

Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health

Victoria P. Belancio, Dale J. Hedges, and Prescott Deiner¹

Tulane Cancer Center and Department of Epidemiology, Tulane University Health Sciences Center, New Orleans, Louisiana 70112, USA

Transposable elements (TEs) have shared an exceptionally long coexistence with their host organisms and have come to occupy a significant fraction of eukaryotic genomes. The bulk of the expansion occurring within mammalian genomes has arisen from the activity of type I retrotransposons, which amplify in a “copy-and-paste” fashion through an RNA intermediate. For better or worse, the sequences of these retrotransposons are now wedded to the genomes of their mammalian hosts. Although there are several reported instances of the positive contribution of mobile elements to their host genomes, these discoveries have occurred alongside growing evidence of the role of TEs in human disease and genetic instability. Here we examine, with a particular emphasis on human retrotransposon activity, several newly discovered aspects of mammalian retrotransposon biology. We consider their potential impact on host biology as well as their ultimate implications for the nature of the TE–host relationship.

With rare exception, transposable elements (TEs) comprise a significant fraction of all eukaryotic organisms for which appreciable genomic sequence is available. As more genomes are sequenced, we are uncovering an ever-increasing diversity of mobile element families as well as a remarkable level of variation in the overall fraction of genomes occupied by these elements. When considering fully sequenced genomes, TE-derived sequence comprise from 30% to over half of mammalian DNA (Lander et al. 2001; Mouse Genome Sequencing Consortium et al. 2002; Lindblad-Toh et al. 2005; Mikkelsen et al. 2007; Pontius et al. 2007). With some exceptions, the majority of this repetitive DNA consists of class I retrotransposon elements, which copy themselves in the genome via an RNA-based intermediate. Within the human genome, the non-long-terminal-repeat (non-LTR) retrotransposon, L1, is the dominant family of elements driving amplification (Fig. 1). As with human genomes, rodent genomes exhibit L1 activity, but they also experience a significant level of LTR retrotransposition (reviewed in Maksakova et al. 2006, and are therefore not covered extensively in this review). Only recently, a third class of autonomous elements, RTE (Fig. 1), was discovered to have been prolific in the marsupial *Monodelphus domestica* (Gentles et al. 2007; Mikkelsen et al. 2007). In addition to autonomous non-LTR elements, mammalian genomes also contain a number of highly successful parasites of L1 elements, referred to as short interspersed elements (SINEs). These include *Alu* elements in humans and other primates (Fig. 1); B1, B2, and ID elements in rodents; along with a diverse assortment of other SINEs across mammalian orders (Kramerov and Vassetzky 2005). Additionally, there is a second class of parasitic element found in hominid primates, designated SVA, which is much less well understood and remains difficult to categorize according to existing schemes due to its chimeric nature (Fig. 1) (Ostertag et al. 2003).

Although not covered in this review, there is now mounting evidence for appreciable levels of class II DNA transposon activity among some mammalian lineages (Pace and Feschotte 2007; Ray

et al. 2007), overturning the long-held perception that these elements were completely extinguished during early mammalian evolution. The diverse composition and activity of TEs both across and within taxa manifests itself in their differential contribution to mutagenesis and disease. The fraction of de novo mutations arising from the insertional activity of TEs varies from >50% in *Drosophila* (Eickbush and Furano 2002), where they represent ~12% of the total genome (Bartolome et al. 2002), to 0.3% in humans (Deiner and Batzer 1999; Kazazian 1999), where almost half of the genome is composed of TEs (Lander et al. 2001). In comparison, TEs are the cause of 10% of all de novo mutations in laboratory mice (for review, see Maksakova et al. 2006).

While the ability of mobile elements to cause disease via the inactivation of genes by insertional mutagenesis has long been appreciated (Kazazian 1998, 2004), there has been an increasing amount of speculation regarding the overall impact of mobile-element activity on genome evolution. Although the vast majority of TE insertions are either neutral or deleterious to their host, on rare occasions new insertions have led to some form of advantageous—or otherwise noteworthy—phenotypic variation. Some of the most highly publicized discoveries of this nature are the species-specific restriction of HIV infection in owl monkeys (Sayah et al. 2004), the generation of merle coat patterns in dogs (Clark et al. 2006), and the creation of new varieties of grapes with altered pigmentation (Kobayashi et al. 2004). There is also an example of a SVA element carrying downstream genic sequences to new locations, generating a new functional gene family (Xing et al. 2006). As we discuss below, these and other such discoveries have resulted in a shift for some authors from the characterization of TEs as primarily “parasitic” to one wherein TEs are more or less cultivated in the genome for their beneficial possibilities. It remains questionable, however, whether such a shift in perspective is ultimately justified based on either current empirical evidence or theoretical considerations.

Autonomous retrotransposons

LINE-1

The most active autonomous non-LTR element identified within currently sequenced mammalian genomes is the LINE-1 (L1) el-

¹Corresponding author.

E-mail pdeinin@tulane.edu; fax (504) 988-5516.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.5558208>. Freely available online through the *Genome Research* Open Access option.

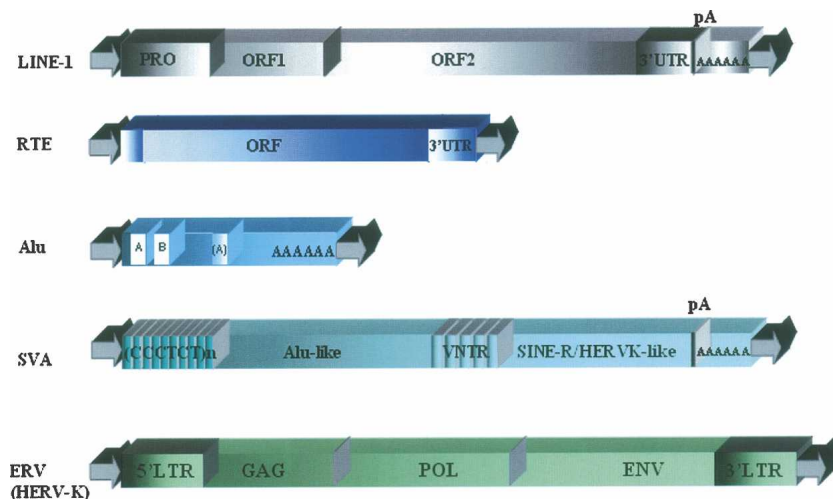


Figure 1. Schematic representation of the genome organization of mammalian retroelements. LINE-1, RTE-1 (retrotransposable element-1), *Alu* (human SINE), and SVA form a group of non-LTR retroelements of which only LINE-1 and RTE-1 are autonomous (i.e., they produce proteins that are required for their retrotransposition); the rest of the members of this group parasitize LINE-1 retrotransposition machinery. LINE-1 elements are composed of the 5' UTR that contains an internal RNA polymerase II (pol II) promoter (PRO) and two open reading frames (ORF1 and ORF2) that are separated in mouse and human L1s but are overlapping in the rat LINE-1. Adjacent to ORF2 is the 3' UTR that contains a polyadenylation (pA) signal and ends in a stretch of adenine residues (pA tail) of variable length. The majority of the disease-causing L1 inserts contain rather long poly(A) tails. Gray arrows bordering each full-length element represent target-site duplications (TSD) that result from amplification of the integration site during each retrotransposition event. RTE elements contain a single ORF that encodes endonuclease and reverse transcriptase activities but in contrast to L1 ORF2 has no *cis*-domain. The RTE ORF is flanked by the 5' and 3' UTRs. *Alu* elements do not code for any proteins. They are driven by a RNA polymerase III promoter, shown as boxes A and B, contain a middle A-rich region (A), and end in a poly(A) tail. SVA elements are a recently discovered hominid-specific group of retroelements that are most likely transcribed by a pol II promoter. They contain a variable number of (CCCTCT) repeats followed by an *Alu*-like sequence, a VNTR region, and a sequence of retroviral origin (HERV-K) that was historically designated "SINE-R"; hence the name SVA. At the very 3' end, SVA contains a poly(A) signal preceding a poly(A) tail. Little is known about their mechanism of mobility. ERVs are endogenous retroviruses that contain long terminal repeats (5' and 3' LTRs) flanking sequences that produce proteins necessary for mobilization (gag, pol, env, etc.).

ement. Transcription of L1 generates a retrotranspositionally competent, full-length L1 mRNA (Skowronski and Singer 1985; Dombroski et al. 1991) and a spectrum of processed L1-related RNA products, the majority of which are not capable of retrotransposition (Fig. 2) (Perepelitsa-Belancio and Deiner 2003; Belancio et al. 2006). The full-length L1 mRNA (6 kb and 7.5 kb long for human and mouse L1s, respectively) is bicistronic (containing open reading frames 1 [ORF1] and 2 [ORF2]) (Skowronski et al. 1988), a feature that is uncharacteristic for eukaryotic gene expression. L1 transcription, which is driven by an unconventional RNA polymerase II promoter residing in the beginning of the 5' untranslated region (UTR) (Swergold 1990; Severynse et al. 1992), is influenced by the upstream genomic sequences (Lavie et al. 2004). The presence of the L1 promoter within the RNA coding region is reminiscent of an RNA polymerase III promoter (Kurose et al. 1995). However, its length, coding capacity, and processing are strong indicators that RNA polymerase II is the primary vehicle for L1 expression. Experimental evidence indicates that at least some portion of L1 RNAs is capped (Athaniar et al. 2004). Capped, 5' UTR-containing L1 mRNAs support levels of translation initiation similar to those detected for the highly efficient beta-actin 5' UTR (Dmitriev et al. 2007). Polyadenylation at the end of the L1 3' UTR seems to be relatively independent of the downstream genomic sequences, but its efficiency improves with increased length of the poly(A) tail (Belancio et al. 2007).

The product of the first ORF possesses nucleic acid chaperone activity and forms trimers that bind to ~50 nt of L1 mRNA in vitro to form ribonucleoprotein (RNP) complexes (Martin et al. 2003, 2005; Basame et al. 2006). RNPs are considered to be retrotranspositional intermediates. Some forms of 3'-end containing L1 RNAs, as well as L1 ORF1 and ORF2 proteins, cofractionate with the polyribosomal portion of the cytoplasm, suggesting that an early form of the RNP is most likely generated at the polyribosome (Kulpa and Moran 2005). While these RNP complexes are determined to be necessary—but not sufficient—for retrotransposition (Kulpa and Moran 2005), there is currently very little information on how these RNPs dissociate from the ribosomes and what form they take in returning to the nucleus (Kubo et al. 2006). RNP formation may serve multiple functions, such as protection of L1 mRNA from degradation, termination of translation, removal of the L1 mRNA from the ribosomes to allow entry into the nucleus, and/or prevention of reverse transcription initiation within internal L1 sequences instead of at the 3' end (Kulpa and Moran 2006).

Even though the requirement of ORF2 for the retrotransposition process has been established (Moran et al. 1996), the mechanism of ORF2 expression from the full-length L1 mRNA is still not completely understood, and there may be differences between the mouse and human elements (Alisch et al. 2006; Li et al. 2006; Dmitriev et al. 2007). Human L1 ORF2 protein can also potentially be made from the splice product that lacks the entire ORF1 sequence but retains the intact ORF2 (Fig. 2) (Belancio et al. 2006). While ORF2 expression from a transcript other than the full-length L1 mRNA would most likely have no influence on the L1 retrotransposition due to the *cis*-preference (Moran et al. 2000), it may be able to drive SINE mobilization and contribute to DNA damage (Gasior et al. 2006). The ORF2 protein contains endonuclease (EN) (Feng et al. 1996) and reverse transcriptase (RT) (Martin et al. 1998) activities, as well as a Cys-rich domain, which are absolutely required for retrotransposition in wild-type cell lines (Moran et al. 1996). In contrast, some DNA repair deficient cell lines allow endonuclease-independent integration of mutant L1 elements (Morrish et al. 2002, 2007). However, simply increasing the amount of the double-strand breaks (DSBs) or nicks in the cellular DNA does not favor EN-independent insertions (El Sawy et al. 2005; Farkash et al. 2006) even though endonuclease-independent insertions—or at least noncanonical insertions that have the appearance of being endonuclease-free—occur with detectable frequency in nature (Sen et al. 2007).

