REVIEW ARTICLE

Progress in Understanding the Biology of the Human Mutagen LINE-1

Daria V. Babushok and Haig H. Kazazian, Jr.*

Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania

Communicated by David Cooper

Long interspersed nucleotide element (LINE)-1 retrotransposon (L1) has emerged as the largest contributor to mammalian genome mass, responsible for over 35% of the human genome. Differences in the number and activity levels of L1s contribute to interindividual variation in humans, both by affecting an individual's likelihood of acquiring new L1-mediated mutations, as well as by differentially modifying gene expression. Here, we report on recent progress in understanding L1 biology, with a focus on mechanisms of L1-mediated disease. We discuss known details of L1 lifecycle, including L1 structure, transcriptional regulation, and the mechanisms of translation and retrotransposition. Current views on cell type specificity, timing, and control of retrotransposition are put forth. Finally, we discuss the role of L1 as a mutagen, using the latest findings in L1 biology to illuminate molecular mechanisms of L1-mediated gene disruption. Hum Mutat 0, 1–13, 2007. Published 2007 Wiley-Liss, Inc.†

KEY WORDS: genome evolution; LINE-1; L1; retrotransposition; antisense transcription; RNAi; bicistronic transcript; termination/reinitiation; human mutation; gene disruption

REPETITIVE CONTENT LEADS TO MAJOR DIFFERENCES IN GENOME SIZE

Genome size tells us surprisingly little about the complexity of the organism. While there is a general increase of total genomic DNA with increasing complexity in lower eukaryotes—1.2×10^7 bp for the single cell *S. cerevisiae*, 1.0×10^8 for *C. elegans*, 1.3×10^8 for *D. melanogaster*, and 2.7×10^8 bp for *A. gambiae*—the correlation breaks down with increasing genome size. For instance, another mosquito, *A. aegypti*, has a genome of 1.4×10^9 bp, larger than that of lower vertebrates fugu (3.9×10^8 bp) and chicken (1.0×10^9 bp), and while mammalian genome sizes are three to eight times larger, they fail to correlate with the apparent increase in organismal complexity, a phenomenon referred to as the C-value paradox [Cavalier-Smith, 1985].

The explanation lies largely in the fact that only a small fraction of genomic DNA, roughly 5% of mammalian genomes [Lander et al., 2001; Waterston et al., 2002; Gibbs et al., 2004; Lindblad-Toh et al., 2005], is subject to evolutionary constraint and, presumably, contains genes and gene regulatory regions. In contrast, over 46% of the human genome and 39% of the mouse genome are recognized as interspersed repeats that are largely nonfunctional [Lander et al., 2001; Waterston et al., 2002]. The actual contribution of repeats to mammalian genomes is significantly larger, as the older, ancestral repeats have diverged beyond the current recognition limit of ~37% divergence [Smit, 1999; Waterston et al., 2002]. This is further supported by the reduced ability to detect ancestral repeat sequences in the mouse (as compared to the human) due to its higher mutation rate—the recognition limits are approximately 100–120 million years (Myr) for mouse and 150–200 Myr for human [Waterston et al., 2002]. For comparison, the chicken genome lacks the abundant mammalian short interspersed nucleotide elements (SINEs) and L1 elements and contains only ~9% repetitive DNA, largely explaining its approximately three-fold smaller genome size [Hillier et al., 2004]. Importantly, difficulties with cloning, sequencing, and assembly of highly repetitive regions have led to the underrepresentation and, sometimes, intentional exclusion of repetitive sequences from genome databases [Lander et al., 2001].

AUTONOMOUS LONG INTERSPERSED NUCLEOTIDE ELEMENTS (LINE)-1 ELEMENTS ARE THE MAJOR DRIVING FORCE BEHIND GENOME EXPANSION

DNA sequences broadly classified as repeats include: 1) tandem repeats of the centromeric and telomeric regions, which are often excluded from genome assemblies; 2) segmental duplications, which arose by recombinational processes and comprise ~5% of the human genome, largely located in the pericentromeric and subtelomeric regions [Eichler, 2001; Zhang, 2003; Stankiewicz et al., 2004]; 3) simple sequence repeats (SSRs), which are tandem, sometimes imperfect repetitions of 1 to 13 nt that arose by DNA polymerase slippage during replication, make up ~3% of the human genome [Lander et al., 2001], and are frequently used

Received 1 November 2006; accepted revised manuscript 22 December 2006.

*Correspondence to: Haig H. Kazazian, Jr., Department of Genetics, University of Pennsylvania School of Medicine, 564 Clinical Research Building, 415 Curie Blvd, Philadelphia, PA 19104-6145. E-mail: kazazian@mail.med.upenn.edu

Grant sponsor: National Institutes of Health.

DOI 10.1002/humu.20486

Published online in Wiley InterScience (www.interscience.wiley.com).

as genetic markers because of their high levels of length polymorphism [Weissenbach, 1993]; and 4) transposable elements, also known as interspersed repeats, which are the largest single component of mammalian genomes and probably the most important determinant of global genome evolution [Lander et al., 2001; Kazazian, 2004].

Based on their mechanism of replication, transposable elements, in turn, are categorized into four classes (Fig. 1): 1) DNA transposons, which move by a "cut-and-paste" mechanism using an encoded transposase gene and inverse terminal repeats (ITRs) [Reznikoff, 2003]; 2) long terminal repeat (LTR) retrotransposons, which contain LTRs, encode *gag* and *pol* genes, and mobilize through an RNA intermediate by a retroviral-like reverse transcription mechanism; 3) LINEs (long interspersed nucleotide elements, also known as non-LTR elements) which mobilize by reverse transcription through an ancient mechanism shared with the group II introns of mitochondria and eubacteria [Eickbush and Malik, 2002] (discussed below); and 4) SINEs (short interspersed nucleotide elements) which lack the ability to retrotranspose autonomously, but can co-opt LINE-1 machinery for their replication [Kazazian, 2004]. Their respective contributions to the human genome sequence are DNA transposons (3%), LTR elements (9%), but over 35% for LINES (21% by self-replication and over 14% indirectly by replication of SINEs and processed pseudogenes) [Lander et al., 2001; Waterston et al., 2002]. The high degree of divergence and lack of presence-absence polymorphisms among the DNA transposons and LTR elements strongly suggest that these elements have been extinct in the human lineage for over 40 Myr [Lander et al., 2001]. In contrast, the predominant mammalian LINE element, LINE-1 or L1, underwent a peak of retrotranspositional activity around 40 Myr ago [Ohshima et al., 2003]. While the activity of L1 has since declined, the average human being carries ~80 to 100 potentially active L1 elements in his/her diploid genome [Brouha et al., 2003]. The presence-absence polymorphism of L1s in the human population, combined with a great variability in the activity levels of L1 alleles make L1 a key contributor to interindividual variation [Dombroski et al., 1991; Konkel et al., in press; Seleme et al., 2006].

STRUCTURE OF AN L1 RETROTRANSPOSON

The full-length human L1 retrotransposon is ~6 kb, and is comprised of a 910-bp 5′ untranslated region (5′UTR), two open reading frames (ORF1 and ORF2) separated by a 63-bp inter-ORF linker region, and a 205-bp 3′UTR containing a functional polyadenylation signal. The L1 element terminates with a variable-length polyA tail, and is typically flanked by short direct repeats (target site duplications [TSDs]) of 2–20 bp (Fig. 2) [Scott et al., 1987; Dombroski et al., 1991; Szak et al., 2002]. Both ORF1 and ORF2 are required for retrotransposition [Moran et al., 1996].

ORF1 encodes a ~40-kDa protein (p40 or ORF1p) rich in basic residues that contains a leucine zipper domain [Holmes et al., 1992]. ORF1p was found to form ribonucleoprotein particles (RNPs) in human and mouse teratocarcinoma cell lines [Martin, 1991; Hohjoh and Singer, 1996], and its direct RNA-binding ability was further confirmed in vitro, with data suggesting that RNA binding is not sequence-specific at least for the mouse ORF1p [Hohjoh and Singer, 1997; Kolosha and Martin, 1997, 2003; Martin et al., 2000]. Both human and mouse ORF1p possess nucleic acid chaperone activity, which is independent of their RNA-binding affinity and is required for L1 retrotransposition. ORF1p accelerates complementary strand annealing, and facilitates melting and strand displacement in mispaired DNA duplexes [Martin and Bushman, 2001; Kulpa and Moran, 2005; Martin et al., 2005a; Martin, 2006].

