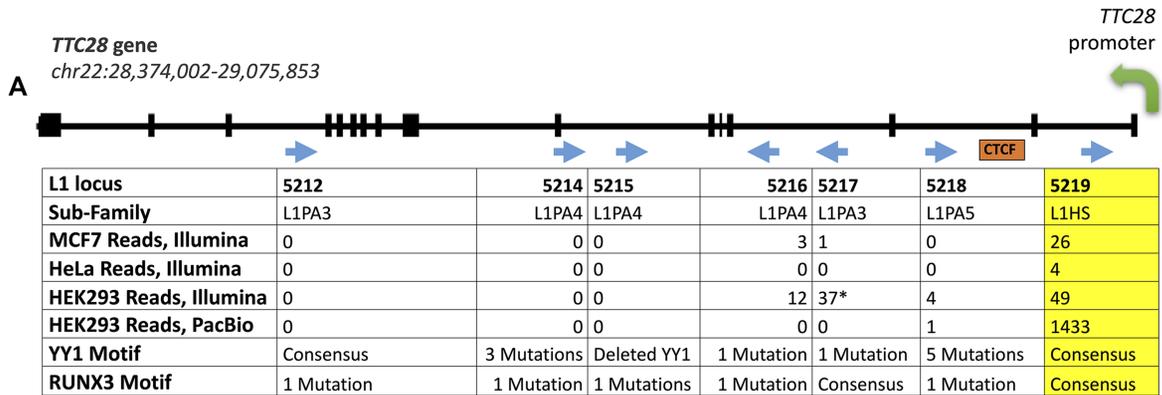
Chromosomal locations (hg19) and analysis of the seven L1 loci in the *TTC28* gene for chromatin accessibility, activating histone marks, and percentage of methylated CpGs (Supplemental Table S1) determined that the promoter region of L1-5219 associates with an ATAC peak in all three cell lines (MCF7, HeLa, and HEK 293) while the other loci do not overlap with an ATAC peak in any cell line. Additionally, despite having the highest number of CpGs (31) in its promoter sequence, L1-5219 has the lowest percentage of methylated CpGs in MCF7 (0%), HeLa (14.29%), and HEK293 cells (18.33%). L1-5219 has a variable percentage of activating histone marks: 33.33% in HEK293, 75% in HeLa and 75% in MCF7 cells. Some of the lack of histone activation mark detection may represent poor mappability in the vicinity of the L1-5219 promoter.

3D interactions in the *TTC28* region using CHIA-PET data show direct RNA pol II interactions only with L1-5219 promoter that connects this locus with a putative enhancer (Figure 7B). The associated enhancer is approximately 75 kb away, between two other genes. This RNA pol II loop is fully within a CTCF loop mediated by a CTCF binding site located within the intronic region of *TTC28* 62 kb upstream of L1-5219. None of the other *TTC28* L1 elements fall within the same loop, possibly insulating them from the influence of the enhancer. Additionally, numerous DNase I hypersensitive site (DHS) linkages, a marker for open chromatin, overlap specifically with L1-5219.

Epigenetic analysis of L1Hs loci may help identify actively expressed unmappable or poorly mappable L1 elements

A major shortcoming associated with mapping RNA-Seq reads uniquely to L1 loci is that many repetitive elements are often indistinguishable from one another (22,32,111). Furthermore, their expression may not be clearly established when these loci are located in the same orientation as an expressed gene (22). These almost identical L1 loci tend to be the youngest L1 elements, L1Hs, that have the strongest potential for retrotransposition (21). In an effort to determine which of these unmappable L1 loci are most likely expressed by using the knowledge gained from our epigenetic analysis of mappable L1 loci, we analyzed the epigenetic features of 246 L1Hs loci present in the MCF7 genome that are also annotated in the human genome. We determined that the L1 loci associating with an ATAC peak ($n = 18$) have a significantly higher number of mapped reads that could map uniquely than loci without an ATAC peak ($n = 228$) (5.1 reads versus 0.5 reads, $P < 0.05$, Student's *t*-test) (Figure 8A). Additionally, as some L1 loci generate longer transcripts resulting from a read-through of the poly-A tail into a more mappable downstream region (112), we also found a significant difference in the number of mapped reads per locus between loci with and without an ATAC peak when including reads mapped within the 2 kb downstream of each L1 locus (14.4 reads versus 1.7 reads, $P < 0.0001$, Student's *t*-test). We also observed significant differences between loci with and without an ATAC peak when comparing both number of activating histone marks (4.5 marks versus 0.3 marks, $P < 0.0001$) and percent of methylated CpGs (44.7% versus 68.4%, $P < 0.0001$) (Figure 8B and C). Of the 18 L1Hs loci that overlap with an ATAC peak in MCF7 cells, three loci are expressed (above our threshold level), three loci have only 1–19 mapped reads and are therefore likely to be expressed at some level, and 12 loci have 0 mapped reads within the L1, although six of these have a few reads mapping downstream that may suggest some level of transcription with readthrough at the 3' end. The six loci with no signs at all of transcription are consistent with the 'transitional' definition. By comparing the epigenetic context of individual loci, we are able to identify a subset of poorly mappable L1Hs L1 loci that are likely to be expressed in MCF7 cells based on favorable epigenetic conditions.

L1-0521 is an example of a poorly mappable L1 locus that is likely expressed in MCF7 cells based on its epigenetic features (Supplemental Figure S7A). This locus has only 3 mapped reads (11 when including 2 kb downstream) due to its sequence being similar to the L1 Hs consensus. Based on the few number of reads, this locus cannot be unambiguously identified as expressed during manual curation. However, we are able to identify this locus as likely expressed due to its positive association with an ATAC peak, high level of activating histone marks (10/12 marks), and low percentage of methylated CpGs (1.33%). Additionally, this locus is located within a CTCF loop and directly associates with 5 pol II loops (Supplemental Figure S7B). Although none of these pol II loops connect to an enhancer region containing a high level of activating histone marks, the L1 locus contains many activating histone marks (Supplemental Figure S7B). These results demonstrate that we are able to use the



*These reads were manually determined not to primarily originate from an L1 promoter

