

# Locus-resolution analysis of L1 regulation and retrotransposition potential in mouse embryonic development

Patricia Gerdes,<sup>1</sup> Dorothy Chan,<sup>1</sup> Mischa Lundberg,<sup>1,2,3</sup> Francisco J. Sanchez-Luque,<sup>1,4,5</sup> Gabriela O. Bodea,<sup>1,6</sup> Adam D. Ewing,<sup>1</sup> Geoffrey J. Faulkner,<sup>1,6</sup> and Sandra R. Richardson<sup>1</sup>

<sup>1</sup>Mater Research Institute - University of Queensland, TRI Building, Woolloongabba, Queensland 4102, Australia; <sup>2</sup>The University of Queensland Diamantina Institute, The University of Queensland, Woolloongabba, Queensland 4102, Australia; <sup>3</sup>Translational Bioinformatics, Commonwealth Scientific and Industrial Research Organisation, Sydney, New South Wales 2113, Australia; <sup>4</sup>GENYO. Centre for Genomics and Oncological Research (Pfizer-University of Granada-Andalusian Regional Government), PTS Granada, 18016, Spain; <sup>5</sup>MRC Human Genetics Unit, Institute of Genetics and Cancer (IGC), University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; <sup>6</sup>Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia

Mice harbor ~2800 intact copies of the retrotransposon Long Interspersed Element 1 (L1). The in vivo retrotransposition capacity of an L1 copy is defined by both its sequence integrity and epigenetic status, including DNA methylation of the monomeric units constituting young mouse L1 promoters. Locus-specific L1 methylation dynamics during development may therefore elucidate and explain spatiotemporal niches of endogenous retrotransposition but remain unresolved. Here, we interrogate the retrotransposition efficiency and epigenetic fate of source (donor) L1s, identified as mobile in vivo. We show that promoter monomer loss consistently attenuates the relative retrotransposition potential of their offspring (daughter) L1 insertions. We also observe that most donor/daughter L1 pairs are efficiently methylated upon differentiation in vivo and in vitro. We use Oxford Nanopore Technologies (ONT) long-read sequencing to resolve L1 methylation genome-wide and at individual L1 loci, revealing a distinctive “smile” pattern in methylation levels across the L1 promoter region. Using Pacific Biosciences (PacBio) SMRT sequencing of L1 5' RACE products, we then examine DNA methylation dynamics at the mouse L1 promoter in parallel with transcription start site (TSS) distribution at locus-specific resolution. Together, our results offer a novel perspective on the interplay between epigenetic repression, L1 evolution, and genome stability.

[Supplemental material is available for this article.]

Retrotransposons are major contributors to ongoing mutagenesis in mammalian genomes. The autonomous non-long terminal repeat (non-LTR) retrotransposon Long Interspersed Element 1 (LINE-1 or L1) is actively mobilizing in both humans and mice, and L1 sequences occupy ~17% of human DNA and ~18% of mouse DNA (International Human Genome Sequencing Consortium 2001; Waterston et al. 2002). While humans contain a single active L1 subfamily, the mouse genome harbors three active L1 subfamilies, termed T<sub>F</sub>, G<sub>F</sub>, and A (Wincker et al. 1987; Schichman et al. 1993; Casavant and Hardies 1994; DeBerardinis et al. 1998; Naas et al. 1998; Saxton and Martin 1998; Hardies et al. 2000; Goodier et al. 2001; Mears and Hutchison 2001) which are each further divided into several sublineages, for example, T<sub>FI</sub>, T<sub>FII</sub>, and T<sub>FIII</sub> (Sookdeo et al. 2013). Although T<sub>F</sub> and G<sub>F</sub> elements descended from the old and now inactive F subfamilies, A elements evolved independently. However, generating a single phylogenetic tree describing the relation of the subfamilies with each other is

difficult because of frequent recombination among elements (Sookdeo et al. 2013). The vast majority of the ~600,000 L1 copies in the mouse genome reference are 5' truncated and mutated (Voliva et al. 1983; Waterston et al. 2002), leaving approximately 2800 full-length L1s (Penzkofer et al. 2017). Ongoing L1 activity has generated substantial variation in L1 content among inbred strains, as well as interindividual variation in L1 content within strains (Akagi et al. 2008; Nellåker et al. 2012; Richardson et al. 2017; Schauer et al. 2018; Gerdes et al. 2022; Ferraj et al. 2023). L1 insertions also are responsible for several spontaneous mouse mutants, driven by L1 T<sub>F</sub> elements in all cases in which the L1 subfamily can be identified (Gagnier et al. 2019).

A full-length mouse L1 is ~6–7 kb long and begins with a 5' untranslated region (5' UTR) containing an internal RNA polymerase II promoter (Severynse et al. 1991; DeBerardinis and Kazazian 1999). The mouse L1 5' UTR has a distinctive structure, wherein a variable number of tandemly repeated ~200 bp monomer units are situated upstream of a nonmonomeric region (Adey et al. 1994; Kong et al. 2022). Each monomer contributes

**Corresponding authors:** [faulknergj@gmail.com](mailto:faulknergj@gmail.com), [sandra.richardson@mater.uq.edu.au](mailto:sandra.richardson@mater.uq.edu.au)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.278003.123>. Freely available online through the *Genome Research* Open Access option.

© 2023 Gerdes et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

additive promoter activity (DeBerardinis and Kazazian 1999). Individual monomers of young L1 subfamilies generally comprise sufficient CpG dinucleotides to qualify as CpG islands (Lee et al. 2010), and also contain several transcription factor binding sites, including for YY1 transcription factor (YY1) (DeBerardinis and Kazazian 1999; Lee et al. 2010). The YY1 binding site is required for accurate L1 transcription initiation (Athanihar et al. 2004; Lee et al. 2010), and an intact YY1 binding site also is important for methylation of the human L1 promoter during cellular differentiation (Sanchez-Luque et al. 2019).

The L1 5' UTR is followed by two open reading frames encoding the proteins ORF1p and ORF2p, and a 3' UTR incorporating a polyadenylation signal (Scott et al. 1987; Skowronski et al. 1988; Dombroski et al. 1991). ORF1p is ~40 kD and harbors RNA binding and chaperone activities (Holmes et al. 1992; Hohjoh and Singer 1996; Martin and Bushman 2001; Khazina and Weichenrieder 2009, 2018) whereas the ~150 kD ORF2p has showed endonuclease (EN) and reverse transcriptase (RT) activities (Mathias et al. 1991; Feng et al. 1996; Ergün et al. 2004; Doucet et al. 2010; Taylor et al. 2013). Both proteins are required for L1 mobilization through reverse transcription of an RNA intermediate in a process termed target-site primed reverse transcription (TPRT) (Scott et al. 1987; Holmes et al. 1992; Luan et al. 1993; Feng et al. 1996; Moran et al. 1996). While L1 retrotransposition can occur in nondividing cells (Kubo et al. 2006; Macia et al. 2017), a growing body of evidence has emerged linking L1 retrotransposition to DNA replication during the S phase of the cell cycle (Mita et al. 2018, 2020; Flasch et al. 2019). Hallmarks of L1 integration by TPRT include flanking target site duplications (TSDs), and the incorporation of a 3' poly(A) tract which reflects the necessity of L1 mRNA polyadenylation for efficient retrotransposition (Grimaldi et al. 1984; Doucet et al. 2015).

Unchecked retrotransposition presents a threat to genome stability, and is countered by a variety of host defense mechanisms (Bourc'his and Bestor 2004; Goodier 2016; MacLennan et al. 2017; Liu et al. 2018; Deniz et al. 2019; Greenberg and Bourc'his 2019; Mita et al. 2020; Tristán-Ramos et al. 2020; Senft and Macfarlan 2021). While repressive histone marks such as H3K9me3 play a major role in silencing older mouse L1 subfamilies (Tan et al. 2013; Castro-Diaz et al. 2014; Jacobs et al. 2014), younger L1s are typically silenced by methylation of the CpG islands in their promoters (Furano et al. 1988; Hata and Sakaki 1997; Lee et al. 2010; de la Rica et al. 2016; Gerdes et al. 2022). During embryonic development, the epigenome undergoes reprogramming including phases of global DNA demethylation (Hajkova et al. 2002; Seki et al. 2005; Abe et al. 2011; Saitou et al. 2012; Seisenberger et al. 2012; Smith et al. 2012; Cantone and Fisher 2013). The developmental methylation dynamics of L1 promoters have been examined using subfamily-specific and whole genome bisulfite sequencing (WGBS) (Hajkova et al. 2002; Kuramochi-Miyagawa et al. 2008; Watanabe et al. 2008; Popp et al. 2010; Saitou et al. 2012; Seisenberger et al. 2012; Smith et al. 2012; Molaro et al. 2014; Schöpp et al. 2020; Zoch et al. 2020). However, because of the repetitive structure and variable length of the mouse L1 promoter, as well as the high sequence identity among young L1 copies in the genome, assignment of short internal reads to specific mouse L1 loci is challenging (Lanciano and Cristofari 2020), and assessment of L1 methylation status *en masse* may mask individual L1s whose methylation dynamics differ from those of their subfamily. Indeed, studies of human L1s suggest certain loci can “escape” methylation and thus contribute to somatic retrotransposition throughout development and in cancer (Pitkänen et al. 2014; Tubio et al. 2014; Paterson et al.

2015; Philippe et al. 2016; Scott et al. 2016; Gardner et al. 2017; Nguyen et al. 2018; Schauer et al. 2018; Salvador-Palomeque et al. 2019; Sanchez-Luque et al. 2019; Ewing et al. 2020). Locus-specific resolution of murine developmental L1 methylation, however, remains largely unexplored. Heritability of locus-specific retrotransposon methylation has been found in the context of “metastable epialleles” mostly involving variably methylated young IAP elements (VM-IAPs) (Bertozzi and Ferguson-Smith 2020). However, genome-wide screens have not revealed evidence of this phenomenon for L1s (Kazachenka et al. 2018; Elmer et al. 2021).

We previously characterized five *de novo* (daughter) L1 insertions arising in pedigrees of inbred mice (Richardson et al. 2017), and identified their source (donor) elements through unique L1 3' transductions (Holmes et al. 1994; Moran et al. 1999; Goodier et al. 2000; Pickeral et al. 2000; Xing et al. 2006; Beck et al. 2010). In this study, we use the mosaic tissues from the animals in which these insertions arose and their heterozygous insertion-bearing offspring, to investigate the retrotransposition potential and epigenetic fate of *de novo* and retrotransposition-competent L1 copies *in vivo*. We also use *in vitro* differentiation of mouse embryonic stem cells (mESCs) as a model to explore developmental L1 methylation dynamics at locus-specific resolution and genome-wide. We use cell culture-based L1 retrotransposition assays, locus-specific bisulfite sequencing, Oxford Nanopore Technologies (ONT) long-read DNA sequencing and methylation profiling, and Pacific Biosciences (PacBio) SMRT sequencing of L1 5' RACE cDNAs to explore the relationship between developmental DNA methylation and L1 expression and retrotransposition capacity at single-locus resolution.

## Results

### L1 retrotransposition efficiency is diminished by ongoing promoter shortening

To evaluate the retrotransposition potential of *de novo* and polymorphic daughter elements relative to their donors, we amplified via polymerase chain reaction (PCR), cloned and capillary sequenced five donor/daughter pairs (Richardson et al. 2017) to derive the exact nucleotide sequence of each element (Sanchez-Luque et al. 2019) (see Methods). All 10 L1s contained at least one T<sub>F</sub> monomer unit and encoded intact open reading frames (ORFs), and each daughter L1 contained between 0.6 and 1.8 fewer monomer units than the corresponding donor L1 (Table 1; Fig. 1A; Supplemental Fig. S1A). The remaining sequence of each daughter L1 was identical to its donor, with the exception of Insertion 2 which had a single nonsynonymous substitution in ORF1 (V303A) (Fig. 1A; Supplemental Fig. S1A). This result, representing a single nucleotide substitution among 33,374 reverse transcribed bases, is consistent with a high fidelity for mouse L1 reverse transcriptase activity *in vivo*.

The loss of daughter element 5' UTR sequence could be explained either by 5' truncation during retrotransposition (Ostertag and Kazazian 2001; Symer et al. 2002; Zingler et al. 2005), or by the use of TSSs within internal monomers of the donor element promoter (DeBerardinis and Kazazian 1999). We analyzed the putative initiator dinucleotide (−1,+1) (Carninci et al. 2006) for the 10 elements under study (Fig. 1A), as well as 10 additional L1 T<sub>F</sub> elements (Supplemental Fig. S1B) comprising five likely donor/daughter pairs (Table 1). Under the assumption that the most 5' position of each element represents the first transcribed nucleotide,