ORF2 encodes a ~150-kDa protein, ORF2p, with conserved endonuclease (EN) [Feng et al., 1996], reverse transcriptase (RT) [Mathias et al., 1991], and C-terminal cysteine-rich [Fanning and Singer, 1987] domains. The EN domain of L1 is thought to originate from the host cells apurinic/apyrimidinic endonuclease (APE) in early eukaryotes, and is conserved among a large subset of non-LTR retrotransposons [Martin et al., 1995; Feng et al., 1996; Malik et al., 1999]. Analyses of L1 and Alu integration sites from genomic databases, as well as studies of de novo integrants, identified a loose EN cleavage consensus of 5′ TTTT′ AA 3′, which facilitates EN interactions with the DNA minor groove [Feng et al., 1996; Jurka, 1997; Cost et al., 2002; Morrish et al., 2002; Berry et al., 2006]. L1 elements lacking a functional EN

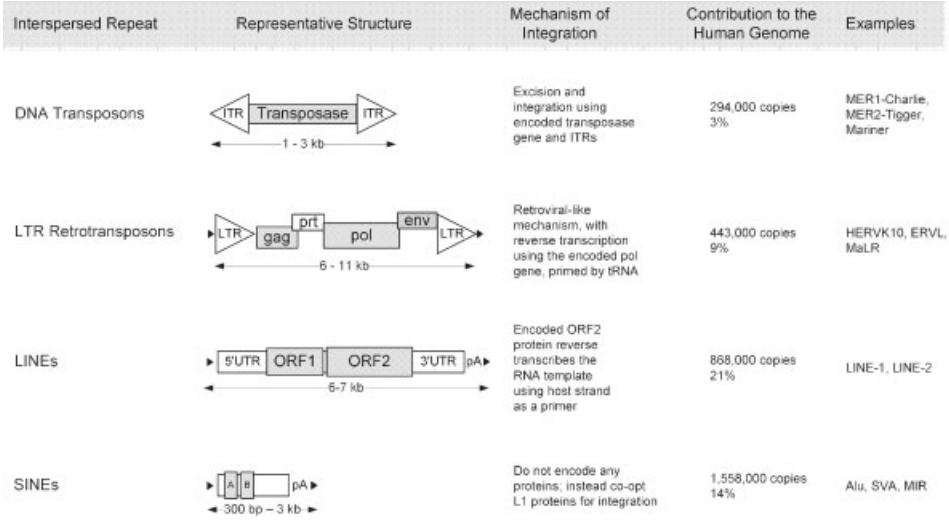

FIGURE 1. **Classes of interspersed repeats. ITR, inverted terminal repeats; LTR, long terminal repeats; *gag*, group specific antigen; *prt*, protease; *pol*, polymerase; *env*, envelope; A and B are the conserved regions of the RNA polymerase III promoter in Alu element. The *env* gene is nonfunctional in LTR-retrotransposons.**

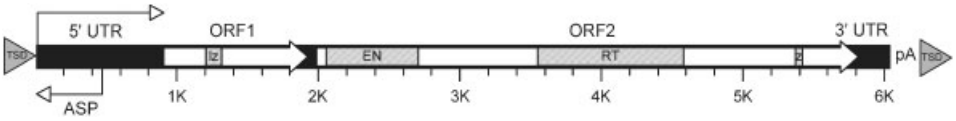

FIGURE 2. **Structure of the human L1 retrotransposon. The full-length human L1 retrotransposon is 6 kb, and contains: 1) a 910-bp 5′UTR region with bidirectional promoter activity; 2) ORF1 region, which encodes a 40-kDa basic RNA-binding protein with a leucine zipper domain (lz); 3) ORF2 region, which encodes a 150-kDa protein with endonuclease (EN) and reverse transcriptase (RT) activities, and a conserved C-terminal zinc knuckle domain (z); and 4) a 3′UTR region dispensable for retrotransposition activity. The 3′UTR contains a functional polyadenylation signal, which occasionally may be bypassed in favor of a stronger downstream signal [Moran et al., 1999; Goodier et al., 2000; Pickeral et al., 2000; Belancio et al., in press]. L1 terminates with a poly(A) sequence (pA), and is flanked by 2-bp to 20-bp target site duplications (TSDs).**

domain were shown to integrate at atypical sites in DNA-repair deficient cell lines, suggesting that L1 may utilize preexisting nicks for integration. However, endonuclease-independent integration is significantly less efficient than integration mediated by L1 endonuclease [Morrish et al., 2002] and is rarely found in vivo [Chen et al., 2005a; Babushok et al., 2006].

The RT domain, defined by the presence of 11 conserved sequence blocks, is present in all non-LTR retrotransposons, and has been used to date the emergence of non-LTR elements to the Precambrian era, making them as old as eukaryotes [Malik et al., 1999]. Biochemical studies of L1 RT using a Ty1/L1 ORF2 hybrid vector and L1 ORF2 overexpressed in baculovirus-infected insect cells demonstrated that L1 RT possesses both RNA- and DNA-dependent DNA polymerase activities, which are Mg^{2+}-dependent, highly processive, and require an intact catalytic motif FXDD [Mathias et al., 1991; Dombroski et al., 1994; Piskareva et al., 2003; Piskareva and Schmatchenko, 2006]. Furthermore, reverse transcriptase activity was detected in ultracentrifugation-concentrated whole cell lysates containing L1 RNA, ORF1p, and ORF2p derived from cultured cells transfected with an L1 expression vector [Kulpa and Moran, 2005, 2006].

The C-terminal cysteine-rich domain of L1 contains a CCHC zinc knuckle-like region [Fanning and Singer, 1987] and is conserved throughout the LINE-element clade [Duvernell and Turner, 1998; Goodwin et al., 2001; Ostertag and Kazazian, 2001a; Casaregola et al., 2002]. Mutations in the conserved cysteine and histidine residues reduce L1 retrotransposition activity in a cultured cell assay, without adversely affecting reverse transcriptase activity [Moran et al., 1996]. The zinc knuckle region is hypothesized to mediate ORF2-DNA interactions during L1 integration [Moran et al., 1996; Ostertag and Kazazian, 2001a], or function as an intramolecular processivity factor facilitating extended polymerization by the L1 RT [Piskareva and Schmatchenko, 2006].

CELL TYPE SPECIFICITY OF L1 RETROTRANSPOSITION

Only a few studies have focused on L1 activity in somatic tissues; nevertheless, it is clear that in some circumstances L1 can be expressed and retrotranspositionally-active somatically. Endogenous mouse L1 transcripts and ORF1 protein were found in somatic subsets of cells in the testes and in steroidogenic tissues [Branciforte and Martin, 1994; Trelogan and Martin, 1995]; endogenous human ORF1 and ORF2 proteins were detected in somatic subsets of testicular cells (Sertoli, Leydig, testicular covering cells, and epididymal columnar epithelial cells), placental syncytiotrophoblasts, and vascular endothelial cells [Ergun et al., 2004]; endogenous rat L1 RNA and protein were detected in cardiac tissue [Lucchinetti et al., 2006]; Miki et al. [1992] and, possibly, Morse et al. [1998] found somatic L1 insertions in tumor tissue in human patients; and human L1 was found to retrotranspose in neuronal precursor cells in transgenic mice [Muotri et al., 2005]. In addition to a wealth of evidence of L1 retrotransposition in transformed somatic cells in culture [Moran et al., 1996; Morrish et al., 2002; Han et al., 2004], L1 retrotransposition in primary human cells and primary neuronal rat cells was recently demonstrated [Muotri et al., 2005; Kubo et al., 2006; Shi et al., 2007].

The frequency and cell-type requirements of retrotransposition are thought to be largely determined by methylation and subsequent repression of the CpG-rich L1 5′UTR [Yu et al., 2001], with general repression in normal somatic tissues, and with aberrant hypomethylation allowing for expression of L1 in some human tumors [Dante et al., 1992; Asch et al., 1996; Florl et al., 1999; Santourlidis et al., 1999; Roman-Gomez et al., 2005]. While several transcription factors have been implicated in transcriptional regulation of L1 (discussed below), none were demonstrated so far to determine the cell type specificity of L1 expression. Two other potential types of L1 regulators were recently proposed: L1-specific siRNAs (discussed below), as well as several APOBEC3 human cytidine deaminases. Several APOBEC3 family members (APOBEC3A, 3B, 3C, and 3F) were found to inhibit L1 retrotransposition independent of their deaminase activity; their mechanism of L1 inhibition remains unclear. APOBEC3 enzymes have been proposed to mediate cytoplasmic sequestration of L1 RNA or proteins, or to directly inhibit L1 enzymatic activity [Bogerd et al., 2006; Chen et al., 2006a; Muckenfuss et al., 2006; Stenglein and Harris, 2006]. APOBEC3G was recently found to inhibit Alu, but not L1 retrotransposition, by sequestering Alu RNAs in cytoplasmic complexes away from the L1 retrotransposition machinery [Turelli et al., 2004; Chiu et al., 2006; Hulme et al., in press].

Although L1 transcription and retrotransposition may occur in a variety of cell types at the level of an individual, only events in primordial germ cells, germline, or the early embryo can be incorporated into the germline lineage and contribute to future generations. Of these, only germline retrotransposition has been demonstrated: de novo L1 insertion consistent with early meiotic retrotransposition was reported in a human patient [Brouha et al., 2002], endogenous full-length L1 transcripts were detected in prepubescent spermatocytes [Branciforte and Martin, 1994], and germline L1 transcription and retrotransposition were observed in transgenic mouse models of human L1 retrotransposition [Ostertag et al., 2002].

