



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

L1 Mosaicism in Mammals

Citation for published version:

Faulkner, GJ & Garcia-Perez, JL 2017, 'L1 Mosaicism in Mammals: Extent, Effects, and Evolution', *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2017.07.004>

Digital Object Identifier (DOI):

[10.1016/j.tig.2017.07.004](https://doi.org/10.1016/j.tig.2017.07.004)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Trends in Genetics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



L1 mosaicism in mammals: extent, effects and evolution

Geoffrey J. Faulkner^{1,2,*} and Jose L. Garcia-Perez^{3,4}

¹Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia.

²Mater Research Institute - University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia.

³Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom.

⁴Department of Genomic Medicine, Centre for Genomics and Oncology (Pfizer—University of Granada and Andalusian Regional Government), PTS Granada, 18016 Granada, Spain.

*Correspondence: G.J.F. (faulknergj@gmail.com)

Keywords: LINE-1; L1; retrotransposon; mosaicism, embryogenesis; neurogenesis.

Abstract: The retrotransposon LINE-1 (L1) is a transposable element that has extensively colonized the mammalian germline. L1 retrotransposition can also occur in somatic cells, causing genomic mosaicism, as well as in cancer. However, the extent of L1-driven mosaicism arising during ontogenesis is unclear. Here we discuss recent experimental data which, at a minimum, fully substantiate L1 mosaicism in early embryonic development and neural cells, including post-mitotic neurons. We also consider the possible biological impact of somatic L1 insertions in neurons, the existence of donor L1s that are highly active ('hot') in specific spatiotemporal niches, and the evolutionary selection of donor L1s driving neuronal mosaicism.

A mosaic of genomes

Barbara McClintock discovered *Ac/Ds* transposition as the genetic basis for maize kernel variegation nearly 70 years ago [1,2]. In this remarkable work, McClintock simultaneously identified mobile DNA and its transposition in somatic cells, hence explaining the observed mosaic kernel phenotype. Various forms of somatic genome mosaicism have since been described [3] in normal and disease contexts, in developing and adult tissues, involving DNA changes ranging from a single nucleotide to entire chromosomes, and, in some cases, being central to critical biological processes [4]. The mobile DNA field founded by McClintock has gone on to identify numerous transposable element (TE) families, which are arguably the preeminent feature of most eukaryotic genomes sequenced to date [5] and are a major source of genetic diversity and regulatory innovation [6,7]. However, despite an ongoing emphasis on mammalian genomics, and the instructive effects of somatic transposition on plant biology revealed by McClintock and others, our understanding of TE mobilization in mammalian somatic cells remains in its infancy. In this Review, we focus on recent reports of Long INterspersed Element 1 (LINE-1, or L1) retrotransposition during murine and human embryogenesis and neurogenesis, discuss the potential biological significance of somatic L1 insertions, and consider how L1 mosaicism may be subject to evolutionary selection.

L1 retrotransposons

Retrotransposition is a molecular “copy-and-paste” process where an RNA template is reverse transcribed and integrated into the host genome, hence duplicating the donor DNA sequence from which the RNA was transcribed [8]. In humans, more than 500,000 L1 copies occupy ~17% of the genome [9]. An intact, full-length L1 is 6kb in length and initiates transcription

1
2
3
4 from a canonical 5' sense promoter (**Figure 1a**). The L1 mRNA encodes two proteins (ORF1p
5 and ORF2p) that catalyze L1 retrotransposition *in cis* [10]. The reverse transcriptase and
6
7 endonuclease activities of ORF2p are indispensable to efficient retrotransposition [11-13]. L1
8
9 also encodes an antisense peptide, ORF0, which may assist L1 mobility [14]. Most new L1
10
11 copies are rendered immobile by 5' truncation or internal mutation, leaving only 80-100
12
13 potentially mobile L1s per individual human genome [15,16]. Of these, fewer than 10 are
14
15 expected to mobilize efficiently if tested *in vitro* and are therefore described as “hot” L1s [15,17-
16
17
18 19]. The vast majority of hot L1s belong to one subfamily (L1-Ta, for Transcribed-Active)
19
20 [15,17]. Although L1 is the only remaining mobile, autonomous human TE, the non-autonomous
21
22 retrotransposon families *Alu* (a Short INterspersed Element, or SINE) and SVA (a composite
23
24 element incorporating SINE-R, a variable number of GC-rich tandem repeats, *Alu*, and a 5'
25
26 hexamer) can be retrotransposed *in trans* by the L1 protein machinery, as can be other
27
28 polyadenylated mRNAs, generating processed pseudogenes [20-24]. In mice, ~3000 L1 copies
29
30 representing three subfamilies (T_F, G_F, A) remain retrotransposition-competent (**Figure 1b**) as do
31
32 multiple endogenous retroviruses (ERVs) and non-autonomous SINE families [25-29]. As
33
34 opposed to an estimate of 1 new L1 insertion per 100 human births [30], at least 1 in 8 mice
35
36 harbor a new L1 insertion [31], despite the similar percentages of identifiable L1 sequences in
37
38 the mouse (~19% from 600,000 copies) and human genomes [9,32].
39
40
41
42
43
44
45
46
47

48 L1-mediated retrotransposition typically occurs via a mechanism called target-primed
49
50 reverse transcription (TPRT) [33] (**Box 1, Figure 1c**). As a result of TPRT, new L1 insertions
51
52 typically incorporate i) an L1 endonuclease motif, ii) target-site duplications (TSDs), and iii) a
53
54 poly(A) tail [34]. Due to cis-preference [10], L1 insertions usually arise from a
55
56 retrotransposition-competent donor L1, which in some cases can be identified by L1 flanking
57
58
59
60
61
62
63
64
65

1
2
3
4 transductions [35-37] or diagnostic internal mutations [38]. Crucially, the hallmarks of TPRT can
5
6 be utilized to discriminate genuine L1 retrotransposition events from other genetic or molecular
7
8 events involving L1 sequences [13,33,34,39-41]. A new L1 insertion can greatly impact gene
9
10 structure and function, through insertional mutagenesis of exons [38,42,43] and regulatory
11
12 elements [44], disruption of RNA polymerase II processivity [45], premature polyadenylation
13
14 [46], provision of alternative promoters [14,47-51] (**Figure 1a**) and various other functional
15
16 consequences [5,6]. L1 insertions are, likely as a result of evolutionary selection, not randomly
17
18 distributed on the genome and are depleted from exons and introns [52]. This mutagenic
19
20 potential also means the L1 5' promoter, if present in a new insertion, is heavily repressed by the
21
22 host genome in most spatiotemporal contexts [48,53-56] (**Figure 1d**). Even in situations where
23
24 full-length L1 transcripts are detected, these are usually generated by a limited number of L1
25
26 copies [38,50]. As a result, the L1 5' promoter is a major battleground in what has often been
27
28 described as an “arms race” pitting L1’s interest to replicate against the host genome’s interest to
29
30 mitigate deleterious L1 mutations [57,58]. Beyond transcriptional repression, the host genome
31
32 has developed multiple strategies to limit ongoing retrotransposition [for reviews, see [59-62]].
33
34
35
36
37
38
39
40
41
42

43 **Methods to detect L1 retrotransposition**

44
45 Two core strategies are available to resolve the spatial and temporal extent of L1
46
47 retrotransposition: L1 reporter assays and high-throughput sequencing. In 1996, an L1 reporter
48
49 assay [13] was adapted from an existing but ingenious design [8,63], and tagged human donor
50
51 L1s with an intron-containing neomycin antibiotic resistance cassette [64] that was made
52
53 functional by retrotransposition. In this assay, neomycin resistant foci function as a readout of L1
54
55 retrotransposition efficiency (**Figure 2a, left**). Remarkably, frequent L1 retrotransposition events
56
57
58
59
60
61
62
63
64
65

1
2
3
4 carrying TPRT hallmarks were observed in human and mouse cells (**Box 1**) [13]. As an
5
6 alternative approach, an enhanced green fluorescent protein (EGFP) based cassette was then
7
8 developed, yielding an L1-EGFP construct where EGFP was made functional by
9
10 retrotransposition (**Figure 2a, right**) [65]. This approach facilitated the use of fluorescence-
11
12 based microscopy and flow cytometry to measure L1 retrotransposition efficiency, including for
13
14 transgenic animals *in vivo* [66]. In all, these reporter L1s, and their derivatives (e.g. [67]), have
15
16 underpinned numerous studies elucidating retrotransposon biology over the past two decades,
17
18 and remain commonly used and effective tools [for a review, see [68]].
19
20
21
22

23
24 Alongside engineered L1 systems, high-throughput sequencing has massively increased
25
26 our ability to characterize DNA variation in human populations [52] and cancer genomes [69].
27
28 L1 insertions are, in this regard, just one type of DNA structural variant and can be studied *en*
29
30 *masse*, either as part of a whole-genome sequencing (WGS) approach, or via targeted sequencing
31
32 of L1-genome junctions (**Figure 2b**). Either strategy requires careful computational analysis and
33
34 experimental validation to confirm true L1 insertions [39,61,70] and typically leverage L1
35
36 polymorphism catalogs [71-73] to discriminate known and unknown L1 insertions. The
37
38 bioinformatic identification of new L1 insertions from WGS data [52,74-76] is advantageous in
39
40 that it can reveal the 5' and 3' L1-genome junctions of an insertion, allowing substantial
41
42 characterization of TPRT hallmarks *a priori*. As a result, WGS analyses tend to report fewer
43
44 false positives and flexibly encompass more variations of TPRT (e.g. 3' transductions [35-37]
45
46 and 5' inversions [42,77]) than can be discerned using targeted methods [30,69,78,79] analyzing
47
48 only one (usually the 3') L1-genome junction. Some targeted methods do however attempt to
49
50 analyze both L1-genome junctions simultaneously [50,80,81] and, importantly, WGS remains far
51
52 more expensive than targeted approaches. Both general strategies can be applied to “bulk” DNA
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 extracted from tissue or pooled cells, and to DNA amplified from individual cells [82,83]. High-
5
6 throughput sequencing has greatly expanded our overall capacity to study endogenous L1s *in*
7
8 *vivo*, as opposed to the considerable caveats of introducing a transgenic L1 into a new epigenetic
9
10 landscape [53,56,66,84,85]. If, however, congruent experimental data are obtained from an L1
11
12 reporter and high-throughput sequencing applied to a common biological system, such as
13
14 cultured stem cells [86,87], the conclusions are likely to be robust.
15
16
17
18
19
20