The current model for canonical L1 insertion events, which requires functional ORF1 and ORF2 proteins (Moran et al. 1996; Martin et al. 2005), features the introduction of the first-strand

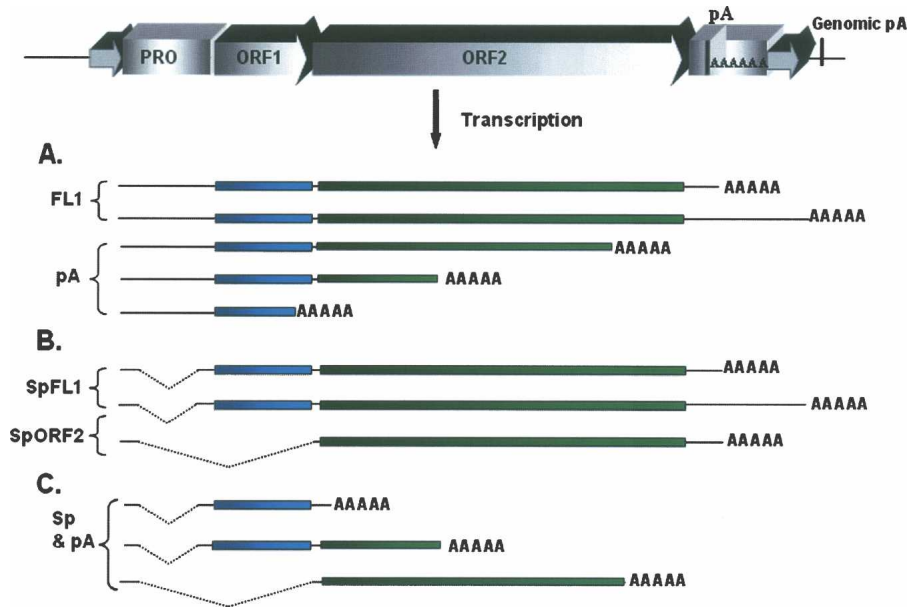


Figure 2. Processing of LINE-1 transcripts. (A) Schematic of two variants of the full-length retrotranspositionally competent L1 (FL1) mRNAs: One ends at the L1 encoded poly(A) signal, the other terminates at a poly(A) site within genomic DNA located downstream of the L1 sequence. These mRNAs represent a small fraction of the products that are made during L1 transcription due to premature polyadenylation of L1 transcripts at the internally positioned pA sites (pA products). (B) In addition to premature polyadenylation, L1 transcripts are also extensively spliced. This processing results in L1 mRNAs that contain both ORFs (SpFL1) and can retrotranspose at low frequency. L1 mRNA splicing also creates a transcript (SpORF2) that has a potential to produce only functional ORF2 protein independent of the full-length L1 mRNA. (C) The majority of the L1 transcripts are differentially spliced and prematurely polyadenylated (Sp & pA products).

nick by the EN activity of ORF2 at the L1 recognition site (Fig. 3). The exposed 3' end of the cellular DNA is proposed to base-pair with the L1 poly(A) tail (Symer et al. 2002) and to serve as a primer for the first-strand DNA synthesis by the L1 ORF2 RT activity. This process is also referred to as target-primed reverse transcription, TPRT (Luan et al. 1993). The origin of the second nick in the cellular DNA and the final steps of L1 integration remain elusive. Transient transfection of functional L1 elements into mammalian cells results in the production of DSBs in an L1 EN-dependent manner (Belgnaoui et al. 2006; Farkash et al. 2006; Gasior et al. 2006) indicating that the second-strand nick might be a product of L1 EN activity.

Some aspects of L1 retrotransposition resemble the integration steps of the *Bombyx mori* R2 element, which serves as an important model system for unraveling of mechanistic nuances of non-LTR retroelements. Recent studies have demonstrated that R2 elements require the association of two identical subunits of the R2 protein with different DNA sequences upstream and downstream of the insertion site for successful retrotransposition (Christensen and Eickbush 2005; Christensen et al. 2005). The authors proposed that L1 ORF2 protein might also form a dimer that is involved in the second-strand DNA cleavage. However, there are currently no experimental data excluding the possibility that the second-strand cleavage required for L1 insertion may arise from alternative cellular sources.

The understanding of the integration process was further advanced by the discovery that, in addition to driving RNA-dependent DNA synthesis, the R2 ORF protein also possesses strand-displacement and DNA-dependent DNA polymerase ac-

tivities (Kurzynska-Kokorniak et al. 2007). This finding suggests a possibility that L1 ORF2 protein may possess similar properties.

L1 retrotransposition generates a combination of full-length, 5'-truncated, spliced and/or partially rearranged copies of the parental element (Ostertag and Kazazian 2001b; Belancio et al. 2006). Truncated insertion events are proposed to be the result of either low L1 RT processivity or base-pairing between the insertion site and internal L1 sequences (Ostertag and Kazazian 2001b; Symer et al. 2002; Zingler et al. 2005), although active disruption of the L1 integration by host proteins cannot presently be excluded (Gilbert et al. 2005). The resolution of full-length element insertions appears to be less dependent on microhomology than does the resolution of 5' truncated elements, suggesting that the full-length integrations employ a somewhat different insertion process (Zingler et al. 2005).

RTE

In addition to L1, an element belonging to the RTE family of autonomous retrotransposons has been reported in mammals (Malik and Eickbush 1998), most recently in the marsupial *M. domestica*.

Originally characterized in *Caenorhabditis elegans* (Youngman et al. 1996), the structure of the RTE consists of a single ORF, which codes for a protein with endonuclease and reverse transcriptase activity. The RTE ORF appears most closely related to the corresponding ORF of the CR1 autonomous element, which is found in avian and reptile genomes (Fig. 1) (Malik and Eickbush 1998).

As opposed to L1 elements, the 5' and 3' UTR sequences of RTE appear to be shorter in length, and the 3' UTR exhibits more variation than is found in other non-LTR retrotransposons. In contrast, the target-site duplications associated with RTE insertion events—ranging from 18 to nearly 1400 bp—are considerably longer than those found with other characterized non-LTR retrotransposons (Malik and Eickbush 1998). During its proliferation, the RTE clade of elements have given rise to numerous nonautonomous SINE lineages, many of which appear to be derived from truncated forms of the autonomous element (Malik and Eickbush 1998). Little is currently known about the molecular biology surrounding RTE expression or retrotransposition.

Nonautonomous retrotransposons

SINEs

SINEs, or short interspersed elements, are represented by *Alu* in primates, B1, B2, and ID elements in rodents, as well as a myriad of other families in other mammalian and non-mammalian genomes (for reviews, see Deininger and Batzer 1993; Kramerov and Vassetzky 2005). SINEs were originally defined purely by their length (75–500 bp) and interspersed nature, but over time it

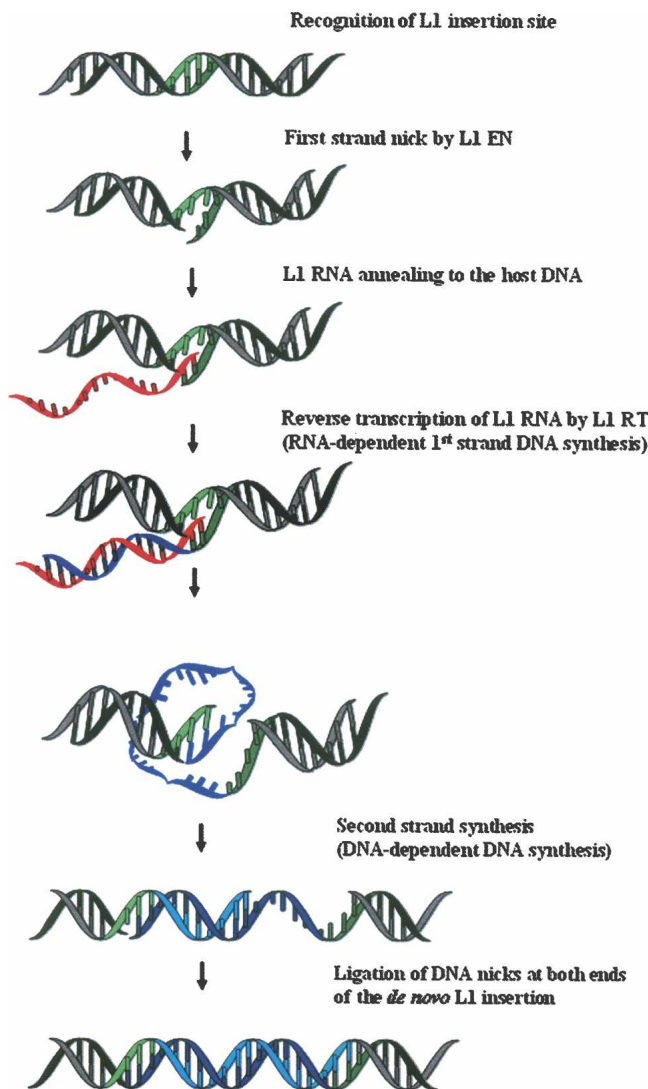


Figure 3. Steps of the LINE-1 integration process. The L1 endonuclease domain encoded by the ORF2 protein loosely recognizes a consensus 5'-TTTTAA-3' sequence (shown in green) in the genomic DNA and introduces a first-strand nick between the T and A nucleotides of the minus strand. The resulting free 3' end of the host DNA is proposed to base-pair with the poly(A) tail of the L1 mRNA (shown in red) and serves as a primer for the first-strand cDNA synthesis (shown in blue) by the L1 reverse transcriptase that uses L1 mRNA as the template. This process is known as a target-primed reverse transcription (TPRT). Mechanistic details of the rest of the L1 integration process are not well defined yet. At some point during L1 integration, either L1 ORF2 or a cellular activity introduces a nick into the plus strand and the structure is resolved to utilize the 3' end as a primer for the second-strand DNA synthesis (shown in light blue) by an unknown polymerase activity. Finally, the two nicks in the cellular DNA are repaired to complete the L1 integration event.

is has also become generally accepted that they are further characterized by RNA polymerase III transcription (Deininger et al. 2003); this latter distinction serves to differentiate them from other short interspersed repeats, such as microsatellites and MITEs. *Alu* elements are composed of two related monomers, ancestrally derived from 7SL RNA, which are separated by a middle A-rich region. B1 repeats in rodents are also derived from 7SL RNA. Although one example of a 5S ribosomal derived SINE has been observed in zebrafish (Kapitonov and Jurka 2003), the

vast majority of SINEs in both mammalian and nonmammalian genomes are derived from tRNA genes. Despite their ancestry, however, the relationship to a particular tRNA sequence cannot always be discerned, and the original clover-leaf secondary structure of the tRNA gene does not generally appear to be retained (for review, see Kramerov and Vassetzky 2005).

At their 3' end, the majority of mammalian SINEs end in a run of simple sequence, most frequently a poly(A) tail, which is required for retrotransposition (Fig. 1) (Roy-Engel et al. 2002; Dewannieux and Heidmann 2005). In addition, there is evidence that a subset of nonmammalian SINEs makes use of alternative sequences at this position (Kajikawa and Okada 2002), which are homologous to the 3' regions of LINE elements in their host species. The 5' portion of SINEs contains an internal RNA polymerase III (Pol III) promoter whose activity can be enhanced by upstream genomic sequences (Chesnokov and Schmid 1996; Roy et al. 2000). SINE transcripts are typically heterogeneous in size because they possess no internal Pol III termination sequences and instead use termination sequences (four or more T residues) located at random distances downstream from the elements.

Because of the heterogeneous nature of *Alu* transcripts, there are only limited data on their transcription, and these are primarily from transformed cell lines (Paulson and Schmid 1986; Sinnott et al. 1992; Liu et al. 1994; Shaikh et al. 1997; Tang et al. 2005). The majority of *Alu* transcripts in mammalian cells are generated by older *Alu* subfamilies (Sinnott et al. 1992; Shaikh et al. 1997) that demonstrate little or no capacity for retrotransposition, as gauged by the de novo disease insertion characteristics (Johanning et al. 2003). Although the abundant representation of older *Alu* subfamilies among detected *Alu* transcripts is most likely due to the high copy numbers of these families, the disparity between the *Alu* diversity observed among transcripts and that detected among actual retrotransposition events has generally been interpreted as evidence that RNA levels per se are not the predominant limiting factor for SINE retrotransposition. While these data do suggest post-transcriptional selection on transcripts (Sinnott et al. 1992), we presently have very limited knowledge concerning the diversity of *Alu* transcripts in the appropriate germline or early embryonic context.

It was originally proposed based on bioinformatic analysis (Roy-Engel et al. 2002; Odom et al. 2004), and later confirmed experimentally (Dewannieux and Heidmann 2005), that the efficiency of *Alu* and other SINE retrotransposition is influenced by the length of the poly(A) tail. Current evidence (Dewannieux and Heidmann 2005) does not support this mechanism as the primary explanation for the several orders of magnitude difference in retrotransposition ability between old and young *Alu* subfamilies (Roy-Engel et al. 2002), implying that additional factors must control the efficiency of SINE retrotransposition.

Although *Alu* and the other SINEs are clearly dependent on L1 elements for their activity, they do not appear to require the assistance of ORF1 (Dewannieux et al. 2003). Expression of L1 elements with mutant ORF1 or production of ORF2 by itself allows relatively active *Alu* mobilization in ex vivo assays (Dewannieux et al. 2003). We note, however, that no studies to date have completely accounted for the possible influence of endogenously expressed ORF1 in these experimental systems. Additionally, the possibility that ORF1 may augment the efficiency of *Alu* retrotransposition has not been explored.

