LINE-1 Evasion of Epigenetic Repression in Humans

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,*]

[1]Mater Research Institute - University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia.

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

[3]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom.

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

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

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

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

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

[9]Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia.

[10]Lead contact.

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

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 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.

HIGHLIGHTS

- Single-cell genomic analysis of hippocampal neurons revealed a somatic L1 insertion.
- The donor L1 was slightly 5′ truncated and lacked a conserved YY1 binding site.
- Young L1s with truncated or mutated YY1 binding sites are hypomethylated.
- L1 is able to mobilize in the brain due to locus-specific exceptions to repression.

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 ~6kb long, polyadenylated, 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 due to 5′ truncation and 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; Evrony et al., 2015; Macia et al., 2017; Muotri et al., 2005; Upton et al., 2015) and in tumor cells (Evrony et al., 2012; Ewing et al., 2015; Iskow et al., 2010; 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 100nt 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 YY1, RUNX3 and SOX2 (Athanikar et al., 2004; Coufal et al., 2009; Yang et al., 2003). L1 methylation is established during 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 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, 18yrs) without evidence of neurological disease (CTRL-36). For each nucleus, we then performed whole genome amplification (WGA) via multiple displacement amplification (MDA), followed by ~47× 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, while 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× and 49× 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 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 5.4Kb L1-Ta insertion, with a 5′ inversion/deletion (Ostertag and Kazazian, 2001), and carrying a 24nt 3′ transduction (Goodier et al., 2000; Moran et al., 1999; Pickeral et al., 2000) followed by a >140nt pure polyA tract (Figure 1D). The insertion presented a degenerate L1 endonuclease cleavage site (5′-CTTT/CC) and yielded a 20nt 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, while the 3′ junction was detected in 4 additional neurons (Figure S1B-D). In an additional 24 MDA-amplified neurons, the 5′ and 3′ junctions were each found only in neuron-#36 (Figure S1B-D). In a third set of 24 neurons, amplified via the MALBAC protocol (Zong et al., 2012), either the 5′ or 3′ junction was found in 5 neurons (Figure S1E). The L1 insertion polyA tract length varied among the neurons where it was detected, and followed a bimodal distribution, clustering around ~130nt and ~65nt (Figure S1C, E and F), corroborating reports of L1 polyA 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 (Figure S1G,J, Table S2). We observed a paucity of long, pure polyA 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). Those heterozygous L1s with pure polyA 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 polyA tracts, could influence detection sensitivity for somatic L1 insertions. As well, while the 3′ junction of the heterozygous L1 with the longest (90nt) polyA tract could be PCR amplified in ~80% of the expanded panel of 48 MDA-amplified neurons (Figure S1J), the filled site was detected in only ~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 empty/filled site PCR validated in neuron-#15, but was likely present in ~25% of the hippocampal neurons analyzed with our integrated single-cell genomic approach.

A somatically-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 31nt (Figure 1D). Strikingly, this donor L1 (named here Chr13Δ31$_{L1}$) gave rise to a somatic L1 insertion found in the cortex of another individual (Evrony et al., 2015) and was inactive when previously tested 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 Chr13Δ31$_{L1}$ (Figures 2A and S2A). Chr13Δ31$_{L1}$ was present in 7/8 individuals (Figure 2A). CTRL-36 was heterozygous for Chr13Δ31$_{L1}$, and carried only allele 1. Allele 1 encoded intact ORF1 and ORF2 sequences, whereas alleles 2 and 3 respectively carried stop codon (C5164T/Q1059Ø) 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 Chr13Δ31$_{L1}$ 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, Chr13Δ31$_{L1}$ alleles 2 and 3 were totally or nearly immobile, while allele 1 retrotransposed at ~40% and ~20% of a hot L1 (L1.3) positive control (Sassaman et al., 1997) in HeLa and HEK239T cells, respectively (Figures 2B and S2B). Restoration of the 31nt 5′ truncated sequence to allele 1 elevated its activity above that of L1.3, as did the presence of CMVp (Figure 2B). We then tested each Chr13Δ31$_{L1}$