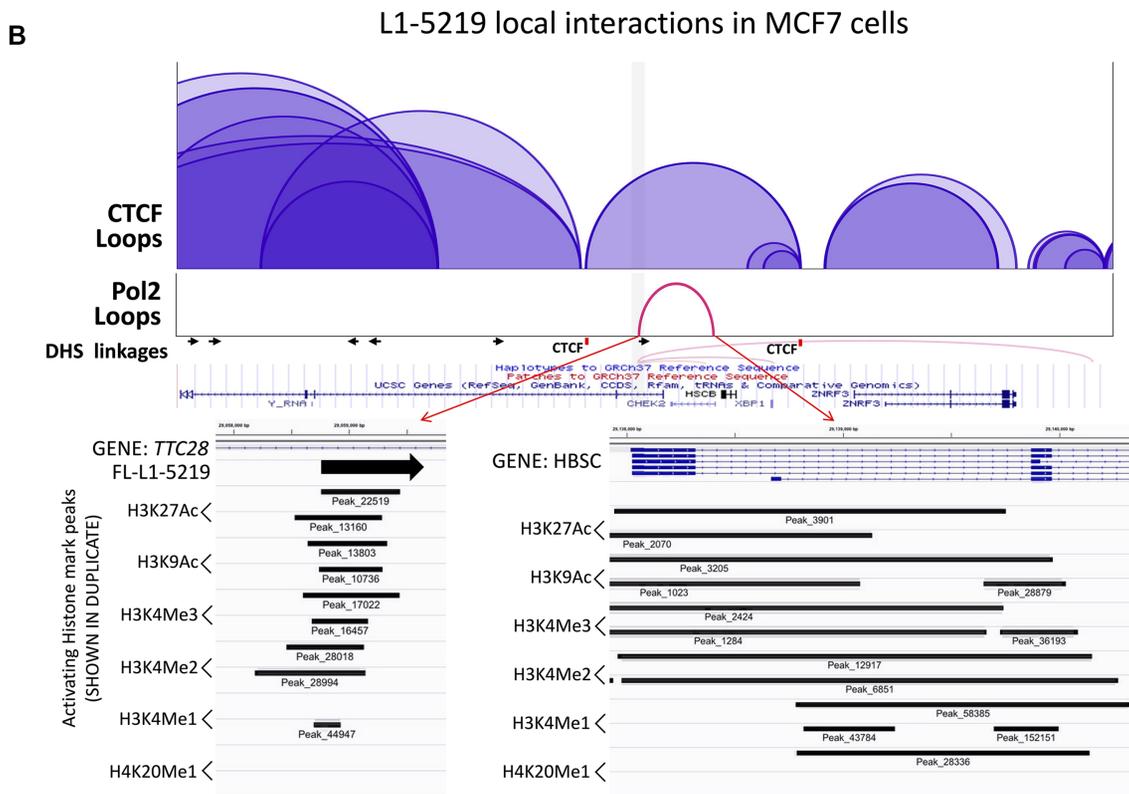


Figure 7. Comparative analysis of full-length L1 loci identified within the *TTC28* gene. (A) Schematic of the *TTC28* gene (black) and exons (heavy black bars) transcription start site (green arrow), seven full-length L1 loci (blue arrows reflecting L1 orientation relative to the gene), and a CTCF binding site (orange box). The chart below the schematic contains information regarding each of the seven loci: L1-5212, L1-5214, L1-5215, L1-5216, L1-5217, L1-5218 and L1-5219. The information provided is L1 sub-family, RNA-Seq reads for expressed loci (Illumina or PacBio technologies (22)), and sequence status of YY1 and RUNX3 binding sites (55). Reads mapped to L1-5217 in HEK293 cells were manually determined not to primarily originate from an L1 promoter. (B) 3D interactions and epigenetic marks identified at L1 loci within the *TTC28* gene in MCF7 cells. Loops indicate CTCF binding sites (purple, shaded) within 500 kb of the L1-5219 start site (black arrow). RNA polymerase II (Pol II) loops (red) are only shown if the pol II binding site directly overlaps with an L1 start site. DNase Hypersensitivity Site (DHS) linkages (light red lines) and the surrounding genomic sequences (blue) are shown as well (92). A CTCF binding site 5' of the L1-5219 is marked in red. Red arrows indicate magnification of the L1-5219 element and the associated enhancer region, respectively, showing the presence of activating histone marks at each genomic region as determined by previously described CHIP-Seq data (Figure 3A). The status of each histone mark is shown from two experiments.

Table 1. Summary of Long-distance interactions of 10 expressed, transitional, and unexpressed L1-loci in MCF7 cells. Data regarding 3D chromatin interactions for all expressed, transitional, and unexpressed L1 loci analyzed in Figure 6 and Supplemental Figure S6 is summarized. The data listed for each L1 locus (L1 locus column) includes number of mapped RNA-Seq reads (Reads column), status of an ATAC peak (ATAC peak column), position within a CTCF loop (CTCF loop column), RNA pol II association (pol II loop column), number of activating histone marks overlapping the L1 locus (histone marks on L1 column), and number of activating histone marks overlapping with the putative enhancer region (histone marks on enhancer column)

Expressed loci							
Figure	L1 locus	Reads	ATAC peak	CTCF loop	Pol II loop	Histone marks on L1	Histone marks on enhancer
6A	L1-2830	484	Yes	Yes	Yes	4	11
Supp. 6A	L1-0728	449	Yes	Yes	Yes	9	0
Supp. 6B	L1-3682	322	Yes	Yes	Yes	7	11
Supp. 6C	L1-1867	272	Yes	Yes	Yes	8	11
Supp. 6D	L1-2476	242	Yes	No	Yes	0	10
Supp. 6E	L1-3165	239	Yes	Yes	Yes	10	0
Supp. 6F	L1-0029	226	Yes	No	Yes	9	10
Supp. 6G	L1-3455	224	Yes	Yes	Yes	10	12
Supp. 6H	L1-1685	216	Yes	Yes	Yes	5	11
Supp. 6I	L1-3239	210	Yes	No	No	9	No Enhancer (0)
	AVERAGE	288.4	Yes (10/10)	Yes (7/10)	Yes (9/10)	7.1 (STDEV: 3.2)	7.6 (STDEV: 5.3)
Transitional loci							
Figure	L1 locus	Reads	ATAC peak	CTCF loop	Pol II loop	Histone marks on L1	Histone marks on enhancer
6B	L1-0650	0	Yes	Yes	No	0	No Enhancer (0)
Supp. 6J	L1-2855	0	Yes	Yes	Yes	6	10
Supp. 6K	L1-0225	0	Yes	Yes	Yes	5	12
Supp. 6L	L1-0986	0	Yes	Yes	Yes	0	3
Supp. 6M	L1-4910	0	Yes	Yes	No	5	No Enhancer (0)
Supp. 6N	L1-5151	0	Yes	Yes	No	4	No Enhancer (0)
Supp. 6O	L1-1469	0	Yes	Yes	No	1	No Enhancer (0)
Supp. 6P	L1-0482	0	Yes	Yes	No	1	No Enhancer (0)
Supp. 6Q	L1-3525	0	Yes	Yes	Yes	9	0
Supp. 6R	L1-4180	0	Yes	Yes	No	2	No Enhancer (0)
	AVERAGE	0	Yes (10/10)	Yes (10/10)	No (4/10)	3.3 (STDEV: 3)	2.5 (STDEV: 4.6)
Unexpressed Loci							
Figure	L1 locus	Reads	ATAC Peak	CTCF loop	Pol II loop	Histone marks on L1	Histone marks on Enhancer
Fig. 6C	L1-0238	0	No	Yes	No	0	No Enhancer (0)
Supp. 6S	L1-0518	0	No	Yes	No	0	No Enhancer (0)
Supp. 6T	L1-1501	0	No	Yes	No	0	No Enhancer (0)
Supp. 6U	L1-1960	0	No	Yes	No	0	No Enhancer (0)
Supp. 6V	L1-4228	0	No	Yes	No	0	No Enhancer (0)
Supp. 6W	L1-4249	0	No	Yes	No	0	No Enhancer (0)
Supp. 6X	L1-4821	0	No	Yes	No	0	No Enhancer (0)
Supp. 6Y	L1-4938	0	No	Yes	Yes	0	0
Supp. 6Z	L1-5330	0	No	Yes	No	0	No Enhancer (0)
Supp. 6AA	L1-5742	0	No	Yes	No	0	No Enhancer (0)
	AVERAGE	0	No (0/10)	Yes (10/10)	No (1/10)	0 (STDEV: 0)	0 (STDEV: 0)

epigenetic context of L1Hs loci to predict which of these poorly mappable L1 elements are likely expressed.