It has also become clear that SINEs are frequently exapted by host genomes for use in important functional roles. A recent analysis indicates that as much as 20% of known human endog-

Table 1. Insertions and disease

Insertion	Locus	Chr	Disease
Alu	3 × F8 ^{a,b}	X	Hemophilia B
	2 × F9 ^b	X	Hemophilia A
	2 × CLCN5 ^b	X	Dent's disease
	2 × BTK ^b	X	X-linked agammaglobulinemia
	IL2RG ^a	X	X-linked severe combined immunodeficiency
	GK ^a	X	Glycerol kinase deficiency
	ABCD1 ^b	X	Adrenoleukodystrophy
	ATP7A ^c	X	Menkes disease
	CD40LG ^d	X	Hyper-immunoglobulin M syndrome
	CRB1 ^b	1	Retinal blinding
	SERPINC1 ^b	1	Type 1 antithrombin deficiency
	ZEB2 ^b	2	Muckle-Wells syndrome
	MSH2 ^b	2	Hereditary non-polyposis colorectal cancer
	CASR ^a	3	Hypocalciuric hypercalcemia and hyperparathyroidism
	BCHE ^a	3	Cholinesterase deficiency
	HESX1 ^b	3	Aplasia anterior pituitary
	MLV12 ^a	5	Associated with leukemia
	2 × APC ^a	5	Hereditary desmoid disease
	NT5C3 ^e	7	Chronic hemolytic anemia
	2 × CFTR ^f	7	Cystic fibrosis
	EYA1 ^a	8	Branchio-oto-renal syndrome
	LPL ^g	8	Lipoprotein lipase deficiency
	CHD7 ^h	8	CHARGE syndrome
	POMT1 ⁱ	9	Walker Warburg syndrome
	FAS ^b	10	Autoimmune lymphoproliferative syndrome
	3 × FGFR2 ^a	10	Apert syndrome
	SERPING1 ^a	11	Complement deficiency
	HMBS ^b	11	Acute intermittent porphyria
CMAS ^j	12	Human-specific evolutionary change	
MLV12	12	Mucopolidosis type II	
2 × BRCA2 ^a	13	Breast cancer	
BRCAT ^a	17	Breast cancer	
NF1 ^a	17	Neurofibromatosis	
L1	CHM ^b	X	Choroideremia
	2 × CYBB ^b	X	Chronic granulomatous disease
	4 × DMD ^{b,k}	X	X-linked Duchenne muscular dystrophy
	3 × F8 ^b	X	Hemophilia A
	2 × F9 ^b	X	Hemophilia B
	RP2 ^b	X	X-linked retinitis pigmentosa
	RPS6KA3 ^b	X	Coffin-Lowry syndrome
	APC ^b	5	Colon cancer
	FKTN ^b	9	Fukuyama-type congenital muscular dystrophy
	HBB ^b	11	Beta-thalassemia
	PDHX ^l	11	Pyruvate dehydrogenase complex deficiency
SVA	BTK ^b	X	X-linked agammaglobulinemia
	LDLRAP1 ^b	1	Autosomal recessive hypercholesterolemia
	SPTA1 ^b	1	Hereditary elliptocytosis and hereditary pyropoikilocytosis
	FKTN ^b	9	Fukuyama-type congenital muscular dystrophy

^aDeininger and Batzer 1999; ^bChen et al. 2006; ^cGu et al. 2007; ^dApoil et al. 2007; ^eManco et al. 2006; ^fChen et al. 2008; ^gOkubo et al. 2007; ^hUdaka et al. 2007; ⁱBouchet et al. 2007; ^jHayakawa et al. 2001; ^kOstertag and Kazazian 2001a; ^lMiné et al. 2007.

enous microRNAs are driven by *Alu* Pol III promoters (Borchert et al. 2006). There are also accumulating examples of evolutionary conservation of portions of ancient SINE insertions both in non-coding and coding regions of the genome (Bejerano et al. 2006; Kamal et al. 2006; Nishihara et al. 2006). In one notable instance, a B2 SINE element was found to serve as a boundary element for transcriptional regulation during organogenesis (Lunyak et al. 2007). This example highlights a critical conceptual distinction, however, between the exaptation of an *individual* TE insertion at a *particular* locus for a functional role and the maintenance of an entire TE lineage's activity for the benefit of the organism. While the exaptation of TE instances has been definitively demonstrated, maintenance of TE *lineage* activity for the sake of the

organism's fitness is currently lacking in both empirical and theoretical support.

SVA

SVA elements represent another class of non-autonomous elements active in the human genome. As with *Alu*, they appear to rely on the L1 retrotransposition machinery for mobilization (Ostertag et al. 2003). There are currently only ~3000 copies of SVA in the human genome, which, combined with their phylogenetic distribution, suggests their relatively young, hominid-specific (Wang et al. 2005) age compared to L1 and *Alu* elements (Lander et al. 2001; Ostertag et al. 2003). The overall structure of the SVA element is chimeric in nature: beginning at the 5' end, SVA elements are composed of a (CCCTCT)_n hexamer repeat region; a region of two antisense *Alu* fragments adjacent to additional sequence from an unidentified source; a variable-number tandem repeat (VNTR) region made of copies of a 35- to 50-bp sequence; a sequence under 500 bp long derived from the 3' end of the *env* gene and the 3' long terminal repeat (LTR) of the endogenous retrovirus HERV-K10; and a poly(A) tail positioned downstream of the predicted conserved polyadenylation signal hexamer AATAAA (Fig. 1) (Ostertag et al. 2003). The VNTR region of SVA elements varies in length from 48 to 2300 bp (Wang et al. 2005). The exact mode of SVA transcription is not known, but it is speculated to be mediated by RNA polymerase II based on the presence of the poly(A) signal as well as numerous runs of Ts distributed throughout the sequence that ordinarily exclude read-through by RNA polymerase III. Similar to L1 elements (Perepelitsa-Belancio and Deininger 2003; Belancio et al. 2006), the SVA sequence contains several predicted internal polyadenylation signals and splice sites; however, no experimental evidence exists to date that confirms their function or explores their potential impact on gene expression. Much like L1 and *Alu* elements, SVA elements demonstrate ongoing activity in humans and sporadically cause disease by randomly inserting into mammalian genes (Table 1) (Ostertag et al. 2003; Wang et al. 2005). Similar to L1 elements (Moran et al. 1999; Babushok et al. 2007), SVA elements transduce genomic sequences adjacent to their 3' ends, a process that is reported to occasionally create new genes (Xing et al. 2006).

Host impact

Insights from comparative genomic studies

The ever-increasing availability of large-scale genomic sequence data has provided considerable insight concerning the important

role TEs have played in building and shaping eukaryotic genomes. The sequence from a representative genome is at best a snapshot of what is ultimately a dynamic process of current and historical mobile-element activity in a lineage. The vast majority of insertion events will be lost from the host population over time and not detected in any extant genome; under neutrality, only $1/(2N_e)$ insertions are expected to reach fixation in the population (N_e represents the effective population size). Those inserts bearing negative fitness consequences will be purged more readily from the population. As a result of unequal selective pressures across the genome, the patterns of element distribution observed via genome sequencing will not necessarily coincide with the initial pattern of insertions. For example, although L1 elements in humans are generally found in gene-poor AT-rich regions, data from cell culture suggest that this may be an effect of post-insertional processes rather than an initial insertion preference dictated by L1 biology (Ovchinnikov et al. 2001; Gilbert et al. 2002; Graham and Boissinot 2006; Gasior et al. 2007). Similar results were obtained when L1 integration events were characterized in transgenic animals (An et al. 2006; Babushok et al. 2006), suggesting that the discrepancy in the distribution between new and old L1 integration events is likely to arise from post-insertional selection pressures. Even though examined de novo TE insertions associated with disease alleles, as well as those acquired through cell culture and/or in vivo assays, have their own share of potential biases, the majority of these data appears to indicate that, prior to significant evolutionary selection, L1 insertions exhibit little to no favoritism for any particular genomic composition in their distribution, at least at scales beyond the immediate region surrounding the insertion site. Thus, the 2000 L1 elements and 7000 *Alu* elements that are specific for humans only represent the distributions of those elements that have undergone significant selection and genetic drift (Hedges et al. 2004; Chimpanzee Sequencing and Analysis Consortium 2005; Mills et al. 2006).

Non-LTR distribution and genomic burden

From a genomics perspective, the most obvious consequence of TE activity during mammalian evolution is the enormous amount of genetic real estate they have come to occupy. Mammals are not unique in this regard; TEs also inhabit substantial portions of the genomes of several distantly related taxa, including wheat (*Triticum monococcum* L.) and rice (*Oryza sativa*) (Wicker et al. 2001; International Rice Genome Sequencing Project 2005). Among those mammals whose genomic sequence has been obtained thus far, *M. domestica* appears to have the largest genomic TE content, approaching 52% of the genome (Mikkelsen et al. 2007), while human estimates approach 45% (Lander et al. 2001). On the lower end of the mammalian spectrum, the recently sequenced cat genome exhibits ~30% repetitive sequence (Pontius et al. 2007). All indications, however, point to these various estimates being fairly conservative, with considerable amounts of older repetitive sequences being undetectable due to their divergence from consensus sequences used for detection.

While a number of views exist regarding the accumulation and associated consequences of copious amounts of repetitive sequence in many higher eukaryotes, a strong case has been made that the observed increases in both TE and intron content with decreasing population sizes in eukaryotes is a consequence of reduced effective population size (Lynch 2002; Lynch and

Conery 2003). In this scenario, the decline in selection efficiency is the direct result of lower population sizes for organisms occupying higher trophic levels. Smaller population sizes decrease the ability of purifying selection to remove extraneous genomic material conferring small deleterious effects. This latter argument is in accord with the “parasitic” paradigm of TE–host dynamics that currently prevails in the mobile-element field. Further corroborating this view is the reported correlation shown between genome size and risk of extinction among plant taxa (Vinogradov 2003), although the picture for vertebrates is less straightforward (Vinogradov 2004). That said, it is also likely that some repeat-laden taxa, such as mammals, have developed highly effective mechanisms for dealing with excess genomic repeat content, and these mechanisms (discussed below) may help to explain why a clear relationship between genome size and extinction risk is difficult to establish across vertebrates in general.

While the presence of TEs represents a unifying theme among eukaryotic genomes, considerable differences exist in both the type and frequency of various TE lineages across divergent taxa, suggesting that several possible “genome ecologies” may exist (Brookfield 2005; Le Rouzic et al. 2007). Regarding L1 retrotransposons in particular, the mammalian complement stands out from other vertebrate taxa examined thus far in that it is both more numerous in copy number and less diverse in nature (Furano et al. 2004; Pritham and Feschotte 2007). Compared to *Danio rerio*, for instance, there are significantly fewer long-lived L1 lineages in mammals, although the copy numbers present in the latter taxa are greatly increased (Furano et al. 2004). It currently remains unclear as to what point during tetrapod evolution this dramatic shift in TE ecology occurred, as well as whether population dynamics, changes in host cell biology, or both, precipitated the transition. Analysis of available sequence data from a number of Deuterostomia genomes suggests that much of the observed loss of diversity took place during synapsid evolution (Kordis et al. 2006).

The task of drawing large-scale evolutionary trends in vertebrate TE diversity remains hampered by insufficient sequence sampling for many relevant taxa. The potential remains that additional sequencing results may substantially alter our understanding of mobile-element diversity and distribution. For example, despite a long-held view that DNA transposons have exhibited little activity in extant mammals, recent analyses have revealed extensive amplification of type II DNA transposons in the bat genus *Myotis* (Pritham and Feschotte 2007; Ray et al. 2007), as well as within early primate evolution (Pace and Feschotte 2007). Furthermore, analysis of the first marsupial genome (*M. domestica*) indicates a more diverse TE complement than is evident in humans, rodents, or dogs (Gentles et al. 2007). Thus the distribution of TE among mammals is by no means monolithic, and additional remnants of ancestral TE diversity may yet thrive in unexplored genomes.

In contrast to mammals as a whole, the recent history of L1 evolution in the rodent and primate lineages has been well documented. In primates, L1 diversity seems to have been trimmed down from three ancestral lineages to a single lineage (Furano et al. 2004). Within this single remaining primate lineage, a curious pattern of one L1 family begetting a single lineage before becoming inactive clearly emerges from the genomic data. This phenomenon has been proposed to result from the competition among L1 instances for scarce host factors necessary for retrotransposition (Khan et al. 2006). However, it also seems plausible that there may exist, at any given time, such a low level of active

elements in the population (Brouha et al. 2003) that, coupled with lower host organism population size among land vertebrates than their marine predecessors, they are subjected to profound bottleneck effects that effectively extinguish many TE lineages. Once older (and more diverse) lineages are lost in this fashion, the accumulation of diversity over time would be required to begin anew.

As might be expected, SINE diversity generally tracks LINE diversity within a given taxa, and there are instances reported of SINEs declining to extinction along with their associated LINE sequence (Casavant et al. 2000; Grahn et al. 2005; Rinehart et al. 2005). The most detailed information regarding recent SINE activity thus far comes from the human and chimpanzee genomes, where the sequence substructure of *Alu* elements is understood in considerable detail. Within these taxa, there appears to be a two- to threefold increase in the overall amplification and/or accumulation rate of *Alu* SINE elements in human vs. chimpanzee lineages (Hedges et al. 2004; Mills et al. 2006). In addition, the diversity and relative activity rates of active *Alu* subfamilies differ appreciably between these closely related taxa (Hedges et al. 2004; Mills et al. 2006). Human–chimpanzee genome comparisons confirmed earlier studies based on transcript and polymorphism analysis (Salem et al. 2003; Otieno et al. 2004), which indicated that only a tiny fraction of the potential *Alu* repertoire in the genome exhibits retropositional competence.

