Annu. Rev. Genet. 2001. 35:501–38


# BIOLOGY OF MAMMALIAN L1 RETROTRANSPOSONS

Eric M. Ostertag and Haig H. Kazazian Jr
*Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104; e-mail: ostertag@mail.med.upenn.edu, kazazian@mail.med.upenn.edu*

**Key Words** transposable elements, reverse transcription, genome structure, human mutation, mouse mutation

■ **Abstract** L1 retrotransposons comprise 17% of the human genome. Although most L1s are inactive, some elements remain capable of retrotransposition. L1 elements have a long evolutionary history dating to the beginnings of eukaryotic existence. Although many aspects of their retrotransposition mechanism remain poorly understood, they likely integrate into genomic DNA by a process called target primed reverse transcription. L1s have shaped mammalian genomes through a number of mechanisms. First, they have greatly expanded the genome both by their own retrotransposition and by providing the machinery necessary for the retrotransposition of other mobile elements, such as Alus. Second, they have shuffled non-L1 sequence throughout the genome by a process termed transduction. Third, they have affected gene expression by a number of mechanisms. For instance, they occasionally insert into genes and cause disease both in humans and in mice. L1 elements have proven useful as phylogenetic markers and may find other practical applications in gene discovery following insertional mutagenesis in mice and in the delivery of therapeutic genes.

## CONTENTS



0066-4197/01/1215-0501$14.00 

## INTRODUCTION

Preliminary analysis of the human genome sequence has already provided several major surprises. For one, the human genome contains less than twice as many genes as the fly, worm, and Arabidopsis genomes. Equally surprising is the observation that 45% of the human genome consists of transposable elements, a much greater percentage than the 3% to 10% observed in the genomes of the three other organisms (107). Since sequences that have been in the genome longer than 200 My have diverged to the point where they are unidentifiable, it is likely that even more than 45% of the human genome is composed of transposable elements (107, 191). Transposable elements have contributed greatly to what we now realize is a highly dynamic genome.

Although this review concentrates on a particular type of transposable element, the L1 retrotransposon, a general introduction to the topic is warranted. Mammalian transposable elements consist of DNA transposons and retrotransposons (Figure 1). DNA transposons have structures similar to bacterial transposons. They have inverted terminal repeats and encode a transposase activity. They generally move by a "cut and paste" mechanism utilizing the transposase (143, 192). Although roughly 3% of the human genome is composed of DNA transposons, they are remnants or fossils of ancient elements, and it is unlikely that any remain transpositionally active (107). In contrast to DNA transposons, retrotransposons encode a reverse transcriptase (RT) activity and move by a "copy and paste" process involving an RNA intermediate. The original retrotransposon is maintained *in situ* where it is transcribed. The transcript is then reverse transcribed and integrated into a new genomic location (115, 217). Approximately 42% of the human genome is composed of retrotransposons (107), and, although most of these elements are inactive, some retain the ability to retrotranspose (172).

Retrotransposable elements can be classified as either autonomous or nonautonomous. Elements are considered autonomous if they encode certain activities necessary for their mobility. However, it is unlikely that they are strictly autonomous because host proteins, such as DNA repair enzymes, are probably also required for their retrotransposition. There are two classes of autonomous retrotransposons: LTR (long terminal repeat) retrotransposons and non-LTR retrotransposons. Mammalian LTR retrotransposons are structurally similar to retroviruses, but they lack a functional *env* gene. These retrotransposons include elements

such as mouse intracisternal A-particles (IAPs) (103) and human endogenous retroviruses (HERVs) (13, 191, 219), both of which are unlikely to include autonomously active elements. About 8% of the human genome is composed of defective endogenous retroviruses and solitary LTRs derived from recombination between the 5′ LTR and the 3′ LTR of these elements (107). The non-LTR retrotransposon class contains LINEs (long interspersed nucleotide elements), which include inactive elements, such as L2 in humans, and active elements, such as L1 in humans and mice (107, 120, 191). Approximately 21% of the human genome is composed of autonomous non-LTR retrotransposons (107).

In addition to the autonomous retrotransposons, there are a large number of nonautonomous retrotransposons in mammalian genomes. These elements do not encode any proteins. Therefore, they require activities encoded by other autonomous retrotransposons for their mobility. The most prominent members of this class are Alu elements in humans and their B1 counterparts in mice (169). The greater than 1 million Alu elements in the human genome account for about 11% of its mass, whereas roughly 100,000 B1 elements populate the mouse genome (107, 191). Other nonautonomous retrotransposons in the human genome include processed pseudogenes and SVA elements. Thus, transposable elements and transposon-derived sequences make up about 45% of the total mass of the human genome, or 40 times the 1.1% of the genome that is composed of protein-coding sequences (107).

L1 elements are the master retrotransposons in mammalian genomes. Besides duplicating themselves, they likely have been responsible for the genomic expansion of nonautonomous retrotransposons, specifically Alu elements, processed pseudogenes, and SVA elements in the human genome. Over evolutionary time they have not only expanded greatly in number, but also have acquired other roles, some of which are quite useful to the organism, whereas others are detrimental to individual members of the species (90). Many of these roles of L1 retrotransposons are discussed in this review.

## MECHANISM OF L1 RETROTRANSPOSITION

In mammals, the great majority of L1 retrotransposons are inactive, defective elements, owing to 5′ truncation, inversion, or point mutations. Of the 520,000 L1 sequences in humans, only about 3000–5000 represent full-length elements (66, 107). The discovery of a full-length mouse L1 element with intact open reading frames (112), and the creation of a consensus alignment of many human L1 sequences (177), helped to elucidate the anatomy of the full-length, 6-kb element. The consensus sequence revealed that L1 elements have a 5′ untranslated region (UTR) with internal promoter activity, two open reading frames (ORFs), a 3′ UTR that ends in an AATAAA polyadenylation signal, and a polyA tail (Figure 1). The discovery that several full-length, retrotranspositionally active elements had the predicted ORFs validated the consensus sequence (147).

In some cases, careful scrutiny of L1 structure has offered insight into the mechanism of L1 retrotransposition. In other cases, an understanding of similar retrotransposons in other organisms has suggested hypotheses for the L1 mechanism. The development of limited functional assays has strengthened some of these hypotheses (41, 52, 127a). Furthermore, the development of a cell culture–based retrotransposition assay was instrumental in demonstrating L1 functions necessary for retrotransposition (147) (Figure 2). However, many aspects of the retrotransposition mechanism remain unknown. The general steps of retrotransposition include transcription, RNA processing, mRNA export, translation, posttranscriptional modifications and RNP formation, return to the nucleus, and reverse transcription and integration (Figure 3). Here we summarize what is known and what is theorized regarding the mechanism of L1 retrotransposition.

## Transcription

The 5′ UTR of human L1 contains internal promoter activity independent of upstream sequences (199), but the machinery responsible for transcribing L1 in vivo remains undetermined. A reporter gene driven by an L1 5′ UTR fragment apparently can be transcribed by RNA polymerase III (Pol III) (105). However, inconsistent with Pol III–mediated transcription, the L1 transcript is much larger than a typical Pol III transcript and encodes proteins. Moreover, the presence of internal Pol III termination sequences in L1 argues against Pol III–mediated transcription. With few exceptions, Pol III in higher eukaryotes terminates within clusters of four or more consecutive T residues in the noncoding DNA strand (16, 58a, 101). Pol III transcripts rarely contain four consecutive internal U residues and almost never contain five consecutive U residues. However, the noncoding strand of the consensus sequence of active human L1 elements contains a stretch of six T residues and a stretch of seven T residues. Furthermore, the active L1 consensus element contains a functional AATAAA polyadenylation (polyA) signal (146). The AATAAA polyA signal is required for RNA polymerase II (Pol II) termination (160), and proper cleavage and polyadenylation after the AATAAA polyA signal requires Pol II (77, 128), suggesting that Pol II transcribes L1. Lastly, the 5′ UTR of both mouse and human L1 elements can be replaced functionally with a heterologous Pol II promoter in the cultured cell assay for retrotransposition (147, 151).

Pol II and its associated transcription factors are recruited to a core promoter element, typically the TATA motif located 25–30 base pairs upstream of the transcription initiation site. Early steps in transcription include binding of TATA Binding Protein (TBP), a component of TFIID, to the TATA motif and subsequent binding of TFIIB to TFIID. Other transcription factors and Pol II bind this complex to form a complete transcription complex (232). However, some Pol II–dependent promoters lack TATA sequences, instead containing initiator (Inr) elements capable of independently directing transcription initiation (215). Pol II transcription from an internal promoter was first demonstrated for *jockey*, a retrotransposon

from *Drosophila melanogaster* (142). It has since been shown that *jockey*, as well as the related Drosophila retrotransposons, I factor, F, and Doc, contain conserved downstream promoter elements (DPEs) that direct transcription initiation at an Inr-like sequence by directly recruiting TFIID (24, 141). A similar arrangement may allow Pol II transcription from an internal promoter in human L1 elements.

Several groups have demonstrated protein binding to a downstream core promoter element essential for human L1 transcription (127, 140). This protein has been identified as the Yin Yang-1 (YY-1) transcription factor (12, 105), and the conserved protein binding sequence, base pairs +13 to +21 of the L1 5′ UTR coding strand, is a perfect match to the YY-1 core binding sequence (83, 188). YY-1 can act as a transcriptional activator, repressor, or initiator (179, 186). Interestingly, YY-1 can initiate transcription in vitro in the absence of TFIID (179, 206), binding TFIIB directly and directing Pol II to the transcription initiation site (207). The human L1 YY-1 binding site may be analogous to the Drosophila DPE, and YY-1 binding may either recruit TFIID or bind TFIIB directly during the formation of a Pol II preinitiation complex. The TATA-less human DNA polymerase $\beta$ gene promoter has an Inr with an overlapping YY-1 binding site. Careful mutational analysis of this promoter demonstrates that YY-1 binding is not required for transcription, but rather may serve to position the transcription complex or regulate promoter activity (216). Such a mechanism cannot be ruled out for L1 transcription, but limited mutational analysis of the L1 promoter suggests that YY-1 binding is very important for efficient transcription (12, 140). Additional promoter studies including further mutational analysis should help to elucidate the mechanism of L1 transcription.

Unlike humans, rats and mice have several distinct L1 subfamilies that are defined by differences in their 5′ structures. Most murine L1 subfamilies have 5′ UTRs notably different from the human 5′ UTR in that they contain a variable number of tandemly repeated units of 205 to 210 bp, called monomers, followed by a short non-monomeric region (51, 55, 112, 163). The V subfamily has no identifiable monomers, and, because its members diverge significantly in sequence from each other, it is the oldest subfamily (4, 87). The F subfamily has a large number of inactive members, and it is the oldest L1 subfamily with monomers evident at its 5′ end (163, 174, 220). A consensus F monomer has been resurrected and shown to have promoter activity (4). The A subfamily (112, 182) contains about 50,000 truncated and 6000 to 8000 full-length members per diploid mouse genome (173). These elements have monomers (112) and share about 95%–97% sequence similarity (174). The recently discovered $T_F$ subfamily (34, 151) contains about 3000 full-length members per diploid genome (151, 173). These elements were called $T_F$ because most are transposable and their monomers are 70% identical in sequence to the F monomers (151). $T_F$ L1s share greater than 99% sequence similarity (most are greater than 99.6% identical to each other) (34). The most recently discovered L1 subfamily is the $G_F$ subfamily. These L1s are 93–99+% identical to each other and contain monomers that, like $T_F$, are about 70% identical to F monomers. Although both $G_F$ and $T_F$ monomers differ from F monomers by about

30%, they also differ from each other by 33%. There are roughly 1500 full-length $G_F$ elements in the diploid mouse genome (59).

Experiments using cultured cells transiently transfected with various regions of the mouse $T_F$ 5′ UTR fused to a reporter gene have revealed that the promoter activity lies within the monomers and that promoter strength is proportional to the number of monomers. L1s of the $T_F$ subfamily contain a conserved YY-1 binding site within each monomer. In the genome, many of these elements begin within or near the YY-1 site, supporting a possible role for YY-1 in positioning of the transcription complex (35). However, $G_F$ L1s contain a YY-1 binding site that differs from consensus by one nucleotide and, in the genome, these elements tend not to begin near this site (59). L1s of the A subfamily do not contain a YY-1 binding site, suggesting that the mechanism of transcription may vary among L1 elements. The mechanism by which monomers are created and maintained is also an interesting, albeit unsolved, puzzle.

Evidence to date suggests that L1 expression is germ line specific. Full-length, sense-stranded L1 transcripts have been detected in prepuberal spermatocytes, but are rare in normal somatic tissues (18). Recently, a mouse model of retrotransposition was created using a tagged human L1 element under the control of its endogenous promoter. Strand-specific RT-PCR designed to detect the full-length tagged transcript demonstrated expression in the male and female germ line, but not in multiple somatic tissues. Studies in the transgenic mouse model indicate that retrotransposition of the tagged human L1 element under the control of its endogenous promoter occurs in late-meiotic and post-meiotic male germ cells (161). The transcription factors that determine the germ line specificity of L1 transcription have not been defined. However, indirect evidence suggests that the SOX family of transcription factors may be involved (201). Many groups have also suggested that transcription of L1 elements may be controlled by methylation at CpG dinucleotides in the L1 5′ untranslated region (72, 156, 202, 221, 230).

If L1 were a genetic parasite, as hypothesized, then one would expect a germ line–specific pattern of L1 expression. L1 retrotransposition in the germ line is likely to lead to an expansion in the number of L1 elements in the genome, whereas L1 retrotransposition events in somatic tissues cannot be passed on to future generations and are likely to be detrimental to the host. For example, unchecked retrotransposition in somatic tissues could result in an insertion into a tumor suppressor gene that could ultimately promote oncogenesis. Interestingly, in a cultured cell assay, retrotransposition is detected in a large variety of transformed cells but not in primary cell lines or ES cells (J.L. Goodier, E.T. Prak & H.H. Kazazian Jr, unpublished data). The initial mutations that cause a cell to become precancerous might occasionally activate L1 transcription. The resultant somatic retrotransposition events could increase the likelihood of accumulating additional mutations that would ultimately produce cancer. In fact, there is at least one example of an authentic retrotransposition event contributing to cancer; an L1 element inserted into the APC gene of tumor cells in a patient who had developed colon cancer (139).