**Table 1.** Characterization of L1 donor/daughter pairs

L1 donor/ daughter pairs	Coordinates <sup>a</sup>	Location	L1 subfamily	Intact ORFs	Monomer #	5' truncation point <sup>b</sup>	TSS	References
Donor 2	Chr 19: 28,149, 495–28,156,131	Intergenic	T <sub>FI</sub>	Y	3.59	85 (0)	CT	Richardson et al. 2017
Insertion 2	Chr 2: 101,251, 997–101,252,012	Intergenic	T <sub>FI</sub>	Y	1.8	41 (–38)	CC	Richardson et al. 2017
Donor 5	Chr 1: 7,328, 682–7,335,740	Intergenic	T <sub>FI</sub>	Y	5.61	82 (0)	CA	Richardson et al. 2017
Insertion 5	Chr 10: 89,133, 398–89,133,411	Intronic ( <i>Ano4</i> )	T <sub>FI</sub>	Y	5.01	208 (–83)	CG	Richardson et al. 2017
Donor 7	Chr 6: 95,658, 065–95,664,762	Intronic ( <i>Sucg2</i> )	T <sub>FI</sub>	Y	3.82	37 (–41)	CG	Richardson et al. 2017
Insertion 7	Chr 18: 85,067, 455–85,067,469	Intergenic	T <sub>FI</sub>	Y	2.61	82 (0)	CA	Richardson et al. 2017
Donor 3	Chr 16: 65,055, 909–65,062,518	Intronic ( <i>Htr1f</i> )	T <sub>FI</sub>	Y	3.49	108 (+21)	TA	Richardson et al. 2017
polyL1Tf_3	Chr 6: 108,844, 479–108,844,492	Intronic ( <i>Edem1</i> )	T <sub>FI</sub>	Y	2.69	64 (–15)	CA	Richardson et al. 2017
Donor 4	Chr 7: 58,421, 728–58,428,359	Intergenic	T <sub>FI</sub>	Y	3.41	125 (+38)	CA	Richardson et al. 2017
polyL1Tf_4	Chr 17: 41,334, 114–41,334,129	Intergenic	T <sub>FI</sub>	Y	2.61	82 (0)	CA	Richardson et al. 2017
Donor L1_Lama2	Chr 13: 4,065, 522–4,076,041	Intronic ( <i>Akr1c14</i> )	T <sub>FI</sub>	Y <sup>c</sup>	4.56	93 (+6)	GG	Gagnier et al. 2019
L1_Lama2	Chr 10: 27,124, 144–27,124,153	Intronic ( <i>Lama2</i> )	T <sub>FI</sub>	N	3.22	166 (+79)	CA	Besse et al. 2003
Donor L1_HCC	Chr 1: 86,584, 921–86,584,935	Intronic ( <i>Cops7b</i> )	T <sub>FI</sub>	Y	3.57	92 (+5)	CG	Schauer et al. 2018
L1_HCC	Chr 1: 12,998, 087–12,998,074	Intergenic	T <sub>FI</sub>	n/d	2.55	96 (+9)	AC	Schauer et al. 2018
Donor L1 <sub>spa</sub>	Chr 12: 91,936, 221–91,943,749	Intergenic	T <sub>FI</sub>	Y	7.83	36 (–43)	CC	Not previously described
L <sub>spa</sub>	Chr 3: 80,861, 509–80,861,524	Intronic ( <i>G1rb</i> )	T <sub>FI</sub>	Y	7.51	103 (+16)	GA	Kingsmore et al. 1994; Naas et al. 1998
Donor L1_Chr6	Chr 18: 10,875, 468–10,882,471	Intergenic	T <sub>FI</sub>	Y	5.35	137 (+50)	TG	Not previously described.
L1_Chr6	Chr 6: 82,256, 067–82,262,701	Intergenic	T <sub>FI</sub>	N	3.61	82 (0)	CA	Not previously described
Donor L1_Chr1	Chr 6: 24,673, 352–24,678,785	Intergenic	T <sub>FI</sub>	Y	2.65	75 (–4)	CG	Not previously described
L1_Chr1	Chr 1: 32,542, 809–32,542,822	Intronic ( <i>Khdrbs2</i> )	T <sub>FI</sub>	N <sup>d</sup>	1.33	142 (+55)	CA	Not previously described

<sup>a</sup>Coordinates for nonreference insertions are the first and the last nucleotides of the TSD; coordinates for the reference insertions are the first and the last nucleotides delimiting the L1 in the mm10 reference genome.

<sup>b</sup>5' truncation position in the 5'-end monomer, numbers in brackets indicate how far upstream (–) or downstream (+) of YY1 core motif (GCCATCTT) truncation occurred, “0” indicates truncation within YY1 core motif.

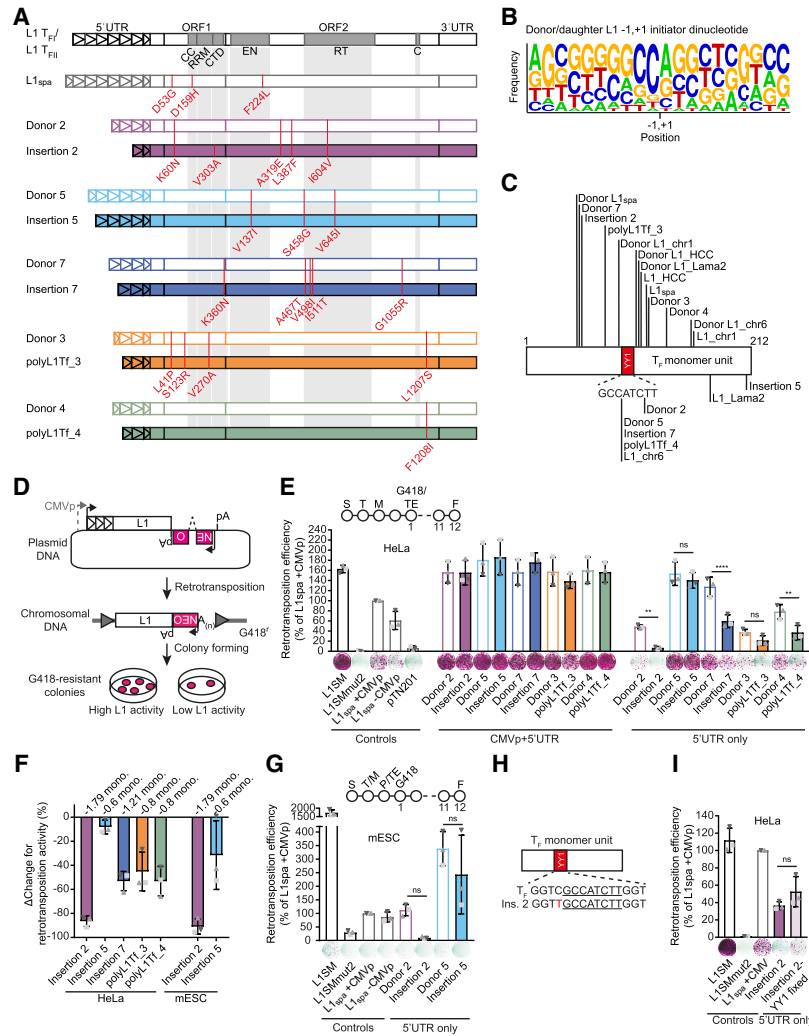
<sup>c</sup>IAP sequence inserted into ORF2 leading to a slightly truncated ORF2p with intact EN and RT domains.

<sup>d</sup>Sequence obtained from ONT sequencing could be inaccurate.

transcription of 14/20 elements initiated at the preferred mammalian PolII initiator pyrimidine/purine dinucleotide (Carninci et al. 2006) (Fig. 1B; Supplemental Fig. S1C; Table 1). 15 of the 20 elements analyzed here were 5' truncated within the first 108 nt of the 5'-most T<sub>F</sub> monomer. 5/20 elements truncated within, and an additional 7/20 truncated in close proximity ( $\leq 21$  bp) to, the YY1 core binding motif (GCCATCTT) (Fig. 1C; Table 1) as previously observed for mouse L1s (Shehee et al. 1987; DeBerardinis and Kazazian 1999; Zhou and Smith 2019). The observed clustering of 5' truncation points, and their coincidence with the Py/Pu initiator dinucleotide, are consistent with the daughter L1 promoters being shortened due to transcription initiation internal to the 5' UTR of the donor element.

To quantify the impact of monomer loss on daughter element mobility, we evaluated the five donor/daughter pairs identified in our previous study (Richardson et al. 2017) in a cultured cell L1 retrotransposition assay (Moran et al. 1996; Wei et al. 2000). We tested each element driven either by a cytomegalovirus promoter and

the native L1 promoter (CMVp + 5' UTR), or by the native L1 promoter only (5' UTR only), and quantified their activity relative to L1<sub>spa</sub> (Fig. 1D; Kingsmore et al. 1994; Naas et al. 1998). In HeLa cells, when driven by CMVp + 5' UTR, all elements mobilized efficiently (~160% of L1<sub>spa</sub>), with donor and daughter elements showing similar activity. Notably, Insertion 2 which has an amino acid change in ORF1p retrotransposed with the same efficiency as its donor (~160%) when transcribed from the CMVp, indicating that the mutation does not influence retrotransposition efficiency. In contrast, when driven by the 5' UTR alone, each daughter element mobilized less efficiently than its donor (Fig. 1E,F). This trend was most pronounced for Insertion 2, which retrotransposed at 8% of L1<sub>spa</sub> + CMVp, compared to 62% for Donor 2, and reached statistical significance for three of the five donor/daughter pairs (One-way ANOVA followed by Sidak's post-hoc test,  $P=0.0060$ , 0.7632, <0.0001, 0.5652, 0.0066 for donor/daughter pairs 2, 5, 7, 3 and 4, respectively; Fig. 1E,F). A similar trend was observed when donor/daughter pairs 2 and 5 (5' UTR only) were tested in



**Figure 1.** L1 donor/daughter pairs retrotranspose efficiently in vitro. (A) Amino acid changes in ORF1 and ORF2 compared to the L1<sub>T<sub>FI</sub></sub> and L1<sub>T<sub>FI</sub></sub> consensus sequences (Sookdeo et al. 2013) are annotated in red. L1<sub>spa</sub> refers to the published disease-causing insertion used in later experiments (Naas et al. 1998). Functional domains in ORF1 and ORF2 are shown: CC = coiled-coiled, RRM = RNA recognition motif, CTD = C-terminal domain, EN = endonuclease, RT = reverse transcriptase, C = cysteine-rich motif. Triangles within the 5' UTR represent monomer units. For nucleotide substitutions in promoters see Supplemental Figure S1. (B) Sequence logo (Crooks et al. 2004) of putative transcription initiation start sites for all ten donor and daughter L1 pairs. Sequence represents transcription initiation dinucleotide in the center ± 9 nucleotides upstream and downstream. -1,+1 indicates transcription initiator dinucleotide. The first nucleotide of L1 sequence corresponds to the second nucleotide in transcription initiator dinucleotide. The position of the first base of each daughter L1 relative to its donor element, and the first base of each donor L1 relative to the L1<sub>T<sub>FI</sub></sub>/L1<sub>T<sub>FI</sub></sub> consensus sequences was analyzed. (C) Donor and daughter L1 5' truncation points. Sequences or locations of donors and daughters were previously published as indicated in Table 1 (Kingsmore et al. 1994; Naas et al. 1998; Besse et al. 2003; Richardson et al. 2017; Schauer et al. 2018; Gagnier et al. 2019). Lines indicate truncation points of elements in the 5'-most monomer. YY1 binding site (GCCATCTT) is shown in red. (D) Rationale of a cultured cell retrotransposition assay (Moran et al. 1996; Wei et al. 2000). Constructs used in this study were previously published (Moran et al. 1996; Goodier et al. 2001; Han and Boeke 2004) or generated by modifying the pTN201 construct [L1<sub>spa</sub> (Naas et al. 1998)]. An antisense orientated neomycin-resistance (NEO<sup>r</sup>) reporter cassette interrupted by a sense-oriented intron is inserted into a mouse L1 3' UTR. The mouse L1 is driven by its native 5' UTR promoter or a CMV promoter (CMVp). Cells harboring a retrotransposition event become neomycin (G418) resistant. The colony number reflects the relative activity of the L1 construct. (E) Comparison of L1 donor/daughter pair retrotransposition efficiency in HeLa cells. The retrotransposition assay timeline is shown in the top (S: seeding, T: transfection, M: change of media, G418: start of G418 selection, TE: measurement of transfection efficiency, F: Fixing and staining of colonies). Constructs: L1SM (positive control), L1SMmut2 (negative control), pTN201, L1<sub>spa</sub> +CMVp/-CMVp, L1 donor/daughter pairs +CMVp/-CMVp. Colony counts were normalized to L1<sub>spa</sub> + CMVp and are shown as mean ± SD of three independent biological replicates, each of which comprised three technical replicates. (\*)  $P \leq 0.0332$ , (\*\*)  $P \leq 0.0021$ , (\*\*\*)  $P \leq 0.0002$ , (\*\*\*\*)  $P < 0.0001$ , ns = not significant (One-way ANOVA followed by Sidak's post-hoc test,  $P = 0.0060, 0.7632, < 0.0001, 0.5652, 0.0066$  for donor/daughter pairs -CMVp from left to right). Representative well pictures are shown below each construct.  $5 \times 10^3$  cells were plated per well in a six-well plate. (F) Percentage change ( $\Delta$ Change) in retrotransposition activity between L1 donor/daughter pairs. Shown is the decrease of retrotransposition efficiency per daughter L1 compared to its respective donor L1. Data is shown as mean ± SD of three independent biological replicates. (G) Comparison of L1 donor/daughter pair retrotransposition efficiency in mESCs. The retrotransposition assay timeline is shown at the top (S: seeding, T: transfection, M: change of media 8 h after transfection, P: passaging of cells into 10 cm plates, TE: measurement of transfection efficiency, G418: start of G418 selection, F: Fixing and staining of colonies). Constructs as described in (E). Colony counts were normalized to L1<sub>spa</sub> + CMVp and are shown as mean ± SD of three independent biological replicates, each of which comprised two technical replicates. (\*)  $P \leq 0.0332$ , (\*\*)  $P \leq 0.0021$ , (\*\*\*)  $P \leq 0.0002$ , (\*\*\*\*)  $P < 0.0001$ , ns = not significant (One-way ANOVA followed by Sidak's post-hoc test,  $P = 0.2851, 0.3305$  for donor/daughter pairs -CMVp from left to right). Representative well pictures are shown below each construct.  $4 \times 10^5$  cells were plated per well in a six-well plate. (H) Schematic of an L1 monomer unit. The YY1 binding site is indicated as red rectangle. The extended YY1 binding motif sequence is shown below. The core YY1 binding motif sequence is underlined. A mutation in the extended YY1 binding motif sequence adjacent to the core motif in Insertion 2 is indicated in red. (I) Comparison of retrotransposition efficiency of Insertion 2 and Insertion 2 with intact YY1 binding sites (Insertion 2-YY1 fixed) in retrotransposition assay in HeLa cells. Constructs as described in (D). Colony counts were normalized to L1<sub>spa</sub> + CMVp and are shown as mean ± SD of three independent biological replicates, each of which comprised three technical replicates. (\*)  $P \leq 0.05$ , ns = not significant (two-tailed t-test,  $P = 0.1607$ ). Representative well pictures are shown below each construct.  $1 \times 10^4$  cells were plated per well in a six-well plate. Note: L1SM retrotransposed very efficiently, leading to cell colony crowding in wells, and a likely underestimate of retrotransposition.

mESCs (MacLennan et al. 2017), but did not reach statistical significance for either pair (One-way ANOVA followed by Sidak's post-hoc test,  $P=0.2851$ ,  $0.3305$  for pairs 2 and 5, respectively; Fig. 1F,G). We also noted that Insertion 2 had a single nucleotide mutation (1115C>T) in the extended YY1 binding motif (GGTCGCCATCTTGGT) in its second monomer (Fig. 1H) which could decrease YY1 binding affinity (Kim and Kim 2009). To determine whether the 1115C>T mutation impacts the retrotransposition efficiency of this element, we restored the YY1 binding site (Insertion 2-YY1 fixed), and observed a ~15% increase in retrotransposition activity which was not statistically significant (two-tailed  $t$ -test,  $P=0.1607$ ; Fig. 1I). Together, these results are consistent with previous luciferase reporter assays showing mouse L1 promoter strength is proportionate to monomer number (DeBerardinis and Kazazian 1999). Moreover, we find that monomer loss consistently diminishes the retrotransposition potential of de novo mouse L1 insertions relative to their donor elements.

