Heritable L1 Retrotransposition Events During Development: Understanding Their Origins

Examination of heritable, endogenous L1 retrotransposition in mice opens up exciting new questions and research directions

Sandra R. Richardson and *Geoffrey J. Faulkner**

The retrotransposon Long Interspersed Element 1 (LINE-1 or L1) has played a major role in shaping the sequence composition of the mammalian genome. In our recent publication, "Heritable L1 retrotransposition in the mouse primordial germline and early embryo," we systematically assessed the rate and developmental timing of de novo, heritable endogenous L1 insertions in mice. Such heritable retrotransposition events allow L1 to exert an ongoing influence upon genome evolution. Here, we place our findings in the context of earlier studies, and highlight how our results corroborate, and depart from, previous research based on human patient samples and transgenic mouse models harboring engineered L1 reporter genes. In parallel, we outline outstanding questions regarding the stage-specificity, regulation, and functional impact of embryonic and germline L1 retrotransposition, and propose avenues for future research in this field.

1. Introduction

1.1. L1 in Mammalian Genomes

L1 retrotransposon sequences account for approximately 17% of human genomic DNA and approximately 18% of mouse genomic DNA.[1,2] While the vast majority of L1 copies are

Dr. S. R. Richardson, Prof. G. J. Faulkner
Mater Research Institute–University of Queensland
Woolloongabba, Queensland 4102, Australia
E-mail: sandra.richardson@mater.uq.edu.au; faulknergj@gmail.com

Prof. G. J. Faulkner
Queensland Brain Institute
University of Queensland
Brisbane, Queensland 4072, Australia

Prof. G. J. Faulkner
School of Biomedical Sciences
University of Queensland
Brisbane, Queensland 4072, Australia

molecular fossils rendered incapable of further retrotransposition by 5′ truncation, internal mutations and structural rearrangements, a subset of L1s retain the ability to mobilize. A typical human genome harbors around 100 of these mobile L1s,[3] while mice are estimated to contain approximately 3000 such elements.[4] A retrotransposition-competent L1 is about 6 kb (human) or 7 kb (mouse) in length,[5,6] and contains a 5′ untranslated region (5′UTR) harboring an internal promoter,[7] two intact open reading frames (ORF1 and ORF2), a 3′ untranslated region (3′UTR), and terminates in a poly(A) tract.[5] In mice, the 5′UTR is variable in length and comprises ≈200 bp repetitive units, or monomers.[8] The sequence of these monomer units is the primary distinguishing characteristic between three presently active L1 subfamilies in the mouse genome: L1 T_F, G_F, and A.[4,9,10]

A round of L1 retrotransposition initiates with transcription of a retrotransposition-competent donor, or source, L1 in the genome, translation of the L1-encoded proteins ORF1p and ORF2p, and association of ORF1p and ORF2p with their encoding L1 mRNA in a phenomenon known as cis-preference.[11] The resultant L1 ribonucleoprotein particle (RNP)[12,13] enters the nucleus, where the L1 ORF2p endonuclease activity initiates the formation of a new L1 insertion by catalyzing a single-strand nick in genomic DNA at the loose consensus sequence 5′-TTTT/AA-3′. This activity liberates a free 3′ hydroxyl residue that serves as a primer from which the L1 ORF2p reverse transcriptase activity initiates synthesis of the first-strand L1 cDNA on the L1 mRNA template, with this process referred to as target-primed reverse transcription (TPRT).[14–17] Second-strand genomic DNA cleavage typically occurs downstream of first-strand cleavage, resulting in variable length target-site duplications (TSDs) flanking the newly-integrated L1 insertion.[18–21]

Although cis-preference is the general rule for L1 retrotransposition, the L1-encoded proteins can occasionally act in trans to mobilize non-autonomous retrotransposons known as short interspersed elements (SINEs). In humans, active SINEs include

the 7SL-derived element *Alu* and the SINE-R-VNTR-*Alu* (SVA) composite retrotransposon.[22–24] In mice, active SINEs include the 7SL-derived B1 element and the tRNA-derived B2 element.[25] Cellular mRNAs can also be mobilized in trans by the L1 protein machinery, generating processed pseudogenes.[11,26,27] A uniting feature of the L1 mRNA and other RNAs that can be trans-mobilized by L1 is a 3′ poly(A) tract and, indeed, recent work has definitively shown that a 3′ poly(A) tract is required for L1 retrotransposition.[28]

During retrotransposition, structural changes can occur in the L1 sequence, and in the proximal target site genomic DNA. Nascent L1 insertions often undergo 5′ truncation, as well as internal inversion/deletion events, rendering the new L1 copy incapable of subsequent retrotransposition.[1,21,29,30] Alterations to target site DNA include deletions of genomic sequence and, occasionally, severe aberrations such as chromosomal trans-locations.[20,31] During transcription of a retrotransposition-competent L1, read-through of the L1 polyadenylation signal in favor of a downstream polyadenylation signal can ultimately result in the retrotransposition of flanking genomic sequence along with the nascent L1 insertion, in a process known as 3′ transduction.[32–34] Less commonly, 5′ transductions arise when transcription of a donor L1 initiates from an upstream genomic promoter rather than the L1 internal promoter, and 5′ flanking genomic sequence is incorporated into the new insertion.[1] L1 insertions into exons of genes can ablate gene function,[21,35,36] and intronic insertions are associated with mis-splicing, premature polyadenylation, disruption of regulatory elements, provision of new regulatory elements, and reduced RNA polymerase processivity leading to subtle changes in gene expression.[37–41]

L1-mediated retrotransposition events have been found to be responsible for 130 cases of human genetic disease[42] and six spontaneous mouse mutants have been attributed to de novo L1 insertions.[43,44] L1 mobilization clearly can have a negative impact on genome integrity and, indeed, L1 activity is curtailed by a variety of host factors (for reviews, see Refs. [45,46]). New L1 insertions also provide a continuing supply of genetic diversity. Based upon the frequency of polymorphic retrotransposon insertions identified in the human population, one new L1 insertion is estimated to arise per 100–200 live births, and one new *Alu* insertion per 20 live births.[47–52] The rate of new L1 insertions arising in mice was until recently estimated to be 30-fold higher, at one insertion per 2–3 live births,[53,54] due provisionally to the much larger number of potentially active L1 copies in the mouse genome.[4]

2. Examining L1 Retrotransposition In Vivo

2.1. Transgenic L1 Reporter Systems Provide Critical Insights

Cultured cell assays[55,56] for L1 retrotransposition have elucidated the mechanism by which L1 mobilizes, as well as potential consequences for the host genome (reviewed in Ref. [57]). In these systems, a retrotransposition-competent L1 is tagged in its 3′UTR with a selectable (e.g., neomycin phosphotransferase[55]) or screenable (e.g., enhanced green fluorescent protein (EGFP)[58]) reporter gene, complete with a heterologous promoter and polyadenylation signal, that will only be expressed if the tagged L1 undergoes a successful round of retrotransposition (**Figure 1**A). Engineered L1s have also been incorporated into transgenic rodent models to study the dynamics of L1 retrotransposition in vivo[59–66] (**Table 1**). These systems have considerably informed our understanding of L1 activity in the germline and early embryo. The advantages of transgenic models include highly sensitive detection, quantitation, and mapping of integrated L1 retrotransposition events (Figure 1B), readily distinguished from existing genomic copies by the presence of the spliced reporter cassette. Expression of the EGFP reporter gene can also facilitate detection and identification of cells harboring engineered retrotransposition events by flow cytometry and fluorescence microscopy (Figure 1B), although silencing of the integrated reporter cassette in some physiological contexts may hinder detection.[67,68]

Other considerations regarding transgenic L1 animal models include the genomic location of the transgene, as randomly integrated reporters may not faithfully recapitulate the activity of endogenous L1s. The choice of L1 element in a transgenic animal model is also important: at least in cultured cells, highly-active human L1s retrotranspose more efficiently than the mouse L1s that have been studied to date,[4,10,69,70] making them an appealing choice to observe high-frequency L1 mobilization in vivo. However, introducing a human L1 into a rodent genome raises the question of whether it will be regulated similarly to endogenous rodent L1s in vivo. Heterologous promoters have been used in addition to the endogenous L1 promoter,[59,60,62] or in some cases in its place,[64,71] to drive high levels of L1 transgene expression which may not recapitulate endogenous L1 regulation.

A synthetic, codon-optimized mouse L1, termed ORFeus_Mm, has provided an intriguing solution to these problems.[72] The amino acid sequence of ORFeus_Mm is that of a native mouse L1, but the nucleotide sequence has been altered for optimal codon usage in mammalian cells, resulting in a substantial increase in GC content. ORFeus_Mm is highly expressed compared to native mouse L1s and retrotransposes extremely well in cultured cells.[72] This element has been used with heterologous promoters[63,64] as well as a native mouse L1 promoter[66] to facilitate efficient L1 retrotransposition in vivo. Whether the synthetic nucleotide sequences of the ORFeus_Mm ORFs affect its regulation in vivo remains to be determined.