## RNA Processing and Nuclear Export

RNAs transcribed by Pol II are typically modified by cleavage and addition of a polyA tail, by addition of a 7-methylguanosine cap, and by splicing of introns (204). As mentioned previously, L1 elements contain a functional AATAAA polyadenylation signal and likely use the cleavage and polyadenylation machinery typical of Pol II transcripts. However, there are two unusual features of the L1 polyadenylation signal. First, the AATAAA polyadenylation signal of L1 elements is immediately followed by the presumed polyA tail. This observation can be interpreted in one of two ways: Either cleavage and polyadenylation of L1 elements occur immediately after the AATAAA signal, as opposed to the usual 10–25 nucleotides downstream (26), or they occur at the typical number of bases downstream of the AATAAA signal and the A residues found between the polyA signal and the polyA tail are encoded. If the latter is true, it begs the question why L1 elements have evolved to contain a stretch of A residues after the polyA signal. One possible answer is that the A-rich region may be positively selected by the L1 integration process (see discussion of TPRT below). Analysis of mouse L1 elements in the genome database suggests that the A-rich region is encoded because many elements have the sequence AATGG A(n) following the polyA signal (J.L. Goodier, personal communication). The second unusual feature of the L1 polyadenylation process is that L1s are usually lacking important sequences downstream of the polyadenylation site. These conserved, GU-rich sequences are 20–60 nucleotides 3′ of the polyadenylation signal and promote efficient cleavage and polyadenylation (129, 130). The sequence 20–60 nucleotides downstream of retrotransposed L1 elements depends upon the insertion site and is highly variable (often this sequence is the polyA tail). One would predict that subsequent retrotransposition events would therefore polyadenylate inefficiently after the AATAAA signal. In fact, this appears to be the case. L1 elements frequently bypass their own polyA signal and use a downstream signal (146). This process results in the retrotransposition of genomic sequence 3′ of the L1 element and is called L1-mediated transduction (see section on L1-mediated transduction below).

Additional modifications of L1 RNA are not known. For example, it is not known whether L1 transcripts are modified by the addition of a 7-methylguanosine cap. L1 transcripts do not contain introns and therefore do not require splicing. Although nearly all mammalian mRNAs contain introns, there are notable exceptions, such as members of the human G-protein–coupled receptor genes and type I interferon genes (57, 167). Interestingly, in a cultured cell assay, tagged L1 transcripts containing an intron are spliced appropriately and are able to retrotranspose (147).

Recent experiments are beginning to elucidate the mechanism of mRNA export from the nucleus in vertebrates. Cells have evolved a mechanism to prevent export of unspliced mRNAs, presumably because the export of unspliced RNAs would be inefficient and could result in protein products that are deleterious to

the cell. Unspliced RNAs contain splice sites that are bound by splicing factors called commitment factors and retained in the nucleus (33). Splicing and nuclear export are therefore coupled. It has been suggested that L1 mRNA might contain *cis*-acting elements required for its nuclear export (159). This is a reasonable speculation based upon the observations that some mRNAs expressed from transfected cDNAs lacking intron sequences are not exported efficiently, and that several viruses have evolved *cis*-acting elements to facilitate export of unspliced RNA. However, mRNAs expressed from some intronless cDNAs are exported well. In addition, unlike retroviruses (32), L1 does not face the problem of exporting an unspliced RNA containing splice sites, an RNA species that is normally retained in the nucleus by commitment factors. One would not expect L1 RNA to be retained because it normally does not contain splice sites, and therefore, may not need *cis*-acting elements for nuclear export.

## Translation

L1 mRNAs are atypical of mammalian mRNAs because they are bicistronic. In humans, the two ORFs are in frame and separated by a 63-bp noncoding spacer region that contains stop codons in all three reading frames. In mice, the two ORFs are also nonoverlapping but in different reading frames, whereas in rats, the ORFs are overlapping. The mechanism of translation remains largely a mystery. Experiments suggest that the ORF1 protein is translated by ribosomal initiation at the 5′ UTR followed by ribosomal scanning (131). Whether initiation is cap-dependent is unknown. Indirect evidence suggests that the human ORF2 is not translated by termination and reinitiation (131). Frameshifting is unnecessary in humans and has been ruled out in rats (84). In addition, there is no evidence of an ORF1/ORF2 fusion protein in either species. Therefore, translation of ORF2 may occur by some form of internal ribosomal entry such as by the use of an internal ribosomal entry site (IRES) or by ribosomal shunting.

ORF1 protein is apparently translated much more efficiently than ORF2 protein. ORF1 protein has been detected in the cytoplasm of a number of human testicular germ cell tumors and in breast carcinoma and medulloblastoma (9, 19–21). ORF1 has also been detected in mouse embryonal carcinoma cell lines, male and female mouse germ cells, Leydig cells of embryonic mouse testis, theca cells of adult mouse ovary, and a large variety of transformed mouse and human cell lines (18, 124, 205). However, ORF2 has escaped detection despite efforts using several antibodies that detect either Baculovirus-produced or bacterial-produced ORF2 protein.

ORF1 encodes an approximately 40-kDa protein (p40) with RNA binding activity (80, 98, 126). The exact size of the ORF1 protein varies among species and occasionally within species. For example, in the mouse, the ORF1 protein has a length polymorphism region (LPR) that causes a length difference of up to 28 amino acids (3, 59, 132). The amino acid sequence of the COOH-terminal half of the ORF1

protein is relatively well conserved across species, whereas the $NH_2$-terminal half of the protein varies considerably. However, the ability of the $NH_2$-terminal half of the protein to form an $\alpha$-helical structure is conserved. Human p40 contains a leucine zipper motif, rabbit ORF1 protein contains a coiled-coil domain, and rat and mouse proteins each have unique $\alpha$-helical structures (39, 79, 82). It was hypothesized that the conserved COOH-terminal end is involved in RNA binding and the conserved NH-terminal $\alpha$-helical structure is involved in protein-protein interaction (see section on RNP formation below). A series of elegant experiments using mouse ORF1 protein strongly support this hypothesis (126). In addition to these functions, mouse ORF1 protein has nucleic acid chaperone activity in vitro (125) (see section on TPRT below).

ORF2 encodes an approximately 150-kDa protein with three conserved domains, an $NH_2$-terminal endonuclease (EN) domain (52), a central reverse transcriptase (RT) domain (127a), and a COOH-terminal zinc knuckle-like domain (50). The L1 EN domain cleaves one strand of double-stranded DNA at a large number of genomic sites characterized by the loose consensus sequence, AA|TTTT (28, 52, 88). Cleavage site preference may be affected by the local chromatin structure (30). The EN domain is evolutionarily related in a subset of non-LTR retrotransposons and shares critical amino acids at positions corresponding to the catalytic sites of Exonuclease III, an endonuclease of *Escherichia coli* (52, 65, 120, 122, 144).

The L1 RT domain is related in all non-LTR retrotransposons (120). Non-LTR retrotransposon RT shares sequence similarities with more distantly related RTs from LTR retrotransposons and retroviruses, yet functions in a very different way (225). Retroviral and LTR retrotransposon RTs function in the cytoplasm within particles, use a tRNA primer, and carry out reverse transcription through a complex process requiring a number of steps (217). On the other hand, non-LTR retrotransposon RTs are thought to function in the nucleus, use genomic DNA as a primer, and carry out reverse transcription through the relatively simple process of target primed reverse transcription (TPRT) (115).

A third conserved domain of ORF2 is a COOH-terminal, cysteine-rich domain (50). This domain is conserved in all known mammalian L1 elements. Comparison with *Swimmer 1* and Zorro, related L1-like non-LTR retrotransposons from teleost fish and *Candida albicans*, respectively, suggests that it may be a conserved CCHC zinc knuckle structure (45, 61). CCHC zinc knuckles are present in all retroviral nucleocapsid proteins, except spumaviruses (14, 95, 198), and are found in other proteins that bind single-stranded RNA (8, 11). These observations suggest a possible role for ORF2 in protein-nucleic acid interaction, specifically the interaction of ORF2 protein with L1 RNA during the formation of retrotransposition intermediates. Interestingly, in addition to a role in nucleic acid binding, mutational analysis of various retroviral CCHC zinc knuckles suggests they are important for reverse transcription, perhaps required for unfolding structured RNA (62, 63, 68). Such a role cannot be ruled out for the L1 ORF2 protein.

## Posttranslational Modifications and Ribonucleoprotein Formation

Posttranslational modifications or protein processing of ORF1 and ORF2 proteins are currently unknown. It is believed that the ORF1 protein, ORF2 protein, and L1 RNA associate to form ribonucleoprotein (RNP) particles that are intermediates in retrotransposition. L1 RNA has been found associated with ORF1 protein in RNP particles in human and mouse teratocarcinoma cells (79, 123). However, ORF2 protein has not yet been detected in these particles. Both mouse and human ORF1 have been demonstrated to form higher-order homomultimers. The protein-protein interaction is likely mediated by the leucine zipper in humans and may be stabilized by interchain disulfide bonds (79). ORF1 proteins from mouse and other mammals lack the leucine zipper, but likely use their $\alpha$-helices for protein-protein interaction (126).

## Entry into the Nucleus

As a consequence of the mechanism of reverse transcription and integration (see section on TPRT below), ORF2 protein and L1 RNA must both gain access to genomic DNA. Proteins larger than approximately 60 kDa are too large to enter the nucleus by passive diffusion through the nuclear pore (64). The ORF2 protein alone is predicted to be about 150 kDa. Therefore, access to genomic DNA must either occur by energy-dependent, active transport through a nuclear pore, or by entry during nuclear membrane breakdown at mitosis or meiosis. Although several mechanisms are known for the active nuclear import of proteins, the classical pathway is mediated by proteins called importins (also called karyopherins), which bind to specific amino acid sequences called nuclear localization signals (64, 153). Experiments suggest that the ORF1 protein does not contain any functional nuclear localization signals (E.M. Ostertag & H.H. Kazazian, Jr, unpublished data) and, as mentioned previously, ORF1 protein has not been detected in the nucleus by immunostaining techniques. ORF2 protein may contain one or more nuclear localization signals, but evidence for their function is lacking because of the difficulty in detecting full-length ORF2 protein. If the ORF2 protein does encode a functional nuclear localization signal, it would be an interesting case of an RNA gaining access to the nucleus by encoding its own nuclear import protein. Another possibility is that the ORF1 or ORF2 proteins bind an additional protein that itself contains a nuclear localization signal or that the L1 RNA is required for nuclear import. Also, retrotransposition may depend upon nuclear breakdown. If retrotransposition takes place only in dividing cells, this would be a second factor, along with reduced transcription, in greatly reducing insertions in differentiated, rarely dividing tissues.

## Target Primed Reverse Transcription (TPRT)

L1 elements are likely reverse transcribed and integrated into the genome by a coupled reverse transcription/integration process called target primed reverse

transcription (TPRT). TPRT was originally demonstrated for the R2 element, a site-specific, non-LTR retrotransposon found in arthropods (115). R2 retrotransposons have a single ORF that encodes a protein with Type II restriction endonuclease (228) and reverse transcriptase activity. Elegant in vitro experiments using a bacterially produced R2 protein demonstrated that the endonuclease domain of the protein cleaves the noncoding strand of its target site, a sequence in the 28S rRNA gene. The reverse transcriptase domain of the R2 protein then uses the free 3′-OH at the DNA nick as a primer and the R2 RNA as a template for the reverse transcription reaction. Reverse transcription of the RNA is followed by cleavage of the coding strand and integration. TPRT produces a perfect duplication of the original target site, which flanks the newly inserted element (115) (Figure 4).

There are several reasons to believe that L1 elements use TPRT as their mechanism of reverse transcription and integration, although it has not yet been demonstrated definitively. First, recent in vitro experiments using Baculovirus-produced, full-length L1 ORF2 protein produce limited TPRT reactions (29). Second, L1 elements in the genome are often flanked by perfect 7- to 20-bp target site duplications, a typical consequence of the TPRT reaction. Lastly, the nucleotides at the predicted cleavage site are often T-rich, which are complementary to the polyA tail at the 3′ end of an L1 element, suggesting that they could indeed be used as a primer for reverse transcription of the L1 RNA.

The vast majority of L1 insertions in vivo are highly truncated at the 5′ end such that the average insertion length is only about 1 kb, or one sixth that of a full-length element (107). L1 truncation has long been explained by an inability of the L1 reverse transcriptase to copy the entire L1 RNA before disassociating from the RNA. Truncation may also be due to the action of a cellular RNAse H competing with L1 reverse transcriptase. Digestion of the RNA before the completion of reverse transcription followed by integration would result in an insertion truncated at the 5′ end.

Roughly 25% of recent L1 insertions also contain an inversion of a few hundred to fifteen hundred nucleotides of L1 sequence (161a). The inversion always involves the 5′ terminal end of the L1 element and is 5′ truncated itself. In other words, if L1 sequence is 5′-A-B-C-D-E-3′, then an inversion-containing insertion may be 5′-C-B-D-E-3′. The point of inversion may contain a deletion, a duplication, or neither. We suggest that inversion is a consequence of the L1 TPRT mechanism and describe a proposed model here.

If cleavage of the second DNA strand occurs before reverse transcription has been completed, an additional 3′ hydroxyl would be available for the priming of reverse transcription. This potential primer could invade the L1 RNA internally and prime reverse transcription at a site distinct from the reverse transcription occurring at the 3′ end of the L1 RNA. The L1 RNA template would therefore be primed by two different primers at two separate locations, a possibility that we call twin priming. Resolution of the RNA/cDNA structure that has undergone twin priming would produce a typical L1 inversion with a 5′ truncation (Figure 5). Depending on the extent of reverse transcription from the primer at the 3′ end of the L1 RNA, the point of inversion would contain a deletion, a duplication, or neither. This model predicts all of the typical L1 inversion structures

that are found in the genome database and does not predict structures that are not found (such as internal inversions).

If this model is correct, then the bases at the 3′ end of the internal primer should complement the bases on the L1 RNA template just proximal to the point of inversion (the bases at the orange arrow in Figure 5). This prediction is strongly supported by analysis of recent L1 insertions found in the genome database (161a). Additionally, cleavage of the second DNA strand must occur before reverse transcription has been completed. During in vitro experiments on the R2 TPRT process, the cleavage of the second DNA strand occurs after reverse transcription (115). The fact that R2 elements do not undergo L1-like inversions supports the possibility that the R2 and L1 TPRT mechanisms differ in this regard (T.H. Eickbush, personal communication).