### Donor and daughter L1 insertions are largely methylated in adult tissues

Having established the retrotransposition competence of the donor and daughter elements in vitro, we next used mouse L1 locus-specific bisulfite sequencing (Schauer et al. 2018) to evaluate the methylation status of each L1 in the somatic tissues and gonads of the mosaic animal in which the daughter insertion arose, and in subsequent generations of heterozygous insertion-bearing animals (Fig. 2A–D). For comparison, we analyzed the genome-wide methylation of the  $T_{FI}$  and  $T_{FII}$  subfamily monomers using primers internal to the monomer sequence (Schauer et al. 2018). Because of their high sequence similarity, it was not possible to design primers specific to the  $T_{FI}$  or  $T_{FII}$  subfamily. Overall, the  $T_{FI}/T_{FII}$  subfamily monomer sequence was >80% methylated in adult tissues, although a few demethylated reads were observed across animals and tissues (Fig. 2E; Supplemental Fig. S2A).

The three de novo daughter L1s (Insertions 2, 5, and 7) (Table 1) were >80% methylated in the somatic tissues and gonads of adult mice, including mosaic animals E (Insertion 5), CD14 (Insertion 7), and their heterozygous F1, F2, and F3 descendants (Fig. 2C,F; Supplemental Fig. S2B). Notably, Insertion 2 originated as germline-restricted mosaic in mouse B and was transmitted only to F1 animal AB15, which was harvested as a postimplantation embryo. Insertion 2 was highly methylated in the adult testis of mouse B, but partially demethylated in embryo AB15 (Fig. 2C,F). As the genomic DNA of embryo AB15 was analyzed in bulk, the demethylated sequences may potentially correspond to primordial germ cells (PGCs) or multipotent stem cells. Consistently, the  $T_{FI}/T_{FII}$  subfamily monomer sequence and Donor 2 (see below) also showed partial demethylation in embryo AB15 (Fig. 2E). We also analyzed the methylation status of unfixed polymorphic insertions polyL1Tf\_3 and polyL1Tf\_4 (Table 1). Insertion polyL1Tf\_3 was nearly 100% methylated across all adult tissues examined (Fig. 2G; Supplemental Fig. S2B). However, methylation of polyL1Tf\_4 was more relaxed (<90%), with a tendency to be especially demethylated in liver (Fig. 2G; Supplemental Fig. S2B). This variability may reflect the influence of genomic location and physiological context on L1 element methylation (Salvador-Palomeque et al. 2019; Sanchez-Luque et al. 2019; Ewing et al. 2020). Together, these results indicate that de novo L1 insertions arising during embryonic development are likely silenced by DNA methylation during later embryogenesis. This methylation is maintained in subsequent generations, with an average methylation

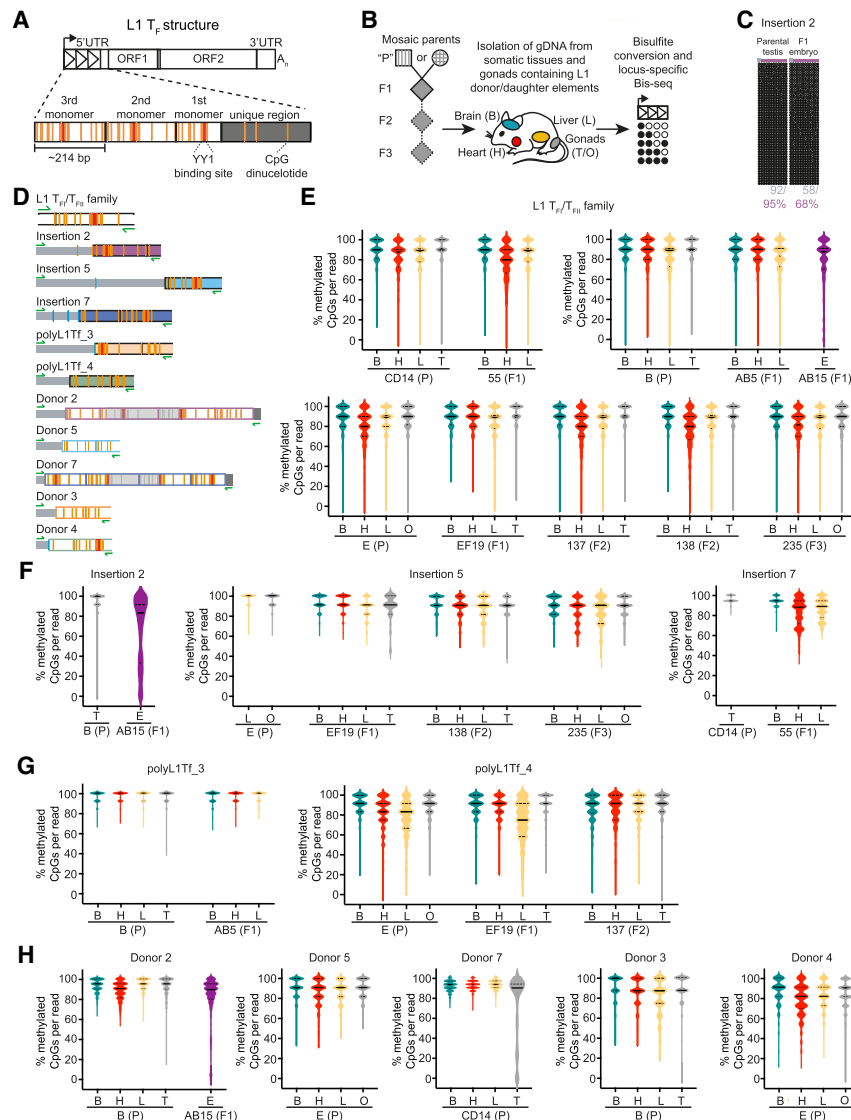
level of 93% in brain, 89% in heart and 86% in liver for daughter insertions and 91% in brain, 87% in heart and 88% in liver for the donor L1s.

We viewed donor L1s active during embryonic development as candidate “escapee” loci in mice, potentially akin to specific human L1 loci that evade epigenetic repression in differentiated cells (Scott et al. 2016; Salvador-Palomeque et al. 2019; Sanchez-Luque et al. 2019; Ewing et al. 2020). We therefore assessed methylation of Donor 2, Donor 5, and Donor 7 in somatic tissues and germ cells of their mosaic founder animals (mouse B, mouse E, and mouse CD14, respectively). We also analyzed methylation of Donor 3 and Donor 4 in animals that carried the respective polymorphic daughter insertions polyL1Tf\_3 and polyL1Tf\_4 (Fig. 2B,D). Nearly all donor elements showed >80% methylation in somatic tissues and gonads (Fig. 2H; Supplemental Fig. S2C). The exception was Donor 7 which, although it was completely methylated in the somatic tissues of founder mouse CD14, was hypomethylated in the germ cell fraction of mouse CD14 testis. This is a notable departure from the genome-wide state of  $T_{FI}/T_{FII}$  monomer sequences, which we found to be largely methylated in adult gonads (Fig. 2E,H; Supplemental Fig. S2A,C). Thus, Donor 7 may represent an L1 that is refractory to methylation during germline development and therefore privileged for heritable retrotransposition.

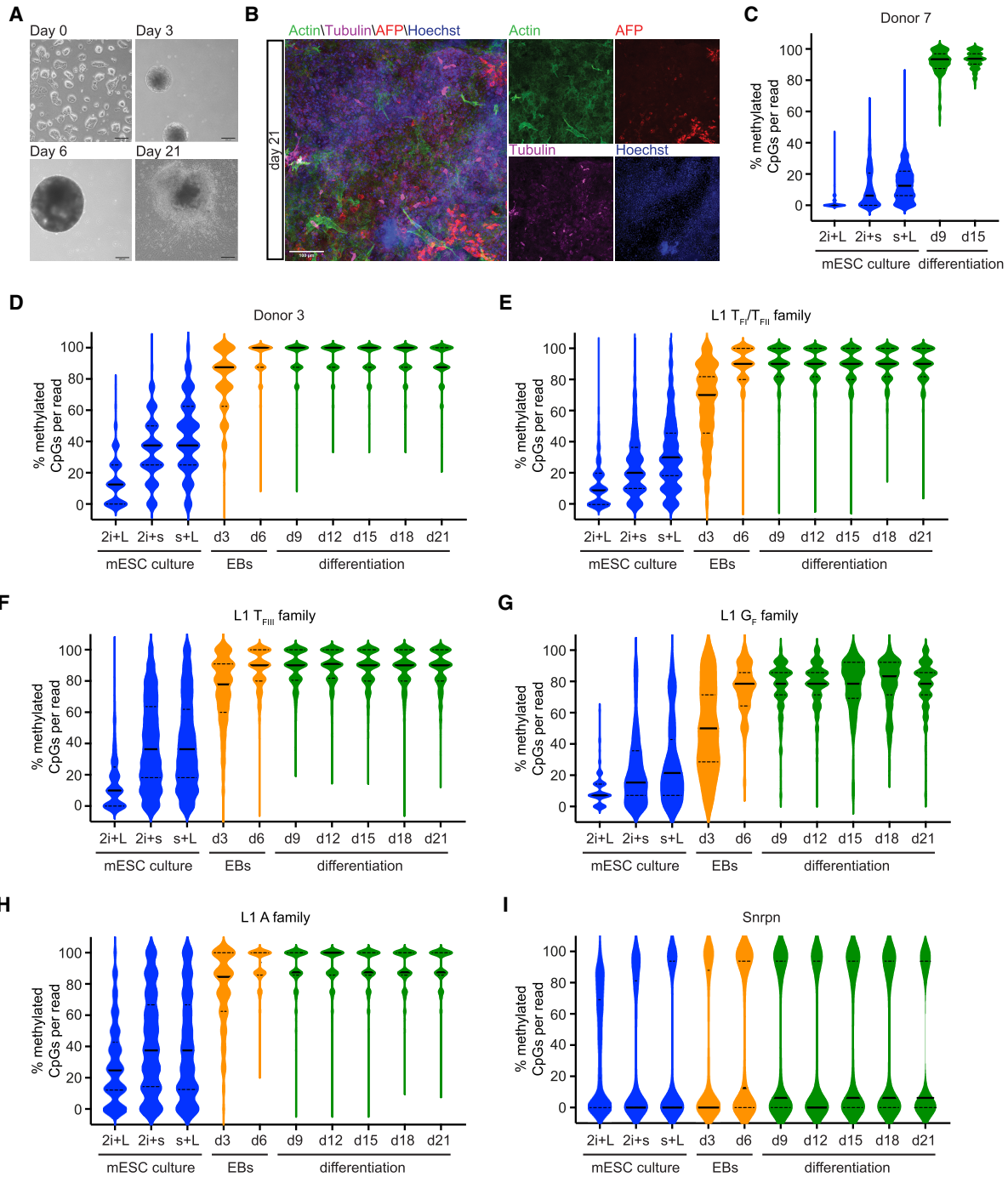
### Young L1s are rapidly methylated during in vitro mESC differentiation

As adult tissues reveal only the end point of developmental L1 methylation dynamics, we next analyzed L1 methylation at genome-wide and locus-specific resolution during cellular differentiation. To model various states of pluripotency, we cultured feeder-free E14 mESCs in three conditions: serum complemented with leukemia inhibitory factor (serum + LIF), which generates a heterogeneous population of mESCs in a pluripotent state (Smith et al. 1988; Williams et al. 1988); two small kinase inhibitors + LIF (2i + LIF), which maintains the mESCs in a naive ground state similar to that of the inner cell mass (ICM) (Silva et al. 2008); and under 2i + serum conditions, which are shown to support engineered mouse L1 retrotransposition (MacLennan et al. 2017). To recapitulate specification and differentiation of cells into the three germ lineages (ectoderm, mesoderm, and endoderm), we collected genomic DNA over a time course, from differentiation induction of serum + LIF mESCs to embryoid bodies (EBs) for 6 d, and through subsequent differentiation over 15 d (Fig. 3A,B; Supplemental Fig. S3A,B; Behringer et al. 2016). L1 methylation state was assessed using genome-wide and locus-specific L1 bisulfite sequencing in the three mESC culture conditions and on days 3, 6, 9, 12, 15, 18 and 21 of differentiation.

The mobile L1 subfamilies  $T_F$ ,  $G_F$  and A each showed their lowest methylation levels in 2i + LIF ground-state conditions, and reached maximal methylation by day 6 of differentiation (Fig. 3E–H; Supplemental Fig. S4C–F). Notably, the  $G_F$  subfamily was less methylated (<80%) than the other subfamilies (>80%) both during differentiation and in fully differentiated EBs (Fig. 3G; Supplemental Fig. S4F). The 129/Ola genetic background from which E14 mESCs are derived contained two of the donor L1s analyzed above, Donor 3 and Donor 7. Both loci were largely demethylated (<50%) across all mESC culture conditions (Fig. 3C,D; Supplemental Fig. S4A,B). Donor 3 showed 80% methylation at day 3 of EB differentiation, and >90% at day 6. Donor 7 showed >90% methylation at day 9 and day 21 (Fig. 3C; Supplemental Fig. S4A). These results were in line with our



