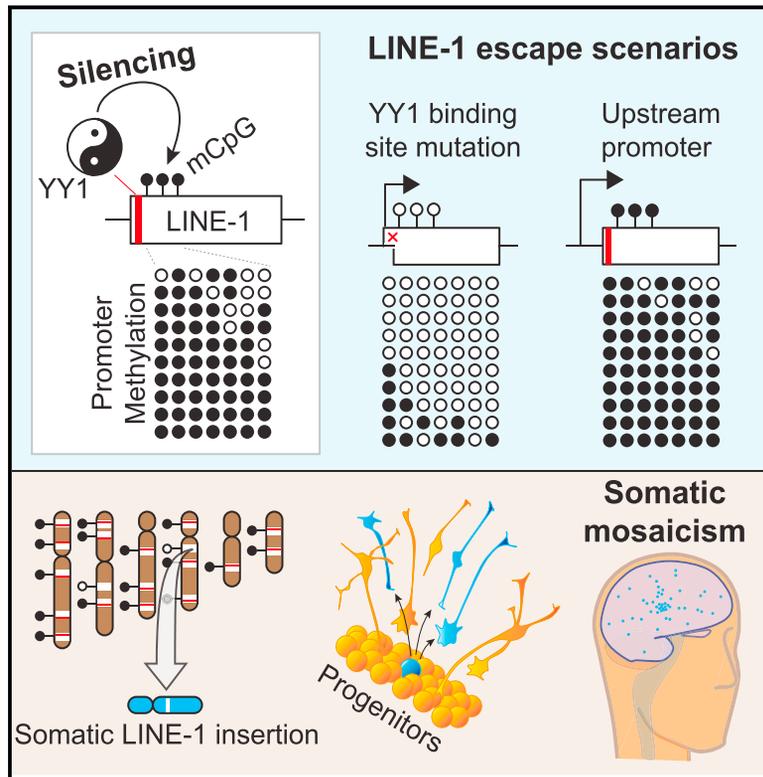


LINE-1 Evasion of Epigenetic Repression in Humans

Graphical Abstract



Authors

Francisco J. Sanchez-Luque,
Marie-Jeanne H.C. Kempen,
Patricia Gerdes, ...,
Jose L. Garcia-Perez, Adam D. Ewing,
Geoffrey J. Faulkner

Correspondence

francisco.sanchezluque@
mater.uq.edu.au (F.J.S.-L.),
faulknergj@gmail.com (G.J.F.)

In Brief

Somatic genome mosaicism occurs in human neurons. Sanchez-Luque et al. show that a conserved YY1 binding site mediates methylation of young L1 retrotransposons. L1s lacking an intact YY1 site escape methylation in embryonic stem cells and can retrotranspose during hippocampal neurodevelopment.

Highlights

- Single-cell genomic analysis of hippocampal neurons reveals a somatic L1 insertion
- The donor L1 is slightly 5' truncated and lacks a conserved YY1 binding site
- Young L1s with truncated or mutated YY1 binding sites are globally hypomethylated
- L1 is able to mobilize in the brain because of locus-specific exceptions to repression



LINE-1 Evasion of Epigenetic Repression in Humans

Francisco J. Sanchez-Luque,^{1,2,*} Marie-Jeanne H.C. Kempen,^{1,3} Patricia Gerdes,¹ Dulce B. Vargas-Landin,^{4,5} Sandra R. Richardson,¹ Robin-Lee Troskie,¹ J. Samuel Jesuadian,¹ Seth W. Cheetham,¹ Patricia E. Carreira,¹ Carmen Salvador-Palomeque,¹ Marta García-Cañadas,² Martin Muñoz-Lopez,² Laura Sanchez,² Mischa Lundberg,¹ Angela Macia,⁶ Sara R. Heras,^{2,7} Paul M. Brennan,⁸ Ryan Lister,^{4,5} Jose L. Garcia-Perez,^{2,3} Adam D. Ewing,¹ and Geoffrey J. Faulkner^{1,9,10,*}

¹Mater Research Institute, University of Queensland, TRI Building, Woolloongabba, QLD 4102, Australia

²GENYO Centre for Genomics and Oncological Research, Pfizer University of Granada, Andalusian Regional Government, Avda Ilustración, 114, PTS Granada 18016, Spain

³MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK

⁴Australian Research Council Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, the University of Western Australia, Perth, WA 6009, Australia

⁵Harry Perkins Institute of Medical Research, Perth, WA 6009, Australia

⁶Department of Pediatrics/Rady Children's Hospital San Diego, School of Medicine, University of California, San Diego, La Jolla, CA, USA

⁷Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, University of Granada, Campus Universitario de Cartuja, 18071 Granada, Spain

⁸Edinburgh Cancer Research Centre, Western General Hospital, Edinburgh, EH4 2XR, UK

⁹Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia

¹⁰Lead Contact

*Correspondence: francisco.sanchezluque@mater.uq.edu.au (F.J.S.-L.), faulknergj@gmail.com (G.J.F.)

<https://doi.org/10.1016/j.molcel.2019.05.024>

SUMMARY

Epigenetic silencing defends against LINE-1 (L1) retrotransposition in mammalian cells. However, the mechanisms that repress young L1 families and how L1 escapes to cause somatic genome mosaicism in the brain remain unclear. Here we report that a conserved Yin Yang 1 (YY1) transcription factor binding site mediates L1 promoter DNA methylation in pluripotent and differentiated cells. By analyzing 24 hippocampal neurons with three distinct single-cell genomic approaches, we characterized and validated a somatic L1 insertion bearing a 3' transduction. The source (donor) L1 for this insertion was slightly 5' truncated, lacked the YY1 binding site, and was highly mobile when tested *in vitro*. Locus-specific bisulfite sequencing revealed that the donor L1 and other young L1s with mutated YY1 binding sites were hypomethylated in embryonic stem cells, during neurodifferentiation, and in liver and brain tissue. These results explain how L1 can evade repression and retrotranspose in the human body.

INTRODUCTION

Retrotransposons are mobile genetic elements that must evade host genome defenses to replicate and survive (Kazazian and Moran, 2017). Long interspersed element 1 (LINE-1 or L1) is the only extant autonomous human retrotransposon (Mills et al., 2007). A full-length L1 mRNA is ~6 kb long, polyadeny-

lated, and encodes two proteins (ORF1p and ORF2p) that catalyze retrotransposition via target-primed reverse transcription (TPRT) (Feng et al., 1996; Luan et al., 1993; Moran et al., 1996; Figure 1A). Nearly all L1 copies are immobile because of 5' truncation and open reading frame (ORF)-disabling mutations. Of 500,000 reference genome L1s, only ~100 are full-length with intact ORFs, and fewer than 10 per individual hold significant retrotransposition potential (Beck et al., 2010; Brouha et al., 2003). These "hot" donor (source) L1s are almost all members of the L1-Ta family and, together, generate one new germline insertion per ~150 births (Brouha et al., 2003; Ewing and Kazazian, 2010). Heritable L1 insertions arise in the early embryo or germline and can cause sporadic genetic disease (Richardson et al., 2017; van den Hurk et al., 2007). Somatic L1 retrotransposition has been observed in the neuronal lineage (Baillie et al., 2011; Coufal et al., 2009; Erwin et al., 2016; Evrony et al., 2012, 2015; Macia et al., 2017; Muotri et al., 2005; Upton et al., 2015) and in tumor cells (Ewing et al., 2015; Iskow et al., 2010; Lee et al., 2012; Nguyen et al., 2018; Scott et al., 2016; Tubio et al., 2014) but is of unresolved biological significance (Burns, 2017; Faulkner and Garcia-Perez, 2017; Scott and Devine, 2017).

Epigenetic and transcriptional silencing guard against L1-mediated mutagenesis (Castro-Diaz et al., 2014; de la Rica et al., 2016; Muotri et al., 2010; Walter et al., 2016), causing L1 to engage in an evolutionary arms race with repressive host factors (Goodier, 2016; Jacobs et al., 2014). The L1 5' UTR is pivotal in this conflict. Its initial 100 nt contains an internal promoter driving L1 mRNA transcription initiation (Swergold, 1990). DNA methylation of an adjacent CpG island regulates this promoter (Hata and Sakaki, 1997; Muotri et al., 2010), as do various transcription factors, including Yin Yang 1 (YY1), RUNX3, and SOX2 (Athaniar et al., 2004; Coufal et al., 2009; Yang et al., 2003). L1 methylation is established during



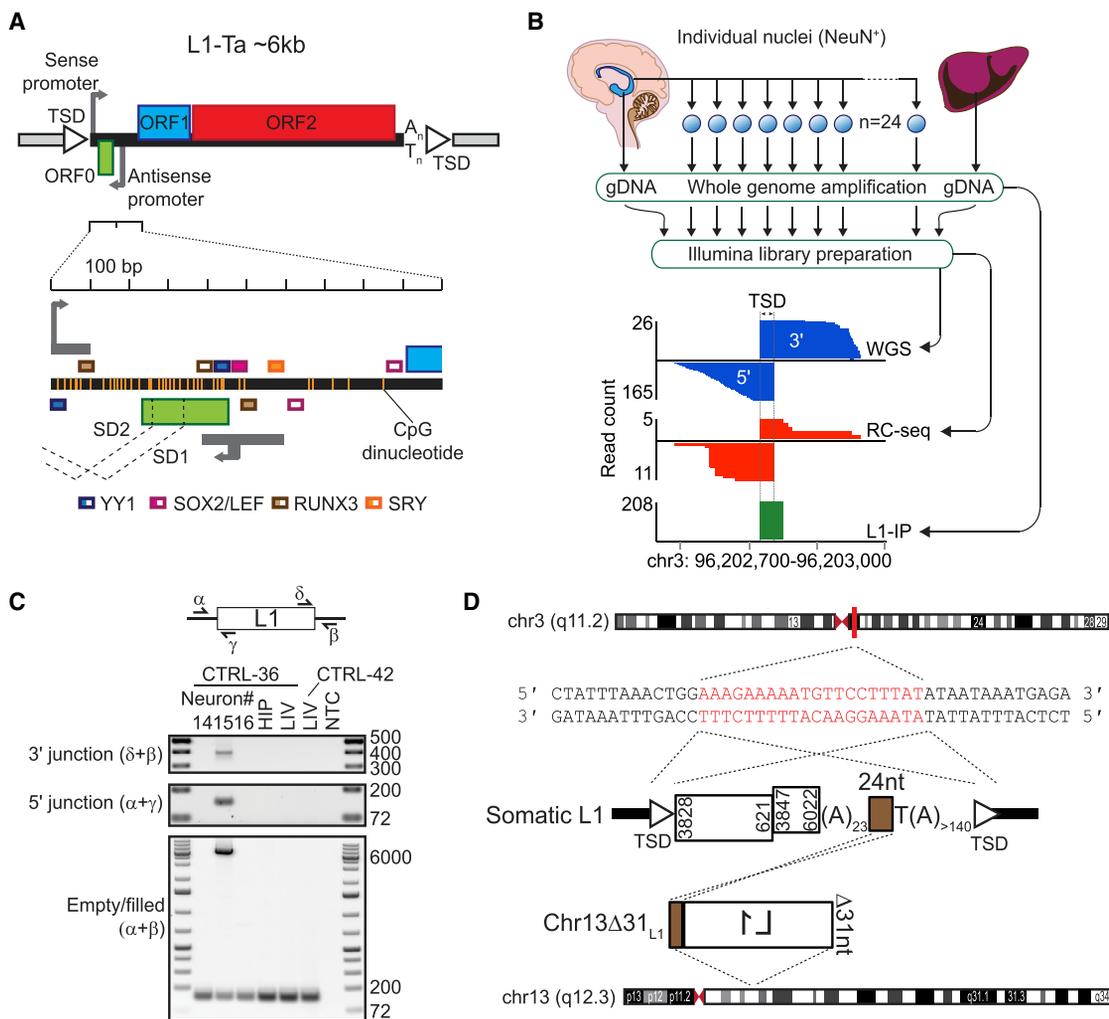


Figure 1. Somatic L1 Insertion Detection and Characterization

(A) Human L1-Ta features. In the magnified 5' UTR view (bottom), SD1 and SD2 represent splice donor sites within ORF0 that can splice into upstream antisense transcript exons. Transcription factor binding sites are represented as boxes above (sense) or below (antisense) the 5' UTR. Solid boxes represent experimentally validated sites. Orange strokes represent CpG dinucleotides.

(B) An integrated genomic approach to detect somatic L1 insertions in hippocampal neurons. Bulk DNA from the hippocampus and liver and from 24 MDA-amplified hippocampal neurons was analyzed with Illumina WGS, RC-seq, and L1-IP. A somatic L1 insertion was found on chromosome 3 in neuron 15 by each approach. Reads spanning the 5' or 3' L1-genome junctions of this event are shown.

(C) PCR validation of a somatic L1 insertion found in CTRL-36 neuron 15. Primers flanking the L1 boundaries (α , δ , γ , and β) were used to amplify the L1 3' junction ($\delta+\beta$), 5' junction ($\alpha+\gamma$), and complete sequence ($\alpha+\beta$). CTRL-36 templates included WGA material from neurons 14–16 as well as bulk hippocampus (HIP) and liver (LIV) genomic DNA (gDNA). Reactions involving CTRL-42 liver gDNA and no-template control (NTC) were also performed.

(D) Complete characterization of the somatic L1 insertion via capillary sequencing. Integration-site nucleotides highlighted in red correspond to the target site duplication (TSD). The L1 was 5' truncated and inverted, as represented by L1-Ta consensus position numerals inside the L1. A 3' transduction (brown box) indicated a donor L1 on chromosome 13 (Chr13 Δ 31_{L1}).

See [Figure S1](#) and [Tables S1](#) and [S2](#) for further PCR validation details.

embryogenesis (Castro-Diaz et al., 2014; de la Rica et al., 2016) and is strongly maintained in somatic tissues (Coufal et al., 2009; Macia et al., 2017; Schauer et al., 2018; Shukla et al., 2013). Given this repression, it is unclear how L1 achieves retrotransposition in the neuronal lineage.

Here we find that a highly conserved YY1 binding site mediates L1 promoter DNA methylation. Exceptions to this repression during neurodifferentiation and in mature tissues appear to govern which L1s mobilize in the brain. Our results suggest

that the YY1 binding site has guarded against L1 retrotransposition over at least the last 70 million years of human evolution.

RESULTS

An Integrated Single-Cell Genomic Analysis of Human Hippocampal Neurons

To identify somatic L1 insertions, we isolated 24 single NeuN⁺ neuronal nuclei from the post-mortem hippocampus of an

individual (female, 18 years old) without evidence of neurological disease (CTRL-36). For each nucleus, we then performed whole-genome amplification (WGA) via multiple displacement amplification (MDA), followed by $\sim 47 \times$ Illumina whole-genome sequencing (WGS), retrotransposon capture sequencing (RC-seq), and L1 insertion profiling (L1-IP) (Table S1). RC-seq employs sequence capture to enrich Illumina libraries for reads spanning L1-Ta 5' and 3' genomic junctions, whereas L1-IP uses PCR to amplify the 3' genomic flank of L1-Ta copies prior to Illumina library preparation (Evrony et al., 2012; Ewing and Kazazian, 2010; Upton et al., 2015). Bulk hippocampus and liver genomic DNA from CTRL-36 were analyzed with $94 \times$ and $49 \times$ WGS, respectively, as well as with RC-seq and L1-IP. Candidate L1 insertions robustly identified by WGS, RC-seq, and L1-IP in at least one neuron, but absent from the liver, were annotated as somatic events (Figure 1B). Following these requirements, we detected one somatic L1 insertion on chromosome 3 in neuron 15 (Figure 1C; Table S2). Capillary sequencing of the entire integration site revealed a 5.4-kb L1-Ta insertion with a 5' inversion (Ostertag and Kazazian, 2001) and carrying a 24 nt 3' transduction (Goodier et al., 2000; Moran et al., 1999; Pickeral et al., 2000) followed by a more than 140 nt pure poly(A) tract (Figure 1D). The insertion presented a degenerate L1 endonuclease cleavage site (5'-CTTT/CC) and yielded a 20 nt target site duplication (TSD). These features were consistent with TPRT-mediated L1 retrotransposition (Jurka, 1997; Luan et al., 1993).

We next attempted to PCR amplify and capillary sequence the entire somatic L1 insertion (empty-filled site reaction), its 5' L1-genome junction, and its 3' transduction-genome junction (Figure S1A) in an extended panel of CTRL-36 hippocampal neurons. In the 24 MDA-amplified neurons subjected to genomic analysis, the filled site was only detected in neuron 15, whereas the 3' junction was detected in 4 additional neurons (Figures S1B–S1D). In an additional 24 MDA-amplified neurons, the 5' and 3' junctions were found only in neuron 36 (Figures S1B–S1D). In a third set of 24 neurons, amplified via the Multiple Annealing and Looping Based Amplification Cycles (MALBAC) protocol (Zong et al., 2012), either the 5' or 3' junction was found in 5 neurons (Figure S1E). The L1 insertion poly(A) tract length varied among the neurons where it was detected and followed a bimodal distribution, clustering around ~ 130 nt and ~ 65 nt (Figures S1C, S1E, and S1F), corroborating reports of L1 poly(A) tract shortening during cell division (Evrony et al., 2015; Grandi et al., 2013; Richardson et al., 2017). The somatic L1 insertion was therefore present in many CTRL-36 hippocampal neurons and likely arose in a neuronal progenitor cell.

To assess the sensitivity of our single-cell genomic analysis, we capillary sequenced the 3' junction of 42 heterozygous germline L1s carried by CTRL-36 (Figures S1G and S1J; Table S2). We observed a paucity of long, pure poly(A) tracts (Figure S1H) that usually accompany new L1 insertions (Evrony et al., 2015; Richardson et al., 2017; Scott et al., 2016). On average, 71.8% and 22.2% of the heterozygous L1s were detected by WGS applied to bulk liver and each single neuron, respectively (Figure S1I), at the detection thresholds we applied to call somatic L1 insertions (≥ 8 reads at each 5' and 3' L1-genome junction). Heterozygous L1s with pure poly(A) tracts were only detected by single-cell WGS with an average sensitivity of 15.3%, a rate

significantly lower than for the remaining heterozygous L1s (24.1%) ($p < 0.0055$, Fisher's exact test). Single-cell WGA and, to a lesser extent, pure poly(A) tracts could influence detection sensitivity for somatic L1 insertions. Although the 3' junction of the heterozygous L1 with the longest (90 nt) poly(A) tract could be PCR amplified in $\sim 80\%$ of the expanded panel of 48 MDA-amplified neurons (Figure S1J), the filled site was detected in only $\sim 33\%$ of the MDA-amplified neurons (Figure S1K). The false negative rate of detection and PCR validation at this standard of evidence may therefore be relatively high. Overall, the somatic L1 insertion was detected and the empty-filled site was PCR validated in neuron 15 but was likely present in $\sim 25\%$ of the hippocampal neurons analyzed with our integrated single-cell genomic approach.