The roughly 200-bp 3′UTR of human L1 appears to lack sequences that are important for reverse transcription, even though the 3′ 250 bps of R2 are critical for the reverse transcriptase activity of that element (114). Nearly all of this sequence can be deleted from human L1 elements with little effect on retrotransposition in HeLa cells (147). In addition, there are now many examples of retrotransposition of sequences flanking the 3′ ends of L1 elements in which these flanking sequences bear no resemblance to the L1 3′ UTR (60) (see section on L1-mediated transduction below). The L1 TPRT model predicts that the L1 polyA tail interacts with the RT domain of ORF2 protein during the initiation of reverse transcription, but the evidence on this point is indirect. The necessary pairing of the A-rich sequence at the 3′ end of an L1 element with the T-rich primer created at the integration site might explain the presence of A residues immediately after the L1 polyadenylation signal (see section on L1 RNA processing above). The A-rich regions might be positively selected over time if they are occasionally used during the priming of reverse transcription. The human SVA element, a nonautonomous retrotransposon thought to use the same L1 TPRT mechanism, also contains an A-rich region immediately following the presumed polyA signal (E.M. Ostertag & H.H. Kazazian Jr, unpublished data) (see section on genomic expansion sponsored by L1 retrotransposons below).

It has recently been demonstrated that mouse p40 (ORF1 protein) contains nucleic acid chaperone activity in vitro. Specifically, p40 promoted annealing of complementary DNA strands and aided strand exchange to form the most stable hybrids, facilitating the melting of imperfect duplexes and stabilizing perfect duplexes (125). The authors suggested that p40 protein might play a role in strand transfer during L1 reverse transcription. However, p40 protein, although readily detected in the cytoplasm, has yet to be detected within the nucleus.

## IMPACT OF L1 RETROTRANSPOSONS ON THE MAMMALIAN GENOME

L1 retrotransposons have affected the genome in numerous ways, some beneficial and others detrimental (Figure 6). However, the total effect of these actions has been major structural remodeling of the genome, occasionally altering gene expression.

Here we present what is known about L1 and the genome, along with what has been proposed and remains unproven.

## Genomic Expansion Sponsored by L1 Retrotransposons

There is considerable evidence that L1-encoded proteins preferentially act upon the L1 element that encoded them (*cis* preference). Strong indirect evidence comes from the following facts. Only about 1% of full-length L1s are active (172). If the retrotransposition machinery from the few active L1 elements could retrotranspose RNA from the many defective elements *in trans*, then one might expect most precursors of recent insertions to be inactive elements. However, this is not the case. Indeed, the precursors of three *de novo* human L1 insertions into the factor VIII, dystrophin, and CYBB genes have been isolated, and they are all active elements (42, 81, 135). Moreover, two full-length disease-causing insertions in humans, $L1_{\beta\text{-thal}}$ and $L1_{RP}$, and two full-length disease-causing insertions in mice, $L1_{spa}$ and $L1_{orl}$, are all active L1 elements (93, 151). Direct evidence of *cis* preference comes from the work of two groups using the retrotransposition assay (49, 214). Wei et al. have found that mutations in either ORF1 or ORF2 can be *trans* complemented by an active L1 at less than 1% of control levels. How the nascent proteins interact with L1 RNA or each other is a mystery. Perhaps the ribonucleoprotein particle forms as the proteins come off the ribosome, thereby limiting their availability to other L1 RNAs. In any case, *cis* preference in L1 retrotransposition is important in greatly limiting genomic expansion by defective L1s. On the other hand, preferential propagation of active L1s increases the likelihood that L1 elements will remain active in mammalian genomes (107, 214). An exception to the *cis* preference rule is the expansion of Alu elements, SVA elements, and processed pseudogenes, which almost certainly results from low-level *trans* complementation by L1 endonuclease and reverse transcriptase activities.

The 1.1 million Alu elements in the human genome contain roughly 300 bps and are composed of two similar 150-bp segments, the 3′ half of which ends in a polyA tail (169) (Figure 1). A subset of human-specific Alus from four closely related subfamilies, Alu Y, Ya5, Ya8, and Yb8, are retrotranspositionally competent (38). Most remaining Alu subfamilies are either not transcribed or weakly transcribed (180). Alu elements are concentrated in GC-rich regions of the human genome, but young Alus have a more uniform distribution across the genome (107). This latter fact has suggested that there might be positive selection for Alu sequences in regions of substantial gene expression, i.e., GC-rich DNA. Schmid has suggested that an increase in Alu transcription promotes general translation of proteins under conditions of cellular stress (175). His group has shown that Alu-mediated inhibition of PKR (double-stranded RNA-regulated protein kinase) activation results in an increase in translation (27). Under this hypothesis, Alus could be under positive selection in the readily transcribed, open chromatin near genes. This could explain the large number of "old" Alus in gene-rich GC-rich regions.

Considerable circumstantial evidence suggests that L1 elements provide key enzymatic activities for Alu insertion. (*a*) Alu sequences end in a polyA tail, which

is thought to be required for L1-mediated TPRT. (*b*) Both types of elements are usually flanked by a target site duplication of 7–20 bp, which is probably created by TPRT. (*c*) The insertion sites of Alu elements have the same general consensus sequence as the consensus sequence for L1 endonuclease (88). It was previously difficult to explain why Alu elements are concentrated in GC-rich DNA, while L1s are concentrated in AT-rich DNA, if the L1 machinery was responsible for the retrotransposition of both elements (100). The observation that recently inserted Alus and L1s are both concentrated in similar genomic sites, i.e., in AT-rich DNA, and the finding that the GC-rich distribution of Alu elements likely represents post-insertion selection, eliminates this argument against the role of L1 machinery in Alu retrotransposition (107).

Processed pseudogenes are DNA copies of RNA polymerase II-derived mRNAs that have been inserted into the genome at locations that resemble L1 target sites. These pseudogenes lack intronic RNA, usually have polyA tails, and are flanked by 7–20-bp target site duplications (209). Their sequences make up roughly 0.5% of the genome (44, 107). Two groups have presented evidence that cotransfection with an active L1 element in addition to a cDNA sequence tagged with a retrotransposition marker cassette can lead to insertion of the tagged cDNA at a low frequency (49, 214). The inserted sequences characterized by one of these groups appeared quite similar to those of endogenous processed pseudogenes (214). In addition, mutations in the L1 ORF1 or ORF2 proteins eliminated processed pseudogene formation, providing strong evidence for the role of both L1 proteins in this process (49, 214).

SVA elements are nonautonomous retrotransposons present in 2000 to 5000 copies in the human genome (158, 183; E.M. Ostertag & H.H. Kazazian Jr, unpublished data). At their 5′ ends, full-length SVA elements have up to 40 hexameric (CCCTCT) repeats. This region is followed by (*a*) a region containing antisense Alu sequence, (*b*) a VNTR region containing multiple copies of a 35–50-bp repeat, (*c*) a SINE-R sequence with similarity to the 3′ end of an endogenous retrovirus, and (*d*) a polyadenylation signal and polyA tail (183) (Figure 1). Since full-length SVA elements are all $>89\%$ similar in sequence to each other (E.M. Ostertag & H.H. Kazazian Jr, unpublished data) and are present only in hominoid primates (92), these elements are quite young by evolutionary standards, probably $<15$ My old. Many characteristics of SVA insertions are reminiscent of L1 insertions. Some insertions are 5′ truncated, they end in a polyA tail directly following a polyA signal, and they are flanked by target site duplications that are similar in length and sequence to L1 TSDs (E.M. Ostertag & H.H. Kazazian Jr, unpublished data).

Alus, processed pseudogenes, and SVA elements share features that make them candidates for retrotransposition by the L1 machinery. They are all likely transcribed in germ line cells where L1 elements are expressed and are able to retrotranspose. Furthermore, they all end in a polyA tail and are flanked by L1-like TSDs, suggesting that they are all inserted into the genome by TPRT. Of the three elements, only Alus and SVAs (see section on human disease below) have resulted in *de novo* insertions, indicating that they are currently retrotranspositionally

active. Interestingly, Alus and SVAs both have Alu sequence components. Perhaps the Alu sequences are important in the *trans*-complementation by L1, placing the element RNA in close proximity to the L1 machinery either on the ribosome or within a ribonucleoprotein particle (15).

## L1 Retrotransposons Can Cause Human Disease

L1 retrotransposons can cause human disease by a number of mechanisms, including promoting unequal homologous recombination, direct L1 insertion into genes, and providing the machinery for insertion of other retrotransposons into genes (Figure 6). Homologous recombination due to mispairing of repeated sequences has been rarely observed for L1 elements and more commonly seen for Alu elements (37). L1 mispairing and unequal crossing over has caused three recent deletions producing disease (25, 178; R Gatti, personal communication). It also produced an ancient duplication having important evolutionary consequences—the duplication of $\gamma$-globin genes that occurred in New World monkeys (54). Over 40 instances of Alu mispairing and crossing over leading to deletion have been reported in various disease states (37).

Several explanations have been proposed as to why homologous recombination is observed more frequently among Alu elements than it is among L1 elements (37). First, Alus contain sequences that are recombinogenic in other contexts (170); however, their role in Alu/Alu recombination is debatable (184). Second, the average genomic distance between two L1s is greater than that between two Alus (107), making L1/L1 events both less likely to occur and more likely to result in lethal mutations. Third, L1s tend to reside in more AT-rich DNA, while Alus reside in GC-rich DNA (107). Since AT-rich DNA is relatively gene poor, L1/L1 homologous recombination events may occur more frequently than suspected, but rarely result in deletion of gene sequences.

Recent insertions of retrotransposons are associated with 35 isolated cases of a variety of disease states in human beings (Table 1). Of these, 13 are L1 insertions, 19 are Alu insertions, 2 are SVA insertions, and 1 is an insertion of a sequence transduced by an SVA element. Another very recent L1 insertion is a normal variant in a family segregating hemophilia A (JH-25 in Table 1). There have not been recent insertions of processed pseudogenes, although some of these sequences are polymorphic as to presence within human populations and are less than 100,000 years old (6).

Although 16 of the 35 total insertions (44%) have occurred into the X chromosome, most of the excess of X chromosome insertions is due to L1 elements. Eleven of 14 recent L1 insertions (79%) are into X chromosomal genes, whereas 6 of 19 Alu insertions (32%) and 1 of 2 SVA insertions are into the X chromosome. Some L1 insertions causing X-linked disease are *de novo* events, i.e., mothers of affected males are noncarriers. A detection bias clearly exists for genes whose disruption by a single hit causes disease, i.e., X-linked and autosomal dominant disorders.

**TABLE 1**

**Human insertions**
**1. Non-LTR retrotransposons**

**L1 insertions**

| Inserted element | Disrupted gene | Insertion size | 3′ Transduction (yes or no) | Insertion site | Orientation of insertion | Reference |
|---|---|---|---|---|---|---|
| JH-27 | Factor VIII | 3.8 kb | No | Exon | Sense | (91) |
| JH-28 | Factor VIII | 2.2 kb | No | Exon | Sense and rearranged | (91) |
| JH-25 | Factor VIII | 681 nts | No | Intron | Sense | (222) |
| APC | APC | 538 nts | Yes | Exon | — | (139) |
| Dystrophin | Dystrophin | 608 nts | No | Exon | Sense | (154) |
| Dystrophin | Dystrophin | 878 nts | No | Exon | Sense | (E Bakker & G van Omenn, personal communication) |
| JH-1001 | Dystrophin | 2.0 kb | Yes | Exon | Sense and rearranged | (81) |
| $L1_{\beta\text{-thal}}$ | $\beta$-Globin | 6.0 kb | No | Intron | Antisense | (40) |
| $L1_{XLCDM}$ | Dystrophin | 524 nts | No | Exon | Antisense | (231) |
| $L1_{RP}$ | RP2 | 6.0 kb | No | Intron | Antisense | (176) |
| $L1_{CYB}$ | CYBB | 1.7 kb | Yes | Exon | Sense and rearranged | (134, 135) |
| $L1_{CYB}$ | CYBB | 940 nts | No | Intron | Sense | (133) |
| $L1_{FCMD}$ | Fukutin | 1.1 kb | No | Intron | Sense No target site duplication–7 nts deletion | (99) |
| $L1_{FIX}$ | FIX | 520 bp (460 bp of L1) | No | Exon | Sense | (226) |

**2. Nonautonomous retrotransposons**

**Alu insertions**

| Gene | Disorder | Alu subfamily | Insertion site | Orientation | De novo (yes or no) | Reference |
|---|---|---|---|---|---|---|
| NF1 | Neurofibromatosis | Ya5 | Intron | Antisense | Yes | (212) |
| BCHE | Acholinesterasemia | Yb8 | Exon | Sense | No | (149) |
| F9 | Hemophilia B | Ya5 | Exon | Sense | Yes | (211) |
| CASR | Familial hypocalciuric hypercalcemia | Ya4 | Exon | Antisense | No | (85) |
| BRCA2 | Breast cancer | Y | Exon | Sense | — | (138) |
| APC | Hereditary desmoid disease | Yb8 | Exon | Sense | No | (69) |
| BTK | X-linked agammaglobulinemia | Y | Exon | Antisense | Yes | (109) |
| IL2RG | X-linked severe combined immunodeficiency | Ya5 | Intron | Antisense | No | (109) |
| EYA1 | Branchio-oto-renal syndrome | Ya5 | Exon | Antisense | Yes | (1) |
| FGFR2 | Apert syndrome | Ya5 | Intron | Antisense | Yes | (157) |
| FGFR2 | Apert syndrome | Yb8 | Exon | Antisense | Yes | (157) |
| ADD1 | Huntington disease | — | Intron | Sense | No | (58) |
| GK | Glycerol kinase deficiency | Ya5 | Intron | Antisense | — | (233) |
| C1NH | C1 inhibitor deficiency | Y | Intron | Sense | No | (197) |
| PBGD | Acute intermittent porphyria | Ya5 | Exon | Antisense | No | (150) |
| MIVI-2 | Associated with leukemia | Ya5 | ? | ? | Yes (somatic?) | (46) |
| FIX | Hemophilia B | Ya3al | Exon | Antisense | No | (226) |
| FIX | Hemophilia B | — | Exon | Sense | No | (224) |
| FVIII | Hemophilia A | Yb8 | Exon | Antisense | No | |

(*Continued*)

**TABLE 1** (*Continued*)

**SVA-related insertions**

| Disrupted gene | Insertion type | Reference |
|---|---|---|
| FcMB | Full-length SVA | (96) |
| BTK | 5′ Truncated SVA (SINE R) | (168) |
| $\alpha$-spectrin | SVA-mediated transduction | (71) |