While there is some evidence of retrotranspositional activity in early development or in later somatic tissues from transgenic studies, these results were obtained with either a native human L1 or a synthetic mouse L1 element driven by a heterologous promoter in a mouse host. De novo retrotransposition in early embryogenesis was described by Prak et al. [2003] in the transgenic founder using a human L1 driven by the mouse RNA

pol II promoter in addition to the L1 5′UTR. Subsequently, high frequency somatic insertions were observed in multiple mouse lines from an integrated human L1 transgene driven by the mouse heat shock (Hsp70-2) promoter [Babushok et al., 2006], as well as in transgenic lines containing a synthetic mouse L1 element driven by the chicken β-actin promoter [An et al., 2006]. Future studies of L1 transcribed from its endogenous 5′UTR will determine whether somatic retrotransposition occurs in vivo.

TRANSCRIPTIONAL REGULATION OF L1 RETROTRANSPOSON

The 5′UTR of human L1 contains an internal promoter located within its first 100–150 bp, with additional sequences in the first 670 bp contributing to transcriptional activation [Swergold, 1990; Minakami et al., 1992; Singer et al., 1993]. This internal promoter initiates transcription within a small region from −9 to +4 of L1 [Lavie et al., 2004]. Internal promoters are typically associated with short noncoding RNAs produced by RNA polymerase III (pol III) [Geiduschek and Kassavetis, 2001]; and pol III-, but not pol II-mediated transcription of L1 was found in an in vitro transcription study using L1 promoter driving transcription of a short L1-region [Kurose et al., 1995]. However, the assayed template lacked long coding ORF sequences and several polyT stretches normally contained in L1 that would likely terminate pol III transcription [Geiduschek and Kassavetis, 2001]. In fact, several characteristics of L1 are more consistent with pol II transcription. The L1 transcript is long (6 kb), protein-coding, and contains a functional polyadenylation signal, as is typical for pol II-transcribed mRNAs. Additionally, from the prevalence of extra 5′ G residues consistent with reverse transcription of the m7G cap, as well as from the bimodal distribution of L1 insertion lengths, suggestive of a protected 5′ end of the full-length elements, we and others have postulated the presence of a 5′m7G cap [Lavie et al., 2004; Babushok et al., 2006], a characteristic feature of pol II transcripts. Finally, RNA polymerase II binding to the L1 5′UTR was detected in chromatin precipitation experiments [Yang and Kazazian, 2006]. Thus, the wealth of suggestive data is most consistent with RNA polymerase II-driven transcription of full-length L1s, or, possibly, with transcription by a combination of pol II and pol III. The efficiency of L1 transcription is greatly diminished due to an L1 length-dependent transcriptional elongation defect [Han et al., 2004]; furthermore, premature polyadenylation limits the amount of full-length L1 transcripts available for retrotransposition [Perepelitsa-Belancio and Deininger, 2003].

In addition to the internal promoter activity driving L1 expression, Speek [2001] demonstrated that L1 5′UTR contains antisense promoter (ASP) activity, which allows for cotranscription of L1 5′UTR and its upstream genomic region. He mapped the antisense promoter activity to nucleotides +400 to +600, with the minimal promoter region located at +399 to +467, and additional 3′ nucleotides contributing to ASP activity. Further analyses of expressed sequence tag (EST) databases revealed that ASP drives transcription of a number of genes, is active across a range of tissues, and may contribute to epigenetic silencing of cellular genes [Nigumann et al., 2002; Matlik et al., 2006] as well as the L1 elements themselves [Soifer and Rossi, 2006; Yang and Kazazian, 2006]. Using a combination of transcription start database analyses [Matlik et al., 2006], as well as 5′RACE transcription start mapping [Yang and Kazazian, 2006], heterogeneous ASP transcription initiation sites were localized to two regions at +378–431 and +480–497 bp. Yang and Kazazian [2006] further hypothesized that L1 retrotransposons themselves may be suppressed by the RNA interference effect caused by bidirectional transcripts from the L1 5′UTR. Using Northern analysis with probes specific for the predicted region of double-stranded RNA, they found the presence of ∼21-bp small RNAs complementary to the L1 5′UTR. Consistent with the RNAi-mediated suppression of L1, both the deletion of the ASP promoter and knock-down of Dicer-1 increased L1 retrotransposition activity in cell culture [Yang and Kazazian, 2006]. While the mouse L1 element is not known to have an antisense promoter, 20–24 nt siRNAs complementary to mouse L1 sequences were also found in the mouse germline, suggesting that RNAi could be a common mechanism of L1 suppression [Watanabe et al., 2006].

Studies of transcription-factor binding to the L1 promoter region identified a number of proteins that bind to sequences in the L1 5′UTR and modulate promoter activity. Yin Yang-1 (YY1) transcription factor was found to bind to bases +13 to +21 and increase promoter activity [Minakami et al., 1992; Becker et al., 1993; Singer et al., 1993; Kurose et al., 1995]; subsequent detailed mutagenesis studies of the YY1-binding site further defined its function to directing transcript initiation to the +1 position of the L1 element [Athanikar et al., 2004]. Two binding sites for SOX family members at +472 to +477 and at +572 to +577 of the L1 5′UTR were shown to be functional, leading to transactivation of up to 10-fold in cotransfection experiments with Sox11 transcription factor and the 5′UTR driving a luciferase reporter [Tchenio et al., 2000]. Furthermore, RUNX3 transcription factor binding sites were identified at +83 to +101 and +526 to +508, the first of which was demonstrated to activate L1 transcription and retrotransposition in RUNX3 overexpression experiments; mutations in the second RUNX3 site were found to modulate the activity of the L1 antisense promoter [Yang et al., 2003; Yang and Kazazian, 2006].

L1 ORF1 AND ORF2 ARE TRANSLATED BY DIFFERENT MECHANISMS FROM A BICISTRONIC TRANSCRIPT

A scanning model of translation, accounting for translation of the majority of eukaryotic mRNAs, postulates 40S ribosome binding to the m7G cap at the 5′ end of the transcript, followed by linear scanning until the first AUG in the appropriate Kozak initiation context [Kozak, 1989]. From the sharp decrease in translation observed when a stable hairpin structure was introduced at +661 of L1 5′UTR, human ORF1 translation is most consistent with a standard ribosomal scanning model [McMillan and Singer, 1993]. Consistent with standard translation, ORF1p can be readily detected in a variety of cells. Human ORF2 is located downstream of ORF1, and the two ORFs are separated by a 63-bp interORF region that contains two stop codons and has the potential to encode a short, 6–amino acid ORF [Dombroski et al., 1991]. In contrast to ORF1p, human ORF2p could not be detected at significant levels from overexpressed intact L1 constructs; however, other factors such as ORF2-mediated toxicity or low antibody sensitivity could have contributed to the apparent lack of ORF2 translation [Ergun et al., 2004; Goodier et al., 2004].

While it is formally possible that ORF2 protein is transcribed from a post-transcriptionally processed L1 transcript with the spliced out leader 5′UTR and ORF1 [Belancio et al., 2006] or from a shorter monocistronic transcript transcribed using a

hypothetical cryptic promoter in ORF1, several lines of evidence argue against this possibility. There is no evidence of transcription of L1 from downstream sites. Retrotransposed monocistronic transcripts are not found in the human genome [Belancio et al., 2006]. Furthermore, strong *cis*-preference of ORF2p for the RNA from which it was translated (see below) makes it unlikely that ORF2p translated from a monocistronic transcript could mediate retrotransposition of a full-length L1 RNA. Most probably, ORF2 protein is produced from a full-length bicistronic L1 transcript [Skowronski et al., 1988; Perepelitsa-Belancio and Deininger, 2003]. Translation from a single bicistronic transcript was recently demonstrated for another eukaryotic non-LTR retrotransposon, SART1, in silkworm [Kojima et al., 2005].

Translation of ORF2 by leaky m7G cap-dependent ribosomal scanning is very unlikely, because of the long leader sequence containing 11 in-frame AUG codons upstream of ORF2 [Alisch et al., 2006]. Similarly, ORF2 translation by nonsense codon suppression, ribosomal frameshifting, or ribosomal hopping is also unlikely, because there is no evidence of ORF1-ORF2 fusion protein when probed with anti-ORF1 antisera [Leibold et al., 1990; Bratthauer and Fanning, 1992, 1993; McMillan and Singer, 1993; Branciforte and Martin, 1994; Asch et al., 1996; Hohjoh and Singer, 1996; Ishida and Leder, 1999; Goodier et al., 2004; Kulpa and Moran, 2005]. Ergun et al. [2004] recently reported ORF2p in immunoprecipitation experiments in a restricted subset of testicular and placental cells. In agreement with separate translation of ORF1 and ORF2, no ORF1-ORF2 fusion protein was detected with anti-ORF2 antibodies in these experiments.