Endogenous L1 elements can be transiently transactivated to lead to an increase in exogenous Alu mobilization

In order to determine whether expression of L1 loci could be manipulated by altering chromatin state without changing DNA sequence, we utilized a CRISPR/Cas9 activator- (CRISPR/Cas9a) and CRISPR/Cas9 inhibitor- (CRISPR/Cas9i) based gene regulation system (113,114) with guide RNAs (gRNAs) targeting multiple locations on both strands within the L1 5' UTR (Figure 9A). The CRISPR/Cas9a approach combines the guide RNAs, with CRISPR expression constructs fused to VP64 activating do-

main, while the CRISPR/Cas9i approach fuse the CRISPR with the KRAB repressor domain. Several guide RNAs complementary to the L1 5'UTR nucleotide positions 76, 83, 190, 285, 480, 677 and 892 in the sense and positions 72, 111, 366, 551, 765, 873 and 880 in the antisense orientation were tested alone, or as pools, for their ability to alter expression of the firefly luciferase reporter mRNA transcribed from the L1 5' UTR in HeLa cells. While individual gRNAs were variably effective at specifically stimulating transcription of the firefly luciferase, pool 2 (containing five gRNAs targeting positions 111, 190, 285, 480 and 880, Figure 9B) was the most effective relative to the negative control transfection of a L1 5' UTR reporter plasmid and gRNA targeting AAV ($P < 0.01$, $P < 0.0001$, Student's *t*-test). When replacing the Cas9-VP64 fusion with a Cas9-KRAB fusion

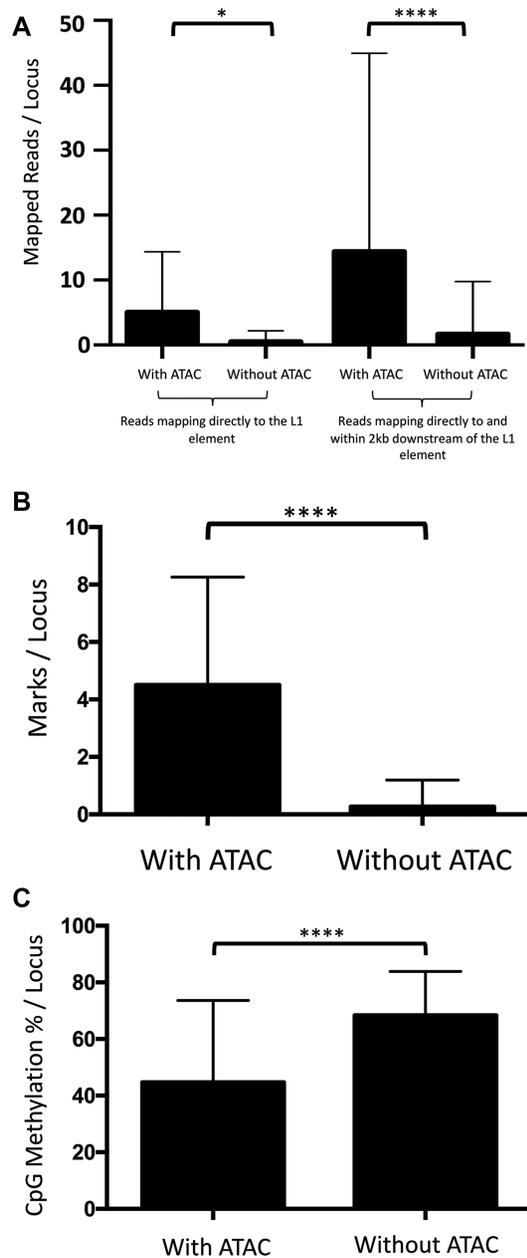


Figure 8. Epigenetic analysis identifies a subset of poorly mappable L1Hs loci that are likely expressed. (A) Reads mapping directly to the L1 element: RNA-Seq reads uniquely mapped to annotated L1Hs loci identified in MCF7 cells with and without an ATAC peak in MCF7 (AVG = 5.1 reads, $n = 18$ versus AVG = 0.5 reads, $n = 228$; $*P < 0.05$). Reads mapping directly to and within 2 kb downstream of the L1 element: Number of RNA-Seq mapped reads per locus with and without an ATAC peak when including 2 kb downstream of each locus is shown (AVG = 14.4 reads, $n = 18$ versus AVG = 1.7 reads, $n = 228$; ****, $P < 0.0001$). Student's *t*-test was used to analyze for significance. (B) Number of activating histone marks per L1Hs locus with and without an ATAC peak is shown (AVG = 4.5 marks, $n = 18$ versus AVG = 0.3 marks, $n = 228$; ****, $P < 0.0001$). Student's *t*-test was used to analyze for significance. (C) Percentage of CpG methylation per L1Hs locus with and without an ATAC peak is shown (AVG = 44.7%, $n = 18$ versus AVG = 68.4%, $n = 228$; ****, $P < 0.0001$). Student's *t*-test was used to analyze for significance.

in this system (CRISPR/Cas9i), we see a significant reduction in the L1 5' UTR-driven reporter-gene expression when co-transfecting a pool of gRNAs as opposed to a negative control gRNA targeting AAV ($P < 0.001$, Student's *t*-test) (Supplemental Figure S 8A).

As a first approach to determine the influence of CRISPR/Cas9a on stimulation of expression of endogenous L1 elements, we used a biologically relevant Alu retrotransposition assay to measure mobilization of an exogenous Alu element by endogenously expressed L1s as previously described (21,115,116) (Supplemental Figure S8B). Alu elements are completely dependent on L1 ORF2p expression for retrotransposition (99). G418-resistant colonies generated in the absence of transfected L1 are produced by Alu retrotransposition driven by endogenously expressed L1 elements. In parallel, any toxicity from the treatment was measured by replacing the Alu reporter plasmid with a constitutive G418 resistance (neo^R) selectable marker. Co-transfection of the Alu reporter plasmid, Cas9-VP64 plasmid, and a pool of gRNAs targeting the L1 5'UTR resulted in 3.2 times as many colonies as were observed in the presence of control gRNA targeting AAV ($P = 0.0359$, Student's *t*-test) (Figure 9C). In contrast, when a Cas9-KRAB fusion plasmid was used in this system, we observed a non-significant decrease in the number of G418-resistant colonies ($P = 0.112$) (Figure 9D) and in neither experiment did we detect any increased cellular toxicity (Figure 9C,D).

Only a subset of endogenous L1 loci can be transactivated by the CRISPR/Cas9a approach

In order to identify which endogenous L1 loci are transactivated by the CRISPR/Cas9a system and the gRNAs targeting L1 sequences using RNA-Seq, we transfected HeLa cells with the CRISPR/Cas9a-expression plasmid along with the pool 2 plasmids expressing gRNAs shown in Figure 9B. As a control, transfection with an empty gRNA plasmid was performed in parallel. Paired-end, stranded RNA-Seq was performed using cytoplasmic RNA isolated from each of these transfected cell lines. Alignment of RNA-Seq reads, manual curation, and quantitation of reads mapped to individual full-length L1 loci was carried out as previously described (22,88). Transfecting HeLa cells with the pool 2 of gRNAs resulted in an approximately 1.5X increase in reads corresponding to expressed L1 loci, from 138.9 reads per million detected in the control to 198.5 reads per million upon activation.