**Figure 2.** L1 donor/daughter elements are methylated in somatic tissues of adult mice. (A) Schematic of CpG dinucleotides in a mouse L1  $T_F$  element. Triangles in 5' UTR represent monomer units. A magnification of the 5' UTR is shown below. Black boxes represent monomer units. Dark gray box represents unique (nonmonomeric) region within 5' UTR. Orange strokes represent CpG dinucleotides. Red boxes represent YY1 binding sites. (B) Experimental design of mouse L1 locus-specific bisulfite sequencing. Genomic DNA was extracted from tissues of C57BL6/J mice harboring previously identified donor and daughter L1 insertions (Richardson et al. 2017). The parental generation "P" (square = male, circle = female) is mosaic for the de novo daughter L1 insertion (represented by stripes). F1-F3 generations are heterozygous for L1 insertions (filled diamond = male or female). F2 and F3 generation animals were only available for Insertion 5 and polyL1Tf\_4. Tissues for analysis of donor elements were only collected from the P generation. DNA was isolated from brain (green), heart (red), liver (orange), and gonads (gray) if available. After bisulfite conversion, the 5' monomeric region of each L1 was PCR amplified. Amplicons were pooled and sequenced as  $2 \times 300$ -mer Illumina reads. Circles represent methylated (black circles) and unmethylated (white circles) CpG dinucleotides in L1 5' UTR. (C) Methylation of a de novo L1 promoter sequence (Insertion 2) shown in the germline mosaic parental testis (animal B) and in the following F1 generation embryonic tissue (animal AB15). Displayed are 50 nonidentical sequences extracted at random from a much larger pool of available Illumina reads. Each cartoon panel corresponds to an amplicon (black circle, methylated CpG; white circle, unmethylated CpG; x, mutated CpG). Colored line above the cartoon represents amplicon (gray = genomic sequence, colored = L1 sequence). The overall percentage of methylated CpG dinucleotides is indicated below each cartoon. Gray letters indicate methylation of CpG dinucleotides in genomic sequence. Colored letters indicate methylation of CpG dinucleotides in L1 sequence. (D) Locus-specific methylation analysis schematic representation for L1  $T_F$  monomer, 3 full-length de novo L1 insertions (Insertion 2, 5, 7), 2 polymorphic L1 insertions (polyL1Tf\_3 and polyL1Tf\_4) and their 5 respective donor elements (Donor 2, 5, 7, 3, 4). 5' monomeric sequences of each L1 were PCR amplified using primer pairs (green arrows) specific to that locus. Orange strokes indicate L1 CpG dinucleotides covered by the assay. Blue strokes represent covered genomic CpG dinucleotides. Gray strokes in the gray shaded area represent CpG dinucleotides not reached by Illumina sequencing. Red boxes indicate YY1 binding sites. Colored boxes represent L1 monomer units. (E) Genome-wide methylation of L1  $T_F/T_{FI}$  promoter sequence shown in all animals. Animals are labeled E, EF19, 137, 138, 235, B, AB5, AB15, CD14 and 55 in the bottom part of the x-axis label, with the generation (P, F1, F2 or F3) indicated in brackets. Each graph contains animals from the same family. The different tissues used for DNA extraction and bisulfite sequencing in each animal is indicated in the top part of the x-axis labeled as: brain, B; heart, H; liver, L; testis, T; ovaries, O; and embryonic tissues, E. Displayed are 1000 nonidentical sequences extracted at random from a much larger pool of available Illumina reads. The violin plots represent the methylation distribution as per Supplemental Figure S2. The black line and dashed lines show the distribution median and quartiles, respectively. The percentage of methylated CpGs per read is indicated on the y-axis. (F) As for (E) but for de novo L1 promoter sequences shown in the mosaic P generation in which each de novo L1 insertion was identified (animals E and CD14) and in following F1-F3 generations if available (animals EF19, 138, 235 and 55). Displayed are 1000 nonidentical sequences (if available) extracted at random from a much larger pool of available Illumina reads (exceptions: Insertion 5: 138 H, 235 B, H, L, O, EF19 T [368, 328, 337, 352, 958, 282 reads, respectively]; Insertion 7: 55 B, H, L, CD14 T [598, 885, 890, 662 reads, respectively]). (G) As for (E,F) but for polymorphic L1 insertions. (H) As for (E,F) but showing methylation of 5 donor L1 elements in the animal they mobilized in P generation).



**Figure 3.** Dynamic methylation of L1 elements during differentiation of mESCs. (A) Differentiation of mESCs to cells of all three germ layers using a standard differentiation protocol (Behringer et al. 2016). Undifferentiated E14 mESCs are grown on gelatin (day 0). Embryoid bodies (EBs) are generated by “hanging drop culture” (day 3) and are grown in suspension culture (day 6). After 6 d, EBs are plated and differentiated for 2 wk (day 21). Scale bar, 200  $\mu$ m. (B) Immunofluorescence image of mesodermal (Actin, green), endodermal (AFP, red) and ectodermal (Tubulin, violet) lineage markers in differentiated E14 mESCs on day 21. Nuclei were stained with Hoechst (blue). Scale bar, 100  $\mu$ m. (C) Methylation of Donor 7 promoter sequence shown in the mESCs cultured in three different conditions (2i+L=2i+LIF, 2i+s=2i+serum, s+L=serum+LIF) and on differentiation day 9 (d9) and day 15 (d15). Because of the technical challenge posed by PCR amplification of long bisulfite-treated fragments, sufficient material was generated to assess Donor 7 methylation only at day 9 and day 15 of differentiation. Displayed are 1000 nonidentical sequences (if available) extracted at random from a much larger pool of available Illumina reads (exceptions: Donor 7: 2i+LIF, 2i+serum, serum+LIF, d9, d15 [874, 837, 876, 110, 124 reads, respectively]). The violin plots represent the methylation distribution as per Supplemental Figure S4. The black line and dashed lines show the distribution median and quartiles, respectively. (D–I) As for (C) but for Donor 3 (D), L1  $T_{FI}/T_{FII}$  family (E), L1  $T_{FI}$  family (F), L1  $G_F$  family (G), L1 A family (H), and the imprinted gene *Snrpn* (I). Primers for L1 families are within the L1 promoter sequence. Shown is methylation in three different mESC culture conditions, during EB culture and during differentiation. Displayed are 1000 nonidentical sequences (if available) extracted at random from a much larger pool of available Illumina reads (exceptions: L1  $G_F$  family: 2i+LIF, 2i+serum, serum+LIF, d3, d6, d9, d12, d15, d18, d21 [86, 127, 119, 120, 168, 162, 161, 155, 155, 175 reads, respectively]).

observation that both donors were completely methylated in somatic tissues of adult mice (Fig. 2H) and is consistent with methylation occurring during differentiation in embryonic development *in vivo*. As an internal control, analysis of the maternally imprinted *Snrpn* gene revealed the expected bimodal distribution of methylation in pluripotent cells and across the differentiation time course (Fig. 3I; Supplemental Fig. S4G). Together, these results show rapid remethylation of L1 sequences during cellular differentiation, with subtle but notable variability between active L1 subfamilies and among individual loci.

### Methylation fluctuates within mouse L1 promoters

Although locus-specific bisulfite sequencing allows base-pair resolution of individual L1 methylation, it is limited to the 5'-most portion of each L1 promoter. To attain complete methylation profiles of L1 loci without bisulfite conversion, we performed PCR-free ONT sequencing of mESCs in serum + LIF (day 0; d0), day 3 (d3) EBs, and day 21 (d21) differentiated cells. We achieved ~15–20× genome-wide depth using an ONT PromethION platform, and surveyed CpG methylation via the methylartist package (Ewing et al. 2020; Cheetham et al. 2022). Evaluated by ONT sequencing, DNA methylation increased during differentiation (Fig. 4A). Examining the methylation profiles of each L1 subfamily averaged across their 5' UTRs, we found that in d0 mESCs the 5' UTRs of full-length  $T_{FI}$ ,  $G_F$  and A subfamily L1s were less methylated compared to the older L1 F subfamily (Fig. 4A). Average methylation of all L1 subfamilies rapidly increased to ~80% by day 3 of differentiation; however,  $T_{FI}$  and  $T_{FII}$  elements were less methylated ( $\leq 80\%$ ) in d3 EBs compared to other L1 subfamilies (Fig. 4A). B1, B2, and BC1 SINES were only slightly demethylated in d0 mESCs (~80%), and fully methylated (>90%) in d3 EBs and d21 fully differentiated cells (Fig. 4B). While IAP retrotransposon LTRs were generally almost 100% methylated at all three differentiation time points, the IAP subfamily IAPEY4 LTRs showed an average methylation level of <40% in d0 mESCs and only ~70% methylation in d3 EBs and d21 differentiated cells. MERVL/MT2 LTRs were >80% methylated at all three time points. Etn elements showed an average methylation level of only ~60% in d0 mESCs but >80% in d3 EBs and d21 differentiated cells (Fig. 4C).

Next, we quantified the methylation of active L1 subfamily promoters independently of the L1 body (unique 5' UTR region, ORFs, 3' UTR). We binned promoters genome-wide based on monomer count, with the majority of young L1 subfamily members containing between 2 and 5 monomer units (Fig. 4D), consistent with previous analyses of the mouse reference genome (Zhou and Smith 2019). On the whole, L1 promoter methylation was distributed between ~0% and 90% in the d0 mESCs, indicating variability among loci even in pluripotent serum + LIF culture conditions. Methylation of L1 loci in d3 EBs was higher than in d0 mESCs but a substantial proportion of L1s were still <80% methylated. Notably, none of the L1 loci displayed here appeared to be completely demethylated in d3 EBs. Methylation of the majority of L1s was re-established (>80%) in d21 differentiated cells (Fig. 4D). However, some loci appeared to be <70% methylated even in d21 differentiated cells. We examined three of these methylation “escapees” and found for each a mixture of methylated and demethylated reads at d21 potentially belonging to specific cell types in our mixed population of differentiated cells (Supplemental Fig. S5) indicating again a likely lineage or cell type specificity for L1 methylation “escapees” (Salvador-Palomeque et al. 2019; Sanchez-Luque et al. 2019; Ewing et al. 2020). Based on the mouse genome reference sequence all three L1s con-

tain intact ORFs and multiple (4–7) monomers, indicating their potential retrotransposition competence.

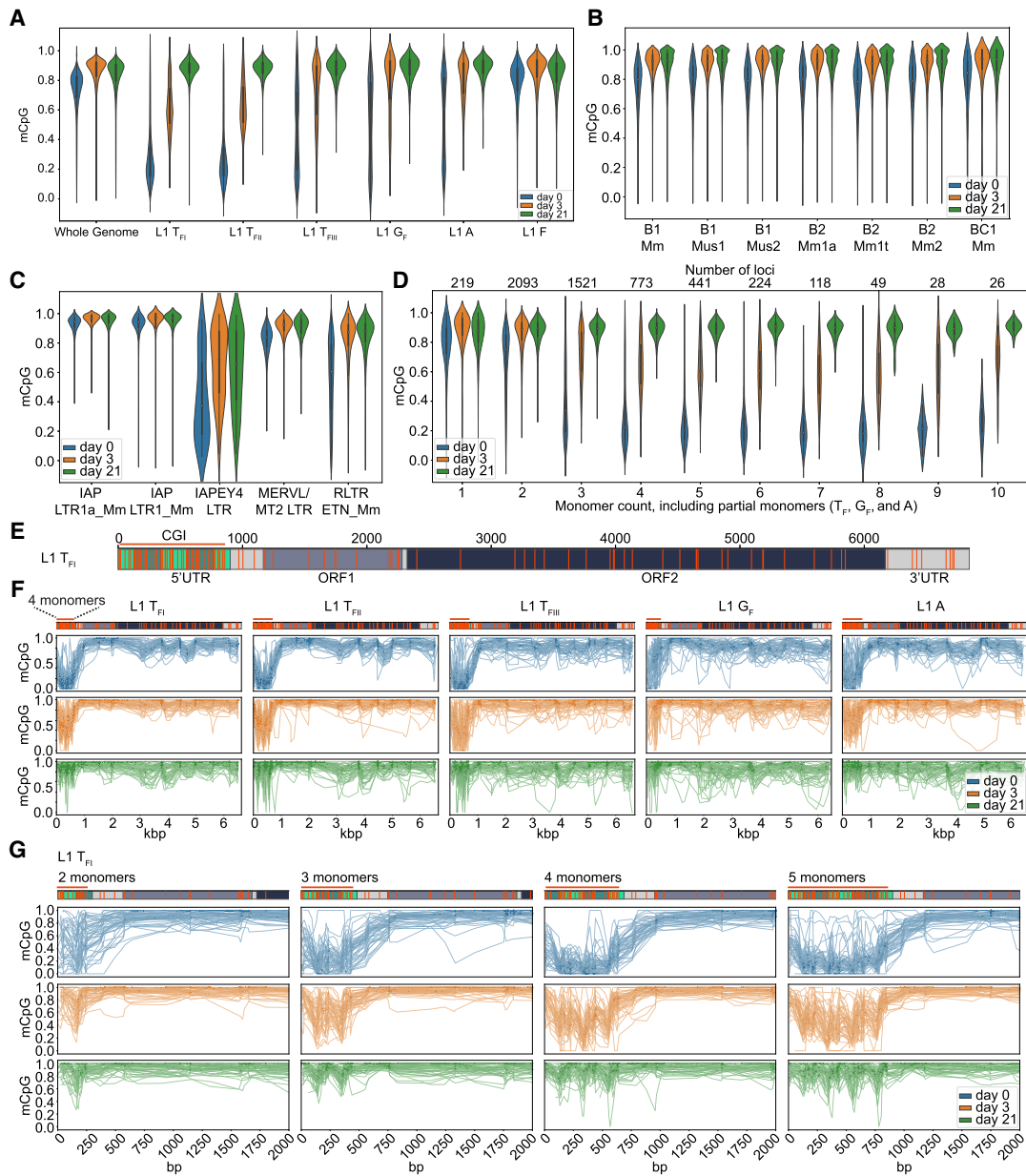
We next assessed composite methylation profiles covering the previously inaccessible interiors of full-length mouse L1s and in particular the entire mouse L1 promoter (Fig. 4E). We observed a consistent methylation trough in the 5' UTR promoter region in d0 mESCs and d3 EBs, whereas the L1 body was consistently hypermethylated at all three time points (Fig. 4F). Zooming in on the promoter region of the composite L1 methylation profiles revealed a consistent “smile” pattern across the 5' UTR, with the innermost monomers less methylated compared to the 5'-most and 3'-most monomers (Fig. 4G; Supplemental Fig. S6). This methylation pattern was most pronounced in  $T_{FI}$  and  $T_{FII}$  elements, and in elements containing three or more monomer units (Fig. 4G; Supplemental Fig. S6A–D). Elements with two monomers appeared highly variable in their methylation status in d0 mESCs (Fig. 4G; Supplemental Fig. S6A–D). Our analyses also revealed peaks and valleys of methylation along mouse L1 promoters, with a periodicity corresponding to the monomer units. Taking advantage of the locus-specific resolution offered by ONT sequencing, we examined methylation of the two donor L1s (Donor 3 and Donor 7) present in E14 mESCs (Supplemental Fig. S7). These elements recapitulated the methylation trough observed in the composite plots in d0 mESCs and d3 EBs. As an internal control, we readily identified the differentially methylated regions (DMRs) of two imprinted genes, *Snrpn* and *Impact* (Supplemental Fig. S8).

To elucidate the pattern of DNA methylation within the L1  $T_{FI}$  5' UTR at single nucleotide resolution, we determined the median percent methylation at each CpG dinucleotide, as well as the average percent methylation for all CpG dinucleotides (%mCpG) within each monomer unit, among at least 20 individual L1  $T_{FI}$  loci containing 2, 3, 4, or 5 monomers, (Fig. 5; Supplemental Fig. 9A–C). This analysis allowed us to quantify the “smile” pattern described above. For example, at day 3 among L1 loci containing five monomer units, the average %mCpG per monomer from the genome-proximal monomer inwards was 68.1%, 51.5%, 38.7%, 50.3%, and 73.5% (Fig. 5, middle panel); a similar pattern was observed for L1 loci containing 2, 3, and 4 monomers (Supplemental Fig. S9). Although methylation was highly variable at most CpG positions, the CpGs flanking the monomer-monomer borders were consistently hypomethylated relative to the CpGs internal to the monomer units. This trend was consistent across L1  $T_{FI}$  elements regardless of monomer count (Fig. 5; Supplemental Fig. S9). We also observed that methylation levels peak at the CpGs flanking the YY1 binding site within each monomer unit, with a slight dip in methylation around the core YY1 binding site. In sum, our ONT methylation analysis of full-length L1s provides unprecedented resolution of interior mouse promoter methylation dynamics.