Based on in vitro translation experiments, Ilves et al. [1992] originally proposed that rat L1 ORF2 translation must occur by either translational termination/reinitiation or an internal ribosome entry mechanism. A similar conclusion was reached from translation studies of human ORF2 fused with a lacZ reporter in cultured cells [McMillan and Singer, 1993]. Understanding of human ORF2 translation was further refined in recent studies by Alisch et al. [2006]. They found that while the ORF2 AUG is the translation start site for ORF2p and is present in all known active L1 retrotransposons, it can be substituted by any other non-stop codon with preservation of 10 to 70% of wild-type retrotransposition activity. In contrast, replacement of the AUG by a stop codon virtually eliminates ORF2 translation; this effect can be rescued by a single-nucleotide insertion after the stop codon that reinstates the reading frame. The authors further found that ORF2 translation requires the presence of an upstream ORF; however, the sequence of the ORF and the interORF region are not important for maintaining translation. In fact, replacement of ORF1 by its inactive mutant, but not by a prematurely truncated mutant, allows for ORF2 translation. Similarly, translation of ORF2 is preserved if the upstream ORF1 is replaced in its entirety by a heterologous ORF (such as the EGFP gene).

Because any upstream ORF will produce ORF2 translation while the ORF1 and interORF sequences are dispensable, it is unlikely that ORF2 is translated from an internal ribosome entry site (IRES) in these regions. Instead, ORF2 is most probably translated by a termination/reinitiation mechanism, whereby the 40S subunit translates an upstream ORF in an m7G-cap dependent manner, and then scans the interORF region until it reaches the 5′ end of the ORF2; the immediate sequence at the 5′ end of ORF2 is likely used to position the ribosomes at or near ORF2 AUG codon [Alisch et al., 2006]. In agreement with this hypothesis, Alisch et al. [2006] observed a drastic drop in ORF2p translation when a thermostable hairpin loop was introduced into the interORF region; in contrast, introducing an unstructured 131-bp sequence did not affect ORF2 translation.

The termination/reinitiation seen in human L1 may be related to the translational coupling mechanism responsible for production of ORF2p from a bicistronic transcript of SART1 retrotransposon [Kojima et al., 2005]. However, it appears that the translation mechanisms of human and rodent L1 proteins are distinct. While the human ORF1 is thought to be translated by classic ribosomal scanning, and ORF2 by a translation termination/reinitiation mechanism [McMillan and Singer, 1993; Alisch et al., 2006], Li et al. [2006] reported IRES sequences guiding the translation of both mouse ORF1 and ORF2 proteins; neither of the two identified IRES regions are well-conserved in mammalian LINE elements. Internal ribosomal entry was also hypothesized to mediate rat ORF2 translation [Ilves et al., 1992]. Future work in this area will undoubtedly uncover further details of the ORF2 translation mechanism among LINEs and other non-LTR retrotransposons.

The low amounts of ORF2p, with possibly as few as one ORF2p molecule present in the L1 ribonucleoprotein (RNP) complex [Wei et al., 2001; Alisch et al., 2006], indicate that either ORF2 translation is extremely inefficient or that ORF2p is quickly eliminated post-translationally. The resulting paucity of ORF2p likely contributes to the phenomenon of L1 *cis*-preference, whereby the nascent ORF2 protein is thought to associate with its encoding L1 mRNA in an RNP particle, which then serves as a retrotransposition intermediate [Kulpa and Moran, 2006].

The *cis*-preference model of L1 retrotransposition is supported by four lines of evidence: 1) disease-causing L1 insertions in humans are identical in sequence to their active progenitor L1 elements despite the vast majority of genomic elements being rearranged and inactive [Dombroski et al., 1991; Holmes et al., 1994]; 2) the low activity of transfected L1 elements with inactive ORF1 or ORF2 cannot be rescued in *trans* by the intact endogenous cellular L1s or cotransfected wild-type L1s [Moran et al., 1996; Wei et al., 2001]; 3) L1s mobilize cellular mRNAs at much lower frequency (~0.01–0.05%) than L1 RNAs [Esnault et al., 2000; Wei et al., 2001]; and 4) the RT activity of L1 RNP complexes preferentially acts on L1 RNA templates [Kulpa and Moran, 2006].

L1 INTEGRATES INTO THE GENOME BY TARGET-PRIMED REVERSE TRANSCRIPTION, LIKELY INVOLVING TWO ORF2 COMPLEXES

The integration of L1 retrotransposon occurs by the target-primed reverse transcription (TPRT) mechanism, which was originally described for the related non-LTR retrotransposon R2 in *Bombyx mori* [Luan et al., 1993]. In this mechanism (Fig. 3), the L1-encoded endonuclease (EN) within the ORF2 protein nicks the host chromosome. The free 3′-hydroxyl generated by the nick is then used as the primer by the L1 RT to polymerize the cDNA copy of L1 directly onto the host DNA. Subsequent biochemical studies using purified ORF2 demonstrated a TPRT reaction of human L1 in vitro [Cost et al., 2002]. Analyses of endogenous L1s in the human genome [Jurka, 1997; Lander et al., 2001; Szak et al., 2002] further confirmed that L1 integrants carry characteristic features of TPRT integration: L1 sequences generally integrate at 5′ TTTT′AA 3′ sites, contain an intact 3′ end region ending in a polyA, are variably truncated at the 5′ end, and are flanked by 2 to 20-bp target site duplications.

Unlike the initial TPRT steps of L1 integration, subsequent second-strand cleavage, L1 RNA removal, second-strand synth-

esis, and nick resolution are poorly understood. From studies of the R2 retrotransposon, a second R2 protein was shown to mediate second-strand cleavage, and predicted to mediate second-strand synthesis; second-strand synthesis by R2p has not yet been demonstrated experimentally [Christensen and Eickbush, 2005]. Evidence of L1-mediated cleavage of both strands is provided by recent demonstrations of γ-H2AX foci indicative of frequent double-strand breaks in cultured cells transiently transfected with an L1 construct; L1-induced double-strand breaks disappeared over time, consistent with DNA repair and integration resolution [Belgnaoui et al., 2006; Gasior et al., 2006].

Similar to the integration of the R2 element, the second RT subunit is also thought to be involved in the integration of at least some L1 molecules; this is predicted from the "twin priming" mechanism that explains frequent inversions of mammalian L1 elements (Fig. 3) [Ostertag and Kazazian, 2001b]. In twin-priming, the second DNA strand is cleaved during the TPRT reaction, releasing a second 3′ hydroxyl group, which is then used for synthesis of the upstream region; subsequent resolution produces an element containing a 3′ direct region joined to a 5′ inverted region, with the whole structure flanked by target site duplications [Ostertag and Kazazian, 2001b].

The attachment of the 5′end of the R2 element is postulated to occur by R2 RT template jumping from the 5′ end of the R2 transcript onto the host DNA strand [George et al., 1996; Burke et al., 1999; Eickbush et al., 2000; Bibillo and Eickbush, 2002]. Efficient template jumping by the R2 RT from the 5′ end of the RNA template to the 3′ end of single-stranded DNA and RNA regions has been demonstrated in vitro [Bibillo and Eickbush, 2002, 2004]. R2 template jumping is often accompanied by the addition of a few extra nucleotides to the end of the nascent cDNA at the site of the jump, which may facilitate annealing between the nascent cDNA strand and the new template [Bibillo and Eickbush, 2004].

Several features of de novo and recent genomic L1 insertions indicate that L1 5′ end attachment could also be mediated by similar L1 RT template jumping from the 5′ end of the L1 RNA to the single-stranded region of the host DNA at the integration site

