

Published in final edited form as:
Front Biosci. ; 17: 1329–1344.

Expressing genes do not forget their LINEs: transposable elements and gene expression

Kristine J. Kines¹ and Victoria P. Belancio¹

¹Department of Structural and Cellular Biology, Tulane University School of Medicine, Tulane University Cancer Center and Tulane Center for Aging

1. ABSTRACT

Historically the accumulated mass of mammalian transposable elements (TEs), particularly those located within gene boundaries, was viewed as a genetic burden potentially detrimental to the genomic landscape. This notion has been strengthened by the discovery that transposable sequences can alter the architecture of the transcriptome, not only through insertion, but also long after the integration process is completed. Insertions previously considered harmless are now known to impact the expression of host genes via modification of the transcript quality or quantity, transcriptional interference, or by the control of pathways that affect the mRNA life-cycle. Conversely, several examples of the evolutionary advantageous impact of TEs on the host gene structure that diversified the cellular transcriptome are reported. TE-induced changes in gene expression can be tissue- or disease-specific, raising the possibility that the impact of TE sequences may vary during development, among normal cell types, and between normal and disease-affected tissues. The understanding of the rules and abundance of TE-interference with gene expression is in its infancy, and its contribution to human disease and/or evolution remains largely unexplored.

Keywords

Transposable elements; LINE-1; L1; SINE; Alu; SVA; HERVs; Gene Expression; Splicing; Polyadenylation; Review

2. INTRODUCTION

Mammalian transposable elements have been restructuring their host genomes for millions of years, to both deleterious and advantageous effects. From the time of their discovery, considerable effort has been devoted to the investigation of the mechanistic aspects of TE mobilization: the process by which they perturb genome integrity. While the disruption of normal gene function by transposable elements upon integration into exonic regions is obvious, their post-insertional effects on gene expression have not received much attention. The appreciation for TE modification of mammalian gene expression gained significant interest after the discoveries that the majority of transposable elements either carry *cis*-acting elements in their sequence that are recognized by the mammalian transcriptional or RNA processing machineries or have high propensity for accrual of these *cis*-signals via mutations long after the completion of the integration process. The potential for TEs to modify normal gene expression even when they are located outside of the gene boundaries is consistent with associations between non-coding DNA and human disease (47,74). This

Send correspondence to: Victoria P. Belancio, Tulane University, Department of Structural and Cellular Biology, Tulane School of Medicine, Tulane Cancer Center and Tulane Center for Aging, SL-49 1430 Tulane Ave., New Orleans, LA, 70112, Tel: 504-988-4506, Fax: 504-988-1687, vperepe@tulane.edu.

review highlights the current understanding of the post-insertional effects of TEs on mammalian gene function with a heavy focus on human TEs.

While the total number of cellular genes has remained relatively conserved in the course of mammalian evolution, the genomic mass occupied by transposable elements populating mammalian genomes has grown up to as much as 52% (58,65,79,94,119,126). The proportion of identifiable TE sequences in the human genome is reported at 44% (58). These numbers are likely an underrepresentation of the true TE contribution to modern genomic composition because a significant fraction of TE sequences is expected to have deteriorated beyond recognition through genomic drift. The forces behind this variation are not fully understood, although one of the obvious explanations is a potential difference in the rate of TE buildup through mobilization in any given genome and loss of accumulated TE sequences due to recombination.

Transposable elements are classified into two major categories based on the specific steps of their amplification method. Class II or DNA transposons, mobilize through a “cut-and-paste” mechanism. This method of mobilization relies on a transposon-encoded enzyme, transposase, which excises the parental copy from its genomic location by recognizing transposon-specific sequence and introduces the removed DNA into a new genomic location. The activity of DNA transposons appears to have ceased in the human genome based on evolutionary analysis and the absence of reports supporting their contribution to human disease.