2.2. High-Throughput Sequencing Allows Detection of Endogenous Retrotransposition Events

Endogenous L1 activity in mammalian genomes can be described in unprecedented resolution by high-throughput DNA sequencing approaches. Such methods typically exploit discordant paired-end sequencing reads, or split-reads that span L1-genome junctions, to identify L1 sequences at loci where an L1 is absent from the reference genome sequence (as reviewed in Refs. [73,74]). L1 insertions can be identified through analysis of
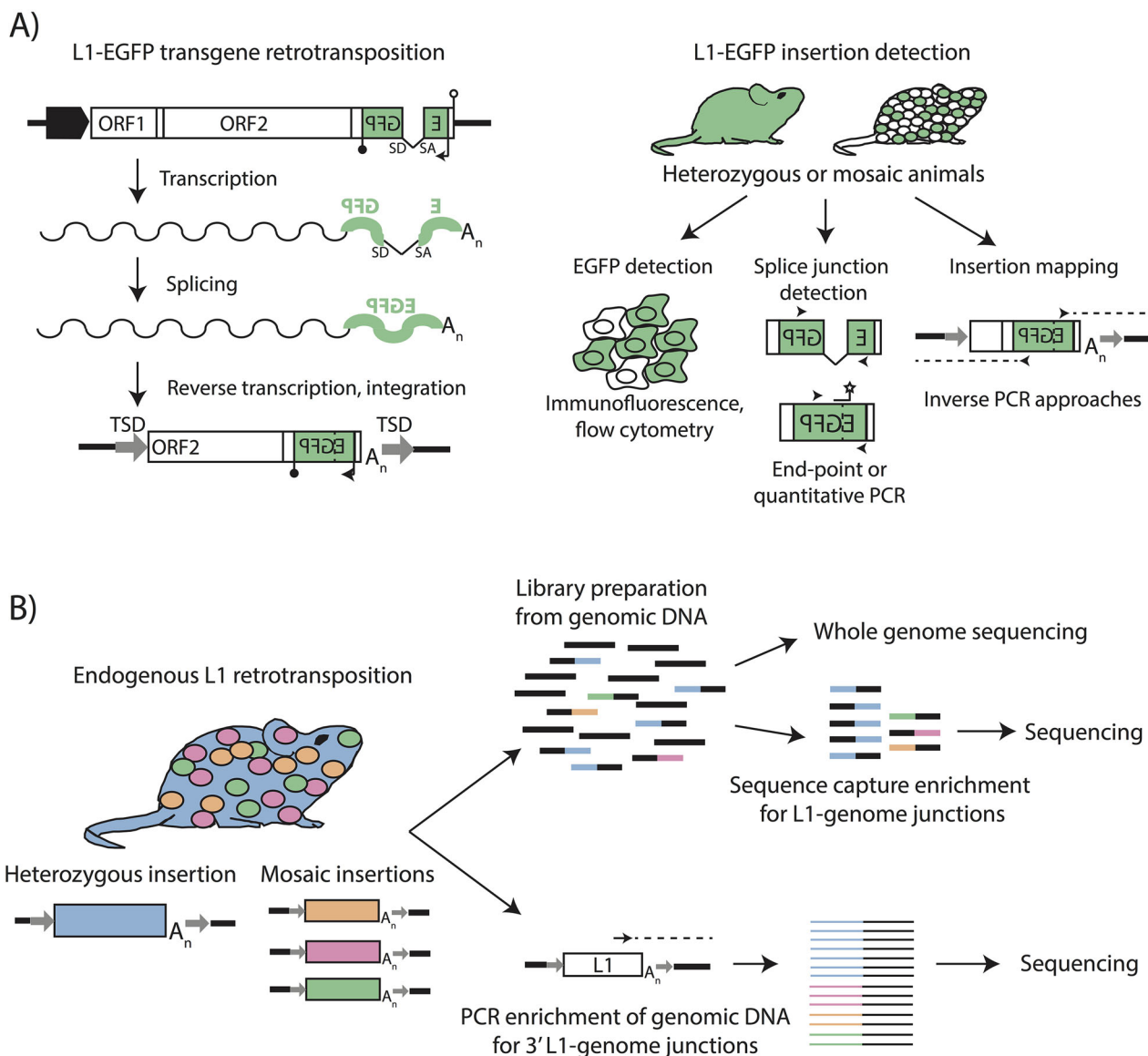
Figure 1. Methods for detecting engineered and endogenous L1 insertions. A) L1-EGFP transgene retrotransposition. An engineered L1-EGFP transgene consists of a retrotransposition-competent L1 element with intact ORFs (white boxes) driven by the native L1 promoter, a heterologous promoter, or both (broad black arrow). The L1 3′UTR contains an enhanced green fluorescent protein (EGFP) reporter gene (green) in antisense orientation to the L1, interrupted by an intron with splice donor (SD) and splice acceptor (SA) sites in sense orientation to the L1. The EGFP reporter is driven by a heterologous promoter (thin black arrow) and terminates in a polyadenylation signal (black lollipop). The EGFP reporter is followed by a polyadenylation signal in sense orientation to the L1 (white lollipop). Upon L1 mRNA transcription (wavy line), the intron in the EGFP gene is spliced. Reverse transcription and integration of the L1 mRNA during target-primed reverse transcription delivers an intact copy of the EGFP reporter into the genome. B) L1-EGFP insertion detection. Animals harboring engineered L1-EGFP insertions can be heterozygous for an insertion (filled green mouse) or mosaic for an insertion (white and green mouse). L1-EGFP insertions can be detected through expression of the EGFP gene (flow cytometry, fluorescence microscopy), or by PCR-mediated detection of the spliced EGFP cassette. The genomic locations and structural characteristics of L1-EGFP insertions can be identified using inverse PCR strategies with primers specific to the EGFP cassette. C) Detection of endogenous L1 retrotransposition. At left, an insertion for which a mouse is consummately heterozygous is represented in blue, and mosaic insertions present at <1 copy per cell are represented in orange, pink, and green. Strategies for detecting endogenous L1 insertions include preparation of Illumina sequencing libraries from genomic DNA (above), which can be directly used for whole genome sequencing (WGS) or enriched for L1-genome junctions using sequence capture probes, as in RC-seq. Enrichment for 3′ L1-genome junctions can also be achieved using PCR-based approaches (below) in which outward-facing primers are used to selectively amplify the 3′ termini of active L1 subfamilies, and amplicons subjected to high-throughput sequencing. In both cases, heterozygous insertions are better represented than mosaic insertions when genomic DNA from bulk tissues is assayed.

Table 1. Transgenic mouse models used to study heritable L1 retrotransposition events.

Study	L1 element used	Promoter driving L1 transgene	Reporter cassette	Transgene copy number	Developmental origin of identified insertions	Insertions transmitted?
Ostertag et al.[59]	Human L1$_{RP}$	Native L1$_{RP}$ 5′UTR	pAC-EGFP	n/d	Male germline	n/d
		pPolII + native L1$_{RP}$ 5′UTR	pAC-EGFP	n/d	Male germline	Yes
Prak et al.[60]	Human L1$_{RP}$	pPolII + native L1$_{RP}$ 5′UTR	pCMV-EGFP	n/d	Early embryonic in founder mouse	Yes
Babushok et al.[62]	Human LRE3	pHSP70-2 + native LRE3 5′UTR	Markerless cassette	n/d	Early embryonic	No
An et al.[63]	Mouse synthetic ORFeus_Mm	pCAG	pRSV LTR-EGFP	310[64]	n/d	Yes
An et al.[64]	Mouse synthetic ORFeus_Mm	pCAG; conditionally activated by Cre-mediated recombination	pRSV LTR-EGFP	Single-copy	n/d	n/d
Kano et al.[65]	Human LRE3 flanked by chicken β-globin insulators	Native LRE3 5′UTR	γ-globin intron in 3′UTR	n/d	Early embryonic	No
	Human L1$_{RP}$[61]	Native L1$_{RP}$ 5′UTR	pCMV-EGFP	n/d		
	Human L1$_{RP}$ (tested in rat)	Native L1$_{RP}$ 5′UTR	Markerless cassette	n/d		
	Mouse TG$_F$21	Native TG$_F$21 5′UTR	pCMV-EGFP	n/d		
Newkirk et al.[66]	Mouse synthetic ORFeus_Mm	Native L1$_{spa}$ 5′UTR	pCMV-EGFP	Single-copy	Male germline	n/d

From left to right, columns indicate the study cited, the species of origin and name of the L1 element, the promoter(s) driving L1 transgene expression, the promoter driving reporter cassette expression, the developmental origins of insertions examined by each study, and whether germline transmission of insertions to subsequent generations was observed. pPolII, mouse RNA polymerase II promoter; pHSP70-2, mouse heat shock protein 70-2 promoter; pAC, preproacrosin promoter; pCMV, human cytomegalovirus immediate-early enhancer and promoter; pCAG, CMV immediate-early enhancer/modified chicken β-actin promoter; pRSV LTR, Rous sarcoma virus LTR promoter. n/d; not determined. In Kano et al.,[65] the L1$_{RP}$ element was used in both transgenic mouse and transgenic rat models; the L1$_{RP}$ transgenic mice were first reported by Muotri et al.[61]

whole genome sequencing (WGS) data,[48,75–80] and targeted enrichment for L1 insertions can be achieved by sequence capture[81] or PCR-based methods in which L1-genome junctions are selectively amplified using outward-facing primers specific to an active L1 family (Figure 1C).[47,82–85] In each case, putative L1 insertions must be validated by PCR and capillary sequencing to exclude chimeric molecules generated during sample preparation and high-throughput sequencing. One obvious advantage of high-throughput sequencing strategies is that they allow detection of endogenous L1 insertions, eliminating caveats posed by the somewhat artificial nature of transgenic systems. By contrast, per observed L1 integration event, high-throughput sequencing is far more expensive at present than tracking engineered L1 insertions. In addition, sequencing-based approaches more readily detect insertions that are present at a high prevalence in a given sample. When bulk tissue samples are used, insertions that occurred in cells that have undergone extensive clonal expansion have an advantage over insertions that occurred in cells with fewer mitotic descendants, or insertions that are unique to a single cell. Whole genome amplification (WGA) of single cells can facilitate detection of such low prevalence insertions,[86,87] but the WGA process represents an additional source of chimeric molecules that can introduce false positive results,[88] making thorough PCR validation especially crucial. Together, transgenic and high-throughput sequencing based approaches provide an often consistent window into L1 biology in vivo, and both are valuable

tools for understanding the dynamics of L1 retrotransposition in vivo.

3. Heritable L1 Retrotransposition Is an Ongoing Process

3.1. How Frequently Do Heritable L1 Insertions Occur in Mice?

The contribution of retrotransposon activity to mouse genetic diversity has been made clear by comparing the genomic sequences of commonly-used inbred mouse strains. In a 2008 study, Akagi et al. identified structural variants (SVs) present in the C57BL/6J reference genome but absent from WGS reads generated from four additional inbred mouse strains (A/J, DBA/2J, 129S1/SvImJ, and 129 × 1/SvJ).[75,89] L1 polymorphisms indicative of relatively recent retrotransposition events accounted for the majority of these variants, with 6723 putative L1 insertions present in the C57BL/6J reference genome and absent from at least one other inbred strain. Most polymorphic L1s identified were members of the young, presently active T$_F$, G$_F$, and A subfamilies. Similarly, Quinlan et al. carried out a detailed comparison of SV content in the C57BL/6J and DBA/2J genomes[78] and found that transposable element sequences accounted for approximately 2/3 of the 7196 SVs detected. Of these SVs, 4412 consisted of a single TE annotation, consistent with retrotransposition rather than

another mechanism of structural variation, and L1 sequences were responsible for nearly half of insertions. Subsequent WGS applied to 17 inbred mouse strains confirmed that transposable element sequences accounted for the majority of SVs[76,90] and totaled 103 798 such events, including 40 074 L1 insertions.[77] Together, these studies affirmed L1 activity as a major force that continues to shape murine genomes.