**Mouse insertions**

**1. LTR-retrotransposons**

**Intracisternal A-particle (IAP) insertions**

| Locus | Allele or disorder name | Reference |
|---|---|---|
| Agouti | $A^{iy}$ | (43) |
| Agouti | $A^{vy}$ | (43) |
| Agouti | $A^{vapy}$ | (137) |
| Agouti | $A^{nvy}$ | (7) |
| Mgca-mahogany | Mg | (67) |
| Mgca-mahogany | $mg^{L}$ | (67) |
| Pale ear | Hermansky-Pudlak | (56) |
| Vibrator | Vibrator | (70) |
| LamB3 | Epidermolysis bullosa | (106) |
| Tyrosinase | Somatic mosaicism | (223) |
| Eyal | BOR syndrome | (86) |
| Fused | Fused | (210) |
| Fused | Knobby | (210) |

**Early transposon (Etn) insertions**

| Locus | Allele or disorder name | Reference |
|---|---|---|
| ob | $ob^{2j}$ | (145) |
| T | Mesoderm formation | (75) |
| Muscle chloride channel | Myotonic | (196) |
| Fas | Ipr | (2) |
| MIP | Cataract | (187) |
| Adenylyl cyclase (Adcy1) Type 1 | Barrelless | (108) |
| Fidgetin | Fidget | (31) |
| Nude (whn) | nu-BC | (78) |
| Tyrosinase (Tyr) | c-3BC | (78) |
| Gli 3 | Polydactyly Nagoya | (203) |
| Stargazin (neuronal $Ca^{2+}$-channel $\gamma$ subunit) | Stargazer | (110) |

**MaLR insertions**

| Locus | Allele or disorder name | Reference |
|---|---|---|
| Cholesterol homeostasis gene | Niemann Pick C (NPC1) | (113) |

## 2. Non-LTR retrotransposons

**L1 insertions**

| Gene | Disorder or mouse name | Insertion length | Subfamily | Reference |
|---|---|---|---|---|
| Glycine receptor $\beta$-subunit | Spastic | Full-length (7.5 kb) | $T_F$ | (94, 148) |
| Reelin | Orleans reeler | Full-length | $T_F$ | (200) |
| $S_{cn}8a$ | Med | $<$100 bp | Unknown | (97) |
| Beige ($\beta\gamma$) | Chediak-Higashi | 1.1 kb | $T_F$ | (165) |
| Mitf [mi-bw] | Black-eyed white | Full-length | $T_F$ | (227) |

The recent L1 insertions have certain characteristics that mirror those of older L1s residing in the human genome. All but one of these L1 insertions is flanked by typical target site duplications. Twelve of the 14 insertions (86%) contain truncated L1 elements ranging from 500 bp to 3.8 kb in length, while two are full-length with intact ORFs. Three of the 14 insertions are associated with transduction of 3′ flanking sequences (81, 134, 139) (see section on L1-mediated transduction below).

Of the 13 disease-producing L1 insertions, 9 are into gene exons and presumably introduce nonsense codons into the coding sequence or produce skipping of the disrupted exon. The four disease-producing L1 insertions into introns cause exon skipping, decreased transcription, or decreased stability of the primary transcript. For example, an L1 insertion into an intron of the fukutin gene in Fukuyama-type muscular dystrophy patients resulted in alternative splicing (99), as did an L1 insertion into an intron of the CYBB gene in a patient with chronic granulomatous disease (133). Full-length insertions into a $\beta$-globin gene intron in a patient with $\beta$-thalassemia and into the *RP2* gene in a retinitis pigmentosa patient caused low or absent mRNA levels without aberrant splicing (40, 176).

While 13 of the L1 insertions occurred either in the germ line or very early in development, an L1 insertion into an exon of the adenomatous polyposis coli (APC) gene in a colon cancer (mentioned previously) was a somatic event in dedifferentiated cells. The insertion clearly occurred in the cancer tissue since it was not present in normal colonic tissue of the patient (139).

A striking observation concerning recent human L1 insertions is that nearly all arise from a single, relatively small, subset of human L1s called the Ta subset. This subset is characterized by substitution of ACA for GAG 92–94 bp upstream of the polyA tail (172, 189). Of the 14 recent L1 insertions, 13 (93%) are Ta subset elements. The Ta subset has recently been subdivided into Ta-0 and Ta-1 subgroups based on nucleotide sequence (17). The Ta-1 subset is younger and presently accounts for the majority of Ta elements; about 70% of the Ta-1 insertions are polymorphic as to presence in the human population, whereas only about 30% of the Ta-0 insertions are polymorphic (17). The fourteenth recent insertion is a pre-Ta element containing ACG instead of ACA at the diagnostic trinucleotide (JH-28 in Table 1) (91). Blot hybridization estimates placed the number of full-length Ta elements in the diploid human genome at about 200 copies (172). Sassaman et al. isolated a number of full-length Ta elements from a genomic library and found that 50% had two intact ORFs and roughly one half of these, or about one quarter of the total, were retrotranspositionally active in the cell culture assay. This led to an estimate of 30 to 60 active L1s in the human genome (172). In an update of the recent analysis of the working draft sequence covering roughly 90% of the human genome (the euchromatic portion), there were 57 full-length Ta elements and 22 full-length pre-Ta elements with intact ORFs (107; R.M. Badge & J.V. Moran, personal communication). If we assume that 50% of the 57 Ta elements and 10% of the pre-Ta elements are retrotranspositionally active, then there are 30 active L1s in roughly 90% of the haploid genome, or about 65 in the full diploid genome, an estimate that is similar to that of Sassaman et al. (172).

We can use the total number of recent insertions along with the number of non-recurrent mutations in the Human Gene Mutation Database (102) (http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html) to estimate the fraction of human mutations that are retrotransposition events. There are reasons why this estimate may be too low, e.g., the inability of investigators to detect all insertions greater than 1 kb in length by PCR, or too high, e.g., the failure of the database to count recurrent mutations. However, the estimate calculated in this way is roughly 1/600, with 1/1100 as the estimate for Alu insertions and 1/1500 as the estimate for L1 insertions (89).

The frequency of retrotransposition events per individual has also been estimated in a number of ways, using mutation rates in specific genes and overall mutation rates in germ cells. All estimates range between 1 retrotransposon insertion in every 4 individuals and 1 insertion in every 100 individuals (36, 89, 226). For the L1 retrotransposon, the estimate translates into 1 insertion in every 10 to 250 individuals.

## L1 Retrotransposition Can Cause Mouse Disease

Although it is likely that in the mouse unequal homologous recombination events between either non-LTR retrotransposons or LTR-retrotransposons exist, none has yet been reported. On the other hand, retrotransposon insertions account for a substantial proportion of disease-producing mutations in the mouse. Mice have many more active retrotransposons of different types than humans do, and have a correspondingly much larger fraction of spontaneous mutations due to insertion of retrotransposons.

In contrast to humans, mice are burdened by insertions of LTR retrotransposons whose origins derive from endogenous retroviruses. These are intracisternal A particles (IAPs), early transposons (Etns), and mammalian LTR-retrotransposons (MaLRs) (Table 1). The estimated 1000–2000 IAPs in the mouse genome are defective retroviral-like elements with *gag*-, *pol*-, and *env*-like similarity regions in their sequence (103). Most IAPs, but not all, lack intact ORFs and nearly all have major deletions in their *env* genes. Although low-level retrotransposition of a defective IAP element has been demonstrated in cultured cells (74), no autonomous retrotransposable IAP has been isolated and characterized to date. There have been at least 13 instances of IAP insertions in spontaneous mouse disease, all of which involve defective IAPs. Presumably IAP RNA is reverse transcribed within cytoplasmic particles by a retroviral reverse transcriptase, but the source of the enzymes and other machinery necessary for IAP mobilization is unknown.

The origin of Etns has recently been elucidated (118). These elements contain two LTRs with roughly 500 bp of sequence unrelated to retroviral sequence between the LTRs (194). Mager & Freeman have shown that the LTRs and a portion of the intervening sequence are derived from murine Type D retroviruses. At one point, Type D retroviruses underwent deletion of most of their internal sequences and later acquired new genomic sequences. Presumably, present-day Etn elements use the reverse transcriptase of their ancestor, the Type D retrovirus, to carry out

their own retrotransposition (118). Etn insertions account for at least 11 instances of mouse disease.

MaLRs are the largest family of mouse retroviral-like elements. They have an unusual mosaic structure consisting of an origin region repeat (ORR1) and a mouse transposon element (MT). They have an open reading frame of 1.3–1.6 kb internal to LTRs. The ORF does not encode reverse transcriptase or any other apparent protein (190). A single insertion of a partially deleted MaLR has been found (113). Although there are less than 5000 IAPs, Etns, and MaLRs in toto in the mouse genome, insertions of these elements account for roughly 7% to 8% of all mutations in the mouse.

L1 insertions make up another 2% to 3% of mutations in the mouse. L1 retrotransposition events are responsible for spontaneous disease in five mouse lines, the spastic mouse, the Orleans reeler mouse, the black-eyed white mouse, the beige mouse, and the med mouse (Table 1). Three of these five disorders are due to insertion of full-length L1s, $L1_{spa}$, $L1_{orl}$, and $L1_{bw}$. Of the five L1 insertions, four are derived from $T_F$ elements, a young and expanding subfamily, and the fifth is too short to classify (59). Many $T_F$ elements have retrotransposed so recently that they are highly polymorphic as to presence in various mouse subspecies and lab strains. *Mus spretus* and *Mus musculus* appear to have similar numbers of $T_F$ elements in their genomes, but the data suggest that many, if not all, of these elements are present at different locations in the genomes of these two subspecies (34).

Although at least 4 of 5 recent L1 insertions in the mouse belong to the $T_F$ subfamily, the A and $G_F$ subfamilies also contain a number of active elements. The number of active mouse L1 elements has been estimated by the cell culture assay to be 1800 $T_F$ (34), 900 A, and 400 $G_F$ (59), for a total of 3100. This number is 50 to 60 times the estimate of active human L1s, and is close to the excess of the proportion of L1 insertions causing mouse disease compared to that producing human disease (2.5% vs. 0.07%, or 35-fold).

## L1-Mediated Transduction

Because cleavage of L1 transcripts at the polyA site is often inconsistent, sequences flanking L1 3′ ends may be carried along in a retrotransposition event (146). These "stowaway" sequences are called transduced sequences. Holmes et al. first recognized an L1 transduction event, a 3′ transduction of over 500 bp included with an L1 insertion into the human dystrophin gene (81). Since then two other 3′ transductions have been recognized in association with recent L1 disease-producing insertions in humans (134, 139). Moran et al. demonstrated 3′ transduction experimentally in the cell culture assay. They placed the retrotransposition marker cassette 3′ of the L1 polyA signal, and showed that L1s are able to retrotranspose sequences from their 3′ flanks to new genomic sites. Further, L1s could retrotranspose a promoterless marker cassette into a transcribed gene, leading to formation of new fusion proteins (146). The results indicated that exons downstream of active L1s could be shuffled into new sites, thereby creating new genes. When human

genome databases were analyzed, it turned out that roughly 20% of L1 insertions contain 3′ transduced sequences (60, 166). In humans, these transductions range from 30 bp to 1 kb and may account for 25 Mbp or about 0.8% of the haploid genome. In mice, 3′ transductions have been observed in about 10% of L1 insertion events, ranging from 500 bp to 3 kb (60). Because there is usually substantial truncation of L1 sequence associated with L1 retrotranspositions, 3′ transductions could easily produce insertions completely lacking in L1 sequence (Figure 6). In addition, inversions within the transduced sequence create the possibility of a wide variety of inserted sequences originating from a single 3′ flank of a retrotransposon.

Since retrotransposons can be transcribed from upstream promoters, sequences flanking their 5′ ends can also be observed following retrotransposition. A few 5′ transductions have now been found associated with insertions of L1s in the human genome (107).

## Effects of Retrotransposon Insertions on Gene Expression

L1 sequences alter expression of some human genes (Figure 6). Three examples are a proposed enhancer activity for the apolipoprotein Lp(a) gene (229), an age-regulatory activity for the factor IX gene (104), and a locus control region activity for the growth hormone gene cluster (185). However, all of the L1s involved have considerable sequence differences from the L1 consensus sequence. Thus, L1s in general do not appear to contain sequence involved in gene regulation.

Although most L1 elements lack enhancer activity, pol II promoter activity has recently been discovered on the antisense strand of L1 DNA between nucleotides 400 to 600 within the 5′ UTR (195) (Figure 6). Fifteen cDNAs were isolated from a human teratocarcinoma cell line that contained L1 5′ UTRs spliced to the sequences of known genes or non-protein coding sequences. Four selected chimeric transcripts were found in total RNA of other cell lines. The author suggests that many human L1s contain an antisense promoter that is capable of interfering with the normal expression of neighboring genes, and that this type of transcriptional control may be quite common. Dispersion of L1 elements may have provided many opportunities for pol II transcription at new genomic locations. In an analogous situation, the mouse B2 SINE element contains pol II promoter activity, which is responsible for the transcription of at least one gene, *Lama 3* (53).

It has recently been proposed that low-level transcription of L1 elements in a subset of somatic cells may affect the expression of neighboring genes. Individual variation in the proportion of cells with retrotransposon transcription, in the location of the retrotransposons, and in the level of transcription could, in theory, lead to individual variation in susceptibility to oncogenesis or complex diseases (218). Although no experimental evidence supports this hypothesis, it would be interesting to study the effect that polymorphic L1 elements may have on neighboring gene expression.

## A Proposed Role for L1 Retrotransposons in X Chromosome Inactivation

In humans, the density of L1 sequences on the X chromosome is twice that of the average density of L1 sequences on autosomes (26% of total sequence versus 13%) (10). In every somatic cell of the mammalian female, one of the two X chromosomes is mostly inactivated (73). Lyon has proposed that L1 elements serve as "booster stations," helping to propagate the signal transmitted by *Xist* RNA (116, 117). *Xist* RNA is thought to play an important role in X inactivation because it is expressed from (22), and interacts specifically with (164), the inactive X chromosome. The evidence for the "Lyon repeat hypothesis," and an alternative version in which X chromosome heterochromatization spreads from one L1 to another through physical interaction(121), is circumstantial. (*a*) There is significant clustering of L1s around the X inactivation center (10). (*b*) L1s are in short supply in regions of the X chromosome that escape inactivation (10). (*c*) At sites of X-autosome translocations in mice, there is a positive correlation between the number of L1 elements on the autosome and the extent of heterochromatization of autosomal genes (116).