21 **Heritable and somatic L1 retrotransposition during early development**

22
23 How has L1 colonized nearly one-fifth of the human and mouse genomes? Heritable L1
24
25 insertions must, by definition, occur in a germ cell, or an embryonic cell contributing to the germ
26
27 line. A landmark 1988 study reported L1 mutagenesis of the factor VIII gene of two hemophilia
28
29 patients [42]. These results established that heritable *de novo* L1 insertions were still occurring in
30
31 humans and that these mutations could cause disease. Nonetheless, the developmental origin of
32
33 *de novo* L1 retrotransposition remained unclear [42]. Subsequent murine studies reported full-
34
35 length L1 mRNA and L1 ORF1p expression in blastocysts, male and female germ cells and,
36
37 interestingly, placental syncytiotrophoblast cells [88-91]. Differential L1 expression was
38
39 observed during germ cell specification; for example, L1 ORF1p was detected in primordial
40
41 spermatogonia, as well as the leptotene and zygotene stages of spermatogenesis, but not in
42
43 mature spermatids [88,91]. Together with later transgenic L1 mouse experiments [66,85,92-94],
44
45 recovery of endogenous L1 insertions from human germ cells [95] and studies of human X-
46
47 linked disease-causing L1 mutations [42,96,97], these reports strongly suggested endogenous L1
48
49 mobilization could occur in germ cells and the early embryo.
50
51
52
53
54
55
56

57
58 Of highest relevance here is a study [97] that reported an L1 mutation associated with
59
60
61
62
63
64
65

1
2
3
4 choroideremia, a rare recessive X-linked condition, in an affected male proband. Notably, his
5
6 mother was a somatic and germline mosaic for the L1 insertion. This example irrefutably
7
8 demonstrated that endogenous L1 retrotransposition could occur early in human embryogenesis.
9
10 In addition, the *de novo* L1 insertion carried a 3' transduction, allowing the authors to trace a
11
12 full-length donor L1 and prove it mobilized efficiently *in vitro* using the L1 reporter assay
13
14 [65,97]. As corollary, human embryonic stem cells (hESCs) support strong full-length L1 mRNA
15
16 and L1 ORF1p expression [49,86,87,98-100], as do human induced pluripotent stem cells
17
18 (hiPSCs) [87,100,101], human embryonic carcinoma cells [49,84], mouse embryonic stem cells
19
20 (mESCs) [102] and mouse induced pluripotent stem cells (miPSCs) [101]. Consistently, the L1-
21
22 EGFP reporters mobilize in hESCs, hiPSCs and embryonic carcinoma (PA-1) cells [87,100,101],
23
24 indicating that embryonic cells are likely a natural habitat for L1 retrotransposition.
25
26
27
28
29
30

31 In a recent analysis, targeted sequencing was performed on multiple cultured hESC and
32
33 hiPSC lines, followed by PCR validation of candidate *de novo* insertions in multiple laboratories
34
35 [87]. hiPSCs were reprogrammed from multiple parental cell types using a variety of approaches,
36
37 again in several different laboratories. Eleven *de novo* L1, *Alu* and SVA insertions were PCR
38
39 validated. These data confirmed that L1 was activated by reprogramming [100,103], a process
40
41 known to involve wholesale epigenomic changes [104]. Interestingly, *de novo* L1 insertions
42
43 identified in hiPSCs appeared unusually likely to be full-length, as found previously for L1-
44
45 EGFP insertions in hiPSCs [100] but not in hESCs [86]. The characteristics of L1 activity may
46
47 therefore be different in hiPSCs and hESCs, although an as yet unrealized catalog of endogenous
48
49 L1 insertions in cultured hESCs would be required to test this possibility.
50
51
52
53
54

55 In comparing the rate of endogenous L1 mobilization in hiPSCs versus hESCs, we strongly
56
57 urge consideration of how heterogeneous each cell population is. Methodological factors, such as
58
59
60
61
62
63
64
65

1
2
3
4 stem cell culture conditions, population bottlenecks in cultured cells, bioinformatic parameters,
5
6 and how candidate L1 insertions are validated, if at all, can drastically influence results [39]. For
7
8 example, a recent study [105] applied WGS to 9 hiPSC lines and did not identify any *de novo*
9
10 retrotransposon insertions, and far fewer mutations overall when compared to earlier studies
11
12 [106,107]. Another report found 7 possible *de novo* L1 insertions in 2 hiPSC lines using targeted
13
14 L1 sequencing but could not PCR validate or fully characterize the genomic integration sites of
15
16 these events [103]. A further study that analyzed 3 miPSC lines with medium coverage (10-12x
17
18 depth) WGS detected no *de novo* L1 insertions, and concluded that retroelement stability was the
19
20 rule in miPSCs [108]. Given the accumulated evidence for L1 expression and mobilization in
21
22 pluripotent cells, including retrotransposition of a codon-optimized L1 T_F element reporter [109-
23
24 111] in mESCs (Garcia-Canadas et al., unpublished data), the lack of *de novo* L1 insertions in
25
26 miPSCs is perhaps surprising. There are, however, fundamental differences in how miPSCs and
27
28 hiPSCs are generated and cultured and, as well, distinct retrotransposon families appear to be
29
30 more active depending on which mouse strain is analyzed [26,29,31]. Overall, we conclude that
31
32 reprogramming offers L1 a dynamic but consistent relaxation of repression, and that L1 also
33
34 encounters relaxed host genome control in pluripotent cells obtained directly from embryonic
35
36 material [60,101,102,112]. Embryogenesis therefore provides a favorable niche for L1
37
38 retrotransposition [31].
39
40
41
42
43
44
45
46
47

48 With this in mind, WGS and targeted sequencing was recently applied to 85 mouse
49
50 genomes obtained from 3 multigenerational C57BL/6J mouse pedigrees [31]. The developmental
51
52 timing of new L1 insertions identified in progeny was then traced in parental mice, via PCR and
53
54 quantitative PCR (qPCR) targeting the 5' L1-genome junction of individual insertions. In total,
55
56 11 *de novo* insertions were identified, with all of these being full-length (≥ 1 monomer) and
57
58
59
60
61
62
63
64
65

1
2
3
4 belonging to the T_F subfamily, indicating a rate of at least 1 new L1 insertion per 8 births. Most
5
6 heritable L1 insertions arose in the early embryo prior to germ cell specification, or in early
7
8 primordial germ cells (PGCs). For L1 insertions traced to the early embryo and early PGCs,
9
10 transmission to multiple offspring was routinely observed, suggesting that more than one allele
11
12 of a given event may be produced in one generation due to DNA replication errors and poly(A)
13
14 tail shortening post-integration [113]. TE diversity within inbred strains is therefore common and
15
16 adds to inter-strain variation [26,29].
17
18
19
20

21 Importantly, this study also identified major depletion of the 3' L1-genome junction for the
22
23 active mouse L1 families in Illumina sequencing data [31], which was attributed to obstruction
24
25 by an extensive G-quadruplex region [114,115]. To our knowledge, this issue was not identified
26
27 by previous genomic analyses of mouse L1 insertions using WGS [108,116] and is potentially
28
29 problematic for TE discovery and sensitivity calculations. For this reason, we consider the
30
31 abovementioned figure of 1/8 to be conservative [31]. As well, data obtained from transgenic
32
33 animals suggest that most engineered L1 retrotransposition occurs in the soma and are not
34
35 inherited [56,92,93,117]. Hence, heritable L1 insertions are likely far outnumbered by
36
37 endogenous L1 insertions occurring in the embryo and later during ontogenesis and lineage
38
39 specification.
40
41
42
43
44
45
46
47

48 **Do mature neurons support L1 retrotransposition?**

49

50 Over the last decade, the L1-EGFP reporter system, alongside other approaches, has been used to
51
52 elucidate engineered L1 mobilization in neural progenitors arising during fetal and adult
53
54 neurogenesis (**Box 2**), suggesting the brain may be a L1 mosaicism hotspot [53,56,84]. However,
55
56 it remains unclear whether mature neurons, or other cell lineages, also accommodate L1 activity.
57
58
59
60
61
62
63
64
65

1
2
3
4 Recently, a human L1-EGFP reporter was introduced into hESC derived neuronal precursor cells
5
6 (NPCs) and, as seen previously [53,56], observed efficient retrotransposition was observed [99].
7
8 The authors then exploited a hybrid L1 adenoviral vector [118] to transduce NPCs with a
9
10 modified L1-EGFP reporter, overcoming limitations associated with plasmid transfection, and
11
12 again found L1 retrotransposition. Finally, the authors differentiated NPCs for 31 days to force
13
14 neuronal maturation, then introduced either the adenoviral or plasmid L1-EGFP reporter along
15
16 with 5-Bromo-2'-deoxyuridine (BrdU), a marker of cycling cells, and found that EGFP⁺ neuronal
17
18 cells were present and not stained with an anti-BrdU antibody. These results suggested that non-
19
20 dividing neuronal cells supported extensive engineered L1 mobilization [118]. The authors
21
22 performed parallel experiments in isogenic hematopoietic and mesenchymal stem cells and,
23
24 compared to NPCs, observed very low L1 expression and L1-EGFP activity. Through infection
25
26 with the adenoviral L1-EGFP reporter, and the use of qPCR to measure integrated EGFP copies,
27
28 it was determined that the rate of L1-EGFP insertions in mature neurons was at least as high as in
29
30 NPCs. These conclusions relied heavily on PCR and qPCR detection of the spliced EGFP
31
32 cassette, and normalization to a plasmid or adenovirus [53]. If taken together, this study and
33
34 previous studies focused on L1 in NPCs [53,56,99], lead us to conclude that engineered L1
35
36 activity, in the cell types and physiological conditions tested thus far, is largely restricted to the
37
38 neuronal lineage, including post-mitotic neurons.
39
40
41
42
43
44
45
46
47
48
49