The L1 integrants identified in the above studies were clearly the result of evolutionarily recent L1 mobilization, as evidenced by their differential presence/absence among inbred strains. However, they present only a snapshot of L1-mediated variation that has accumulated during the divergence of inbred strains, rather than a "real-time" account of ongoing retrotransposition in individual genomes. In a recent publication, we studied the genesis of de novo L1 insertions in mice with unprecedented resolution.[91] By analyzing the genomes of individual, related C57BL/6J animals, we directly quantified the frequency of endogenous, de novo L1 retrotransposition across multiple generations of mice, and deduced the developmental timing of many of these events.

Our experimental strategy entailed generating two- and three-generation pedigrees of wild-type C57BL/6J mice and performing mouse retrotransposon capture sequencing (mRC-seq) and WGS (Figure 1C) on parental mice and their offspring. Among 85 mouse genomes analyzed, we identified 11 de novo L1 insertions, providing an estimate of one new L1 insertion per 8 births. Of these 11 insertions, we could explicitly demonstrate that 9 were heritable, indicating a frequency of at least one heritable retrotransposition event per 10 births. We regard this figure as a conservative estimate, as the mouse L1 3′ end is refractory to sequencing on Illumina platforms due to the presence of a GC-rich tract which may form a G-quadruplex structure[92,93] and this technical hurdle may have prevented the detection of 5′ truncated L1 insertions by mRC-seq. All 11 insertions belonged to the L1 T_F subfamily, which is consistent with reports of spontaneous disease-causing L1 insertions in mice,[43,44] although examination of de novo retrotransposition events in different inbred mouse strains may reveal strain-specific activity of G_F and A subfamily elements. Indeed, we have very recently uncovered a tumor-specific L1 G_F insertion in a mouse model of hepatocellular carcinoma on a congenic FVB.129 Sv background.[94]

3.2. What Are the Consequences of Heritable L1 Retrotransposition?

Among 11 de novo mouse L1 insertions uncovered in our study, 10 were intergenic and the 11th intronic insertion had no significant impact upon expression of the disrupted gene, at least in adult tissues from heterozygous animals.[91] In addition, insertions were transmitted from heterozygous animals to their offspring at the expected Mendelian ratios. Thus, the vast majority of de novo L1 insertions in live-born mammals likely produce no immediate phenotype or selective disadvantage, and instead represent new genetic diversity that could ultimately provide the raw material for genome evolution through exaptation of L1 sequences. At the other end of the spectrum, insertions with catastrophic impacts on developmentally essential genes are predicted to result in sterility or embryonic lethality, and hence would not be observed in live-born individuals. The prevalence and impact of such insertions upon mammalian reproductive outcomes remains an open question. Intriguingly, increased L1 expression and reverse transcriptase activity has been implicated in fetal oocyte attrition in mice, although the role of mutagenic L1 integrants in this process remains to be explored.[95] In the future, the contribution of L1 insertions to infertility and embryonic lethality may be resolved, at least in part, by the application of single-cell genomic approaches to identify mutagenic L1 insertions in early embryonic and germ cells.

3.3. L1 Insertions Are an Ongoing Source of New Genetic Diversity

A distinguishing aspect of L1-mediated mutagenesis is that retrotransposition is a replicative process. Thus, following a round of retrotransposition, the original donor L1 remains in its genomic position and retains the capacity to generate more L1 insertions. In addition, when a new L1 insertion is retrotransposition-competent, this daughter element also has the potential to generate retrotransposition events.[69] It follows that, in our study, we were able to determine the complete sequence of nine full-length de novo L1 insertions, and all nine contained intact open reading frames. Two were tested in a cultured cell retrotransposition assay[55,96] and retained the ability to mobilize, highlighting the potential contribution of new L1 insertions to genetic diversity. Furthermore, 3′ transductions carried by three de novo insertions (and two additional insertions that were polymorphic with respect to presence among our mice, indicating relatively recent retrotransposition) allowed the identification of five discrete donor L1s in the mouse reference genome that generated heritable endogenous insertions. As the retrotransposition capacity of a donor L1 depends on its epigenetic regulation in a particular cell, in addition to the enzymatic efficiency of its encoded proteins, it will be interesting to determine whether de novo daughter insertions in vivo are subjected to the same epigenetic and host factor regulation as pre-existing genomic elements, or whether there is some delay in establishing epigenetic marks upon newly-integrated L1 copies. Recent studies employing engineered L1 elements in mouse embryonic stem cells and transgenic mice suggest the former scenario is more likely.[97,98]

Examination of shared 3′ transductions carried by polymorphic L1 insertions in human genomes has indicated that certain donor L1s tend to produce new heritable insertions, and have propagated "hot" L1 lineages.[3,99,100] In addition, examination of 3′ transductions and distinguishing internal mutations borne by tumor specific L1 insertions in humans has likewise revealed particular donor L1s that are recurrently active in the context of cancer.[36,101,102] These findings raise the prospect of specific donor L1s being highly active in mammalian somatic cells.[74,103] Although all five transductions identified in our study were unique, future studies potentially uncovering common transductions in mouse L1s will reveal whether particular donor L1s are presently creating "hot" lineages by being especially active in the early embryo, and thereby recurrently contributing to mouse genomic diversity.
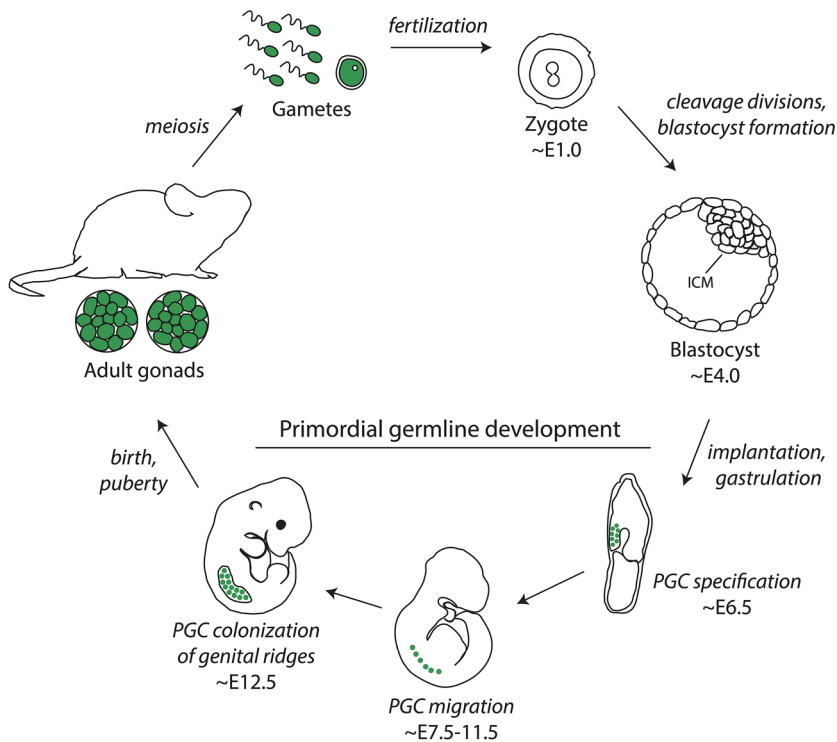
Figure 2. Simplified timeline of mouse embryonic and germline development. Cells of the germ lineage are depicted in green. Clockwise from upper left: fertilization of a mature oocyte by a sperm cell gives rise to the single-cell zygote (embryonic day 1.0 [E1.0], which progresses through cleavage divisions and the morula stage to form the blastocyst [≈E4.0]). The inner cell mass (ICM) of the blastocyst is composed of pluripotent cells that give rise to the embryo. During the process of gastrulation, these pluripotent embryonic cells differentiate to form the three germ layers (endoderm, ectoderm, and mesoderm). At ≈E6.5, a founding population of primordial germ cells (PGCs) are specified, then migrate (≈E7.5–11.5) to colonize the genital ridges by ≈E12.5, followed by formation of the embryonic gonads. Following birth and sexual maturation, completion of meiosis in adult animals generates mature sperm and oocytes.

3.4. At What Points During Development Do Heritable L1 Insertions Occur?

To cause genetic disease and contribute to inter-individual diversity, L1 insertions must occur in cells with the potential to contribute their genetic material to subsequent generations. Previous studies relying on transgenic L1 reporter mice or rare disease-causing insertions in human patients had indicated that heritable L1 insertions can occur in the pluripotent cells of the early embryo, prior to germline specification, and in cells of the germline proper (**Figure 2**).[59,60,63,65,70,104]

3.5. Human Patients and Cell Culture Models Provide Insight Into Embryonic Retrotransposition

That new L1 insertions, and retrotransposition intermediates, can arise in pluripotent embryonic cells was established by back-to-back studies in 2007.[104,105] Van den Hurk et al. traced the causative mutation in a patient with X-linked choroideremia to a full-length L1 insertion in the CHM gene.[104] By genotyping the patient's mother and siblings for the insertion and flanking heterozygous single nucleotide polymorphisms (SNPs), Van den Hurk et al. deduced that the patient's mother was both a somatic and germline mosaic for the insertion, consistent with L1 mobilization during her embryonic development prior to the segregation of the germ lineage from the soma.[104] In the accompanying study, Garcia-Perez et al. utilized L1 reporter constructs to demonstrate engineered L1 retrotransposition in human embryonic stem cells (ESCs), which generated new L1 insertions bearing the structural hallmarks of TPRT.[105] Engineered L1 retrotransposition was subsequently found to occur in mouse ESCs[98] and human induced pluripotent stem cells (iPSCs).[106] Endogenous L1, *Alu*, and SVA retrotransposition has also been identified in human ESCs and iPSCs via retrotransposon capture sequencing (RC-seq).[107] By contrast, evidence for endogenous L1 retrotransposition in mouse iPSCs has been lacking to date[108] and the reasons for this apparent discrepancy are unclear.