## Non-LTR Retrotransposon RT and Cellular Telomerase

The relationship of the RT of non-LTR retrotransposons to the catalytic subunit of telomerase is quite striking (111, 136). In a mechanism similar to the TPRT reaction of non-LTR retrotransposons, telomerase adds deoxyribonucleotides to the ends of chromosomes using the 3′OH end of a DNA strand as primer and a telomerase-associated RNA as template (48). Telomerase has a number of sequence domains similar to those of retrotransposon RTs, along with an additional domain not found in the RTs (111). Phylogenetic analysis suggests a close relationship between these enzymes (120), but there is controversy as to whether the catalytic subunit of telomerase is derived from retrotransposon RT or vice versa (48, 152). We favor the idea that eukaryotic cells recruited retrotransposon RT to acquire telomerase activity. The evolutionary age of non-LTR retrotransposons goes along with the very early eukaryotic origins of telomerase. A phylogenetic tree of eukaryotic RTs rooted by prokaryotic mobile elements also suggests that telomerase RT was derived from retrotransposon RT (120). There are other interesting examples of human proteins that have either evolved from transposable element proteins or have incorporated transposable element protein domains during their evolution (107, 191), notably the RAG proteins, which are responsible for V(D)J recombination (5, 76).

## A Proposed Role for L1s in DNA Repair

In the late 1980s, Edgell and colleagues proposed that double-strand break (DSB) repair is a major role for L1 insertions (47). Such a role for L1s would maintain the integrity of the genome and have important evolutionary consequences. One

would expect that L1 insertions into genomic double-stranded breaks would not be flanked by the perfect target site duplications that are created by the action of the L1 endonuclease during the TPRT process. Since many L1 elements in the genome have perfect target site duplications and because mutations in the active site residues of the L1 endonuclease greatly reduce retrotransposition capability in the cell culture assay, it is not likely that L1 elements play a major role in DSB repair in humans. However, a minor endonuclease-independent pathway may exist for L1 insertions. Indeed, examples both in humans and in mice may represent DSB repair mediated by endonuclease-independent retrotransposition of L1 elements (23, 119, 208).

## APPLICATIONS OF L1 RETROTRANSPOSONS

1. Phylogenetic markers. The presence of L1 retrotransposons in the genome for at least several hundred million years, their continuous and recent retrotransposition activity, and their stable integration are properties of L1 elements that make them excellent phylogenetic markers. Old L1 insertions can be used to perform phylogenetic analysis between species (155) and recent L1 insertions, which are polymorphic as to presence or absence in human populations, can be used to study recent human population dynamics (171, 181). Alu elements share many of the same desirable properties for use as phylogenetic markers and have been used successfully to study human diversity (213).
2. Random mutagenesis system. Weak target site preference means that L1 elements retrotranspose relatively randomly throughout the genome, and there is no bias against L1s inserting into gene sequences (146). The ability to disrupt genes randomly and stably makes L1 elements potentially very attractive for use in a random mutagenesis system in mouse. The recent development of an enhanced green fluorescent protein (EGFP)-based retrotransposition cassette that can detect single-cell retrotransposition events in vivo opens the door for such a system (162). Additionally, recent experiments demonstrate that L1 elements are able to retrotranspose in the mouse germ line at a frequency of greater than 1 in 100 sperm (161; E.M. Ostertag, unpublished data). Incorporation of gene-trapping technology may create a powerful and simple system for making mouse mutants without the requirement for embryonic stem cell-based strategies.
3. Gene delivery vector. The ability to stably integrate into the genome, the ability to carry 3′ sequences via L1-mediated transduction, and the lack of proteins that are not endogenous to the genome (and therefore potentially immunogenic) are the L1 properties that have created interest in using L1 elements as gene delivery vehicles. In fact, an L1 element packaged in a gutted adenoviral vector has been used to deliver marker genes to transformed cells in culture (193).

## ACKNOWLEDGMENTS

We thank B.L. Brouha, J.L. Goodier, E.T. Luning Prak, and M.C. Seleme for helpful review of this manuscript and A.B. Downend for help in its preparation. E.M. Ostertag is supported by a Howard Hughes Predoctoral Fellowship and H.H. Kazazian Jr is supported by grants from the NIH.



## LITERATURE CITED

1. Abdelhak S, Kalatzis V, Heilig R, Compain S, Samson D, et al. 1997. Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of EYA1. *Hum. Mol. Genet.* 6:2247–55
2. Adachi M, Watanabe-Fukunaga R, Nagata S. 1993. Aberrant transcription caused by the insertion of an early transposable element in an intron of the Fas antigen gene of lpr mice. *Proc. Natl. Acad. Sci. USA* 90:1756–60
3. Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, Hutchison CA. 1994. Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Mol. Biol. Evol.* 11:778–89
4. Adey NB, Tollefsbol TO, Sparks AB, Edgell MH, Hutchison CA. 1994. Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc. Natl. Acad. Sci. USA* 91:1569–73
5. Agrawal A, Eastman QM, Schatz DG. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394:744–51
6. Anagnou NP, Antonarakis SE, O'Brien SJ, Modi WS, Nienhuis AW. 1988. Chromosomal localization and racial distribution of the polymorphic human dihydrofolate reductase pseudogene (DHFRP1). *Am. J. Hum. Genet.* 42:345–52
7. Argeson AC, Nelson KK, Siracusa LD. 1996. Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. *Genetics* 142:557–67
8. Arning S, Gruter P, Bilbe G, Kramer A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* 2:794–810
9. 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.* 8:239–47
10. Bailey JA, Carrel L, Chakravarti A, Eichler EE. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl. Acad. Sci. USA* 97:6634–39
11. Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev.* 11:1703–16
12. Becker KG, Swergold GD, 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.* 2:1697–702
13. Benit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. 1999. ERV-L elements: a family of endogenous

retrovirus-like elements active throughout the evolution of mammals. *J. Virol.* 73:3301–8
14. Berg JM. 1990. Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J. Biol. Chem.* 265:6513–16
15. Boeke JD. 1997. LINEs and Alus—the polyA connection. [letter; comment]. *Nat. Genet.* 16:6–7
16. Bogenhagen DF, Brown DD. 1981. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell* 24:261–70
17. Boissinot S, Chevret P, Furano AV. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* 17:915–28
18. Branciforte D, Martin SL. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol. Cell. Biol.* 14:2584–92
19. Bratthauer GL, Cardiff RD, Fanning TG. 1994. Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* 73:2333–36
20. Bratthauer GL, Fanning TG. 1992. Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* 7:507–10
21. Bratthauer GL, Fanning TG. 1993. LINE-1 retrotransposon expression in pediatric germ cell tumors. *Cancer* 71:2383–86
22. Brown CJ, Lafreniere RG, Powers VE, Sebastio G, Ballabio A, et al. 1991. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* 349:82–84
23. Browning VL, Chaudhry SS, Planchart A, Dixon MJ, Schimenti JC. 2001. Mutations of the mouse Twist and sy (Fibrillin 2) genes induced by chemical mutagenesis of ES cells. *Genomics* 73:291–98
24. Burke TW, Kadonaga JT. 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* 10:711–24
25. Burwinkel B, Kilimann MW. 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J. Mol. Biol.* 277:513–17
26. Chen F, MacDonald CC, Wilusz J. 1995. Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res.* 23:2614–20
27. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW. 1998. Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell. Biol.* 18:58–68
28. Cost GJ, Boeke JD. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* 37:18081–93
29. Cost GJ, Feng Q, Boeke JD. 2000. *Initiation of L1 transposition* in vitro. Presented at Keystone Symp. Transposition and Other Genome Rearrangements, Santa Fe, NM. (Abstr. 119)
30. Cost GJ, Golding A, Schlissel MS, Boeke JD. 2001. Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* 29:573–77
31. Cox GA, Mahaffey CL, Nystuen A, Letts VA, Frankel WN. 2000. The mouse fidgetin gene defines a new role for AAA family proteins in mammalian development. *Nat. Genet.* 26:198–202
32. Cullen BR. 1998. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* 249:203–10
33. Cullen BR. 2000. Nuclear RNA export pathways. *Mol. Cell. Biol.* 20:4181–87
34. DeBerardinis RJ, Goodier JL, Ostertag EM, Kazazian HH Jr. 1998. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat. Genet.* 20:288–90
35. DeBerardinis RJ, Kazazian HH Jr. 1999. Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics* 56:317–23

36. Deininger PL, Batzer MA. 1995. SINE master genes and population biology. In *The Impact of Short, Interspersed Elements* (*SINEs*) *on the Host Genome*, ed. R Maraia, pp. 43–60. Georgetown, TX: Landes
37. Deininger PL, Batzer MA. 1999. Alu repeats and human disease. *Mol. Genet. Metab.* 67:183–93
38. Deininger PL, Batzer MA, Hutchison CA, Edgell MH. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* 8:307–11
39. Demers GW, Brech K, Hardison RC. 1986. Long interspersed L1 repeats in rabbit DNA are homologous to L1 repeats of rodents and primates in an open-reading-frame region. *Mol. Biol. Evol.* 3:179–90
40. Divoky V, Indrak K, Mrug M, Brabec V, Huisman THJ, Prchal JT. 1996. A novel mechanism of beta thalassemia: the insertion of L1 retrotransposable element into beta globin IVS II. *Blood* 88:148a (Abstr. 580)
41. Dombroski BA, Feng Q, Mathias SL, Sassaman DM, Scott AF, et al. 1994. An in vivo assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 14:4485–92
42. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH. 1991. Isolation of an active human transposable element. *Science* 254:1805–8
43. Duhl DM, Vrieling H, Miller KA, Wolff GL, Barsh GS. 1994. Neomorphic agouti mutations in obese yellow mice. *Nat. Genet.* 8:59–65
44. Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, et al. 1999. The DNA sequence of human chromosome 22. *Nature* 402:489–95. Erratum. 2000. *Nature* 404(6780):904
45. 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.* 15:1791–93
46. Economou-Pachnis A, Tsichlis PN. 1985. Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. *Nucleic Acids Res.* 13:8379–87
47. Edgell MH, Hardies SC, Loeb DD, Shehee WR, Padgett RW, et al. 1987. The L1 family in mice. *Prog. Clin. Biol. Res.* 251:107–29
48. Eickbush TH. 1997. Telomerase and retrotransposons: Which came first? [letter; comment]. *Science* 277:911–12
49. Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24:363–67
50. Fanning T, Singer M. 1987. The line-1 DNA-sequences in 4 mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res.* 15:2251–60
51. Fanning TG. 1983. Size and structure of the highly repetitive BAM HI element in mice. *Nucleic Acids Res.* 11:5073–91
52. Feng Q, Moran JV, Kazazian HH, Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87:905–16
53. Ferrigno O, Virolle T, Djabari Z, Ortonne JP, White RJ, Aberdam D. 2001. Transposable B2SINE elements can provide mobile RNA polymerase II promoters. *Nat. Genet.* 28:77–81
54. Fitch DH, Bailey WJ, Tagle DA, Goodman M, Sieu L, Slightom JL. 1991. Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proc. Natl. Acad. Sci. USA* 88:7396–400
55. Furano AV, Robb SM, Robb FT. 1988. The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. *Nucleic Acids Res.* 16:9215–31
56. Gardner JM, Wildenberg SC, Keiper NM, Novak EK, Rusiniak ME, et al. 1997. The mouse pale ear (ep) mutation

is the homologue of human Hermansky-Pudlak syndrome. *Proc. Natl. Acad. Sci. USA* 94:9238–43
57. Gentles AJ, Karlin S. 1999. Why are human G-protein-coupled receptors predominantly intronless? *Trends Genet.* 15: 47–49
58. Goldberg YP, Rommens JM, Andrew SE, Hutchinson GB, Lin B, et al. 1993. Identification of an Alu retrotransposition event in close proximity to a strong candidate gene for Huntington's disease. *Nature* 362:370–73
58a. Goodier JL, Maraia RJ. 1998. Terminator-specific recycling of a B1-Alu transcription complex by RNA polymerase III is mediated by the RNA terminus-binding protein La. *J. Biol. Chem.* 273:26110–16
59. Goodier JL, Ostertag EM, Du K, Kazazian HH Jr. 2001. Characterization of a novel active L1 retrotransposon subfamily in the mouse. *Genome Res.* In press
60. Goodier JL, Ostertag EM, Kazazian HH Jr. 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 9:653–57
61. Goodwin TJD, Ormandy JE, Poulter RTM. 2001. L1-like non-LTR retrotransposons in the yeast *Candida albicans*. *Curr. Genet.* 39:83–91
62. Gorelick RJ, Fu W, Gagliardi TD, Bosche WJ, Rein A, et al. 1999. Characterization of the block in replication of nucleocapsid protein zinc finger mutants from moloney murine leukemia virus. *J. Virol.* 73:8185–95
63. Gorelick RJ, Gagliardi TD, Bosche WJ, Wiltrout TA, Coren LV, et al. 1999. Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. *Virology* 256:92–104
64. Gorlich D, Kutay U. 1999. Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* 15:607–60
65. Gorman MA, Morera S, Rothwell DG, de La Fortelle E, Mol CD, et al. 1997. The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *EMBO J.* 16:6548–58
66. Grimaldi G, Skowronski J, Singer MF. 1984. Defining the beginning and end of KpnI family segments. *EMBO J.* 3:1753–59
67. Gunn TM, Miller KA, He L, Hyman RW, Davis RW, et al. 1999. The mouse mahogany locus encodes a transmembrane form of human attractin. *Nature* 398:152–56
68. Guo J, Wu T, Anderson J, Kane BF, Johnson DG, et al. 2000. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* 74:8980–88
69. Halling KC, Lazzaro CR, Honchel R, Bufill JA, Powell SM, et al. 1999. Hereditary desmoid disease in a family with a germline Alu I repeat mutation of the APC gene. *Hum. Hered.* 49:97–102
70. Hamilton BA, Smith DJ, Mueller KL, Kerrebrock AW, Bronson RT, et al. 1997. The vibrator mutation causes neurodegeneration via reduced expression of PITP alpha: positional complementation cloning and extragenic suppression. *Neuron* 18:711–22
71. Hassoun H, Coetzer TL, Vassiliadis JN, Sahr KE, Maalouf GJ, et al. 1994. A novel mobile element inserted in the alpha spectrin gene: spectrin dayton. A truncated alpha spectrin associated with hereditary elliptocytosis. *J. Clin. Invest.* 94:643–48
72. Hata K, Sakaki Y. 1997. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 189:227–34
73. Heard E, Clerc P, Avner P. 1997. X-chromosome inactivation in mammals. *Annu. Rev. Genet.* 31:571–610
74. Heidmann O, Heidmann T. 1991. Retrotransposition of a mouse IAP sequence