50 **Extent of endogenous L1 mobilization in the brain**

51
52 Despite ongoing debate regarding the various types of mosaic DNA variation found in the brain
53
54 [70,81,119-121], an unequivocal consensus view, based on genomic analysis of bulk brain tissue
55
56 [53,80], individual cells [41,81-83] and clonal cell lines derived from individual neurons [116]
57
58
59
60
61
62
63
64
65

1
2
3
4 has formed: endogenous L1 retrotransposition can occur in the neuronal lineage, in line with
5
6 foundational data obtained from engineered L1 reporter systems [53,55,56,99]. Estimates of L1
7
8 mobilization rate have nonetheless varied widely in each of the relevant studies, which have used
9
10 various analytical approaches (Table 1). The earliest, and most approximate, calculations of per
11
12 cell somatic L1 insertion count were based on a L1 copy number variation (CNV) assay [53],
13
14 and retrotransposon capture sequencing (RC-seq) [39,80] applied to bulk hippocampal tissue,
15
16 setting what appear now to be the extrema values of 80 and 0.04, respectively. In a 2012
17
18 methodological tour de force [82], multiple displacement amplification (MDA) was coupled to
19
20 an earlier targeted 3' L1-genome sequencing approach [30,78] to create L1 insertion profiling
21
22 (L1-IP). Application of L1-IP to 300 individual pyramidal neurons from cortex and caudate
23
24 nucleus revealed a single somatic L1 insertion, which carried a 5' transduction and could be PCR
25
26 amplified and capillary sequenced in its entirety (an “empty/filled” assay, which we consider to
27
28 be the highest validation standard). Another 4 events were detected by L1-IP but could be PCR
29
30 amplified only at their 3' L1-genome junction. Subsequent WGS [83] applied to 16 of the MDA-
31
32 amplified cortical neurons analyzed by L1-IP, including the neuron where the archetypal
33
34 neuronal L1 insertion [82] was found, re-identified that event as well as another somatic L1
35
36 insertion flanked by a 614nt 3' transduction that was, for this reason, initially overlooked by L1-
37
38 IP [82]. Two additional single-cell studies of hippocampal and cortical neurons, via MDA
39
40 followed by somatic L1-associated variant sequencing (SLAV-seq) [41], and through multiple
41
42 annealing and looping-based amplification cycles (MALBAC) followed by RC-seq [81], also
43
44 identified and PCR validated multiple somatic L1 variants. Hence, single-cell genomic analyses
45
46 have consistently found strong evidence for endogenous L1 mobilization in the neuronal lineage
47
48 (Table 1).
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 As a discipline still in its infancy, single-cell genomic analysis can lead to conclusions
5 heavily influenced by technical considerations. For example, a major signature of somatic C>T
6 mutations reported by one analysis of MDA amplified neurons [119] was suggested to be an
7 artifact by another study employing a different genome-wide amplification approach [120]. In
8 these circumstances, cooperation and consensus building are essential and, fortunately, the field
9 is moving in this direction [70]. Nonetheless, discovery and characterization of somatic L1
10 insertions found in a handful of cells, or even one cell, via single-cell genomics remains
11 technically challenging [39] because whole-genome amplification and sequencing library
12 preparation can each generate molecular artifacts, or chimeras, that obscure real L1 insertions.
13 Sophisticated bioinformatic strategies tailored to the underlying single-cell genomic approach are
14 hence required to distinguish signal and noise. For example, variant discovery with the three
15 targeted L1 sequencing methods used thus far to analyze neuronal genomes has filtered
16 candidate *de novo* L1 insertions primarily based on read count (L1-IP), L1 integration site
17 sequence features (single-cell RC-seq), or using a combination of both read count and sequence
18 features (SLAV-seq) [41,81,82]. If the analysis approach suitable to one technique is applied to
19 another (e.g. applying lessons learned from single-cell RC-seq to L1-IP [81], or the reciprocal
20 application of a read count filter suitable for L1-IP to single-cell RC-seq data already filtered
21 based on sequence features [121]), the resulting L1 mobilization rate estimate can be very
22 different, necessitating method standardization [39,70]. The common ground shared by all of
23 these techniques is their high false positive rates, a need for rigorous and time-consuming PCR
24 validation, and their assumption that heterozygous L1 variants in single-cell genomic analysis
25 are equivalent to somatic L1 variants [41,81,82]. This latter consideration is central to the
26 estimation of false negatives. In this regard, it should be noted that the poly(A) tails (91nt and
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 107nt in length, on average) of the two somatic L1 insertions validated to date by the
5
6 empty/filled PCR assay and presenting clear TSDs [83] are significantly longer and more
7
8 adenine pure than those carried by the vast majority of heterozygous L1 insertions [17], due to
9
10 rapid intra-individual and intergenerational poly(A) tail shortening [83,113], and this
11
12 phenomenon is even more evident for older L1 insertions [113]. Illumina sequencing is known to
13
14 have issues with long homopolymer tracts [122] and it is unclear how very long poly(A)
15
16 sequences fare during whole genome amplification. Moreover, it is interesting that engineered
17
18 L1 insertions have been shown to accumulate mainly in post-mitotic neurons [99], whereas the
19
20 two somatic L1 insertions referred to above were each detected in multiple neurons [83]. These
21
22 considerations lead us to ask whether the false negative rate has been consistently
23
24 underestimated when assessing the degree of L1 mosaicism in the brain with single-cell
25
26 genomics, whilst acknowledging that accurate false positive rate calculations are essential
27
28 [81,121]. Finally, it must be noted that single-cell genomic analyses of L1 mobilization have
29
30 been performed on very few human brain samples thus far, and on broad neuronal types, leaving
31
32 open the possibility that some individuals, brain regions and neuronal subtypes may support
33
34 more endogenous L1 activity than others, and thus contribute to disparate somatic L1
35
36 retrotransposition frequency estimates.
37
38
39
40
41
42
43
44

45
46 To our knowledge, no single-cell analysis of endogenous L1 mobilization in the mouse
47
48 brain has been reported to date. However, in an elegant study, somatic cell nuclear transfer
49
50 (SCNT) was used to reprogram mESCs with neuronal nuclei obtained from the mouse olfactory
51
52 bulb, followed by clonal expansion and bulk WGS to identify *de novo* TE insertions and other
53
54 somatic variants [116]. This approach provided an excellent and robust alternative to whole-
55
56 genome amplification and eliminated errors associated with the latter technique, although also
57
58
59
60
61
62
63
64
65

1
2
3
4 potentially selecting neuronal nuclei with a lower burden of DNA damage (including from L1)
5
6 [116]. In 6 reprogrammed neuronal clones, 4 *de novo* L1 insertions were validated through
7
8 junction-specific PCR and capillary sequencing, revealing in each case hallmark features of
9
10 TPRT. Based on a false negative rate of approximately ~50%, the analyzed neurons likely each
11
12 contained ~1.3 somatic L1 insertions, on average. Interestingly, this rate estimate differed quite
13
14 dramatically from the extrema values of 0.04 [82] and 13.7 [81] obtained from single-cell
15
16 genomic analysis of human neurons, although L1 appears to be more active in the mouse than in
17
18 the human germline [30,31]. It is unclear how much, if at all, the 3' L1-genome junction
19
20 depletion observed recently in WGS and RC-seq data [31] affected the false negative rate
21
22 calculation of this study, given that the WGS analysis appeared to group all TE families together
23
24 when calculating false negative rate, and the 3' depletion observed elsewhere was L1-specific
25
26 [31]. More generally, it is unknown how much L1 activity varies in the brains of different
27
28 species, or different inbred animal strains, or for that matter how much ageing and senescence
29
30 impact TE mosaicism in species with very dissimilar lifespans [123-126]. It is nonetheless
31
32 remarkable that L1 mosaicism may be very common in the mouse brain, and conserved in
33
34 Mammalia, based on the conservative estimate that olfactory neurons contain at least one
35
36 somatic L1 insertion, on average [116].
37
38
39
40
41
42
43
44
45
46
47

48 **When does L1 jump in brain development?**

49
50 As noted above, engineered L1 insertions occur throughout fetal and adult neurogenesis, as well
51
52 as in mature neurons [53,56,99]. With regards to endogenous L1 activity, one study detected two
53
54 somatic L1 insertions, each in 2/16 neurons assayed by WGS [83]. By lineage tracing, the
55
56 authors found that one of these events was timed to occur in the developing cortex and the other
57
58
59
60
61
62
63
64
65

1
2
3
4 likely arose early in central nervous system development, and perhaps even earlier [83]. The
5
6 latter circumstance reconciles well with embryonic events elucidated in mouse [31,92]. By
7
8 contrast, two other studies found that most of these events appeared to arise later in neurogenesis
9
10 [41,81], agreeing with reports of engineered L1 mobilization in post-mitotic neurons [99].
11
12 Interestingly, studies of engineered and endogenous L1 retrotransposition in brain tissues and
13
14 neural cells have recurrently found L1 insertions in neuronal genes [41,53,56,80,81] and
15
16 enhancers active in NSCs [81], raising the prospect of integration patterns specific to the
17
18 neuronal epigenetic landscape, or post-integration selection. Pyramidal and other neuronal
19
20 subtypes have been shown to contain somatic L1 insertions [81-83], as have, in far fewer
21
22 instances, glia [41,81]. It therefore remains unclear whether most somatic L1 insertions found in
23
24 the brain arose in the embryo, during neurogenesis, in mature post-mitotic neurons, or, as is
25
26 possible, in each of these scenarios, leading to complex neuronal mosaicism.
27
28
29
30
31
32
33
34
35

36 **A model for evolutionary selection of somatic L1 retrotransposition**

37
38 Is an L1 mosaic brain functionally distinct to an L1 homogenous brain? We note here only that i)
39
40 neuronal circuitry is highly interconnected and exquisitely sensitive to perturbation [127], ii)
41
42 intragenic L1 insertions can grossly impact gene function [42,45], iii) despite this, the potential
43
44 roles of L1 mosaicism in learning and cognition remain almost entirely theoretical [39,128,129],
45
46 and iv) abnormal somatic L1 activity in neurological disorders, including Rett syndrome (RTT)
47
48 [55,130], schizophrenia [131] and ataxia telangiectasia [132], has also been considered
49
50 extensively but, apart from RTT, the related etiological contribution of L1 to disease is very
51
52 unclear. Even for RTT, where MeCP2, a major L1 transcriptional repressor [130], is mutated and
53
54 L1 mRNA, L1 protein and L1-EGFP transgene activity are all elevated [55], MeCP2 conditional
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 rescue can restore apparently normal neurobiological function in mice [133], meaning L1
5
6 mosaicism is unlikely to be a major component of RTT neuronal phenotype. Hence, although L1
7
8 insertions can impact phenotype in the context of disease [38,42,43], it remains wholly unclear
9
10 whether this applies to normal or abnormal neurobiology.
11
12