3.6. Transgenic Animal Models Exhibit Embryonic Retrotransposition

Analysis of transgenic mouse models likewise supports the case for early embryonic L1 retrotransposition (Table 1). Prak et al. generated a transgenic mouse harboring a human L1-EGFP reporter gene, and identified an engineered insertion arising during early embryonic development of the founder mouse.[60] An et al.[63] also observed heritable engineered insertions in an L1-EGFP transgenic mouse model, although individual insertions were not traced to a discrete developmental timeframe.

Two subsequent studies using transgenic mouse models produced results consistent with retrotransposition occurring post-fertilization. Babushok et al. generated transgenic mice using a human "markerless" reporter cassette, consisting only of an intron rather than an expressed reporter gene.[62] While frequent de novo retrotransposition events were detected in F1 and F2 offspring from a transgenic founder animal, these insertions were not transmitted in Mendelian fashion and were present in <1 copy per cell. Together, these results suggested that the L1 insertions frequently contributed to somatic, but not germline, mosaicism.

In a 2009 study, Kano et al. verified this result more robustly by generating multiple transgenic mouse and rat lines harboring human and mouse L1-EGFP reporter transgenes driven by the corresponding native L1 promoter (Table 1).[65] L1 insertions were identified in the offspring of transgenic males and females crossed with wild-type animals, and in

some cases the offspring containing an insertion lacked the transgene. This result is consistent with retrotransposition in the parental germline, prior to the end of meiosis I, and subsequent segregation of the new insertion away from the transgene during meiosis II. In such a scenario, insertion-positive, transgene-negative offspring are expected to be heterozygous for the insertion, and transmit it to approximately 50% of their offspring (**Figure 3**A). Unexpectedly, insertion-positive, transgene-negative animals uniformly failed to transmit the insertion to their offspring, and

harbored the insertion at <1 copy per cell. This result led to a model in which an L1 mRNA is transcribed in the parental germline and, perhaps complexed in an RNP, can be carried over during fertilization to undergo retrotransposition in a cell of the resultant F1 embryo.[65] The failure of such insertions to be inherited by F2 offspring suggests that these insertions contributed to somatic but not germline mosaicism, consistent with retrotransposition after specification of the germ lineage, or in pluripotent cells which did not contribute to the germline (Figure 3A).[65]
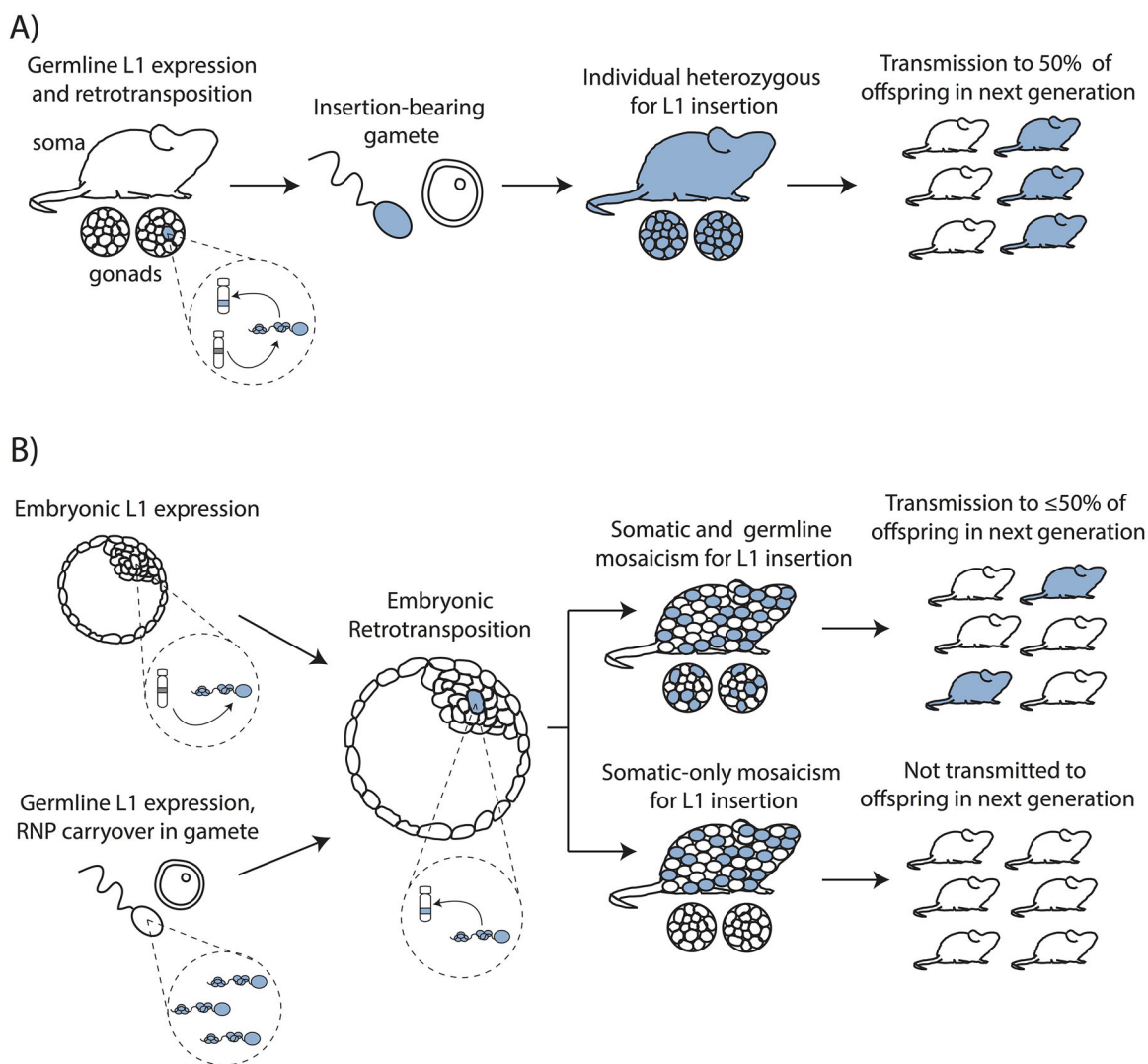

Figure 3. Germline and embryonic L1 retrotransposition. A) Germline L1 retrotransposition. A retrotransposition-competent L1 (gray) undergoes retrotransposition at some point during germline development (gonads represented by circles under mouse) to generate a new insertion (blue). If a gamete harboring this insertion (blue sperm) undergoes fertilization, the resultant animal will be heterozygous for the insertion (blue mouse) and is expected to transmit the insertion to ≈50% of its offspring. B) Embryonic L1 retrotransposition. From left: L1 expression and RNP formation may occur in a cell of the early embryo (above), or an L1 RNP formed in the parental germline may be carried over in a gamete during fertilization (below). Completion of retrotransposition in a pluripotent embryonic cell will give rise to an insertion for which the embryo is mosaic (initial insertion-harboring cell shown in blue). Descendants of the insertion-harboring cell may contribute to the somatic and germ cell lineages (above), rendering the insertion heritable by up to ≈50% of offspring. Alternatively, insertion-harboring cells may contribute only to somatic cell lineages (below), in which case the insertion is not heritable by offspring.

3.7. Endogenous Heritable Retrotransposition Events Occur in Early Mouse Embryos

Our recent results elucidating endogenous heritable L1 retrotransposition events share several consistencies with the studies described above, but also feature important differences.[91] In our study, we were able to trace the developmental timing of multiple endogenous L1 insertions, by using insertion-specific PCR genotyping to detect each insertion in the various somatic and germ tissues of the mouse of origin and of related mice. Among 11 de novo L1 insertions, we traced 6 to pluripotent embryonic cells, as evidenced by mosaicism for the insertion among the tissues of the originating mouse detected by conventional and quantitative genotyping PCR. Among these six insertions, we demonstrated transmission of four insertions from the mosaic animal of origin to its offspring, consistent with the findings of Van den Hurk et al. in humans and Prak et al. in a transgenic mouse (Figure 3B).[60,104] The remaining two L1 insertions could be detected in multiple tissues from the originating mosaic animal, but were not transmitted to the next generation. One explanation for this result is that the number of offspring produced by the originating mosaic animals was insufficient to observe infrequent transmission. The other possibility is that these two events contributed to somatic but not germline mosaicism during embryonic development of the originating animal, consistent with the results of Babushok et al. and Kano et al. (Figure 3B).[62,65]

3.8. What Proportion of Early Embryonic Insertions Are Heritable?

While more than half of the early embryonic insertions we identified contributed to the germline and were demonstrably heritable, the vast majority of engineered L1-EGFP insertions reported by earlier studies arose in early embryonic cells of transgenic animals and contributed to somatic mosaicism but were not transmitted to subsequent generations.[60,62,65] The likely explanation for this discrepancy is that our approach favored the detection of transmissible insertions. All four heritable embryonic insertions initially were identified by mRC-seq in bulk tissues from heterozygous offspring, and were subsequently detected at low prevalence in the corresponding maternal tissues. Thus, many additional somatic-restricted L1 insertions, such as those occurring in the neuronal lineage,[61,86,87,109,110] may have been present in our animals but fell below the detection capacity of mRC-seq.[74] Indeed, a transgenic mouse model allows much more sensitive detection of such "late" or low-prevalence insertions a priori, since the spliced reporter cassette is much more readily detected by PCR than a low-prevalence endogenous insertion is by high-throughput sequencing. Nevertheless, it is quite striking that Kano et al. never detected transmission of early embryonic insertions, despite using mouse and human L1 transgenes driven by their native promoters, and examining 170 F2 progeny. It is possible that the L1 transgenes preferentially generated insertions slightly later during embryonic development, restricted to cells committed to the somatic lineages, while insertions arising from endogenous donor L1s were more likely to generate earlier insertions in pluripotent cells with the capacity to contribute to the germline. Future studies employing single-cell genomics have the potential to resolve this discrepancy by elucidating the rate of endogenous embryonic retrotransposition in an unbiased manner, in totipotent, pluripotent, and lineage-committed cells during early embryonic development.