Insertional mutagenesis and disease

The ongoing potential within mammalian genomes for the de novo insertion of mobile elements into both coding and regulatory regions brings with it obvious deleterious consequences to the host organism. While the total fitness cost of various levels of mobile-element activity is difficult to quantify due to a general lack of knowledge concerning the actual distribution of mutational fitness effects (e.g., fraction of dominant vs. negative mutations), the deleterious consequences of insertional mutagenesis are readily apparent in numerous disease phenotypes. It appears that an intricate and possibly tenuous balance between the activity of mobile elements and the ability of the host genome to tolerate insertional and recombinational mutagenesis has been established over evolutionary time. When this balance is disturbed, such as in transgenic mice engineered to ubiquitously express the highly active mobile element, Sleeping Beauty, mice exhibited markedly elevated levels of embryonic lethality and postnatal cancer development (Dupuy et al. 2005).

Examples of human diseases caused by L1-driven mobile elements continue to accumulate with over 50 reported instances to date (Table 1). The level of mobile-element insertions observed to cause human disease has been suggested to represent ~0.3% of all human mutations (Deininger and Batzer 1999; Kazazian 1999). This indicates approximately one insertion in every 20–100 live births (Deininger and Batzer 1999; Kazazian 1999; Cordaux et al. 2006). Examples of mobile-element insertions in mice have been previously reviewed, demonstrating that they experienced relatively low levels of L1- and SINE-related insertions compared to insertions by endogenous retroviruses (Ostertag and Kazazian 2001a; Maksakova et al. 2006). There remains a great deal of uncertainty concerning the number of retrotranspositionally competent L1s in the mouse genome, with estimates ranging from 12 to 3000 (DeBerardinis et al. 1998; Mouse Genome Sequencing Consortium et al. 2002). Overall, however, mice have a 100 times higher rate of spontaneous mutations

caused by mobile elements (primarily LTR elements) (Maksakova et al. 2006) than humans (all non-LTR elements).

One of the more striking observations regarding the characteristics of L1 disease-causing insertions—and to a lesser extent the *Alu* and SVA disease insertions (Table 1)—is the high proportion of these insertions found on the X chromosome. In humans this represents 13 out of 17 disease-generating L1 insertions. One expects a certain ascertainment bias due to the fact that genetic defects in the X chromosome are often dominant in males; bias also no doubt stems from the number of X chromosome-related diseases have been heavily scrutinized for causal mutations. L1 elements are present at a higher density on the X chromosome than any other chromosome (Lander et al. 2001), and it remains possible that some of this density increase may represent insertion preference. However, there was no detectable enrichment of L1 element insertions on the X chromosome in tissue culture studies that directly measures de novo L1 insertions (Graham and Boissinot 2006; Gasior et al. 2007), and the enrichment on the X chromosome has largely been ascribed to lower recombination rates and corresponding reduction in the efficiency of negative selection among the sex chromosomes (Boissinot et al. 2001). In contrast, 31 endogenous retroviral insertions in mice (Ostertag and Kazazian 2001a; Maksakova et al. 2006) have been detected, and none of those insertions was found to reside on the X chromosome. Perhaps more surprising, out of six L1 insertions causing mouse mutations (Ostertag and Kazazian 2001a; Chen et al. 2006), there are also no insertions on the X chromosome. There remains the possibility that additional biases are involved in the mutational screening process and/or inbred mouse lines may be part of the skewed representation for mouse L1s.

In order to help assess the extent an X-chromosomal ascertainment bias might have impacted the number of insertion detections in humans, we analyzed the portion of all mutations in the Human Genome Mutation Database (HGMD) occurring on the X chromosome (Krawczak et al. 2000). We find that roughly 17.9% of HGMD mutations were detected on the X chromosome (out of a total of 73,411 recorded mutations). When considering the fact that 5.1% of the total DNA in women (2.6% in men) consists of the X chromosome, and given a roughly 4% sex-averaged X-chromosomal DNA content, the HGMD data suggest a 4.5-fold detection bias in X chromosome disease mutations overall. A similar result is obtained when considering total known gene counts. However, we observe a 20-fold enrichment of L1 disease insertions on the X compared to what is expected. Thus, even taking into consideration the expected elevation of X chromosome mutation detections, we cannot fully account for enrichment of L1 disease insertions. One possible explanation for this phenomenon is that the reduced recombination rate on the X chromosome decreases the number of catastrophic megabase-scale rearrangements that might occur for L1 insertions on the autosomes (see “Post-insertional mutagenesis,” below). Thus, while many additional disease-causing insertions do occur on the autosomes, a large fraction are ultimately lost due to recombination events which convert them from recessive alleles to lethal large-scale rearrangements, which are then immediately purged from the population. We wish to be careful here to distinguish the above-mentioned process from the Muller’s ratchetlike scenario previously proposed (Boissinot et al. 2001, 2006) to account for the increased L1 accumulation on the sex chromosomes. While both phenomena are related to recombination frequency, they are distinct (but not mutually exclusive).

Insertion-mediated deletions caused by *Alu* and L1

In addition to direct disruption of genetic information through mobile-element insertion, a related phenomenon has also been observed, wherein the insertion of mobile elements is associated with the deletion of adjacent genomic sequence. In some cases these events, designated *Alu*-retrotransposition-mediated deletions (ARMD) and L1 insertion-mediated deletions (LIMD), can lead to the removal of entire exons (Callinan et al. 2005; Han et al. 2005). In the case of the *Alus*, ~33 deletion instances were detected between the humans and chimpanzees lineages. For L1, ~50 such deletion events were discovered, collectively removing 18 and 15 kb from the human and chimpanzee genomes, respectively (Han et al. 2005). Of course, for both *Alu* and L1, the amount of DNA removed via these processes represents only a fraction of that added by the retrotransposition process itself. Nevertheless, along with recombination-mediated deletion and other deletion processes, these deletions may assist in curbing the expansion of genome size associated with the retrotransposition process.

Post-insertional mutagenesis

One could argue that the most negative fitness consequence arising from the proliferation of TE sequences within mammalian genomes involves the increased potential for mutagenic nonallelic homologous recombination (NAHR) events. Such mutagenic recombinations can result in a diverse array of genetic rearrangements ranging from small-scale genomic deletions, to chromosomal inversions, to interchromosomal translocations (for review, see Hedges and Deininger 2007). A much larger fraction of TE-related diseases in humans results from recombinational mutations than from insertional mutations (Deininger and Batzer 1999).

Although NAHR events occur at appreciable rates and have been implicated in a number of human genetic diseases, the overall level of ectopic recombinations and associated phenomena appears to be lower among TEs in mammals than in other taxa, such as yeast (Johnson and Jasin 2000). This may be the direct consequence of the modulation of DNA repair pathways in the mammalian lineage to avoid using nonallelic (i.e., TE) templates for repair. Unlike yeast, mammals more readily employ non-recombination-based repair mechanisms, which while often resulting in the loss of some genetic information, nevertheless avoid the potentially calamitous effects of large-scale rearrangements associated with homologous repair-associated pathways (Johnson and Jasin 2000). It is currently unclear whether this shift in DNA repair mechanism preference over eukaryotic evolution has been a capitulation to, or the instigation of, the expansion of genomic repeats.

Evidence for the deleterious effects of TE-related NAHR comes from multiple sources. As indicated above, >0.3% of human genetic diseases are attributed to nonallelic recombinations (Deininger and Batzer 1999). To date, all such cases have been associated with *Alu*-*Alu* recombination events deleting or duplicating genic regions. Comparative genomic data and population studies, on the other hand, provide the clearest case for negative selection against L1 ectopic recombination events (Song and Boissinot 2007). Full-length L1 elements show a skewed distribution toward non-recombining sex chromosomes, with shorter insertions being under less-stringent selection pressures than longer L1 inserts (Boissinot et al. 2001, 2006). A subsequent analysis examining the population frequency of full-length vs.

near-full-length elements showed that near-full-length elements (lacking promoter sequences crucial to transcription) were in fact indistinguishable from the full-length L1 elements in terms of their fitness effects (Song and Boissinot 2007), lending further support to the notion that the potential for ectopic recombination is one of the major factors behind negative selection against these segregating loci. It is also possible that longer, nearly full-length L1 inserts that are present within introns may interfere with normal gene expression, contributing to the observed negative selection against longer L1 elements.

The apparent disparity between strong ectopic recombination-related selection against L1 (observed through population-based genomic studies) and the exclusion of L1 among reported cases of TE-recombination-related disease mutations may at first seem paradoxical. It is most likely the case that the larger number of *Alus*, as well as their tendency to cluster in and around genic regions, makes the possibility of small-scale, nonlethal (yet ultimately pathological) mutations more likely. The less-prevalent L1 sequences, on the other hand, while more prone to recombine due to their greater lengths of homology (Boissinot et al. 2006; Song and Boissinot 2007), may also be more likely to engage in large-scale genomic rearrangements that result in nonviable offspring. As a consequence, they are not well represented among documented diseases.

In addition to the above data suggesting genomic instability associated with TEs, human-chimpanzee comparisons have also shed light on the background level of neutral recombination events that may be occurring among genomic TEs. A comparison of the human and chimpanzee genomes has indicated ~492 *Alu*-*Alu* recombination events in humans that have collectively removed some 400 kb of sequence, including three exons that are functional in the chimpanzee (Sen et al. 2006). The median length of these deletions was 486 bp, although, as discussed above, these observed genomic data are likely heavily influenced by prior evolutionary filtering.

Post-insertional alterations to mobile elements

We have already discussed how elements may continue to contribute to recombinational mutagenesis long after their initial insertion into the genome. There are also several examples of more subtle genetic disruptions caused by extant elements. The A tails of LINEs and SINEs, for instance, are one of the major sources for the creation of microsatellites in mammalian genomes (Arcot et al. 1995). In at least two reported cases, A-rich regions of human *Alu* elements have evolved into unstable repeats: a triplet causing Friedreich's ataxia (Campuzano et al. 1996) and an unstable pentanucleotide causing spinocerebellar ataxia 1 (Kurosaki et al. 2006).

Alu elements have also been found to contain sequences that can be converted to active splice signals by single point mutations long after their insertion. These point mutations lead to splicing alterations that sometimes cause disease (Vervoort et al. 1998) and certainly cause at least partial disruption of proper splicing of a number of genes, a process termed *Alu* exonization (Sorek et al. 2002). Similar observations have been made for mouse and human L1 elements, which can contribute splice sites that are already existing in their consensus sequences (for more details, see the section "LINE-1 expression," below) (Belancio et al. 2006; Mätlik et al. 2006; Zemojtel et al. 2007). Various scenarios of post-insertional interference of mobile elements with gene expression are summarized in Figure 4 and reviewed in Ka-

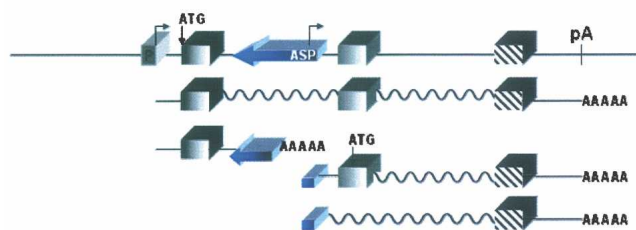
zazian (2004). *Alu* elements present in the introns of unspliced transcripts are, however, subject to extensive RNA editing (Athanasias et al. 2004; Kim et al. 2004). This modification is proposed to serve as a protective measure against the usage of cryptic splice sites and/or nuclear export of unspliced transcripts.