A Somatic Active Hot Donor L1

We traced the somatic L1 insertion 3' transduction to an intergenic L1-Ta located on chromosome 13 and 5' truncated by 31 nt (Figure 1D). Strikingly, this donor L1 (named here Chr13 Δ 31_{L1}) gave rise to a somatic L1 insertion found in the cortex of another individual (Evrony et al., 2015) and was inactive when tested previously for retrotransposition *in vitro* (Brouha et al., 2003). Among CTRL-36 and 7 other unrelated people, we characterized three allelic variants (numbered 1–3) of Chr13 Δ 31_{L1} (Figures 2A and S2A). Chr13 Δ 31_{L1} was present in 7 of 8 individuals (Figure 2A). CTRL-36 was heterozygous for Chr13 Δ 31_{L1} and carried only allele 1. Allele 1 encoded intact ORF1 and ORF2 sequences, whereas alleles 2 and 3, respectively, carried stop codon (C5164T/Q1059 \emptyset) and missense (A2036G/N16S) mutations likely to disable ORF2p activity (Moran et al., 1996; Weichenrieder et al., 2004).

To test the retrotransposition efficiency of each Chr13 Δ 31_{L1} allele, we employed two L1 reporter assays based on the activation of an antibiotic resistance or fluorescence cassette upon retrotransposition, with L1 transcription driven by its native promoter or a cytomegalovirus promoter (CMVp) (Moran et al., 1996; Ostertag et al., 2000). In these assays, Chr13 Δ 31_{L1} alleles 2 and 3 were totally or nearly immobile, whereas allele 1 retrotransposed at $\sim 40\%$ and $\sim 20\%$ of a hot L1 (L1.3) positive control (Sassaman et al., 1997) in HeLa and HEK293T cells, respectively (Figures 2B and S2B). Restoration of the 31 nt 5' truncated sequence to allele 1 elevated its activity above that of L1.3, as did the presence of the CMVp (Figure 2B). We then tested each Chr13 Δ 31_{L1} allele in PA-1 embryonic carcinoma cells, which silence newly mobilized L1 reporter cassettes unless treated with trichostatin A (Garcia-Perez et al., 2010). Allele 1 was not active in PA-1 cells unless the 31 nt truncated sequence was restored (Figure 2C). Consistently, a luciferase promoter reporter assay indicated that all three Chr13 Δ 31_{L1} alleles were transcriptionally active in HeLa and HEK293T cells and not in PA-1 cells (Figures 2D and S2C). The endogenous Chr13 Δ 31_{L1} promoter was thus active in some cell types despite its 5' truncation, providing potential for the retrotransposition-competent allele 1 to mobilize *in vivo*.

Slightly 5'-Truncated L1s Evade DNA Methylation

We hypothesized that incomplete epigenetic repression enabled Chr13 Δ 31_{L1} somatic retrotransposition. We therefore developed

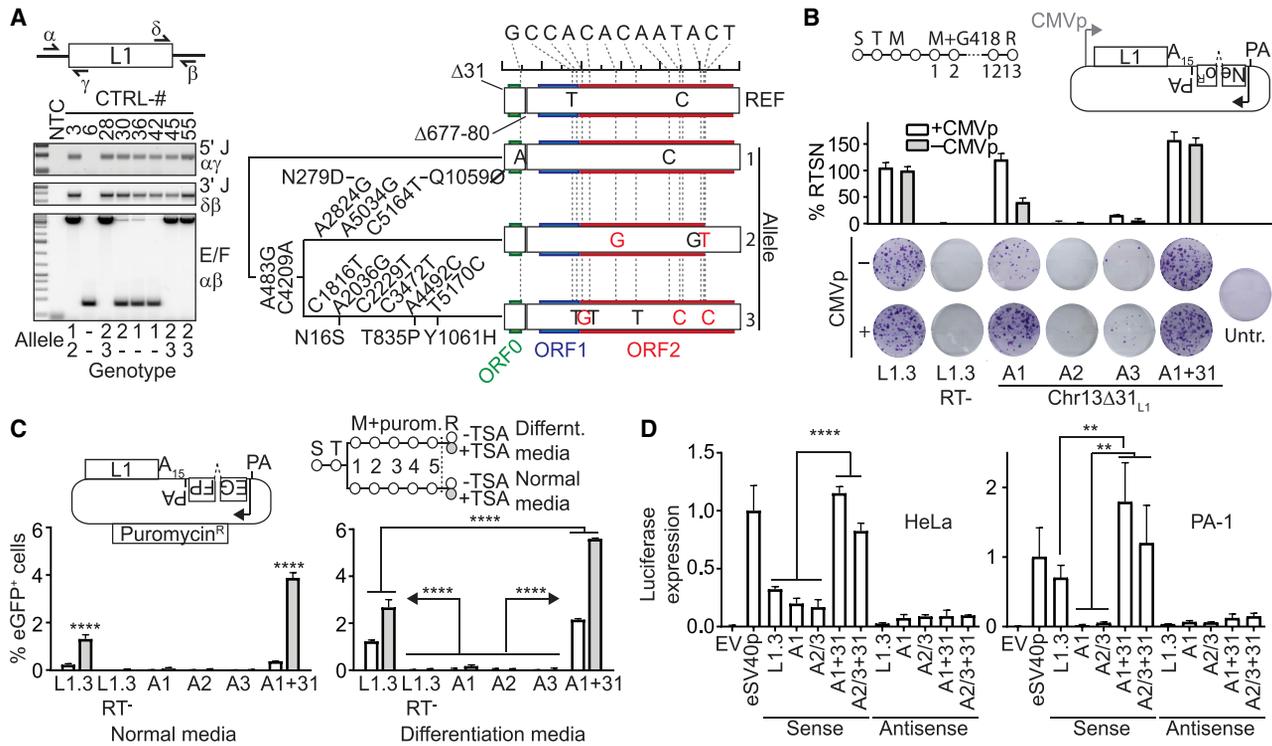


Figure 2. Chr13 Δ 31 L_1 Allele Retrotransposition Activity

(A) Chr13 Δ 31 L_1 genotype among 8 individuals (CTRL-#). Three Chr13 Δ 31 L_1 alleles in this cohort were resolved by capillary sequencing. Their relationship based on sequence similarity is shown in the cladogram. Nucleotide variants among the three alleles and the reference genome (REF) allele, compared with the L1-Ta consensus (top), are shown. Non-synonymous mutations are highlighted in red.

(B) Chr13 Δ 31 L_1 alleles in a cultured HeLa cell retrotransposition assay (Moran et al., 1996). The experimental approach involving neomycin (G418) selection is summarized at the top (S, seeding; T, transfection; M, change of media; R, result analysis; PA, polyadenylation signal; CMVp, CMV promoter; numbers represent days of treatment with antibiotic). Elements were tested for retrotransposition efficiency (RTSN) with and without CMVp and included positive (L1.3) and negative controls (L1.3 RT-), Chr13 Δ 31 L_1 alleles 1–3 (A1, A2, and A3), and allele 1 with its 5' truncation restored (A1+31). Histogram values were normalized to L1.3 (+CMVp). Representative well pictures, including an untransfected control, are shown.

(C) Chr13 Δ 31 L_1 allele retrotransposition, assayed as in (B) except using an EGFP-based reporter system with puromycin selection (Ostertag et al., 2000), in differentiating and non-differentiating PA-1 cells. Gray and white bars represent cells treated or not treated, respectively, with trichostatin A (TSA), which is known to release the EGFP reporter from silencing (Garcia-Perez et al., 2010).

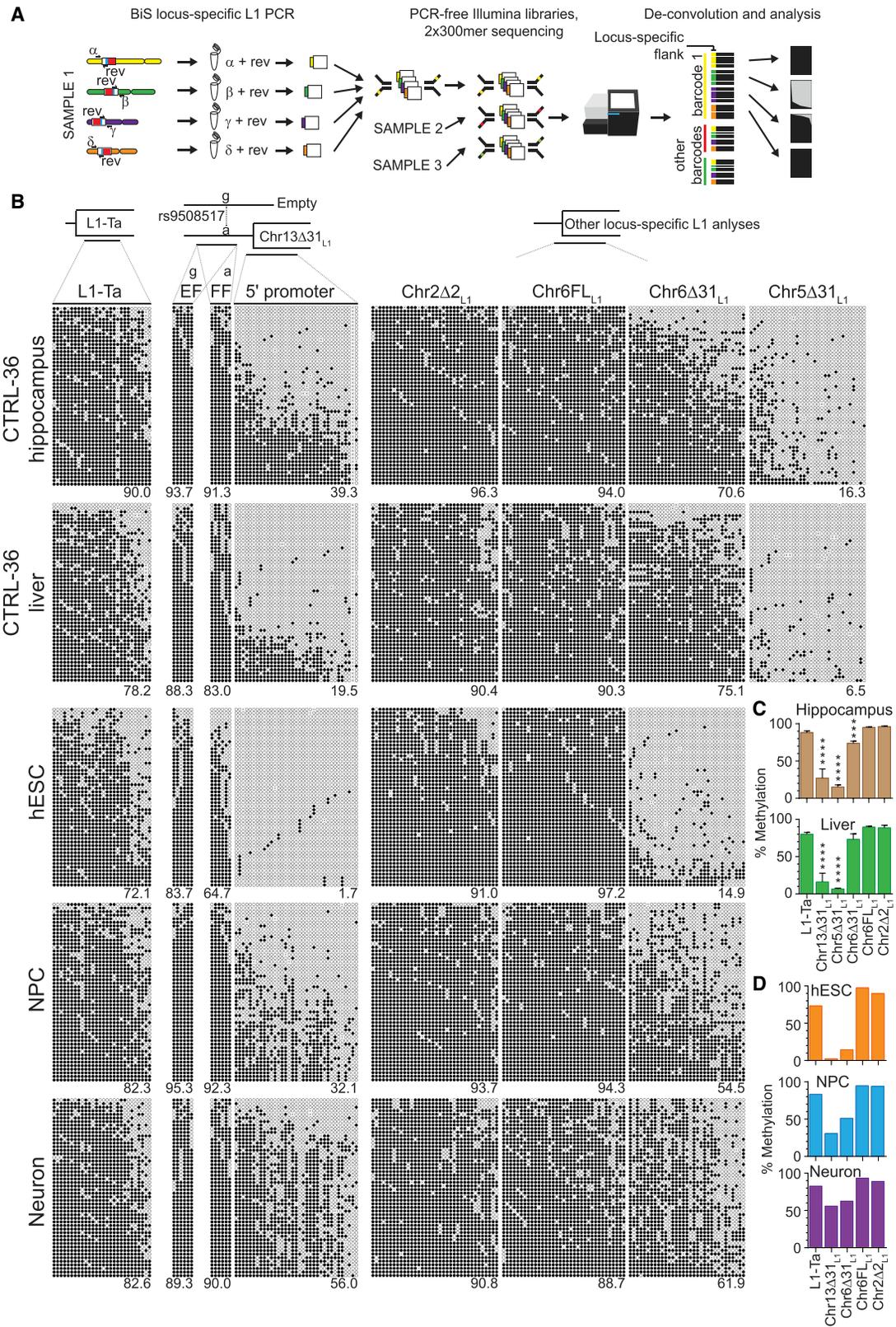
(D) Dual-luciferase promoter reporter assay for Chr13 Δ 31 L_1 alleles in sense and antisense orientation in HeLa and PA-1 cell lines. Histogram values were normalized to the positive control enhanced SV40 promoter (eSV40p). EV, empty vector; A1, Chr13 Δ 31 L_1 allele 1; A2/3, alleles 2 and 3 (identical sequences); A1+31 and A2/3+31, alleles with 5' truncation restored.

In (B)–(D), data are represented as mean \pm SD (** p < 0.01 and **** p < 0.0001). See Figure S2 for Chr13 Δ 31 L_1 genotyping and L1 reporter assays in HEK293T cells.

a PCR-free bisulfite sequencing strategy to measure L1 locus-specific DNA methylation as well as L1-Ta family methylation genome-wide (Figure 3A). Paired-end 300-mer Illumina sequencing allowed higher throughput and wider resolution of the L1 5' UTR CpG island compared with prior approaches (Coufal et al., 2009; Klawitter et al., 2016; Scott et al., 2016; Tubio et al., 2014; Wissing et al., 2012). We found 90.0% and 78.2% L1-Ta family methylation in the CTRL-36 hippocampus and liver, respectively (Figure 3B). By contrast, the Chr13 Δ 31 L_1 promoter was 39.3% and 19.5% methylated in the hippocampus and liver, respectively, with numerous fully demethylated sequences in each tissue (Figure 3B). The only two other CTRL-36 germline L1-Ta copies (Chr5 Δ 31 L_1 and Chr6 Δ 31 L_1) 5' truncated by 31 nt were almost entirely demethylated (Chr5 Δ 31 L_1) or fully demethylated in 5%–10% of cells (Chr6 Δ 31 L_1) (Figure 3B; Data S1A; Table S2). However, two heterozygous, intergenic full-length

germline L1-Ta insertions (Chr6FL L_1 and Chr2 Δ 2 L_1) were almost completely methylated (Figure 3B; Data S1A). We also observed this contrasting pattern in the hippocampus and liver and, where available, in cortex tissue obtained from the remainder of our cohort (Figures S3A–S3D; Data S1A–S1F). Chr13 Δ 31 L_1 was strongly (p < 0.0001, one-way ANOVA with Tukey's multiple comparisons test) demethylated compared with the L1-Ta family in all 7 carrier individuals (Figures 2A, 3C, and S3B). These results suggested that 5'-truncated L1s were hypomethylated in mature human tissues.

Embryonic development witnesses dramatic increases in genome-wide L1 DNA methylation (Castro-Diaz et al., 2014; Coufal et al., 2009; de la Rica et al., 2016; Macia et al., 2017; Salvador-Palomeque et al., 2019; Walter et al., 2016). To assess Chr13 Δ 31 L_1 methylation during neurodevelopment *in vitro*, we conducted L1 bisulfite sequencing on pluripotent H1 human



(legend on next page)

embryonic stem cells (hESCs) as well as H1-derived neuronal progenitor cells (NPCs) and neurons. Genotyping via 43× WGS (Table S1) revealed that Chr13Δ31_{L1}, Chr6FL_{L1}, Chr2Δ2_{L1}, and Chr6Δ31_{L1} were heterozygous in H1 cells, whereas Chr5Δ31_{L1} was absent (Table S2). Overall, the L1-Ta family was 72.1% methylated in hESCs and more strongly methylated in neurons (82.6%), as expected (Coufal et al., 2009; Macia et al., 2017; Salvador-Palomeque et al., 2019; Figures 3B and 3D; Data S1G). The full-length elements Chr6FL_{L1} and Chr2Δ2_{L1} were ~90% methylated in hESCs and during neurodifferentiation (Figure 3B). By contrast, the 5'-truncated elements Chr13Δ31_{L1} and Chr6Δ31_{L1} were 1.7% and 14.9% methylated, respectively, in hESCs and only partially remethylated (~60%) in neuronal cells (Figure 3B). Both DNA strands of the Chr13Δ31_{L1} promoter remained fully unmethylated in ~5% of neurons (Figure 3B; Data S1H). Next we identified a SNP (rs9508517) only present in the 5' genomic flank of each Chr13Δ31_{L1} allele. Bisulfite sequencing of this flank and regions further upstream indicated that it was highly methylated in NPCs, neurons, and brain tissue regardless of whether Chr13Δ31_{L1} was present (Figures 3B, S3C, and S3D; Data S1I). In hESCs, moderate demethylation of the flanking region extended up to 500 bp away from Chr13Δ31_{L1}, when the L1 was present, and formed a methylation "sloping shore" (Figure 3B; Data S1I) observed previously adjacent to retrotransposed CpG islands (Grandi et al., 2015). Overall, these data depicted an element-specific failure to repress Chr13Δ31_{L1} in mature tissues and during neurodevelopment.

A YY1 Binding Site Enables L1 Locus-Specific Promoter Methylation

Given the distinct but consistent DNA methylation patterns observed for full-length and 31 nt 5' truncated L1s, we investigated the degree of 5' truncation required for L1 hypomethylation. We assembled a panel of 28 germline L1-Ta insertions that were full length or 5' truncated up to 31 nt and present in CTRL-36 or the H1 genome (Table S2). We then performed L1 bisulfite sequencing using genomic DNA from CTRL-36 liver and the H1 neurodifferentiation time course. At least ~60%, but generally more than 80%, methylation was observed for L1s that were full length or truncated by less than 14 nt (Figures 4A and 4B; Data S2). Among this group, three highly active full-length L1s, Chr22FL_{L1}-L1.2, ChrXFL_{L1}, and Chr22FL_{L1}-TTC28,

tended to be the least methylated, consistent with prior results (Philippe et al., 2016; Tubio et al., 2014; Wissing et al., 2012). Conversely, of the L1s truncated by 14 nt or more, all apart from Chr6Δ31_{L1} were less than 20% methylated in liver tissue (Figure 4A), and all except Chr1Δ21_{L1}-LRE2 were less than 15% methylated in hESCs (Figure 4B). Almost every fully or near-fully unmethylated sequence was found in elements truncated by 14 nt or more (Figures 4A and 4B), and even Chr1Δ21_{L1}-LRE2 was fully unmethylated in some hESCs, in line with its capacity to mobilize in the germline (Holmes et al., 1994). Further examination revealed frequent non-canonical CpH (H = A/C/T) methylation in hESCs at L1-Ta position +44 (Figures 4C and S4A) in sequences exhibiting high CpG methylation, consistent with *de novo* DNA methyltransferase activity (Gowher and Jeltsch, 2001; Liao et al., 2015). A 5' truncation of 14 nt or more thus demarcated methylated and hypomethylated L1s.

YY1 is a zinc-finger protein (ZFP) that has been shown biochemically to bind L1-Ta positions +12 to +20 and direct transcription initiation to position +1 (Athanihar et al., 2004; Becker et al., 1993). L1s truncated by 14 nt or more therefore lacked at least three nucleotides of this YY1 binding site (Figure 4B), which is conserved in almost all primate L1 lineages found in the human genome (Table S3; Khan et al., 2006). To assess the potential effect of YY1 site sequence variation, we used L1 bisulfite sequencing to analyze methylation of full-length L1-Ta and L1PA2 elements (the latter family becoming only recently immobile in humans; Mills et al., 2007) that carried point mutations in their YY1 motif. Likely because of YY1 site conservation, few such examples were available. However, an L1PA2 copy on chromosome 17 that harbored two YY1 site mutations was found to be far less methylated in hESCs and during neurodifferentiation than the L1PA2 family overall (Figures 4D and S4B). We also found fully unmethylated promoter sequences for two L1PA2 and L1-Ta elements, located, respectively, on chromosomes 5 and 1, carrying single YY1 site mutations (Figure S4B). These examples, alongside our other results, suggested YY1 binding site perturbation via either point mutation or 5' truncation coincided with L1 hypomethylation.

Genome-Wide Young L1 Repression Mediated by YY1

Distinct regulatory programs may repress newly emerged and older L1 families. For example, KAP1 (TRIM28) binds L1 in

Figure 3. Chr13Δ31_{L1} Is Hypomethylated in Human Tissues and Neuronal Lineage Cells

(A) Schematic illustration of the locus-specific, high-throughput analysis of L1 promoter CpG methylation. For each L1, depicted on four different chromosomes, a bisulfite-converted 5' UTR and genomic flank are PCR amplified using a specific primer (α , β , γ , or δ) matched with a common L1 reverse primer (rev). Independent PCR products are combined into a barcoded Illumina library, mixed with libraries similarly generated for other samples, and analyzed via 2× 300-mer sequencing. Genomic flanks are colored to match their chromosome of origin.