3.9. Carry-Over of L1 RNPs: Spatial and Temporal Separation of L1 Expression and Insertion?

Whether endogenous L1 RNPs arising in the maternal or paternal germline can be carried over during fertilization to generate early embryonic insertions, as has been proposed for L1 RNPs arising from engineered transgenes,[65] remains an open question. Our study employed inbred C57BL/6J mice, so the vast majority of potential donor L1s responsible for insertions and flanking SNPs were homozygous. Thus, we could not use 3′ transductions to discern whether de novo embryonic L1 insertions arose from a maternal or paternal donor L1, nor whether the mosaic daughter insertion occurred independently of the donor allele, indicative of endogenous L1 RNP carryover during fertilization. Future studies employing a more genetically diverse mouse model, in which the parent of origin and progenitor L1 element responsible for embryonic insertions can potentially be discerned, may allow formal demonstration of endogenous L1 RNP carryover during fertilization.

If L1 expression and TPRT can be spatially and temporally separated in this extreme scenario, it also seems likely that the transcription of a full-length L1 mRNA, and its translation and subsequent incorporation into an RNP, could occur in a given cell, and the RNP then persist during multiple rounds of cell division, and perhaps differentiation, to undergo retrotransposition in a descendant cell at a later developmental stage (**Figure 4**). Indeed, endogenous and engineered L1 proteins and mRNA have been demonstrated to form perinuclear aggregates,[111–113] which we hypothesize could facilitate the long-term protection and persistence of L1 RNPs. In such a scenario, we speculate that L1 expression could occur in pluripotent cells where L1 sequences are demethylated and transcriptionally active, and that L1 RNPs could persist to carry out retrotransposition at a stage where L1 sequences may be highly methylated but intracellular conditions (i.e., the presence or absence of certain host factors) allow TPRT to be completed. The hypothesis of mitotic L1 RNP carryover thus represents an exciting avenue for future research into L1 and host factor dynamics during mammalian development.

4. L1 Retrotransposition Occurs in the Germline

Germ cells represent a clear opportunity for the generation of heritable retrotransposition events, and developmentally regulated L1 expression in the mammalian germline suggests that germ cells may accommodate L1 retrotransposition. Full-length endogenous L1 mRNAs and ORF1p are expressed in prepuberal mouse testis at postnatal day 10, in leptotene and zygotene
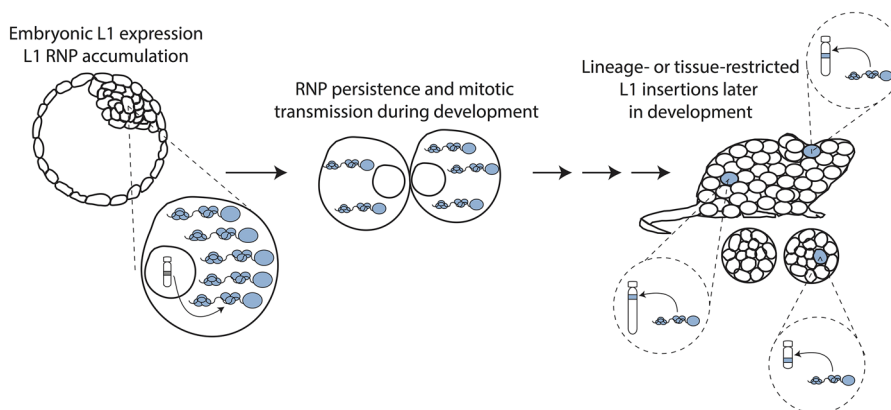
Figure 4. Hypothetical scenario for mitotic transmission of L1 RNPs during development. From left: L1 expression in a permissive cell type (e.g., a pluripotent embryonic cell) could lead to L1 RNP accumulation. If these RNPs persist during subsequent rounds of cell division and differentiation, they may complete retrotransposition much later in development, generating lineage-, or tissue-restricted insertions.

spermatocytes during the first wave of meiosis.[114] In adult testis, ORF1p is expressed in round and elongating spermatids, and can be faintly and rarely detected in spermatocytes.[114,115] L1 ORF1p can also be detected in fetal male and female germ cells from embryonic day 15.5.[94,116]

4.1. Does L1 Retrotransposition Occur in Human Germ Cells?

Evidence for L1 retrotransposition in the human germline is limited. Brouha et al. identified an L1 insertion bearing a 3′ transduction that had occurred into the X-linked CYBB gene of a male patient, resulting in a case of chronic granulomatous disease.[70] The 3′ transduction allowed the donor L1 for the disease-causing insertion to be identified. Moreover, a polymorphism within this transduced sequence made it possible to determine that the disease causing insertion had been inherited by the patient independently of its donor allele. This result is consistent with retrotransposition in the maternal germline prior to the completion of meiosis I. However, the observation of transgene-independent early embryonic insertions in mice resulting from L1 RNP carryover during fertilization, as reported by Kano et al.,[65] provides an equally likely explanation for this result. Indeed, while human oocytes have been demonstrated to accommodate the retrotransposition of L1-EGFP reporter constructs,[117] analysis of human sperm for endogenous L1 insertions using a PCR-based strategy uncovered no insertions and resulted in an estimate of fewer than one de novo insertion per 400 sperm.[118]

4.2. Transgenic Animal Models Undergo Germline L1 Retrotransposition

Results from transgenic animal models have provided some support for germline L1 retrotransposition. Ostertag et al. generated transgenic L1-EGFP mouse lines in which expression of the L1 was driven by the human L1 5′UTR with and without a heterologous mouse RNA polymerase (pPol II) promoter.[59] Two L1-EGFP insertions were detected among 135 offspring of male

pPol II-L1 EGFP mice mated to wild-type females. One insertion was inherited along with the transgene, and therefore could have occurred either in the male germline or post-fertilization in the developing embryo. The second insertion was inherited independently of the transgene, strongly suggesting that retrotransposition had occurred in the male germline prior to the end of meiosis I. Whether this insertion arose during meiosis or at an earlier developmental stage could not be determined.

Very recently, Newkirk et al.[66] developed a transgenic mouse model using the synthetic mouse L1 ORFeus_Mm, driven by a native L1 T_F promoter[10] and bearing an EGFP retrotransposition indicator cassette.[58] In contrast to most previous transgenic L1 mouse models (Table 1), the authors performed their study using a mouse line containing a single copy of the transgene, avoiding the complication of repeat-induced transgene silencing observed for multi-copy transgene concatamers.[119] Indeed, methylation of this transgene recapitulated the dynamics of endogenous L1 T_F copies in pre- and post-natal male germ cells and somatic cells. Insertions arising from the L1-EGFP transgene were detected using a highly sensitive digital droplet PCR (ddPCR) strategy targeting the EGFP splice junction. Newkirk et al. applied this system to mice deficient in *Mov10l1*, an RNA helicase with an essential role in the piRNA pathway, an adaptive transposon defense mechanism active in male gonadal primordial germ cells.[120,121] Their results clearly demonstrated that deficiency in a piRNA pathway effector results not only in an increase in L1 expression, but also an increase in retrotransposition in germ cells during meiosis.

Although Newkirk et al. primarily focused on the increase in germline L1 retrotransposition in *Mov10l1−/−* mice, their findings have interesting implications for the dynamics of heritable retrotransposition in wild-type animals. Based upon a frequency of 0.00024 engineered ORFeus_Mm insertions per cell as detected by ddPCR in adult control (*Mov10l1+/−*) testes, the estimation of 3000 active L1 copies per diploid mouse genome,[4] and the assumption that ORFeus_Mm retrotransposes 200-fold more efficiently than the average active L1 copy,[72] the authors arrive at an estimated frequency of one insertion per 278 cells in the mouse germline. As the authors note, their methodology does not distinguish between an insertion arising de novo specifically in a germ cell, and an insertion that arose in

a progenitor cell at an earlier stage of development. Indeed, their data reflect a degree of variability between individual animals in the prevalence of insertions in adult tissues and also in staged spermatogenic cells. This inter-individual variability could arise from different levels of de novo retrotransposition in individual somatic and germ cells, or from differences in the relative germline prevalence of insertions that occurred earlier during the development of each mouse, either prior to germline specification or during earlier germline development. The developmental timing of engineered insertions in control animals bearing this experimental system therefore remains to be determined. Indeed, development of a transgenic mouse model that allows the distinction of individual engineered L1 insertions could be an invaluable tool for lineage-tracing experiments that establish the developmental timing of L1 retrotransposition events.

4.3. How Frequently Does L1 Retrotranspose in Mouse Germ Cells?

In our examination of de novo endogenous L1 insertions in mice, we uncovered a single L1 insertion that was uniformly heterozygous in all tissues of the originating mouse, and could not be detected in the somatic or germ tissues of either parental animal (Figure 3A).[91] We attributed this insertion to germline development, although the exact timing of this event and the parental origin for the insertion could not be established. Determination of endogenous L1 retrotransposition frequency during later stages of germline development will likely require single-cell analysis applied to staged spermatogenic cells or individual developing oocytes.

5. L1 Retrotransposition Occurs in Early Primordial Germ Cells

Mouse primordial germline development begins in the posterior primitive streak of the post-implantation embryo at embryonic day 6.5 (E6.5) with a discrete founding population of around 40 PGCs in the proximal epiblast. PGCs then migrate through the developing hindgut from E7.5–E8.5, and then along the midline of the embryo through the hindgut epithelium and dorsal mesentary at about E9.0–E9.5. The PGCs then form two clusters as they move laterally to colonize the genital ridges from E10.0–E12.5 (Figure 2).[122] In our recent study,[91] three out of the 11 de novo L1 insertions identified showed a distinctive pattern of inheritance: among the offspring of a given male mouse, 3–5% of siblings were heterozygous for a new L1 insertion that was not present in the somatic tissues of either parent. However, the insertion was present at <1 copy per cell in both testicles of the paternal mouse. Together, these results indicated germline-restricted mosaicism for the insertion, and that the insertion had originated very early during germline development, most likely in an early PGC prior to colonization of the genital ridges and the formation of two testicles.