13
14 Somatic mutations are, of course, not inherited. However, donor L1s causing *de novo* L1
15
16 insertions in somatic cells are carried through the germline, and are therefore subject to selection
17
18 because they can simultaneously cause germline and somatic mosaicism [31,38]. Moreover, if a
19
20 particular donor L1 is very active in somatic cells it may affect the immediate evolutionary
21
22 fitness of the carrier individual, through disease [38,43] or even positive developmental or
23
24 neurological consequences, if they exist [129]. That some L1s are apparently more mobile in
25
26 somatic cells than elsewhere is supported by the identification of donor L1s that are far more
27
28 active in tumors than would be expected by their activity in the germline, such as an oft-
29
30 transduced donor L1 in the *TTC28* gene of numerous cancer genomes [76]. Reciprocally, some
31
32 donor L1s are sufficiently active to give rise to multiple donor L1 progeny in the human
33
34 germline but have not been found to be particularly active in cancer [17,76,134]. Finally, some
35
36 donor L1s are highly active in both the germline and tumors [17,38].
37
38
39
40
41
42

43 Donor L1s can each have multiple alleles, which can in turn mobilize at very different
44
45 rates, even in the same context [18,19]. As well, the same donor L1 may mobilize well in one
46
47 context and not another [50]. For example, the donor L1 found previously to generate a 3'
48
49 transduction-flanked neuronal L1 insertion [83] putatively mobilized during brain development
50
51 but, when tested with an L1 reporter assay, did not retrotranspose in cultured osteosarcoma cells
52
53 [15]. It follows that, as more active donor L1s generate longer new L1 insertions [135], they
54
55 have a higher chance of generating retrotransposition-competent L1 insertions that can be easily
56
57
58
59
60
61
62
63
64
65

1
2
3
4 traced back to their donor L1s by 5' or 3' transduced sequences, as was the case for both of the
5
6 neuronal L1 insertions referred to above [83]. For these reasons, we hypothesize that donor L1s
7
8 that are “hot” for retrotransposition [15] in certain somatic contexts *in vivo* (**Figure 3, Key**
9
10 **Figure**) exist in the human population. This possibility is further supported by context-specific
11
12 donor L1 activity in cultured cell lines [50], and a recent colorectal cancer study [38] that found a
13
14 tumorigenic L1 insertion in the *APC* gene and traced that mutation to a polymorphic donor L1
15
16 that was demethylated in the tumor, but also the matched normal colon [17]. If other
17
18 polymorphic donor L1s are highly active in the brain, and L1 mosaicism is ultimately found to
19
20 impact neurobiology, we predict that donor L1s, the regulatory elements they carry (e.g.
21
22 antisense promoters [14,48]) and the relevant host defense factors, may undergo genetic selection
23
24 due to their activity in the soma. Despite somatic L1 insertions not being inherited, this model
25
26 could lead to varying rates of L1 mosaicism amongst individuals, and subject the phenomenon to
27
28 natural selection.
29
30
31
32
33
34
35
36
37

38 **Concluding remarks and future directions**

39
40 Endogenous L1 retrotransposition occurs in the embryo and during neurogenesis, and causes
41
42 somatic genome mosaicism in neurons. The character of this mosaicism, in terms of complexity
43
44 and impact, remains largely undefined. However, as the average human brain contains 80-100
45
46 billion neurons [136], even the most conservative estimates of neuronal L1 mosaicism
47
48 extrapolate to a very extensive catalogue of L1-driven variation within any individual. We would
49
50 also expect that some neuronal subtypes support more L1 activity than others, perhaps as a
51
52 function of when during life those neurons arise, their spatial distribution in the brain, or their
53
54 neurobiological function, and in those neurons the potential for L1 insertions to drive phenotypic
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 diversity is arguably higher than in cells that carry few or no somatic L1 insertions. TE
5
6 mobilization in somatic cells is, of course, not restricted to mammals, with McClintock's maize
7
8 [1,2], silk worm [137] and fruit fly [125,138] each providing examples of mosaicism caused by
9
10 mobile DNA. Major advances in single-cell genomic analysis and high-throughput sequencing
11
12 therefore leave the field well placed to further define somatic genome mosaicism, and its
13
14 potential functional consequences, in different species and biological contexts (see Outstanding
15
16 Questions).

23 **Acknowledgements**

24
25
26 G.J.F. acknowledges the support of a CSL Centenary Fellowship, NHMRC Project Grants
27
28 (GNT1106206, GNT1125645, GNT1126393) and the Mater Foundation. J.L.G.-P. acknowledges
29
30 funding from CICE-FEDER-P12-CTS-2256, Plan Nacional de I+D+I 2013-2016 (FIS-FEDER-
31
32 PI14/02152), PCIN-2014-115-ERA-NET NEURON II, the European Research Council (ERC-
33
34 Consolidator ERC-STG-2012-233764), an International Early Career Scientist grant from the
35
36 Howard Hughes Medical Institute (IECS-55007420) and The Wellcome Trust-University of
37
38 Edinburgh Institutional Strategic Support Fund (ISFF2). We thank Francisco Sanchez-Luque for
39
40 helpful discussions and figure design, and Sandra Richardson and Adam Ewing for critical
41
42 review of the manuscript.
43
44
45
46
47
48
49

50 **References**

- 51
52
53 1 McClintock, B. (1951) Chromosome organization and genic expression. *Cold Spring Harb*
54 *Symp Quant Biol* 16, 13-47
55
56 2 McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U*
57 *S A* 36, 344-355
58
59
60
61
62
63
64
65

- 1
2
3
4 3 Campbell, I.M., *et al.* (2015) Somatic mosaicism: implications for disease and transmission
5 genetics. *Trends Genet* 31, 382-392
6
7
8 4 Hozumi, N. and Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin
9 genes coding for variable and constant regions. *Proc Natl Acad Sci U S A* 73, 3628-3632
10
11 5 Levin, H.L. and Moran, J.V. (2011) Dynamic interactions between transposable elements and
12 their hosts. *Nat Rev Genet* 12, 615-627
13
14
15 6 Chuong, E.B., *et al.* (2017) Regulatory activities of transposable elements: from conflicts to
16 benefits. *Nat Rev Genet* 18, 71-86
17
18 7 Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. *Science* 165,
19 349-357
20
21 8 Boeke, J.D., *et al.* (1985) Ty elements transpose through an RNA intermediate. *Cell* 40, 491-
22 500
23
24
25 9 Lander, E.S., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409,
26 860-921
27
28 10 Wei, W., *et al.* (2001) Human L1 retrotransposition: cis preference versus trans
29 complementation. *Mol Cell Biol* 21, 1429-1439
30
31
32 11 Feng, Q., *et al.* (1996) Human L1 retrotransposon encodes a conserved endonuclease required
33 for retrotransposition. *Cell* 87, 905-916
34
35 12 Mathias, S.L., *et al.* (1991) Reverse transcriptase encoded by a human transposable element.
36 *Science* 254, 1808-1810
37
38
39 13 Moran, J.V., *et al.* (1996) High frequency retrotransposition in cultured mammalian cells.
40 *Cell* 87, 917-927
41
42 14 Denli, A.M., *et al.* (2015) Primate-Specific ORF0 Contributes to Retrotransposon-Mediated
43 Diversity. *Cell* 163, 583-593
44
45 15 Brouha, B., *et al.* (2003) Hot L1s account for the bulk of retrotransposition in the human
46 population. *Proc Natl Acad Sci U S A* 100, 5280-5285
47
48
49 16 Mills, R.E., *et al.* (2007) Which transposable elements are active in the human genome?
50 *Trends Genet* 23, 183-191
51
52 17 Beck, C.R., *et al.* (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141,
53 1159-1170
54
55 18 Lutz, S.M., *et al.* (2003) Allelic heterogeneity in LINE-1 retrotransposition activity. *Am J*
56 *Hum Genet* 73, 1431-1437
57
58
59
60
61
62
63
64
65

- 1
2
3
4 19 Seleme, M.C., *et al.* (2006) Extensive individual variation in L1 retrotransposition capability
5 contributes to human genetic diversity. *Proc Natl Acad Sci U S A* 103, 6611-6616
6
7
8 20 Esnault, C., *et al.* (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat*
9 *Genet* 24, 363-367
10
11 21 Doucet, A.J., *et al.* (2015) A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Mol*
12 *Cell* 60, 728-741
13
14
15 22 Hancks, D.C., *et al.* (2011) Retrotransposition of marked SVA elements by human L1s in
16 cultured cells. *Hum Mol Genet* 20, 3386-3400
17
18
19 23 Dewannieux, M., *et al.* (2003) LINE-mediated retrotransposition of marked Alu sequences.
20 *Nat Genet* 35, 41-48
21
22 24 Ahl, V., *et al.* (2015) Retrotransposition and Crystal Structure of an Alu RNP in the
23 Ribosome-Stalling Conformation. *Mol Cell* 60, 715-727
24
25
26 25 Ostertag, E.M. and Kazazian, H.H., Jr. (2001) Biology of mammalian L1 retrotransposons.
27 *Annu Rev Genet* 35, 501-538
28
29 26 Nellaker, C., *et al.* (2012) The genomic landscape shaped by selection on transposable
30 elements across 18 mouse strains. *Genome Biol* 13, R45
31
32
33 27 Sookdeo, A., *et al.* (2013) Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob*
34 *DNA* 4, 3
35
36 28 Goodier, J.L., *et al.* (2001) A novel active L1 retrotransposon subfamily in the mouse.
37 *Genome Res* 11, 1677-1685
38
39
40 29 Maksakova, I.A., *et al.* (2006) Retroviral elements and their hosts: insertional mutagenesis in
41 the mouse germ line. *PLoS Genet* 2, e2
42
43 30 Ewing, A.D. and Kazazian, H.H., Jr. (2010) High-throughput sequencing reveals extensive
44 variation in human-specific L1 content in individual human genomes. *Genome Res* 20, 1262-
45 1270
46
47
48 31 Richardson, S.R., *et al.* (2017) Heritable L1 retrotransposition in the mouse primordial
49 germline and early embryo. *Genome Res* DOI: 10.1101/gr.219022.116
50
51 32 Waterston, R.H., *et al.* (2002) Initial sequencing and comparative analysis of the mouse
52 genome. *Nature* 420, 520-562
53
54
55 33 Luan, D.D., *et al.* (1993) Reverse transcription of R2Bm RNA is primed by a nick at the
56 chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595-605
57
58 34 Jurka, J. (1997) Sequence patterns indicate an enzymatic involvement in integration of
59 mammalian retroposons. *Proc Natl Acad Sci U S A* 94, 1872-1877
60
61
62
63
64
65