### DNA methylation during differentiation impacts mouse L1 TSS distribution

The fluctuating L1 promoter methylation patterns during mESC differentiation led us to investigate the influence of DNA methylation on L1  $T_{FI}$  TSS usage. We performed 5' rapid amplification of cDNA ends (5' RACE) on total RNA isolated from d0, d3, and d21 differentiated mESCs, followed by PCR with an L1  $T_{FI}$  ORF1-specific primer in an approach similar to that used by Deininger et al. to study human L1 expression (Fig. 6A; Deininger et al. 2017). Purified 5' RACE products were sequenced on a PacBio SMRT flow cell. Reads were filtered to retain those aligned to only one

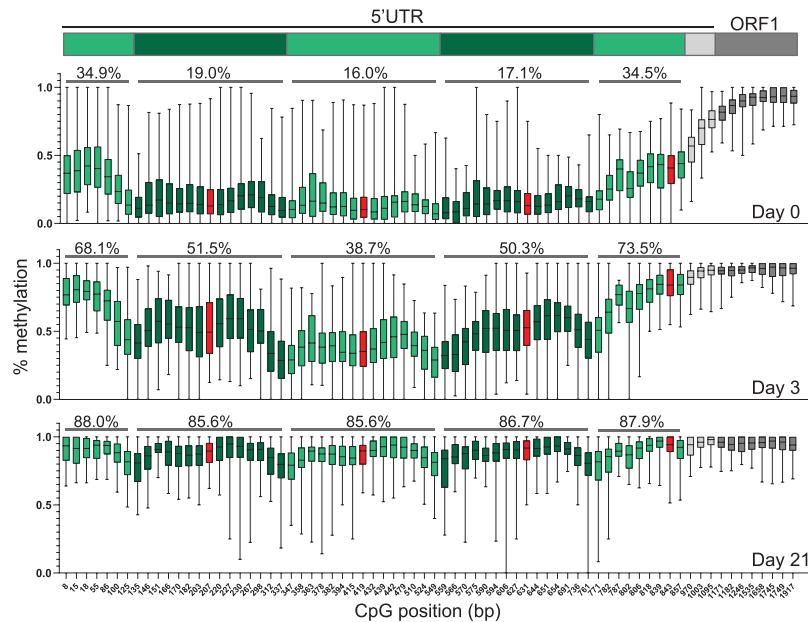




**Figure 4.** ONT CpG methylation profiles of TEs. (A) Violin plots are showing methylated CpG fraction for the whole genome (6 kbp windows), L1 5' UTRs belonging to the active L1  $T_{FI}$ ,  $T_{FIIL}$ ,  $T_{FIIL}$ ,  $G_F$  and A subfamilies and the evolutionary older and inactive L1 F subfamily at three time points of differentiation: d0 (undifferentiated mESCs in serum + LIF), d3 (EBs on day 3 of differentiation) and on d21 (completely differentiated cells). (B) As for (A), but for B1, B2 and BC1 SINE subfamilies. (C) As for (A), but for IAP LTR1a\_Mm and LTR1\_Mm, IAPEY4 LTR, MERV-L/MT2 LTR and RLTR ETn\_Mm copies. (D) As for (A), but for violin plots showing the methylated CpG fraction of active L1 subfamily ( $T_F$ ,  $G_F$ , and A together) promoters (monomers only) depending on the number of monomers (including partial monomers). Only elements with a minimum coverage of five reads across the whole 5' UTR were included in the plot. The number of loci represented in each bin is shown in the top. (E) Annotated full-length L1  $T_{FI}$  consensus showing the monomer units in green, unique region in light gray, ORF1 in dark gray, ORF2 in dark green, and 3' UTR in light gray. CpG dinucleotides throughout the whole element are displayed as orange strokes. The promoter CpG island (CGI) is indicated as an orange line. Number of bp are shown above the element. (F) Data is shown for full-length L1s ( $T_{FI}$ ,  $T_{FIIL}$ ,  $T_{FIIL}$ ,  $G_F$  and A) containing four monomers in the promoter at three time points of differentiation: d0 (undifferentiated mESCs in serum + LIF), d3 (EBs on day 3 of differentiation), and on d21 (completely differentiated cells). Each graph displays up to 50 methylation profiles for the specified L1 subfamily. Annotated consensus sequences as per (E) are shown at top including CpG positions. (G) As for (F), but for promoters of L1  $T_{FI}$  subfamily members containing between 2 and 5 monomers at three time points of differentiation (d0, d3, and d21).

reference genome L1  $T_{FI}$  element, leveraging internal L1 sequence polymorphisms to identify unique TSSs supported by at least one read. At all three time points, TSSs within L1  $T_{FI}$  monomers clustered around the YY1 binding site as previously reported, consistent with our sequence analyses of donor/daughter L1 pairs

(Figs. 1C, 6B; DeBerardinis and Kazazian 1999). The most prominent peak at all three time points was at position 65 of the  $T_F$  monomer consensus sequence, 15 bp upstream of the YY1 binding site (positions 80–87) (Fig. 6B). This finding recalls the situation in the human L1 5' UTR, in which YY1 directs transcription initiation



**Figure 5.** Methylation of individual CpG dinucleotides across the L1  $T_F$  promoter. Box-and-whisker plots display the median percent methylation determined by ONT sequencing for individual CpG dinucleotides across the L1  $T_F$  promoter for at least 20 individual L1 loci containing five monomer units, at d0 (top), d3 (middle), and d21 (bottom) of differentiation. The CpG positions along the x-axis are derived from a representative five monomer L1  $T_F$  sequence used in our analysis (Supplemental Table S1). The central line represents median percent CpG methylation; box indicates interquartile range. Whiskers represent the top and bottom quartiles. Alternating green shading indicates CpGs belonging to each monomer unit, corresponding to the schematic of the L1  $T_F$  5' UTR, above. The CpG dinucleotide partially encompassed by the core YY1 binding site is shown in red. Above each monomer unit for each box plot is shown the average percent methylation among all  $\geq 20$  L1 loci across all CpGs present within the monomer unit.

upstream of the YY1 binding site, near the +1 site of the L1 5' UTR (Athaniar et al. 2004). Plotting the putative initiator dinucleotide at the  $-1/+1$  position indicated preference for Py/Pu at all three time points (Supplemental Fig. S10A). While in d3 EBs L1  $T_{FI}$  transcripts remained abundant, the total number of TSSs corresponding to L1  $T_{FI}$  elements diminished at d21 (Fig. 6A,B), despite the similar sequencing depth applied to each sample. This potential reduction in L1  $T_{FI}$  transcription at d21 would reconcile with the observed global increase in L1 5' UTR methylation at this time point (Figs. 4, 5). At all three time points, we identified 5' RACE products that initiated in genomic sequence upstream of an L1  $T_{FI}$  element; analyzed separately these reads also displayed a preference for a Py/Pu dinucleotide at the  $-1/+1$  position (Supplemental Fig. S10B). Notably, the proportion of upstream initiating transcripts increased substantially in fully differentiated cells at day 21, and the number of loci with reads corresponding to upstream TSSs was 1669 at day 0, 1603 at day 3, 1334 at day 21 (Fig. 6C). The distance between upstream TSSs and their L1s ranged from 1 bp to 420 kb, and the median distance was 163 bp at d0, 115 bp at d3 and 1039 bp at d21. Distances greater than the PacBio read length (mean 1341 bp) likely reflect splicing of the mRNA molecule as previously observed for human L1s (Sanchez-Luque et al. 2019).

We selected two L1 loci with upstream TSSs to examine in greater detail. We analyzed 5' RACE reads uniquely mapping to Donor 7 (Chr 6: 95,658,065–95,663,747) and observed upstream TSSs at all three time points, with most TSSs located within a SINE B2 element (Supplemental Fig. S11A). Upstream TSSs comprise 30% (20/70) of the reads mapping to this locus at d0, 17% (3/18)

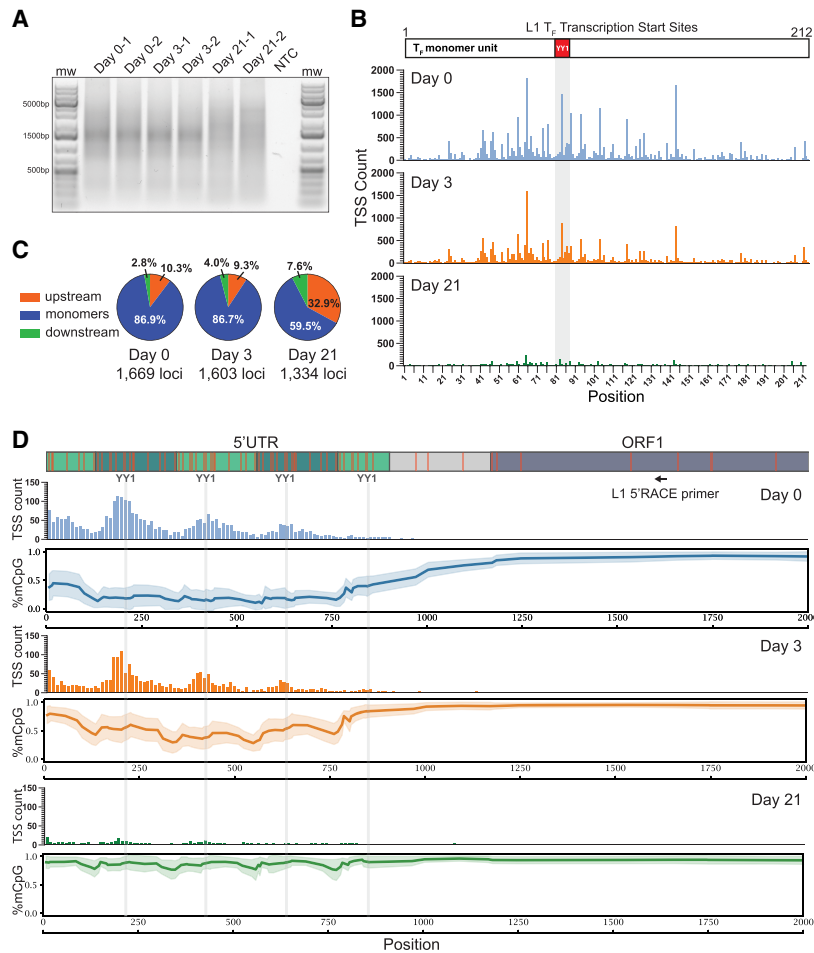
at d3, and 63% (5/8) at d21, exemplifying the general shift to upstream TSSs during differentiation. We also examined 5' RACE reads uniquely mapping to a full-length L1  $T_F$  element with intact open reading frames (Chr 6: 22,125,162–22,131,805) situated in sense orientation within an intron of the gene *Cped1* (Supplemental Fig. S11B). For this locus, we observed 71% (34/48) upstream TSSs at d0 and 68% (17/25) at d3. At d21, 97% (37/38) of 5' RACE reads for this locus supported upstream TSS usage, with only a single read initiating within the L1 5' UTR. At d0 and d3, most upstream TSSs were located within microsatellite repeats directly adjacent to the L1 element, whereas at d21 most TSSs were located further upstream. Two d21 upstream initiating reads contained exonic *Cped1* sequence and showed evidence of splicing (Supplemental Fig. S11B). Together, these examples suggest that the potential of individual L1 copies to become expressed and retrotranspose is influenced by both their DNA methylation status and their unique genomic environment.

We next generated TSS profiles for the 5' UTRs of L1  $T_{FI}$  elements containing 2, 3, 4, and 5 monomer units (Fig. 6D; Supplemental Fig. S12A,B). As observed for L1  $T_{FI}$  monomer sequences in general (Fig. 6B), the TSS distribution peaked in the vicinity of the YY1 binding site with-

in each monomer. Notably, we did not observe a TSS peak in the “short” monomer proximal to the nonmonomeric region of the 5' UTR, despite the presence of a YY1 binding site in this region (Fig. 6D; Supplemental Fig. S12), consistent with recent analysis of L1  $T_{FI}$  promoter activity (Kong et al. 2022). The TSS peaks along the L1  $T_{FI}$  5' UTR corresponded to the periodic spikes in DNA methylation, a trend most clearly visible at the day 3 time point (Figs. 4G, 6D; Supplemental Fig. S12). Together, our results detail the dynamics of TSS distribution during establishment of CpG methylation across the mouse L1  $T_{FI}$  promoter.

## Discussion

In this study, we evaluate the mutational capacity of new heritable L1 copies, in terms of their inherent retrotransposition potential and their epigenetic status during development and in adult tissues. We find that loss of 5' monomers by daughter elements relative to their donors consistently diminishes daughter element retrotransposition efficiency. While for individual examples we cannot determine whether this promoter shortening arose because of 5' truncation of the L1 cDNA during retrotransposition (Ostertag and Kazazian 2001; Symer et al. 2002; Zingler et al. 2005), or because of the use of a TSS internal to the donor L1 promoter, analysis of putative transcription initiator dinucleotides supports the latter general hypothesis (Fig. 1B). Indeed, L1  $T_{FI}$  TSS distribution analyzed via 5' RACE suggests that L1 transcription tends to initiate in the vicinity of the YY1 binding site within each monomer, similar to the clustering observed for the first



**Figure 6.** DNA methylation during differentiation impacts mouse L1 TSS distribution. (A) Gel image of 5' RACE products. Total RNA was extracted from two independent replicate cultures at d0 (pluripotent mESCs), d3 of differentiation, and d21 of differentiation and subjected to 5' RACE NTC; no template control. mw; molecular weight marker. (B) Position of TSSs within the L1 T<sub>F</sub> monomer sequence. The TSS count is shown on the y-axis. The position within the 212 bp L1 T<sub>F</sub> monomer consensus sequence is displayed on the x-axis, and a schematic of the L1 T<sub>F</sub> monomer consensus with the YY1 binding site highlighted in red is shown at the top. Blue, d0; orange, d3; green, d21. (C) Percentage of TSSs upstream of L1 T<sub>F</sub> (orange), within L1 T<sub>F</sub> monomers (blue) and downstream of monomeric region (green) for d0, d3, and d21. The number of loci represented with upstream TSSs at each time point is indicated. (D) Above, schematic of the first 2000 bp of an L1 T<sub>F</sub> element containing five monomer units. Alternating green shading represents monomer units. Light gray shading represents the nonmonomeric region of the L1 T<sub>F</sub> promoter. Dark gray represents ORF1. Orange lines show the position of CpG dinucleotides. The positions of the YY1 binding sites are labeled and represented as vertical light gray lines extending down the figure panel. The position of the L1-specific 5' RACE primer is indicated. Below, TSS counts and mean CpG methylation are shown for time points d0 (blue), d3 (orange), and d21 (green). The histograms in the upper plots show TSS count for 137 L1 T<sub>F</sub> elements with 5 monomer units, with each bar representing a 10 bp bin. The lower plots display composite DNA methylation profiles with mean (thick line) and standard deviation (shaded region) indicated.

nucleotide of L1 insertions analyzed here and elsewhere (Fig. 1C; DeBerardinis and Kazazian 1999).