L1 retrotransposition at this very early stage of germline development represents a largely unexplored aspect of L1 biology in mammals. Indeed, previous studies have focused on L1

expression and activity in later germ cells, particularly during meiosis. Thus, numerous questions remain to be answered: is L1 expression activated at the onset of PGC specification, or afterwards, as PGCs migrate toward the genital ridges? Genome-wide demethylation occurs progressively during these processes, including on L1 copies,[123] but what other epigenetic mechanisms regulate L1 in early PGCs? Notably, a recent study by Kim et al. demonstrated L1 expression is upregulated in E11.5–E12.5 mouse PGCs with conditional deletion of the histone arginine methyltransferase PRMT5, although a corresponding increase in L1 copy number was not observed.[124]

The low number of PGCs per mouse, particularly in the PGC founder population, renders PGC specification and early development a particularly difficult process to study in vivo. Ethical considerations limit the extent to which PGC development can be studied in humans. However, recently-developed protocols for differentiating mouse and human cultured pluripotent cells to PGC-like cells may be more tractable to gain insight into the dynamics of endogenous L1 regulation during PGC specification, and to high-throughput sequencing approaches to identify endogenous L1 retrotransposition events occurring during PGC specification in vitro.[125,126] Such cultured cell systems may also be amenable to the use of engineered L1 reporter constructs. Future studies will no doubt elucidate the dynamics of L1 regulation and activity in this important developmental niche.

6. Conclusions and Outlook

The generation of heritable L1 insertions is critically important for the evolutionary success of L1 as a selfish genetic element, and has had a profound effect on the landscape of mammalian genomes. Rare cases where the developmental origins of disease-causing L1 insertions can be traced, in addition to innovative cultured cell and transgenic mouse model systems allowing the detection of engineered L1 retrotransposition events, have provided invaluable insight regarding L1 dynamics in the mammalian germline and early embryo. Our recent work took advantage of high-throughput sequencing approaches to identify and track the developmental origins of heritable, endogenous L1 insertions.[91] Our results build upon and strengthen the conclusions from previous studies, but also reveal important differences between the results obtained from transgenic model systems and via mapping endogenous L1 retrotransposition events. Going forward, the integration of high-throughput sequencing approaches, including single-cell genomics, with engineered reporter systems in cultured cells and transgenic animals, should provide the means to answer remaining questions regarding the rate and timing of L1 retrotransposition during mammalian development, the fate of embryonic cells carrying de novo L1 mutations, and the consequences of this activity on the host genome.

1700189 (10 of 14)

Conflict of Interest

The authors declare no conflict of interest.

[1] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, *Nature* **2001**, *409*, 860.

[2] R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, E. S. Lander, *Nature* **2002**, *420*, 520.

[3] B. Brouha, J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran, H. H. Kazazian, Jr., *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5280.

[4] J. L. Goodier, E. M. Ostertag, K. Du, H. H. Kazazian, Jr., *Genome Res.* **2001**, *11*, 1677.

[5] B. A. Dombroski, S. L. Mathias, E. Nanthakumar, A. F. Scott, H. H. Kazazian, Jr., *Science* **1991**, *254*, 1805.

[6] T. G. Fanning, *Nucl. Acids Res.* **1983**, *11*, 5073.

[7] G. D. Swergold, *Mol. Cell Biol.* **1990**, *10*, 6718.

[8] D. D. Loeb, R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, C. A. Hutchison, 3rd, *Mol. Cell Biol.* **1986**, *6*, 168.

[9] R. J. DeBerardinis, J. L. Goodier, E. M. Ostertag, H. H. Kazazian, Jr., *Nat. Genet.* **1998**, *20*, 288.

[10] T. P. Naas, R. J. DeBerardinis, J. V. Moran, E. M. Ostertag, S. F. Kingsmore, M. F. Seldin, Y. Hayashizaki, S. L. Martin, H. H. Kazazian, *EMBO J.* **1998**, *17*, 590.

[11] W. Wei, N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke, J. V. Moran, *Mol Cell Biol* **2001**, *21*, 1429.

[12] H. Hohjoh, M. F. Singer, *EMBO J.* **1996**, *15*, 630.

[13] S. L. Martin, *Mol. Cell Biol.* **1991**, *11*, 4804.

[14] Q. Feng, J. V. Moran, H. H. Kazazian, Jr., J. D. Boeke, *Cell* **1996**, *87*, 905.

[15] G. J. Cost, Q. Feng, A. Jacquier, J. D. Boeke, *EMBO J.* **2002**, *21*, 5899.

[16] D. D. Luan, M. H. Korman, J. L. Jakubczak, T. H. Eickbush, *Cell* **1993**, *72*, 595.

1700189 (11 of 14)

[17] D. A. Kulpa, J. V. Moran, *Nat. Struct. Mol. Biol.* **2006**, *13*, 655.

[18] G. Grimaldi, J. Skowronski, M. F. Singer, *EMBO J.* **1984**, *3*, 1753.

[19] N. Gilbert, S. Lutz-Prigge, J. V. Moran, *Cell* **2002**, *110*, 315.

[20] N. Gilbert, S. Lutz, T. A. Morrish, J. V. Moran, *Mol. Cell Biol.* **2005**, *25*, 7780.

[21] H. H. Kazazian, Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, S. E. Antonarakis, *Nature* **1988**, *332*, 164.

[22] M. Dewannieux, C. Esnault, T. Heidmann, *Nat. Genet.* **2003**, *35*, 41.

[23] J. Raiz, A. Damert, S. Chira, U. Held, S. Klawitter, M. Hamdorf, J. Lower, W. H. Stratling, R. Lower, G. G. Schumann, *Nucl Acids Res.* **2012**, *40*, 1666.

[24] D. C. Hancks, J. L. Goodier, P. K. Mandal, L. E. Cheung, H. H. Kazazian, Jr., *Hum. Mol. Genet.* **2011**, *20*, 3386.

[25] M. Dewannieux, T. Heidmann, *J. Mol. Biol.* **2005**, *349*, 241.

[26] C. Esnault, J. Maestre, T. Heidmann, *Nat. Genet.* **2000**, *24*, 363.

[27] A. D. Ewing, T. J. Ballinger, D. Earl, C. C. Harris, L. Ding, R. K. Wilson, D. Haussler, *Genome Biol.* **2013**, *14*, R22.

[28] A. J. Doucet, J. E. Wilusz, T. Miyoshi, Y. Liu, J. V. Moran, *Mol. Cell.* **2015**, *60*, 728.

[29] J. S. Myers, B. J. Vincent, H. Udall, W. S. Watkins, T. A. Morrish, G. E. Kilroy, G. D. Swergold, J. Henke, L. Henke, J. V. Moran, L. B. Jorde, M. A. Batzer, *Am. J. Hum. Genet.* **2002**, *71*, 312.

[30] E. M. Ostertag, H. H. Kazazian, Jr., *Genome Res.* **2001**, *11*, 2059.

[31] N. Gilbert, J. M. Bomar, M. Burmeister, J. V. Moran, *Hum. Mutat.* **2004**, *24*, 9.

[32] J. V. Moran, R. J. DeBerardinis, H. H. Kazazian, Jr., *Science* **1999**, *283*, 1530.

[33] O. K. Pickeral, W. Makalowski, M. S. Boguski, J. D. Boeke, *Genome Res.* **2000**, *10*, 411.

[34] J. L. Goodier, E. M. Ostertag, H. H. Kazazian, Jr., *Hum. Mol. Genet.* **2000**, *9*, 653.

[35] Y. Miki, I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K. W. Kinzler, B. Vogelstein, Y. Nakamura, *Cancer Res.* **1992**, *52*, 643.

[36] E. C. Scott, E. J. Gardner, A. Masood, N. T. Chuang, P. M. Vertino, S. E. Devine, *Genome Res.* **2016**, *26*, 745.

[37] V. Perepelitsa-Belancio, P. Deininger, *Nat. Genet.* **2003**, *35*, 363.

[38] V. P. Belancio, A. M. Roy-Engel, P. Deininger, *Gene* **2008**, *411*, 38.

[39] J. S. Han, S. T. Szak, J. D. Boeke, *Nature* **2004**, *429*, 268.

[40] R. Shukla, K. R. Upton, M. Munoz-Lopez, D. J. Gerhardt, M. E. Fisher, T. Nguyen, P. M. Brennan, J. K. Baillie, A. Collino, S. Ghisletti, S. Sinha, F. Iannelli, E. Radaelli, A. Dos Santos, D. Rapoud, C. Guettier, D. Samuel, G. Natoli, P. Carninci, F. D. Ciccarelli, J. L. Garcia-Perez, J. Faivre, G. J. Faulkner, *Cell* **2013**, *153*, 101.

[41] A. M. Denli, I. Narvaiza, B. E. Kerman, M. Pena, C. Benner, M. C. Marchetto, J. K. Diedrich, A. Aslanian, J. Ma, J. J. Moresco, L. Moore, T. Hunter, A. Saghatelian, F. H. Gage, *Cell* **2015**, *163*, 583.

[42] H. H. Kazazian, Jr., J. V. Moran, *N. Eng. J. Med.* **2017**, *377*, 361.

[43] E. M. Ostertag, H. H. Kazazian, Jr., *Annu. Rev. Genet.* **2001**, *35*, 501.

[44] J. Chen, A. Rattner, J. Nathans, *Hum. Mol. Genet.* **2006**, *15*, 2146.

[45] J. H. Crichton, D. S. Dunican, M. Maclennan, R. R. Meehan, *Cell. Mol. Life Sci.* **2014**, *71*, 1581.

[46] J. L. Goodier, *Mobile DNA* **2016**, *7*, 16.

[47] A. D. Ewing, H. H. Kazazian, Jr., *Genome Res.* **2010**, *20*, 1262.

[48] J. Xing, Y. Zhang, K. Han, A. H. Salem, S. K. Sen, C. D. Huff, Q. Zhou, E. F. Kirkness, S. Levy, M. A. Batzer, L. B. Jorde, *Genome Res.* **2009**, *19*, 1516.

[49] H. H. Kazazian, Jr., *Nat. Genet.* **1999**, *22*, 130.

[50] R. Cordaux, D. J. Hedges, S. W. Herke, M. A. Batzer, *Gene* **2006**, *373*, 134.

[51] X. Li, W. A. Scaringe, K. A. Hill, S. Roberts, A. Mengos, D. Careri, M. T. Pinto, C. K. Kasper, S. S. Sommer, *Hum. Mutat.* **2001**, *17*, 511.

[52] D. C. Hancks, H. H. Kazazian, Jr., *Curr. Opin. Genet. Dev.* **2012**, *22*, 191.

[53] H. H. Kazazian, Jr., JV. Moran, *Nat. Genet.* **1998**, *19*, 19.

[54] H. H. Kazazian, Jr., *Science* **2000**, *289*, 1152.