- 1
2
3
4 35 Goodier, J.L., *et al.* (2000) Transduction of 3'-flanking sequences is common in L1
5 retrotransposition. *Hum Mol Genet* 9, 653-657
6
7
8 36 Moran, J.V., *et al.* (1999) Exon shuffling by L1 retrotransposition. *Science* 283, 1530-1534
9
10 37 Pickeral, O.K., *et al.* (2000) Frequent human genomic DNA transduction driven by LINE-1
11 retrotransposition. *Genome Res* 10, 411-415
12
13 38 Scott, E.C., *et al.* (2016) A hot L1 retrotransposon evades somatic repression and initiates
14 human colorectal cancer. *Genome Res* 26, 745-755
15
16
17 39 Richardson, S.R., *et al.* (2014) L1 retrotransposons and somatic mosaicism in the brain. *Annu*
18 *Rev Genet* 48, 1-27
19
20 40 Monot, C., *et al.* (2013) The specificity and flexibility of l1 reverse transcription priming at
21 imperfect T-tracts. *PLoS Genet* 9, e1003499
22
23 41 Erwin, J.A., *et al.* (2016) L1-associated genomic regions are deleted in somatic cells of the
24 healthy human brain. *Nat Neurosci* 19, 1583-1591
25
26
27 42 Kazazian, H.H., Jr., *et al.* (1988) Haemophilia A resulting from de novo insertion of L1
28 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164-166
29
30 43 Miki, Y., *et al.* (1992) Disruption of the APC gene by a retrotransposal insertion of L1
31 sequence in a colon cancer. *Cancer Res* 52, 643-645
32
33 44 Shukla, R., *et al.* (2013) Endogenous retrotransposition activates oncogenic pathways in
34 hepatocellular carcinoma. *Cell* 153, 101-111
35
36
37 45 Han, J.S., *et al.* (2004) Transcriptional disruption by the L1 retrotransposon and implications
38 for mammalian transcriptomes. *Nature* 429, 268-274
39
40
41 46 Perepelitsa-Belancio, V. and Deininger, P. (2003) RNA truncation by premature
42 polyadenylation attenuates human mobile element activity. *Nat Genet* 35, 363-366
43
44 47 Wheelan, S.J., *et al.* (2005) Gene-breaking: a new paradigm for human retrotransposon-
45 mediated gene evolution. *Genome Res* 15, 1073-1078
46
47
48 48 Faulkner, G.J., *et al.* (2009) The regulated retrotransposon transcriptome of mammalian cells.
49 *Nat Genet* 41, 563-571
50
51
52 49 Macia, A., *et al.* (2011) Epigenetic control of retrotransposon expression in human embryonic
53 stem cells. *Mol Cell Biol* 31, 300-316
54
55 50 Philippe, C., *et al.* (2016) Activation of individual L1 retrotransposon instances is restricted to
56 cell-type dependent permissive loci. *Elife* 5, e13926
57
58
59 51 Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of
60 adjacent cellular genes. *Mol Cell Biol* 21, 1973-1985
61
62
63
64
65

- 1
2
3
4 52 Ewing, A.D. and Kazazian, H.H., Jr. (2011) Whole-genome resequencing allows detection of
5 many rare LINE-1 insertion alleles in humans. *Genome Res* 21, 985-990
6
7
8 53 Coufal, N.G., *et al.* (2009) L1 retrotransposition in human neural progenitor cells. *Nature*
9 460, 1127-1131
10
11 54 Hata, K. and Sakaki, Y. (1997) Identification of critical CpG sites for repression of L1
12 transcription by DNA methylation. *Gene* 189, 227-234
13
14
15 55 Muotri, A.R., *et al.* (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature*
16 468, 443-446
17
18 56 Muotri, A.R., *et al.* (2005) Somatic mosaicism in neuronal precursor cells mediated by L1
19 retrotransposition. *Nature* 435, 903-910
20
21
22 57 Jacobs, F.M., *et al.* (2014) An evolutionary arms race between KRAB zinc-finger genes
23 ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516, 242-245
24
25 58 Imbeault, M., *et al.* (2017) KRAB zinc-finger proteins contribute to the evolution of gene
26 regulatory networks. *Nature* 543, 550-554
27
28 59 Friedli, M. and Trono, D. (2015) The developmental control of transposable elements and the
29 evolution of higher species. *Annu Rev Cell Dev Biol* 31, 429-451
30
31
32 60 Gerdes, P., *et al.* (2016) Transposable elements in the mammalian embryo: pioneers surviving
33 through stealth and service. *Genome Biol* 17, 100
34
35
36 61 Goodier, J.L. (2016) Restricting retrotransposons: a review. *Mob DNA* 7, 16
37
38 62 Thompson, P.J., *et al.* (2016) Long Terminal Repeats: From Parasitic Elements to Building
39 Blocks of the Transcriptional Regulatory Repertoire. *Mol Cell* 62, 766-776
40
41
42 63 Heidmann, T., *et al.* (1988) An indicator gene to demonstrate intracellular transposition of
43 defective retroviruses. *Proc Natl Acad Sci U S A* 85, 2219-2223
44
45 64 Freeman, J.D., *et al.* (1994) A modified indicator gene for selection of retrotransposition
46 events in mammalian cells. *Biotechniques* 17, 46, 48-49, 52
47
48
49 65 Ostertag, E.M., *et al.* (2000) Determination of L1 retrotransposition kinetics in cultured cells.
50 *Nucleic Acids Res* 28, 1418-1423
51
52 66 Ostertag, E.M., *et al.* (2002) A mouse model of human L1 retrotransposition. *Nat Genet* 32,
53 655-660
54
55
56 67 Xie, Y., *et al.* (2011) Characterization of L1 retrotransposition with high-throughput dual-
57 luciferase assays. *Nucleic Acids Res* 39, e16
58
59 68 Rangwala, S.H. and Kazazian, H.H., Jr. (2009) The L1 retrotransposition assay: a
60 retrospective and toolkit. *Methods* 49, 219-226
61
62
63
64
65

- 1
2
3
4 69 Iskow, R.C., *et al.* (2010) Natural mutagenesis of human genomes by endogenous
5 retrotransposons. *Cell* 141, 1253-1261
6
7
8 70 McConnell, M.J., *et al.* (2017) Intersection of diverse neuronal genomes and neuropsychiatric
9 disease: The Brain Somatic Mosaicism Network. *Science* 356
10
11 71 Mir, A.A., *et al.* (2015) euL1db: the European database of L1HS retrotransposon insertions in
12 humans. *Nucleic Acids Res* 43, D43-47
13
14
15 72 Sudmant, P.H., *et al.* (2015) An integrated map of structural variation in 2,504 human
16 genomes. *Nature* 526, 75-81
17
18 73 Wang, J., *et al.* (2006) dbRIP: a highly integrated database of retrotransposon insertion
19 polymorphisms in humans. *Hum Mutat* 27, 323-329
20
21
22 74 Helman, E., *et al.* (2014) Somatic retrotransposition in human cancer revealed by whole-
23 genome and exome sequencing. *Genome Res* 24, 1053-1063
24
25 75 Lee, E., *et al.* (2012) Landscape of somatic retrotransposition in human cancers. *Science* 337,
26 967-971
27
28
29 76 Tubio, J.M., *et al.* (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive
30 DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345, 1251343
31
32
33 77 Ostertag, E.M. and Kazazian, H.H. (2001) Twin priming: a proposed mechanism for the
34 creation of inversions in L1 retrotransposition. *Genome Res* 11, 2059-2065
35
36 78 Badge, R.M., *et al.* (2003) ATLAS: a system to selectively identify human-specific L1
37 insertions. *Am J Hum Genet* 72, 823-838
38
39 79 Rodic, N., *et al.* (2015) Retrotransposon insertions in the clonal evolution of pancreatic ductal
40 adenocarcinoma. *Nat Med* 21, 1060-1064
41
42
43 80 Baillie, J.K., *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the
44 human brain. *Nature* 479, 534-537
45
46 81 Upton, K.R., *et al.* (2015) Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* 161, 228-
47 239
48
49
50 82 Evrony, G.D., *et al.* (2012) Single-neuron sequencing analysis of L1 retrotransposition and
51 somatic mutation in the human brain. *Cell* 151, 483-496
52
53 83 Evrony, G.D., *et al.* (2015) Cell lineage analysis in human brain using endogenous
54 retroelements. *Neuron* 85, 49-59
55
56
57 84 Garcia-Perez, J.L., *et al.* (2010) Epigenetic silencing of engineered L1 retrotransposition
58 events in human embryonic carcinoma cells. *Nature* 466, 769-773
59
60
61
62
63
64
65