To determine which endogenous L1 loci were stimulated under these experimental conditions, we manually curated the L1 loci with 5 + mapped reads in both the CRISPR/Cas9a-treated cells and the empty gRNA plasmid control cells according to their individual expression level. All L1 loci with expression levels identified as 5+ reads in either the CRISPR/Cas9a-treated or control samples that pass manual curation are used for this comparison. We reported the expression level for each of these loci in both the activated and control cells (Figure 10, the loci expressed in the control cells (blue) versus the loci expressed in the CRISPR/Cas9a-treated cells (black; Additional File, page 2). The L1 loci with the highest levels of mRNA expres-

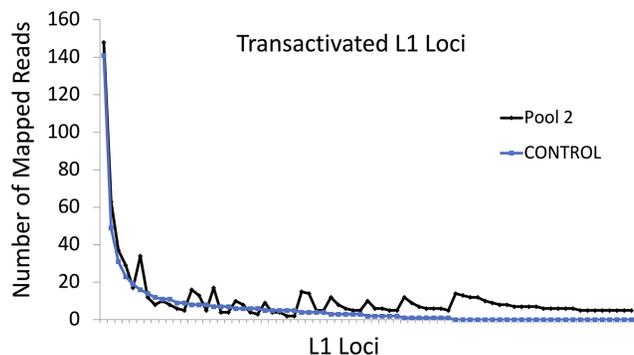


Figure 10. A CRISPR/Cas9a-based regulatory system activates endogenous unexpressed L1 elements. HeLa cells were transfected with a plasmid expressing CAS9-VP64 domain as well as plasmid expressing either pool 2 gRNAs (Figure 9A) (Pool 2) or empty gRNA control (CONTROL). PE RNA-Seq was performed and all L1 loci with 5+ mapped reads identified by RNA-Seq pipeline and manually validated as expressed in either the pool 2 (black) or control (blue) samples are displayed in the order from high to low number of reads as determined for Pool 2. HeLa cells transfected with pool 2 gRNAs had 65 loci with 5+ mapped reads and HeLa cells transfected with empty control gRNAs had 27 loci with 5+ mapped reads. Detailed information regarding these loci can be found on page 2 of the Additional File. The total number of mapped reads from RNA-Seq BAM files in the pool 2 and control HeLa cells is 17,927,082 and 16,581,681, respectively.

L1 loci is also observed *in vivo* (33). The number of expressed L1 loci is typically <1% of full-length L1 loci in most cell lines (22). Even in a cell line like MCF7 that supports some of the highest levels of L1 mRNA expression, only 3% of L1 loci express moderately (Figure 1B). Based on the finding that methylation can suppress L1 promoter function (38,39), a large number of previous studies have often assumed that global hypomethylation and histone activating marks at L1 promoters correlate with L1 expression (36,38,48,64,117–119). Our findings demonstrate that the sensitivity of those studies is often insufficient to correlate epigenetic marks with L1 expression because they were investigating the state of the small group of expressed L1 elements (~50–100) pooled together with a large set of thousands of unexpressed L1 loci. For instance, we find that both the percent of global CpG methylation (Supplemental Figure S4B) and percent of CpG methylation per L1 locus (Supplemental Figure S4C) that are significantly higher in MCF7 than in HeLa and HEK293 cells do not correlate with the highest level of L1 mRNA expression detected in MCF7 cells (Figure 1A). However, significant hypomethylation of L1 promoters is detected when only expressed L1 loci are considered in our study (Figure 4). This finding highlights that global analysis of epigenetic state of L1 promoters does not sensitively reflect L1 mRNA expression. Thus, there is still a relatively poor understanding of what epigenetic features are characteristic of expressed L1 loci and studies focusing on analysis of expressed and unexpressed L1 loci as separate groups are needed to understand how L1 expression may be regulated during development, differentiation, and transformation.

We have focused our studies on individual highly expressed L1 loci in MCF7 cells. We find that there is a strong correlation between L1 locus expression status and their

promoter's association with epigenetic signals that are typically found at the promoters of expressed cellular genes. Specifically, the presence of an ATAC peak over the L1 promoter, histone activating marks around the L1 promoter region, and promoter hypomethylation all show significant enrichment in L1 loci expressed in MCF7 cells compared to unexpressed L1 loci (Figures 2–4). In HeLa and HEK293 cells, we also find enrichment for ATAC peaks (Figure 2) and activating histone marks (Supplemental Figure S3A–C) at promoters of the expressed L1 loci. In contrast to the findings in MCF7 cells, we were unable to detect hypomethylation at L1 loci in expressed HeLa and HEK293 cells (Supplemental Figure S4A), probably reflecting that these cells have relatively few and poorly expressed loci (Figure 1). When comparing the epigenetic status of loci expressed and unexpressed in MCF7 cells within HeLa and HEK293 cells, we find that loci expressed in MCF7 retain their associations with activating histone marks (Supplemental Figure S5A,B) and ATAC peaks (Supplemental Figure S5D) as a group in both HeLa and HEK293. This is likely due to the fact that 41 of the 162 L1 loci are expressed in HeLa cells and 53 of the 162 L1 loci are 'transitional' (Figure 5). As expected, within this group of 162 L1 loci expressed in MCF7 cells, there are fewer L1 loci that have activating histone marks in HeLa cells (Figure 3A and Supplemental Figure S5A), a likely reflection of the L1 loci that are not expressed in HeLa cells. In contrast to ATAC and activating histone marks, methylation status of the 162 L1 loci in HeLa and HEK 293 cells does not distinguish them from the unexpressed L1 loci in the manner it does in MCF7 cells (Supplemental Figure S5C and Figure 4). These findings indicate a role for open chromatin, activating histone marks, and CpG methylation in contributing to cell-line specific expression of individual L1 loci.

ATAC, Chip-Seq, and bisulfite sequencing analyses of expressed and unexpressed L1 loci determined that no single epigenetic factor is sufficient to clearly predict expressed L1 loci (Figures 2–4). However, these approaches have identified a subset of L1 loci that are not expressed, but have an ATAC peak (Figure 2), often contain histone activating marks (Figure 3B), and are relatively hypomethylated (Figure 4). We refer to these as 'transitional' L1 loci, which we consider to have some of the necessary features for expression, but still are missing at least one of the necessary factors to be expressed. This consideration is based on the finding that stimulation of endogenous L1 mRNA expression using the CRISPR/Cas9a system results in the upregulation of the poorly expressed and 'transitional' loci (Supplemental Figure S9), suggesting that these loci are susceptible to reaching the critical mass of activating factors to launch their expression.

In addition to identifying differences in the local epigenetic factors at the promoters of expressed and unexpressed L1 loci, we discovered that promoters of expressed L1 loci appear to associate with an active transcription enhancer. This is best seen using the three-dimensional ChIA-PET data that demonstrate an RNA polymerase II-driven association between the promoters of expressed L1 loci and an enhancer with strong histone activating marks (Table 1, Figure 6, Supplemental Figure S6). This seems to be one of the major factors that distinguishes the expressed loci

from the ‘transitional’ L1 loci, although there are exceptions (Table 1). It seems likely that the three-dimensional L1 promoter association with an enhancer is also influenced by the formation of specific CTCF-mediated interactions. This is supported by the finding that L1 elements near a highly expressed L1 element, like those nearby L1-5219, are not expressed because they are not localized within the CTCF/cohesin loop(s) that also contains RNA Pol II-mediated loops with strong enhancers (Figure 7B). This is further supported by the CRISPR/Cas9a experiments demonstrating that a targeted recruitment of pol II to ‘transitional’ L1 loci results in their expression (Figure 10, Supplemental Figure S9).