[55] J. V. Moran, S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, H. H. Kazazian, Jr., *Cell* **1996**, *87*, 917.

[56] S. H. Rangwala, H. H. Kazazian, Jr., *Methods* **2009**, *49*, 219.

[57] S. R. Richardson, A. J. Doucet, H. C. Kopera, J. B. Moldovan, J. L. Garcia-Perez, J. V. Moran, *Microbiol. Spectr.* **2015**, *3*, MDNA3-0061-2014.

[58] E. M. Ostertag, E. T. Prak, R. J. DeBerardinis, J. V. Moran, H. H. Kazazian, Jr., *Nucl. Acids Res.* **2000**, *28*, 1418.

[59] E. M. Ostertag, R. J. DeBerardinis, J. L. Goodier, Y. Zhang, N. Yang, G. L. Gerton, H. H. Kazazian, Jr., *Nat. Genet.* **2002**, *32*, 655.

[60] E. T. Prak, A. W. Dodson, E. A. Farkash, H. H. Kazazian, Jr., *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1832.

[61] A. R. Muotri, V. T. Chu, M. C. Marchetto, W. Deng, J. V. Moran, F. H. Gage, *Nature* **2005**, *435*, 903.

[62] D. V. Babushok, E. M. Ostertag, C. E. Courtney, J. M. Choi, H. H. Kazazian, Jr., *Genome Res.* **2006**, *16*, 240.

[63] W. An, J. S. Han, S. J. Wheelan, E. S. Davis, C. E. Coombes, P. Ye, C. Triplett, J. D. Boeke, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18662.

[64] W. An, J. S. Han, C. M. Schrum, A. Maitra, F. Koentgen, J. D. Boeke, *Genesis* **2008**, *46*, 373.

[65] H. Kano, I. Godoy, C. Courtney, M. R. Vetter, G. L. Gerton, E. M. Ostertag, H. H. Kazazian, Jr., *Genes. Dev.* **2009**, *23*, 1303.

[66] S. J. Newkirk, S. Lee, F. C. Grandi, V. Gaysinskaya, J. M. Rosser, N. Vanden Berg, C. A. Hogarth, M. C. N. Marchetto, A. R. Muotri, M. D. Griswold, P. Ye, A. Bortvin, F. H. Gage, J. D. Boeke, W. An, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E5635.

[67] J. L. Garcia-Perez, M. Morell, J. O. Scheys, D. A. Kulpa, S. Morell, C. C. Carter, G. D. Hammer, K. L. Collins, K. S. O'Shea, P. Menendez, J. V. Moran, *Nature* **2010**, *466*, 769.

[68] A. Macia, T. J. Widmann, S. R. Heras, V. Ayllon, L. Sanchez, M. Benkaddour-Boumzaouad, M. Munoz-Lopez, A. Rubio, S. Amador-Cubero, E. Blanco-Jimenez, J. Garcia-Castro, P. Menendez, P. Ng, A. R. Muotri, J. L. Goodier, J. L. Garcia-Perez, *Genome Res.* **2017**, *27*, 335.

[69] M. L. Kimberland, V. Divoky, J. Prchal, U. Schwahn, W. Berger, H. H. Kazazian, Jr., *Hum. Mol. Genet.* **1999**, *8*, 1557.

[70] B. Brouha, C. Meischl, E. Ostertag, M. de Boer, Y. Zhang, H. Neijens, D. Roos, H. H. Kazazian, Jr., *Am. J. Hum. Genet.* **2002**, *71*, 327.

[71] W. An, J. D. Boeke, *Genome Biol.* **2005**, *6*, 361.

[72] J. S. Han, J. D. Boeke, *Nature* **2004**, *429*, 314.

[73] A. D. Ewing, *Mobile DNA* **2015**, *6*, 24.

[74] G. J. Faulkner, J. L. Garcia-Perez, *Trends Genet.* **2017**, *33*, 802.

[75] K. Akagi, J. Li, R. M. Stephens, N. Volfovsky, D. E. Symer, *Genome Res.* **2008**, *18*, 869.

[76] T. M. Keane, L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. A. Guerra-Assuncao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint, D. J. Adams, *Nature* **2011**, *477*, 289.

[77] C. Nellaker, T. M. Keane, B. Yalcin, K. Wong, A. Agam, T. G. Belgard, J. Flint, D. J. Adams, W. N. Frankel, C. P. Ponting, *Genome Biol.* **2012**, *13*, R45.

[78] A. R. Quinlan, R. A. Clark, S. Sokolova, M. L. Leibowitz, Y. Zhang, M. E. Hurles, J. C. Mell, I. M. Hall, *Genome Res.* **2010**, *20*, 623.

[79] E. Lee, R. Iskow, L. Yang, O. Gokcumen, P. Haseley, L. J. Luquette, 3rd, J. G. Lohr, C. C. Harris, L. Ding, R. K. Wilson, D. A. Wheeler, R. A. Gibbs, R. Kucherlapati, C. Lee, P. V. Kharchenko, P. J. Park, *Science* **2012**, *337*, 967.

[80] J. L. Hazen, G. G. Faust, A. R. Rodriguez, W. C. Ferguson, S. Shumilina, R. A. Clark, M. J. Boland, G. Martin, P. Chubukov, R. K. Tsunemoto, A. Torkamani, S. Kupriyanov, I. M. Hall, K. K. Baldwin, *Neuron* **2016**, *89*, 1223.

[81] J. K. Baillie, M. W. Barnett, K. R. Upton, D. J. Gerhardt, T. A. Richmond, F. De Sapio, P. M. Brennan, P. Rizzu, S. Smith, M. Fell, R. T. Talbot, S. Gustincich, T. C. Freeman, J. S. Mattick, D. A. Hume, P. Heutink, P. Carninci, J. A. Jeddeloh, G. J. Faulkner, *Nature* **2011**, *479*, 534.

[82] F. M. Sheen, S. T. Sherry, G. M. Risch, M. Robichaux, I. Nasidze, M. Stoneking, M. A. Batzer, G. D. Swergold, *Genome Res.* **2000**, *10*, 1496.

[83] I. Ovchinnikov, A. B. Troxel, G. D. Swergold, *Genome Res.* **2001**, *11*, 2050.

[84] R. M. Badge, R. S. Alisch, J. V. Moran, *Am. J. Hum. Genet.* **2003**, *72*, 823.

[85] R. C. Iskow, M. T. McCabe, R. E. Mills, S. Torene, W. S. Pittard, A. F. Neuwald, E. G. Van Meir, P. M. Vertino, S. E. Devine, *Cell* **2010**, *141*, 1253.

[86] K. R. Upton, D. J. Gerhardt, J. S. Jesuadian, S. R. Richardson, F. J. Sanchez-Luque, G. O. Bodea, A. D. Ewing, C. Salvador-Palomeque, M. S. van der Knaap, P. M. Brennan, A. Vanderver, G. J. Faulkner, *Cell* **2015**, *161*, 228.

[87] G. D. Evrony, X. Cai, E. Lee, L. B. Hills, P. C. Elhosary, H. S. Lehmann, J. J. Parker, K. D. Atabay, E. C. Gilmore, A. Poduri, P. J. Park, C. A. Walsh, *Cell* **2012**, *151*, 483.

[88] S. R. Richardson, S. Morell, G. J. Faulkner, *Annu. Rev. Genet.* **2014**, *48*, 1.

[89] R. J. Mural, M. D. Adams, E. W. Myers, H. O. Smith, G. L. Miklos, R. Wides, A. Halpern, P. W. Li, G. G. Sutton, J. Nadeau, S. L. Salzberg, R. A. Holt, C. D. Kodira, F. Lu, L. Chen, Z. Deng, C. C. Evangelista, W. Gan, T. J. Heiman, J. Li, Z. Li, G. V. Merkulov, N. V. Milshina, A. K. Naik, R. Qi, B. C. Shue, A. Wang, J. Wang, X. Wang, X. Yan, J. Ye, S. Yooseph, Q. Zhao, L. Zheng, S. C. Zhu, K. Biddick, R. Bolanos, A. L. Delcher, I. M. Dew, D. Fasulo, M. J. Flanigan, D. H. Huson, S. A. Kravitz, J. R. Miller, C. M. Mobarry, K. Reinert, K. A. Remington, Q. Zhang, X. H. Zheng, D. R. Nusskern, Z. Lai, Y. Lei, W. Zhong, A. Yao, P. Guan, R. R. Ji, Z. Gu, Z. Y. Wang, F. Zhong, C. Xiao, C. C. Chiang, M. Yandell, J. R. Wortman, P. G. Amanatides, S. L. Hladun, E. C. Pratts, J. E. Johnson, K. L. Dodson, K. J. Woodford, C. A. Evans, B. Gropman, D. B. Rusch, E. Venter, M. Wang, T. J. Smith, J. T. Houck, D. E. Tompkins, C. Haynes, D. Jacob, S. H. Chin, D. R. Allen, C. E. Dahlke, R. Sanders, K. Li, X. Liu, A. A. Levitsky, W. H. Majoros, Q. Chen, A. C. Xia, J. R. Lopez, M. T. Donnelly, M. H. Newman, A. Glodek, C. L. Kraft, M. Nodell, F. Ali, H. J. An, D. Baldwin-Pitts, K. Y. Beeson, S. Cai, M. Carnes, A. Carver, P. M. Caulk, A. Center, Y. H. Chen, M. L. Cheng, M. D. Coyne, M. Crowder, S. Danaher, L. B. Davenport, R. Desilets, S. M. Dietz, L. Doup, P. Dullaghan, S. Ferriera, C. R. Fosler, H. C. Gire, A. Glueecksmann, J. D. Gocayne, J. Gray, B. Hart, J. Haynes, J. Hoover, T. Howland, C. Ibegwam, M. Jalali, D. Johns, L. Kline, D. S. Ma, S. MacCawley, A. Magoon, F. Mann, D. May, T. C. McIntosh, S. Mehta, L. Moy, M. C. Moy, B. J. Murphy, S. D. Murphy, K. A. Nelson, Z. Nuri, K. A. Parker, A. C. Prudhomme, V. N. Puri, H. Qureshi, J. C. Raley, M. S. Reardon, M. A. Regier, Y. H. Rogers, D. L. Romblad, J. Schutz, J. L. Scott, R. Scott, C. D. Sitter, M. Smallwood, A. C. Sprague, E. Stewart, R. V. Strong, E. Suh, K. Sylvester, R. Thomas, N. N. Tint, C. Tsonis, G. Wang, G. Wang, M. S. Williams, S. M. Williams, S. M. Windsor, K. Wolfe, M. M. Wu, J. Zaveri, K. Chaturvedi, A. E. Gabrielian, Z. Ke, J. Sun, G. Subramanian, J. C. Venter, C. M. Pfannkoch, M. Barnstead,

L. D. Stephenson. Barnstead, L. D. Stephenson, *Science* **2002**, *296*, 1661.

[90] B. Yalcin, K. Wong, A. Agam, M. Goodson, T. M. Keane, X. Gan, C. Nellaker, L. Goodstadt, J. Nicod, A. Bhomra, P. Hernandez-Pliego, H. Whitley, J. Cleak, R. Dutton, D. Janowitz, R. Mott, D. J. Adams, J. Flint, *Nature* **2011**, *477*, 326.