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 31nt truncated sequence was restored (Figure 2C). Consistently, a luciferase promoter reporter assay indicated all three Chr13Δ31$_{L1}$ alleles were transcriptionally active in HeLa and HEK293T cells, and not in PA-1 cells (Figures 2D and S2C). The endogenous Chr13Δ31$_{L1}$ 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 Chr13Δ31$_{L1}$ somatic retrotransposition. We therefore developed 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 300mer Illumina sequencing allowed higher throughput and wider resolution of the L1 5′UTR CpG island compared to 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 CTRL-36 hippocampus and liver, respectively (Figure 3B). By contrast, the Chr13Δ31$_{L1}$ promoter was 39.3% and 19.5% methylated in hippocampus and liver, respectively, with numerous fully demethylated sequences in each tissue (Figure 3B). The only two other CTRL-36 germline L1-Ta copies (Chr5Δ31$_{L1}$ and Chr6Δ31$_{L1}$) 5′ truncated by 31nt were almost entirely demethylated (Chr5Δ31$_{L1}$) or fully demethylated in 5-10% of cells (Chr6Δ31$_{L1}$) (Figures 3B and S3E, Table S2). However, two heterozygous, intergenic full-length germline L1-Ta insertions (Chr6FL$_{L1}$ and Chr2Δ2$_{L1}$) were almost completely methylated (Figures 3B and S3E). We also observed this contrasting pattern in hippocampus, liver and, where available, cortex tissue obtained from the remainder of our cohort (Figure S3A-J). Chr13Δ31$_{L1}$ was strongly ($P<0.0001$, one-way ANOVA with Tukey's multiple comparison test) demethylated compared to the L1-Ta family in all 7 carrier individuals (Figures 3C and S3B,E). These results suggested 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; Walter et al., 2016). To assess Chr13Δ31$_{L1}$ methylation during neurodevelopment *in vitro*, we conducted L1 bisulfite sequencing on pluripotent H1 human 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) (Figures 3B,D and S3K). 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 (Figures 3B and S3L). Next, we identified a single nucleotide polymorphism (rs9508517) only present in the 5′ genomic flank of each Chr13Δ31$_{L1}$ allele. Bisulfite sequencing of this flank, and regions further upstream, indicated it was highly methylated in NPCs, neurons and brain tissue, regardless of whether Chr13Δ31$_{L1}$ was present (Figures 3B and S3C,D,M). In hESCs, moderate demethylation of the flanking region extended up to 500bp away from Chr13Δ31$_{L1}$, when the L1 was present, and formed a methylation "sloping shore" (Figures 3B and S3M) previously observed 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

Provided the distinct but consistent DNA methylation patterns observed for full-length and 31nt 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 31nt, 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 the L1s that were full-length or truncated by <14nt (Figures 4A,B and S4A,B). 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 ≥14nt, all apart from Chr6Δ31$_{L1}$ were <20% methylated in liver tissue (Figure 4A) and all except Chr1Δ21$_{L1}$-LRE2 were <15% methylated in hESCs (Figure 4B). Almost every fully or near-fully unmethylated sequence was found in elements truncated by

≥14nt (Figure 4A,B), 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 S4C) 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 ≥14nt 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 (Athanikar et al., 2004; Becker et al., 1993). L1s truncated by ≥14nt 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 impact 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 due to 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 S4D). 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 S4D). 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 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 L1-Ta and L1PA2 elements 5′ truncated by ≥14nt 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 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, if YY1 mediated L1 promoter methylation, loss of its binding site would principally impact young L1 families.

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 300nt of germline L1s found in the reference genome, where methylation was typically higher than further 3′ in individual L1 promoters (Figure 3B), and offered lower resolution than our locus-specific approach. Nonetheless, we determined that full-length members of each L1 family were >90% methylated (Figure 5B), in agreement with prior results (de la Rica et al., 2016). By contrast, L1-Ta and L1PA2 elements truncated by ≥14nt were significantly less methylated than full-length L1s from the same families ($P<0.001$, one-way ANOVA with Dunn's multiple comparison test), while 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 ≥14nt truncated L1-Ta and L1PA2 sequences (Figure 5B). As 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 ~10%) among each L1 family. The level of hmC was not significantly different among ≥14nt 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.