Our results highlight the benefit of analyzing individual expressed L1 loci as a group to demonstrate the complexity and variability of factors involved in expression of different L1 loci. In addition, they highlight the power of epigenetic measures to guide and confirm expression analysis of individual L1 loci. For instance, ATAC analysis alone limits the number of L1 loci to be considered for potential expression by ~10-fold (from 5,000 to ~500). Combining ATAC or activating histone marks with RNA polymerase II association with an enhancer sequence in MCF7 cells provides even more power (Table 1). Building on our findings, epigenetic analysis may help identify which L1Hs loci are expressed but cannot be unambiguously detected with RNA-Seq due to poor mappability. Our data demonstrate that the 18 (out of 246) L1Hs loci present in MCF7 cells that have an ATAC peak contain a higher number of activating histone marks and lower levels of CpG methylation compared to L1Hs loci without an ATAC peak (Figure 8, Supplemental Figure S7). As only three of these 18 loci are expressed above our threshold in MCF7 cells (Additional File, page 3), we predict that some, but not all, of the other 15 loci are likely expressed at some level due to their favorable epigenetics conditions below threshold levels of expression. Consistent with ATAC performance on mappable L1 loci, it provides a greater than 10-fold reduction in the number of L1Hs loci to be considered for expression. As L1Hs loci are younger elements that are more likely to be highly expressed and retrotranspositionally-competent (21,120,121), being able to identify which of these loci are expressed may have important prognostic implications regarding the mutagenic effects of L1 expression (122,123).

Our analysis of epigenetic marks at individual expressed and unexpressed L1 loci establishes a roadmap for manipulation of L1 loci expression either through a CRISPR/Cas9a system or knockout of specific sequences, as well as for studies of normal changes occurring at L1 promoters in different biological settings. We would predict that the ‘transitional’ loci in a tissue would be most prone to activation during development or through the oncogenic transformation. In fact, combined with cell type specific L1 mRNA expression (5), our findings demonstrate that any biological process associated with perturbations of epigenetic and/or three-dimensional genomic landscape, such as cancer, aging, environmental exposures, and even normal developmental processes, would be expected to involve epigenetic features that are identified in this study to alter L1 expression patterns. Cancer cells, for instance, typically have much higher levels of L1 expression than do normal tissues

(27,124–126). One previously used explanation was that hypomethylation of L1 promoters that may occur in the tumor could have a broad impact on L1 loci expression. However, as expression of individual L1 elements is highly dependent on cellular context, as seen in various mouse tissues (33) and human cell lines (22,32) (Figure 1), our data demonstrate that a wide array of cancer-specific changes, including sequence rearrangements repositioning CTCF binding sites, individual interactions between L1 elements and their enhancers, or changes in RNA Pol II recruitment through transcriptional factors, influence expression of individual L1 loci.

Perhaps the most general finding in our studies is that the L1 element by itself is probably insufficient for significant expression without help from its local genomic environment. Some level of open chromatin, allowing detection of an ATAC signal at the promoter, as well as activating histone marks is clearly necessary, but not sufficient for L1 expression as demonstrated by the presence of ‘transitional’ L1 loci. A dominant feature of expressed L1 loci also seems to be the presence of a three-dimensional interaction between the L1 promoter and a strong enhancer. Unlike a typical gene whose entire genomic environment evolves as a unit or a virus that brings its own enhancer, a new L1 insert is placed in a completely unique sequence and epigenetic environment. The distinct features of individual environments will immediately have specific, but likely different, influences on expression for newly inserted L1 loci. Thus, following integration, each L1 locus will evolve as a completely independent gene. This is supported by the results of our CRISPRa stimulation of L1 expression (Figures 9 and 10). The unique feature of this approach is that essentially the same promoter is stimulated at different locations with different results depending on those environments. This is a unique situation compared to genes. It is also an important technical observation in that it shows investigators a working approach to stimulate endogenous L1 expression to look for the influence of changes in L1 expression in a manner that is much more physiological than transfecting in an exogenous L1. Based on these findings, it might be expected that a new L1 insert that landed in the genomic environment that is permissive of its expression to high levels in many cell types (i.e. similar to a housekeeping gene) would experience high negative evolutionary pressure and be more likely to be eliminated by selection. This is consistent with the finding that full-length L1 loci are deleterious and have been under purifying selection (127–129). This may well explain the strong cell-type and tissue-specific expression patterns from individual L1 loci (22,32,130) because expressing an L1 locus ubiquitously would be more deleterious.

Overall, in this study we identify key local (chromatin state, methylation status, activating histone marks at individual L1 promoters) and long-distance (PolII interactions, CTCF loops, activating histone marks at relevant enhancers) epigenetic features that are different between expressed and unexpressed L1 loci. We demonstrate their relative power in distinguishing expressed from unexpressed L1 loci. We demonstrate that our discoveries can be used for predicting expression of poorly mapped, but the most disease relevant, L1Hs loci, expression of which cannot be unambiguously determined by RNA-Seq alone. Based

on their epigenetic and expression status we identify a new category of L1 loci we term ‘transitional’ because they have several epigenetic features unique to expressed L1 loci even though they are not expressed. Using L1-specific CRISPR/Cas activation and inhibition approaches, we demonstrate as a proof-of-principle that expression of individual endogenous L1 loci can be manipulated, but the three classes of L1 loci (expressed, unexpressed and transitional) respond differently to this manipulation with transitional L1 loci being the most responsive to the induced activation. Finally, we demonstrate that activation of endogenous L1 mRNA expression is biologically relevant and measurable because we observe a significant increase in Alu retrotransposition in response to L1 activation.

DATA AVAILABILITY

All data will be made available to the public upon publication. The manuscript’s meta-data is available for viewing via SRA at <https://dataview.ncbi.nlm.nih.gov/object/PRJNA724674?reviewer=hs0mv88e2vgn8br01t2pbm139r>.

Publicly-available data can be accessed at the following accession numbers:

- RNA-Seq:
 - MCF7: ENCF000HSC and ENCF000HSK (ENCODE)
 - HeLa: Data can be found at link provided in AVAILABILITY section (SRA)
 - HEK293: SRR1275413 (GEO)
- ATAC-Seq: Data for MCF7, HeLa and HEK293 can be found at link provided in AVAILABILITY section (SRA)
- WGBS: Data for MCF7, HeLa and HEK293 can be found at link provided in AVAILABILITY section (SRA)
- Histone Marks CHIP-Seq (ENCODE):
 - MCF7
 - H3K27Ac
 - ENCSR752UOD
 - H3K9Ac
 - ENCSR056UBA
 - H3K4Me3
 - ENCSR985MIB
 - H3K4Me2
 - ENCSR875KOJ
 - H3K4Me1
 - ENCSR493NBY
 - H4K20Me1
 - ENCSR639RHG
 - HeLa
 - H3K27Ac
 - ENCSR000AOC
 - H3K9Ac
 - ENCSR000AOH
 - H3K4Me3
 - ENCSR340WQU
 - H3K4Me2
 - ENCSR000AOE
 - H3K4Me1
 - ENCSR000APW
 - H4K20Me1

- ENCSR000AOI
- HEK293
- H3K27Ac
- ENCSR000FCH
- H3K4Me3
- ENCSR000DTU
- H3K4Me1
- ENCSR000FCG

Additional information regarding data used in this manuscript can be found in the Materials and Methods section.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank members of the Consortium of Mobile Elements at Tulane (COMET) for critical discussions of the data.