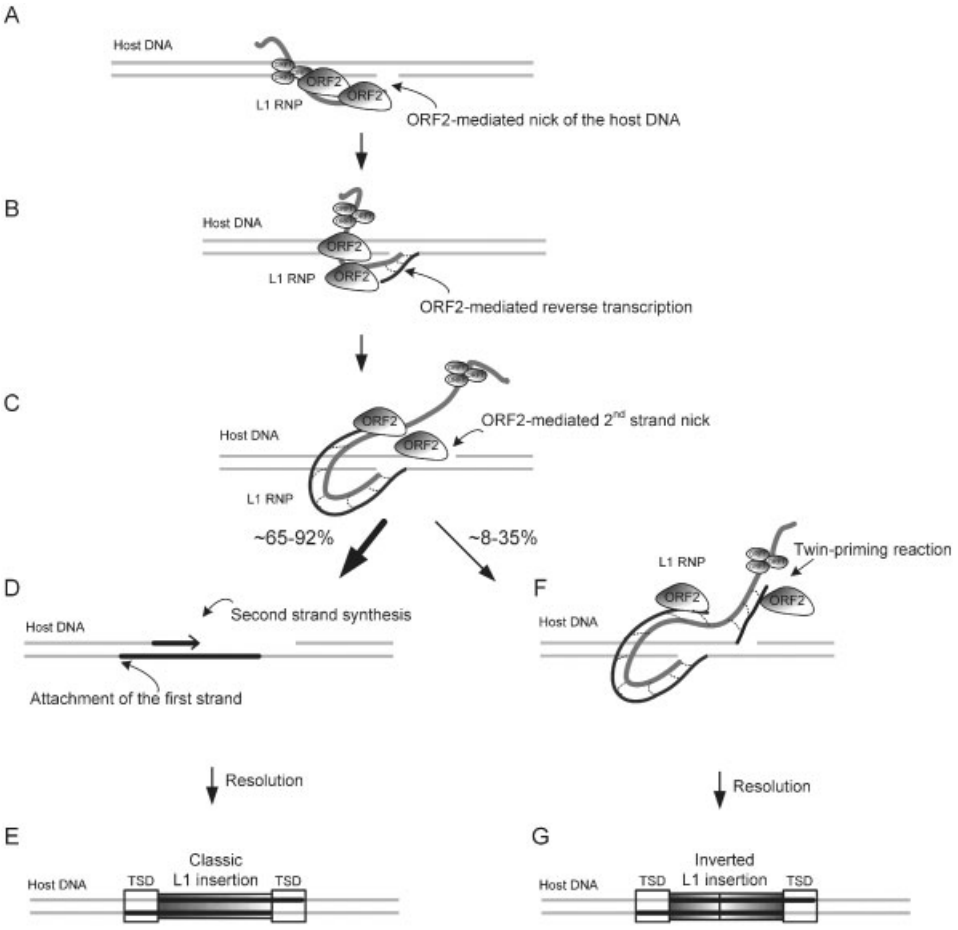

FIGURE 3. **Model of L1 integration reaction. L1 RNA is associated in the L1 ribonucleoprotein complex (L1 RNP) with the ORF1 chaperone protein and ORF2 protein containing the endonuclease (EN) and reverse transcriptase (RT) activities. The L1 RNP is the active intermediate in the target-primed reverse transcription reaction (TPRT). A:** In TPRT, L1 ORF2p first nicks the bottom strand of host DNA, releasing a 3′ hydroxyl. **B:** L1 ORF2p then uses L1 RNA (thick gray line) as a template, and the nicked host strand as a primer to reverse transcribe the L1 transcript directly into the host genome (black line shows nascent L1 retrocopy). **C:** L1 ORF2p nicks the top strand of host DNA. **D:** The majority of insertions first integrate the newly copied bottom strand, and then synthesize the complementary second strand of DNA. **E:** Resolution of the integrant results in a classic L1 insertion, which is flanked by target site duplications (TSDs, boxes), and contains a variable length direct fragment of L1 sequence beginning at the 3′ end, and often truncated at the 5′ end. **F:** In up to 35% of L1 insertions, after the nick of the top strand of host DNA, a second ORF2 molecule initiates a second reverse transcription reaction, copying an upstream region of the L1 transcript; this reaction is called twin-priming. **G:** Resolution of a twin-priming integrant produces an inverted L1 insertion, which contains a direct fragment with an intact 3′ end and a variably-truncated 5′ end joined to an inverted copy of an upstream region of L1. The whole insertion is flanked by target site duplications (TSDs).

(Fig. 4) [Babushok et al., 2006]. First, occasional template jumping from L1 RNA to other RNA templates during L1 integration is known to give rise to chimeric insertions (e.g., L1-U6) [Hayward et al., 1997; Buzdin et al., 2002, 2003; Gilbert et al., 2002, 2005]. Second, similar to the R2 element [George et al., 1996; Eickbush et al., 2000], de novo L1 insertions occasionally contain extra nucleotide stretches at the 5′ junction with genomic DNA. Many of these are comprised of only one to a few nucleotides, and, as was proposed for R2 element, these could be added by the RT during the template jump [Bibillo and Eickbush, 2004]. Other extra nucleotide additions duplicate host DNA sequences immediately upstream of the integration site [Symer et al., 2002; Babushok et al., 2006], and these are most consistent with L1 RT jumping templates from the L1 RNA onto the upstream host DNA region. Third, consistent with the ability to use host DNA as a template for DNA synthesis, L1 RT has efficient DNA-dependent DNA polymerase activity [Mathias et al., 1991; Piskareva et al., 2003; Piskareva and Schmatchenko, 2006]. Finally, the majority of characterized de novo integrants in cultured cells and in transgenic mice as well as a large number of endogenous L1s in the human and mouse genomes [Gilbert et al., 2002, 2005; Symer et al., 2002; Martin et al., 2005b; Zingler et al., 2005; An et al., 2006; Babushok et al., 2006] contain microhomologies between the 5′ end of L1 RNA and the flanking host DNA. These microhomologies could be produced by annealing between the 3′ end of nascent L1 cDNA and the single-stranded region of the top host strand, facilitating the L1 RT template jump [Babushok et al., 2006].

Both the initial TPRT model, as well as the template jumping model invoke cellular enzymes for final resolution and nick repair [Luan et al., 1993; Ostertag and Kazazian, 2001b; Babushok et al., 2006]. One DNA repair enzyme, ataxia telangiectasia mutated (ATM), was recently shown to be required for L1-induced double-strand break repair and L1 retrotransposition [Gasior et al., 2006]. Further studies will be required to evaluate the template-jumping model of L1 integration and to determine the involvement of host enzymes in the L1 lifecycle.

L1 RETROTRANSPOSON CONTRIBUTES TO HUMAN DISEASE

While the majority of L1 and other L1-mediated insertions (e.g., Alu and SVA insertions) land in intergenic and intronic sequences with little or no consequence to their host, occasional insertions have disrupted gene expression and caused human disease [Kazazian et al., 1988; Deininger and Batzer, 1999; Batzer and Deininger, 2002; Chen et al., 2005b, 2006b; Ostertag and Kazazian, 2005; Hulme et al., 2006; Mine et al., 2006; Musova et al., 2006]. Because these insertions are typically identified in patients presenting with overt disease, the majority of characterized cases are X-linked or dominant single-gene genetic disorders, such as hemophilia A (MIM# 306700) or Apert syndrome (MIM# 101200). Occasional cases of recessive inactivating mutations have been found on the background of incidental inherited mutation in the other allele (e.g., pyruvate dehydrogenase complex deficiency; MIM# 608769) [Mine et al., 2006], in consanguineous populations (e.g. gyrate atrophy of the choroid and retina; MIM# 258870) [Mitchell et al., 1991], as well as in cases where second-hit mutations are more frequent and likely to produce disease (e.g., insertions into tumor suppressor genes, such as the APC gene; MIM# 175100) [Miki et al., 1992] (reviewed in Ostertag and Kazazian [2005]). Of 53 known disease-causing insertions, 17 were caused by L1 itself, while L1-mediated integrations of Alu and SVA elements caused another 29 and four cases, respectively; three additional insertions were caused by L1-mediated insertions of simple polyA repeats.

Of the 17 L1 insertions, 16 were caused by L1 elements belonging to the younger, human-specific subset of L1s (Ta, transcribed group a subset) [Skowronski et al., 1988; Boissinot et al., 2000], whereas only one is derived from the older, pre-Ta subset [Kazazian et al., 1988]. The predominance of disease-causing Alu insertions is in part due to the underdiagnosis of L1 integrants: L1 insertions are generally longer than Alu elements and may be more difficult to identify by PCR-based approaches [Chen et al., 2006b]. Importantly, endogenous Alus appear to integrate at efficiencies similar to L1 elements, as estimated by evolution- and mutation-based analyses: roughly one insertion in every 20 births for Alu [Cordaux et al., 2006] vs. one insertion in every eight births for L1 [Kazazian, 1999]; the estimated in vivo Alu integration rates are ~10 times higher than those in cultured cells using transfected Alu elements [Dewannieux et al., 2003]. While experimental mobilization of SVA elements has not been demonstrated so far, SVAs appear to be evolutionarily younger than both L1 and Alu elements, and are thought to have relatively high retrotransposition activity [Ostertag et al., 2003; Wang et al., 2005].

Despite the fact that exonic regions comprise only a small fraction of eukaryotic genes, the majority of published disease-causing insertions (10/17 for L1, 20/29 for Alu, and 2/4 for SVA) disrupt the coding regions of genes. The predominance of exonic insertions reflects the fact that they are most likely to disrupt a gene by causing frameshifts and subsequent premature stop codons or exon-skipping. Rarely, an integration event may be accompanied by a large genomic deletion at the insertion site that may eliminate coding sequence [Gilbert et al., 2002, 2005; Symer et al., 2002; Chen et al., 2005b, 2006b; Mine et al., 2006]; in one recent case, an intronic L1 insertion in the PDHX gene was accompanied by a ~46-kb deletion eliminating exons 3 to 9 of the PDHX gene and causing pyruvate dehydrogenase complex deficiency (MIM# 608769) [Mine et al., 2006]. The remainder of disease-causing insertions disrupt introns, or the 5′ or 3′UTRs of genes; in these cases gene disruption has been attributed to missplicing, exon skipping, or a general decrease in mRNA levels, but in the majority of cases the molecular mechanism leading to decrease in expression has not been characterized [Ostertag and Kazazian, 2005; Chen et al., 2006b]. Surprisingly, several independent disease-causing retrotransposon insertions integrated in the same gene at the same nucleotide position, indicating that certain chromatin environments may be more susceptible to L1 integration [Chen et al., 2005b; Conley et al., 2005].