Chr13Δ31$_{L1}$ 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 Chr13Δ31$_{L1}$ ASP activity, we designed primers to target an annotated RNA (NR_135320) antisense to Chr13Δ31$_{L1}$, as well as RNAs initiated from the Chr13Δ31$_{L1}$ ASP and spliced into exons more than 30kb away (Figure S5B). Using RT-PCR and RNA extracted from differentiating PA-1 cells, we identified various transcripts initiated by the Chr13Δ31$_{L1}$ ASP (Figure S5B,C). We then targeted a commonly-used splice junction and detected Chr13Δ31$_{L1}$ antisense transcripts expressed in hippocampus or liver tissue from each Chr13Δ31$_{L1}$ carrier in our cohort (Figure S5D). TaqMan RT-qPCR indicated Chr13Δ31$_{L1}$ antisense transcript abundance and DNA methylation were inversely correlated during hESC neurodifferentiation *in vitro*, including a ~10-fold reduction in expression upon differentiation to NPCs (Figures 5D and S5E). These experiments demonstrated Chr13Δ31$_{L1}$ expression coincident 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 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 near full-length L1-Ta (Chr8Δ3$_{L1}$) located intronic to the *KCBN2* gene was earlier identified as the source of a cortical neuron somatic L1 insertion that carried a 101nt 5′ transduction (Evrony et al., 2012; Evrony et al., 2015). In our cohort, Chr8Δ3$_{L1}$ was present only in CTRL-28 and CTRL-42, and as a heterozygous polymorphism in each individual. Locus-specific L1 bisulfite sequencing indicated Chr8Δ3$_{L1}$ 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 the region upstream of Chr8Δ3$_{L1}$ was heavily demethylated in brain tissue, but not liver (Figures 6A and S6D). A transcript (DA461809) spliced shortly upstream of the Chr8Δ3$_{L1}$ 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 Chr8Δ3$_{L1}$

(Figure 6A). We therefore propose the genomic location of Chr8Δ3$_{L1}$, in a gene expressed in brain and downstream of a strong promoter element, enabled transcription and retrotransposition of a chimeric DA461809-Chr8Δ3$_{L1}$ mRNA, despite methylation of the Chr8Δ3$_{L1}$ promoter.

Another element, Chr22FL$_{L1}$-TTC28, is a fixed germline L1-Ta (Gardner et al., 2017) located antisense to the first intron of *TTC28*, which 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, Chr1Δ21$_{L1}$-LRE2 was abnormally methylated in hESCs and neuronal cells, compared to the remaining 5′ truncated L1s (Figures 4B and S4B). An L1PA13 element was located ~2.7kb upstream of Chr1Δ21$_{L1}$-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 Chr1Δ21$_{L1}$-LRE2 (Figure 4B and S4B). Together, Chr8Δ3$_{L1}$, Chr22FL$_{L1}$-TTC28 and Chr1Δ21$_{L1}$-LRE2 highlight how YY1-mediated repression may be supplanted occasionally by locus-specific regulatory mechanisms.

DISCUSSION

Our experiments indicate 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). Genome-wide analyses suggest 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 these young L1s (Castro-Diaz et al., 2014). A general lack of KAP1-associated deposition of repressive H3K9me3 upon 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. 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 involve YY1-dependent DNA methylation.

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 (Athanikar et al., 2004; Khan et al., 2006). As an activator and repressor, YY1 (Yin Yang 1) 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 ~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 Chr13Δ31$_{L1}$ 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 enrichment 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 Chr13Δ31$_{L1}$ allele 1 was found in 3/8 members of our cohort, as well as another individual where it generated a cortical neuron L1 insertion (Evrony et al., 2015), suggests many people carry this hot L1 allele, and that it is recurrently mobile in the neuronal lineage. Another element lacking a YY1 binding site due to 5′ truncation, Chr1Δ21$_{L1}$-LRE2, was discovered as 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 Chr13Δ31$_{L1}$, Chr1Δ21$_{L1}$-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, due to their genomic location. For example, the heavily methylated element Chr8Δ3$_{L1}$ mobilized in 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 due to 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 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 they arose in a neuronal lineage progenitor cell. However, owing to 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 mouse brain, without genomic analysis requiring WGA, suggesting 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 impact is 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 Chr13Δ31$_{L1}$ alleles resolves a prior discrepancy whereby 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 Chr13Δ31$_{L1}$ 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; Evrony et al., 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

sensitivity for heterozygous germline and somatic L1 insertions is similar in single-cell genomic analyses (Erwin et al., 2016; Evrony et al., 2012; Evrony et al., 2015; Upton et al., 2015). Notably, somatic L1 insertions carry long, pure polyA tails, while heterozygous L1s rarely do. Despite deep (47×) single-cell WGS, our sensitivity for somatic L1 insertions was, at most, ~15%, even without accounting for PCR validation false negatives. 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 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.