The other major category of transposons, class I or RNA transposons, more often referred to as retrotransposons, amplify through a “copy-and-paste” mechanism using a RNA intermediate. Retroelements are further subdivided into Long-Terminal Repeat (LTR)-containing or non-LTR elements (Figure 1). The LTR retroelement group includes endogenous retroviruses, whose genome organization parallels that of retroviruses (reviewed in (68)). While this group of retroelements exhibits relatively high activity in the mouse genome (based on the frequency of disease-causing germ-line mutations), there are no reports of congenital human diseases associated with the insertion of the human endogenous LTR elements. The non-LTR group of transposable elements encompasses three families of retroelements: Long Interspersed Element-1 (LINE-1 or L1), Short Interspersed Elements (SINEs) represented by Alu elements in the human genome, and SVA (SINE-VNTR-Alu) (55,88,96,117). Out of the three, only L1 elements have coding capacity. L1 is therefore classified as an autonomous retrotransposon. The two open reading frames (ORFs) present within the L1 sequence encode for the proteins that are absolutely essential for L1, and likely SVA, mobilization (83,88). In contrast, only one of these proteins, ORF2, is necessary in *trans* for Alu retrotransposition (32). Although the ORF1 protein is not required for Alu mobilization, retrotransposition of Alu elements is significantly enhanced in the presence of ORF1 (116). Combined, Alu and SVA elements form a non-autonomous group of non-LTR retrotransposons because they must parasitize on the L1 retrotransposition machinery for their mobilization. In contrast to the LTR transposons that appear to be dormant in the human genome, all three types of non-LTR retroelements are currently active as evidenced by their contribution to human germ-line disease (reviewed in (9)).

The retrotransposition process involves transcription of mRNA molecules that generate functional proteins required for mobilization by LINE-1 elements or non-coding Alu and SVA RNAs that hijack L1 retrotransposition machinery (83,108). After translation, a ribonucleoprotein (RNP) complex composed of the RNA and proteins is formed (70). One of the retroelement-produced proteins, ORF2, encodes an endonuclease that introduces a nick in the AT-rich target site within genomic DNA (37,53). The next step is proposed to involve base-pairing of the retroelement mRNA polyA tail with the genomic DNA (109)

that serves as a primer for cDNA generation by the retroelement-encoded reverse transcriptase (RT) (71), a process known as target-primed reverse transcription (TPRT) (28,66). Despite the majority of the steps beyond cDNA synthesis and the proteins critical for their completion being unknown, the structure of the final *de novo* integration product is well characterized. It is usually flanked by target site duplications (TSD) and has a polyA-stretch present upstream of the 3' TSD, both of which are hallmarks of retrotransposition (53). One of the most significant differences between the “cut-and-paste” and “copy-and-paste” methods of transposition is that the latter almost always results in an increase in copy number, while the former only rarely does so. Additionally, a single active retroelement can give rise to a substantial number of offspring elements to form a subfamily of retrotransposons (102), leading to a non-linear mode of copy accumulation.