Plotting DNA methylation levels determined by ONT sequencing across the L1 5' UTR revealed a "smile" pattern, wherein the inner monomers tend to be less methylated than the genome-proximal and L1-proximal monomers. This pattern was observed for individual L1 loci (Supplemental Fig. S7) and for the composite methylation profiles of L1s genome-wide, and was most prominent in d3 EBs actively undergoing de novo DNA methylation (Figs. 4G, 6D, 5; Supplemental Figs. S5, S6, S9, S11). Why methylation of the L1 T<sub>F</sub> promoter apparently proceeds in this "outside-

in" fashion is an intriguing topic for future investigations. Our nucleotide resolution analysis of CpG methylation across mouse L1 promoters showed a local peak in CpG methylation in the vicinity of each YY1 binding site (Fig. 5; Supplemental Fig. S9), consistent with a potential role for YY1 in directing the de novo DNA methylation machinery to mouse L1 promoters. Notably, L1 T<sub>F</sub> TSSs identified by analysis of 5' RACE products also cluster around the YY1 binding site (Fig. 6B,D; Supplemental Fig. S11), as do the putative TSSs of full-length L1 T<sub>F</sub> insertions (Fig. 1C; DeBerardinis and Kazazian 1999). Thus, as we speculated previously for human L1s (Sanchez-Luque et al. 2019), YY1 or its binding site(s) within the L1 promoter may be required for both L1 transcription initiation, and, paradoxically, the epigenetic repression of L1 in somatic cells.

Our analysis of L1 TSS distributions revealed that in d21 differentiated cells, the proportion of L1 T<sub>F</sub> TSSs mapping upstream of the L1 5' UTR increased compared to d0 and d3, even as the absolute TSS count was diminished, consistent with silencing by DNA methylation (Fig. 6B,C). We speculate that L1 loci residing downstream of promoters that retain activity in differentiated cells are more likely to produce somatic L1 insertions in cells in which the native L1 5' UTR promoter is heavily methylated. Together with L1 loci that evade 5' UTR DNA methylation during differentiation (Supplemental Fig. S5), these elements may comprise a cohort of "escapee" loci capable of expression and retrotransposition to generate somatic mosaicism in differentiated cells. Indeed, in our previous studies we encountered somatic L1 insertions that likely arose because of both scenarios (Salvador-Palomeque et al. 2019; Sanchez-Luque et al. 2019; Billon et al. 2022).

Our locus-specific bisulfite sequencing interrogation of donor and daughter L1 methylation in multiple generations of adult tissues, and in differentiating mESCs in vitro, revealed that developmentally active donor L1s and the resultant daughter insertions are generally remethylated in concert with L1 elements genome-wide. De novo daughter L1s were methylated even in the somatic and germ tissues of the mosaic animals in which they arose (Fig. 2), likely because of abundant expression of de novo DNA methyltransferases in pluripotent cells and early postimplantation embryos (Okano et al. 1998; Chen et al. 2003). This result broadly agrees with a previous study analyzing the methylation status of transgene-derived engineered L1 insertions in cultured cells and transgenic animals (Kannan et al. 2017). It should be noted, however, that Kannan et al.

queried the methylation status of heterologous promoters or GFP reporter cassette sequences internal to engineered insertions, rather than the mouse L1 5' UTR. Our ONT analysis across full-length endogenous L1s during differentiation (Fig. 4) showed that the L1 body remains methylated even in pluripotent cells, with the 5' UTR undergoing dynamic methylation changes likely to affect the expression of full-length L1 transcripts and therefore de novo retrotransposition events. In addition, the “smile” pattern in methylation across the mouse L1 5' UTR could also lead to an overestimation of mouse L1 promoter methylation, and may impede approaches that analyze only the 5'-most monomers from identifying VM-L1s. Indeed, polyL1Tf\_4 is less methylated (<80%) in liver compared to other tissues (>80%) within the same animal, and this methylation pattern is re-established in the F1 but not F2 animal (Fig. 2G). Therefore, it is possible that polyL1Tf\_4 is a tissue-specific VM-L1 (Tubio et al. 2014; Schauer et al. 2018; Sanchez-Luque et al. 2019).

In sum, we conclude that the majority of the insertions analyzed here likely arose as a consequence of opportunity provided by genome-wide epigenetic reprogramming during embryonic development (Hajkova et al. 2002; Seki et al. 2005; Tachibana et al. 2007; Abe et al. 2011; Saitou et al. 2012; Seisenberger et al. 2012; Cantone and Fisher 2013). Some exceptional L1s do however escape methylation in a proportion of fully differentiated cells (Supplemental Fig. S5) or, as for Donor 7, appear methylated in somatic tissues but unmethylated in a large fraction of adult germ cells (Fig. 2E,H; Supplemental Fig. S2A,C). Indeed, a recent study examining the somatic L1 epigenetic and retrotransposition landscape in human colorectal epithelial cell clones concluded that L1 methylation escape occurs postgastrulation during the early stages of organogenesis (Nam et al. 2023). Future studies are likely to reveal additional tissue and developmental stage-specific mouse L1 “escapee” loci, with the capacity to evade genome-wide methylation by virtue of their sequence content or surrounding genomic environment.

## Methods

### Mice

All mouse work was performed in compliance with the guidelines set forth by the University of Queensland Animal Ethics Committee. Tissues from animals generated for Richardson et al. (2017) were used for this study.

### Cell culture

HeLa-JVM cells were maintained at 37°C and 5% CO<sub>2</sub> in Dulbecco's Modified Eagle's Media (DMEM) (Thermo Fisher Scientific) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Thermo Fisher Scientific), 1% L-glutamine (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Thermo Fisher Scientific). The cells were passaged every 3–4 d after they achieved confluency of 80%–90% using Trypsin 0.25% EDTA (Thermo Fisher Scientific).

E14Tg2a mESCs (ATCC CRL-1821) were cultured on gelatinized tissue culture plates and maintained at 37°C and 5% CO<sub>2</sub>. Cells were maintained in serum + LIF, 2i + serum, and 2i + LIF conditions and were passaged at 70%–80% confluency every 2–3 d using Trypsin 0.25% EDTA (Thermo Fisher Scientific), with media changed every day. The details of mESC culture conditions and media composition are provided in the Supplemental Methods. E14Tg2a mESCs were differentiated into EBs using the hanging drop method as previously described (Behringer et al. 2016).

Details of the mESC differentiation protocol can be found in the Supplemental Materials.

### Identification of L1 donor/daughter pairs

We previously identified 11 de novo endogenous L1 T<sub>F</sub> insertions among 85 C57BL6/J mice belonging to multigeneration pedigrees, as well as six unfixed polymorphic L1 T<sub>F</sub> insertions absent from the C57BL6/J reference genome and differentially present/absent among these pedigrees (Richardson et al. 2017). We traced the majority of heritable insertions to pluripotent embryonic cells, evidenced by shared somatic/germline mosaicism of the founder mouse, and early primordial germ cells (PGCs), evidenced by germline-restricted mosaicism across both testes of male founder mice. Analysis of unique L1 3' transduced sequences (Holmes et al. 1994; Moran et al. 1999; Goodier et al. 2000; Pickeral et al. 2000; Xing et al. 2006) allowed us to identify the source (donor) L1 loci responsible for three offspring (daughter) de novo insertions (one early embryonic, two early PGC) and two unfixed polymorphic daughter insertions.

### DNA extraction

Genomic DNA from mouse tissue was extracted as previously described (Richardson et al. 2017).

Genomic DNA from cultured cells was extracted using DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's protocol. The DNA concentration was determined by Qubit 3.0 Fluorometer (Invitrogen) using the Qubit dsDNA HS Assay Kit (Invitrogen) following the manufacturer's instructions.

High molecular weight (HMW) genomic DNA from cultured cells was extracted using the Nanobind CBB Big DNA Kit (Circulomics) following the manufacturer's instructions.

### Generation of mouse L1 reporter constructs

Mouse L1 reporter constructs were generated via PCR amplification of donor/daughter mouse L1s from genomic DNA, followed by capillary sequencing of multiple clones to identify PCR errors. Constructs containing the full, correct L1 sequence were built from individual PCR clones as described previously (Sanchez-Luque et al. 2019). A modified version of the previously described pTN201 construct (Naas et al. 1998) in which the L1 3' UTR poly-purine tract is located downstream, rather than upstream, of the NEO indicator cassette was used as a backbone to generate L1 reporter constructs (Richardson et al. 2022). The molecular details of the cloning strategy are presented in the Supplemental Methods.

### Immunostaining

Immunostaining was performed on day 7 (24 h after plating of EBs) and day 21 of mESC differentiation. EBs were plated on gelatinized coverslips in twelve-well plates. Immunostaining was performed using primary antibodies against  $\beta$ -tubulin (Rabbit IgG; Sigma-Aldrich, #T2200), A-fetoprotein (AFP), (Goat IgG; R&D Systems AF5369), and smooth muscle actin (Mouse IgG; Thermo Fisher Scientific 14976080). Secondary antibodies used were Alexa Fluor 647 Donkey Anti-Rabbit IgG (Jackson ImmunoResearch 711-606-152); Cy3 Donkey Anti-Goat IgG (Jackson ImmunoResearch 715-165-150), and Alexa Fluor 488 Donkey Anti-Mouse IgG (Jackson ImmunoResearch 715-546-150). Details of the immunostaining and imaging procedures are supplied in the Supplemental Methods.

### Retrotransposition assay

Retrotransposition assays in HeLa-JVM cells were performed as previously described (Kopera et al. 2016) with some minor modifications. HeLa-JVM cells were plated at the appropriate density for each construct and transfected the following day with 1 µg of plasmid DNA per well using FuGene-HD (Promega). Transfection efficiency was determined in parallel using L1 reporter construct and pCEP4-eGFP co-transfection. Geneticin/G418 (400 µg/mL) (Thermo Fisher Scientific) selection was started 3 d post-transfection and performed for 12 d, after which the cells were fixed and stained with crystal violet. Colonies were counted and retrotransposition efficiency determined by normalizing colony counts to transfection efficiency for each construct. The full details of the HeLa-JVM retrotransposition assay protocol are provided in the Supplemental Methods.

L1 retrotransposition assays in E14Tg2a mESCs (ATCC CRL-1821) were performed as previously described (MacLennan et al. 2017) in antibiotic-free 2i + serum conditions using 1 µg of plasmid DNA per well of a six-well dish and Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 24 h after transfection, the mESCs were passaged into 10 cm dishes. G418 (Invitrogen) selection (200 µg/mL) was started 24 h after passaging and continued for 12 d. Drug-resistant colonies were fixed, stained and counted as described for HeLa-JVM cells. The full details of the mESC retrotransposition assay protocol are provided in the Supplemental Methods.

### Locus-specific bisulfite sequencing

Bisulfite conversion was performed using the EZ DNA Methylation-Lightning Kit (Zymo Research), following the manufacturer's instructions. Primers used for target amplification are listed in Supplemental Table S1. PCRs were performed using MyTaq HS DNA Polymerase (Bioline). PCR products were gel purified using the MinElute Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. Illumina libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) following the manufacturer's instructions. Barcoded libraries were pooled equimolar and sequenced on an Illumina MiSeq platform using a MiSeq Reagent Kit v3 (600-cycle). Analysis was performed as previously described (Schauer et al. 2018). Unconverted amplicon sequences are listed in Supplemental Table S1. Per sample, up to 1000 reads per violin plot were randomly extracted from the bisulfite sequencing files using the ExSeq tool ([https://github.com/MischaLundberg/extract\\_sequences](https://github.com/MischaLundberg/extract_sequences)). Selected reads were analyzed using QUMA (QUantification tool for Methylation Analysis) (Kumaki et al. 2008). Full details of the locus-specific bisulfite sequencing analysis can be found in the Supplemental Methods.

### Nanopore sequencing

Purity of HMW DNA was determined using a NanoDrop One Spectrophotometer (Thermo Fisher Scientific) according to the manufacturer's instructions. The DNA concentration was determined by Qubit 3.0 Fluorometer (Invitrogen) using the Qubit dsDNA HS Assay Kit (Invitrogen) and on a TapeStation System (Agilent Technologies). ONT sequencing libraries were created using 1D Ligation (SQK-LSK109), sheared to create an average fragment size of ~10 kb and sequenced at the Kinghorn Centre for Clinical Genomics at the Garvan Institute of Medical Research (Darlinghurst, NSW, Australia) on an ONT PromethION platform.

Bases were called using Guppy version 3.2.10 (Oxford Nanopore Technologies) and aligned to the reference genome build mm10 using minimap2 version 2.17 (Li 2018) and SAMtools version 1.3 (Li et al. 2009). Reads were indexed and

per-CpG methylation calls generated using nanopolish version 0.13.2 (Simpson et al. 2017). Methylation likelihood data were sorted by position and indexed using tabix version 1.10.2 (Li 2011).

Reference L1 locations were derived from the RepeatMasker (<https://www.repeatmasker.org>) .OUT track files available for mm10 from the UCSC Genome Browser (Kent 2002). As full-length mouse L1s are often broken into multiple adjacent annotations when present on the - strand of the genome, we merged adjacent similarly oriented L1s before analysis. L1s were considered if annotated as >6000 bp in length. Monomers were counted using a Python script which considers the best alignment between a library of known monomer sequences and a target transposable element using exonerate (Slater and Birney 2005), records and masks the monomer alignment, and repeats the process until no further monomer alignments are present. Methylation results were annotated per-subfamily and per-monomer count and plotted using these categories. Methylation statistics for reference L1s were generated using the "segmeth" function in methylartist (<https://github.com/adamewing/methylartist>) and plotted using the "segplot" function. Reads mapping completely within L1s were excluded from the reference L1 methylation analysis via the "--exclude\_ambiguous" option in methylartist segmeth to negate the contribution of ambiguous mappings. Plots categorized by monomer count (Fig. 4D) were limited to reads spanning the entire segment with the addition of the "--spanning\_only" argument to segmeth. Methylation plots for individual L1s and differentially methylated regions (DMRs) (Fig. 5) were created using the methylartist "locus" function. Composite per-element methylation plots (Fig. 4F,G; Supplemental Figs. S5, S7) were created using the methylartist "composite" function, with individual CpG statistics obtained via the -output\_table and -meanplot\_cutoff 20 parameters.