Authors’ contributions: All authors contributed to data collection. B.F., T.W., P.D. and V.B. conceived ideas, analyzed data, and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

National Institutes of Health [R01 GM121812 to P.D., R01 AG057597 to V.B.]; Brown Foundation (to V.B.); Louisiana State Board of Regents Graduate Research Fellowship (to B.F.). Funding for open access charge: National Institutes of Health [R01 AG057597].

Conflict of interest statement. None declared.

REFERENCES

1. Cordaux, R. and Batzer, M. A. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.*, **10**, 691–703.
2. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
3. Khan, H., Smit, A. and Boissinot, S. (2006) Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.*, **16**, 78–87.
4. Ostertag, E. M. and Kazazian, H. H. Jr (2001) Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.*, **35**, 501–538.
5. Swergold, G. D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.*, **10**, 6718–6729.
6. Khazina, E., Truffault, V., Buttner, R., Schmidt, S., Coles, M. and Weichenrieder, O. (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat. Struct. Mol. Biol.*, **18**, 1006–1014.
7. Martin, S. L. (2010) Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol.*, **7**, 706–711.
8. Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D. and Kazazian, H. H. Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
9. Feng, Q., Moran, J. V., Kazazian, H. H. Jr and Boeke, J. D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
10. Leibold, D. M., Swergold, G. D., Singer, M. F., Thayer, R. E., Dombroski, B. A. and Fanning, T. G. (1990) Translation of LINE-1 DNA elements in vitro and in human cells. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 6990–6994.

11. McMillan, J.P. and Singer, M.F. (1993) Translation of the human LINE-1 element, L1Hs. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 11533–11537.
12. Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D. and Moran, J.V. (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.*, **21**, 1429–1439.
13. Hohjoh, H. and Singer, M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **15**, 630–639.
14. Kulpa, D.A. and Moran, J.V. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.*, **13**, 655–660.
15. Freeman, B.T., Sokolowski, M., Roy-Engel, A.M., Smither, M.E. and Belancio, V.P. (2019) Identification of charged amino acids required for nuclear localization of human L1 ORF1 protein. *Mob. DNA*, **10**, 20.
16. Dai, L., LaCava, J., Taylor, M.S. and Boeke, J.D. (2014) Expression and detection of LINE-1 ORF-encoded proteins. *Mob. Genet. Elements*, **4**, e29319.
17. Alisch, R.S., Garcia-Perez, J.L., Muotri, A.R., Gage, F.H. and Moran, J.V. (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.*, **20**, 210–224.
18. Cost, G.J., Feng, Q., Jacquier, A. and Boeke, J.D. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J.*, **21**, 5899–5910.
19. Whitcomb, J.M. and Hughes, S.H. (1992) Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell Biol.*, **8**, 275–306.
20. Luan, D.D., Korman, M.H., Jakubczak, J.L. and Eickbush, T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.
21. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian, H.H. Jr (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 5280–5285.
22. Deininger, P., Morales, M.E., White, T.B., Baddoo, M., Hedges, D.J., Servant, G., Srivastav, S., Smither, M.E., Concha, M., DeHaro, D.L. et al. (2017) A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.*, **45**, e31.
23. Goodier, J.L. (2016) Restricting retrotransposons: a review. *Mob. DNA*, **7**, 16.
24. Bannister, A.J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.
25. Dahlet, T., Argueso Lleida, A., Al Adhami, H., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet, J., Auclair, G., Bardet, A.F. et al. (2020) Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat. Commun.*, **11**, 3153.
26. Helman, E., Lawrence, M.S., Stewart, C., Sougnez, C., Getz, G. and Meyerson, M. (2014) Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.*, **24**, 1053–1063.
27. Solyom, S., Ewing, A.D., Rahrman, E.P., Doucet, T., Nelson, H.H., Burns, M.B., Harris, R.S., Sigmon, D.F., Casella, A., Erlanger, B. et al. (2012) Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.*, **22**, 2328–2338.
28. Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J. 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K. et al. (2012) Landscape of somatic retrotransposition in human cancers. *Science*, **337**, 967–971.
29. Tubio, J.M.C., Li, Y., Ju, Y.S., Martincorena, I., Cooke, S.L., Tojo, M., Gundem, G., Pipinikas, C.P., Zamora, J., Raine, K. et al. (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science*, **345**, 1251343.
30. Kaul, T., Morales, M.E., Sartor, A.O., Belancio, V.P. and Deininger, P. (2020) Comparative analysis on the expression of L1 loci using various RNA-Seq preparations. *Mob. DNA*, **11**, 2.
31. Scott, E.C., Gardner, E.J., Masood, A., Chuang, N.T., Vertino, P.M. and Devine, S.E. (2016) A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.*, **26**, 745–755.
32. Philippe, C., Vargas-Landin, D.B., Doucet, A.J., van Essen, D., Vera-Otarola, J., Kuciak, M., Corbin, A., Nigumann, P. and Cristofari, G. (2016) Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife*, **5**, e13926.
33. Stow, E.C., Kaul, T., DeHaro, D.L., Dem, M.R., Beletsky, A.G., Morales, M.E., LaRosa, A.J., Yang, H., Smither, E., Baddoo, M. et al. (2021) Organ-, sex-, and age-dependent patterns of endogenous L1 mRNA expression at a single locus resolution. *Nucleic Acids Res.*, **49**, 5813–5831.
34. de la Rica, L., Deniz, O., Cheng, K.C., Todd, C.D., Cruz, C., Houseley, J. and Branco, M.R. (2016) TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.*, **17**, 234.
35. Walter, M., Teissandier, A., Perez-Palacios, R. and Bourc'his, D. (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife*, **5**, e11418.
36. Sanchez-Luque, F.J., Kempen, M.H.C., Gerdes, P., Vargas-Landin, D.B., Richardson, S.R., Troskie, R.L., Jesuadian, J.S., Cheetham, S.W., Carreira, P.E., Salvador-Palomeque, C. et al. (2019) LINE-1 evasion of epigenetic repression in humans. *Mol. Cell*, **75**, 590–604.
37. Crowther, P.J., Doherty, J.P., Linsenmeyer, M.E., Williamson, M.R. and Woodcock, D.M. (1991) Revised genomic consensus for the hypermethylated CpG island region of the human L1 transposon and integration sites of full length L1 elements from recombinant clones made using methylation-tolerant host strains. *Nucleic Acids Res.*, **19**, 2395–2401.
38. Thayer, R.E., Singer, M.F. and Fanning, T.G. (1993) Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene*, **133**, 273–277.
39. Hata, K. and Sakaki, Y. (1997) Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene*, **189**, 227–234.
40. Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P. and Warren, W.D. (1997) Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J. Biol. Chem.*, **272**, 7810–7816.
41. Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.
42. Rollins, R.A., Haghghi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J. and Bestor, T.H. (2006) Large-scale structure of genomic methylation patterns. *Genome Res.*, **16**, 157–163.
43. Garcia-Perez, J.L., Morell, M., Scheys, J.O., Kulpa, D.A., Morell, S., Carter, C.C., Hammer, G.D., Collins, K.L., O'Shea, K.S., Menendez, P. et al. (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature*, **466**, 769–773.
44. Rosser, J.M. and An, W. (2012) L1 expression and regulation in humans and rodents. *Front. Biosci. (Elite Ed.)*, **4**, 2203–2225.
45. Singer, H., Walier, M., Nusgen, N., Meesters, C., Schreiner, F., Woelfle, J., Fimmers, R., Wienker, T., Kalscheuer, V.M., Becker, T. et al. (2012) Methylation of L1Hs promoters is lower on the inactive X, has a tendency of being higher on autosomes in smaller genomes and shows inter-individual variability at some loci. *Hum. Mol. Genet.*, **21**, 219–235.
46. Grandi, F.C., Rosser, J.M., Newkirk, S.J., Yin, J., Jiang, X., Xing, Z., Whitmore, L., Bashir, S., Ivics, Z., Izsvák, Z. et al. (2015) Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Res.*, **25**, 1135–1146.
47. Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Héroult, Y., Guillou, F. and Bourc'his, D. (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, **354**, 909–912.
48. Salvador-Palomeque, C., Sanchez-Luque, F.J., Fortuna, P.R.J., Ewing, A.D., Wolvetang, E.J., Richardson, S.R. and Faulkner, G.J. (2019) Dynamic methylation of an L1 transduction family during reprogramming and neurodifferentiation. *Mol. Cell. Biol.*, **39**, e00499-18.
49. Molaro, A., Malik, H.S. and Bourc'his, D. (2020) Dynamic evolution of de novo DNA methyltransferases in rodent and primate genomes. *Mol. Biol. Evol.*, **37**, 1882–1892.