3. INSERTIONAL MUTAGENESIS

Even though new integration events are largely constrained to the sites recognizable by the L1-encoded endonuclease (the consensus for which is 5'-TTAAAA-3', although permutations are readily tolerated (37)), due to the promiscuity of the enzyme and millions of sites available within any given genome, the distribution of *de novo* integration events is fairly indiscriminate (39). The randomness of the integration process occasionally introduces L1, Alu, and SVA insertions into the coding regions of genes leading to mutations via insertional mutagenesis (reviewed in (9)). Even though all three currently active human retrotransposons rely on the activity of the L1-produced proteins for their integration (32,83,88), the frequencies of insertional mutagenesis estimated for L1, Alu, and SVA vary drastically between these elements (26,54,122). To date, Alu elements are responsible for over two-thirds of the TE integration-induced germ-line human diseases (over 43 reported cases compared to 16 by L1 and 4 by SVA) (reviewed in (9)). Multiple lines of evidence provide valid reasons for explaining the existing discrepancy in the disease-causing rate among the active human retrotransposons. Among the likely reasons are a significant difference in the time requirement for the completion of the retrotransposition process between Alu and L1 elements reported in tissue culture (57) and possible excess of L1 ORF2 for Alu mobilization produced by spliced L1 mRNA generated by the L1 loci maintaining functional ORF2 protein (8,11). While it is important to account for the L1, Alu, and SVA mutagenesis separately, it is also equally as important to keep in mind that L1 is the driving force of Alu and SVA mobilization. Human diseases caused by L1, Alu, and SVA insertional mutagenesis range from hemophilia and X-linked Duchenne muscular dystrophy to cystic fibrosis and breast cancer (reviewed in (9)), supporting the random nature of the integration process. It is worth noting, however, the dominating presence of the X chromosome-linked diseases (29 out of 63) in the reported group of retroelement integration-associated human illnesses (reviewed in (9)). This enrichment is likely due to the ascertainment bias associated with the haploid state of the mutations in X-linked genes in male carriers.

4. POSTINSERTIONAL INTERFERENCE WITH GENE EXPRESSION BY TEs

While *de novo* integration events generated in tissue culture and in the transgenic L1 mouse model are relatively randomly dispersed throughout the genome (3,6,39,42,89), the portrait of TE distribution in the human genome is far from being random (58,77). These findings suggest that evolutionary forces acting upon TE integration events postinsertionally, rather than the biology of their integration process, have a profound effect on the TEs distribution profiles in the genome. L1 elements, particularly full-length L1s inserted into introns in the forward orientation, are poorly tolerated (14,20) and as a result they are significantly underrepresented not only within genes, but also in the 5 kb regions flanking human gene boundaries (58,77). In contrast, Alu elements are considerably enriched within human gene

boundaries (58) sometimes representing over 40% of the gene's sequence as is in the case of the BRCA1 gene (104). The specific evolutionary forces promoting this inverted distribution are not very well known. However, the observed bias against L1 elements within genes is likely to be a result of the selective loss of L1-containing alleles that cause embryonic lethality and/or reduced fitness of their carriers. The apparent enrichment of Alu elements within genes potentially reflects their less devastating effect on normal gene expression following integration into introns.

Insertions of transposable elements within intronic sequences can interfere with normal gene expression through the introduction of functional (i) promoters and their regulatory elements, (ii) polyadenylation (pA) signals, and (iii) splice donor (SD) and acceptor (SA) sites. Besides the effect of TEs on the expression or function of a single gene through direct insertional interference, some TE integration events can also alter gene or cellular pathway function through indirect mechanisms such as regulation of miRNA expression. The following sections will provide a detailed discussion of the specific differences between transposable elements that likely contribute to the indirect effects they have on human gene expression after the integration process is completed.

4.1. LINE-1 elements

Human L1 elements are about 6 kb long (Figure 1), which is remarkably small compared to the size of the majority of human genes that can be as long as hundreds of kilobases. Yet, their functional structure includes most of the features found in generally larger human genes such as a promoter, 5' and 3' UTRs, open reading frames, and cis-acting signals for mRNA processing. L1 uses a polymerase II (pol II) promoter (sense promoter) located within the L1 5' UTR to drive expression of a bicistronic mRNA that encodes two open reading frames that are absolutely required for L1 retrotransposition (108) (Figure 2). The antisense L1 promoter (anti), also present within the 5' UTR, is demonstrated to drive expression of sequences located upstream of the L1 elements (86,106). The biological significance of the antisense promoter is not well established. One of the compelling hypotheses is that its role is to interfere with the transcription initiated within upstream sequences to secure transcription from the sense L1 promoter. Alternatively, the L1 antisense promoter is implicated in the production of small interfering RNAs that limit L1 expression (123). Both of these promoters can modify the normal gene expression. Independent of the orientation of the L1 insert (forward or reverse relative to gene expression) they have the potential for "gene breaking" by generating 5'-truncated genomic transcripts (121) (Figure 3).