### 5' RACE

Total RNA was extracted from E14Tg2a mESCs differentiated to embryoid bodies as described above, at day 0, day 3, and day 21 of the differentiation protocol using the RNeasy Mini kit (Qiagen, #74104). 5' RACE was performed using the 5' RACE module of the SMARTer 5'/3' RACE Kit (Takara Bio, #634858) according to the manufacturer's protocol. 1 µg total RNA was used as input for each reaction, and two reactions from biologically independent samples were performed per time point. PCR amplification of 5' RACE cDNAs was performed using the L1 gene specific primer 5'-catctctgtattctgtgtgctgatgctcaa-3' and the 5' RACE universal primer provided in the SMARTer 5'/3' RACE Kit (Takara Bio). 25 PCR cycles were performed as follows: 94°C for 30 sec, 67 °C for 30 sec, 72 °C for 3 min. PCR products were visualized on a 1% agarose gel. Products ranging from 700–5000 bp were excised, and PCR fragments were purified using the MinElute Gel Extraction Kit (Qiagen, # 28604). Iso-Seq template preparation using the Iso-Seq Express kit followed by PacBio SMRT Cell sequencing on a PacBio Sequel II platform was performed by the Australian Genome Research Facility. Six samples (two replicates per time points d0, d3, and d21) were multiplexed on a single SMRT cell, generating 3,699,664 CCS reads in total.

Reads were aligned to the mouse reference genome (mm10) using minimap2 (Li 2018) and parameters (-t 96 -N 1000 -p 0.95 -ax splice:hq -uf) and sorted with SAMtools (Li et al. 2009). Uniquely mapped reads, that is, those that aligned to one genomic position only at their best alignment score and as a primary alignment, were retained if an L1-specific primer and a 5' RACE universal primer were located at each of their termini. Multimapping reads were discarded to enable unambiguous read assignment to a given monomer-count resolved L1. The majority (>75%) of reads

aligned to an L1 also aligned uniquely to the genome. Reads were then assigned to the full-length L1 T<sub>FI</sub> elements they overlapped, with alignments required to terminate within the L1 body. The start positions of these alignments within the 5' UTR or upstream of the L1 were then recorded as putative TSSs supported by at least one read. These positions were used to generate TSS distributions relative to monomer coordinates or to L1 T<sub>FI</sub> 5' UTRs composed of 2, 3, 4 or 5 monomers. Sequence logos surrounding these TSSs were generated using WebLogo (Crooks et al. 2004).

## Data access

All Sanger sequencing data generated in this study have been submitted to the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession numbers OQ856744–OQ856753.

The Oxford Nanopore Technologies and PacBio data generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA763783.

New scripts TSSprofile.py and intersect.py used to analyse TSS distribution across L1 T<sub>F</sub> elements with 2, 3, 4, and 5 monomers have been uploaded as Supplemental Code.

## Competing interest statement

The authors declare no competing interests.

## Acknowledgments

The authors thank Dr. John V. Moran, Dr. Ian R. Adams and members of the Richardson and Faulkner labs for helpful advice and discussion. HeLa-JVM cells were a gift from Dr. John V. Moran. The L1SM and L1SMmut2 constructs were a gift from Dr. Jef D. Boeke, and the L1<sub>spa</sub> construct was a gift from Dr. Haig Kazazian. This research was supported by an Australian Government Research Training Program Scholarship and a Mater Research Frank Clair Scholarship awarded to P.G., a University of Queensland Research Training Program Scholarship and Commonwealth Scientific and Industrial Research Organisation Postgraduate Top-Up Scholarship awarded to M.L., the People Programme (Marie Skłodowska-Curie Actions) of the European Union Seventh Framework Program (FP7/2007-2013) under REA grant agreement P10F-GA-2013-623324 awarded to F.J.S.-L. and the Australian National Health and Medical Research Council (NHMRC) and Australian Research Council (ARC) Dementia Research Development Fellowship GNT1108258 awarded to G.O.B. This study was funded by the Australian NHMRC (GNT1125645, GNT1138795, and GNT1173711 to G.J.F.; GNT1173476 to S.R.R.), The ARC (DP200102919 to G.J.F. and S.R.R.), an Australian Department of Health Medical Frontiers Future Fund (MRFF) (MRF1175457) grant awarded to A.D.E., a CSL Centenary Fellowship to G.J.F., an Advance Queensland Women's Academic Fund Maternity Funding award to S.R.R., the Mater Research Strategic Grant for Outstanding Women to S.R.R., and the Mater Foundation (Equity Trustees / AE Hingeley, QFC Thomas George and KC BM Thomson Trusts). We acknowledge the TRI flow cytometry core for technical assistance and equipment. We acknowledge QBI Advanced Microscopy Facility for technical assistance and equipment, supported by ARC LIEF grant LE130100078.

**Author contributions:** P.G., D.C., F.J.S.-L., G.O.B., G.J.F., and S.R.R. designed and performed experiments. P.G., M.L., A.D.E., and G.J.F. performed bioinformatics analyses. P.G., G.J.F., and S.R.R. conceived the study. P.G., G.J.F., and S.R.R. wrote the manuscript. All authors commented on and approved the final manuscript.

## References

- Abe M, Tsai SY, Jin S-G, Pfeifer GP, Szabó PE. 2011. Sex-specific dynamics of global chromatin changes in fetal mouse germ cells. *PLoS One* **6**: e23848. doi:10.1371/journal.pone.0023848
- Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, Hutchison CA. 1994. Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Mol Biol Evol* **11**: 778–789. doi:10.1093/oxfordjournals.molbev.a040158
- Akagi K, Li J, Stephens RM, Volfovsky N, Symer DE. 2008. Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Res* **18**: 869–880. doi:10.1101/gr.075770.107
- Athanikar JN, Badge RM, Moran JV. 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* **32**: 3846–3855. doi:10.1093/nar/gkh698
- Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV. 2010. LINE-1 Retrotransposition activity in human genomes. *Cell* **141**: 1159–1170. doi:10.1016/j.cell.2010.05.021
- Behringer R, Gertsenstein M, Nagy KV, Nagy A. 2016. Differentiating mouse embryonic stem cells into embryoid bodies by hanging-drop cultures. *Cold Spring Harb Protoc* **2016**: 1073–1076. doi:10.1101/pdb.prot092429
- Bertozzi TM, Ferguson-Smith AC. 2020. Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Semin Cell Dev Biol* **97**: 93–105. doi:10.1016/j.semcdb.2019.08.002
- Besse S, Allamand V, Vilquin J-T, Li Z, Poirier C, Vignier N, Hori H, Guénet J-L, Guicheney P. 2003. Spontaneous muscular dystrophy caused by a retrotransposon insertion in the mouse laminin  $\alpha 2$  chain gene. *Neuromuscul Disord* **13**: 216–222. doi:10.1016/s0960-8966(02)00278-x
- Billon V, Sanchez-Luque FJ, Rasmussen J, Bodea GO, Gerhardt DJ, Gerdes P, Cheetham SW, Schauer SN, Ajjikuttira P, Meyer TJ, et al. 2022. Somatic retrotransposition in the developing rhesus macaque brain. *Genome Res* **32**: 1298–1314. doi:10.1101/gr.276451.121
- Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**: 96–99. doi:10.1038/nature02886
- Cantone I, Fisher AG. 2013. Epigenetic programming and reprogramming during development. *Nat Struct Mol Biol* **20**: 282–289. doi:10.1038/nsm.2489
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Sempé CAM, Taylor MS, Engström PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626–635. doi:10.1038/ng1789
- Casavant NC, Hardies SC. 1994. The dynamics of murine LINE-1 subfamily amplification. *J Mol Biol* **241**: 390–397. doi:10.1006/jmbi.1994.1515
- Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D. 2014. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev* **28**: 1397–1409. doi:10.1101/gad.241661.114
- Cheetham SW, Kindlova M, Ewing AD. 2022. Methylartist: tools for visualizing modified bases from nanopore sequence data. *Bioinformatics* **38**: 3109–3112. doi:10.1093/bioinformatics/btac292
- Chen T, Ueda Y, Dodge JE, Wang Z, Li E. 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23**: 5594–605. doi:10.1128/MCB.23.16.5594-5605.2003
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14**: 1188–1190. doi:10.1101/gr.849004
- DeBerardinis RJ, Kazazian HH. 1999. Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics* **56**: 317–323. doi:10.1006/geno.1998.5729
- DeBerardinis RJ, Goodier JL, Ostertag EM, Kazazian HH. 1998. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet* **20**: 288–290. doi:10.1038/3104
- Deininger P, Morales ME, White TB, Baddoo M, Hedges DJ, Servant G, Srivastav S, Smither ME, Concha M, DeHaro DL, et al. 2017. A comprehensive approach to expression of L1 loci. *Nucleic Acids Res* **45**: e31. doi:10.1093/nar/gkw1067
- de la Rica L, Deniz Ö, Cheng KCL, Todd CD, Cruz C, Houseley J, Branco MR. 2016. TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol* **17**: 234. doi:10.1186/s13059-016-1096-8
- Deniz Ö, Frost JM, Branco MR. 2019. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* **20**: 417–431. doi:10.1038/s41576-019-0106-6
- Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. 1991. Isolation of an active human transposable element. *Science* **254**: 1805–1808. doi:10.1126/science.1662412
- Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, et al. 2010.

- Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet* **6**: e1001150. doi:10.1371/journal.pgen.1001150
- Doucet AJ, Wilusz JE, Miyoshi T, Liu Y, Moran JV. 2015. A 3' Poly(A) tract is required for LINE-1 retrotransposition. *Mol Cell* **60**: 728–741. doi:10.1016/j.molcel.2015.10.012
- Elmer JL, Hay AD, Kessler NJ, Bertozzi TM, Ainscough EAC, Ferguson-Smith AC. 2021. Genomic properties of variably methylated retrotransposons in mouse. *Mob DNA* **12**: 6. doi:10.1186/s13100-021-00235-1
- Ergün S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Strätling WH, Schumann GG. 2004. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J Biol Chem* **279**: 27753–27763. doi:10.1074/jbc.M312985200
- Ewing AD, Smits N, Sanchez-Luque FJ, Faivre J, Brennan PM, Richardson SR, Cheetham SW, Faulkner GJ. 2020. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol Cell* **80**: 915–928.e5. doi:10.1016/j.molcel.2020.10.024
- Feng Q, Moran JV, Kazazian HH, Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**: 905–916. doi:10.1016/S0092-8674(00)81997-2
- Ferraj A, Audano PA, Balachandran P, Czechanski A, Flores JI, Radecki AA, Mosur V, Gordon DS, Walawalkar IA, Eichler EE, et al. 2023. Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements. *Cell Genomics* **3**: 100291. doi:10.1016/j.xgen.2023.100291
- Flasch DA, Macia Á, Sánchez L, Ljungman M, Heras SR, García-Pérez JL, Wilson TE, Moran JV. 2019. Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell* **177**: 837–851.e28. doi:10.1016/j.cell.2019.02.050
- Furano AV, Robb SM, Robb FT. 1988. The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. *Nucleic Acids Res* **16**: 9215–9231. doi:10.1093/nar/16.19.9215
- Gagnier L, Belancio VP, Mager DL. 2019. Mouse germ line mutations due to retrotransposon insertions. *Mob DNA* **10**: 15. doi:10.1186/s13100-019-0157-4
- Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Stephen Pittard W, Mills RE, Devine SE. 2017. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res* **27**: 1916–1929. doi:10.1101/gr.218032.116
- Gerdes P, Lim SM, Ewing AD, Larcombe MR, Chan D, Sanchez-Luque FJ, Walker L, Carleton AL, James C, Knaupp AS, et al. 2022. Retrotransposon instability dominates the acquired mutation landscape of mouse induced pluripotent stem cells. *Nat Commun* **13**: 7470. doi:10.1038/s41467-022-35180-x
- Goodier JL. 2016. Restricting retrotransposons: a review. *Mob DNA* **7**: 16. doi:10.1186/s13100-016-0070-z
- Goodier JL, Ostertag EM, Kazazian HH. 2000. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum Mol Genet* **9**: 653–657. doi:10.1093/hmg/9.4.653
- Goodier JL, Ostertag EM, Du K, Kazazian HH. 2001. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* **11**: 1677–1685. doi:10.1101/gr.198301
- Greenberg MVC, Bourc'his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**: 590–607. doi:10.1038/s41580-019-0159-6
- Grimaldi G, Skowronski J, Singer MF. 1984. Defining the beginning and end of KpnI family segments. *EMBO J* **3**: 1753–1759. doi:10.1002/j.1460-2075.1984.tb02042.x
- Hajkova P, Erhardt S, Lane N, Haaf T, El-Maari O, Reik W, Walter J, Surani MA. 2002. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**: 15–23. doi:10.1016/S0925-4773(02)00181-8
- Han JS, Boeke JD. 2004. A highly active synthetic mammalian retrotransposon. *Nature* **429**: 314–318. doi:10.1038/nature02535
- Hardies SC, Wang L, Zhou L, Zhao Y, Casavant NC, Huang S. 2000. Line-1 (L1) lineages in the mouse. *Mol Biol Evol* **17**: 616–628. doi:10.1093/oxfordjournals.molbev.a026340
- Hata K, Sakaki Y. 1997. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* **189**: 227–234. doi:10.1016/S0378-1119(96)00856-6
- Hohjoh H, Singer MF. 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J* **15**: 630–639. doi:10.1002/j.1460-2075.1996.tb00395.x
- Holmes SE, Singer MF, Swergold GD. 1992. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *J Biol Chem* **267**: 19765–19768. doi:10.1016/S0021-9258(19)88618-0
- Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH. 1994. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat Genet* **7**: 143–148. doi:10.1038/ng0694-143
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921. doi:10.1038/35057062
- Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**: 242–245. doi:10.1038/nature13760
- Kannan M, Li J, Fritz SE, Husarek KE, Sanford JC, Sullivan TL, Tiwary PK, An W, Boeke JD, Symer DE. 2017. Dynamic silencing of somatic L1 retrotransposon insertions reflects the developmental and cellular contexts of their genomic integration. *Mob DNA* **8**: 8. doi:10.1186/s13100-017-0091-2
- Kazachenka A, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunning R, Pahita E, Adams S, Adams D, Ferguson-Smith AC. 2018. Identification, characterization, and heritability of murine metastable epialleles: implications for nongenetic inheritance. *Cell* **175**: 1259–1271.e13. doi:10.1016/j.cell.2018.09.043
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* **12**: 656–664. doi:10.1101/gr.229202
- Khazina E, Weichenrieder O. 2009. Non-LTR retrotransposons encode non-canonical RRM domains in their first open reading frame. *Proc Natl Acad Sci* **106**: 731–736. doi:10.1073/pnas.0809964106
- Khazina E, Weichenrieder O. 2018. Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. *eLife* **7**: e34960. doi:10.7554/eLife.34960
- Kim JD, Kim J. 2009. YY1's longer DNA-binding motifs. *Genomics* **93**: 152–158. doi:10.1016/j.ygeno.2008.09.013
- Kingsmore SE, Giros B, Suh D, Bieniarz M, Caron MG, Seldin MF. 1994. Glycine receptor  $\beta$ -subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nat Genet* **7**: 136–142. doi:10.1038/ng0694-136
- Kong L, Saha K, Hu Y, Tschetter JN, Habben CE, Whitmore LS, Yao C, Ge X, Ye P, Newkirk SJ, et al. 2022. Subfamily-specific differential contribution of individual monomers and the tether sequence to mouse L1 promoter activity. *Mob DNA* **13**: 13. doi:10.1186/s13100-022-00269-z
- Kopera HC, Larson PA, Moldovan JB, Richardson SR, Liu Y, Moran JV. 2016. Line-1 cultured cell retrotransposition assay. In *Methods in molecular biology* (ed. Garcia-Perez JL), Vol. 1400, pp. 139–156. Humana Press Inc., New York.
- Kubo S, Seleme MDC, Soifer HS, Perez JLG, Moran JV, Kazazian HH, Kasahara N. 2006. L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci* **103**: 8036–8041. doi:10.1073/pnas.0601954103
- Kumaki Y, Oda M, Okano M. 2008. QUMA: quantification tool for methylation analysis. *Nucleic Acids Res* **36**: W170–W175. doi:10.1093/nar/gkn294
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908–917. doi:10.1101/gad.1640708
- Lanciano S, Cristofari G. 2020. Measuring and interpreting transposable element expression. *Nat Rev Genet* **21**: 721–736. doi:10.1038/s41576-020-0251-y
- Lee S-H, Cho S-Y, Shannon MF, Fan J, Rangasamy D. 2010. The impact of CpG island on defining transcriptional activation of the mouse L1 retrotransposable elements. *PLoS One* **5**: e11353. doi:10.1371/journal.pone.0011353
- Li H. 2011. Tabix: Fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**: 718–719. doi:10.1093/bioinformatics/btq671
- Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi:10.1093/bioinformatics/bty191
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079. doi:10.1093/bioinformatics/btp352
- Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassik MC, Wysocka J. 2018. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**: 228–232. doi:10.1038/nature25179
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for nonLTR retrotransposition. *Cell* **72**: 595–605. doi:10.1016/0092-8674(93)90078-5
- Macia A, Widmann TJ, Heras SR, Ayllon V, Sanchez L, Benkaddour-Boumzouad 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. doi:10.1101/gr.206805.116
- MacLennan M, García-Cañadas M, Reichmann J, Khazina E, Wagner G, Playfoot CJ, Salvador-Palomeque C, Mann AR, Peressini P, Sanchez L,