- 1
2
3
4 85 Prak, E.T., *et al.* (2003) Tracking an embryonic L1 retrotransposition event. *Proc Natl Acad Sci U S A* 100, 1832-1837
5
6
7
8 86 Garcia-Perez, J.L., *et al.* (2007) LINE-1 retrotransposition in human embryonic stem cells.
9 *Hum Mol Genet* 16, 1569-1577
10
11 87 Klawitter, S., *et al.* (2016) Reprogramming triggers endogenous L1 and Alu retrotransposition
12 in human induced pluripotent stem cells. *Nat Commun* 7, 10286
13
14
15 88 Branciforte, D. and Martin, S.L. (1994) Developmental and cell type specificity of LINE-1
16 expression in mouse testis: implications for transposition. *Mol Cell Biol* 14, 2584-2592
17
18 89 Malki, S., *et al.* (2014) A role for retrotransposon LINE-1 in fetal oocyte attrition in mice.
19 *Dev Cell* 29, 521-533
20
21 90 Packer, A.I., *et al.* (1993) A discrete LINE-1 transcript in mouse blastocysts. *Dev Biol* 157,
22 281-283
23
24
25 91 Trelogan, S.A. and Martin, S.L. (1995) Tightly regulated, developmentally specific
26 expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl*
27 *Acad Sci U S A* 92, 1520-1524
28
29
30 92 Kano, H., *et al.* (2009) L1 retrotransposition occurs mainly in embryogenesis and creates
31 somatic mosaicism. *Genes Dev* 23, 1303-1312
32
33 93 An, W., *et al.* (2006) Active retrotransposition by a synthetic L1 element in mice. *Proc Natl*
34 *Acad Sci U S A* 103, 18662-18667
35
36
37 94 Newkirk, S.J., *et al.* (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis.
38 *Proc Natl Acad Sci U S A* DOI: 10.1073/pnas.1701069114
39
40 95 Freeman, P., *et al.* (2011) L1 hybridization enrichment: a method for directly accessing de
41 novo L1 insertions in the human germline. *Hum Mutat* 32, 978-988
42
43
44 96 Brouha, B., *et al.* (2002) Evidence consistent with human L1 retrotransposition in maternal
45 meiosis I. *Am J Hum Genet* 71, 327-336
46
47 97 van den Hurk, J.A., *et al.* (2007) L1 retrotransposition can occur early in human embryonic
48 development. *Hum Mol Genet* 16, 1587-1592
49
50
51 98 Castro-Diaz, N., *et al.* (2014) Evolutionally dynamic L1 regulation in embryonic stem cells.
52 *Genes Dev* 28, 1397-1409
53
54 99 Macia, A., *et al.* (2017) Engineered LINE-1 retrotransposition in nondividing human neurons.
55 *Genome Res* 27, 335-348
56
57
58 100 Wissing, S., *et al.* (2012) Reprogramming somatic cells into iPS cells activates LINE-1
59 retroelement mobility. *Hum Mol Genet* 21, 208-218
60
61
62
63
64
65

- 1
2
3
4 101 Friedli, M., *et al.* (2014) Loss of transcriptional control over endogenous retroelements
5 during reprogramming to pluripotency. *Genome Res* 24, 1251-1259
6
7
8 102 Walter, M., *et al.* (2016) An epigenetic switch ensures transposon repression upon dynamic
9 loss of DNA methylation in embryonic stem cells. *Elife* 5, e11418
10
11 103 Arokium, H., *et al.* (2014) Deep sequencing reveals low incidence of endogenous LINE-1
12 retrotransposition in human induced pluripotent stem cells. *PLoS One* 9, e108682
13
14
15 104 Lister, R., *et al.* (2011) Hotspots of aberrant epigenomic reprogramming in human induced
16 pluripotent stem cells. *Nature* 471, 68-73
17
18 105 Bhutani, K., *et al.* (2016) Whole-genome mutational burden analysis of three pluripotency
19 induction methods. *Nat Commun* 7, 10536
20
21 106 Gore, A., *et al.* (2011) Somatic coding mutations in human induced pluripotent stem cells.
22 *Nature* 471, 63-67
23
24
25 107 Hussein, S.M., *et al.* (2011) Copy number variation and selection during reprogramming to
26 pluripotency. *Nature* 471, 58-62
27
28 108 Quinlan, A.R., *et al.* (2011) Genome sequencing of mouse induced pluripotent stem cells
29 reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell*
30 *Stem Cell* 9, 366-373
31
32
33 109 Goodier, J.L., *et al.* (2007) LINE-1 ORF1 protein localizes in stress granules with other
34 RNA-binding proteins, including components of RNA interference RNA-induced silencing
35 complex. *Mol Cell Biol* 27, 6469-6483
36
37
38 110 Han, J.S. and Boeke, J.D. (2004) A highly active synthetic mammalian retrotransposon.
39 *Nature* 429, 314-318
40
41 111 Naas, T.P., *et al.* (1998) An actively retrotransposing, novel subfamily of mouse L1
42 elements. *EMBO J* 17, 590-597
43
44
45 112 Theunissen, T.W., *et al.* (2016) Molecular Criteria for Defining the Naive Human
46 Pluripotent State. *Cell Stem Cell* 19, 502-515
47
48 113 Grandi, F.C., *et al.* (2013) LINE-1-derived poly(A) microsatellites undergo rapid shortening
49 and create somatic and germline mosaicism in mice. *Mol Biol Evol* 30, 503-512
50
51
52 114 Howell, R. and Usdin, K. (1997) The ability to form intrastrand tetraplexes is an
53 evolutionarily conserved feature of the 3' end of L1 retrotransposons. *Mol Biol Evol* 14, 144-155
54
55 115 Sahakyan, A.B., *et al.* (2017) G-quadruplex structures within the 3' UTR of LINE-1
56 elements stimulate retrotransposition. *Nat Struct Mol Biol* 24, 243-247
57
58
59
60
61
62
63
64
65

- 1
2
3
4 116 Hazen, J.L., *et al.* (2016) The Complete Genome Sequences, Unique Mutational Spectra,
5 and Developmental Potency of Adult Neurons Revealed by Cloning. *Neuron* 89, 1223-1236
6
7
8 117 Babushok, D.V., *et al.* (2006) L1 integration in a transgenic mouse model. *Genome Res* 16,
9 240-250
10
11 118 Kubo, S., *et al.* (2006) L1 retrotransposition in nondividing and primary human somatic
12 cells. *Proc Natl Acad Sci U S A* 103, 8036-8041
13
14
15 119 Lodato, M.A., *et al.* (2015) Somatic mutation in single human neurons tracks developmental
16 and transcriptional history. *Science* 350, 94-98
17
18 120 Chen, C., *et al.* (2017) Single-cell whole-genome analyses by Linear Amplification via
19 Transposon Insertion (LIANTI). *Science* 356, 189-194
20
21 121 Evrony, G.D., *et al.* (2016) Resolving rates of mutation in the brain using single-neuron
22 genomics. *Elife* 5, e12966
23
24
25 122 Quail, M.A., *et al.* (2012) A tale of three next generation sequencing platforms: comparison
26 of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, 341
27
28 123 De Cecco, M., *et al.* (2013) Genomes of replicatively senescent cells undergo global
29 epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell*
30 12, 247-256
31
32
33 124 Krug, L., *et al.* (2017) Retrotransposon activation contributes to neurodegeneration in a
34 *Drosophila* TDP-43 model of ALS. *PLoS Genet* 13, e1006635
35
36
37 125 Li, W., *et al.* (2013) Activation of transposable elements during aging and neuronal decline
38 in *Drosophila*. *Nat Neurosci* 16, 529-531
39
40 126 Van Meter, M., *et al.* (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1
41 but this repression fails with stress and age. *Nat Commun* 5, 5011
42
43 127 Kobayashi, K., *et al.* (2016) Single-Cell Memory Regulates a Neural Circuit for Sensory
44 Behavior. *Cell Rep* 14, 11-21
45
46
47 128 Muotri, A.R., *et al.* (2009) Environmental influence on L1 retrotransposons in the adult
48 hippocampus. *Hippocampus* 19, 1002-1007
49
50
51 129 Singer, T., *et al.* (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal
52 genomes? *Trends Neurosci* 33, 345-354
53
54 130 Yu, F., *et al.* (2001) Methyl-CpG-binding protein 2 represses LINE-1 expression and
55 retrotransposition but not Alu transcription. *Nucleic Acids Res* 29, 4493-4501
56
57
58 131 Bundo, M., *et al.* (2014) Increased L1 retrotransposition in the neuronal genome in
59 schizophrenia. *Neuron* 81, 306-313
60
61
62
63
64
65

- 1
2
3
4 132 Coufal, N.G., *et al.* (2011) Ataxia telangiectasia mutated (ATM) modulates long
5 interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci U*
6 *S A* 108, 20382-20387
7
8
9 133 Guy, J., *et al.* (2007) Reversal of neurological defects in a mouse model of Rett syndrome.
10 *Science* 315, 1143-1147
11
12 134 Macfarlane, C.M., *et al.* (2013) Transduction-specific ATLAS reveals a cohort of highly
13 active L1 retrotransposons in human populations. *Hum Mutat* 34, 974-985
14
15 135 Farley, A.H., *et al.* (2004) More active human L1 retrotransposons produce longer
16 insertions. *Nucleic Acids Res* 32, 502-510
17
18 136 Herculano-Houzel, S. and Lent, R. (2005) Isotropic fractionator: a simple, rapid method for
19 the quantification of total cell and neuron numbers in the brain. *J Neurosci* 25, 2518-2521
20
21 137 Eickbush, M.T. and Eickbush, T.H. (2011) Retrotransposition of R2 elements in somatic
22 nuclei during the early development of *Drosophila*. *Mob DNA* 2, 11
23
24 138 Perrat, P.N., *et al.* (2013) Transposition-driven genomic heterogeneity in the *Drosophila*
25 brain. *Science* 340, 91-95
26
27 139 Boeke, J.D. (2003) The unusual phylogenetic distribution of retrotransposons: a hypothesis.
28 *Genome Res* 13, 1975-1983
29
30 140 Gilbert, N., *et al.* (2005) Multiple fates of L1 retrotransposition intermediates in cultured
31 human cells. *Mol Cell Biol* 25, 7780-7795
32
33 141 Morrish, T.A., *et al.* (2002) DNA repair mediated by endonuclease-independent LINE-1
34 retrotransposition. *Nat Genet* 31, 159-165
35
36 142 Gilbert, N., *et al.* (2002) Genomic Deletions Created upon LINE-1 Retrotransposition. *Cell*
37 110, 315-325
38
39 143 Garcia-Perez, J.L., *et al.* (2007) Distinct mechanisms for trans-mediated mobilization of
40 cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res* 17, 602-611
41
42 144 Kajikawa, M. and Okada, N. (2002) LINES mobilize SINEs in the eel through a shared 3'
43 sequence. *Cell* 111, 433-444
44
45 145 Sugano, T., *et al.* (2006) Isolation and characterization of retrotransposition-competent
46 LINES from zebrafish. *Gene* 365, 74-82
47
48 146 Graham, V., *et al.* (2003) SOX2 functions to maintain neural progenitor identity. *Neuron* 39,
49 749-765
50
51 147 Tchenio, T., *et al.* (2000) Members of the SRY family regulate the human LINE
52 retrotransposons. *Nucleic Acids Res* 28, 411-415
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