L1 endonuclease and double-stranded DNA breaks

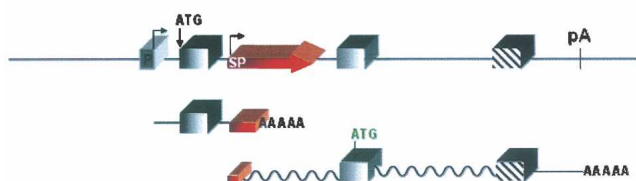
Even though the L1 contribution to genetic instability has long been recognized, insertional and recombinational mutagenesis in the germline have been considered to be the major avenue of impact of human mobile elements on genomic DNA (Kazazian 1998; Deininger and Batzer 1999; Ostertag and Kazazian 2001a; Kazazian and Goodier 2002). Recently it has been confirmed experimentally that the L1 integration process outlined in Figure 3 leads to introduction of double-strand DNA breaks (DSB) (Belg-

naoui et al. 2006; Farkash et al. 2006; Gasior et al. 2006). Double-strand breaks in the cellular DNA result in gamma-phosphorylation of histone 2AX (H2AFX), which can be detected by immunohistochemistry in the form of gamma-H2AFX foci, wherein each focus corresponds to a single DSB (Rogakou et al. 1998). Detection of gamma-H2AFX foci created in response to L1 expression demonstrated that the number of DSBs introduced by transiently transfected L1 element in HeLa cells is at least 10-fold greater than the rate of L1 integration under the same transfection conditions; it was further established that this damage was specific to the L1 endonuclease activity (Gasior et al. 2006). DNA DSBs can be repaired by nonhomologous end joining (NHEJ) or homology-driven repair (HDR) pathways. The main pathway for DSB repair in mammals is NHEJ (for review, see Gorbunova and Seluanov 2005). DSB DNA lesions are highly toxic to mammalian cells, and they often result in mutations when nonconservatively repaired by the cellular DNA repair machinery (Pierce et al. 2001; Vilenchik and Knudson 2003). The discovery of L1-induced DSBs suggests the possibility that L1-associated mutagenesis, in the form of random mutations resulting from the error-prone repair of DSBs, could be greater than insertional mutagenesis. The repaired sites of L1 endonuclease breaks would carry no signs of L1 involvement (such as target site duplications or inserted L1 sequence), and therefore these random mutations would most likely be attributed to the exogenous mutagens, reactive oxygen species, etc. Introduction of DSBs into the cellular DNA may also trigger genetic instability due to nonhomologous recombination. Further compounding this mutagenic potential, recreation of DSBs by L1 elements independent of L1 insertion means that even very low levels of L1 expression (for example, from mutant elements that retain functional endonuclease activity) in germline and somatic cells would result in introduction of damage to cellular DNA that could contribute to cancer initiation and progression as well as age-associated disease. L1 activity is also likely to contribute to cancer via instigation of NAHR among high-density SINE elements, such as *Alu* (Babcock et al. 2003; Gentles et al.

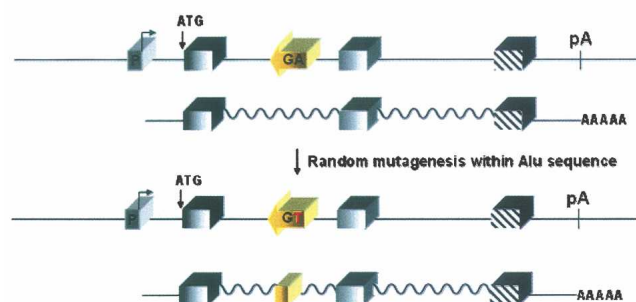
A. LINE-1 insertion in the reverse orientation.



B. LINE-1 insertion in the forward orientation.



C. *Alu* exonization.



D. Interference of L1s inserted upstream of the genes with gene expression.

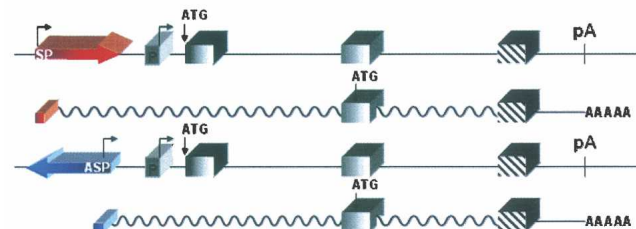


Figure 4. Different effects of mobile-element integration on gene expression. (A) Integration of the full-length L1 element into the intronic region in the opposite orientation relative to the gene transcription can result in premature termination of the cellular transcript due to the usage of the polyadenylation (pA) sites encoded by the L1 antisense strand. Independent of this process, L1 antisense promoter can drive expression of the gene portion located downstream of the L1 insertion site, producing a 5'-truncated mRNA. This transcript may include all of the remaining exons that are accurately spliced, or it may lack some of the exons if the L1 antisense sequence donates a splice donor site that is used with one of the normally utilized splice acceptor sites within the gene. These types of transcripts have the potential to generate proteins by initiating translation at alternative ATG. (B) A similar phenomenon is observed when a full-length L1 element inserts in the same orientation as gene transcription. In this case, usage of the pA sites present in the L1 sense strand leads to the production of prematurely polyadenylated cellular transcripts. At the same time, L1 sense promoter can drive the expression of the remaining portion of the gene resulting in the production of the mRNA that contains L1 5' UTR sequences spliced to the exons located downstream of the L1 insertion site. Functional alternative ATGs in this portion of the RNA can lead to the translation of the truncated protein with a dominant negative effect or a gain of function. (C) *Alu* sequences inserted within introns can be included into mature mRNAs if they gain functional splice donor or acceptor sites via random mutagenesis of their sequence. (D) Integration of the full-length L1 element upstream of the cellular gene may also interfere with the normal gene expression due to the presence of the functional promoter activity (SP, sense promoter; ASP, antisense promoter) and splice sites in both sense and antisense strands of the L1 5' UTR.

2005; for review, see Hedges and Deininger 2007). *Alu*-*Alu* non-allelic recombination has already been established as a significant source of disease-related genomic instability (for review, see Deininger and Batzer 1999). *Alu* amplification relies on L1 retrotransposition machinery, and, as a direct consequence, every *Alu* copy is flanked by a pair of L1 endonuclease recognition sites that have already been successfully utilized by ORF2 at least once in the past. It has been proposed (Babcock et al. 2003; Gasior et al. 2006) that the proximity of these endonuclease targets to homologous interspersed elements predisposes these locations toward involvement in homology-driven DNA repair processes and, consequently, increases the likelihood of genomic instability.

LINE-1 expression

There has been a general assumption that, with few exceptions, endogenous L1 expression was restricted to the germline (Branciforte and Martin 1994; Trelogan and Martin 1995; Ostertag et al. 2002). This was based on early studies which indicated elevated levels of L1 ORF1 protein in mouse germ cells, Leydig cells of embryonic testis, and theca cells of adult ovary, with lesser signals being observed in the few non-germ cell types examined. The only normal cell type reported to be positive for ORF1 expression in early studies was the epithelium of normal mammary glands (Asch et al. 1996). Recently, L1 translation products have been reported at fairly high levels in some somatic tissues, such as the vascular endothelia of human male gonads (Ergun et al. 2004). L1 ORF1 and ORF2 protein expression has also been detected in rat cardiomyocytes and endothelial cells, respectively (Lucchinetti et al. 2006). These reports combined with abundant levels of L1 ESTs detectable across a wide range of somatic tissues (Perepelitsa-Belancio and Deininger 2003) and unpublished data from our own laboratory that support production of endogenous L1 mRNAs in normal human tissues suggest that somatic cells are likely to support varying levels of L1 expression.

Recent studies with transgenic mice have also shown that transgenic L1 elements driven by their endogenous promoter are capable of expression and retrotransposition in neuronal stem cells (Muotri et al. 2005), and it has been suggested that these L1 integration events may alter differentiation plasticity of neuronal stem cells. The latest studies in transgenic mouse models demonstrate that when L1 is expressed in somatic cells, they are capable of supporting very high levels of retrotransposition activity (An et al. 2006; Babushok et al. 2006). Somatic and germline mosaicism for the L1 insertion has also been reported in humans (van den Hurk et al. 2007). However, the requirements for L1 mobilization in normal adult cells in tissue culture remains controversial (Kubo et al. 2006; Shi et al. 2007). One of the most important unanswered questions regarding mammalian L1 elements is the extent to which they contribute to somatic damage via insertional mutagenesis, as well as the other forms of damage described above.

There are several thousand full-length L1 elements in human (Lander et al. 2001) and potentially more in mouse (DeBerardinis et al. 1998). Thus, it is not surprising that their expression cannot be completely suppressed across all tissues. Although there is almost certainly some tissue specificity provided by the *trans*-acting factors used by the L1 promoter (Hata and Sakaki 1997; Yu et al. 2001; Yang et al. 2003; Athanikar et al. 2004), it is generally accepted that methylation is the primary mode of repression for mammalian mobile-element expression (Hata and

Sakaki 1997; Bourc'his and Bestor 2004). The internal promoter used by L1 elements encompasses a CpG island, and they are typically highly methylated (Hata and Sakaki 1997). There have been a number of studies correlating methylation levels, particularly in transformed cells, with L1 expression (Bratthauer and Fanning 1992; Bratthauer et al. 1994). In addition, knockout of the *Dnmt3l* gene in mice, which is a key factor in *de novo* methylation, causes loss of methylation of L1 elements and in LTR elements, resulting in high levels of expression of their RNAs and catastrophic failure of the male germline (Bourc'his and Bestor 2004).

In addition to methylation suppressing initiation of transcription, L1 elements are also subject to several levels of post-transcriptional regulation. It has been suggested that the A-richness of L1 elements may contribute to a general slowdown of transcription (Han et al. 2004). It has also been demonstrated that the vast majority of human and mouse L1 transcripts are subject to either premature polyadenylation (Perepelitsa-Belancio and Deininger 2003) that results in transcripts lacking one or both open reading frames, or to splicing events that also eliminate crucial portions of L1 sequence (Fig. 2) (Belancio et al. 2006, 2008). One of the intriguing aspects of the L1 RNA regulation via processing is that some of the spliced and prematurely polyadenylated transcripts can potentially produce functional L1 proteins in addition to or independent from their translation from the full-length L1 mRNA. The use of splicing as one of the mechanisms of down-regulation of L1 expression has also recently been reported in *Danio rerio* (Tamura et al. 2007). Despite multiple mechanisms working in unison to suppress endogenous L1 expression, potentially functional L1 loci are expressed in human embryonic cells and cancer cells (Skowronski et al. 1988; Garcia-Perez et al. 2007). There is also mounting evidence that RNA-based silencing is employed to control retrotransposon activity in mammals. A recent study in mice demonstrated that a PIWI-related silencing complex harbors numerous L1-targeted RNAs, and mutations in the PIWIL2 (formerly MILI) protein lead to loss of methylation of L1 and IAP elements along with a corresponding increase of expression (Aravin et al. 2007). In addition, there appears to be some capability of RNAi inhibition of L1 activity because of antisense transcripts made from the 5' end of L1 (Soifer et al. 2005; Yang and Kazazian 2006). All of these limitations placed on initial expression, in conjunction with secondary controls on retrotransposition exemplified by the APOBEC3 proteins (see section "Cellular responses to mobile-element activity," below), are likely necessary to keep the negative consequences of L1 activities to a tolerable level.

A consequence of the presence of polyadenylation and splicing signals, as well as the influence of A-richness, is that insertion of an L1 element into a gene, even within an intron, can result in truncated or improperly spliced and/or polyadenylated transcripts of the target gene. Examples of these improper transcripts are found in the human and mouse EST libraries (Speek 2001; Nigumann et al. 2002; Wheelan et al. 2005; Belancio et al. 2006; Mätlik et al. 2006; Zemojtel et al. 2007). In combination with the presence of functional sense and antisense promoters (Swergold 1990; Speek 2001) in the L1 5' UTR, splice sites and polyadenylation signals within full-length L1 insertions have even more potential to interfere with the production of cellular mRNAs in numerous ways (Fig. 4). Previously reported bioinformatic analyses of L1 distribution across various mammalian genomes empirically supported the idea that L1 inserts in the forward orientation impose a much higher risk of interference with the normal

gene expression than those integrated in the reverse orientation (Medstrand et al. 2002). This hypothesis has also been demonstrated experimentally (Chen et al. 2006; Ustyugova et al. 2006), suggesting that genic L1 inserts in the forward orientation are most likely eliminated from the genome fairly quickly. The genic insertion orientation bias has been consistent enough across taxa that it has been incorporated as part of a novel gene detection strategy (Glusman et al. 2006).

One of the unique aspects of L1's influence on gene function is the observation that these elements can interfere with the normal gene expression even when positioned outside of the gene boundaries (Speck 2001; Nigumann et al. 2002; Belancio et al. 2006). Perhaps this is one reason why there is a noticeable depletion of the full-length L1 inserts upstream and downstream of mammalian genes (Medstrand et al. 2002). Interference of mobile elements with normal gene expression is further compounded by the presence of different classes of TEs that carry various regulatory elements that can result in the production of the complex chimeric transcripts (Landry et al. 2001). It is difficult to estimate the true impact of TE insertions on gene expression, for even genomewide quantitative surveys of ESTs would be limited to detection of only those hybrid transcripts that are relatively stable and have not been eliminated during embryonic development and/or long-term evolutionary selection.

Cellular responses to mobile-element activity

Mobile elements are effectively parasitic entities in the sense that they require several activities provided by the host cell for their amplification. Above and beyond their obvious mutagenic potential, mobile elements influence numerous cellular processes in the course of interacting with the host cell. These interactions are illustrated by the finding that overexpression of L1 elements has been shown to lead to toxic effects on cells (Goodier et al. 2004; Gasior et al. 2006). The endonuclease activity produced by the ORF2 protein is at least partially, but not completely, responsible for that toxicity. This ORF2 endonuclease toxicity is likely due to the induction of the double-strand break (DSB) repair response, as evidenced by the formation of gamma-H2AFX foci at the sites of the breaks. DSBs can culminate in cell cycle arrest or apoptosis (Lindblad-Toh et al. 2005; Belgnaoui et al. 2006; Gasior et al. 2006). This DSB response almost certainly involves ATM activation, as cells deficient in ATM were found to only poorly support L1 retrotransposition (Gasior et al. 2006). This is similar to studies with DNA transposons, such as Sleeping Beauty, which are found to have their amplification influenced by a number of genes associated with DNA repair (Izsvak et al. 2004).