L1 promoter activity is heavily regulated by epigenetic modifications (46,80). The short- and long-term consequences of L1 integration (particularly the full-length elements) within or in the vicinity of genes on the epigenetic state and chromatin signature of the gene are not known. Some of the hypotheses dealing with potential contribution of TEs to the epigenetic regulation of the mammalian genome were recently reviewed (49). L1 elements have been proposed to potentially influence the selective expression of monoallelically-expressed genes due to the enrichment of evolutionarily more recent LINE-1 elements in the regions surrounding these genes in human and mouse (2). Furthermore, L1 promoters contain binding sites for various transcription factors and regulatory proteins that can alter gene expression in response to various stimuli (45,81,82,124). L1 sequences can exert their influence on host gene expression by altering promoter specificity or strength (7,59,84,99,107). Not all TE interference is deleterious however, TE modification of gene expression may be responsible for genetic differences among species that translate into species-specific patterns of gene expression. The presence of L1 sequences in the vicinity of one of the three alternative promoters in the human flavin-containing mono-oxygenase 1 (FMO1) gene is reported to be responsible for the species-dependent, tissue-specific

difference in the expression of FMO1 gene in humans relative to the expression pattern of the same gene in mice (99). Another example of TE sequences influencing species-specific variation in the expression pattern of mammalian genes comes from the homology comparison of the composition of repetitive elements inserted between the regulatory sequences and the promoter of the human and murine type X collagen gene (7).

The existence of the sense and antisense promoters within the L1 5' UTR is not the only feature of these elements that makes them an unwelcome addition to genomic regions. Similar to their retroviral relatives, L1 sequences carry numerous functional polyadenylation (pA) and splice sites (8,10,90). These pA and splice sites are dispersed throughout the L1 sequence suggesting that all intronic copies of these elements, including the 5'-truncated ones, are potentially capable of interfering with normal gene expression. Active polyadenylation sites are particularly abundant within the L1 ORF sequences perhaps due to their unusually high AT content (90) and several functional pA sites are also present in the antisense strand of L1 elements (43). L1-encoded pA sites are effectively used during L1 transcription to attenuate the production of the full-length retrotranspositionally competent L1 mRNA (90), but they can also interfere with normal gene expression (43,121) (Figure 2). Even the shortest L1 insertions contain polyA sites at the end of their 3' UTRs that can be recognized during transcription (12).

Much like the polyA sites, functional splice sites present within the L1 sequence are used to generate processed L1 mRNAs (8) (Figure 2). In fact, splicing of the L1 transcripts leads to the production of retrotranspositionally incompetent L1 mRNAs and serves as one of the multiple mechanisms that significantly reduce L1-associated damage through limiting the production of full-length L1 mRNA. It is not clear whether processed L1-related RNAs have any biological function in the L1 life-cycle. However, one of the L1 splice products has a potential to produce functional L1 ORF2 protein (11), which has significant implications for Alu mobilization. There appears to be a considerable variation in the efficiency of L1 splice site usage among human tissues (11) suggesting that a tissue-specific repertoire of splicing factors may influence recognition of the L1-encoded splice sites.

The experimental evidence for L1's ability to donate its pA and splice sites has been collected through tissue culture and bioinformatic approaches with a strong agreement between the data from different sources. Full-length (i.e. containing more polyA and splice sites), polymorphic (i.e. likely maintaining functional sense and antisense promoters) L1 insertions are reported to specifically decrease the expression of primary transcripts in human cell lines from the alleles containing these integration events relative to the corresponding alleles without L1 (112). The presence of the full-length mouse L1 in an intron of a mini-gene reporter system convincingly demonstrated the debilitating effect of the full-length L1 inserted in the forward orientation on normal gene expression (up to 10-fold reduction) (20). This observation held in a variety of sequence contexts. In the same experimental system, mouse L1 insertion in the reverse orientation did not significantly impact transcript production. In addition to splice and polyA sites, the unusual AT richness of the L1 coding region is also proposed to interfere with the processing capacity of RNA polymerase II during the transcription of L1 sequences (43).