ACCESSION NUMBERS

WGS, RC-seq, L1-IP, WGBS, and RNA-seq data are available from the European Nucleotide Archive (ENA) using the identifier PRJEB24579.

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.

ACKNOWLEDGEMENTS

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's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement PIOF-GA-2013-623324 awarded to F.J.S-L., and Australian Research Council Discovery Early Career Researcher Award (DE150101117) and Discovery Project (DP170101198) grants awarded 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-2012-233764), 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 the support of the Mater Foundation, a CSL Centenary Fellowship and NHMRC Project Grants (GNT1106206, GNT1125645, GNT1126393, GNT1138795).

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. Mol Cell *31*, 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 U S A *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, O., Cheng, K.C., Todd, C.D., Cruz, C., Houseley, J., and Branco, M.R. (2016). TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. Genome Biol *17*, 234.

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 U S A *90*, 6513-6517.

Doucet-O'Hare, T.T., Sharma, R., Rodic, N., Anders, R.A., Burns, K.H., and Kazazian, H.H., Jr. (2016). Somatically 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., *et al.* (2015). Cell lineage analysis in human brain using endogenous retroelements. Neuron *85*, 49-59.

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.

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.

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*, D84-90.

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.* (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., *et al.* (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., The 1000 Genomes Project Consortium, and Devine, S.E. (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., Izsvak, 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., Eyras, E., Garcia-Perez, J.L., and Caceres, 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., *et al.* (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 retroposons. Proc Natl Acad Sci U S A *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., and Kim, J. (2009). YY1's longer DNA-binding motifs. Genomics *93*, 152-158.

Klawitter, S., Fuchs, N.V., Upton, K.R., Munoz-Lopez, M., Shukla, R., Wang, J., Garcia-Canadas, 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. (Humana Press, New York, NY), 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 U S A *111*, 8131-8136.

Kumaki, Y., Oda, M., and Okano, M. (2008). QUMA: quantification tool for methylation analysis. Nucleic Acids Res *36*, W170-175.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bioGN] *March. 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., Munoz-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., Munoz-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.

Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957-2963.

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., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. Science *283*, 1530-1534.

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.

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., Makalowski, 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., Munoz-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., Tchasovnikarova, 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.

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. (Humana Press, New York, NY), 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., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., *et al.* (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell *153*, 101-111.

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., Stromberg, M.P., Walker, J.A., Konkel, M.K., Stutz, A.M., Urban, A.E., Grubert, F., Lam, H.Y., Lee, W.P., *et al.* (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.* (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., Fenyo, D., and Boeke, J.D. (2018). Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. Proc Natl Acad Sci U S A *115*, E5526-E5535.

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.

Tubio, J.M.C., Li, Y., Ju, Y.S., Martincorena, I., Cooke, S.L., Tojo, M., Gundem, G., Pipinikas, C.P., Zamora, J., Raine, K., *et al.* (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science *345*, 1251343.

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., Sanchez-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., Perez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife *5*, e11418.

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., Munoz-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., Fuchtbauer, A.C., Fuchtbauer, 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.

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.

FIGURE LEGENDS

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 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 (symbols α, δ, γ and β) were used to amplify the L1 3′ junction (δ+β), 5′ junction (α+γ) and complete sequence (α+β). CTRL-36 templates included WGA material from neurons

#14-16, as well as bulk hippocampus (HIP) and liver (LIV) 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 with an inversion/deletion, as represented by L1-Ta consensus position numerals inside the L1. A 3′ transduction (brown box) indicated a donor L1 on chromosome 13 (Chr13Δ31$_{L1}$). Please see Figure S1, and Tables S1 and S2 for further PCR validation details.

Figure 2. Chr13Δ31$_{L1}$ allele retrotransposition activity.

(A) Chr13Δ31$_{L1}$ genotype among 8 individuals (CTRL-#). Three Chr13Δ31$_{L1}$ 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, when compared to the L1-Ta consensus (top), are shown. Non-synonymous mutations are highlighted in red.

(B) Chr13Δ31$_{L1}$ alleles in a cultured HeLa cell retrotransposition assay (Moran et al., 1996). Experimental approach involving neomycin (G418) selection is summarized at 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-), Chr13Δ31$_{L1}$ 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) Chr13Δ31$_{L1}$ 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. Grey 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 Chr13Δ31$_{L1}$ 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, Chr13Δ31$_{L1}$ allele 1; A2/3, alleles 2 and 3 (identical sequences); A1+31 and A2/3+31, alleles with 5′ truncation restored.