(B) Methylation of the overall L1-Ta family, Chr13Δ31_{L1}, two other 31nt 5'-truncated elements (Chr5Δ31_{L1} and Chr6Δ31_{L1}), and two full-length elements (Chr2Δ2_{L1} and Chr6FL_{L1}) in CTRL-36 hippocampus and liver tissues and H1 hESC neurodifferentiation. Each cartoon panel corresponds to an amplicon and displays 50 non-identical sequences (black circle, methylated CpG; white circle, unmethylated CpG; x, mutated CpG) extracted at random from the corresponding and much larger Illumina library. The percentage of methylated CpG is indicated in the lower right corner of each cartoon. The Chr13Δ31_{L1} filled (FF) and empty (EF) allele genomic flanks were discriminated by a linked SNP (rs9508517). Chr5Δ31_{L1} was absent from the H1 genome.

(C) Methylation of the overall L1-Ta family and, where present, Chr13Δ31_{L1}, Chr5Δ31_{L1}, Chr6Δ31_{L1}, Chr2Δ2_{L1}, and Chr6FL_{L1} in the hippocampus and liver of 8 individuals. Data represent the mean percentage of methylation ± SD obtained from 50 random sequences per amplicon and sample (**p < 0.001 and **** < 0.0001).

(D) Methylation of the overall L1-Ta family, Chr13Δ31_{L1}, Chr6Δ31_{L1}, Chr2Δ2_{L1}, and Chr6FL_{L1} during hESC neurodifferentiation, obtained by randomly sampling 1,000 sequences per amplicon and sample.

See Figure S3, Data S1, and Table S2 for additional methylation analysis information.

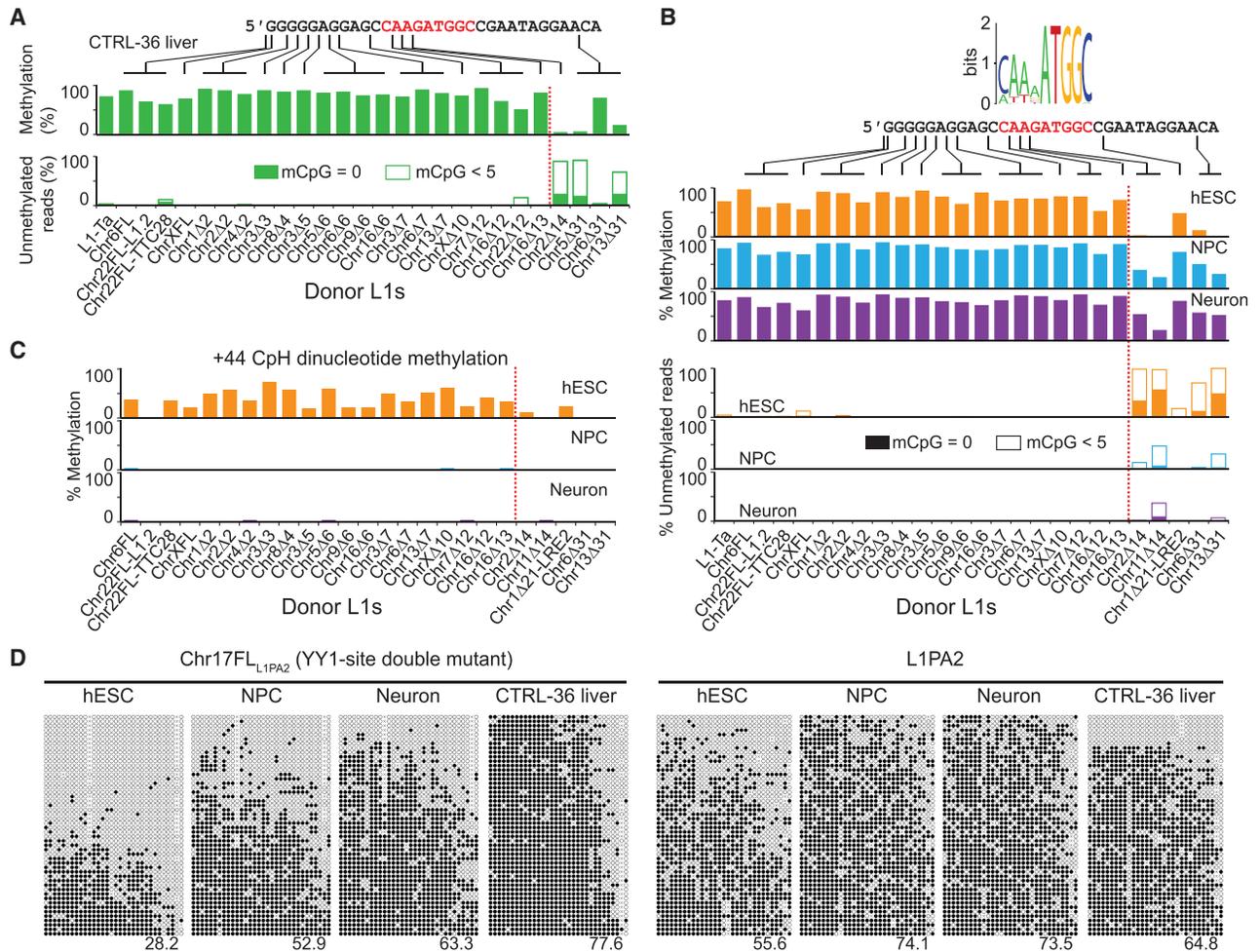


Figure 4. A YY1 Binding Site Mediates L1 Promoter Methylation

(A) Promoter CpG methylation (top graph) and proportion of unmethylated reads (bottom graph) for a cohort of full-length and 5'-truncated L1-Ta elements and the overall L1-Ta family in CTRL-36 liver tissue. Data were obtained via analysis of 50 non-identical random sequences per amplicon. A dotted red line separates L1s 5'-truncated by less than 14 nt or 14 nt or more. The L1 5' end sequence is displayed above the histograms, and the YY1 binding site is shown in red. Chr11 Δ 14_{L1} and Chr1 Δ 21_{L1}-LRE2 were not present in CTRL-36.

(B) As for (A) but displaying data obtained from H1 neurodifferentiation and using 1,000 randomly sampled reads per amplicon, with the exception of Chr22FL_{L1}-TTC28 and Chr1 Δ 21_{L1}-LRE2, which are represented by 50 reads each. A sequence logo for the YY1 binding site (Kim and Kim, 2009) is displayed along with the L1 5' end sequence above the histograms. Chr6 Δ 6_{L1}, Chr22 Δ 12_{L1}, and Chr5 Δ 31_{L1} were not present in the H1 cell line.

(C) CpH methylation level at L1-Ta nucleotide +44 in the 28 L1-Ta elements analyzed in (B) during H1 neurodifferentiation.

(D) Promoter CpG methylation level for the Chr17FL_{L1PA2} YY1 site double mutant and the overall L1PA2 family during hESC neurodifferentiation and in CTRL-36 liver tissue. Each cartoon panel displays 50 non-identical random sequences (black circle, methylated CpG; white circle, unmethylated CpG; x, mutated CpG) matching each amplicon. The percentage of methylated CpG is indicated in the lower right corner of each cartoon.

See Figure 3, Figure S4, Data S2, and Table S4 for supporting L1 methylation data.

hESCs (Figure 5A) and particularly limits expression of the older primate-specific families L1PA3–L1PA6 (Castro-Diaz et al., 2014; Jacobs et al., 2014). YY1 binding, by contrast, is pronounced at the 5' end of the young L1-Ta and L1PA2 families (Figure 5A; Sun et al., 2018), despite conservation of the YY1 motif in older L1 families (Table S3), and is strongly anticorrelated with KAP1 binding ($r = -0.93$, Pearson). As expected, we found that L1-Ta and L1PA2 elements 5' truncated by 14 nt or more were far less bound by YY1 in hESCs than full-length L1s, whereas no difference in KAP1 binding was observed (Figure 5A). Full-length L1s carrying YY1 motif point mutations

were also less likely to bind YY1 than elements with an intact binding site (Figure 5A). We then analyzed published data obtained from HEK293 cells engineered to express GFP-tagged YY1 protein (Schmitges et al., 2016), and again we found that YY1 was heavily bound to L1-Ta and L1PA2 elements (Figure S5A). Consistently, YY1 overexpression in HEK293 cells significantly ($p < 0.05$, two-tailed t test) reduced transcription from only these young L1 families (Figure S5A). These results suggested that if YY1 mediated L1 promoter methylation, then loss of its binding site would principally affect young L1 families.

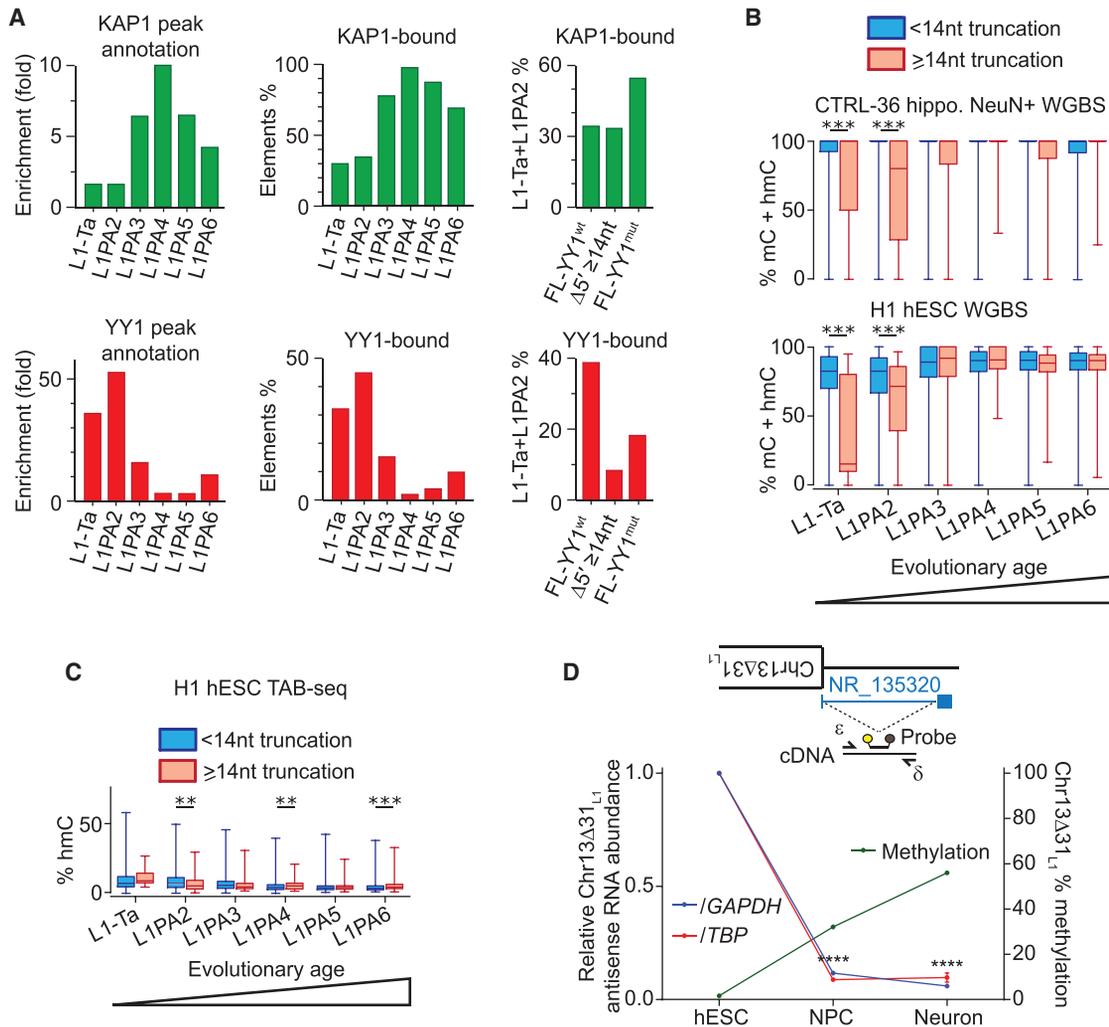


Figure 5. YY1 Mediates Methylation of Young L1 Families

(A) KAP1 binding was enriched across full-length members of older (L1PA3–L1PA6) L1 families, whereas YY1 was bound more strongly to young (L1-Ta and L1PA2) families (left and center). YY1 binding was lower among 5'–truncated young L1s and full-length elements carrying YY1 site mutations than for young L1s with intact YY1 sites (right). KAP1 and YY1 hESC chromatin immunoprecipitation (ChIP-seq) data were obtained from prior studies (ENCODE Project Consortium, 2012; Turelli et al., 2014).

(B) Genome-wide methylcytosine (mC) and hydroxymethylcytosine (hmC) percentages for CpG dinucleotides present in the first 300 bp of L1-Ta and older L1 promoter sequences. Analyses were performed for NeuN⁺ CTRL-36 hippocampal neurons (top) as well as using published H1 hESC data (bottom) (ENCODE Project Consortium, 2012). Boxplots indicate median, quartile, and extrema values for groups of elements 5'–truncated by less than 14 nt or 14 nt or more within each L1 family (****p* < 0.001).

(C) As for (B) but displaying genome-wide hmC percentages obtained using published H1 hESC data (Yu et al., 2012) (***p* < 0.01 and ****p* < 0.001).

(D) Chr13Δ31_{L1} antisense transcript (RefSeq: NR_135320) expression during hESC neurodifferentiation (left y axis) normalized to *GAPDH* (blue) or *TBP* (red). The TaqMan primer-probe design used to quantify NR_135320 abundance is shown above the graph. Primers (ε and δ) flank the probe, which, in turn, spans the (NR_135320) splice junction. Values represent the mean ± SD (***p* < 0.01, ****p* < 0.001, and *****p* < 0.0001). Chr13Δ31_{L1} methylation (green, right y axis) was determined by locus-specific bisulfite sequencing of DNA from the same samples.

See Figure S5 for additional analyses of Chr13Δ31_{L1} antisense transcription and Figure 3 for Chr13Δ31_{L1} bisulfite sequencing results during H1 differentiation.

To test this possibility genome-wide, we performed ~33 × whole-genome bisulfite sequencing (WGBS) on neuronal nuclei isolated from CTRL-36 hippocampal tissue. This analysis encompassed only the initial 300 nt of germline L1s found in the reference genome, where methylation was typically higher than for CpGs further 3' in individual L1 promoters (Figure 3B), and offered a lower resolution than our locus-specific approach. Nonetheless, we determined that full-length members of each

L1 family were more than 90% methylated (Figure 5B), in agreement with prior results (de la Rica et al., 2016). By contrast, L1-Ta and L1PA2 elements truncated by 14 nt or more were significantly less methylated than full-length L1s from the same families (*p* < 0.001, one-way ANOVA with Dunn's multiple comparisons test), whereas older truncated L1s were not hypomethylated (Figure 5B). Repeating this analysis using published H1 hESC WGBS data (ENCODE Project Consortium, 2012), we

again observed widespread methylation of full-length L1s and significant hypomethylation ($p < 0.001$) of only 14 nt or more truncated L1-Ta and L1PA2 sequences (Figure 5B). Because bisulfite sequencing cannot distinguish methylcytosine and hydroxymethylcytosine (hmC), we also analyzed published genome-wide hmC data from H1 cells obtained via Tet-assisted bisulfite sequencing (TAB-seq) (Yu et al., 2012). As reported elsewhere (de la Rica et al., 2016), hmC was low (less than $\sim 10\%$) among each L1 family. The level of hmC was not significantly different among 14 nt or more truncated and full-length L1-Ta copies (Figure 5C) and did not exceed 25% for any individual L1-Ta promoter. Overall, DNA hypomethylation of young L1s with mutant YY1 sites was detected by locus-specific and genome-wide analyses and primarily reflected reduced methylcytosine levels.

Chr13 $\Delta 31_{L1}$ Transcription during Neurodifferentiation

Promoter hypomethylation alone does not demonstrate transcription, and mRNAs transcribed by members of a young L1 family, such as L1-Ta, are difficult to link to a specific L1 copy. However, an antisense promoter (ASP) located at +600 to +400 in the L1 5' UTR can generate chimeric L1 transcripts incorporating unique upstream sequences (Denli et al., 2015; Faulkner et al., 2009; Speek, 2001). L1 ASP activity may therefore serve as a proxy for transcription from the canonical L1 sense promoter (Macia et al., 2011). To assess Chr13 $\Delta 31_{L1}$ ASP activity, we designed primers to target an annotated RNA (RefSeq: NR_135320) antisense to Chr13 $\Delta 31_{L1}$ as well as RNAs initiated from the Chr13 $\Delta 31_{L1}$ ASP and spliced into exons more than 30 kb away (Figure S5B). Using RT-PCR and RNA extracted from differentiating PA-1 cells, we identified various transcripts initiated by the Chr13 $\Delta 31_{L1}$ ASP (Figures S5B and S5C). We then targeted a commonly used splice junction and detected Chr13 $\Delta 31_{L1}$ antisense transcripts expressed in hippocampus or liver tissue from each Chr13 $\Delta 31_{L1}$ carrier in our cohort (Figure S5D). TaqMan qRT-PCR indicated that Chr13 $\Delta 31_{L1}$ antisense transcript abundance and DNA methylation were inversely correlated during hESC neurodifferentiation *in vitro*, including an ~ 10 -fold reduction in expression upon differentiation to NPCs (Figures 5D and S5E). These experiments demonstrated Chr13 $\Delta 31_{L1}$ expression coincident with hypomethylation of its promoter in mature tissues, in hESCs, and during neurodifferentiation *in vitro*.

Locus-Specific Mechanisms of L1 Repression and Escape

Our analyses suggested that a YY1 binding site was generally required for L1-Ta promoter methylation *in vivo*. However, we observed locus-specific exceptions to this pattern. First, a nearly full-length L1-Ta (Chr8 $\Delta 3_{L1}$) located intronic to the *KCNB2* gene was identified earlier as the source of a cortical neuron somatic L1 insertion that carried a 101 nt 5' transduction (Evrony et al., 2012, 2015). In our cohort, Chr8 $\Delta 3_{L1}$ was present only in CTRL-28 and CTRL-42 and as a heterozygous polymorphism in each individual. Locus-specific L1 bisulfite sequencing indicated that Chr8 $\Delta 3_{L1}$ was almost completely methylated in brain and liver tissues (Figures 6A and S6A), consistent with its intact YY1 binding site. *KCNB2* is specifically expressed in the brain (Figure S6B) and was detected here by RNA sequencing (RNA-seq) applied

to hippocampal tissue (Figure S6C). Bisulfite analysis indicated that the region upstream of Chr8 $\Delta 3_{L1}$ was heavily demethylated in brain tissue but not liver (Figures 6A and S6D). A transcript (GenBank: DA461809) spliced shortly upstream of the Chr8 $\Delta 3_{L1}$ 5' end is likely initiated from an annotated promoter (Forrest et al., 2014) in the demethylated flanking region. Crucially, the DA461809 splice junction was used to generate the template RNA for the 5' transduction carried by the cortical L1 insertion traced to Chr8 $\Delta 3_{L1}$ (Figure 6A). We therefore propose that the genomic location of Chr8 $\Delta 3_{L1}$, in a gene expressed in the brain and downstream of a strong promoter element, enabled transcription and retrotransposition of a chimeric DA461809-Chr8 $\Delta 3_{L1}$ mRNA despite methylation of the Chr8 $\Delta 3_{L1}$ promoter.