In addition to the reduction of normally processed mRNA, the presence of functional promoters, splice and pA sites within the L1 sequence (Figure 2) creates multiple opportunities for effective gene breaking. This may occur through termination of cellular transcripts at L1 pA sites and transcription initiation from the L1-encoded promoters to generate 5' truncated transcripts (121). Furthermore, some of the L1 SD and SA sites are located within the L1 5' UTR and are used in combination with genomic sequences during transcription driven by either sense or antisense L1 promoters to produce hybrid L1/cellular

transcripts (8,86). A combination of splice and polyadenylation sites located within the L1 sequence can result in the inclusion of sequences from the 5' truncated L1 loci into cellular mRNAs. For example, a portion of L1 containing a stop codon and a pA site is incorporated as a 3' exon into cellular mRNA to generate a soluble form of human attractin gene (110). Skipping of the L1-containing exon produces a mRNA for the membrane-bound form of the attractin protein. Finally, L1 integration can also block normal splicing by separating functional elements required for the proper splicing event to occur (20).

L1 interference with normal gene expression through splicing and polyadenylation is likely to vary among tissues, developmental stages, and disease states because both of these processes are altered in a development- and tissue-specific manner (63,115). In fact, attenuation of L1 expression by post-transcriptional processing differs significantly among normal human tissues as well as in cancer cell lines (11) strongly indicating that L1-encoded *cis*-acting RNA processing elements respond to tissue-specific environments. Furthermore, both sense- and antisense L1 promoters are reported to exhibit tissue-specificity (73,124). While no biological significance has been reported to date for the majority of the known L1/host gene chimeric mRNAs, cancer-specific L1-driven hybrid transcripts were identified in breast and colon cancer cell lines (29). In addition, the expression of fusion transcripts between L1 and the proto-oncogene cMet has been induced with demethylation agents in colon cancer and leukemia cell lines (120). Thus, L1 *cis*-acting regulatory elements can significantly alter normal gene expression by providing numerous possibilities for the generation of alternatively processed cellular mRNAs, some of which may have potential biological implications.

4.2. Alu elements

Alu elements have accumulated in the human genome to over 1,000,000 copies (58) and are one of the most successful retrotransposons in any genome. Their enrichment within coding regions of human genes (58) provide ample opportunities for these elements to change normal gene expression long after the integration event is completed.

There are a number of ways Alu elements can interfere with or modify normal gene expression. Alu transcription is driven by RNA polymerase III (polIII) and, along with other SINEs, Alu RNA contains no open reading frames. There are no reported examples of direct interference by Alu elements with normal human gene expression through the transcriptional activity of the Alu promoter. However, deamination of CpG islands within Alu sequences inserted near promoters of human genes can generate active transcription binding sites and is known to skew the programmed regulation of native promoters (125). Genome-wide bioinformatic assessment demonstrated that the highest Alu sequence densities are observed within 16kb upstream or downstream of transcript start positions, but are depleted within the 1kb region adjacent to human gene promoters (111), suggesting that such close proximity of Alus to gene promoters may interfere with their normal function. Nevertheless, expression of over 1,000 human genes can be potentially controlled by CpGs that originated from Alus (87). Alu elements have been reported to be able to drive expression of microRNAs (miRNAs) (15), which have various effects on the expression of their target genes. Bioinformatic analysis suggests that many miRNA genes are derived from all four major classes of transposable elements (LINEs, SINEs, LTR containing elements and DNA transposons). Fifty-five experimentally validated human miRNA sequences were found to be derived from TEs, and additional 85 novel miRNA genes are predicted to be originated from TEs (93). Together, these TE-derived miRNAs have the potential to provide a significant network of regulation for multiple human genes.