Note: Panels (B-D) show mean values ± SD (**, p<0.01 and ****<0.0001).

Please see Figure S2 for Chr13Δ31$_{L1}$ genotyping, and L1 reporter assays in HEK293T cells.

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 is 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 2×300mer sequencing. Note: 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; ×, 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). Note: 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 hippocampus and liver of 8 individuals. Data represent the mean percentage 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.

Please see Figure S3 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 the analysis of 50 non-identical random sequences per

amplicon. A dotted red line separates those L1s 5′ truncated by <14nt or ≥14nt. The L1 5′ end sequence is displayed above the histograms and the YY1 binding site is shown in red. Chr11Δ14$_{L1}$ and Chr1Δ21$_{L1}$-LRE2 were not present in CTRL-36.

(B) As for (A), except displaying data obtained from H1 neurodifferentiation and using 1,000 randomly sampled reads per amplicon, with the exception of Chr22FL$_{L1}$-TTC28 and Chr1Δ21$_{L1}$-LRE2, where 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. Chr6Δ6$_{L1}$, Chr22Δ12$_{L1}$ and Chr5Δ31$_{L1}$ 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; ×, mutated CpG) matching each amplicon. The percentage of methylated CpG is indicated in the lower right corner of each cartoon.

Please see Figure 3, Figure S4 and Table S4 for supporting L1 methylation data.

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 middle). 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 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 300bp 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). Box plots indicate median, quartile and extrema values for groups of elements 5′ truncated by <14nt or ≥14nt within each L1 family (***, $p<0.001$).

(C) As for (B), except 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 (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 spanned 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.

Please see Figure S5 for additional analyses of Chr13Δ31$_{L1}$ antisense transcription, and Figures 3 and S3 for Chr13Δ31$_{L1}$ bisulfite sequencing results during H1 differentiation.

Figure 6. Genomic environment influences donor L1 regulation.

(A) Chr8Δ3$_{L1}$ locus methylation. Top: an expressed sequence tag (EST: DA461809, GENCODE), indicated an upstream RNA spliced into Chr8Δ3$_{L1}$, 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). Middle: Chr8Δ3$_{L1}$ promoter and upstream methylation cartoons displaying 50 random, non-identical sequences (black circle, methylated CpG; white circle, unmethylated CpG; ×, mutated CpG). The percentage of methylated CpG is indicated in the lower right corner of each cartoon. Below: average Chr8Δ3$_{L1}$ promoter methylation in hippocampus (HIP) and liver (LIV) tissue from Chr8Δ3$_{L1}$ carrier individuals CTRL-28 and CTRL-42, and the upstream region in all 8 individuals. Values represent the mean methylation ± 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 (HIP) and liver (LIV) tissue. Values represent the mean methylation ± SD in 8 individuals. Chr22FL$_{L1}$-TTC28 was significantly hypomethylated in liver tissues (****, $p<0.0001$).

(C) Methylation of the Chr1Δ21$_{L1}$-LRE2 promoter and a ~2.7kb upstream L1PA13 copy during hESC neurodifferentiation. As indicated, the L1PA13 sequence contains an intact YY1 binding site utilized in H1 cells. Cartoon panels were generated as in (A).

Please see Figures S4 and S6 and Table S2 for supporting L1 methylation data.

(Anders et al., 2015) a (Bolger et al., 2014) a (DeLuca et al., 2012) a (Dobin et al., 2013) a (Dombroski et al., 1993) a (Ewing et al., 2015) a (Ewing and Kazazian, 2011) a (Flicek et al., 2012) a (Guo et al., 2013) a (Helman et al., 2014) a (Heras et al., 2013) a (Hormozdiari et al., 2011) a (Kopera et al., 2016) a (Kuhn et al., 2014) a (Kumaki et al., 2008) a (Li, 2013) a (Magoc and Salzberg, 2011) a (Robinson et al., 2010) a (Sanchez-Luque et al., 2017) a (Stewart et al., 2011) a (Sudmant et al., 2015) a (Tarasov et al., 2015) a (Wang et al., 2006)a (Witherspoon et al., 2010) a (Witherspoon et al., 2013)