tagged with an indicator gene. *Cell* 64:159–70
75. Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H. 1990. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* 343:617–22
76. Hiom K, Melek M, Gellert M. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94:463–70
77. Hirose Y, Manley JL. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395:93–96
78. Hofmann M, Harris M, Juriloff D, Boehm T. 1998. Spontaneous mutations in SELH/Bc mice due to insertions of early transposons: molecular characterization of null alleles at the nude and albino loci. *Genomics* 52:107–9
79. Hohjoh H, Singer MF. 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* 15:630–39
80. Hohjoh H, Singer MF. 1997. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* 16:6034–43
81. Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH. 1994. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.* 7:143–48
82. 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.* 267:19765–68
83. Hyde-DeRuyscher RP, Jennings E, Shenk T. 1995. DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Res.* 23:4457–65
84. 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.* 12:4242–48
85. Janicic N, Pausova Z, Cole DE, Hendy GN. 1995. Insertion of an Alu sequence in the Ca(2+)-sensing receptor gene in familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Am. J. Hum. Genet.* 56:880–86
86. Johnson KR, Cook SA, Erway LC, Matthews AN, Sanford LP, et al. 1999. Inner ear and kidney anomalies caused by IAP insertion in an intron of the Eya1 gene in a mouse model of BOR syndrome. *Hum. Mol. Genet.* 8:645–53
87. Jubier-Maurin V, Cuny G, Laurent AM, Paquereau L, Roizes G. 1992. A new 5′ sequence associated with mouse L1 elements is representative of a major class of L1 termini. *Mol. Biol. Evol.* 9:41–55
88. Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA* 94:1872–77
89. Kazazian HH Jr. 1999. An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* 22:130
90. Kazazian HH Jr. 2000. L1 retrotransposons shape the mammalian genome. *Science* 289:1152–53
91. 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* 332:164–66
92. Kim HS, Takenaka O, Crow TJ. 1999. Cloning and nucleotide sequence of retroposons specific to hominoid primates derived from an endogenous retrovirus (HERV-K). *AIDS Res. Hum. Retroviruses* 15:595–601
93. Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W, Kazazian HH Jr. 1999. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum. Mol. Genet.* 8:1557–60
94. Kingsmore SF, Giros B, Suh D, Bieniarz M, Caron MG, Seldin MF. 1994. Glycine

receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nat. Genet.* 7:136–41

95. Klein DJ, Johnson PE, Zollars ES, De Guzman RN, Summers MF. 2000. The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the C-terminal zinc knuckle. *Biochemistry* 39:1604–12
96. Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, et al. 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 394:388–92
97. Kohrman DC, Harris JB, Meisler MH. 1996. Mutation detection in the med and medJ alleles of the sodium channel Scn8a. Unusual splicing due to a minor class AT-AC intron. *J. Biol. Chem.* 271:17576–81
98. 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* 94:10155–60
99. Kondo-Iida E, Kobayashi K, Watanabe M, Sasaki J, Kumagai T, et al. 1999. Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). *Hum. Mol. Genet.* 8:2303–9
100. Korenberg JR, Rykowski MC. 1988. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* 53:391–400
101. Korn LJ, Brown DD. 1978. Nucleotide sequence of *Xenopus borealis* oocyte 5S DNA: comparison of sequences that flank several related eucaryotic genes. *Cell* 15:1145–56
102. Krawczak M, Cooper DN. 1997. The human gene mutation database. *Trends Genet.* 13:121–22
103. Kuff EL, Lueders KK. 1988. The intracisternal A-particle gene family: structure and functional aspects. *Adv. Cancer Res.* 51:183–276
104. Kurachi S, Deyashiki Y, Takeshita J, Kurachi K. 1999. Genetic mechanisms of age regulation of human blood coagulation factor IX. *Science* 285:739–43
105. 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.* 23:3704–9
106. Kuster JE, Guarnieri MH, Ault JG, Flaherty L, Swiatek PJ. 1997. IAP insertion in the murine LamB3 gene results in junctional epidermolysis bullosa. *Mamm. Genome* 8:673–81
107. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
108. Leong WL, Dobson MJ, Logsdon JM Jr, Abdel-Majid RM, Schalkwyk LC, et al. 2000. ETn insertion in the mouse Adcy1 gene: transcriptional and phylogenetic analyses. *Mamm. Genome* 11:97–103
109. Lester T, McMahon C, Van Regemorter N, Jones A, Genet S. 1997. *X-linked immunodeficiency caused by insertion of Alu repeat sequences*. Presented at Br. Hum. Genet. Conf., York, England
110. Letts VA, Felix R, Biddlecome GH, Arikkath J, Mahaffey CL, et al. 1998. The mouse stargazer gene encodes a neuronal $Ca^{2+}$-channel gamma subunit. *Nat. Genet.* 19:340–47
111. Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR. 1997. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276:561–67
112. Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, et al. 1986. The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. *Mol. Cell. Biol.* 6:168–82

113. Loftus SK, Morris JA, Carstea ED, Gu JZ, Cummings C, et al. 1997. Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science* 277:232–35
114. Luan DD, Eickbush TH. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.* 15:3882–91
115. 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
116. Lyon MF. 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* 80:133–37
117. Lyon MF. 2000. LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? [letter; comment]. *Proc. Natl. Acad. Sci. USA* 97:6248–49
118. Mager DL, Freeman JD. 2000. Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences. *J. Virol.* 74:7221–29
119. Mager DL, Henthorn PS, Smithies O. 1985. A Chinese G gamma + (A gamma delta beta)zero thalassemia deletion: comparison to other deletions in the human beta-globin gene cluster and sequence analysis of the breakpoints. *Nucleic Acids Res.* 13:6559–75
120. Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* 16:793–805
121. Marahrens Y. 1999. X-inactivation by chromosomal pairing events. *Genes Dev.* 13:2624–32
122. 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
123. Martin SL. 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 11:4804–7
124. Martin SL, Branciforte D. 1993. Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 13:5383–92
125. Martin SL, Bushman FD. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* 21:467–75
126. 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
127. Mathias SL, Scott AF. 1993. Promoter binding proteins of an active human L1 retrotransposon. *Biochem. Biophys. Res. Commun.* 191:625–32
127a. Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. *Science* 254:1800–10
128. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, et al. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385:357–61
129. McDevitt MA, Imperiale MJ, Ali H, Nevins JR. 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. *Cell* 37:993–99
130. McLauchlan J, Gaffney D, Whitton JL, Clements JB. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. *Nucleic Acids Res.* 13:1347–68
131. McMillan JP, Singer MF. 1993. Translation of the human LINE-1 element, L1Hs. *Proc. Natl. Acad. Sci. USA* 90: 11533–37

132. Mears ML, Hutchison CA. 2001. The evolution of modern lineages of mouse L1 elements. *J. Mol. Evol.* 52:51–62
133. Meischl C, Boer M, Ahlin A, Roos D. 2000. A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Eur. J. Hum. Genet.* 8:697–703
134. Meischl C, De Boer M, Roos D. 1998. Chronic granulomatous disease caused by line-1 retrotransposons. *Eur. J. Haematol.* 60:349–50
135. Meischl C, de Boer M, Ostertag EM, Zhang Y, Neijens HJ, et al. 2001. Characterization of an active L1 precursor to an insertion in the CYBB gene. In preparation
136. Meyerson M, Counter CM, Eaton EN, Ellisen LW, Steiner P, et al. 1997. hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell* 90:785–95
137. Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP. 1994. Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* 8:1463–72
138. Miki Y, Katagiri T, Kasumi F, Yoshimoto T, Nakamura Y. 1996. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* 13:245–47
139. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, et al. 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52:643–45
140. 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.* 20:3139–45
141. Minchiotti G, Contursi C, Di Nocera PP. 1997. Multiple downstream promoter modules regulate the transcription of the *Drosophila melanogaster* I, Doc and F elements. *J. Mol. Biol.* 267:37–46
142. Mizrokhi LJ, Georgieva SG, Ilyin YV. 1988. jockey, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. *Cell* 54:685–91
143. Mizuuchi K. 1992. Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu. Rev. Biochem.* 61:1011–51
144. Mol CD, Kuo CF, Thayer MM, Cunningham RP, Tainer JA. 1995. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* 374:381–86
145. Moon BC, Friedman JM. 1997. The molecular basis of the obese mutation in ob2J mice. *Genomics* 42:152–56
146. Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* 283:1530–34
147. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* 87:917–27
148. Mulhardt C, Fischer M, Gass P, Simon-Chazottes D, Guenet JL, et al. 1994. The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. *Neuron* 13:1003–15
149. Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, et al. 1991. Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. *Proc. Natl. Acad. Sci. USA* 88:11315–19
150. Mustajoki S, Ahola H, Mustajoki P, Kauppinen R. 1999. Insertion of Alu element responsible for acute intermittent porphyria. *Hum. Mutat.* 13:431–38
151. Naas TP, DeBerardinis RJ, Moran JV, Ostertag EM, Kingsmore SF, et al. 1998. An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.* 17:590–97

152. Nakamura TM, Cech TR. 1998. Reversing time: origin of telomerase. *Cell* 92:587–90
153. Nakielny S, Dreyfuss G. 1999. Transport of proteins and RNAs in and out of the nucleus. *Cell* 99:677–90
154. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Minami R, et al. 1993. Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* 91:1862–67
155. Nikaido M, Rooney AP, Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* 96:10261–66
156. Nur I, Pascale E, Furano AV. 1988. The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. *Nucleic Acids Res.* 16:9233–51
157. Oldridge M, Zackai EH, McDonald-McGinn DM, Iseki S, Morriss-Kay GM, et al. 1999. De novo alu-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. *Am. J. Hum. Genet.* 64:446–61
158. Ono M, Kawakami M, Takezawa T. 1987. A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res.* 15:8725–37
159. Ooi SL, Pope I, Hope TJ, Boeke JD. 2000. *LINE-1 RNA nuclear export signals*. Presented at Keystone Symp. Transposition and Other Genome Rearrangements, Santa Fe, NM (Abstr. 226)
160. Osheim YN, Proudfoot NJ, Beyer AL. 1999. EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell* 3:379–87
161. Ostertag EM, DeBerardinis RJ, Kim K-S, Gerton G, Kazazian HH Jr. 2000. Human L1 retrotransposition in germ cells of transgenic mice. *Am. J. Hum. Genet.* 67:A102 (Abstr.)
161a. Ostertag EM, Kazazian HH Jr. 2001. Twin priming, a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* In press
162. Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV, Kazazian HH Jr. 2000. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* 28:1418–23
163. Padgett RW, Hutchison CA 3rd, Edgell MH. 1988. The F-type 5′ motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. *Nucleic Acids Res.* 16:739–49
164. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996. Requirement for Xist in X chromosome inactivation. *Nature* 379:131–37
165. Perou CM, Pryor RJ, Naas TP, Kaplan J. 1997. The bg allele mutation is due to a LINE1 element retrotransposition. *Genomics* 42:366–68
166. Pickeral OK, Makalowski W, Boguski MS, Boeke JD. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* 10:411–15
167. Roberts RM, Liu L, Guo Q, Leaman D, Bixby J. 1998. The evolution of the type I interferons. *J. Interferon Cytokine Res.* 18:805–16. Erratum. *J. Interferon Cytokine Res*. 1999. 19(4):427
168. Rohrer J, Minegishi Y, Richter D, Eguiguren J, Conley ME. 1999. Unusual mutations in Btk: an insertion, a duplication, an inversion, and four large deletions. *Clin. Immunol.* 90:28–37
169. Rowold DJ, Herrera RJ. 2000. Alu elements and the human genome. *Genetica* 108:57–72
170. Rudiger NS, Gregersen N, Kielland-Brandt MC. 1995. One short well conserved region of Alu-sequences is involved in human gene rearrangements

and has homology with prokaryotic chi. *Nucleic Acids Res.* 23:256–60
171. Santos FR, Pandya A, Kayser M, Mitchell RJ, Liu A, et al. 2000. A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. *Hum. Mol. Genet.* 9:421–30
172. Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, et al. 1997. Many human L1 elements are capable of retrotransposition. *Nat. Genet.* 16:37–43
173. Saxton JA, Martin SL. 1998. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *J. Mol. Biol.* 280:611–22
174. Schichman SA, Severynse DM, Edgell MH, Hutchison CA 3rd. 1992. Strand-specific LINE-1 transcription in mouse F9 cells originates from the youngest phylogenetic subgroup of LINE-1 elements. *J. Mol. Biol.* 224:559–74
175. Schmid CW. 1998. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* 26:4541–50
176. Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, et al. 1998. Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat. Genet.* 19:327–32
177. Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, et al. 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics* 1:113–25
178. Segal Y, Peissel B, Renieri A, de Marchi M, Ballabio A, et al. 1999. LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. *Am. J. Hum. Genet.* 64:62–69
179. Seto E, Shi Y, Shenk T. 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* 354:241–45
180. Shaikh TH, Roy AM, Kim J, Batzer MA, Deininger PL. 1997. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* 271:222–34
181. Sheen FM, Sherry ST, Risch GM, Robichaux M, Nasidze I, et al. 2000. Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome Res.* 10:1496–508
182. Shehee WR, Chao SF, Loeb DD, Comer MB, Hutchison CA 3rd, Edgell MH. 1987. Determination of a functional ancestral sequence and definition of the 5′ end of A-type mouse L1 elements. *J. Mol. Biol.* 196:757–67
183. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, et al. 1994. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. *J. Biol. Chem.* 269:8466–76
184. Shen MR, Deininger PL. 1992. An in vivo assay for measuring the recombination potential between DNA sequences in mammalian cells. *Anal. Biochem.* 205:83–89
185. Shewchuk BM, Cooke NE, Liebhaber SA. 2001. The human growth hormone locus control region mediates long-distance transcriptional activation independent of nuclear matrix attachment regions. *Nucleic Acids Res.* 29:3356–61
186. Shi Y, Lee JS, Galvin KM. 1997. Everything you have ever wanted to know about Yin Yang 1. *Biochim. Biophys. Acta* 1332:F49–66
187. Shiels A, Bassnett S. 1996. Mutations in the founder of the MIP gene family underlie cataract development in the mouse. *Nat. Genet.* 12:212–15
188. Shrivastava A, Calame K. 1994. An analysis of genes regulated by the multifunctional transcriptional regulator Yin Yang-1. *Nucleic Acids Res.* 22:5151–55
189. Skowronski J, Singer MF. 1986. The

abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. *Cold Spring Harbor Symp. Quant. Biol.* 51:457–64