Progress in understanding of L1 and Alu biology can shed light on most reported cases with an unknown mechanism of insertion-related mRNA decrease (Fig. 5). In addition to the direct integration into splicing regulatory regions, intronic L1 insertions may decrease RNA levels of their host gene by: 1) L1 length-dependent transcriptional elongation defects [Han et al., 2004]; 2) transcriptional interference from the L1 bidirectional promoter region [Speek, 2001]; 3) post-transcriptional suppression by RNAi effects [Yang and Kazazian, 2006]; and 4) premature polyadenylation [Perepelitsa-Belancio and Deininger, 2003; Wheelan et al., 2005]. GC-rich Alu, SVA, and the 5′UTR region of L1 may introduce methylation into a gene's regulatory region and repress transcription [Batzer and Deininger, 2002; Schulz et al., 2006]. Both L1 and Alu insertions have been associated with large deletions of target site sequences [Gilbert et al., 2002, 2005; Symer et al., 2002; Chen et al., 2005b, 2006b; Mine et al., 2006], and

both contain functional splice sites, which may interfere with the host gene splicing [Makalowski et al., 1994; Lev-Maor et al., 2003; Ast, 2004; Belancio et al., 2006; Zhang and Chasin, 2006].

In addition to causing disease by de novo insertions, existing genomic L1 and Alu elements can be potent substrates for unequal homologous recombination, causing gain or loss of genomic sequence [Jurka, 2004]. High recombinogenic potential of Alu elements was noted previously in studies of the junctions of segmental duplications [Babcock et al., 2003; Bailey et al., 2003], and indeed the majority of known disease-causing recombination events have occurred between Alu elements. A total of 33 of these occurred in the germline causing disorders such as familial hypercholesterolemia (MIM# 143890), α-thalassemia (MIM# 141800), type VI variant of Ehlers-Danlos syndrome (MIM# 225400), and Tay Sachs disease (MIM# 272800); 16 recombination events occurred somatically to cause a variety of cancers, including ALL1 rearrangement leukemias (MIM# 159555) and BRCA1-associated familial breast cancer (MIM# 113705) [Deininger and Batzer, 1999]. There are also three known cases of recombination between L1s causing Alport syndrome (MIM# 301050), phosphorylase kinase deficiency of liver and muscle (MIM# 261750), and ataxia-telangiectasia (MIM# 208900) [Ostertag and Kazazian, 2005]. Certain genes, such as the LDLR gene that has hosted eight independent recombination events, appear to be preferentially involved in recombination. However, how much of this apparent preference is caused by bias of ascertainment and differences in genetic diagnostic techniques is not clear.

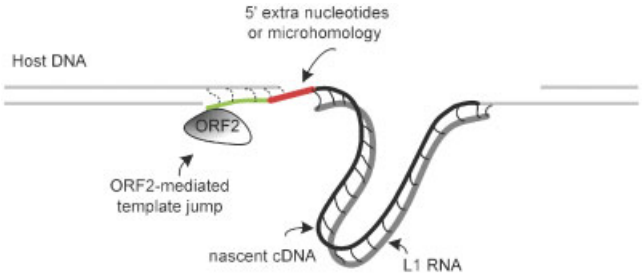

FIGURE 4. **The template jumping model of L1 5′ end attachment. After the initial TPRT reaction and cleavage of the second strand of host DNA (Fig. 3C), L1 ORF2p (containing L1 RT activity) reaches the 5′ terminal end of the L1 RNA and jumps templates onto the upstream overhang of the host DNA strand. Addition of several extra nucleotides or annealing between the 3′ end of nascent cDNA (shown in red) and the overhang of host DNA facilitates this jump. L1 RT then likely continues copying the host strand region (shown in green). The structure is finally resolved by ligation or homologous recombinational repair.**

CONCLUSION

LINE-1 is the only autonomous mobile element known to be currently active in humans and is the largest contributor to mammalian genome mass. Differences in L1 content and total retrotransposition potential contribute to interindividual variation in gene expression and mutagenic load. Continued

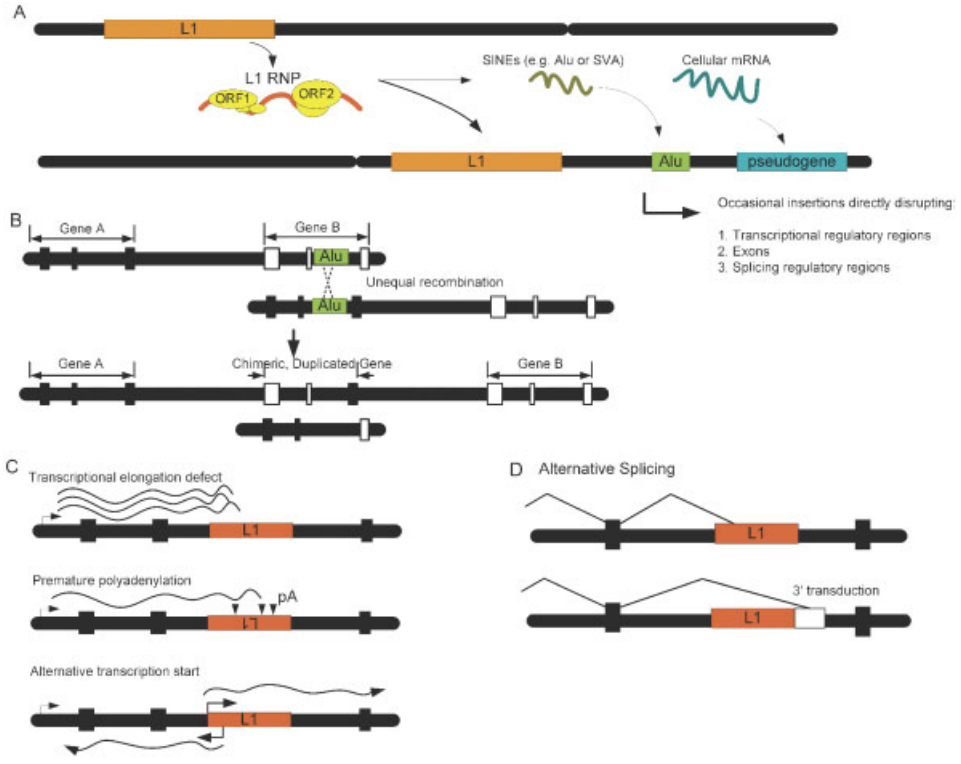

FIGURE 5. **Mechanisms of L1-mediated gene disruption. A: Integration of LINE-1, SINE elements Alu and SVA, and other cellular transcripts by L1 retrotransposition machinery may occasionally cause disease by direct disruption of transcriptional regulatory regions, exons, or splicing regulatory regions. B: Recombination between existing genomic L1 or Alu elements may cause duplications, deletions, and rearrangements of genes and their regulatory regions. C: Intronic L1 insertions (usually in the antisense orientation) may downregulate expression of their host genes by causing transcriptional elongation defects, premature polyadenylation, transcriptional interference, or siRNA-mediated transcript degradation. D: Functional splice sites contained within L1 and Alu elements, or present in genomic sequences cotranscribed and integrated by L1 (3′ transduction) may cause alternative splicing, often followed by host gene transcript degradation.**

progress in L1 biology will allow for better understanding of L1-mediated gene disruption, interspecies variation, and genome evolution.

ACKNOWLEDGMENTS

We thank the members of Kazazian laboratory for critical scientific discussions, and four anonymous referees for their insightful comments and helpful suggestions that have improved this manuscript. This work was supported by a grant from the National Institutes of Health (to H.H.K.).

REFERENCES

Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. 2006. Unconventional translation of mammalian LINE-1 retrotransposons. Genes Dev 20:210–224.

An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD. 2006. Active retrotransposition by a synthetic L1 element in mice. Proc Natl Acad Sci USA 103:18662–18667.

Asch HL, Eliacin E, Fanning TG, Connolly JL, Bratthauer G, Asch BB. 1996. Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. Oncol Res 8:239–247.

Ast G. 2004. How did alternative splicing evolve? Nat Rev Genet 5: 773–782.

Athanikar JN, Badge RM, Moran JV. 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res 32:3846–3855.

Babcock M, Pavlicek A, Spiteri E, Kashork CD, Ioshikhes I, Shaffer LG, Jurka J, Morrow BE. 2003. Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution. Genome Res 13:2519–2532.

Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr. 2006. L1 integration in a transgenic mouse model. Genome Res 16: 240–250.

Bailey JA, Liu G, Eichler EE. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73:823–834.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379.

Becker KG, Swergold G, Ozato K, Thayer RE. 1993. Binding of the ubiquitous nuclear transcription factor YY1 to a *cis* regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet 2: 1697–1702.

Belancio VP, Hedges DJ, Deininger P. 2006. LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic Acids Res 34: 1512–1521.

Belancio VP, Whelton M, Deininger P. Requirements for polyadenylation at the 3′ end of LINE-1 elements. Gene (in press).

Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A. 2006. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. Cancer Cell Int 6:13.