- et al. 2017. Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *eLife* **6**: e26152. doi:10.7554/eLife.26152
- Martin SL, Bushman FD. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol* **21**: 467–475. doi:10.1128/MCB.21.2.467-475.2001
- Mathias S, Scott A, Kazazian H, Boeke J, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. *Science* **254**: 1808–1810. doi:10.1126/science.1722352
- Mears ML, Hutchison CA. 2001. The evolution of modern lineages of mouse L1 elements. *J Mol Evol* **52**: 51–62. doi:10.1007/s002390010133
- Mita P, Wudzinska A, Sun X, Andrade J, Nayak S, Kahler DJ, Badri S, LaCava J, Ueberheide B, Yun CY, et al. 2018. LINE-1 protein localization and functional dynamics during the cell cycle. *eLife* **7**: e30058. doi:10.7554/eLife.30058
- Mita P, Sun X, Fenyö D, Kahler DJ, Li D, Agmon N, Wudzinska A, Keegan S, Bader JS, Yun C, et al. 2020. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat Struct Mol Biol* **27**: 179–191. doi:10.1038/s41594-020-0374-z
- Molaro A, Falcatori I, Hodges E, Aravin AA, Marran K, Raffi S, McCombie WR, Smith AD, Hannon GJ. 2014. Two waves of de novo methylation during mouse germ cell development. *Genes Dev* **28**: 1544–1549. doi:10.1101/gad.244350.114
- Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**: 917–927. doi:10.1016/S0092-8674(00)81998-4
- Moran JV, DeBerardinis RJ, Kazazian HH. 1999. Exon shuffling by L1 retrotransposition. *Science* **283**: 1530–1534. doi:10.1126/science.283.5407.1530
- Naas TP, DeBerardinis RJ, Moran JV, Ostertag EM, Kingsmore SF, Seldin MF, Hayashizaki Y, Martin SL, Kazazian HH. 1998. An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J* **17**: 590–597. doi:10.1093/emboj/17.2.590
- Nam CH, Youk J, Kim JY, Lim J, Park JW, Oh SA, Lee HJ, Park JW, Won H, Lee Y, et al. 2023. Widespread somatic L1 retrotransposition in normal colorectal epithelium. *Nature* **617**: 540–547. doi:10.1038/s41586-023-06046-z
- Nellåker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, Ponting CP. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol* **13**: R45. doi:10.1186/gb-2012-13-6-r45
- Nguyen THM, Carreira PE, Sanchez-Luque FJ, Schauer SN, Fagg AC, Richardson SR, Davies CM, Jesuadian JS, Kempen MJHC, Troskie RL, et al. 2018. L1 retrotransposon heterogeneity in ovarian tumor cell evolution. *Cell Rep* **23**: 3730–3740. doi:10.1016/j.celrep.2018.05.090
- Okano M, Xie S, Li E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**: 219–220. doi:10.1038/890
- Ostertag EM, Kazazian HH. 2001. Twin priming: A proposed mechanism for the creation of insertions in L1 retrotransposition. *Genome Res* **11**: 2059–2065. doi:10.1101/gr.205701
- Paterson AL, Weaver JMJ, Eldridge MD, Tavaré S, Fitzgerald RC, Edwards PAW. 2015. Mobile element insertions are frequent in oesophageal adenocarcinomas and can mislead paired-end sequencing analysis. *BMC Genomics* **16**: 473. doi:10.1186/s12864-015-1685-z
- Penzkofer T, Jäger M, Figlerowicz M, Badge R, Mundlos S, Robinson PN, Zemojtel T. 2017. L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res* **45**: D68–D73. doi:10.1093/nar/gkw925
- Philippe C, Vargas-Landin DB, Doucet AJ, Van Essen D, Vera-Otarola J, Kuciak M, Corbin A, Nigumann P, Cristofari G. 2016. Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *eLife* **5**: e13926. doi:10.7554/eLife.13926
- Pickeral OK, Makalowski W, Boguski MS, Boeke JD. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res* **10**: 411–415. doi:10.1101/gr.10.4.411
- Pitkänen E, Cajuso T, Katainen R, Kaasinen E, Välimäki N, Palin K, Taipale J, Aaltonen LA, Kilpivaara O. 2014. Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. *Oncotarget* **5**: 853–859. doi:10.18632/oncotarget.1781
- Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**: 1101–1105. doi:10.1038/nature08829
- Richardson SR, Gerdes P, Gerhardt DJ, Sanchez-Luque FJ, Bodea GO, Muñoz-Lopez M, Jesuadian JS, Kempen MJHC, Carreira PE, Jeddeloh JA, et al. 2017. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res* **27**: 1395–1405. doi:10.1101/gr.219022.116
- Richardson SR, Chan D, Gerdes P, Han JS, Boeke JD, Faulkner GJ. 2022. Revisiting the impact of synthetic ORF sequences on engineered LINE-1 retrotransposition. bioRxiv doi:10.1101/2022.08.29.505632
- Saitou M, Kagiwada S, Kurimoto K. 2012. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**: 15–31. doi:10.1242/dev.050849
- Salvador-Palomeque C, Sanchez-Luque FJ, Fortuna PRJ, Ewing AD, Wolvetang EJ, Richardson SR, Faulkner GJ. 2019. Dynamic methylation of an L1 transduction family during reprogramming and neurodifferentiation. *Mol Cell Biol* **39**: e00499-18. doi:10.1128/MCB.00499-18
- Sanchez-Luque FJ, Kempen MJHC, Gerdes P, Vargas-Landin DB, Richardson SR, Troskie RL, Jesuadian JS, Cheetham SW, Carreira PE, Salvador-Palomeque C, et al. 2019. LINE-1 evasion of epigenetic repression in humans. *Mol Cell* **75**: 590–604.e12. doi:10.1016/j.molcel.2019.05.024
- Saxton JA, Martin SL. 1998. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *J Mol Biol* **280**: 611–622. doi:10.1006/jmbi.1998.1899
- Schauer SN, Carreira PE, Shukla R, Gerhardt DJ, Gerdes P, Sanchez-Luque FJ, Nicoli P, Kindlova M, Ghisletti S, Dos Santos AD, et al. 2018. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res* **28**: 639–653. doi:10.1101/gr.226993.117
- Schichman SA, Adey NB, Edgell MH, Hutchison CA. 1993. L1 A-monomer tandem arrays have expanded during the course of mouse L1 evolution. *Mol Biol Evol* **10**: 552–570. doi:10.1093/oxfordjournals.molbev.a040025
- Schöpp T, Zoch A, Berrens RV, Auchynnikava T, Kabayama Y, Vasiliauskaitė L, Rappsilber J, Allshire RC, O'Carroll D. 2020. TEX15 is an essential executor of MIWI2-directed transposon DNA methylation and silencing. *Nat Commun* **11**: 3739. doi:10.1038/s41467-020-17372-5
- Scott AF, Schmeckpeper B, Abdelrazik M, Comey C, O'Hara B, Rossiter J, Cooley T, Heath P, Smith K, Margolet L. 1987. Origin of the human L1 elements: Proposed progenitor genes deduced from a consensus DNA sequence. *Genomics* **1**: 113–125. doi:10.1016/0888-7543(87)90003-6
- Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE. 2016. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res* **26**: 745–755. doi:10.1101/gr.201814.115
- Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W, et al. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* **48**: 849–862. doi:10.1016/j.molcel.2012.11.001
- Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y. 2005. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**: 440–458. doi:10.1016/j.ydbio.2004.11.025
- Senft AD, Macfarlan TS. 2021. Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* **22**: 691–711. doi:10.1038/s41576-021-00385-1
- Severynse DM, Hutchison CA, Edgell MH. 1991. Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon. *Mamm Genome* **2**: 41–50. doi:10.1007/BF00570439
- Shehee WR, Chao SF, Loeb DD, Comer MB, Hutchison CA, Edgell MH. 1987. Determination of a functional ancestral sequence and definition of the 5' end of A-type mouse L1 elements. *J Mol Biol* **196**: 757–767. doi:10.1016/0022-2836(87)90402-5
- Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A. 2008. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6**: e253. doi:10.1371/journal.pbio.0060253
- Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**: 407–410. doi:10.1038/nmeth.4184
- Skowronski J, Fanning TG, Singer MF. 1988. Unit-length LINE-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol* **8**: 1385–1397. doi:10.1128/mcb.8.4.1385-1397.1988
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31. doi:10.1186/1471-2105-6-31
- Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**: 688–690. doi:10.1038/336688a0
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**: 339–344. doi:10.1038/nature10960
- Sookdeo A, Hepp CM, McClure MA, Boissinot S. 2013. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* **4**: 3. doi:10.1186/1759-8753-4-3



- Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. 2002. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* **110**: 327–338. doi:10.1016/S0092-8674(02)00839-5
- Tachibana M, Nozaki M, Takeda N, Shinkai Y. 2007. Functional dynamics of H3K9 methylation during meiotic prophase progression. *EMBO J* **26**: 3346–3359. doi:10.1038/sj.emboj.7601767
- Tan X, Xu X, Elkenani M, Smorag L, Zechner U, Nolte J, Engel W, Pantakani DVK. 2013. Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem Cell Res* **11**: 1045–1059. doi:10.1016/j.scr.2013.07.006
- Taylor MS, LaCava J, Mita P, Molloy KR, Huang CRL, Li D, Adney EM, Jiang H, Burns KH, Chait BT, et al. 2013. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155**: 1034–1048. doi:10.1016/j.cell.2013.10.021
- Tristán-Ramos P, Rubio-Roldan A, Peris G, Sánchez L, Amador-Cubero S, Viollet S, Cristofari G, Heras SR. 2020. The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. *Nat Commun* **11**: 5712. doi:10.1038/s41467-020-19430-4
- Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, et al. 2014. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**: 1251343. doi:10.1126/science.1251343
- Voliva CF, Jahn CL, Comer MB, Hutchison CA, Edgell MH. 1983. The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. *Nucleic Acids Res* **11**: 8847–8859. doi:10.1093/nar/11.24.8847
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539–543. doi:10.1038/nature06908
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562. doi:10.1038/nature01262
- Wei W, Morrish TA, Alisch RS, Moran JV. 2000. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal Biochem* **284**: 435–438. doi:10.1006/abio.2000.4675
- Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**: 684–687. doi:10.1038/336684a0
- Wincker P, Jubier-Maurin V, Roizés G. 1987. Unrelated sequences at the 5' end of mouse LINE-1 repeated elements define two distinct subfamilies. *Nucleic Acids Res* **15**: 8593–8606. doi:10.1093/nar/15.21.8593
- Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci* **103**: 17608–17613. doi:10.1073/pnas.0603224103
- Zhou M, Smith AD. 2019. Subtype classification and functional annotation of L1Md retrotransposon promoters. *Mob DNA* **10**: 14. doi:10.1186/s13100-019-0156-5
- Zingler N, Willhoeft U, Brose H-P, Schoder V, Jahns T, Hanschmann K-MO, Morrish TA, Löwer J, Schumann GG. 2005. Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res* **15**: 780–789. doi:10.1101/gr.3421505
- Zoch A, Auchynnika T, Berrens RV, Kabayama Y, Schöpp T, Heep M, Vasiliauskaite L, Pérez-Rico YA, Cook AG, Shkumatava A, et al. 2020. SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature* **584**: 635–639. doi:10.1038/s41586-020-2557-5

Received April 25, 2023; accepted in revised form August 21, 2023.