Several reports have demonstrated that cells produce proteins that inhibit retrotransposition. A family of human proteins, APOBEC3, exhibits various influences on the retrotransposition process as well as on retroviral life cycles. All of the APOBEC3 proteins except 3A inhibit HIV replication, while APOBEC3A, APOBEC3B, and, to a lesser extent, APOBEC3C inhibit human L1 and *Alu* activities (Bogerd et al. 2006; Muckenfuss et al. 2006; Stenglein and Harris 2006). Despite the reliance of *Alu* elements on L1 activity for their amplification, APOBEC3G inhibits *Alu*, but not L1, mobilization by sequestering *Alu* RNAs in cytoplasmic high-molecular-mass A3G complexes away from the nuclear L1 enzymatic machinery (Chiu et al. 2006; Hulme et al. 2007). The various APOBECs also inhibit a spectrum of other mobile elements (for summary, see Hulme et al. 2007). The underlying mechanism of this inhibition is not currently clear. Although the

APOBEC3 proteins possess cytidine deaminase activity, mutants abolishing this activity are equally capable of inhibiting mobile-element retrotransposition (Stenglein and Harris 2006). Surprisingly, the mouse genome contains only a single APOBEC3 gene, while the human genome has seven functional forms (Conticello et al. 2005). It is possible that the diversity of APOBEC forms in primates evolved to more efficiently inhibit retroelements. This may have resulted in the silencing of LTR—and, to a lesser extent, L1—elements in primates, in contrast to their continued activity in mice (for summary, see Mouse Genome Sequencing Consortium et al. 2002). Some of this diversity may also have developed to inhibit *Alu* elements, which, despite their enormous success, have exhibited attenuated activity over the past 40 million years. Although the inhibition of retroelements by APOBEC3 proteins can be quite strong, as indicated in the transient transfection studies described above, it was found that there were no immediate consequences of knocking out the single *ApoBec3* gene in mice (Mikl et al. 2005). Thus, it is possible that the effect of APOBECs may exert only subtle influences on retroelement activity, which only become apparent over extended time periods.

ORF1 expression triggers activation of MAPK13 (formerly p38 δ MAP kinase) (Kuchen et al. 2004). The p38 MAP kinases are a group of serine/threonine protein kinases that participate in the signal transduction cascade of the cellular responses to diverse external stimuli (for review, see Kumar et al. 1997). MAPK13 has been shown to abrogate apoptosis in renal carcinoma cells treated with an inhibitor of cyclic GMP production (Ambrose et al. 2006). Thus, it is quite plausible that L1 expression can influence numerous signaling pathways in cells and that diverse cell types will respond differentially to L1 based on their characteristic signaling pathways.

The ability of environmental stimuli to alter retrotransposition frequencies is also indicative of the complex interactions of these elements with host factors. There is obviously the need for *trans*-acting factors to interact with the element promoters. However, several heavy metals, as well as ionizing radiation, have been shown to be capable of stimulating the L1 retrotransposition process at a stage *downstream* of transcription (Kale et al. 2005, 2006; Farkash et al. 2006). This stimulation appears to be independent of DNA-nicking activity of these factors and, therefore, is likely to involve influences on some other downstream interactions in the integration process. Similarly, there have been several reports of environmental influences on L1 (Stribinskis and Ramos 2006) and *Alu* (Liu et al. 1995) transcription.

The selfish DNA vs. function controversy

Due to the ancient and complex history of mobile elements and their host genomes, the relationship between host and element is not easily designated as “parasitic,” “symbiotic,” or “mutualistic” in nature. Along with the numerous detrimental effects of mobile-element activity described above, it has been increasingly observed that TEs serve as substrates for evolutionary innovation in many taxa. In several cases, such as with *Drosophila* telomere maintenance by TART and HeT-A (for review, see Pardue et al. 2005) and V(D)J recombination in the vertebrate immune system (Hiom et al. 1998), mobile elements appear to have become “domesticated” and commandeered for essential host functions. It has also been shown that members of a SINE family that remains active in the Coelacanth genome have been co-opted in tetrapods to alternatively serve as both highly conserved noncoding

regulatory regions and as components of proteins (Bejerano et al. 2006). Nevertheless, we should be mindful not to make the leap from the several established cases of TE benefits to the thus-far unsubstantiated claim that TEs are maintained over time within taxa for their various evolutionary advantages. Generally speaking, the notion of selection for evolvability in this context remains controversial and lacking in empirical support (Sniegowski and Murphy 2006).

It seems more plausible to the authors that, given their amplification capability and the population and host-response dynamics involved, selection simply need not be invoked to explain general persistence of TEs across many taxa over evolutionary time scales (for example, see Bestor 1999).

Conclusions

A general impression appears to have developed in the scientific community that mobile elements are fascinating evolutionary oddities which, despite lending themselves to very interesting speculations, are not terribly relevant to the immediate health of organisms. The finding of widespread examples of mobile elements contributing to disease, as well as additional avenues for introducing damage to the host genome, has generated some level of concern for the role of mobile elements in human health. Such concern provides incentive for a more thorough exploration of the impact of these elements on somatic cells, as opposed to simply focusing on the germline activities that contribute to their evolution. The wide range of mechanisms by which the host limits mobile-element activity strongly suggests the importance of repressing mobile elements. Only when we more fully understand the mechanistic interactions of mobile elements with cellular components will we be able to make a reliable assessment of whether TEs pose more of a long-term liability or asset to the host organism.

Acknowledgments

This work was supported by grants from USPHS (R01GM45668), NIH (P20 RR020152), National Science Foundation (EPS-0346411), and the State of Louisiana Board of Regents Support Fund.

References

- Alisch, R.S., Garcia-Perez, J.L., Muotri, A.R., Gage, F.H., and Moran, J.V. 2006. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes & Dev.* **20**: 210–224.
- Ambrose, M., Ryan, A., O'Sullivan, G.C., Dunne, C., and Barry, O.P. 2006. Induction of apoptosis in renal cell carcinoma by reactive oxygen species: involvement of extracellular signal-regulated kinase 1/2, p38 δ / γ , cyclooxygenase-2 down-regulation, and translocation of apoptosis-inducing factor. *Mol. Pharmacol.* **69**: 1879–1890.
- An, W., Han, J.S., Wheelan, S.J., Davis, E.S., Coombes, C.E., Ye, P., Triplett, C., and Boeke, J.D. 2006. Active retrotransposition by a synthetic L1 element in mice. *Proc. Natl. Acad. Sci.* **103**: 18662–18667.
- Apoil, P.A., Kuhllein, E., Robert, A., Rubie, H., and Blancher, A. 2007. HIGM syndrome caused by insertion of an AluYb8 element in exon 1 of the *CD40LG* gene. *Immunogenetics* **59**: 17–23.
- Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744–747.
- Arcot, S.S., Wang, Z., Weber, J.L., Deininger, P.L., and Batzer, M.A. 1995. *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* **29**: 136–144.
- Asch, H.L., Eliacin, E., Fanning, T.G., Connolly, J.L., Bratthauer, G., and Asch, B.B. 1996. Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* **8**: 239–247.
- Athanasiadis, A., Rich, A., and Maas, S. 2004. Widespread A-to-I RNA editing of *Alu*-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**: e391. doi: 10.1371/journal.pbio.0020391.
- Athanikar, J.N., Badge, R.M., and Moran, J.V. 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **32**: 3846–3855.
- Babcock, M., Pavlicek, A., Spiteri, E., Kashork, C.D., Ioshikhes, I., Shaffer, L.G., Jurka, J., and Morrow, B.E. 2003. Shuffling of genes within low-copy repeats on 22q11 (LCR22) by *Alu*-mediated recombination events during evolution. *Genome Res.* **13**: 2519–2532.
- Babushok, D.V., Ostertag, E.M., Courtney, C.E., Choi, J.M., and Kazazian Jr., H.H. 2006. L1 integration in a transgenic mouse model. *Genome Res.* **16**: 240–250.
- Babushok, D.V., Ohshima, K., Ostertag, E.M., Chen, X., Wang, Y., Mandal, P.K., Okada, N., Abrams, C.S., and Kazazian Jr., H.H. 2007. A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Res.* **17**: 1129–1138.
- Bartolome, C., Maside, X., and Charlesworth, B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.* **19**: 926–937.
- Basame, S., Wai-lun Li, P., Howard, G., Branciforte, D., Keller, D., and Martin, S.L. 2006. Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *J. Mol. Biol.* **357**: 351–357.
- Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**: 87–90.
- Belancio, V.P., Hedges, D.J., and Deininger, P. 2006. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.* **34**: 1512–1521.
- Belancio, V.P., Whelton, M., and Deininger, P. 2007. Requirements for polyadenylation at the 3' end of LINE-1 elements. *Gene* **390**: 98–107.
- Belancio, V.P., Roy-Engel, A.M., and Deininger, P. 2008. The impact of multiple splice sites in human L1 elements. *Gene* doi: 10.1016/j.gene.2007.12.022.
- Belgnaoui, S.M., Gosden, R.G., Semmes, O.J., and Haoudi, A. 2006. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.* **6**: 13.
- Bestor, T.H. 1999. Sex brings transposons and genomes into conflict. *Genetica* **107**: 289–295.
- Boger, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. 2006. Cellular inhibitors of long interspersed element 1 and *Alu* retrotransposition. *Proc. Natl. Acad. Sci.* **103**: 8780–8785.
- Boissinot, S., Entezam, A., and Furano, A.V. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18**: 926–935.
- Boissinot, S., Davis, J., Entezam, A., Petrov, D., and Furano, A.V. 2006. Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl. Acad. Sci.* **103**: 9590–9594.
- Borchert, G.M., Lanier, W., and Davidson, B.L. 2006. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **13**: 1097–1101.
- Bouchet, C., Vuillaumier-Barrot, S., Gonzales, M., Boukari, S., Le Bizec, C., Fallet, C., Delezoide, A.-L., Moirot, H., Laquerriere, A., Encha-Razavi, F., et al. 2007. Detection of an *Alu* insertion in the POMT1 gene from three French Walker Warburg syndrome families. *Mol. Genet. Metab.* **90**: 93–96.
- Bourc'his, D. and Bestor, T.H. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**: 96–99.
- Branciforte, D. and Martin, S.L. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: Implications for transposition. *Mol. Cell. Biol.* **14**: 2584–2592.
- Bratthauer, G.L. and Fanning, T.G. 1992. Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* **7**: 507–510.
- Bratthauer, G.L., Cardiff, R.D., and Fanning, T.G. 1994. Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* **73**: 2333–2336.
- Brookfield, J.F.Y. 2005. The ecology of the genome—Mobile DNA elements and their hosts. *Nat. Rev. Genet.* **6**: 128–136.
- Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian Jr., H.H. 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci.* **100**: 5280–5285.
- Callinan, P.A., Wang, J., Herke, S.W., Garber, R.K., Liang, P., and Batzer, M.A. 2005. *Alu* retrotransposition-mediated deletion. *J. Mol. Biol.*