Berry C, Hannenhalli S, Leipzig J, Bushman FD. 2006. Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2:e157.

Bibillo A, Eickbush TH. 2002. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J Mol Biol 316:459–473.

Bibillo A, Eickbush TH. 2004. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. J Biol Chem 279: 14945–14953.

Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR. 2006. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci USA 103:8780–8785.

Boissinot S, Chevret P, Furano AV. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol 17: 915–928.

Branciforte D, Martin SL. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol 14:2584–2592.

Bratthauer GL, Fanning TG. 1992. Active LINE-1 retrotransposons in human testicular cancer. Oncogene 7:507–510.

Bratthauer GL, Fanning TG. 1993. LINE-1 retrotransposon expression in pediatric germ cell tumors. Cancer 71:2383–2386.

Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, Roos D, Kazazian HH Jr. 2002. Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet 71:327–336.

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA 100:5280–5285.

Burke WD, Malik HS, Jones JP, Eickbush TH. 1999. The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol Biol Evol 16:502–511.

Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E. 2002. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of l1. Genomics 80: 402–406.

Buzdin A, Gogvadze E, Kovalskaya E, Volchkov P, Ustyugova S, Illarionova A, Fushan A, Vinogradova T, Sverdlov E. 2003. The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. Nucleic Acids Res 31:4385–4390.

Casaregola S, Neuveglise C, Bon E, Gaillardin C. 2002. Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica*. Mol Biol Evol 19:664–677.

Cavalier-Smith T. 1985. The evolution of genome size. New York: Wiley. 534p.

Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN. 2005a. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. Hum Mutat 25:207–221.

Chen JM, Stenson PD, Cooper DN, Ferec C. 2005b. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet 117:411–427.

Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD. 2006a. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol 16:480–485.

Chen JM, Ferec C, Cooper DN. 2006b. LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: mutation detection bias and multiple mechanisms of target gene disruption. J Biomed Biotechnol 2006:56182.

Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, Heidmann T, Greene WC. 2006. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc Natl Acad Sci USA 103: 15588–15593.

Christensen SM, Eickbush TH. 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.

Conley ME, Partain JD, Norland SM, Shurtleff SA, Kazazian HH Jr. 2005. Two independent retrotransposon insertions at the same site within the coding region of BTK. Hum Mutat 25:324–325.

Cordaux R, Hedges DJ, Herke SW, Batzer MA. 2006. Estimating the retrotransposition rate of human Alu elements. Gene 373:134–137.

Cost GJ, Feng Q, Jacquier A, Boeke JD. 2002. Human L1 element target-primed reverse transcription in vitro. EMBO J 21:5899–5910.

Dante R, Dante-Paire J, Rigal D, Roizes G. 1992. Methylation patterns of long interspersed repeated DNA and alphoid repetitive DNA from human cell lines and tumors. Anticancer Res 12:559–563.

Deininger PL, Batzer MA. 1999. Alu repeats and human disease. Mol Genet Metab 67:183–193.

Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35:41–48.

Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. 1991. Isolation of an active human transposable element. Science 254: 1805–1808.

Dombroski BA, Feng Q, Mathias SL, Sassaman DM, Scott AF, Kazazian HH Jr, Boeke JD. 1994. An in vivo assay for the reverse transcriptase of

human retrotransposon L1 in *Saccharomyces cerevisiae*. Mol Cell Biol 14:4485–4492.

Duvernell DD, Turner BJ. 1998. Swimmer 1, a new low-copy-number LINE family in teleost genomes with sequence similarity to mammalian L1. Mol Biol Evol 15:1791–1793.

Eichler EE. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 17:661–669.

Eickbush DG, Luan DD, Eickbush TH. 2000. Integration of *Bombyx mori* R2 sequences into the 28S ribosomal RNA genes of *Drosophila melanogaster*. Mol Cell Biol 20:213–223.

Eickbush TH, Malik HS. 2002. Origins and evolution of retrotransposons. In: Craig NL, editor. Mobile DNA II. Washington, DC: ASM Press. p 1111–1144.

Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Stratling WH, Schumann GG. 2004. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J Biol Chem 279:27753–27763.

Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24:363–367.

Fanning T, Singer M. 1987. The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic Acids Res 15:2251–2260.

Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916.

Florl AR, Lower R, Schmitz-Drager BJ, Schulz WA. 1999. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. Br J Cancer 80:1312–1321.

Gasior SL, Wakeman TP, Xu B, Deininger PL. 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357:1383–1393.

Geiduschek EP, Kassavetis GA. 2001. The RNA polymerase III transcription apparatus. J Mol Biol 310:1–26.

George JA, Burke WD, Eickbush TH. 1996. Analysis of the 5′ junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. Genetics 142:853–863.

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521.

Gilbert N, Lutz-Prigge S, Moran JV. 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell 110:315–325.

Gilbert N, Lutz S, Morrish TA, Moran JV. 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol 25:7780–7795.

Goodier JL, Ostertag EM, Kazazian HH Jr. 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 9:653–657.

Goodier JL, Ostertag EM, Engleka KA, Seleme MC, Kazazian HH Jr. 2004. A potential role for the nucleolus in L1 retrotransposition. Hum Mol Genet 13:1041–1048.

Goodwin TJ, Ormandy JE, Poulter RT. 2001. L1-like non-LTR retrotransposons in the yeast *Candida albicans*. Curr Genet 39:83–91.

Han JS, Szak ST, Boeke JD. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature 429:268–274.

Hayward BE, Zavanelli M, Furano AV. 1997. Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. Genetics 146:641–654.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobbson L, Kerje S, Andersson L, Crooijmans RP, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MD, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WR, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK, Wang J, Liu B, Yu J, Yang H, Nefedov M, Koriabine M, Dejong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yang S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte F, Taylor J, He J, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SM, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Betran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyras E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716.

Hohjoh H, Singer MF. 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J 15:630–639.

Hohjoh H, Singer MF. 1997. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J 16:6034–6043.

Holmes SE, Singer MF, Swergold GD. 1992. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J Biol Chem 267:19765–19768.

Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr. 1994. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat Genet 7:143–148.

Hulme AE, Kulpa DA, Garcia-Perez JL, Moran JV. 2006. The impact of LINE-1 retrotransposition on the human genome. In: Lupski JR, Stankiewicz PT, editors. Genomic disorders: the genomic basis of disease. Totowa, NJ: Humana Press. p 35–55.

Hulme AE, Bogerd HP, Cullen BR, Moran JV. Selective inhibition of Alu retrotransposition by APOBEC3G. Gene (in press).

Ilves H, Kahre O, Speek M. 1992. Translation of the rat LINE bicistronic RNAs in vitro involves ribosomal reinitiation instead of frameshifting. Mol Cell Biol 12:4242–4248.

Ishida Y, Leder P. 1999. RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. Nucleic Acids Res 27:e35.

Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 94:1872–1877.

Jurka J. 2004. Evolutionary impact of human Alu repetitive elements. Curr Opin Genet Dev 14:603–608.

Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332:164–166.

Kazazian HH Jr. 1999. An estimated frequency of endogenous insertional mutations in humans. Nat Genet 22:130.

Kazazian HH. 2004. Mobile elements: drivers of genome evolution. Science 303:1626–1632.

Kojima KK, Matsumoto T, Fujiwara H. 2005. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. Mol Cell Biol 25:7675–7686.

Kolosha VO, Martin SL. 1997. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. Proc Natl Acad Sci USA 94: 10155–10160.

Kolosha VO, Martin SL. 2003. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). J Biol Chem 278:8112–8117.

Konkel MK, Wang J, Liang P, Batzer MA. Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. Gene (in press).

Kozak M. 1989. The scanning model for translation: an update. J Cell Biol 108:229–241.

Kubo S, Seleme MdC, Soifer HS, Perez JLG, Moran JV, Kazazian HH Jr, Kasahara N. 2006. L1 retrotransposition in nondividing and primary human somatic cells. Proc Natl Acad Sci USA 103:8036–8041.

Kulpa DA, Moran JV. 2005. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet 14:3237–3248.

Kulpa DA, Moran JV. 2006. *Cis*-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol 13: 655–660.

Kurose K, Hata K, Hattori M, 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.

Lander TES, Linton LM, Birren B, Nusbaum C, Zody MC, 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.

Lavie L, Maldener E, Brouha B, Meese EU, 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.

Leibold DM, Swergold GD, Singer MF, Thayer RE, Dombroski BA, Fanning TG. 1990. Translation of LINE-1 DNA elements in vitro and in human cells. Proc Natl Acad Sci USA 87:6990–6994.

Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. Science 300:1288–1291.