Another element, Chr22FL $_{L1}$ -TTC28, is a fixed germline L1-Ta (Gardner et al., 2017) located antisense to the first intron of *TTC28*, a gene that is highly expressed in many tissues (Figure S6E). In our cohort, Chr22FL $_{L1}$ -TTC28 was methylated in brain tissues but, despite its intact YY1 binding site, was fully demethylated in a subset of hepatic cells (Figures 6B and S6F). Reciprocally, locus-specific repression may influence young L1s lacking a YY1 binding site. For example, Chr1 $\Delta 21_{L1}$ -LRE2 was abnormally methylated in hESCs and neuronal cells compared with the remaining 5' truncated L1s (Figure 4B; Data S2B). An L1PA13 element was located ~ 2.7 kb upstream of Chr1 $\Delta 21_{L1}$ -LRE2 and incorporated a YY1 binding site (Figure 6C). Methylation of this L1PA13 was complete in hESCs and maintained throughout neurodifferentiation (Figure 6C). We speculate that methylation spreading from the L1PA13 may explain the unusual repression of Chr1 $\Delta 21_{L1}$ -LRE2 (Figures 4B and 6C; Data S2B). Together, Chr8 $\Delta 3_{L1}$, Chr22FL $_{L1}$ -TTC28, and Chr1 $\Delta 21_{L1}$ -LRE2 highlight how YY1-mediated repression may be supplanted occasionally by locus-specific regulatory mechanisms.

DISCUSSION

Our experiments indicate that a highly conserved YY1 binding site is central to L1 repression in pluripotent and differentiated human cells. It is possible that YY1 recruits DNA methyltransferases directly to silence members of the L1-Ta and L1PA2 families (Castro-Diaz et al., 2014; Hervouet et al., 2009; Schlesinger et al., 2013; Tsumura et al., 2006). Genome-wide analyses suggest that YY1 and KAP1 bind distinct L1 families (Castro-Diaz et al., 2014; Sun et al., 2018). KAP1 silences older L1s and other transposable elements by recruiting histone-modifying factors (Castro-Diaz et al., 2014; Ecco et al., 2016; Imbeault et al., 2017; Rowe et al., 2010; Turelli et al., 2014; Wolf et al., 2015; Yang et al., 2017). KAP1 knockdown in hESCs does not significantly alter L1-Ta or L1PA2 expression, whereas knockdown of DNA methyltransferases increases expression of young human and mouse L1s (Castro-Diaz et al., 2014; Tsumura et al., 2006). A general lack of KAP1-associated deposition of repressive H3K9me3 among young L1 families (Castro-Diaz et al., 2014) may explain why YY1 can only access its binding site and mediate DNA methylation of young L1s. Alternative inhibitory pathways (e.g., piwi-interacting RNAs [piRNAs]) may also target the YY1 motif (Aravin et al., 2008; Castro-Diaz et al., 2014; Marchetto et al., 2013). These scenarios are not exclusive, and each involves YY1-dependent DNA methylation.

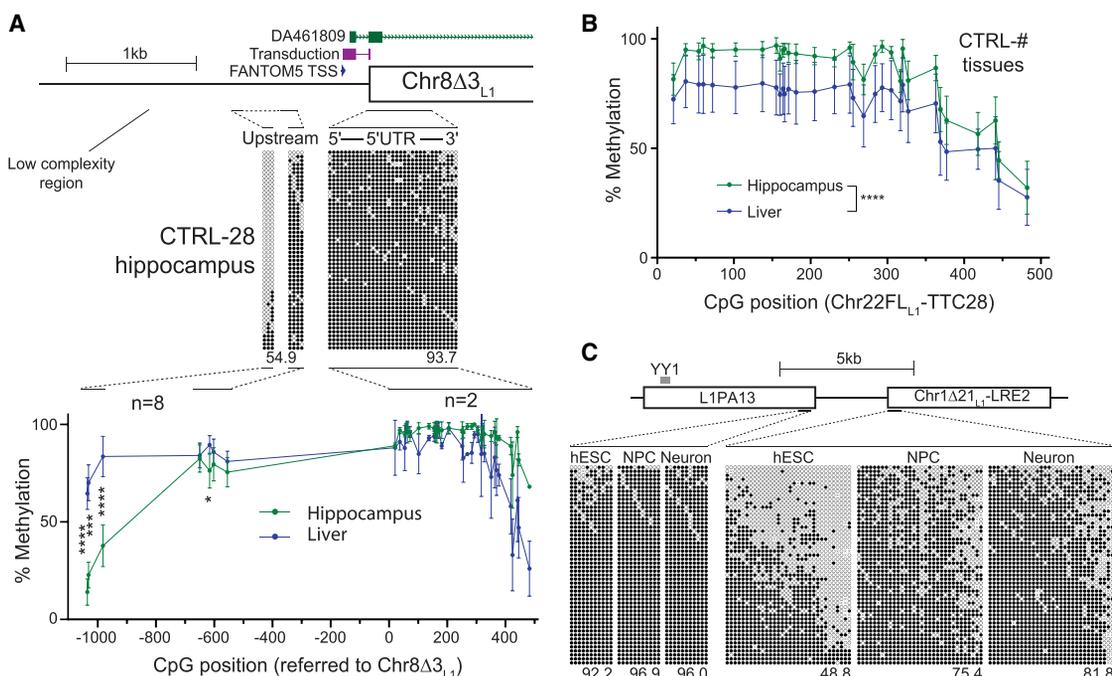


Figure 6. The Genomic Environment Influences Donor L1 Regulation

(A) Chr8 Δ 3_{L1} locus methylation. Top: an expressed sequence tag (EST; GenBank: DA461809) indicated an upstream RNA spliced into Chr8 Δ 3_{L1}, which coincided with a previously reported 5' transduction in a somatic L1 insertion (Evrony et al., 2012). A potential transcription start site (TSS) for the spliced and transduced RNA template was delineated by FANTOM5 (Forrest et al., 2014). Center: Chr8 Δ 3_{L1} promoter and upstream methylation cartoons displaying 50 random, non-identical sequences (black circle, methylated CpG; white circle, unmethylated CpG; x, mutated CpG). The percentage of methylated CpG is indicated in the bottom right corner of each cartoon. Bottom: average Chr8 Δ 3_{L1} promoter methylation in hippocampus and liver tissue from Chr8 Δ 3_{L1} carrier individuals CTRL-28 and CTRL-42 and the upstream region in all 8 individuals. Values represent the mean methylation \pm SD, indicated by 50 random sequences corresponding to each amplicon and sample. Statistical differences were analyzed pairwise between upstream CpG dinucleotides (****p < 0.0001, ***p < 0.001, *p < 0.05).

(B) Chr22FL_{L1}-TTC28 promoter methylation in hippocampus and liver tissue. Data are represented as mean methylation \pm SD in 8 individuals. Chr22FL_{L1}-TTC28 was significantly hypomethylated in liver tissues (****p < 0.0001).

(C) Methylation of the Chr1 Δ 21_{L1}-LRE2 promoter and an \sim 2.7kb upstream L1PA13 copy during hESC neurodifferentiation. As indicated, the L1PA13 sequence contains an intact YY1 binding site utilized in H1 cells. The cartoon panels were generated as in (A).

See Figure S6, Data S2, and Table S2 for supporting L1 methylation data.

YY1 facilitates full-length L1 transcription, and nearly all L1 families active over the last 70 million years of human evolution present a YY1 binding site at their 5' end (Athaniar et al., 2004; Khan et al., 2006). As an activator and repressor, YY1 is an enduring modulator of L1 activity. In turn, L1 is engaged in an evolutionary arms race with host genome defenses. Almost all human L1s have lost this conflict, are immobile, and are controlled by KAP1 and other factors (Imbeault et al., 2017; Jacobs et al., 2014; Liu et al., 2018; Robbez-Masson et al., 2018). Sequence divergence is likely pivotal in L1 eluding complete repression. For example, loss of a 5' UTR binding site for the repressor ZNF93 \sim 12.5 million years ago enabled L1PA3 and younger L1 families to escape from ZNF93 restriction at the cost of a weakened promoter (Jacobs et al., 2014). It is striking, then, that absence of the YY1 site from Chr13 Δ 31_{L1} reduces, but does not abolish, its promoter activity. Numerous L1s lacking the YY1 site may have escaped repression and retrotransposed, as achieved by L1PA3 millions of years ago, and yet failed to spread further in the germline without the YY1 site to provide their progeny with a functional 5' sense promoter. Given enrich-

ment of YY1 bound to young L1 families, despite conservation of the YY1 binding site among much older L1s, we speculate that YY1 has sequentially repressed each new mobile L1 family that has emerged during human evolution, with control passing to KAP1 or other factors as these new L1 families grow older and less likely to mobilize.

Numerous retrotransposition-competent L1s without an intact YY1 binding site could exist in the global population. That the Chr13 Δ 31_{L1} allele 1 was found in 3 of 8 members of our cohort as well as in another individual, where it generated a cortical neuron L1 insertion (Evrony et al., 2015), suggests that many people carry this hot L1 allele and that it is recurrently mobile in the neuronal lineage. Another element lacking a YY1 binding site because of 5' truncation, Chr1 Δ 21_{L1}-LRE2, was discovered to be the source of a pathogenic 3' transduction-carrying L1 insertion (Holmes et al., 1994) and is mobile in the germline and tumors (Gardner et al., 2017; Tubio et al., 2014). It is likely that further retrotransposition of Chr13 Δ 31_{L1}, Chr1 Δ 21_{L1}-LRE2, and other slightly 5'-truncated L1s will be reported in the future. Full-length L1s with intact YY1 binding sites may also

escape repression, by exception, because of their genomic location. For example, the heavily methylated element Chr8 Δ 3_{L1} mobilized in the brain (Evrony et al., 2012) with the assistance of an upstream promoter. Another full-length element, Chr22FL_{L1}-TTC28, is highly mobile and hypomethylated in tumors (Nguyen et al., 2018; Schauer et al., 2018; Tubio et al., 2014). As we found here, Chr22FL_{L1}-TTC28 was also unmethylated in many hepatic cells, perhaps because of its location intronic to a highly expressed gene. It is plausible that full-length and 5'-truncated donor L1s employ context-specific routes to evade YY1-mediated methylation and retrotranspose in both neural and non-neural somatic cells (Doucet-O'Hare et al., 2016; Shukla et al., 2013), generating L1 mosaicism beyond the brain.

Including this study, three somatic L1 insertions have been identified in neurons by single-cell WGS and were PCR amplified across their entire length (Evrony et al., 2015). Each carried a 5' or 3' transduction, which otherwise flank a minority of *de novo* L1 insertions. It is unclear whether WGA favors recovery of these events. That all three somatic L1 insertions were present in multiple neurons suggests that they arose in a neuronal lineage progenitor cell. However, because of the false negative rate of the approach and ascertainment bias, we cannot resolve the predominant neurogenic timing of L1 mobilization. Somatic L1 insertions arising during embryogenesis have, however, been detected in the mouse brain, without genomic analysis requiring WGA, suggesting that early neurodevelopment is a source of neuronal L1 mosaicism in mammals (Richardson et al., 2017). The probability of a somatic L1 insertion influencing phenotype presumably scales with the number of neurons carrying that event. However, a functional effect has yet to be discerned for any neuronal L1 insertion detected to date, and it remains to be seen whether donor L1s mobile in the brain are genetically associated with human neurological traits. Our discovery of three Chr13 Δ 31_{L1} alleles resolves a prior discrepancy where an L1 insertion was detected *in vivo* (Evrony et al., 2015) and arose from a donor L1 considered immobile *in vitro* (Brouha et al., 2003). It is almost certain that different Chr13 Δ 31_{L1} alleles were assayed in these two studies (Brouha et al., 2003; Evrony et al., 2015), highlighting a need to distinguish mobile and immobile donor L1 alleles found at the same genomic location.

To build a consensus view of somatic retrotransposition in the hippocampus, we applied WGS, RC-seq, and L1-IP to MDA-amplified neurons. The proportion of neurons found to harbor a somatic L1 insertion resembled prior estimates based on WGS and targeted L1 sequencing of MDA-amplified cortical (Evrony et al., 2012, 2015) and hippocampal neurons (Erwin et al., 2016) and is lower than that of a previous RC-seq analysis of MALBAC-amplified hippocampal neurons (Upton et al., 2015). False positives can occur in single-cell analyses of L1 insertions and other genomic variants (Faulkner and Garcia-Perez, 2017; McConnell et al., 2017). False negatives are, by contrast, harder to assess. We and others have previously assumed that sensitivity for heterozygous germline and somatic L1 insertions is similar in single-cell genomic analyses (Erwin et al., 2016; Evrony et al., 2012, 2015; Upton et al., 2015). Notably, somatic L1 insertions carry long, pure poly(A) tails, whereas heterozygous L1s rarely do. Despite deep (47 \times) single-cell WGS, our sensitivity for somatic L1 insertions was, at most, \sim 15%, even without ac-

counting for false negatives from PCR validation. These considerations preclude an accurate calculation of L1 mobilization rate. Our results nonetheless demonstrate L1 mosaicism in hippocampal neurons at the most conservative standard of genomic analysis and PCR validation, as shown elsewhere in the cortex (Evrony et al., 2015). More importantly, elucidation of YY1-mediated L1 repression and routes by which it is avoided provides a mechanistic explanation for L1 retrotransposition during neurodevelopment and positions YY1 as a major regulator of L1 activity over the course of human evolution.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Human Tissue Samples
 - Cell Lines
- METHOD DETAILS
 - DNA Extraction
 - RNA Extraction
 - Neuronal Isolation and Whole-Genome Amplification
 - Whole Genome-Sequencing (WGS), Retrotransposon Capture Sequencing (RC-seq), and L1 Insertion Profiling (L1-IP)
 - L1 Insertion Site PCR
 - Capillary Sequencing
 - Chr13 Δ 31_{L1} Allele Identification and Reconstruction for Retrotransposition Assay
 - PA-1 Cell Differentiation
 - hESC Neurodifferentiation
 - Retrotransposition Assay
 - Luciferase Reporter Assay
 - L1 Locus-Specific Methylation Assays
 - Genome-Wide Methylation and Hydroxymethylation Analyses
 - Genome-Wide Analyses of YY1 and KAP1 Binding
 - Detection of Chr13 Δ 31_{L1} Antisense Transcription
 - RNA-Seq Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.molcel.2019.05.024>.

ACKNOWLEDGMENTS

The authors thank the anonymous human subjects of this research for donating their postmortem tissues to the MRC Edinburgh Brain and Tissue Bank. The authors also thank the Translational Research Institute Flow Cytometry Facility, Alicia Barroso-delJesus, Pablo Tristan-Ramos, and Rabina Giri for technical assistance. This study was supported by the People Programme (Marie Curie Actions) of the European Union Seventh Framework Program (FP7/2007-2013) under REA grant agreement P10F-GA-2013-623324 (to F.J.S.-L.), an NHMRC Early Career Fellowship (GNT1161832 to S.W.C.), and

ARC Discovery Early Career Researcher Award (DE150101117) and Discovery Project (DP170101198) grants (to A.D.E.). J.L.G.-P. acknowledges funding from CICE-FEDER-P12-CTS-2256, Plan Nacional de I+D+I 2013–2016 (FIS-FEDER-PI14/02152), PCIN-2014-115-ERA-NET NEURON II, the European Research Council (ERC-Consolidator ERC-STG-2012-309433), an International Early Career Scientist grant from the Howard Hughes Medical Institute (IECS-55007420), and The Wellcome Trust-University of Edinburgh Institutional Strategic Support Fund (ISFF2). G.J.F. acknowledges support from the Mater Foundation, a CSL Centenary Fellowship, and NHMRC Project grants GNT1106206, GNT1125645, GNT1126393, and GNT1138795.

AUTHOR CONTRIBUTIONS

F.J.S.-L., M.-J.H.C.K., P.G., D.B.V.-L., S.R.R., R.-L.T., J.S.J., P.E.C., C.S.-P., M.G.-C., M.M.-L., L.S., A.M., and S.R.H. performed experiments. P.M.B., R.L., J.L.G.-P., and G.J.F. provided resources. D.B.V.-L., S.W.C., M.L., A.D.E., and G.J.F. performed bioinformatic analyses. F.J.S.-L. and G.J.F. conceived the study, designed experiments, generated figures, and wrote the manuscript. G.J.F. directed the study. All authors commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 11, 2018

Revised: April 8, 2019

Accepted: May 15, 2019

Published: June 20, 2019

REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Cell* *37*, 785–799.
- 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.
- Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* *479*, 534–537.
- Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* *141*, 1159–1170.
- 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.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.
- 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. USA* *100*, 5280–5285.
- Burns, K.H. (2017). Transposable elements in cancer. *Nat. Rev. Cancer* *17*, 415–424.
- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* *28*, 1397–1409.
- Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., and Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* *460*, 1127–1131.
- de la Rica, L., Deniz, Ö., 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.
- DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* *28*, 1530–1532.
- Denli, A.M., Narvaiza, I., Kerman, B.E., Pena, M., Benner, C., Marchetto, M.C., Diedrich, J.K., Aslanian, A., Ma, J., Moresco, J.J., et al. (2015). Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* *163*, 583–593.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
- Dombroski, B.A., Scott, A.F., and Kazazian, H.H., Jr. (1993). Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. *Proc. Natl. Acad. Sci. USA* *90*, 6513–6517.
- Doucet-O'Hare, T.T., Sharma, R., Rodić, N., Anders, R.A., Burns, K.H., and Kazazian, H.H., Jr. (2016). Somatic Acquired LINE-1 Insertions in Normal Esophagus Undergo Clonal Expansion in Esophageal Squamous Cell Carcinoma. *Hum. Mutat.* *37*, 942–954.
- Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev. Cell* *36*, 611–623.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.
- Erwin, J.A., Paquola, A.C., Singer, T., Gallina, I., Novotny, M., Quayle, C., Bedrosian, T.A., Alves, F.I., Butcher, C.R., Herdy, J.R., et al. (2016). L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* *19*, 1583–1591.
- Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A., et al. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* *151*, 483–496.
- Evrony, G.D., Lee, E., Mehta, B.K., Benjamini, Y., Johnson, R.M., Cai, X., Yang, L., Haseley, P., Lehmann, H.S., Park, P.J., and Walsh, C.A. (2015). Cell lineage analysis in human brain using endogenous retroelements. *Neuron* *85*, 49–59.
- Ewing, A.D., and Kazazian, H.H., Jr. (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* *20*, 1262–1270.
- Ewing, A.D., and Kazazian, H.H., Jr. (2011). Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Res.* *21*, 985–990.
- Ewing, A.D., Gacita, A., Wood, L.D., Ma, F., Xing, D., Kim, M.S., Manda, S.S., Abril, G., Pereira, G., Makohon-Moore, A., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* *25*, 1536–1545.
- Faulkner, G.J., and Garcia-Perez, J.L. (2017). L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends Genet.* *33*, 802–816.
- Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* *41*, 563–571.
- 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.
- Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2012). Ensembl 2012. *Nucleic Acids Res.* *40*, D84–D90.
- Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM

- Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. *Nature* 507, 462–470.
- 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., and Moran, J.V. (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 466, 769–773.
- Gardner, E.J., Lam, V.K., Harris, D.N., Chuang, N.T., Scott, E.C., Pittard, W.S., Mills, R.E., and Devine, S.E.; 1000 Genomes Project Consortium (2017). The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res.* 27, 1916–1929.
- Goodier, J.L. (2016). Restricting retrotransposons: a review. *Mob. DNA* 7, 16.
- Goodier, J.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2000). Transduction of 3′-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 9, 653–657.
- Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J. Mol. Biol.* 309, 1201–1208.
- Grandi, F.C., Rosser, J.M., and An, W. (2013). LINE-1-derived poly(A) microsatellites undergo rapid shortening and create somatic and germline mosaicism in mice. *Mol. Biol. Evol.* 30, 503–512.
- 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.
- Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.Y., and Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* 14, 774.
- Hata, K., and Sakaki, Y. (1997). Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 189, 227–234.
- 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.
- Heras, S.R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyra, E., Garcia-Perez, J.L., and Cáceres, J.F. (2013). The Microprocessor controls the activity of mammalian retrotransposons. *Nat. Struct. Mol. Biol.* 20, 1173–1181.
- Hervouet, E., Vallette, F.M., and Cartron, P.F. (2009). Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* 4, 487–499.
- 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.
- Hormozdiari, F., Alkan, C., Ventura, M., Hajirasouliha, I., Malig, M., Hach, F., Yorukoglu, D., Dao, P., Bakhshi, M., Sahinalp, S.C., and Eichler, E.E. (2011). Alu repeat discovery and characterization within human genomes. *Genome Res.* 21, 840–849.
- Imbeault, M., Helleboid, P.Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543, 550–554.
- Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261.
- 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.
- Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc. Natl. Acad. Sci. USA* 94, 1872–1877.
- Kazazian, H.H., Jr., and Moran, J.V. (2017). Mobile DNA in Health and Disease. *N. Engl. J. Med.* 377, 361–370.
- 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.
- Kim, J.D., and Kim, J. (2009). YY1’s longer DNA-binding motifs. *Genomics* 93, 152–158.
- Klawitter, S., Fuchs, N.V., Upton, K.R., Muñoz-Lopez, M., Shukla, R., Wang, J., Garcia-Cañadas, M., Lopez-Ruiz, C., Gerhardt, D.J., Sebe, A., et al. (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* 7, 10286.
- Kopera, H.C., Larson, P.A., Moldovan, J.B., Richardson, S.R., Liu, Y., and Moran, J.V. (2016). LINE-1 Cultured Cell Retrotransposition Assay. In *Methods Mol Biol.* J.L. Garcia-Perez, ed. (New York, NY: Humana Press), pp. 139–156.
- Kuhn, A., Ong, Y.M., Cheng, C.Y., Wong, T.Y., Quake, S.R., and Burkholder, W.F. (2014). Linkage disequilibrium and signatures of positive selection around LINE-1 retrotransposons in the human genome. *Proc. Natl. Acad. Sci. USA* 111, 8131–8136.
- Kumaki, Y., Oda, M., and Okano, M. (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.* 36, W170–5.
- 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.; Cancer Genome Atlas Research Network (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, arXiv:13033997, <http://arxiv.org/abs/1303.3997>.
- Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., et al. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* 47, 469–478.
- Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C., and Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* 553, 228–232.
- 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.
- Macia, A., Muñoz-Lopez, M., Cortes, J.L., Hastings, R.K., Morell, S., Lucena-Aguilar, G., Marchal, J.A., Badge, R.M., and Garcia-Perez, J.L. (2011). Epigenetic control of retrotransposon expression in human embryonic stem cells. *Mol. Cell Biol.* 31, 300–316.
- Macia, A., Widmann, T.J., Heras, S.R., Ayllon, V., Sanchez, L., Benkaddour-Boumzaouad, M., Muñoz-Lopez, M., Rubio, A., Amador-Cubero, S., Blanco-Jimenez, E., et al. (2017). Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* 27, 335–348.
- Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Marchetto, M.C., Carroumeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* 143, 527–539.
- Marchetto, M.C.N., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C.M., Desai, K.N., Herai, R.H., Weitzman, M.D., et al. (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* 503, 525–529.
- McConnell, M.J., Moran, J.V., Abyzov, A., Akbarian, S., Bae, T., Cortes-Ciriano, I., Erwin, J.A., Fasching, L., Flasch, D.A., Freed, D., et al. (2017). Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* 356, eaal1641.
- Mills, R.E., Bennett, E.A., Iskow, R.C., and Devine, S.E. (2007). Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191.
- 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.

- Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* 283, 1530–1534.
- Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910.
- Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446.
- Nguyen, T.H.M., Carreira, P.E., Sanchez-Luque, F.J., Schauer, S.N., Fagg, A.C., Richardson, S.R., Davies, C.M., Jesuadian, J.S., Kempen, M.H.C., Troskie, R.L., et al. (2018). L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. *Cell Rep.* 23, 3730–3740.
- Ostertag, E.M., and Kazazian, H.H., Jr. (2001). Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* 11, 2059–2065.
- Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* 28, 1418–1423.
- 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.
- Pickeral, O.K., Makałowski, W., Boguski, M.S., and Boeke, J.D. (2000). Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* 10, 411–415.
- Richardson, S.R., Gerdes, P., Gerhardt, D.J., Sanchez-Luque, F.J., Bodea, G.O., Muñoz-Lopez, M., Jesuadian, J.S., Kempen, M.H.C., Carreira, P.E., Jeddeloh, J.A., et al. (2017). Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res.* 27, 1395–1405.
- Robbez-Masson, L., Tie, C.H.C., Conde, L., Tunbak, H., Husovsky, C., Tchakovnikarova, I.A., Timms, R.T., Herrero, J., Lehner, P.J., and Rowe, H.M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* 28, 836–845.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463, 237–240.
- 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.
- Sanchez-Luque, F.J., Richardson, S.R., and Faulkner, G.J. (2017). Analysis of Somatic LINE-1 Insertions in Neurons. In *Genomic Mosaicism in Neurons and Other Cell Types*, J.M. Frade and F.H. Gage, eds. (New York, NY: Humana Press), pp. 219–251.
- Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., and Kazazian, H.H., Jr. (1997). Many human L1 elements are capable of retrotransposition. *Nat. Genet.* 16, 37–43.
- Schauer, S.N., Carreira, P.E., Shukla, R., Gerhardt, D.J., Gerdes, P., Sanchez-Luque, F.J., Nicoli, P., Kindlova, M., Ghisletti, S., Santos, A.D., et al. (2018). L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res.* 28, 639–653.
- Schlesinger, S., Lee, A.H., Wang, G.Z., Green, L., and Goff, S.P. (2013). Proviral silencing in embryonic cells is regulated by Yin Yang 1. *Cell Rep.* 4, 50–58.
- Schmitges, F.W., Radovani, E., Najafabadi, H.S., Barazandeh, M., Campitelli, L.F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T., et al. (2016). Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res.* 26, 1742–1752.
- Scott, E.C., and Devine, S.E. (2017). The Role of Somatic L1 Retrotransposition in Human Cancers. *Viruses* 9, 131.
- 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.
- Shukla, R., Upton, K.R., Muñoz-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.
- Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell Biol.* 21, 1973–1985.
- Stewart, C., Kural, D., Strömberg, M.P., Walker, J.A., Konkel, M.K., Stütz, A.M., Urban, A.E., Grubert, F., Lam, H.Y., Lee, W.P., et al.; 1000 Genomes Project (2011). A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* 7, e1002236.
- Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H., et al.; 1000 Genomes Project Consortium (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* 526, 75–81.
- Sun, X., Wang, X., Tang, Z., Grivainis, M., Kahler, D., Yun, C., Mita, P., Fenyö, D., and Boeke, J.D. (2018). Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl. Acad. Sci. USA* 115, E5526–E5535.
- Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell Biol.* 10, 6718–6729.
- Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034.
- Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R., et al. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11, 805–814.
- 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.; ICGC Breast Cancer Group; ICGC Bone Cancer Group; ICGC Prostate Cancer Group (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345, 1251343.
- Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* 24, 1260–1270.
- Upton, K.R., Gerhardt, D.J., Jesuadian, J.S., Richardson, S.R., Sánchez-Luque, F.J., Bodea, G.O., Ewing, A.D., Salvador-Palomeque, C., van der Knaap, M.S., Brennan, P.M., et al. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* 161, 228–239.
- van den Hurk, J.A., Meij, I.C., Seleme, M.C., Kano, H., Nikopoulos, K., Hoefsloot, L.H., Sistermans, E.A., de Wijs, I.J., Mukhopadhyay, A., Plomp, A.S., et al. (2007). L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* 16, 1587–1592.
- Walter, M., Teissandier, A., Pérez-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.
- Wang, J., Song, L., Grover, D., Azrak, S., Batzer, M.A., and Liang, P. (2006). dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* 27, 323–329.
- Weichenrieder, O., Repanas, K., and Perrakis, A. (2004). Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. *Structure* 12, 975–986.
- Wissing, S., Muñoz-Lopez, M., Macia, A., Yang, Z., Montano, M., Collins, W., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2012). Reprogramming somatic cells into IPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* 21, 208–218.

- Witherspoon, D.J., Xing, J., Zhang, Y., Watkins, W.S., Batzer, M.A., and Jorde, L.B. (2010). Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. *BMC Genomics* 11, 410.
- Witherspoon, D.J., Zhang, Y., Xing, J., Watkins, W.S., Ha, H., Batzer, M.A., and Jorde, L.B. (2013). Mobile element scanning (ME-Scan) identifies thousands of novel Alu insertions in diverse human populations. *Genome Res.* 23, 1170–1181.
- Wolf, G., Yang, P., Füchtbauer, A.C., Füchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* 29, 538–554.
- 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.
- 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.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149, 1368–1380.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
- Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
One Shot TOP10 Electrocomp <i>E. coli</i> cells	Invitrogen	C404052
One Shot TOP10 Chemically Competent cells	Invitrogen	C404006
Biological Samples		
Snap frozen hippocampus, liver, and cortex tissue from eight post-mortem individuals.	Edinburgh Sudden Death Brain and Tissue Bank	As listed in Table S1 .
Chemicals, Peptides, and Recombinant Proteins		
TRIzol	Invitrogen	15596026
Trichostatin A (from <i>Streptomyces</i> sp.)	Sigma-Aldrich	T8552
All <i>trans</i> -retinoic acid	Sigma-Aldrich	302-79-4
Matrigel membrane matrix	Corning, Thermo Fisher Scientific	CB-40234
N-2 Supplement (100X)	GIBCO, Thermo Fisher Scientific	17502048
Y-27632 dihydrochloride	Tocris Bioscience, Thermo Fisher Scientific	12-541-0
Dorsomorphin dihydrochloride	Tocris Bioscience, Thermo Fisher Scientific	30-931-0
Stemolecule SB431542	ReproCell	04-0010-10
B-27 Supplement, Serum Free	GIBCO, Thermo Fisher Scientific	17-504-044
Recombinant human Fibroblast Growth Factor, basic (bFGF)	R&D Systems	4114-TC
Poly-L-ornithine hydrobromide	Sigma-Aldrich	P3655
Laminin mouse protein, natural	Thermo Fisher Scientific	23017015
Critical Commercial Assays		
Dual Luciferase Reporter Assay System	Promega	E1910
TruSeq Nano DNA Library Preparation Kit	Illumina	FC-121-4001/2
TruSeq DNA PCR-free Library Preparation Kit	Illumina	FC-121-3001/2
TruSeq Stranded mRNA Library Preparation Kit	Illumina	RS-122-2101/2
MiSeq Reagent Kit (600-cycle)	Illumina	MS-102-3003
Agencourt AMPure XP beads	Beckman Coulter	A63882
EZ DNA Methylation Lightning Kit	Zymo Research	D5030
TaqPath 1-Step RT-qPCR Master Mix	Applied Biosystems	A15299
Human TBP (20X) Pre-Developed TaqMan Assay Reagent. Endogenous Control (FAM/MGB probe, non-primer limited)	Applied Biosystems by Life Technologies	4333769F
Human GAPDH (20X) Pre-Developed TaqMan Assay Reagent. Endogenous Control (FAM/MGB probe, non-primer limited)	Applied Biosystems by Life Technologies	4333764F
Deposited Data		
WGS, RC-seq and L1-IP for CTRL-36 hippocampal neurons, bulk hippocampus, and bulk liver. WGS also for 7 other control individual liver samples, and H1 cell line	This paper	ENA: PRJEB24579
Hippocampus RNA-seq for 8 individuals	This paper	ENA: PRJEB24579
WGBS CTRL-36 hippocampal neurons	This paper	ENA: PRJEB24579
H1 TAB-seq	Gene Expression Omnibus	GEO: GSM882245
H1 WGBS	Gene Expression Omnibus	GEO: GSM432685
H1 YY1 ChIP-seq	Gene Expression Omnibus	GEO: GSM803513
H1 KAP1 ChIP-seq	Gene Expression Omnibus	GEO: GSM1399258
HEK293 YY1 ChIP-seq	Gene Expression Omnibus	GEO: GSE76494

(Continued on next page)

Continued		
REAGENT OR RESOURCE	SOURCE	IDENTIFIER
HEK293 RNA-seq	Gene Expression Omnibus	GEO: GSE76495
Unprocessed gel images	https://doi.org/10.17632/jm9mr476vn.1	This work
Experimental Models: Cell Lines		
HeLa-JVM	Prof. John V. Moran	N/A
PA-1	ATCC	CRL-1572
HEK293T	ATCC	CRL-3022
H1-hESC	WiCell Research Institute, Inc.	WA01
Oligonucleotides		
Oligonucleotide sequences are shown in Table S4 .	Integrated DNA Technologies, Inc. (Iowa, US)	N/A
Probe for RT-qPCR: [VIC]CCAAGCCCTAATTAA[MGB]	Applied Biosystems by Life Technologies	This work
Recombinant DNA		
pJM-L1.3	Prof. John V. Moran	N/A
pJM-L1.3ΔCMV	Prof. John V. Moran	N/A
99-gfp-L1.3	Dr. J.L.G-P. lab.	N/A
pGL3-Basic vector	Promega	GenBank: U47295
Software and Algorithms		
TEBreak	https://github.com/adamewing/tebreak	(Ewing et al., 2015)
CpH-methylation	https://github.com/MischaLundberg/ch-methylation	This work
Quantification tool for Methylation Analysis (QUMA)	http://quma.cdb.riken.jp/	(Kumaki et al., 2008)
GraphPad Prism version 7.00 for Windows	GraphPad Software, La Jolla California USA, https://www.graphpad.com	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Geoffrey J. Faulkner (faulknergj@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Tissue Samples

Snap frozen hippocampus and liver tissue from eight post-mortem individuals (see [Table S1](#) for gender and age information) without neurological disease, as well as frontal cortex from three of these individuals, was provided to P.M.B. by the Edinburgh Sudden Death Brain and Tissue Bank with ethical approval to be used as described in the study (East of Scotland Research Ethics Service, Reference: LR/11/ES/0022). Further ethics approvals were provided by the Mater Health Services Human Research Ethics Committee (Reference: HREC-15-MHS-52) and the University of Queensland Medical Research Review Committee (Reference: 2014000221).

Cell Lines

HeLa cells (HeLa-JVM) ([Moran et al., 1996](#)) were kindly provided by Prof. John V. Moran (University of Michigan) and were originally derived from the cervical adenocarcinoma of a 31yr old human female. HeLa cells were maintained in high glucose Dulbecco's Modified Eagle Medium (DMEM, Life Technologies), 10% heat-inactivated Fetal Bovine Serum (FBS, Life Technologies), 2mM L-Glutamine (Life Technologies) and 100U/mL Penicillin-Streptomycin solution (Pen-Strep, Life Technologies). PA-1 cells were purchased from the American Type Culture Collection (ATCC). These are an ovarian teratocarcinoma cell line from a 12yr old human female. PA-1 cells were maintained in Minimum Essential Medium (MEM) with GlutaMAX supplement (Life Technologies), 10% heat-inactivated FBS, 1X non-essential amino acids (NEAA, Life Technologies) and 100U/mL Pen-Strep solution (Life Technologies). HEK293T cells were purchased from ATCC, are a human embryonic kidney cell line, and were maintained in DMEM high glucose, 10% heat-inactivated FBS, 2mM L-Glutamine and 100U/mL Pen-Strep. For passaging, HeLa, PA-1 and HEK293T cell lines were washed using Dulbecco's Phosphate Buffered Saline (DPBS, Life Technologies) and raised using Trypsin 0.25% EDTA (Life Technologies). H1 is a human embryonic stem cell (hESC) line derived from a human embryo (male) donated for research at the University of Wisconsin, and acquired for this study from the WiCell Research Institute. H1 cells were cultured on plates coated with Matrigel (Corning) in mTESR1 medium (StemCell Technologies). All cell lines were grown in a tissue culture incubator at 37°C, 5% CO₂, ~95% humidity, with media replaced daily.

METHOD DETAILS

DNA Extraction

Genomic DNA was obtained from tissue via phenol-chloroform extraction. For each sample, ~50mg of snap frozen tissue was shaved on dry ice with a scalpel blade and dissolved in 500 μ L of lysis buffer (1X TE, 2% SDS, 100 μ g/mL Proteinase K, New England Biolabs) in a 1.5mL tube. Samples were incubated at 65°C and 600rpm in a thermomixer for ~1hr, or until the tissue was completely dissolved, followed by the addition of RNase A (Thermo Fisher Scientific) to 20 μ g/mL final concentration and further incubated at 37°C and 600rpm for 30min. One volume of phenol (Sigma Aldrich) was then added to the sample and mixed by vortexing. Phases were separated by centrifugation at 12,000 g for 10min, with the aqueous phase transferred to a fresh tube. Successive steps with one volume of phenol:chloroform:isoamyl alcohol (25:24:1; Sigma Aldrich) and one volume of chloroform:isoamyl alcohol (24:1; Sigma Aldrich) were then performed. The final aqueous phase was transferred to a fresh tube and DNA was then ethanol precipitated by adding 3 volumes of molecular grade absolute ethanol (Sigma Aldrich), 0.1 volume of 3M Sodium Acetate (Sigma Aldrich) and 5ng of glycogen (Ambion, Life Technologies). Samples were incubated at –80°C for 30min and then spun down at 4°C and 12,000 g for 30min. Supernatant was discarded by inversion. The pellet was then washed by adding 1mL of 75% ethanol and spun down for 5min at 4°C and 12,000 g. Supernatant was again discarded by inversion. The resulting DNA was air-dried and resuspended in TE buffer.

Genomic DNA was obtained from cell lines as follows: PA-1 adherent cells were grown in 6-well plates (NUNC) until confluent. Media were aspirated, and cells washed by adding 2mL of DPBS, which was removed by aspiration. Cells were then raised by adding 0.5mL of Trypsin 0.25% EDTA and incubated at 37°C for 5min. Trypsinization was stopped by adding 0.5mL of DPBS in 10% FBS. Cells were pipetted up and down for complete dissociation, transferred to a new 1.5mL tube and pelleted by centrifugation at 3,000 g at room temperature for 5min. Supernatant was removed by aspiration. The cell pellet was then resuspended in 0.5–1mL of lysis buffer. DNA extraction was then performed as described above for tissue samples. H1 hESCs and derived cultures were grown in 10cm plates. Media were aspirated and replaced with PBS, and cells were raised using a cell scraper. Resuspended cells were transferred to 15mL tubes and centrifuged (170 g for 3min for hESCs and 300 g for 5min for NPCs and neurons). DNA extractions were performed using the resulting cell pellets, which contained $\sim 3 \times 10^6$ to 1×10^8 cells, and a NucleoSpin Tissue kit (Macherey-Nagel).