- 348:** 791–800.
- Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. 1996. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**: 1423–1427.
- Casavant, N.C., Scott, L., Cantrell, M.A., Wiggins, L.E., Baker, R.J., and Wichman, H.A. 2000. The end of the LINE?: Lack of recent L1 activity in a group of South American rodents. *Genetics* **154**: 1809–1817.
- Chen, J., Rattner, A., and Nathans, J. 2006. Effects of L1 retrotransposon insertion on transcript processing, localization, and accumulation: Lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements. *Hum. Mol. Genet.* **15**: 2146–2156.
- Chen, J.M., Masson, E., Macek Jr., M., Raguene, O., Piskackova, T., Fercot, B., Fila, L., Cooper, D.N., Audrezet, M.P., and Ferec, C. 2008. Detection of two *Alu* insertions in the *CFTR* gene. *J. Cyst. Fibros.* **7**: 37–43.
- Chesnokov, I. and Schmid, C.W. 1996. Flanking sequences of an *Alu* source stimulate transcription in vitro by interacting with sequence-specific transcription factors. *J. Mol. Evol.* **42**: 30–36.
- Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.
- Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T., and Greene, W.C. 2006. High-molecular-mass APOBEC3G complexes restrict *Alu* retrotransposition. *Proc. Natl. Acad. Sci.* **103**: 15588–15593.
- Christensen, S.M. and Eickbush, T.H. 2005. R2 target-primed reverse transcription: Ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol. Cell. Biol.* **25**: 6617–6628.
- Christensen, S.M., Bibillo, A., and Eickbush, T.H. 2005. Role of the *Bombyx mori* R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res.* **33**: 6461–6468.
- Clark, L.A., Wahl, J.M., Rees, C.A., and Murphy, K.E. 2006. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proc. Natl. Acad. Sci.* **103**: 1376–1381.
- Coticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., and Neuberger, M.S. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **22**: 367–377.
- Cordaux, R., Hedges, D.J., Herke, S.W., and Batzer, M.A. 2006. Estimating the retrotransposition rate of human *Alu* elements. *Gene* **373**: 134–137.
- DeBerardinis, R.J., Goodier, J.L., Ostertag, E.M., and Kazazian Jr., H.H. 1998. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat. Genet.* **20**: 288–290.
- Deininger, P.L. and Batzer, M.A. 1993. Evolution of retrotransposons. In *Evolutionary biology* (eds. M.K. Heckh et al.), Vol. 27, pp. 157–196. Plenum Publishing, New York.
- Deininger, P.L. and Batzer, M.A. 1999. *Alu* repeats and human disease. *Mol. Genet. Metab.* **67**: 183–193.
- Deininger, P.L., Moran, J.V., Batzer, M.A., and Kazazian Jr., H.H. 2003. Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* **13**: 651–658.
- Dewannieux, M. and Heidmann, T. 2005. Role of poly(A) tail length in *Alu* retrotransposition. *Genomics* **86**: 378–381.
- Dewannieux, M., Esnault, C., and Heidmann, T. 2003. LINE-mediated retrotransposition of marked *Alu* sequences. *Nat. Genet.* **35**: 41–48.
- Dmitriev, S.E., Andreev, D.E., Terenin, I.M., Olovnikov, I.A., Prassolov, V.S., Merrick, W.C., and Shatsky, I.N. 2007. Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol. Cell. Biol.* **27**: 4685–4697.
- Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., and Kazazian Jr., H.H. 1991. Isolation of an active human transposable element. *Science* **254**: 1805–1808.
- Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. 2005. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**: 221–226.
- Eickbush, T.H. and Furano, A.V. 2002. Fruit flies and humans respond differently to retrotransposons. *Curr. Opin. Genet. Dev.* **12**: 669–674.
- El Sawy, M., Kale, S.P., Dugan, C., Quyen Nguyen, T., Belancio, V., Bruch, H., Roy-Engel, A.M., and Deininger, P.L. 2005. Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. *J. Mol. Biol.* **354**: 246–257.
- Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Strätling, W.H., and Schumann, G.G. 2004. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J. Biol. Chem.* **279**: 27753–27763.
- Farkash, E.A., Kao, G.D., Horman, S.R., and Prak, E.T. 2006. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res.* **34**: 1196–1204.
- Feng, Q., Moran, J.V., Kazazian Jr., H.H., and Boeke, J.D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**: 905–916.
- Furano, A.V., Duvernell, D.D., and Boissinot, S. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* **20**: 9–14.
- Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. 2007. LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16**: 1569–1577.
- Gasior, S.L., Wakeman, T.P., Xu, B., and Deininger, P.L. 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **357**: 1383–1393.
- Gasior, S.L., Preston, G., Hedges, D.J., Gilbert, N., Moran, J.V., and Deininger, P.L. 2007. Characterization of pre-insertion loci of de novo L1 insertions. *Gene* **390**: 190–198.
- Gentles, A.J., Kohany, O., and Jurka, J. 2005. Evolutionary diversity and potential recombinogenic role of integration targets of Non-LTR retrotransposons. *Mol. Biol. Evol.* **22**: 1983–1991.
- Gentles, A.J., Wakefield, M.J., Kohany, O., Gu, W., Batzer, M.A., Pollock, D.D., and Jurka, J. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* **17**: 992–1004.
- Gilbert, N., Lutz-Prigge, S., and Moran, J.V. 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**: 315–325.
- Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* **25**: 7780–7795.
- Glusman, G., Qin, S., El-Gewely, M.R., Siegel, A.F., Roach, J.C., Hood, L., and Smit, A.F. 2006. A third approach to gene prediction suggests thousands of additional human transcribed regions. *PLoS Comput. Biol.* **2**: e18. doi: 10.1371/journal.pcbi.0020018.
- Goodier, J.L., Ostertag, E.M., Engleka, K.A., Selem, M.C., and Kazazian Jr., H.H. 2004. A potential role for the nucleolus in L1 retrotransposition. *Hum. Mol. Genet.* **13**: 1041–1048.
- Gorbulonova, V. and Seluanov, A. 2005. Making ends meet in old age: DSB repair and aging. *Mech. Ageing Dev.* **126**: 621–628.
- Graham, T. and Boissinot, S. 2006. The genomic distribution of L1 elements: The role of insertion bias and natural selection. *J. Biomed. Biotechnol.* **2006**: 75327. doi: 10.1155/JBB/2006/75327.
- Grahn, R.A., Rinehart, T.A., Cantrell, M.A., and Wichman, H.A. 2005. Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. *Cytogenet. Genome Res.* **110**: 407–415.
- Gu, Y., Kodama, H., Watanabe, S., Kikuchi, N., Ishitsuka, I., Ozawa, H., Fujisawa, C., and Shiga, K. 2007. The first reported case of Menkes disease caused by an *Alu* insertion mutation. *Brain Dev.* **29**: 105–108.
- Han, J.S., Szak, S.T., and Boeke, J.D. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**: 268–274.
- Han, K., Sen, S.K., Wang, J., Callinan, P.A., Lee, J., Cordaux, R., Liang, P., and Batzer, M.A. 2005. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.* **33**: 4040–4052.
- Hata, K. and Sakaki, Y. 1997. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* **189**: 227–234.
- Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata N. 2001. *Alu*-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proc. Natl. Acad. Sci.* **98**: 11399–11404.
- Hedges, D.J. and Deininger, P.L. 2007. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.* **616**: 46–59.
- Hedges, D.J., Callinan, P.A., Cordaux, R., Xing, J., Barnes, E., and Batzer, M.A. 2004. Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* **14**: 1068–1075.
- Hiom, K., Melek, M., and Gellert, M. 1998. DNA transposition by the RAG1 and RAG2 proteins: A possible source of oncogenic translocations. *Cell* **94**: 463–470.
- Hulme, A.E., Bogerd, H.P., Cullen, B.R., and Moran, J.V. 2007. Selective inhibition of *Alu* retrotransposition by APOBEC3G. *Gene* **390**: 199–205.
- International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* **436**: 793–800.
- Izsvak, Z., Stüwe, E.E., Fiedler, D., Katzer, A., Jeggo, P.A., and Ivics, Z. 2004. Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. *Mol. Cell*

- 13:** 279–290.
- Johanning, K., Stevenson, C.A., Oyeniran, O.O., Gozal, Y.M., Roy-Engel, A.M., Jurka, J., and Deininger, P.L. 2003. Potential for retroposition by old *Alu* subfamilies. *J. Mol. Evol.* **56**: 658–664.
- Johnson, R.D. and Jasin, M. 2000. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19**: 3398–3407.
- Kajikawa, M. and Okada, N. 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* **111**: 433–444.
- Kale, S.P., Moore, L., Deininger, P.L., and Roy-Engel, A.M. 2005. Heavy metals stimulate human LINE-1 retrotransposition. *Int. J. Environ. Res. Public Health* **2**: 14–23.
- Kale, S.P., Carmichael, M.C., Harris, K., and Roy-Engel, A.M. 2006. The L1 retrotranspositional stimulation by particulate and soluble cadmium exposure is independent of the generation of DNA breaks. *Int. J. Environ. Res. Public Health* **3**: 121–128.
- Kamal, M., Xie, X.H., and Lander, E.S. 2006. A large family of ancient repeat elements in the human genome is under strong selection. *Proc. Natl. Acad. Sci.* **103**: 2740–2745.
- Kapitonov, V.V. and Jurka, J. 2003. A novel class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol.* **20**: 694–702.
- Kazazian Jr., H.H. 1998. Mobile elements and disease. *Curr. Opin. Genet. Dev.* **8**: 343–350.
- Kazazian Jr., H.H. 1999. An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* **22**: 130.
- Kazazian Jr., H.H. 2004. Mobile elements: Drivers of genome evolution. *Science* **303**: 1626–1632.
- Kazazian Jr., H.H. and Goodier, J.L. 2002. LINE drive: Retrotransposition and genome instability. *Cell* **110**: 277–280.
- Khan, H., Smit, A., and Boissinot, S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**: 78–87.
- Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. 2004. Widespread RNA editing of embedded *Alu* elements in the human transcriptome. *Genome Res.* **14**: 1719–1725.
- Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. 2004. Retrotransposon-induced mutations in grape skin color. *Science* **304**: 982.
- Kordis, D., Lovsin, N., and Gubensek, F. 2006. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. *Syst. Biol.* **55**: 886–901.
- Kramerov, D.A. and Vassetzky, N.S. 2005. Short retrotransposons in eukaryotic genomes. *Int. Rev. Cytol.* **247**: 165–221.
- Krawczak, M., Ball, E.V., Fenton, I., Stenson, P.D., Abeyasinghe, S., Thomas, N., and Cooper, D.N. 2000. Human gene mutation database—A biomedical information and research resource. *Hum. Mutat.* **15**: 45–51.
- Kubo, S., Seleme, M.C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian Jr., H.H., and Kasahara, N. 2006. L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci.* **103**: 8036–8041.
- Kuchen, S., Seemayer, C.A., Rethage, J., von Knoch, R., Kuenzler, P., Beat, A.M., Gay, R.E., Gay, S., and Neidhart, M. 2004. The L1 retroelement-related p40 protein induces p38delta MAP kinase. *Autoimmunity* **37**: 57–65.
- Kulpa, D.A. and Moran, J.V. 2005. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* **14**: 3237–3248.
- Kulpa, D.A. and Moran, J.V. 2006. *cis*-Preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* **13**: 655–660.
- Kumar, S., McDonnell, P.C., Gum, R.J., Hand, A.T., Lee, J.C., and Young, P.R. 1997. Novel homologues of CSBP/p38 MAP kinase: Activation, substrate specificity and sensitivity to inhibition by pyridinyl imidazoles. *Biochem. Biophys. Res. Commun.* **235**: 533–538.
- Kurosaki, T., Ninokata, A., Wang, L., and Ueda, S. 2006. Evolutionary scenario for acquisition of CAG repeats in human SCA1 gene. *Gene* **373**: 23–27.
- Kurose, K., Hata, K., Hattori, M., and Sakaki, Y. 1995. RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. *Nucleic Acids Res.* **23**: 3704–3709.
- Kurzynska-Kokorniak, A., Jamburuthugoda, V.K., Bibillo, A., and Eickbush, T.H. 2007. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J. Mol. Biol.* **374**: 322–333.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Landry, J.R., Medstrand, P., and Mager, D.L. 2001. Repetitive elements in the 5' untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. *Genomics* **76**: 110–116.
- Lavie, L., Maldener, E., Brouha, B., Meese, E.U., and Mayer, J. 2004. The human L1 promoter: Variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.* **14**: 2253–2260.
- Le Rouzic, A., Dupas, S., and Capy, P. 2007. Genome ecosystem and transposable elements species. *Gene* **390**: 214–220.
- Li, P.W., Li, J., Timmerman, S.L., Krushel, L.A., and Martin, S.L. 2006. The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: Implications for retrotransposition. *Nucleic Acids Res.* **34**: 853–864.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas 3rd, E.J., Zody, M.C., et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- Liu, W.M., Marais, R.J., Rubin, C.M., and Schmid, C.W. 1994. *Alu* transcripts: Cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res.* **22**: 1087–1095.
- Liu, W.M., Chu, W.M., Choudary, P.V., and Schmid, C.W. 1995. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.* **23**: 1758–1765.
- Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* **72**: 595–605.
- Lucchinetti, E., Feng, J., Silva, R., Tolstonog, G.V., Schaub, M.C., Schumann, G.G., and Zaugg, M. 2006. Inhibition of LINE-1 expression in the heart decreases ischemic damage by activation of Akt/PKB signaling. *Physiol. Genomics* **25**: 314–324.
- Lunyak, V.V., Prefontaine, G.G., Núñez, E., Cramer, T., Ju, B.-G., Ohgi, K.A., Hutt, K., Roy, R., García-Díaz, A., Zhu, X., et al. 2007. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**: 248–251.
- Lynch, M. 2002. Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci.* **99**: 6118–6123.
- Lynch, M. and Conery, J.S. 2003. The origins of genome complexity. *Science* **302**: 1401–1404.
- Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. 2006. Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLoS Genet.* **2**: e2. doi: 10.1371/journal.pgen.0020002.
- Malik, H.S. and Eickbush, T.H. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol. Biol. Evol.* **15**: 1123–1134.
- Manco, L., Relvas, L., Silva, P.C., Pereira, J., Almeida, A.B., and Ribeiro, M.L. 2006. Molecular characterization of five Portuguese patients with pyrimidine 5'-nucleotidase deficient hemolytic anemia showing three new P5'N-I mutations. *Haematologica* **91**: 266–267.
- Martin, S.L., Li, J., Epperson, L.E., and Lieberman, B. 1998. Functional reverse transcriptases encoded by A-type mouse LINE-1: Defining the minimal domain by deletion analysis. *Gene* **215**: 69–75.
- Martin, S.L., Branciforte, D., Keller, D., and Bain, D.L. 2003. Trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl. Acad. Sci.* **100**: 13815–13820.
- Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S.C., Hodges, R.S., and Williams, M.C. 2005. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.* **348**: 549–561.
- Mätlik, K., Redik, K., and Speek, M. 2006. L1 antisense promoter drives tissue-specific transcription of human genes. *J. Biomed. Biotechnol.* **2006**: 71753. doi: 10.1155/JBB/2006/71753.
- Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Res.* **12**: 1483–1495.
- Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**: 167–177.
- Mikl, M.C., Watt, I.N., Lu, M., Reik, W., Davies, S.L., Neuberger, M.S., and Rada, C. 2005. Mice deficient in APOBEC2 and APOBEC3. *Mol. Cell. Biol.* **25**: 7270–7277.
- Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S., and Devine, S.E. 2006. Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* **78**: 671–679.
- Miné, M., Chen, J.M., Brivet, M., Desguerre, I., Marchant, D., de Lonlay, P., Bernard, A., Férec, C., Abitbol, M., Ricquier, D., et al. 2007. A large genomic deletion in the *PDHX* gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Hum. Mutat.* **28**: 137–142.
- Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and