148 Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* 10, 6718-6729

149 Athanikar, J.N., *et al.* (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* 32, 3846-3855

150 Yang, N., *et al.* (2003) An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res* 31, 4929-4940

1
2
3
4 **Text Boxes**
5

6 **Box 1. Target-primed reverse transcription (TPRT).**
7

8
9 This seminal mechanistic model was proposed by the Eickbush laboratory [33], based on
10 experimental data obtained from the silk moth R2 LINE-like retrotransposon, which provided a
11 tractable system as it preferentially inserts into 28S rDNA genes [33,137]. Briefly, for murine
12 and primate L1s, TPRT involves the transcription and translation of a full-length, capped and
13 polyadenylated L1 mRNA [21,139] followed by the association *in cis* of the L1 mRNA, ORF1p
14 and ORF2p to generate a cytoplasmic ribonucleoprotein particle (RNP) [10]. The L1 RNP can
15 next access the nucleus [118], where the endonuclease activity of ORF2p [11] cleaves one
16 genomic DNA strand at a degenerate 5'-TTTT/AA site [34] and then the ORF2p reverse
17 transcriptase [12] initiates reverse transcription from the exposed 3' hydroxyl group using the
18 accompanying L1 mRNA as a template, generating a new L1 copy primed from the cleavage site
19 [40]. After this process, the second DNA strand is also cleaved, presumably by the same ORF2p
20 endonuclease activity, near to the first cleavage site which, after the nascent L1 insertion is
21 resolved by DNA synthesis, usually leads to the formation of target-site duplications (TSDs)
22 flanking the newly synthesized DNA. Retrotransposition can also occur through variations of the
23 fundamental TPRT model [140-142] including occasional *trans* mobilization of mutant L1
24 mRNAs that do not encode intact ORFs [10,143]. The TPRT mechanism is likely to be
25 conserved in all vertebrates as, for example, eel and zebrafish LINEs retrotranspose in human
26 cells [144,145].
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53
54
55 **Box 2. Engineered L1 mobilization in neural progenitor cells.**
56

57 A 2005 study [56] discovered *in vivo* L1-EGFP mobilization in transgenic mouse neurons, and in
58
59
60
61
62
63
64
65

1
2
3
4 cultured rat neural stem cells (NSCs) and neuronal precursor cells (NPCs), providing
5
6 foundational evidence of an L1 mosaic mammalian brain. Amongst various key findings, the
7
8 authors elucidated Sox2, a transcription factor required to maintain NSC identity [146], as a
9
10 repressor of the L1 5'UTR that is downregulated to complete neuronal maturation, hence
11
12 providing a scenario for L1 mobilization [56,147]. A significant caveat of this work was that it
13
14 depended on a human L1 tagged with EGFP and integrated into the foreign epigenetic landscape
15
16 of another species. As a follow on, a subsequent study [53] showed that the L1-EGFP reporter
17
18 mobilized in human NPCs derived from fetal brain and hESCs *in vitro*, and that the CpG island
19
20 at the core of the L1-Ta promoter [148] was partially demethylated in fetal brain when compared
21
22 to non-neural tissues, further explaining L1 activation during neurogenesis. Through an L1
23
24 qPCR-based copy number variation (CNV) assay, they determined that more L1 copies were
25
26 found in brain tissues than non-brain tissues [53]. Interestingly, both studies observed that neural
27
28 cells known to carry retrotransposed L1-EGFP copies could be negative for EGFP expression
29
30 [53,56]. An epigenetic mechanism for transcriptional silencing of integrated L1-EGFP copies
31
32 was discovered in a follow up paper [84], suggesting that the rates of *in vitro* and *in vivo*
33
34 engineered L1 retrotransposition observed by the earlier works were likely conservative.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 **Figure Legends**
5

6 **Figure 1. Mammalian retrotransposons.** (a) Mobile human retrotransposon families. L1, Long
7 INterspersed Element 1; *Alu*, a family of Short INterspersed Element (SINE); SVA, a composite
8 of SINE-R, Variable number of tandem repeats (VNTR), *Alu*, and 5' hexamer sequences; EN,
9 endonuclease; RT, reverse transcriptase. Pol II and Pol III promoters are represented by solid and
10 empty arrows, respectively. (b) As for a), except detailing mouse L1, SINE B1, SINE B2 and
11 IAP (intracisternal A particle) endogenous retrovirus (ERV) families. LTR, long terminal repeat;
12 GAG, group-specific antigen; POL, polymerase; ENV, envelope; IN, integrase; RH, RNase H.
13 (c) Mechanism of target-primed reverse transcription (TPRT). First and second strand cleavage
14 positions are depicted by red and green arrowheads, respectively. (d) Factors activating and
15 repressing the human L1 CpG island-centric 5'UTR promoter. CpG dinucleotides, including
16 those assayed by two studies [53,54], are represented with vertical orange strokes. Validated
17 transcription factor binding sites are represented by horizontal red lines [55,56,130,149,150].
18 Activator and repressor TFs are represented above and below the diagram, respectively. HDAC,
19 histone deacetylase.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 **Figure 2. Methods to identify engineered and endogenous L1 insertions.** (a) Schematic of an
44 L1 reporter system. Retrotransposition from an exogenous construct carrying an L1 tagged with
45 a spliced fluorescent reporter (e.g. EGFP [65]) or antibiotic resistance (e.g. neomycin [13])
46 activates the cassette, enabling downstream analysis of L1 retrotransposition efficiency. (b)
47 Targeted sequencing approaches to map an endogenous L1 insertion. Genomic DNA can be
48 enriched for L1-genome junctions via sequence capture [80], PCR or adaptor ligation [30,50,79],
49 sequenced and computationally analyzed to reveal the *de novo* L1 variant.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7 **Figure 3. Key Figure: Context-specific donor L1 activity.** Schematic representation of donor
8
9 L1s and their distinct impacts on germline and somatic mosaicism, in two individuals. Each
10
11 donor L1 locus (numbered from 1 to 13) can be empty (black), or contain either a
12
13 retrotransposition-competent (red color) or incompetent (grey) L1. Donor L1s can be
14
15 heterozygous or homozygous. Locus-specific L1 activation can be restricted by tissue,
16
17 developmental stage or cell type (E, embryo; O, oncological processes; Gl, germline; B, brain;
18
19 Lv, liver; Co, colon), or can be unrestricted (asterisk). In the anatomic panels of the respective
20
21 individuals, colors and numbers represent the potential contexts with somatic L1 variants and
22
23 their matched donor L1s.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Tables

Table 1. Summary of endogenous L1 mobilization rates in mammalian neurons.

Study	Species	Tissues	Purified neurons?	Amplification strategy ^a	Core L1 analysis method ^b	Estimated somatic L1 insertions per neuron	PCR validation details ^c	Notes	Reference
Coufal et al. (2009)	Human	Hippocampus, cerebellum	No	None (bulk)	L1 qPCR	80	N/A	Rate normalized to plasmid spike in	[53]
Baillie et al. (2011)	Human	Hippocampus, caudate nucleus	No	None (bulk)	RC-seq	0.04	Junction specific	Very approximate <i>post hoc</i> rate estimate [39]	[80]
Evrony et al. (2012)	Human	Cortex, caudate nucleus	Yes	MDA	L1-IP	0.04	Empty/filled	Validated 1 somatic L1 insertion carrying a 5' transduction with empty/filled PCR, and 4 additional L1 insertions via junction specific PCR	[82]
Evrony et al. (2015)	Human	Cortex	Yes	MDA	WGS	0.32	Empty/filled	Found somatic L1 insertion flanked by a 3' transduction	[83]
Upton et al. (2015)	Human	Hippocampus, cortex	Yes	MALBAC	RC-seq	13.7	Junction specific	Amplification method unsuitable for empty/filled PCR validation	[81]
Hazen et al. (2016)	Mouse	Olfactory bulb	Yes	SCNT	WGS	1.3	Junction specific	L1 insertion sequences and families not provided	[116]
Erwin et al. (2016)	Human	Hippocampus, cortex	Yes	MDA	SLAV-seq	0.58-1	Empty/filled	Also identified putative somatic L1-associated deletions	[41]

^aMDA, multiple displacement amplification; MALBAC, multiple annealing and looping-based amplification cycles; SCNT, somatic cell nuclear transfer.

^bRC-seq, retrotransposon capture sequencing; L1-IP, L1Hs insertion profiling; WGS, whole genome sequencing; SLAV-seq, somatic L1-associated variant sequencing.

^cN/A, not applicable; junction specific, PCR targeting a 5' or 3' L1-genome junction; empty/filled, PCR targeting the complete L1 insertion via amplification using primers positioned on either flank of the L1 insertion, followed by capillary sequencing (the gold standard approach).

Trends Box

- L1 retrotransposons can mobilize during embryogenesis, and in the neuronal lineage, causing somatic genome mosaicism.
- Genomic analysis of endogenous L1 mobilization in mouse pedigrees, and transgenic L1 rodents, has revealed the early embryo, prior to germ cell specification, as the primary niche for the accumulation of new, heritable L1 insertions.
- Neuronal progenitor cells and post-mitotic neurons accommodate engineered L1 retrotransposition, but other cell lineages support limited or no L1 activity, in the physiological conditions tested to date.
- L1 retrotransposition clearly occurs in the brain, based on data obtained from engineered L1 reporter systems and single-cell genomic analysis, but the relevant techniques and estimated L1 mobilization rates vary considerably.
- Donor L1s can be differentially active in germline and somatic cells, potentially influencing evolutionary selection of donor L1s that are highly active in the brain.