RNA Extraction

Total RNA was extracted from tissue using TRIzol Reagent (Invitrogen) and following the manufacturer's instructions with minimal modifications. ~50mg of snap frozen tissue was shaved with a scalpel blade and dissolved in 1mL of TRIzol. Tissue shaving was done on a dry ice bed and shaved tissues were kept in 1.5mL tubes on dry ice until the entire batch of samples had been shaved, prior to dissolving them in TRIzol. TRIzol suspensions were centrifuged for 5min at 12,000 g at 4°C to pellet cell debris. The solution was then transferred to a fresh tube and incubated for 5min at room temperature. 200 μ L of chloroform (Sigma-Aldrich) was added to each tube, shaken by hand for 15sec, and incubated at room temperature for 3min. Samples were then centrifuged at 12,000 g at 4°C for 15min. The aqueous phase was transferred to a fresh tube, where 500 μ L of isopropanol (Sigma-Aldrich) was immediately added. Samples were mixed by inversion, incubated at 4°C for 30min and then centrifuged at 12,000 g at 4°C for 30min. Supernatant was then discarded by inversion and pellets were washed by adding 1mL of 75% ethanol, preventing pellet dislodgment. Tubes were centrifuged again at 12,000 g at 4°C for 10min and ethanol was removed by inversion. Samples were left upside down on a paper cloth (Kimwipes, Kimtech) to dry for 5min. RNA was resuspended immediately in 45 μ L of nuclease free water. DNase treatment was performed using a TURBO DNase-free kit (Thermo Fisher Scientific) following manufacturer's instructions. DNA removal was tested by amplification of a 127bp L1 DNA fragment directly from the extracted RNA sample. Briefly, a PCR reaction with MyTaq HS DNA Polymerase (BioLine) in a 1X MyTaq buffer, 20pmol of primer L1_5819_Fw, 20pmol of primer L1_5945_Rv, 1U MyTaq DNA polymerase and 1 μ L of RNA sample in 20 μ L final volume reaction (50ng of human gDNA was used as a positive control). PCR cycling conditions were (95°C, 1min) \times 1; (95°C, 15sec; 52°C, 15sec; 72°C, 10sec) \times 35; (72°C, 5min; 4°C, hold) \times 1 and the reaction was resolved in a 2% agarose gel electrophoresis in 1X TAE, 1×10^{-4} (v/v) SYBR Safe DNA gel stain (Life Technologies). RNA integrity was also visually checked by electrophoresis of 500ng of RNA in a 2% agarose gel in 1X TAE, 1×10^{-4} (v/v) SYBR Safe DNA gel stain. RNA concentration was measured using NanoDrop Lite Microlitre Spectrophotometer (Thermo Fisher Scientific), with aliquots then kept at –80°C.

Total RNA was extracted from cultured cells using the TRIzol protocol described for tissue, with the following modifications: PA-1 cell media were removed from 6-well plates by aspiration when cultures were ~80% confluent and washed with 2mL of DPBS. After removing the DPBS by aspiration, cells were resuspended and lysed in 2mL of TRIzol by pipetting up and down. The TRIzol lysate was split into two 1.5mL tubes (1mL each) for further processing, and later the two RNA extractions from each sample were combined into the same tube when resuspended in 45 μ L of nuclease free water. RNA extraction from H1 hESCs and derived cultures was performed using pellets obtained as described above for DNA extraction. Cell pellets were lysed by resuspension in 1mL of TRIzol.

Neuronal Isolation and Whole-Genome Amplification

NeuN⁺ neuronal nuclei were sorted from CTRL-36 hippocampal tissue, and individually isolated using a micromanipulator, as described previously (Sanchez-Luque et al., 2017). Single nuclei were placed in separate 0.2mL PCR tubes and subjected to whole genome amplification (WGA) by multiple displacement amplification (MDA) (Evrony et al., 2012), with the following minor modifications: phi29 enzyme inactivation after the 16hr incubation step was omitted. Instead, the WGA material was diluted to twice its volume

with molecular grade water (40 μ L final volume) and a clean-up with 1:1.3 (v/v) ratio Agencourt AMPure XP beads (Beckman Coulter) was performed immediately before the de-branching step. Multiplexed PCR quality control was performed after the de-branching step, using MyTaq HS DNA Polymerase (Bioline). The PCR reaction comprised: 2 μ L of a 1:50 dilution of the de-branched DNA as template, 1X reaction buffer, 100pmol of each primer (listed in Table S4), 1U of MyTaq, and molecular grade H₂O added to a final volume of 20 μ L. PCR cycling conditions were as follows: (94°C, 3min) \times 1; (94°C, 1min; 68°C -1° C/cycle, 1min; 72°C, 1min) \times 13; (92°C, 1min; 55°C, 1min; 72°C, 1min) \times 27; (72°C, 10min; 4°C, hold) \times 1. Only single nuclei that produced 3 or 4 bands in the quality control PCR after WGA were retained. For more a detailed protocol, please see (Sanchez-Luque et al., 2017).

Whole Genome-Sequencing (WGS), Retrotransposon Capture Sequencing (RC-seq), and L1 Insertion Profiling (L1-IP)

Illumina libraries were prepared using DNA from 24 single MDA-amplified CTRL-36 hippocampal neurons, bulk CTRL-36 hippocampus, bulk liver from CTRL-36 and 7 other individuals, H1 hESCs, H1-derived neuronal progenitor cells (NPCs), and H1-derived neurons. Library preparation involved a TruSeq Nano DNA Library Preparation Kit (Illumina). CTRL-36 MDA-amplified neuron, bulk hippocampus, and bulk liver libraries were enriched for L1-genome junctions via RC-seq, as described previously (Upton et al., 2015). The 24 MDA-amplified neuron libraries were barcoded using Low Throughput (LT) 6-mer indices and pooled together in a single RC-seq hybridization reaction. Briefly, 1 μ g DNA was diluted to 130 μ L final volume and sheared in a Covaris M220 Focused-Ultrasonicator (peak power 50, duty factor 20, pulses per burst 200) for 120sec in MicroTube AFA Snap-Cap tubes (Covaris). Note that shearing parameters for both MDA-amplified and bulk material were as detailed for bulk genomic DNA in our previous study (Upton et al., 2015) as MDA amplicons are much larger than those obtained via MALBAC. DNA was purified by Ampure XP beads clean up using 1:1 volume of beads and eluting in 60 μ L volume of resuspension buffer. The Illumina library preparation protocol was then followed as indicated by the manufacturer until the tandem clean up after adaptor ligation. Samples were instead suspended in 20 μ L of resuspension buffer and loaded in a 2% high-resolution agarose gel (Sigma-Aldrich) in 1 \times TAE buffer for electrophoresis. Size selection was performed by purifying gel cuts of 270-290bp size, which were eluted using the MiniElute Gel Extraction Kit (QIAGEN). Buffer QG was added at a ratio of 6 μ L per 1mg of gel cut and the agarose was dissolved at room temperature. Elution was performed using 16 μ L of 60°C pre-heated buffer EB twice, ending with \sim 30 μ L of elution volume. Library amplification was performed using 1 \times Phusion High-Fidelity PCR Master Mix (New England Biolabs) with 100 pmol of each TS-F and TS-R Illumina primers in 100 μ L final volume. Cycling conditions were as follows: (98°C, 45sec) \times 1; (98°C, 15sec; 60°C, 30sec; 72°C, 30sec) \times (6 for bulk DNA and 8 for WGA DNA); (72°C, 5min; 4°C, hold) \times 1. Samples were purified by Ampure XP beads clean up using 1:1 volume and eluted in 30 μ L of molecular grade water. Whole-genome amplified and bulk DNA libraries were pooled in equimolar amounts, combining different barcodes up to 1 μ g final DNA, together with 10 μ L of Sequence Capture Developer Reagent, 1nmol of Universal Blocking Oligonucleotide and 1nmol of an equimolecular pool of the blocker oligonucleotides for the adaptor used (Upton et al., 2015). The mixture was split in two halves (one for each end capture) and heat/vacuum dried in a Speed-Vac. Each half was resuspended in 7.5 μ L of 2 \times Hybridization Buffer and 3 μ L of Hybridization Component A from the NimbleGen Sequence Capture Kit (Roche) by vortexing, followed by 10sec of spinning down, and then denatured at 95°C for 5min in a thermoblock with a heat-covered lid to prevent evaporation. Each half of the pool was combined with 4.5 μ L of either 5' or 3' L1 end LNA capture probe (Upton et al., 2015), pre-heated to 95°C and placed in a 95°C heated thermocycler for 3min. Each capture reaction was then transferred to a 47°C thermocycler with 57°C lid for 3 days. Then, the captured library was purified using Dynabeads M-270 Streptavidin (Life Technologies) and NimbleGen Capture Wash Kit (Roche) following the manufacturer's instructions. Briefly, each capture reaction was transferred to a tube with washed, dried and 47°C pre-heated Dynabeads (using 100 μ L of original beads per each capture reaction) and then incubated at 47°C for 30sec. Each tube was then quickly flicked, spun down, and placed at 47°C in a thermocycler for 45min, and resuspended by pipetting every 15min. Capture reactions were then placed in a 47°C pre-warmed magnetic rack and supernatant was aspirated and discarded. The bead pellet was resuspended by pipetting in 100 μ L of 47°C pre-warmed wash buffer I during 10sec in a 47°C thermoblock. The tubes were then placed again in the 47°C pre-warmed magnetic rack, supernatant was aspirated, and 200 μ L of 47°C pre-warmed stringent wash buffer were added to each tube. The beads were resuspended by pipetting up and down 10 times and incubated at 47°C for 5min. This step was repeated once and the buffer was aspirated. 200 μ L of room temperature 1 \times wash buffer I were added and the beads were mixed by vortexing at minimum speed for 2min. This step was repeated with room temperature 1 \times wash buffer II and vortexed for 1min and then with room temperature 1 \times wash buffer III and vortexed for 30sec to be finally resuspended in 50 μ L of molecular grade water. Post-hybridization amplification was performed in 200 μ L final volume with 1 \times Phusion High-Fidelity PCR Master Mix, 200pmol of TS-F primer and 200pmol of TS-R primer using the following cycling conditions: (98°C, 45sec) \times 1; (98°C, 15sec; 60°C, 30sec; 72°C, 30sec) \times (8 for bulk DNA and 10 for WGA DNA); (72°C, 5min; 4°C, hold) \times 1 (note that each reaction was split in two tubes with 100 μ L each for the amplification which were combined afterward). The reaction was purified using a MiniElute Gel Extraction Kit, starting with the addition of 1mL of binding buffer to the 200 μ L reaction and loading it into the column in two steps of 600 μ L. The column was incubated with 700 μ L of washing buffer for 2min at room temperature before centrifugation at maximum speed for 1min. Columns were then rotated 180° and spun again for 30sec. Final elution was performed by adding 16 μ L of 60°C pre-warmed elution buffer, incubating the column for 5min and spinning down at maximum speed in a fresh tube. Library concentration was calculated by Qubit dsDNA HS Assay kit (Life Technologies) and combined in a 3:7 ratio for the 5' and 3' LNA captures.

L1-IP library preparation was performed using the same input DNA samples as used for the RC-seq libraries, and as described previously (Evrony et al., 2012), with minor modifications. Briefly, we perform 8 replicate PCR reactions with each WGA and bulk

DNA sample using GoTaq Flexi DNA polymerase (Promega). Each replicate had 200ng of DNA input, 2.5U of GoTaq DNA polymerase and 40pmol of L1SP1A2 primer (Evrony et al., 2012) in 1 × colorless reaction buffer with 2mM of each dNTP, 3mM MgCl₂ and 1% DMSO and was subjected to the following cycling conditions: (95°C, 2min:30sec) × 1; (95°C, 30sec; 58°C, 1min; 72°C, 2min) × 5. The reactions were then paused at 60°C and 4μL of a 5μM solution of Seed#Primer-N5 × 2-BC# primer were added to each. The same barcode (BC) set of primers was used for each set of 8 reactions, using a different seed sequence per replicate. Samples were immediately subjected to the following cycling conditions: (95°C, 30sec; 58°C, 30sec; 72°C, 1min:30sec) × 15; (72°C, 10min; 4°C, hold) × 1. All reactions were purified by clean up using 1.2 volumes of AMPure XP beads and resuspended in 50μL of resuspension buffer. A second round of PCR was performed using GoTaq Green Master Mix (Promega) with 2.5μL of the first PCR reaction as input and 15pmol of each Adap1L1HsG and Adap2Seq1 primers (Evrony et al., 2012) in a 25μL final volume reaction with the following cycling conditions: (95°C, 2min) × 1; (95°C, 30sec; 62°C, 30sec; 72°C, 1min) × 15; (72°C, 5min; 4°C, hold) × 1. All reactions were loaded into a 1.5% agarose gel electrophoresis in 1 × TAE and gel cuts aiming for 200-500bp fragment size were purified using MiniElute Gel Extraction Kit in 26μL of elution buffer. Each reaction was quantified using an Agilent 2100 Bioanalyzer and Agilent DNA 1000 Reagent and DNA chips (Agilent Technologies). Each set of 8 reactions was pooled, aiming to include a total of 15ng of each reaction when possible. Combined samples were topped with elution buffer to a final volume of 76μL and subjected to a cleanup with 1.2 volume of AMPure XP beads and a final resuspension in 21μL of resuspension buffer. Each combined sample was then quantified using an Agilent 2100 Bioanalyzer and pooled together in two sets containing a different group of 12 WGA neurons plus one bulk sample (liver or hippocampus) library, aiming for a total of 100ng DNA in each pool, and distributed as follows: 20% for the bulk sample and 80% of an equimolecular pool of the 12 WGA neurons. Once pooled, both sets were subjected to end polishing by T4 DNA polymerase (New England Biolabs). Each reaction contained the ~100ng DNA library pool, 1.5U of T4 DNA polymerase and 1 × NEB buffer 2, 100μM each dNTP and 100μg/mL BSA in 50μL final volume, and was performed at 12°C for 15min. Pools were finally cleaned up with 1.2 volume of AMPure XP beads and eluted in 21μL of resuspension buffer. We used 14 index sequences corresponding to Illumina barcodes 1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 18 and 19 so that 12 of the barcodes were used for two neuron samples each, and the remaining barcodes were used for the bulk samples. Neurons sharing the same barcode were included in only one of the two pools.

All CTRL-36 samples were analyzed with WGS, RC-seq and L1-IP. The bulk liver tissues of the 7 other individuals and H1 samples were only analyzed with WGS. All libraries were subjected to paired-end 2 × 150-mer sequencing using Illumina HiSeq X Ten instruments (Kinghorn Centre for Clinical Genomics, Australia; Macrogen, South Korea). For WGS experiments, each MDA-amplified neuron and bulk library was sequenced independently on a single HiSeq X Ten lane, excluding the CTRL-36 hippocampus, which was sequenced on two lanes for additional depth. For RC-seq experiments, the post-hybridization pool of 24 MDA-amplified neuron libraries was sequenced on one HiSeq X Ten lane, while CTRL-36 bulk liver and hippocampus RC-seq data were obtained from a prior study (Upton et al., 2015). For L1-IP experiments, two batches of 12 MDA-amplified neuron libraries pooled with either CTRL-36 bulk hippocampus or liver libraries were each sequenced on one HiSeq X Ten lane. Here, the second read in each pair was not retrieved as the L1-IP library design involves single-end sequencing. Overall WGS, RC-seq and L1-IP library statistics are provided in Table S1.

Whole genome sequencing and RC-seq data were analyzed via TEBreak (<https://github.com/adamewing/tebreak>). Briefly, paired-end data were aligned to the hg19 reference genome sequence using BWA-MEM (Li, 2013) with parameters -Y -M and duplicate reads marked with Picard MarkDuplicates (<http://broadinstitute.github.io/picard>). L1-IP data were analyzed as described previously (Ewing et al., 2015). Putative somatic L1 insertions were required to meet or exceed the following read depth thresholds in at least one MDA-amplified neuron library: WGS, each L1-genome junction detected by ≥ 8 reads; RC-seq, at least one L1-genome junction detected by ≥ 8 reads; L1-IP, a peak of ≥ 100 reads. Candidate insertions were then filtered if they were located within 10kb of an annotated reference genome L1-Ta or L1PA2 insertion, were detected in any bulk liver library, or corresponded to a known non-reference germline L1 insertion (Beck et al., 2010; Ewing and Kazazian, 2010, 2011; Helman et al., 2014; Hormozdiari et al., 2011; Iskow et al., 2010; Kuhn et al., 2014; Lee et al., 2012; Shukla et al., 2013; Stewart et al., 2011; Sudmant et al., 2015; Tubio et al., 2014; Wang et al., 2006; Witherspoon et al., 2010, 2013).

A cohort of 46 candidate heterozygous polymorphic L1 insertions present in CTRL-36 were identified using the TEBreak genotyping function, which uses the ratio of reads aligned across the TSD that either support (filled site) or do not support (empty site) an L1-genome junction to estimate the L1 variant allele fraction (VAF). VAFs between 0.2 and 0.8 were considered potentially heterozygous, and their presence in CTRL-36 was confirmed via PCR. The presence of specific full-length and truncated L1-Ta and L1PA2 copies in CTRL-36 and H1 cells (Table S2) was assessed via examination of WGS data with TEBreak and the Integrative Genomics Viewer.

L1 Insertion Site PCR

PCR reactions targeting L1-genome junctions were used to evaluate germline L1 insertion genotype (present/absent) and to validate somatic L1 insertions. Each reaction was performed using MyTaq HS DNA polymerase, with 1X MyTaq Reaction Buffer, 20pmol of each primer flanking the junction, 10ng of template DNA and 1U of enzyme, in a 20μL final volume. PCR cycling conditions were as follows: (95°C, 2min) × 1; (95°C, 30sec; 60°C, 30sec; 72°C, 30sec) × 30; (72°C, 5min; 4°C, hold) × 1. PCR primers were designed following established guidelines (Sanchez-Luque et al., 2017) and are listed in Table S4, together with reaction-specific modifications. Reaction products were loaded and resolved via electrophoresis in a 2% agarose gel composed of 1X TAE and 1 × 10⁻⁴ (v/v) SYBR Safe DNA gel stain.

Additional PCR reactions were designed to simultaneously amplify L1 insertions and their flanking genomic site (empty/filled assay) for germline L1 insertions (absent/heterozygous/homozygous), and to validate somatic L1 insertions. Each reaction was performed with an Expand Long Range dNTPack kit (Sigma Aldrich), using 1X reaction buffer with MgCl₂, 0.5mM of each dNTP, 5% DMSO, 10pmol of each primer, 2.75U of enzyme and 4ng of template DNA in a 25μL final volume. PCR cycling conditions were as follows: (92°C, 2min) × 1; (92°C, 10sec; 58°C, 15sec; 68°C, 6:00min) × 10; (92°C, 10sec; 58°C, 15sec; 68°C, 6min+20sec/cycle) × 30; (68°C, 10min; 4°C, hold) × 1. Products were resolved via electrophoresis in a 1% agarose gel. Gels were scanned in a Typhoon FLA 9500 Scanner (GE Healthcare Life Sciences). Where necessary, bands were excised from gels and purified using a conventional phenol:chloroform DNA extraction approach.