190. Smit AF. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 21:1863–72
191. Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* 9:657–63
192. Smit AF, Riggs AD. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* 93:1443–48
193. Soifer H, Higo C, Kazazian HH Jr., Moran JV, Mitani K, Kasahara N. 2001. Stable integration of transgenes delivered by a retrotransposon-adenovirus hybrid vector. *Hum. Gene Ther.* 11:1417–28
194. Sonigo P, Wain-Hobson S, Bougueleret L, Tiollais P, Jacob F, Brulet P. 1987. Nucleotide sequence and evolution of ETn elements. *Proc. Natl. Acad. Sci. USA* 84:3768–71
195. Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 21:1973–85
196. Steinmeyer K, Klocke R, Ortland C, Gronemeier M, Jockusch H, et al. 1991. Inactivation of muscle chloride channel by transposon insertion in myotonic mice. *Nature* 354:304–8
197. Stoppa-Lyonnet D, Carter PE, Meo T, Tosi M. 1990. Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. *Proc. Natl. Acad. Sci. USA* 87:1551–55
198. Summers MF, South TL, Kim B, Hare DR. 1990. High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* 29:329–40
199. Swergold GD. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10:6718–29

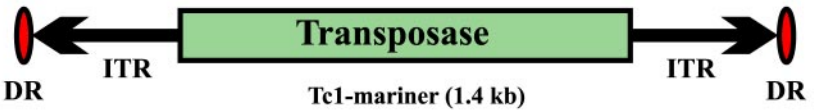
199a. Sukarova E, Dimovski AJ, Tchacarova P, Petkov GI, Efremov GD. 2001. An Alu insert as the cause of a severe form of Hemophilia A. *Acta Haematol.* In press

200. Takahara T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N, et al. 1996. Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. *Hum. Mol. Genet.* 5:989–93
201. Tchenio T, Casella JF, Heidmann T. 2000. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* 28:411–15
202. Thayer RE, Singer MF, Fanning TG. 1993. Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene* 133:273–77
203. Thien H, Ruther U. 1999. The mouse mutation Pdn (Polydactyly Nagoya) is caused by the integration of a retrotransposon into the Gli3 gene. *Mamm. Genome* 10:205–9
204. Tollervey D, Caceres JF. 2000. RNA processing marches on. *Cell* 103:703–9
205. 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* 92:1520–24
206. Usheva A, Shenk T. 1994. TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* 76:1115–21
207. Usheva A, Shenk T. 1996. YY1 transcriptional initiator: protein interactions and association with a DNA site containing unpaired strands. *Proc. Natl. Acad. Sci. USA* 93:13571–6
208. Van de Water N, Williams R, Ockelford P, Browett P. 1998. A 20.7 kb deletion

within the factor VIII gene associated with LINE-1 element insertion. *Thrombosis Haemostasis* 79:938–42
209. Vanin EF. 1985. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* 19:253–72
210. Vasicek TJ, Zeng L, Guan XJ, Zhang T, Costantini F, Tilghman SM. 1997. Two dominant mutations in the mouse fused gene are the result of transposon insertions. *Genetics* 147:777–86
211. Vidaud D, Vidaud M, Bahnak BR, Siguret V, Gispert Sanchez S, et al. 1993. Haemophilia B due to a de novo insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* 1:30–36
212. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. 1991. A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 353:864–66
213. Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, et al. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* 68:738–52
214. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, et al. 2001. Human L1 retrotransposition: cis preference versus *trans* complementation. *Mol. Cell. Biol.* 21:1429–39
215. Weis L, Reinberg D. 1992. Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J.* 6:3300–9
216. Weis L, Reinberg D. 1997. Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell. Biol.* 17:2973–84
217. Whitcomb JM, Hughes SH. 1992. Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell Biol.* 8:275–306
218. Whitelaw E, Martin DIK. 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* 27:361–65
219. Wilkinson DA, Mager DL, Leong JC. 1994. Endogenous human retroviruses. In *The Retroviridae*, ed. JA Levy, pp. 465–535. New York: Plenum
220. Wincker P, Jubier-Maurin V, Roizes G. 1987. Unrelated sequences at the 5′ end of mouse LINE-1 repeated elements define two distinct subfamilies. *Nucleic Acids Res.* 15:8593–606
221. Woodcock DM, Williamson MR, Doherty JP. 1996. A sensitive RNase protection assay to detect transcripts from potentially functional human endogenous L1 retrotransposons. *Biochem. Biophys. Res. Commun.* 222:460–65
222. Woods-Samuels P, Wong C, Mathias SL, Scott AF, Kazazian HH, Antonarakis SE. 1989. Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics* 4:290–96
223. Wu M, Rinchik EM, Wilkinson E, Johnson DK. 1997. Inherited somatic mosaicism caused by an intracisternal A particle insertion in the mouse tyrosinase gene. *Proc. Natl. Acad. Sci. USA* 94:890–94
224. Wulff K, Gazda H, Schroder W, Robicka-Milewska R, Herrmann FH. 2000. Identification of a novel large F9 gene mutation—an insertion of an Alu repeated DNA element in exon e of the factor 9 gene. *Hum. Mutat.* 15: 299
225. Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9:3353–62
226. Xuemin L, Scaringe WA, Hill KA, Roberts S, Mengos A, et al. 2001. Frequency of recent retrotransposition events in the human factor IX gene. *Hum. Mutat.* 17:511–19

227. Yajima I, Sato S, Kimura T, Yasumoto K, Shibahara S, et al. 1999. An L1 element intronic insertion in the black-eyed white (Mitf[mi-bw]) gene: the loss of a single Mitf isoform responsible for the pigmentary defect and inner ear deafness. *Hum. Mol. Genet.* 8:1431–41
228. Yang J, Malik HS, Eickbush TH. 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl. Acad. Sci. USA* 96:7847–52
229. Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R. 1998. Apolipoprotein(a) gene enhancer resides within a LINE element. *J. Biol. Chem.* 273:891–97
230. Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13:335–40
231. Yoshida K, Nakamura A, Yazaki M, Ikeda S, Takeda S. 1998. Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Hum. Mol. Genet.* 7:1129–32
232. Zawel L, Reinberg D. 1993. Initiation of transcription by RNA polymerase II: a multi-step process. *Prog. Nucleic Acid Res. Mol. Biol.* 44:67–108
233. Zhang Y, Dipple KM, Vilain E, Huang BL, Finlayson G, et al. 2000. AluY insertion (IVS4–52ins316alu) in the glycerol kinase gene from an individual with benign glycerol kinase deficiency. *Hum. Mutat.* 15:316–23

# DNA Transposons

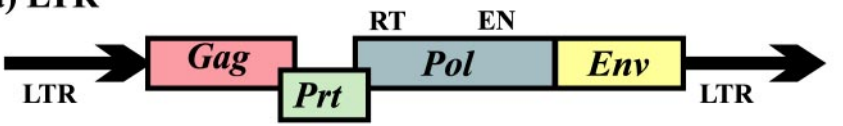



Tc1-mariner (1.4 kb)

# Retrotransposons

## -Autonomous

### a) LTR

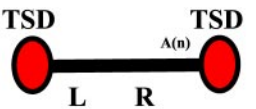



HERV (~9.2 kb)

### b) Non-LTR


L1 Element (6 kb)

## -Non-Autonomous




Alu Element (.3 kb)


SVA Element (>1 kb)

See legend on next page

**Figure 1** (See figure on previous page) Structure of mammalian transposable elements. Mammalian transposable elements consist of DNA transposons and retrotransposons. DNA transposons are flanked by inverted terminal repeats (ITRs) and have a single open reading frame (ORF) that encodes a transposase. They are also flanked by short direct repeats (DRs) created during the integration process. An example of a DNA transposon is the Tc1-mariner transposon. Retrotransposons can be divided into autonomous and nonautonomous elements based upon whether they have ORFs (*colored rectangles*) that encode proteins required for their retrotransposition. Autonomous retrotransposons are classified as (*a*) long terminal repeat (LTR) or (*b*) non-LTR. An example of an LTR retrotransposon is the human endogenous retrovirus (HERV). The LTR retrotransposons are flanked by LTRs and have partially overlapping ORFs for their group-specific antigen (*gag*), protease (*prt*), polymerase (*pol*), and envelope (*env*) genes. Also shown are the reverse transcriptase (RT) and endonuclease (EN) domains of the polymerase protein. An L1 element is an example of a non-LTR retrotransposon. L1s consist of a 5′ untranslated region (5′UTR), two ORFs separated by a short intergenic region, a 3′UTR, a polyA signal (AATAAA), and a polyA tail (A(n)). L1 elements are often flanked by 7–20-bp target site duplications (TSD)s. Shown are the RT and EN domains of the ORF2 protein, as well as a conserved cysteine-rich motif (C). The Alu element and the SVA element are examples of nonautonomous retrotransposons. Alu elements contain two similar sequences, the left monomer (L) and the right monomer (R) and end in a polyA tail. SVA elements consist of CCCTCT hexameric repeats, an antisense Alu-like region, a VNTR region, a region (SINE-R) with homology to the end of a HERV, a polyA signal and a polyA tail. Alus and SVAs are flanked by L1-like TSDs. The approximate size in kilobases (kb) of a full-length element of each example is indicated in parentheses.

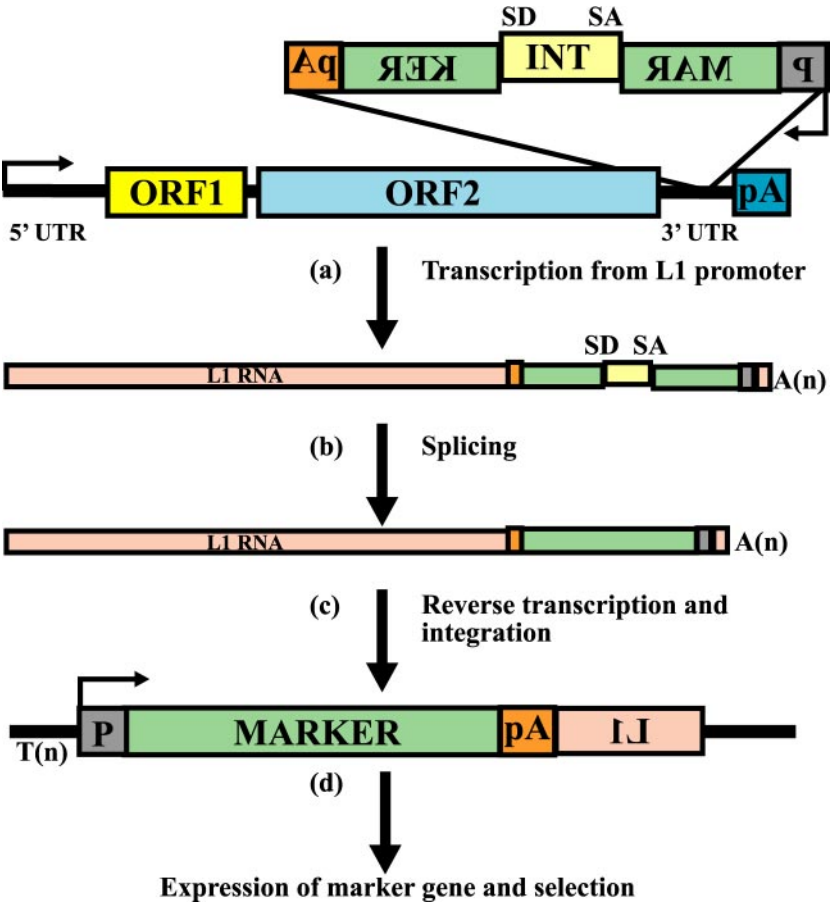



**Figure 2** A cell culture-based retrotransposition assay. In the retrotransposition assay, a full-length L1 element is tagged with a retrotransposition cassette. A retrotransposition cassette consists of a marker gene (*light green rectangle*) interrupted by an intron (*light yellow rectangle*) in the opposite transcriptional orientation. The splice donor (SD) and splice acceptor (SA) sites are indicated. Transcripts directed from the marker's promoter (*light gray rectangle and bent arrow*) cannot remove the intron by splicing and will not produce functional protein. The entire cassette is cloned into the 3′ untranslated region (3′UTR) of an L1 element in the orientation opposite that of the L1 promoter (*5′UTR and bent arrow*). The marker can only become activated (*a*) when a full-length L1 element is transcribed (L1 RNA is represented in *pink*, the color of marker components have been maintained), (*b*) the intron is removed by splicing, and (*c*) the RNA is reverse transcribed and integrated into the genome. (*d*) Transcripts directed from the marker's promoter after a retrotransposition event produce functional protein. The tagged-L1 construct is transiently transfected into cultured cells. Typical markers include the neomycin phosphotransferase (*neo*) gene or the Enhanced Green Fluorescent Protein (EGFP) gene. Positive cells are selected by growth in G418 containing media or by analysis for fluorescence, respectively (147, 162).