Outstanding Questions Box

- Human pluripotent stem cells obtained via reprogramming or from embryonic material consistently support L1 retrotransposition, as do mouse embryonic stem cells. However, it is unclear as to why *de novo* L1 insertions apparently do not occur in miPSCs. Is this a technical issue? Or a result of mouse L1s being less amenable to jumping during reprogramming than human L1s (despite seeming more active in the early embryo)? Moreover, although endogenous L1 retrotransposition is now well demonstrated in the mouse embryo, it is less defined in terms of spatial extent and frequency in early human embryogenesis.
- What is the frequency of endogenous L1 mobilization in the brain? It is accepted that L1 can jump in the brain, however the available rate estimates, and interpretations of the same data, vary widely. A focus on false positives should be complemented by a closer examination of false negatives, and standardization of techniques. L1 insertions are likely to occur in post-mitotic neurons, meaning even a low rate of neuronal L1 mobilization could generate a constellation of L1 variation amongst the $\sim 10^{11}$ neurons found in the human brain. Is this mosaicism, however, variable among different neuronal subtypes?
- What are the immediate and broader functional consequences of somatic L1 insertions in the brain? Transcriptomic and genomic analysis of the same individual neuron could, at least, answer the first question. The impact of L1 mosaicism on neurobiology is a much more challenging and large-scale issue, with little clear evidence produced to date of somatic L1 insertions impacting neurological function, psychiatric disorders or neurodegenerative diseases.
- If, however, L1 mosaicism impacts neurobiology, it is plausible that donor L1s highly

active in the neuronal lineage may undergo evolutionary selection despite their offspring somatic L1 insertions not being heritable. The available experimental data suggest that some donor L1s are unusually active in cancer genomes. The same may be true of donor L1s in normal somatic cells, including neurons. Hence, the donor L1 cohort of individuals, and their haplotypes, may define the level of L1 activity in the embryo and brain.

Figure 1

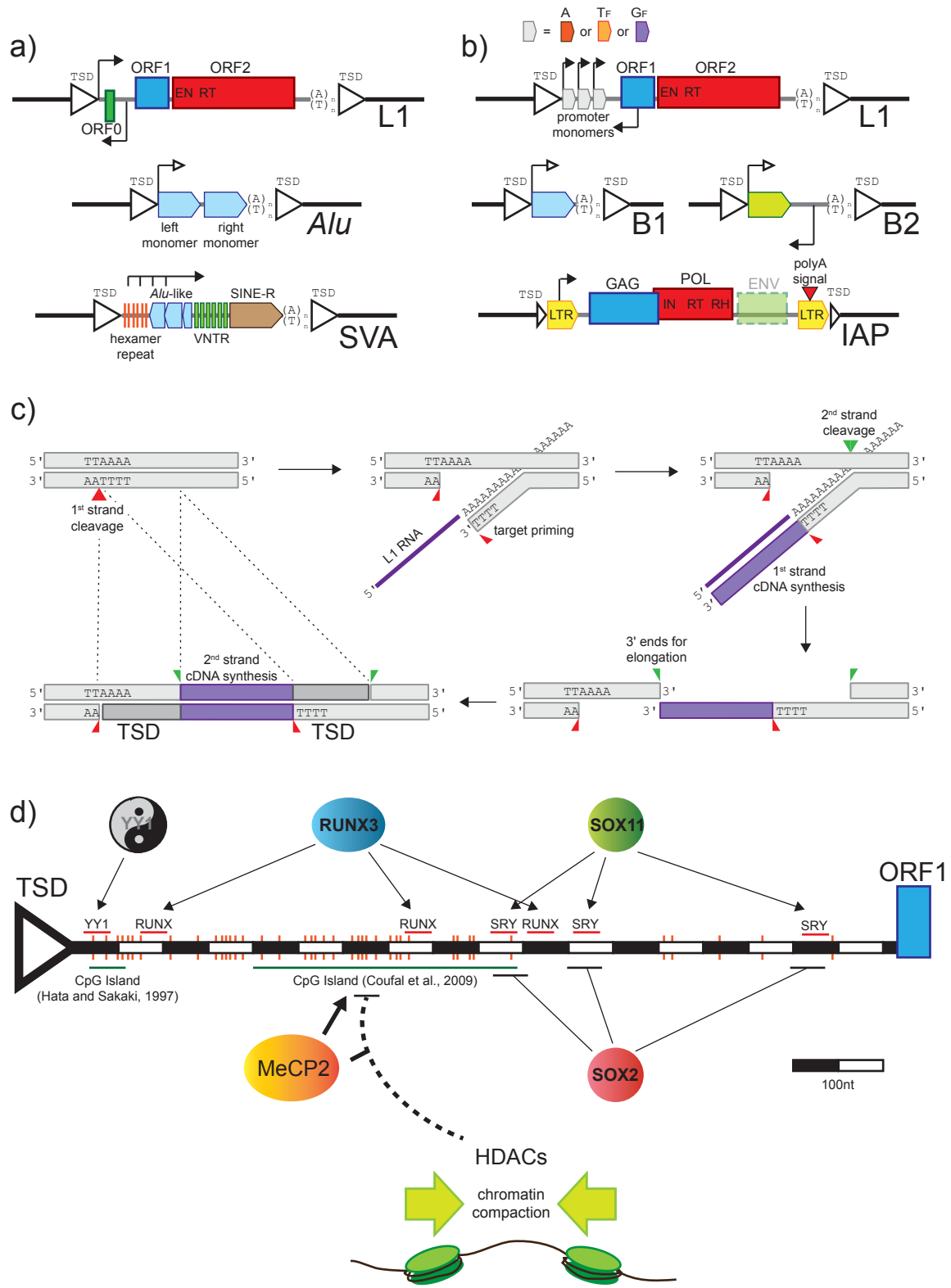


Figure 2

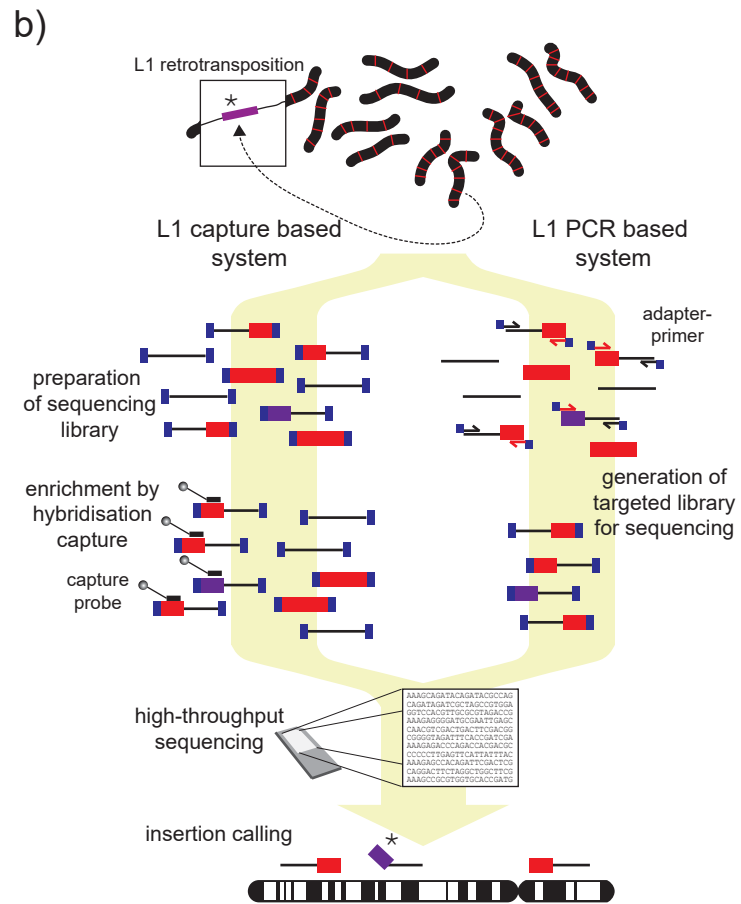
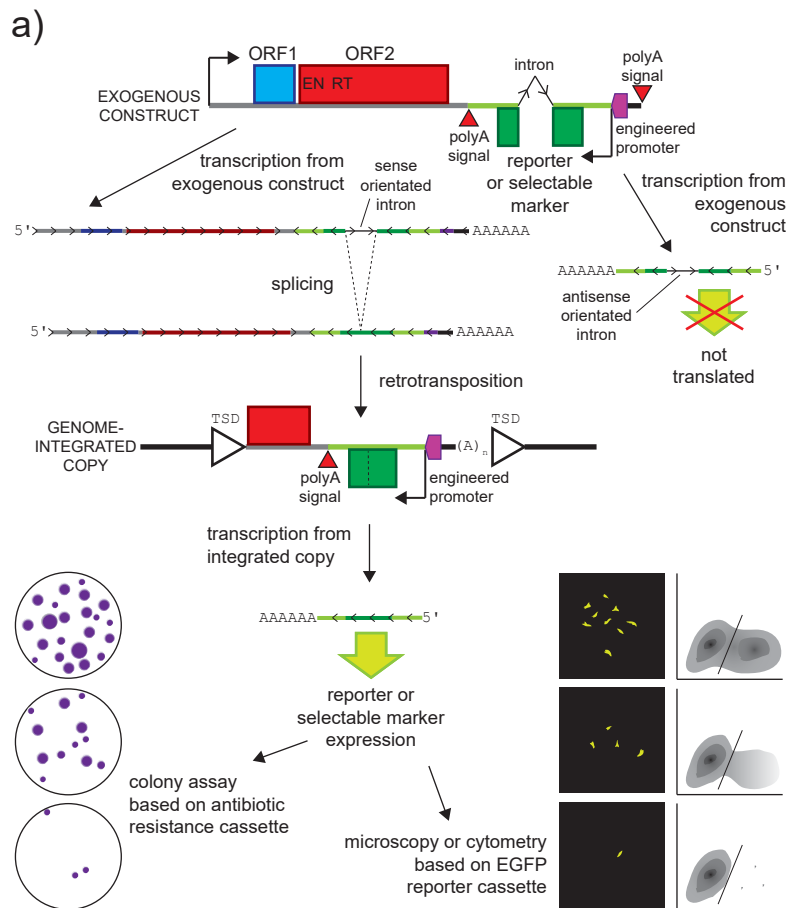


Figure 3

[Click here to download Key Figure Figure 3.ai](#)