[91] S. R. Richardson, P. Gerdes, D. J. Gerhardt, F. J. Sanchez-Luque, G. O. Bodea, M. Munoz-Lopez, J. S. Jesuadian, M. H. C. Kempen, P. E. Carreira, J. A. Jeddeloh, J. L. Garcia-Perez, H. H. Kazazian, Jr., A. D. Ewing, G. J. Faulkner, *Genome Res.* **2017**, *27*, 1395.

[92] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877.

[93] A. B. Sahakyan, P. Murat, C. Mayer, S. Balasubramanian, *Nat. Struct. Mol. Biol.* **2017**, *24*, 243.

[94] S. N. Schauer, P. E. Carreira, R. Shukla, D. J. Gerhardt, P. Gerdes, F. J. Sanchez-Luque, P. Nicoli, M. Kindlova, S. Ghisletti, A. D. Santos, D. Rapoud, D. Samuel, J. Faivre, A. D. Ewing, S. R. Richardson, G. J. Faulkner, L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res* **2018**.

[95] S. Malki, G. W. van der Heijden, K. A. O'Donnell, S. L. Martin, A. Bortvin, *Dev. Cell* **2014**, *29*, 521.

[96] W. Wei, T. A. Morrish, R. S. Alisch, J. V. Moran, *Anal. Biochem.* **2000**, *284*, 435.

[97] M. Kannan, J. Li, S. E. Fritz, K. E. Husarek, J. C. Sanford, T. L. Sullivan, P. K. Tiwary, W. An, J. D. Boeke, D. E. Symer, *Mobile DNA* **2017**, *8*, 8.

[98] M. MacLennan, M. Garcia-Canadas, J. Reichmann, E. Khazina, G. Wagner, C. J. Playfoot, C. Salvador-Palomeque, A. R. Mann, P. Peressini, L. Sanchez, K. Dobie, D. Read, C. C. Hung, R. Eskeland, R. R. Meehan, O. Weichenrieder, J. L. Garcia-Perez, I. R. Adams, *eLife* **2017**, *6*, e26152.

[99] C. R. Beck, P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge, J. V. Moran, *Cell* **2010**, *141*, 1159.

[100] C. M. Macfarlane, P. Collier, R. Rahbari, C. R. Beck, J. F. Wagstaff, S. Igoe, J. V. Moran, R. M. Badge, *Hum. Mutat.* **2013**, *34*, 974.

[101] J. M. Tubio, Y. Li, Y. S. Ju, I. Martincorena, S. L. Cooke, M. Tojo, G. Gundem, C. P. Pipinikas, J. Zamora, K. Raine, A. Menzies, P. Roman-Garcia, A. Fullam, M. Gerstung, A. Shlien, P. S. Tarpey, E. Papaemmanuil, S. Knappskog, P. Van Loo, M. Ramakrishna, H. R. Davies, J. Marshall, D. C. Wedge, J. W. Teague, A. P. Butler, S. Nik-Zainal, L. Alexandrov, S. Behjati, L. R. Yates, N. Bolli, L. Mudie, C. Hardy, S. Martin, S. McLaren, S. O'Meara, E. Anderson, M. Maddison, S. Gamble, I. B. C. Group, I. B. C. Group, I. P. C. Group, C. Foster, A. Y. Warren, H. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, A. G. Lynch, T. Visakorpi, W. B. Isaacs, L. van't Veer, C. Caldas, C. Desmedt, C. Sotiriou, S. Aparicio, J. A. Foekens, J. E. Eyfjord, S. R. Lakhani, G. Thomas, O. Myklebost, P. N. Span, A. L. Borresen-Dale, A. L. Richardson, M. Van de Vijver, A. Vincent-Salomon, G. G. Van den Eynden, A. M. Flanagan, P. A. Futreal, S. M. Janes, G. S. Bova, M. R. Stratton, U. McDermott, P. J. Campbell, *Science* **2014**, *345*, 1251343.

[102] E. J. Gardner, V. K. Lam, D. N. Harris, N. T. Chuang, E. C. Scott, W. S. Pittard, R. E. Mills, C. Genomes Project, S. E. Devine, *Genome Res.* **2017**, *27*, 1916.

[103] E. C. Scott, S. E. Devine, *Viruses* **2017**, *9*, 131.

[104] J. A. van den Hurk, I. C. Meij, M. C. Seleme, H. Kano, K. Nikopoulos, L. H. Hoefsloot, E. A. Sistermans, I. J. de Wijs, A. Mukhopadhyay, A. S. Plomp, P. T. de Jong, H. H. Kazazian, F. P. Cremers, *Hum. Mol. Genet.* **2007**, *16*, 1587.

[105] J. L. Garcia-Perez, M. C. Marchetto, A. R. Muotri, N. G. Coufal, F. H. Gage, K. S. O'Shea, J. V. Moran, *Hum. Mol. Genet.* **2007**, *16*, 1569.

[106] S. Wissing, M. Munoz-Lopez, A. Macia, Z. Yang, M. Montano, W. Collins, J. L. Garcia-Perez, J. V. Moran, W. C. Greene, *Hum. Mol. Genet.* **2012**, *21*, 208.

[107] S. Klawitter, N. V. Fuchs, K. R. Upton, M. Munoz-Lopez, R. Shukla, J. Wang, M. Garcia-Canadas, C. Lopez-Ruiz, D. J. Gerhardt, A. Sebe, I. Grabundzija, S. Merkert, P. Gerdes, J. A. Pulgarin, A. Bock, U. Held, A. Witthuhn, A. Haase, B. Sarkadi, J. Lower, E. J. Wolvetang, U. Martin, Z. Ivics, Z. Izsvak, J. L. Garcia-Perez, G. J. Faulkner, G. G. Schumann, *Nat. Commun.* **2016**, *7*, 10286.

[108] A. R. Quinlan, M. J. Boland, M. L. Leibowitz, S. Shumilina, S. M. Pehrson, K. K. Baldwin, I. M. Hall, *Cell Stem Cell* **2011**, *9*, 366.

[109] G. D. Evrony, E. Lee, B. K. Mehta, Y. Benjamini, R. M. Johnson, X. Cai, L. Yang, P. Haseley, H. S. Lehmann, P. J. Park, C. A. Walsh, *Neuron* **2015**, *85*, 49.

[110] J. A. Erwin, A. C. Paquola, T. Singer, I. Gallina, M. Novotny, C. Quayle, T. A. Bedrosian, F. I. Alves, C. R. Butcher, J. R. Herdy, A. Sarkar, R. S. Lasken, A. R. Muotri, F. H. Gage, *Nat. Neurosci.* **2016**, *19*, 1583.

[111] A. J. Doucet, A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran, N. Gilbert, *PLoS Genet.* **2010**, 6.

[112] J. L. Goodier, P. K. Mandal, L. Zhang, H. H. Kazazian, Jr., *Hum. Mol. Genet.* **2010**, *19*, 1712.

[113] S. L. Martin, D. Branciforte, *Mol. Cell Biol.* **1993**, *13*, 5383.

[114] D. Branciforte, S. L. Martin, *Mol. Cell Biol.* **1994**, *14*, 2584.

[115] S. F. Soper, G. W. van der Heijden, T. C. Hardiman, M. Goodheart, S. L. Martin, P. de Boer, A. Bortvin, *Dev. Cell* **2008**, *15*, 285.

[116] S. A. Trelogan, S. L. Martin, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 1520.

[117] I. Georgiou, D. Noutsopoulos, E. Dimitriadou, G. Markopoulos, A. Apergi, L. Lazaros, T. Vaxevanoglou, K. Pantos, M. Syrrou, T. Tzavaras, *Hum. Mol. Genet.* **2009**, *18*, 1221.

[118] P. Freeman, C. Macfarlane, P. Collier, A. J. Jeffreys, R. M. Badge, *Hum. Mutat.* **2011**, *32*, 978.

[119] D. Garrick, S. Fiering, D. I. Martin, E. Whitelaw, *Nat. Genet.* **1998**, *18*, 56.

[120] A. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, G. J. Hannon, *Mol. Cell.* **2008**, *31*, 785.

[121] K. Zheng, J. Xiol, M. Reuter, S. Eckardt, N. A. Leu, K. J. McLaughlin, A. Stark, R. Sachidanandam, R. S. Pillai, P. J. Wang, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 11841.

[122] K. A. Ewen, P. Koopman, *Mol. Cell. Endocrinol.* **2010**, *323*, 76.

[123] S. Seisenberger, S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, W. Reik, *Mol. Cell.* **2012**, *48*, 849.

[124] S. Kim, U. Gunesdogan, J. J. Zylicz, J. A. Hackett, D. Cougot, S. Bao, C. Lee, S. Dietmann, G. E. Allen, R. Sengupta, M. A. Surani, *Mol. Cell.* **2014**, *56*, 564.

[125] K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, M. Saitou, *Cell* **2011**, *146*, 519.

[126] K. Sasaki, S. Yokobayashi, T. Nakamura, I. Okamoto, Y. Yabuta, K. Kurimoto, H. Ohta, Y. Moritoki, C. Iwatani, H. Tsuchiya, S. Nakamura, K. Sekiguchi, T. Sakuma, T. Yamamoto, T. Mori, K. Woltjen, M. Nakagawa, T. Yamamoto, K. Takahashi, S. Yamanaka, M. Saitou, *Cell Stem Cell* **2015**, *17*, 178.