Capillary Sequencing

Filled site PCR products were cloned for sequencing in a TOPO XL PCR cloning kit (Life Technologies) using One Shot TOP10 Electrocomp *E. coli* cells (Invitrogen). Empty site and L1-genome junction amplicons were cloned using the pGEM-T Easy Vector System (Promega). DNA was obtained from clones using a QIAPrep Miniprep kit (QIAGEN). Filled site PCR products were also directly sequenced, with DNA concentration quantified using a Qubit dsDNA HS Assay kit. For filled site PCR products, stepping primers (Table S4) along the L1 sequence were used for independent capillary reactions covering the whole length of the amplicon/insert. To estimate the lengths of polyA tracts found in the 3' L1-genome junctions of the CTRL-36 heterozygous germline L1 insertions and the somatic L1 insertion, we capillary sequenced at least 3 clones obtained from PCRs using DNA extracted from liver tissue (germline L1s) or each MDA-amplified neuron (somatic L1s). Capillary sequencing was performed by Macrogen (South Korea) and the Australian Genome Research Facility (AGRF, University of Queensland, Australia).

Chr13Δ31_{L1} Allele Identification and Reconstruction for Retrotransposition Assay

Filled site capillary sequences for Chr13Δ31_{L1} from each individual were manually assembled and analyzed with BioEdit Sequence Alignment Editor v7.0.5.2. Where Chr13Δ31_{L1} was homozygous, sequencing electropherograms obtained from PCR amplicons were screened for double peaks, which indicated single nucleotide differences between each allele. Double peaks were not found where Chr13Δ31_{L1} was heterozygous, consistent with the presence of only one template allele. At least 5 TOPO XL vector clones were sequenced from samples where Chr13Δ31_{L1} was homozygous, to ensure at least two clones of each allele was analyzed. These experiments revealed three Chr13Δ31_{L1} alleles among 7/8 individuals included in the study.

To prepare each Chr13Δ31_{L1} allele for testing in a cultured cell retrotransposition assay, we performed filled site-specific PCRs to amplify Chr13Δ31_{L1} using template liver gDNA from the 7 individuals carrying this polymorphic element, following the procedure described above for empty/filled PCRs but replacing the 5' primer with a primer that annealed to the 5' L1-genome junction and included a NotI restriction site. Amplicons were purified and cloned via the TOPO XL vector, as described above. The sequences of at least 3 TOPO XL vector clones from individual CTRL-36 (to obtain Chr13Δ31_{L1} allele 1), CTRL-30 (Chr13Δ31_{L1} allele 2), and from CTRL-28 (Chr13Δ31_{L1} allele 3), were compared against the corresponding amplicon sequences for each allele to identify PCR-induced mutations. We then developed a cloning strategy employing restriction enzymes (New England Biolabs) to obtain contiguous fragments to reassemble each allele without PCR-induced mutations, within pCEP plasmids. These plasmids contained neomycin and hygromycin resistance-based retrotransposition cassettes (Moran et al., 1996), or an eGFP-based retrotransposition cassette with puromycin resistance for plasmid selection (Ostertag et al., 2000). Ligations were performed using T4 DNA ligase (New England Biolabs) and equimolar amounts of each fragment in 1X T4 ligation buffer and 400U of T4 ligase, in a final volume of 10μL, and incubated overnight at 16°C. One Shot TOP10 Electrocomp *E. coli* cells (Invitrogen) were used for transformations.

For each assayed Chr13Δ31_{L1} allele construct, the element was reconstructed in full, including its polyA tract, instead of omitting the fragment upstream of the L1 3'UTR BstZ171 site (Moran et al., 1996; Ostertag et al., 2000). The polyadenylation signal (AATAAA) located immediately upstream of, and contiguous to, the polyA tract was removed, with only 15 adenine nucleotides of the polyA tract retained. Similarly, we generated a 5' 31nt restored Chr13Δ31_{L1} allele 1 series by including this short tract within the NotI primer used for the filled site-specific PCR. All clones were capillary sequenced to confirm their fidelity, using primers listed in Table S4. The neomycin resistance-based series were tested with and without a CMV promoter (Moran et al., 1996). In summary, we prepared three series of plasmids with the following backbones: i) pCEP4_CMVp_L1_mneol (using the pJM series from (Moran et al., 1996) and rebuilding the complete 3'UTR of the element), ii) pCEP4_L1_mneol (the prior plasmid without CMVp), and iii) pCEP4_L1_eGFP1 (without CMVp, as used in (Ostertag et al., 2000), and rebuilding the complete 3'UTR of the element).

PA-1 Cell Differentiation

PA-1 cells were differentiated as described previously (Garcia-Perez et al., 2010). Briefly, cells were seeded at a range of densities (10,000, 8,000, 3,000 and 1,000 cells/well) in 6-well plates, using 2mL of maintenance media per well. The day after seeding, the media were removed by aspiration and replaced with differentiation media, which incorporated Minimum Essential Medium (MEM) with GlutaMAX supplement, 10% Knock Out Serum Replacement (KOSR, Life Technologies), 1X non-essential amino acids, 1μM all *trans*-retinoic acid (Sigma-Aldrich) and 100U/mL of Pen-Strep solution. Differentiation media were replaced daily until wells were nearly confluent (at 4, 7, 10 and 14 days after seeding, respectively, at the densities indicated above).

hESC Neurodifferentiation

H1 hESCs were induced to differentiate to neuronal progenitor cells (NPCs) as described previously (Marchetto et al., 2010). Briefly, 80% confluent hESC cultures were cultured with N2 medium (DMEM/F12 (50/50) (Corning), 1% HEPES (VWR International), 0.5% Penicillin-Streptomycin solution, 1X GlutaMAX supplement and 1% N2 supplement (GIBCO)), supplemented with 10 μ M of Y-27632 dihydrochloride (ROCK inhibitor) (Tocris), 1 μ M of dorsomorphin dihydrochloride (Tocris) and 10 μ M of SB431542 (ReproCell). Cells were lifted after 48hr with a cell scraper and transferred to a 6-well plate and cultured in suspension while subjected to orbital shaking at 95rpm at 37°C, 5% CO₂ and ~95% humidity to yield embryoid bodies (EBs). Media were replaced every other day using N2 medium supplemented with 1 μ M of dorsomorphin dihydrochloride and 10 μ M of SB431542. EBs were dissociated manually on day 7 and seeded on matrigel-coated plates using NB media (DMEM/F12 (50/50), 1% HEPES, 0.5% Pen-Strep solution, 1X GlutaMAX supplement, 0.5% N2 supplement and 1% B-27 supplement (GIBCO)), supplemented with 20ng/ μ L of recombinant human fibroblast growth factor basic (bFGF) (R&D Systems). NPC rosettes grown at day 14 were manually selected under a microscope and dissociated with StemPro Accutase (Life Technologies). NPCs were seeded on 10 μ g/mL poly-L-ornithine hydrobromide (Sigma-Aldrich) and 2.5 μ g/mL laminin mouse protein (Thermo Fisher Scientific) coated plates and cultured in NB media supplemented with 20ng/ μ L of bFGF. Media were replaced every other day. NPC differentiation to neurons was performed as described previously (Macia et al., 2017). Briefly, $\sim 3 \times 10^6$ NPCs were seeded on 10 μ g/mL poly-L-ornithine hydrobromide and 2.5 μ g/mL laminin mouse protein 10cm-coated plates and cultured in NB media supplemented without bFGF, and supplemented with 10 μ M of ROCK inhibitor for 48 hours. Cells were then cultured in NB media for 45 days and collected using a cell scraper. DNA and RNA extractions were performed as described above, after scaling volumes accordingly, and from independent plates.

Retrotransposition Assay

We performed a cultured cell retrotransposition assay in HeLa-JVM cells using plasmids containing each of the three Chr13 Δ 31_{L1} alleles, with and without a CMV promoter, and tagged with a neomycin resistance-based retrotransposition cassette (Kopera et al., 2016). For each allele, we performed an eGFP-based retrotransposition assay using PA-1 and HEK293T cells (Ostertag et al., 2000). Retrotransposition assays were performed as described previously (Garcia-Perez et al., 2010). DNA from each construct was generated with a Plasmid Midi Kit (QIAGEN) following manufacturer's instructions, and resuspended in molecular grade sterile water within a laminar flow tissue culture hood. For the HeLa experiments using the neomycin resistance cassette, we used: L1.3 (GenBank accession L19088.1 (Dombroski et al., 1993)) as a positive control, L1.3 RT⁻ (with the reverse transcriptase activity disabling mutation D702A (Moran et al., 1996)) as a negative control, Chr13 Δ 31_{L1} alleles 1, 2 and 3, Chr13 Δ 31_{L1} allele 1 with its 31nt 5' restored, and finally a well of untransfected cells. For the PA-1 and HEK293T experiments using the eGFP reporter cassette, we assayed the same constructs as tested in HeLa cells, except only testing them without a CMV promoter, and with the addition of a pCAG-eGFP plasmid that expresses eGFP under the CAG promoter as a positive control for eGFP expression (Richardson et al., 2017).

HeLa retrotransposition assays were performed as follows: HeLa cells at ~70% confluence were dissociated using Trypsin-EDTA 0.25% (Life Technologies), incubated at 37°C for 2-5 min, resuspended by gently pipetting, and diluted with DMEM, 10% heat inactivated FBS, 2mM L-Glutamine and 100U/mL Pen-Strep solution. Cells were then counted using Trypan Blue Stain (Life Technologies) and a TC20 Automated Cell Counter (Bio-Rad). Cell seeding solution was prepared in order to plate 5,000 cells/well in a 6-well plate, with a 2mL final volume of complete media used per well. Transfection was performed ~14hr after seeding. To each well, we prepared and added a transfection mix comprising: 1 μ g of DNA diluted in 96 μ L with Opti-MEM (Life Technologies) and 4 μ L of FuGENE-HD (Promega). Plates were shaken gently to homogenize the transfection mix. As technical replicates, we used three wells plated from the same cell suspension and with the same transfection master mix. Cells were incubated at 37°C, 5% CO₂ and ~95% humidity for the course of the experiment. Media were replaced with 2mL of complete media 24hr after transfection, and then replaced with complete media plus 400 μ g/mL of G418 sulfate (Geneticin Selective Antibiotic, Life Technologies) every 48hr for 12 days. On day 14, media were aspirated and each well was washed by adding 1-2mL of DPBS, which was then aspirated. Cells were immediately fixed by adding 1mL of 1X DPBS, 0.2% glutaraldehyde and 2% formaldehyde solution and incubated at room temperature for 20min. The fixing solution was removed by inversion and cells were carefully washed with reverse osmosis-purified (RO) H₂O. Cell colonies were then stained by adding 1mL/well of 0.1% crystal violet solution and incubated at room temperature for 10min. The staining solution was aspirated, and plates were washed with RO H₂O, and air-dried before scanning.

PA-1 and HEK293T retrotransposition assays were performed using eGFP-based constructs. For the PA-1 cell assay in normal media, cells were transfected as for the HeLa experiments described above but with the following differences: we seeded 2×10^6 cells/well, and used PA-1 maintenance media MEM with GlutaMAX supplement, 10% heat-inactivated FBS, 1X NEAA and 100U/mL Pen-Strep solution. Media were replaced every day, starting from the day after transfection. Cells were selected for the presence of the plasmid on the day after transfection with 0.5 μ g/mL puromycin (Puromycin Dihydrochloride, Life Technologies) and 1 μ g/mL final concentration of puromycin each day afterward. Cells were maintained under these conditions for 5 days after transfection and eGFP expression assessed by flow cytometry on day 6. One set of PA-1 cells was incubated with 500nM final concentration of Trichostatin A (Sigma-Aldrich) in media 24hr before flow cytometry. In preparation for flow cytometry, cells were washed with 2mL of DPBS as indicated for HeLa cells, then dissociated by incubating each well with 0.5mL of Trypsin-EDTA 0.25% at 37°C for 2-5min, and then diluted with PBS, 10% heat-inactivated FBS to reach 2mL final volume to stop trypsinization. A single-cell re-suspension was obtained by gentle pipetting, and the 2mL of cell solution was filtered through a cell strainer snap cap 5mL Falcon

tube (Thermo Fisher Scientific) and centrifuged at 4°C, ~450 g for 5min. Supernatant was removed and cells resuspended in 200–300 μ L of 4°C 1X PBS. Cells were analyzed by flow cytometry in a CytoFLEX Flow Cytometer (Beckman Coulter) using 20 μ g/mL propidium iodide (Life Technologies) dead cell stain. Flow cytometry was performed at the TRI Flow Cytometry facility. For retrotransposition assays performed on PA-1 cells under differentiation conditions, normal media were replaced with differentiation media from the day after transfection. PA-1 differentiation media incorporated Minimum Essential Medium with GlutaMAX Supplement, 10% Knock Out Serum Replacement (KOSR, Life Technologies), 1X Non-Essential Amino Acids, 1 μ M all *trans*-Retinoic Acid (Sigma-Aldrich) and 100U/mL of Pen-Strep solution. HEK293T cell retrotransposition assays were performed following the same procedure as for PA-1 cells under normal conditions but using DMEM high glucose, 10% heat-inactivated FBS, 2mM L-Glutamine and 100U/mL Pen-Strep. Trichostatin A treatment was not applied to HEK293T cells.

Plasmid transfection efficiency in each cell line was tested. For each construct, 2×10^4 cells/well were seeded in a 6-well plate and transfected with 0.5 μ g of plasmid and 0.5 μ g of pCAG-eGFP, following the instructions indicated above. This analysis was performed in duplicate for each retrotransposition experiment using the same cell suspension used for seeding. Media were replaced 24hr after transfection with new media without antibiotic (only normal media were used for PA-1 cells) and eGFP was measured by flow cytometry 72hr after transfection. Cells were dissociated and analyzed by flow cytometry as described above for PA-1 and HEK293T cells. For the retrotransposition assay conducted with HeLa cells, the number of colonies counted for each construct was normalized using the transfection efficiency of the corresponding plasmid.

Luciferase Reporter Assay

The promoter sequences of L1.3, Chr13 Δ 31_{L1} allele 1, Chr13 Δ 31_{L1} allele 2 or 3 (which were identical in the relevant sequence), and Chr13 Δ 31_{L1} allele 1 and 2/3 with their 31nt 5' truncated sequence restored, were amplified by PCR from the constructs prepared for retrotransposition assays. PCR reactions were performed with primers (Table S4) incorporating HindIII site overhangs and an Expand Long Range dNTPack (Sigma Aldrich), using 1X reaction buffer with MgCl₂, 0.5mM of each dNTP, 5% DMSO, 10pmol of each primer, 2.75U of enzyme and 4ng of template DNA in a 25 μ L reaction. PCR cycling conditions were as follows: (92°C, 2min) \times 1; (92°C, 10sec; 58°C, 15sec; 68°C, 1min) \times 40; (68°C, 10min; 4°C, hold) \times 1. Amplicons were resolved via electrophoresis on a 2% agarose gel and excised and purified as described above for L1 insertion site PCR products. Promoter amplicons were then cloned into pGL3-basic (Promega), which contains the Firefly luciferase reporter gene, using HindIII enzyme (New England Biolabs). Constructs were generated with each promoter in either sense or antisense orientation. Cloning yielded the following vectors: pGL3_L1.3 s; pGL3_A1s; pGL3_A2/3 s; pGL3_A1+31 s; pGL3_A2/3+31 s; pGL3_L1.3as; pGL3_A1as; pGL3_A2/3as; pGL3_A1+31as; pGL3_A2/3+31as. A1 and A2/3 refer to Chr13 Δ 31_{L1} allele 1 and allele 2/3 5'UTRs, respectively. A1+31 and A2/3+31 refer to the Chr13 Δ 31_{L1} allele 1 and allele 2/3 5'UTRs with their 31nt 5' truncation restored. L1.3 refers to the L1.3 5'UTR. 's' and 'as' indicates the 5'UTR in sense or antisense orientation, respectively. Chr13 Δ 31_{L1} allele 1 contained an A483G substitution. pGL3 with the enhanced SV40 promoter (eSV40p) upstream of the luciferase cassette (pGL3_eSV40p) and pGL3 empty vector (Heras et al., 2013) served as positive and negative controls, respectively. *Renilla* luciferase control vector was used (Heras et al., 2013) to normalize differences in transfection efficiency between wells. Plasmid DNA was prepared using the Plasmid Midi Kit (QIAGEN) according to the manufacturer's instructions and re-suspended in molecular grade sterile water within a laminar flow hood.

Luciferase reporter experiments for each plasmid were performed using the Dual Luciferase Reporter Assay System (Promega). Briefly, we plated HeLa (2×10^4 cells/well), HEK293T (2×10^4 cells/well) and PA-1 (1×10^4 cells/well) cells in 100 μ L of the same media used for retrotransposition assays and in 96-well white plates (Corning). Cells were incubated at 37°C, 5% CO₂ and ~95% humidity for 16hr, then transfected using FuGENE-HD as described above with the following reagents per well: 20ng of each pGL3-derived plasmid, 20ng of *Renilla* control vector, and 0.15 μ L of FuGENE-HD, in a total of 4.5 μ L Opti-MEM. 24hr after transfection, media were aspirated and cells were washed with 100 μ L of DPBS. The wash was removed and cells were lysed by adding 20 μ L of 1X passive lysis buffer, followed by 15min of soft rocking. 100 μ L of Luciferase Assay Reagent II was added to each well with an 11sec delay between wells. Firefly luciferase activity was measured in a PHERAstar Fs microplate reader (BMG Labtech) with a 10sec measuring time and 1sec delay between wells. The reader was programmed to analyze the wells in the same order in which the assay reagent had been added. 100 μ L of Stop & Glo solution was added to each well and mixed by pipetting in the same order with 11sec delay between wells. *Renilla* luciferase activity was measured with the same settings. Firefly and *Renilla* luminescence was corrected by the background luminescence in a blank well with no added cell lysate. The luminescence of each well was then normalized to the corresponding *Renilla* intensity. The intensity ratio of each well was compared to the average luminescence of the pGL3-eSV40p construct. Each experiment was performed in technical triplicate (three wells of cells from the same cell suspension) and independently repeated at least once.