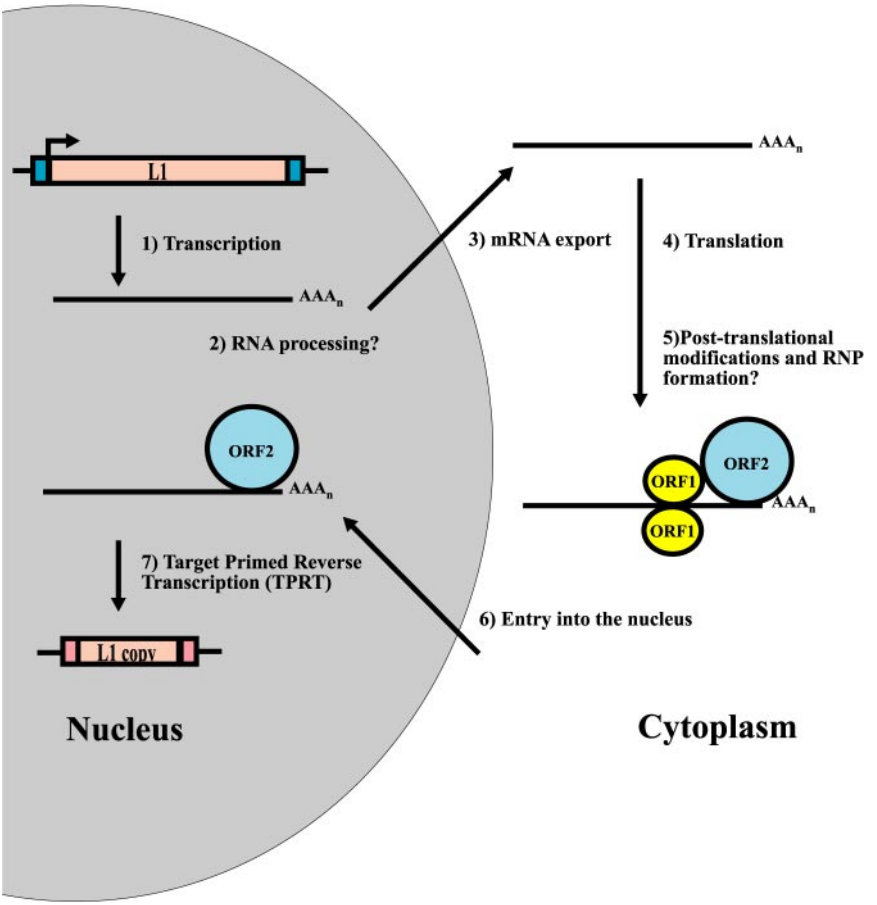



**Figure 3** The steps in L1 retrotransposition. A full-length active L1 element is transcribed from its internal promoter (*bent arrow*) to produce a bicistronic mRNA. It is currently unknown if the RNA undergoes processing or how the RNA is exported from the nucleus. Once in the cytoplasm, the ORF1 and ORF2 proteins are translated and specifically function on the RNA that transcribed them (*cis* preference). At least one L1 RNA molecule, one ORF2 molecule, and one or more ORF1 molecules may assemble into a ribonucleoprotein (RNP) complex that is an intermediate in retrotransposition. Both the ORF2 protein and associated L1 RNA must gain access to the nucleus, where the L1 RNA is reverse transcribed and integrated into a new genomic location by a process called target primed reverse transcription (TPRT). Many L1 elements undergo 5′ truncation or 5′ inversion and truncation during the TPRT process, resulting in an inactive DNA copy of the original element. The TPRT process creates 7–20-bp target site duplications that flank the L1 element (*blue and pink rectangles*).

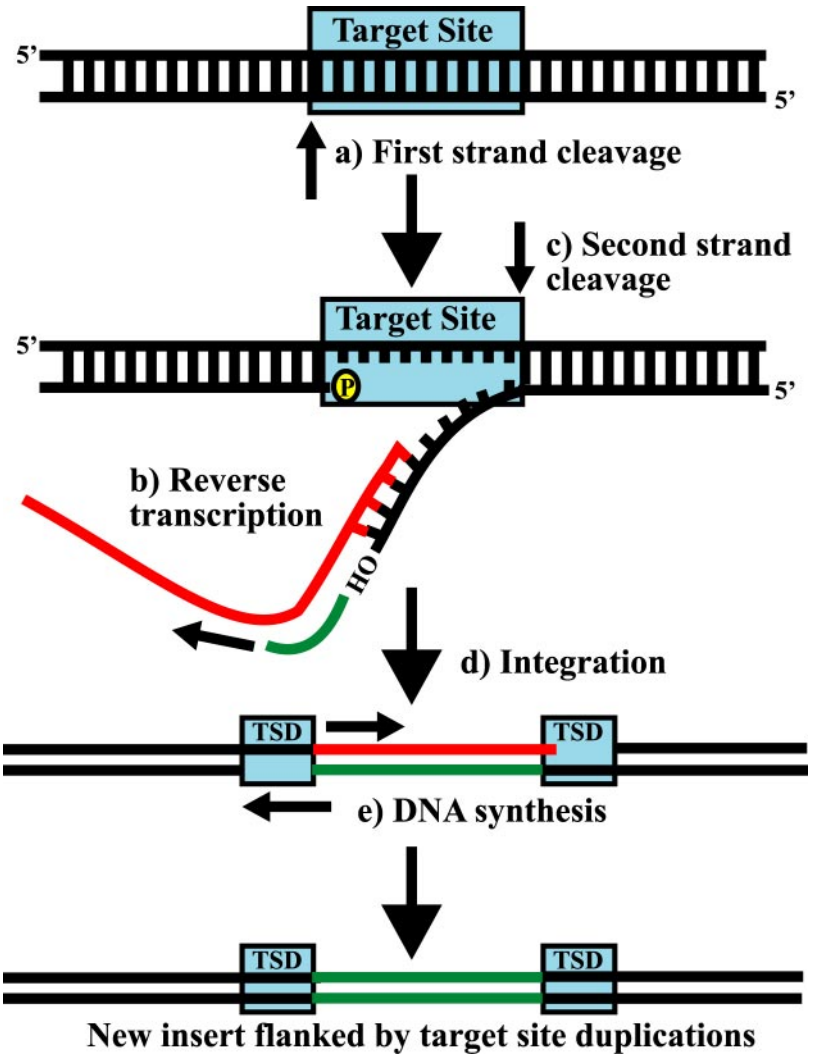



**Figure 4** Target primed reverse transcription (TPRT). The non-LTR retrotransposons are thought to integrate by TPRT. This TPRT model is based upon the mechanism worked out *in vivo* for the R2 retrotransposon (115). (*a*) During TPRT, the retrotransposon's endonuclease cleaves one strand of genomic DNA at its target site (*blue rectangle*), producing a 3′ hydroxyl (OH) at the nick. (*b*) The retrotransposon RNA (*red line*) inserts at the nick and the retrotransposon's reverse transcriptase uses the free 3′OH to prime reverse transcription. Reverse transcription proceeds, producing a cDNA of the retrotransposon RNA (*green line*). (*c*) The endonuclease cleaves the second DNA strand of the target site to produce a staggered break. (*d*) The cDNA inserts into the break by an unknown mechanism. (*e*) Removal of RNA and completion of DNA synthesis produces a complete insertion flanked by target site duplications (TSDs).

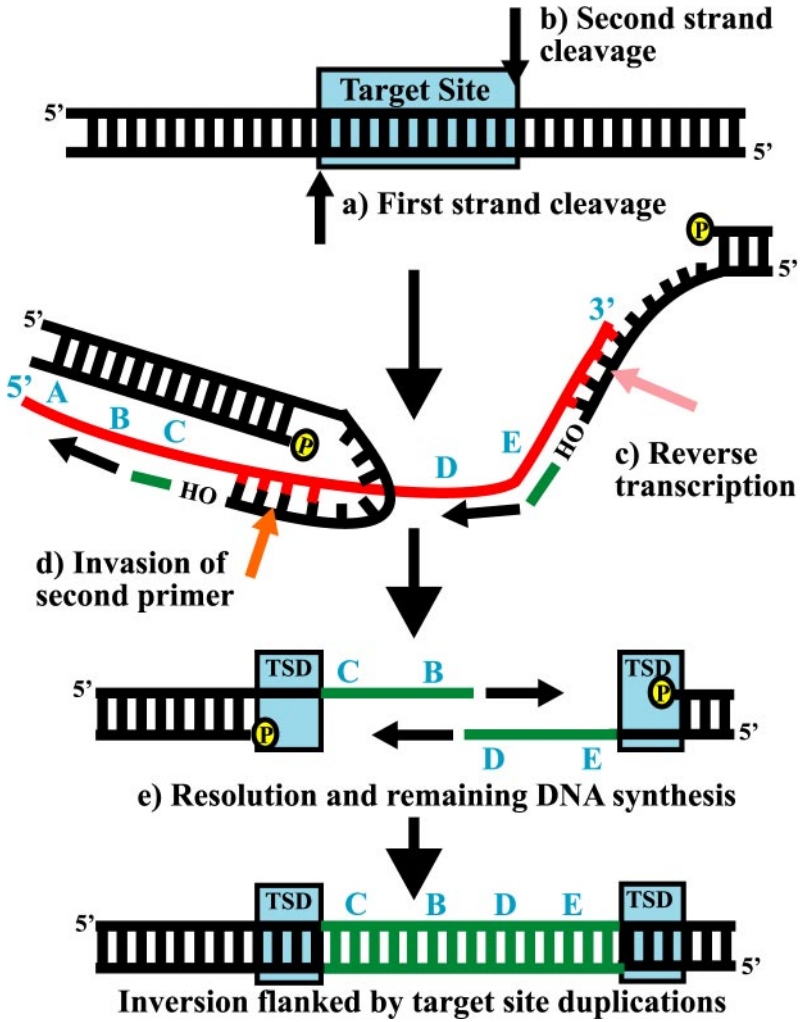
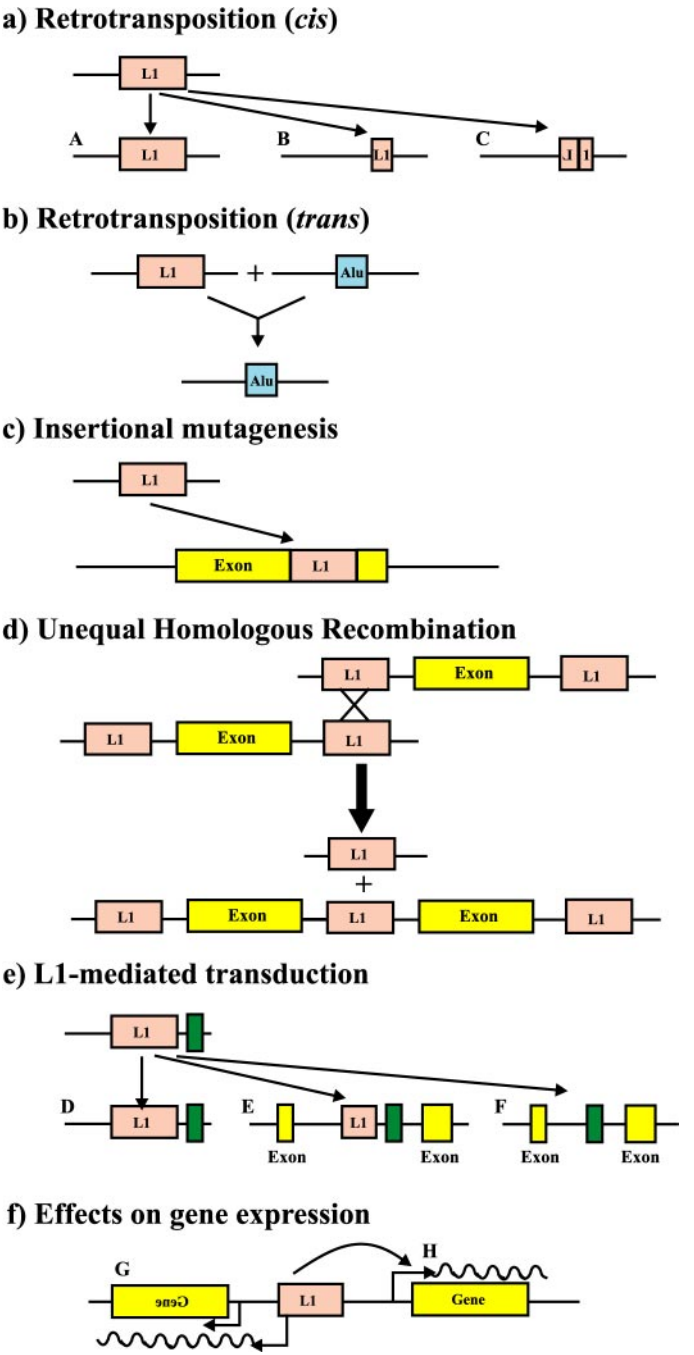



**Figure 5** Twin priming. This model demonstrates our proposed mechanism by which L1 5′ inversions are created during the TPRT process. (*a*) The L1 endonuclease cleaves one strand of genomic DNA at its target site (*blue rectangle*), creating a 3′ hydroxyl (OH) at the nick. (*b*) The endonuclease performs second strand cleavage before reverse transcription has been completed, creating a second 3′ hydroxyl and staggered break. (*c*) The L1 RNA inserts into the break and the L1 reverse transcriptase uses the first 3′OH to initiate reverse transcription (*pink arrow*). (*d*) The second 3′OH invades the RNA internally and is used to prime reverse transcription at a second site (*orange arrow*). (*e*) Resolution of the RNA/cDNA structure and completion of DNA synthesis produces an insertion with a 5′ inversion. The entire insertion is flanked by target site duplications (TSDs). The L1 RNA sequence is represented by 5′-A-B-C-D-E-3′. After the inversion, the insertion sequence is 5′-C-B-D-E-3′.




See legend on next page

**Figure 6** (See figure on previous page) Impact of L1 retrotransposons on the mammalian genome. L1 elements have had a variety of effects on the human genome. (*a*) L1 elements expand the genome by their retrotransposition. L1 elements replicate by a "copy and paste" mechanism. The L1 proteins work preferentially on the RNA that transcribed them (*cis* preference). Therefore, only full-length elements with two open reading frames are active. The active L1 is first transcribed *in situ*, and then the RNA is reverse transcribed and integrated into a new location. L1 elements can integrate as full-length copies (*A*), however, they frequently 5′ truncate (*B*), or 5′ invert and truncate (*C*), producing an inactive copy of the original element. (*b*) Some nonautonomous retrotransposons, such as Alu, are an exception to the *cis* preference rule and are retrotransposed *in trans* by the L1 machinery, further contributing to genomic expansion. (*c*) L1 elements occasionally insert into gene sequences, thereby causing genetic disorders. (*d*) L1 elements can also cause disease by creating deletions and duplications after unequal homologous recombination. (*e*) L1 elements often bypass their own polyA signal and use a downstream signal. This results in the transduction of 3′ sequences, potentially gene exon sequences (*green rectangle*), upon their retrotransposition (*D*). Retrotransposition of a 3′ exon into another gene could result in exon shuffling (*E*). L1 elements could even shuffle exons without leaving evidence of themselves if they severely 5′ truncate during retrotransposition of a 3′ exon (*F*). (*f*) The expression of genes can be affected by the presence of an L1 element. Some L1s have antisense Pol II promoters, which can affect the expression of nearby genes (*G*). Other L1s have acquired an enhancer function and can regulate expression of local genes (*H*).