50. Deaton, A.M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
51. Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
52. Athanikar, J.N., Badge, R.M. and Moran, J.V. (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.*, **32**, 3846–3855.
53. Becker, K.G., Swergold, G.D., Ozato, K. and Thayer, R.E. (1993) Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum. Mol. Genet.*, **2**, 1697–1702.
54. Yang, N., Zhang, L., Zhang, Y. and Kazazian, H.H. Jr (2003) An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.*, **31**, 4929–4940.
55. Lee, S.H., Cho, S.Y., Shannon, M.F., Fan, J. and Rangasamy, D. (2010) The impact of CpG island on defining transcriptional activation of the mouse L1 retrotransposable elements. *PLoS One*, **5**, e11353.
56. Sun, X., Wang, X., Tang, Z., Grivainis, M., Kahler, D., Yun, C., Mita, P., Fenyo, D. and Boeke, J.D. (2018) Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E5526–E5535.
57. Thornburg, B.G., Gotea, V. and Makalowski, W. (2006) Transposable elements as a significant source of transcription regulating signals. *Gene*, **365**, 104–110.
58. Wylie, A., Jones, A.E., D’Brot, A., Lu, W.J., Kurtz, P., Moran, J.V., Rakheja, D., Chen, K.S., Hammer, R.E., Comerford, S.A. *et al.* (2016) p53 genes function to restrain mobile elements. *Genes Dev.*, **30**, 64–77.
59. Tchenio, T., Casella, J.F. and Heidmann, T. (2000) Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.*, **28**, 411–415.
60. Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R. and Haussler, D. (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, **516**, 242–245.
61. Yang, P., Wang, Y. and Macfarlan, T.S. (2017) The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet.*, **33**, 871–881.
62. Rangasamy, D. (2013) Distinctive patterns of epigenetic marks are associated with promoter regions of mouse LINE-1 and LTR retrotransposons. *Mob DNA*, **4**, 27.
63. Pezic, D., Manakov, S.A., Sachidanandam, R. and Aravin, A.A. (2014) piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev.*, **28**, 1410–1428.
64. Whongsiri, P., Goering, W., Lautwein, T., Hader, C., Niegisch, G., Kohrer, K., Hoffmann, M.J. and Schulz, W.A. (2020) Many different LINE-1 retroelements are activated in bladder cancer. *Int. J. Mol. Sci.*, **21**, 9433.
65. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, **40**, 897–903.
66. Kratz, A., Arner, E., Saito, R., Kubosaki, A., Kawai, J., Suzuki, H., Carninci, P., Arakawa, T., Tomita, M., Hayashizaki, Y. *et al.* (2010) Core promoter structure and genomic context reflect histone 3 lysine 9 acetylation patterns. *BMC Genomics*, **11**, 257.
67. Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.J. *et al.* (2017) Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.*, **292**, 14456–14472.
68. Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C. and Harte, P.J. (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes drosophila polycomb silencing. *Development*, **136**, 3131–3141.
69. Creighton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21931–21936.
70. Bae, S. and Lesch, B.J. (2020) H3K4me1 distribution predicts transcription state and poising at promoters. *Front. Cell Dev. Biol.*, **8**, 289.
71. Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J. 3rd, Gingeras, T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
72. Vakoc, C.R., Sachdeva, M.M., Wang, H. and Blobel, G.A. (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell Biol.*, **26**, 9185–9195.
73. He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M.M., Zhou, J., Sun, L., Chang, C., Li, Y. *et al.* (2019) Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.*, **10**, 34.
74. Ren, G., Jin, W., Cui, K., Rodriguez, J., Hu, G., Zhang, Z., Larson, D.R. and Zhao, K. (2017) CTCF-Mediated enhancer-promoter interaction is a critical regulator of Cell-to-Cell variation of gene expression. *Mol. Cell*, **67**, 1049–1058.
75. Aitken, S.J., Ibarra-Soria, X., Kentepozidou, E., Flicek, P., Feig, C., Marioni, J.C. and Odom, D.T. (2019) CTCF maintains regulatory homeostasis of cancer pathways. *Genome Biol.*, **19**, 106.
76. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. and Darzacq, X. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife*, **6**, e25776.
77. Lavie, L., Maldener, E., Brouha, B., Meese, E.U. and Mayer, J. (2004) The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.*, **14**, 2253–2260.
78. An, W., Dai, L., Niewiadomska, A.M., Yetil, A., O’Donnell, K.A., Han, J.S. and Boeke, J.D. (2011) Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. *Mob. DNA*, **2**, 2.
79. Sokolowski, M., Chynces, M., deHaro, D., Christian, C.M. and Belancio, V.P. (2017) Truncated ORF1 proteins can suppress LINE-1 retrotransposition in trans. *Nucleic Acids Res.*, **45**, 5294–5308.
80. Kines, K.J., Sokolowski, M., deHaro, D.L., Christian, C.M. and Belancio, V.P. (2014) Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic Acids Res.*, **42**, 10488–10502.
81. Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell Biol.*, **21**, 1973–1985.
82. Belancio, V.P., Hedges, D.J. and Deininger, P. (2006) LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.*, **34**, 1512–1521.
83. Belancio, V.P., Roy-Engel, A.M. and Deininger, P. (2008) The impact of multiple splice sites in human L1 elements. *Gene*, **411**, 38–45.
84. Belancio, V.P., Whelton, M. and Deininger, P. (2007) Requirements for polyadenylation at the 3’ end of LINE-1 elements. *Gene*, **390**, 98–107.
85. Doucet, A.J., Wilusz, J.E., Miyoshi, T., Liu, Y. and Moran, J.V. (2015) A 3’ poly(a) tract is required for LINE-1 retrotransposition. *Mol. Cell*, **60**, 728–741.
86. Han, J.S. and Boeke, J.D. (2005) LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays*, **27**, 775–784.
87. Criscione, S.W., Theodosakis, N., Micevic, G., Cornish, T.C., Burns, K.H., Neretti, N. and Rodic, N. (2016) Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genomics*, **17**, 463.
88. Kaul, T., Morales, M.E., Smither, E., Baddoo, M., Belancio, V.P. and Deininger, P. (2019) RNA next-generation sequencing and a bioinformatics pipeline to identify expressed LINE-1s at the locus-specific level. *J. Vis. Exp.*, **147**, <https://doi.org/10.3791/59771>.
89. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of chip-Seq (MACS). *Genome Biol.*, **9**, R137.
90. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
91. Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*, **27**, 1571–1572.
92. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Rusczycki, B. *et al.*