Li PW, Li J, Timmerman SL, Krushel LA, Martin SL. 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 CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshatsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolen O, Elong R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M, Foley K, Foley C, Franke A, Friedrich D, Gage D, Garber M, Gearin G, Giannoukos G, Goode T, Goyette A, Graham J, Grandbois E, Gyaltsen K, Hafez N, Hagopian D, Hagos B, Hall J, Healy C, Hegarty R, Honan T, Horn A, Houde N, Hughes L, Hunnicutt L, Husby M, Jester B, Jones C, Kamat A, Kanga B, Kells C, Khazanovich D, Kieu AC, Kisner P, Kumar M, Lance K, Landers T, Lara M, Lee W, Leger JP, Lennon N, Leuper L, LeVine S, Liu J, Liu X, Lokyitsang Y, Lokyitsang T, Lui A, Macdonald J, Major J, Marabella R, Maru K, Matthews C, McDonough S, Mehta T, Meldrim J, Melnikov A, Meneus L, Mihalev A, Mihova T, Miller K, Mittelman R, Mlenga V, Mulrain L, Munson G, Navidi A, Naylor J, Nguyen T, Nguyen N, Nguyen C, Nicol R, Norbu N, Norbu C, Novod N, Nyima T, Olandt P, O'Neill B, O'Neill K, Osman S, Oyono L, Patti C, Perrin D, Phunkhang P, Pierre F, Priest M, Rachupka A, Raghuraman S, Rameau R, Ray V, Raymond C, Rege F, Rise C, Rogers J, Rogov P, Sahalie J, Settipalli S, Sharpe T, Shea T, Sheehan M, Sherpa N, Shi J, Shih D, Sloan J, Smith C, Sparrow T, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Stone S, Sykes S, Tchuinga P, Tenzing P, Tesfaye S, Thoulutsang D, Thoulutsang Y, Topham K, Topping I, Tsamla T, Vassiliev H, Venkataraman V, Vo A, Wangchuk T, Wangdi T, Weiand M, Wilkinson J, Wilson A, Yadav S, Yang S, Yang X, Young G, Yu Q, Zainoun J, Zembek L, Zimmer A, Lander ES. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803–819.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72: 595–605.

Lucchinetti E, Feng J, Silva Rd, Tolstonog GV, Schaub MC, Schumann GG, 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.

Makalowski W, Mitchell GA, Labuda D. 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet 10: 188–193.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805.

Martin SL. 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol 11:4804–4807.

Martin F, Maranon C, Olivares M, Alonso C, Lopez MC. 1995. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. J Mol Biol 247:49–59.

Martin SL, Li J, Weisz JA. 2000. Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. J Mol Biol 304:11–20.

Martin SL, Bushman FD. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol 21:467–475.

Martin SL, Cruceanu M, Branciforte D, Wai-Lun Li P, Kwok SC, Hodges RS, Williams MC. 2005a. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J Mol Biol 348: 549–561.

Martin SL, Li WL, Furano AV, Boissinot S. 2005b. The structures of mouse and human L1 elements reflect their insertion mechanism. Cytogenet Genome Res 110:223–228.

Martin SL. 2006. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. J Biomed Biotechnol 2006:45621.

Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810.

Matlik K, Redik K, Speek M. 2006. L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol 2006:71753.

McMillan JP, Singer MF. 1993. Translation of the human LINE-1 element, L1Hs. Proc Natl Acad Sci USA 90:11533–11537.

Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 52:643–645.

Minakami R, Kurose K, Etoh K, Furuhata Y, Hattori M, Sakaki Y. 1992. Identification of an internal *cis*-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. Nucleic Acids Res 20:3139–3145.

Mine M, Chen JM, Brivet M, Desguerre I, Marchant D, de Lonlay P, Bernard A, Ferec C, Abitbol M, Ricquier D, Marsac C. 2006. 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.

Mitchell GA, Labuda D, Fontaine G, Saudubray JM, Bonnefont JP, Lyonnet S, Brody LC, Steel G, Obie C, Valle D. 1991. Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation. Proc Natl Acad Sci USA 88:815–819.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH. 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87:917–927.

Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. Science 283:1530–1534.

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 31:159–165.

Morse B, Rotherg PG, South VJ, Spandorfer JM, Astrin SM. 1988. Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. Nature 333:87–90.

Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C. 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem 281:22161–22172.

Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435:903–910.

Musova Z, Hedvicakova P, Mohrmann M, Tesarova M, Krepelova A, Zeman J, Sedlacek Z. 2006. A novel insertion of a rearranged L1 element in exon 44 of the dystrophin gene: further evidence for possible bias in retroposon integration. Biochem Biophys Res Commun 347:145–149.

Nigumann P, Redik K, Matlik K, Speek M. 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79:628–634.

Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. 2003. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol 4:R74.

Ostertag EM, Kazazian HH Jr. 2001a. Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538.

Ostertag EM, Kazazian HH Jr. 2001b. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res 11:2059–2065.

Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HHJ. 2002. A mouse model of human L1 retrotransposition. Nat Genet. 32:655–660.

Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet 73:1444–1451.

Ostertag EM, Kazazian HH. J. 2006. Retrotransposition and Human Disorders. In: Encyclopedia of Life Sciences: John Wiley & Sons. Ltd: Chichester http://www.els.net/[doi: 10.1038/npg.els.0005492].

Perepelitsa-Belancio V, Deininger P. 2003. RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet 35:363–366.

Pickeral OK, Makalowski W, Boguski MS, Boeke JD. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 10:411–415.

Piskareva O, Denmukhametova S, Schmatchenko V. 2003. Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells. Protein Expr Purif 28:125–130.

Piskareva O, Schmatchenko V. 2006. DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. FEBS Lett 580:661–668.

Prak ET, Dodson AW, Farkash EA, Kazazian HH Jr. 2003. Tracking an embryonic L1 retrotransposition event. Proc Natl Acad Sci USA 100:1832–1837.

Reznikoff WS. 2003. Tn5 as a model for understanding DNA transposition. Mol Microbiol 47:1199–1206.

Roman-Gomez J, Jimenez-Velasco A, Agirre X, Cervantes F, Sanchez J, Garate L, Barrios M, Castillejo JA, Navarro G, Colomer D, Prosper F, Heiniger A, Torres A. 2005. Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia. Oncogene 24:7213–7223.

Santourlidis S, Florl A, Ackermann R, Wirtz HC, Schulz WA. 1999. High frequency of alterations in DNA methylation in adenocarcinoma of the prostate. Prostate 39:166–174.

Schulz WA, Steinhoff C, Florl AR. 2006. Methylation of endogenous human retroelements in health and disease. Curr Top Microbiol Immunol 310:211–250.

Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, Margolet L. 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1:113–125.

Seleme MC, Vetter MR, Cordaux R, Bastone L, Batzer MA, Kazazian HH Jr. 2006. Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. Proc Natl Acad Sci USA 103:6611–6616.

Shi X, Seluanov A, Gorbunova V. 2007. Cell divisions are required for L1 retrotransposition. Mol Cell Biol 27:1264–1270.

Singer MF, Krek V, McMillan JP, Swergold GD, Thayer RE. 1993. LINE-1: a human transposable element. Gene 135:183–188.

Skowronski J, Fanning TG, Singer MF. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. Mol Cell Biol 8:1385–1397.

Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9:657–663.

Soifer HS, Rossi JJ. 2006. Small interfering RNAs to the rescue: blocking L1 retrotransposition. Nat Struct Mol Biol 13:758–759.

Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 21:1973–1985.

Stankiewicz P, Shaw CJ, Withers M, Inoue K, Lupski JR. 2004. Serial segmental duplications during primate evolution result in complex human genome architecture. Genome Res 14:2209–2220.

Stenglein MD, Harris RS. 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem 281:16837–16841.

Swergold GD. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 10:6718–6729.

Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. 2002. Human l1 retrotransposition is associated with genetic instability in vivo. Cell 110:327–338.

Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. 2002. Molecular archeology of L1 insertions in the human genome. Genome Biol 3:research0052.

Tchenio T, Casella J-F, Heidmann T. 2000. Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res 28:411–415.

Trelogan SA, Martin SL. 1995. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. Proc Natl Acad Sci USA 92:1520–1524.

Turelli P, Vianin S, Trono D. 2004. The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. J Biol Chem 279:43371–43373.

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. 2005. SVA elements: a hominid-specific retroposon family. J Mol Biol 354:994–1007.

Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev 20:1732–1743.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. 2001. Human L1 retrotransposition: *cis* preference versus *trans* complementation. Mol Cell Biol 21:1429–1439.

Weissenbach J. 1993. Microsatellite polymorphisms and the genetic linkage map of the human genome. Curr Opin Genet Dev 3:414–417.

Wheelan SJ, Aizawa Y, Han JS, Boeke JD. 2005. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res 15:1073–1078.

Yang N, Zhang L, Zhang Y, Kazazian HH Jr. 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res 31:4929–4940.

Yang N, Kazazian HH Jr. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol 13:763–771.

Yu F, Zingler N, Schumann G, Stratling WH. 2001. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res 29:4493–4501.

Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298.

Zhang XH, Chasin LA. 2006. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proc Natl Acad Sci USA 103:13427–13432.

Zingler N, Willhoeft U, Brose HP, Schoder V, Jahns T, Hanschmann KM, Morrish TA, Lower J, Schumann GG. 2005. Analysis of 5′ junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5′-end attachment requiring microhomology-mediated end-joining. Genome Res 15:780–789.