L1 Locus-Specific Methylation Assays

Genomic DNA extracted from bulk tissue or cultured cells was bisulfite treated using an EZ DNA Methylation Lightning kit (Zymo Research) with the following specifications: DNA input was 500ng per column, the desulfonation incubation was performed for 20min, and the product was eluted in 25 μ L elution buffer. Primers were designed against individual L1-genome junctions, other loci of interest, and the L1-Ta and L1PA2 consensus sequences (Table S4). Reactions incorporated MyTaq HS DNA Polymerase (Bio-line), and contained 1X reaction buffer, 20pmol of each primer, 2 μ L bisulfite treated DNA input template and 1U of enzyme in a 20 μ L final volume. PCR cycling conditions were as follows: (95°C, 2min) \times 1; (95°C, 30sec; 54°C, 30sec; 72°C, 30sec) \times 40; (72°C, 5min;

4°C, hold) × 1. PCR products were visualized via electrophoresis on a 2% agarose gel, followed by the excision of fragments of the expected size and conventional phenol:chloroform DNA extraction and ethanol precipitation aided by glycogen (Ambion). DNA concentrations were quantified using a Qubit dsDNA HS Assay kit. Amplicons were then pooled in equimolar quantities, with amplicons obtained from the same locus but in different samples kept in different pools. Each pool contained 500ng-1μg DNA in 30μL final volume. PCR-free libraries were prepared using a TruSeq DNA PCR-free Library Prep kit (Illumina) as follows: 20μL End Repair Mix was added to each sample and incubated at 30°C for 30min in a thermocycler (lid at 40°C). Clean-up was performed using 65μL Agencourt AMPure XP beads (Beckman Coulter). Samples were eluted in 17.5μL resuspension buffer. 3' end adenylation was performed by adding 12.5μL A-Tailing Mix to each reaction and incubated at 37°C for 30min, 70°C for 5min and 4°C for > 5min in a thermocycler (lid at 80°C). Each pool was barcoded by ligating TruSeq Illumina adapters. Ligation was performed by adding 2.5μL resuspension buffer, 2.5μL Ligation Mix and 2.5μL adaptor, and incubated at 30°C for 10min in a thermocycler (lid at 40°C). The reaction was stopped by adding 5μL Stop Solution. Two tandem clean-ups using Agencourt AMPure XP beads were performed as indicated in the library prep kit instructions. Samples were then eluted in 15μL resuspension buffer. Library concentration was measured using a KAPA Library Quantification Kit (Kapa Biosystems) with low ROX, following the manufacturer's instructions with minimal modifications. Briefly, two dilutions (1:2,000 and 1:20,000) were measured. 2μL input DNA was used in each reaction and half volumes of kit reagents were added. Quantitative PCR was performed using a ViiA 7 Real-Time PCR system (Thermo Fisher Scientific) with the following cycling conditions: (95°C, 5min) × 1, (95°C, 30sec; 60°C, 45sec*) × 40; followed by a melting curve (asterisk indicates data acquisition step). Fragment size was approximated as 550bp. Libraries were pooled in equimolar quantities prior to sequencing. Paired-end 2 × 300-mer sequencing was performed on a MiSeq platform (Illumina). 50% PhiX was used as a spike-in and cluster density was 800-1000K/mm².

Paired-end reads were assembled into contigs via FLASH (Magoč and Salzberg, 2011) using parameters -x 0.1 -m 20. Contigs with sequences at each of their termini exactly matching the primers of a target amplicon were retained and used to identify which amplicon to analyze each contig against. Primer and unconverted amplicon sequences used for data deconvolution are shown in Table S4. Each contig was then aligned to the corresponding mock bisulfite converted target amplicon with blastn (parameters -m 0 -q 1 -G 2 -E 1 -e 0.00001 -F F). Contigs where non-CpG cytosine bisulfite conversion was < 95%, or ≥ 5% CpG dinucleotides were mutated, or ≥ 5% of adenine and guanine nucleotides were mutated, were removed. For the stricter L1-Ta analysis presented in Data S1G, contigs with any mutated CpG dinucleotides, bisulfite conversion < 98%, or ≥ 2% mutated adenine or guanine nucleotides were excluded. Post-filtering, analysis was performed using the Quantification tool for Methylation Analysis (QUMA, (Kumaki et al., 2008); <http://quma.cdb.riken.jp/>) with default parameters, applying the "strict CpG site check" option and excluding PCR duplicates. 50 random sequences matching these conditions were analyzed per region of interest. For the CpH methylation analysis, QUMA alignment data from 50 random, non-duplicate sequencing reads was analyzed using the custom script 'Portable Calculator for CH methylation (PoCaCH)' (<https://github.com/MischaLundberg/ch-methylation>).

Low complexity regions within two target amplicons required variations to this analysis approach. First, the Chr22FL_{L1}-TTC28 amplicon sequence contained an imperfect polyA tract upstream of the L1-genome junction. The length of this polyA tract within contigs carrying primer sequences that matched the Chr22FL_{L1}-TTC28 amplicon sequence varied significantly, creating numerous mismatches when these contigs were aligned to the amplicon. To avoid these contigs being filtered, the genomic flank upstream of the polyA tract was removed from the unconverted sequence used for QUMA analysis and contigs were manually processed to allow only those that presented exactly the same imperfect polyA tract found in the Chr22FL_{L1}-TTC28 amplicon. QUMA analysis was performed with a maximum of 20 mismatches allowed, rather than the default of 10 mismatches. Second, for the Chr8Δ3_{L1} upstream genomic region, a low diversity region within the unconverted Chr8Δ3_{L1} amplicon sequence was removed and contigs were trimmed of this region and then joined with FLASH prior to alignment and QUMA analysis.

Genome-Wide Methylation and Hydroxymethylation Analyses

Whole genome bisulfite sequencing (WGBS) was performed upon a pool of 1 × 10⁴ NeuN⁺ nuclei isolated from CTRL-36 hippocampal tissue via flow cytometry. Briefly, genomic DNA was extracted using a DNeasy Blood & Tissue Kit (QIAGEN, Cat#69504). Next, 15ng of genomic DNA containing 0.5% (w/w) unmethylated lambda phage DNA (Thermo Fisher Scientific Cat#SD0011) was sheared to a mean length of 350bp using the Covaris S220 Shearing System and bisulfite converted using the EZ-DNA Methylation Gold kit (Zymo Research, Cat#D5005). Libraries for WGBS were prepared using the Accel-NGS Methyl-Seq DNA library kit, following the manufacturer's instructions (Swift Biosciences, Cat#30024). Bisulfite converted DNA was denatured to ensure single-strandedness, followed by adaptase treatment which simultaneously performs end repair, tailing of 3' ends, and ligation of the first truncated adaptor. Subsequently, extension and ligation steps were performed to add the second truncated adaptor to the bottom strand only, followed by PCR amplification to incorporate an index sequence and full-length adaptors with the KAPA HiFi Uracil+ ReadyMix (Roche). PCR cycling conditions were as follows: 95°C for 2min, followed by 9 cycles of 95°C for 30sec, 63°C for 30sec and 68°C for 1min, and then hold at 4°C. Paired-end 2 × 150-mer sequencing was performed on an Illumina NextSeq 500 platform, as per the manufacturer's instructions. After sequencing, reads were computationally processed as follows: adaptors were trimmed with Trimmomatic (Bolger et al., 2014) (options ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10), then polyC and polyG tails were removed before further quality trimming with Trimmomatic (options LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, and MINLEN:25). Reads were aligned to hg19 with Bowtie and BS-Seeker2 (Guo et al., 2013) (option -n 1). PCR duplicates were removed using Sambamba v.0.5.9 before CpG dinucleotide methylation quantification with BSseeker2 (Guo et al., 2013; Tarasov et al., 2015). H1 hESC WGBS and Tet-assisted

bisulfite sequencing (TAB-seq) data were obtained from GEO using identifiers GSM432685 and GSM882245, respectively. From the CTRL-36 NeuN⁺ hippocampal neuron WGBS and H1 datasets, only CpG dinucleotides covered by at least 4 reads were considered for further analysis. GSM882245 TAB-seq data were originally mapped to the hg18 human reference genome, necessitating a conversion to hg19 coordinates using the UCSC Genome Browser Batch Coordinate Conversion (LiftOver) tool. To precisely identify full-length reference genome L1 sequences, we aligned the initial 300bp of each L1 family (L1-Ta, L1PA2, L1PA3, L1PA4, L1PA5, L1PA6) consensus sequence (Khan et al., 2006) to hg19 with BLAT and default parameters, retaining alignments extending to the 3' end of each 300bp sequence and spanning at least 30bp. We then intersected the coordinates of the aligned 5' ends with RepeatMasker genomic coordinates for L1s at least 5700nt in length, retaining the RepeatMasker L1 family annotations and the 5' coordinate indicated by the BLAT alignment. This approach identified 1,846 (278 L1-Ta, 431 L1PA2, 399 L1PA3, 274 L1PA4, 235 L1PA5 and 229 L1PA6) L1s that were full-length or 5' truncated by less than 270bp. The YY1 binding site coordinates provided in Table S3 were used to divide each set into those retaining a YY1 motif (truncated by < 14nt in L1-Ta, or the equivalent position in the other L1 families) and those without a viable YY1 motif (truncated by \geq 14nt). CpG dinucleotide coordinates and methylation levels for each WGBS and TAB-seq dataset were then intersected with the genomic coordinates of the aforementioned 1,846 L1s, generating the results provided in Figures 5B and 5C.

Genome-Wide Analyses of YY1 and KAP1 Binding

ChIP-seq peaks for YY1 (ENCODE Project Consortium, 2012) and KAP1 (Turelli et al., 2014) binding in H1 hESCs, and YY1 binding in HEK293 cells (Schmitges et al., 2016), were intersected with the genomic coordinates of the 1,846 full- or near full-length L1s identified above. Published H1 hESC ChIP-seq peaks for both YY1 and KAP1 were called with the model-based analysis of ChIP-seq (MACS) algorithm (Zhang et al., 2008) and incorporated multiple replicate libraries containing $> 1 \times 10^7$ reads (ENCODE Project Consortium, 2012; Turelli et al., 2014). Figures 5A and S5A report the percentages of full-length L1s from each family that coincided with a peak from at least one of two YY1 and KAP1 ChIP-seq replicates that were used for each analysis. KAP1 and YY1 ChIP-seq peak fold enrichment among L1 families in Figure 5A was calculated by dividing the observed total count of peaks overlapping a member of each L1 family by the count obtained when the genomic positions of the same ChIP-seq peaks were randomized, and averaged over 1,000 iterations. RNA-seq library data were generated by Schmitges et al. from HEK293 cell lines that inducibly expressed 80 different zinc finger proteins, including YY1. As several read lengths were employed by this study, only the initial 50bp of each single-end read was used. Relative L1 expression was obtained by aligning RNA-seq reads directly to the initial 300bp, or equivalent position for slightly truncated elements, of the full-length L1 sequences used for the above ChIP-seq analysis. Alignments were performed via BLAT (-repMatch = 4096 -minScore = 40). Only alignments with \geq 48 matches and \leq 2 errors (mismatch, indel) were considered. If a read aligned to only one L1 family at its best (highest identify) match, then that read was assigned to that family. Counts were then normalized to the number of reads in each library, and then to the average normalized read count of all libraries. As libraries were made with two kits (TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold or TruSeq RNA Library Preparation Kit version 2), normalization was performed within library sets made with the same kit.

Detection of Chr13Δ31_{L1} Antisense Transcription

We used 2μg of total RNA from tissue samples and cultured cells to generate cDNA using Superscript III Reverse Transcriptase (Invitrogen) with Oligo(dT)₂₀ primer (50μM) (Invitrogen) following the manufacturer's instructions. A negative reverse transcriptase sample was generated for each RNA sample by replacing the Superscript III Reverse Transcriptase with H₂O in the reaction mix. RT-PCR to identify Chr13Δ31_{L1} antisense transcript isoforms was performed with primers (Table S4) targeting specific splice junctions in NR_135320 and a longer transcript (Macia et al., 2011). PCR amplification was performed using MyTaq HS DNA Polymerase in a 1X MyTaq buffer, 20pmol of forward primer, 20pmol of reverse primer, 1U MyTaq DNA polymerase and 2μL of cDNA sample in a 20μL final volume reaction. Cycling conditions were as follows: (95°C, 2min) × 1; (95°C, 30sec; 60°C, 30sec; 72°C, 30sec) × 40; (72°C, 5min; 4°C, hold) × 1. We included a PCR no template control by replacing input cDNA with H₂O in the reaction mix. PCR products were visualized via electrophoresis in a 2% agarose gel. Amplicons were excised from the gel and purified, cloned in pGEM-T easy vector, and capillary sequenced.

TaqMan RT-qPCR expression analysis of a Chr13Δ31_{L1} antisense transcript (RefSeq: NR_135320) splice junction was performed using RNA extracted from H1 and PA-1 cells using TaqPath 1-Step RT-qPCR Master Mix (Applied Biosystems). Two normalization genes (*GAPDH* and *TBP*) were used, with three technical replicates per sample and gene. Reactions contained 80ng of RNA, 4pmol of primer ε, 4pmol of primer δ and 1pmol of probe in 1X human *TBP* or *GAPDH* Pre-Developed TaqMan Assay Reagent, 1X reaction buffer and 10μL of final volume. Primer and probe sequences are provided in Table S4. *TBP* and *GAPDH* expression was tested using Pre-Developed TaqMan primer-probe sets (Applied Biosystems). RT-qPCR reactions were performed on a ViiA 7 Real-Time PCR system using MicroAmp Optical 384-well reaction plates with barcode (Applied Biosystems). Cycling conditions were as follows: (25°C, 2min) × 1 for UNG incubation; (50°C, 15min) × 1 for reverse transcription; (95°C, 2min) × 1 for polymerase activation; and (95°C, 3sec; 60°C, 30sec) × 40 for amplification. As the 1-Step kit master mix contains reverse transcriptase, to generate a reaction to test amplification from contaminant gDNA (equivalent to a negative reverse transcriptase test), we prepared a set of reactions in duplicate with the RNA samples pre-treated with RNase A (Invitrogen). RNA samples were treated with RNase A in a ratio of 5μg of enzyme for ~400ng of RNA in 10μL final volume reaction at 37°C for 30min. 80ng RNA-equivalent product of this reaction was then used as template for this negative control reaction. Standard curves were run for each primer-probe set separately, and for reactions

where target and normalization gene primer-probe sets were combined, and an optimal reaction efficiency (~100%) was confirmed in the latter.

TaqMan RT-qPCR failed to reliably detect the NR_135320 splice junction in hippocampal samples, most likely due to the low quality of RNA extracted from snap frozen tissue. Instead, we performed an end-point PCR targeting the NR_135320 splice junction using total RNA reverse transcribed to yield cDNA, as described above. Reactions involved MyTaq HS DNA Polymerase in a 1X MyTaq buffer, 20pmol of primer ϵ , 20pmol of primer δ , 1U MyTaq DNA polymerase and 2 μ L of cDNA sample in 20 μ L final volume reaction. PCR cycling conditions were as follows: (95°C, 2min) \times 1; (95°C, 30sec; 60°C, 30sec; 72°C, 30sec) \times 40; (72°C, 5min; 4°C, hold) \times 1. Reactions were resolved via electrophoresis in a 2% agarose gel. As a positive control for RNA input, we also PCR amplified a splice junction of the β -actin gene using the same conditions as above (see [Table S4](#) for primer sequences).

RNA-Seq Analysis

~4 μ g of total RNA extracted from hippocampal tissue was used to prepare strand-specific polyadenylated RNA-seq libraries using a TruSeq Stranded mRNA Library Preparation kit (Illumina). RNA libraries were prepared in at least triplicate starting from RNA from different extractions from tissue. In total, 46 libraries were generated from 8 individuals. Manufacturer's instructions were followed with the exception of reducing the time for chemical fragmentation of the RNA from 8 to 5min. Each library was barcoded, and then grouped into pools and subjected to paired-end 2 \times 150-mer sequencing on an Illumina HiSeq 2500 platform (Macrogen, South Korea). ~1.7 \times 10¹⁰ paired-end reads were generated per RNA-seq library, on average. RNA-seq reads were aligned to the hg19 reference genome using STAR ([Dobin et al., 2013](#)) in two passes to establish splice sites and subsequent spliced alignment. Quality control metrics were generated via RNA-SeQC ([DeLuca et al., 2012](#)) and reads were counted against the RefSeq gene annotations using HTSeq-count ([Anders et al., 2015](#)). Raw tag counts were imported into R and analyzed using the edgeR package ([Robinson et al., 2010](#)). Normalized counts per million (CPM) values were generated using the TMM method in edgeR.

QUANTIFICATION AND STATISTICAL ANALYSIS

Retrotransposition assays in HeLa-JVM, PA-1, and HEK293T cells presented in [Figures 2](#) and [S2](#) show the results of one representative experiment, with three technical replicates (three wells of cells were plated per assayed construct) represented by their mean and standard deviation (SD). Experimental data from HeLa and PA-1 cells were analyzed via 2-way ANOVA, with one-way ANOVA applied to data from HEK293T cells, followed by Tukey's multiple comparison test.

Luciferase reporter assay data shown in [Figures 2](#) and [S2](#) show one representative experiment with three technical replicates (three wells of cells were plated per assayed construct). These replicates are represented by their mean and SD. Statistical analysis was via a one-way ANOVA with Tukey's multiple comparison test.

Statistical differences in WGS sensitivity for CTRL-36 heterozygous L1s with pure and non-pure polyA tracts were calculated with Fisher's exact test.

Graphs in [Figure S3D](#) and [Data S1F](#) and [S1I](#) indicate CpG dinucleotide methylation percentage within each sequence, with CpGs being placed according to their nucleotide position on the x axis. To allow consistent comparison of CpG dinucleotide methylation levels in different L1s, the x axis positions of CpGs in each L1 were assigned according to the equivalent positions in the L1-Ta consensus, disregarding, for example, locus-specific indels. The graph in [Figure 6A](#) represents the mean and SD percentage methylation among the human cohort analyzed here, for each CpG dinucleotide, and in individuals where the relevant L1 was present. CpG methylation differences in the region upstream of Chr8 Δ 3_{L1}, as displayed in [Figure 6A](#), were assessed for each CpG pair using a t test with Bonferroni correction. The graph in [Figure 3C](#) represents the mean and SD of the methylation levels of Chr13 Δ 31_{L1}, Chr5 Δ 31_{L1}, Chr6 Δ 31_{L1}, Chr6FL_{L1} and Chr2 Δ 2_{L1} in comparison to L1-Ta families across liver and hippocampus in the eight human individuals. Statistical differences of each element to the L1-Ta family are calculated by a non-parametric one-way ANOVA and Tukey's multiple comparison test.

CpG dinucleotide methylation or hydroxymethylation levels for L1 family 5' ends, obtained by WGBS and TAB-seq, respectively, are represented in [Figure 5B](#) and [C](#) by box-and-whisker plots. Groups of elements 5' truncated by < 14nt or \geq 14nt within each L1 family were analyzed by a non-parametric one-way ANOVA (Kruskal-Wallis test) with a post hoc Dunn's multiple comparison test between each pair.

TaqMan RT-qPCR data obtained for Chr13 Δ 31_{L1} antisense transcripts, as shown in [Figure 5D](#) and [S5E](#), were analyzed as follows. Δ Ct was calculated for each well by subtracting the Ct value of *GAPDH* or *TBP* control genes from the Ct value of the target Chr13 Δ 31_{L1} antisense transcript (Δ Ct = Ct_{target} - Ct_{control}). The Δ Ct average of each differentiation stage was normalized to the Δ Ct average of the initial stage (hESC or undifferentiated PA-1). Negative controls generated Ct > 35. Statistical analyses comparing time points were via one-way ANOVA and Tukey's multiple comparison post hoc test.

RNA-seq expression data for 16 tissues from the Illumina Body Map 2.0, as shown in [Figures S6B](#) and [S6E](#), were downloaded from Ensembl ([Flicek et al., 2012](#)). Hippocampus RNA-seq data from 8 individuals, as shown in [Figure S6C](#), were analyzed with edgeR.

Significance values for all statistical tests were as follows: *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. Unless otherwise stated, statistical analyses were performed with GraphPad Prism version 7.00 (GraphPad Software).

DATA AND SOFTWARE AVAILABILITY

WGS, RC-seq, L1-IP, WGBS, and RNA-seq data were deposited in the European Nucleotide Archive (ENA) using the identifier PRJEB24579. TEBreak and instructions for its installation and application are available at <https://github.com/adamewing/tebreak>. The PoCalCH tool used for non CpG methylation analysis is available from: <https://github.com/MischaLundberg/ch-methylation>. Unprocessed gel images presented in this manuscript can be found at: <https://doi.org/10.17632/jm9mr476vn.1>