- (2015) CTCF-Mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
93. Kuhn, R.M., Haussler, D. and Kent, W.J. (2013) The UCSC genome browser and associated tools. *Brief Bioinform*, **14**, 144–161.
 94. Perepelitsa-Belancio, V. and Deininger, P. (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.*, **35**, 363–366.
 95. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.*, **31**, 833–838.
 96. Urrutia, R. (2003) KRAB-containing zinc-finger repressor proteins. *Genome Biol.*, **4**, 231.
 97. El-Sawy, M., Kale, S.P., Dugan, C., Nguyen, T.Q., Belancio, V., Bruch, H., Roy-Engel, A.M. and Deininger, P.L. (2005) Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. *J. Mol. Biol.*, **354**, 246–257.
 98. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via cas9. *Science*, **339**, 823–826.
 99. Dewannieux, M., Esnault, C. and Heidmann, T. (2003) LINE-mediated retrotransposition of marked alu sequences. *Nat. Genet.*, **35**, 41–48.
 100. Penzkofer, T., Jager, M., Figlerowicz, M., Badge, R., Mundlos, S., Robinson, P.N. and Zemojtel, T. (2017) L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res.*, **45**, D68–D73.
 101. Gasior, S.L., Wakeman, T.P., Xu, B. and Deininger, P.L. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.*, **357**, 1383–1393.
 102. Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D. et al. (2020) Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.*, **52**, 306–319.
 103. Buenrostro, J.D., Wu, B., Chang, H.Y. and Greenleaf, W.J. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.*, **109**, 21.29.1–21.29.9.
 104. Luo, R.X. and Dean, D.C. (1999) Chromatin remodeling and transcriptional regulation. *J. Natl. Cancer Inst.*, **91**, 1288–1294.
 105. Petty, E. and Pillus, L. (2013) Balancing chromatin remodeling and histone modifications in transcription. *Trends Genet.*, **29**, 621–629.
 106. Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., Kumar, S., Ongen, H., Popadin, K., Marbach, D. et al. (2019) Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science*, **364**, eaat8266.
 107. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M. et al. (2018) The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.*, **19**, 151.
 108. Hansen, A.S., Cattoglio, C., Darzacq, X. and Tjian, R. (2018) Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*, **9**, 20–32.
 109. Kentepozidou, E., Aitken, S.J., Feig, C., Stefflova, K., Ibarra-Soria, X., Odom, D.T., Roller, M. and Flicek, P. (2020) Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.*, **21**, 5.
 110. Pitkanen, E., Cajuso, T., Katainen, R., Kaasinen, E., Valimaki, N., Palin, K., Taipale, J., Aaltonen, L.A. and Kilpivaara, O. (2014) Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. *Oncotarget*, **5**, 853–859.
 111. Lanciano, S. and Cristofari, G. (2020) Measuring and interpreting transposable element expression. *Nat. Rev. Genet.*, **21**, 721–736.
 112. Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D. and Kazazian, H.H. Jr (1994) A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.*, **7**, 143–148.
 113. Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E.P.R.I., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J. et al. (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*, **12**, 326–328.
 114. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
 115. Cordaux, R., Hedges, D.J., Herke, S.W. and Batzer, M.A. (2006) Estimating the retrotransposition rate of human alu elements. *Gene*, **373**, 134–137.
 116. Jin, Y., Tam, O.H., Paniagua, E. and Hammell, M. (2015) TETranscripts: a package for including transposable elements in differential expression analysis of RNA-Seq datasets. *Bioinformatics*, **31**, 3593–3599.
 117. Alves, G., Tatro, A. and Fanning, T. (1996) Differential methylation of human LINE-1 retrotransposons in malignant cells. *Gene*, **176**, 39–44.
 118. Jonsson, M.E., Ludvik Brattas, P., Gustafsson, C., Petri, R., Yudovich, D., Pircs, K., Verschuere, S., Madsen, S., Hansson, J., Larsson, J. et al. (2019) Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.*, **10**, 3182.
 119. Sultana, T., van Essen, D., Siol, O., Bailly-Bechet, M., Philippe, C., Zine El Aabidine, A., Pioger, L., Nigumann, P., Sacconi, S., Andrau, J.C. et al. (2019) The landscape of L1 retrotransposons in the human genome is shaped by Pre-insertion sequence biases and Post-insertion selection. *Mol. Cell*, **74**, 555–570.
 120. Konkel, M.K., Wang, J., Liang, P. and Batzer, M.A. (2007) Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. *Gene*, **390**, 28–38.
 121. Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A. and Swergold, G.D. (2000) Reading between the LINES: human genomic variation induced by LINE-1 retrotransposition. *Genome Res.*, **10**, 1496–1508.
 122. Beck, C.R., Garcia-Perez, J.L., Badge, R.M. and Moran, J.V. (2011) LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.*, **12**, 187–215.
 123. Payer, L.M. and Burns, K.H. (2019) Transposable elements in human genetic disease. *Nat. Rev. Genet.*, **20**, 760–772.
 124. Belancio, V.P., Roy-Engel, A.M., Pochampally, R.R. and Deininger, P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.*, **38**, 3909–3922.
 125. Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S. et al. (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell*, **153**, 101–111.
 126. Scott, E.C. and Devine, S.E. (2017) The role of somatic L1 retrotransposition in human cancers. *Viruses*, **9**, 131.
 127. Boissinot, S., Entezam, A. and Furano, A.V. (2001) Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.*, **18**, 926–935.
 128. Boissinot, S. and Furano, A.V. (2001) Adaptive evolution in LINE-1 retrotransposons. *Mol. Biol. Evol.*, **18**, 2186–2194.
 129. Boissinot, S., Davis, J., Entezam, A., Petrov, D. and Furano, A.V. (2006) Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 9590–9594.
 130. Stow, E.C., Kaul, T., deHaro, D.L., Dem, M.R., Beletsky, A.G., Morales, M.E., Du, Q., LaRosa, A.J., Yang, H., Smither, E. et al. (2021) Organ-, sex- and age-dependent patterns of endogenous L1 mRNA expression at a single locus resolution. *Nucleic Acids Res.*, **49**, 5813–5831.