- Kazazian Jr., H.H. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**: 917–927.
- Moran, J.V., DeBerardinis, R.J., and Kazazian Jr., H.H. 1999. Exon shuffling by L1 retrotransposition. *Science* **283**: 1530–1534.
- Moran, J.V., Gilbert, N., Boeke, J., Kazazian, H., Ostertag, E., Loon, S., and Wei, W. 2000. Human L1s retrotransposition: *cis*-Preference vs. *trans*-complementation. *Am. J. Hum. Genet.* **67**: 199.
- Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31**: 159–165.
- Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., Taccioli, G.E., Sekiguchi, J., and Moran, J.V. 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446**: 208–212.
- Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G.G., and Münk, C. 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.* **281**: 22161–22172.
- Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**: 903–910.
- Nigumann, P., Redik, K., Mätlik, K., and Speck, M. 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* **79**: 628–634.
- Nishihara, H., Smit, A.F.A., and Okada, N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **16**: 864–874.
- Odom, G.L., Robichaux, J.L., and Deininger, P.L. 2004. Predicting mammalian SINE subfamily activity from A-tail length. *Mol. Biol. Evol.* **21**: 2140–2148.
- Okubo, M., Horinishi, A., Saito, M., Ebara, T., Endo, Y., Kaku, K., Murase, T., and Eto, M. 2007. A novel complex deletion-insertion mutation mediated by *Alu* repetitive elements leads to lipoprotein lipase deficiency. *Mol. Genet. Metab.* **92**: 229–233.
- Ostertag, E.M. and Kazazian Jr., H.H. 2001a. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**: 501–538.
- Ostertag, E.M. and Kazazian Jr., H.H. 2001b. Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **11**: 2059–2065.
- Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., and Kazazian Jr., H.H. 2002. A mouse model of human L1 retrotransposition. *Nat. Genet.* **32**: 655–660.
- Ostertag, E.M., Goodier, J.L., Zhang, Y., and Kazazian Jr., H.H. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **73**: 1444–1451.
- Otieno, A.C., Carter, A.B., Hedges, D.J., Walker, J.A., Ray, D.A., Garber, R.K., Anders, B.A., Stoilova, N., Laborde, M.E., Fowlkes, J.D., et al. 2004. Analysis of the human *Alu* Ya-lineage. *J. Mol. Biol.* **342**: 109–118.
- Ovchinnikov, I., Troxel, A.B., and Swergold, G.D. 2001. Genomic characterization of recent human LINE-1 insertions: Evidence supporting random insertion. *Genome Res.* **11**: 2050–2058.
- Pace, J.K. and Feschotte, C. 2007. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **17**: 422–432.
- Pardue, M.L., Rashkova, S., Casacuberta, E., DeBaryshe, P.G., George, J.A., and Traverse, K.L. 2005. Two retrotransposons maintain telomeres in *Drosophila*. *Chromosome Res.* **13**: 443–453.
- Paulson, K.E. and Schmid, C.W. 1986. Transcriptional inactivity of *Alu* repeats in HeLa cells. *Nucleic Acids Res.* **14**: 6145–6158.
- Perepelitsa-Belancio, V. and Deininger, P. 2003. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* **35**: 363–366.
- Pierce, A.J., Stark, J.M., Araujo, F.D., Moynahan, M.E., Berwick, M., and Jasim, M. 2001. Double-strand breaks and tumorigenesis. *Trends Cell Biol.* **11**: S52–S59.
- Pontius, J.U., Mullikin, J.C., Smith, D.R., Agencourt Sequencing Team, Lindblad-Toh, K., Gnerre, S., Clamp, M., Chang, J., Stephens, R., Neelam, B., et al. 2007. Initial sequence and comparative analysis of the cat genome. *Genome Res.* **17**: 1675–1689.
- Pritham, E.J. and Feschotte, C. 2007. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc. Natl. Acad. Sci.* **104**: 1895–1900.
- Ray, D.A., Pagan, H.J., Thompson, M.L., and Stevens, R.D. 2007. Bats with hATs: Evidence for recent DNA transposon activity in genus *Myotis*. *Mol. Biol. Evol.* **24**: 632–639.
- Rinehart, T.A., Grahn, R.A., and Wichman, H.A. 2005. SINE extinction preceded LINE extinction in sigmodontine rodents: Implications for retrotranspositional dynamics and mechanisms. *Cytogenet. Genome Res.* **110**: 416–425.
- Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. 1998. DNA double-stranded breaks induce histone H2AFX phosphorylation on serine 139. *J. Biol. Chem.* **273**: 5858–5868.
- Roy, A.M., West, N.C., Rao, A., Adhikari, P., Alemán, C., Barnes, A.P., and Deininger, P.L. 2000. Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* **302**: 17–25.
- Roy-Engel, A.M., Salem, A.H., Oyeniran, O.O., Deininger, L., Hedges, D.J., Kilroy, G.E., Batzer, M.A., and Deininger, P.L. 2002. Active *Alu* element “A-tails”: Size does matter. *Genome Res.* **12**: 1333–1344.
- Salem, A.H., Kilroy, G.E., Watkins, W.S., Jorde, L.B., and Batzer, M.A. 2003. Recently integrated *Alu* elements and human genomic diversity. *Mol. Biol. Evol.* **20**: 1349–1361.
- Sayah, D.M., Sokolskaja, E., Berthoux, L., and Luban, J. 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**: 569–573.
- Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A., Dyer, M., Cordaux, R., Liang, P., and Batzer, M.A. 2006. Human genomic deletions mediated by recombination between *Alu* elements. *Am. J. Hum. Genet.* **79**: 41–53.
- Sen, S.K., Huang, C.T., Han, K., and Batzer, M.A. 2007. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res.* **35**: 3741–3751.
- Severynse, D.M., Hutchison III, C.A., and Edgell, M.H. 1992. Identification of transcriptional regulatory activity within the 5′ A-type monomer sequence of the mouse LINE-1 retroposon. *Mamm. Genome* **2**: 41–50.
- Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A., and Deininger, P.L. 1997. cDNAs derived from primary and small cytoplasmic *Alu* (scAlu) transcripts. *J. Mol. Biol.* **271**: 222–234.
- Shi, X., Seluanov, A., and Gorbunova, V. 2007. Cell divisions are required for L1 retrotransposition. *Mol. Cell Biol.* **27**: 1264–1270.
- Sinnett, D., Richer, C., Deragon, J.M., and Labuda, D. 1992. *Alu* RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. *J. Mol. Biol.* **226**: 689–706.
- Skowronski, J. and Singer, M.F. 1985. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci.* **82**: 6050–6054.
- Skowronski, J., Fanning, T.G., and Singer, M.F. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell Biol.* **8**: 1385–1397.
- Sniegowski, P.D. and Murphy, H.A. 2006. Evolvability. *Curr. Biol.* **16**: R831–R834.
- Soifer, H.S., Zaragoza, A., Peyvan, M., Behlke, M.A., and Rossi, J.J. 2005. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res.* **33**: 846–856.
- Song, M. and Boissinot, S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **390**: 206–213.
- Sorek, R., Ast, G., and Graur, D. 2002. *Alu*-containing exons are alternatively spliced. *Genome Res.* **12**: 1060–1067.
- Speck, M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell Biol.* **21**: 1973–1985.
- Stenglein, M.D. and Harris, R.S. 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.* **281**: 16837–16841.
- Stribinskis, V. and Ramos, K.S. 2006. Activation of human long interspersed nuclear element 1 retrotransposition by benzo(a)pyrene, an ubiquitous environmental carcinogen. *Cancer Res.* **66**: 2616–2620.
- Swergold, G.D. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell Biol.* **10**: 6718–6729.
- Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. 2002. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* **110**: 327–338.
- Tamura, M., Kajikawa, M., and Okada, N. 2007. Functional splice sites in a zebrafish LINE and their influence on zebrafish gene expression. *Gene* **390**: 221–231.
- Tang, R.B., Wang, H.Y., Lu, H.Y., Xiong, J., Li, H.H., Qiu, X.H., and Liu, H.Q. 2005. Increased level of polymerase III transcribed *Alu* RNA in hepatocellular carcinoma tissue. *Mol. Carcinog.* **42**: 93–96.
- Trelogan, S.A. and Martin, S.L. 1995. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc. Natl. Acad. Sci.* **92**: 1520–1524.
- Udaka, T., Okamoto, N., Aramaki, M., Torii, C., Kosaki, R., Hosokai, N., Hayakawa, T., Takahata, N., Takahashi, T., and Kosaki, K. 2007. An

- Alu* retrotransposition-mediated deletion of *CHD7* in a patient with CHARGE syndrome. *Am. J. Med. Genet. A* **143**: 721–726.
- Ustyugova, S.V., Lebedev, Y.B., and Sverdlov, E.D. 2006. Long L1 insertions in human gene introns specifically reduce the content of corresponding primary transcripts. *Genetica* **128**: 261–272.
- van den Hurk, J.A., Meij, I.C., Seleme, M.C., Kano, H., Nikopoulos, K., Hoefloot, L.H., Sistermans, E.A., de Wijs, I.J., Mukhopadhyay, A., Plomp, A.S., et al. 2007. L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* **16**: 1587–1592.
- Vervoort, R., Gitzelmann, R., Lissens, W., and Liebaers, I. 1998. A mutation (IVS8+0.6kdelTC) creating a new donor splice site activates a cryptic exon in an *Alu*-element in intron 8 of the human beta-glucuronidase gene. *Hum. Genet.* **103**: 686–693.
- Vilenchik, M.M. and Knudson, A.G. 2003. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci.* **100**: 12871–12876.
- Vinogradov, A.E. 2003. Selfish DNA is maladaptive: Evidence from the plant Red List. *Trends Genet.* **19**: 609–614.
- Vinogradov, A.E. 2004. Genome size and extinction risk in vertebrates. *Proc. Biol. Sci.* **271**: 1701–1705.
- Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A., and Batzer, M.A. 2005. SVA elements: A hominid-specific retroposon family. *J. Mol. Biol.* **354**: 994–1007.
- Wheelan, S.J., Aizawa, Y., Han, J.S., and Boeke, J.D. 2005. Gene-breaking: A new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* **15**: 1073–1078.
- Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E., and Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J.* **26**: 307–316.
- Xing, J., Wang, H., Belancio, V.P., Cordaux, R., Deininger, P.L., and Batzer, M.A. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl. Acad. Sci.* **103**: 17608–17613.
- Yang, N. and Kazazian Jr., H.H. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* **13**: 763–771.
- Yang, N., Zhang, L., Zhang, Y., and Kazazian Jr., H.H. 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* **31**: 4929–4940.
- Youngman, S., van Luenen, H.G., and Plasterk, R.H. 1996. Rte-1, a retrotransposon-like element in *Caenorhabditis elegans*. *FEBS Lett.* **380**: 1–7.
- Yu, F., Zingler, N., Schumann, G., and Strätling, W.H. 2001. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not *Alu* transcription. *Nucleic Acids Res.* **29**: 4493–4501.
- Zemojtel, T., Penzkofer, T., Schultz, J., Dandekar, T., Badge, R., and Vingron, M. 2007. Exonization of active mouse L1s: A driver of transcriptome evolution? *BMC Genomics* **8**: 392. doi: 10.1186/1471-2164-8-392.
- Zingler, N., Willhoeft, U., Brose, H.P., Schoder, V., Jahns, T., Hanschmann, K.M., Morrish, T.A., Löwer, J., and Schumann, G.G. 2005. Analysis of 5' junctions of human LINE-1 and *Alu* retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res.* **15**: 780–789.



Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health

Victoria P. Belancio, Dale J. Hedges and Prescott Deininger

Genome Res. 2008 18: 343-358

Access the most recent version at doi:[10.1101/gr.5558208](https://doi.org/10.1101/gr.5558208)

References This article cites 199 articles, 61 of which can be accessed free at:
<http://genome.cshlp.org/content/18/3/343.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

License Freely available online through the Genome Research Open Access option.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