In contrast to L1 elements, Alu sequences do not contain inherent RNA polIII transcription processing sites such as polyadenylation or splice sites. Thus, *de novo* Alu integrations

within the introns of human genes do not immediately result in the production of alternatively processed transcripts due to utilization of Alu-encoded *cis*-acting sequences. However, as these elements accumulate random mutations, they have a tendency to accrue both polyadenylation and splice sites that can effectively interfere with gene expression (Figure 4). Bioinformatic analysis of the origin of polyA sites within human genes demonstrated that a large proportion of alternatively used pA sites can be traced to Alu elements. Out of the 1.1 million Alu copies in the human genome, 10,000 are found in the 3'UTRs of genes, 107 of which are reported to contribute active polyadenylation sites with a very significant bias toward originating from Alus inserted in the forward orientation (19).

In addition to accumulation of functional polyadenylation sites, multiple examples of exonization of Alu-derived sequences into the human transcriptome have been reported (105,114). Bioinformatic analysis demonstrated that 5.2% of the alternatively spliced internal exons, but none of the constitutively spliced exons in the human transcriptome, had a significant BLAST hit to an Alu sequence (105). The inclusion of Alu sequences in the coding region of mature transcripts resulted in a frame-shift or a termination codon most of the time (84%) (105). It was also determined that there is a bias against the inclusion of the Alu-containing exons into the mature transcripts (105). Alu exonization appears to be influenced by the sequences within Alu itself, for example the deletion of the left arm of Alu shifts the inclusion of the Alu right arm from alternative to constitutive splicing (38). In addition, Alu splicing exhibits tissue- and species-specificity suggesting that evolutionary pressure acts to refine the process of Alu exonization (64).

Alu elements can also interfere with normal RNA splicing by triggering exon skipping (61) (Figure 4). This process heavily depends on the orientation and sequence homology of Alu elements within introns and relies on the base-pairing of Alu sequences that serve as substrates for RNA editing enzymes (61). Single intronic Alu elements inserted in the reverse orientation can also trigger exon skipping in the minigene system perhaps through a different mechanism (61). Bioinformatic analysis revealed that more Alus flank alternatively spliced exons than constitutively spliced ones (61). In addition to promoting exon-skipping through the formation of hairpin-like structures by inverted Alus within primary transcripts, similar double-strand RNAs can be formed by Alu elements often found within the 3'UTRs of human RNAs (21). At least 333 human genes are identified that contain Alu sequences that can form duplex structures in their 3'UTRs (22). Indeed, the presence of inverted Alu repeats is reported to interfere with the nuclear export of Alu-containing mRNAs resulting in the translational silencing of these transcripts through nuclear retention (22). Additionally, there are several cases of single *de novo* Alu insertions in exons that cause exon skipping due to altered splicing, for example in cases of Duchenne muscular dystrophy (4) and Dent's disease (24). It is likely that due to the multiple ways Alu elements can interfere with normal gene expression in a tissue- and tumor-specific manner, the overall effect of TEs on the human transcriptome, as demonstrated by a bioinformatic prediction analysis, is significantly higher than on its mouse relative (78).

4.3. SVA elements

SVA elements are a more recently formed retrotranspositionally active family of elements that have contributed to human disease through insertional mutagenesis (88,117). The relative novelty of these elements comes with little understanding of their basic biology. Much like Alu, SVA elements are thought to parasitize the L1 retrotransposition machinery based on the presence of the L1-mediated retrotransposition signature (TSD and polyA tail) in identified SVA loci (88). In drastic contrast to their LINE and SINE relatives, it appears that at least some SVA expression relies on acquisition of a functional promoter at the site of insertion (30). Based on this report, it is likely that SVA elements would not interfere with cellular gene expression by contributing promoter activity common to L1 and Alu elements.

However, SVA elements contain cryptic splice sites that are used to form hybrid transcripts with host genes and to assist with the expression and retrotransposition of these elements (30,44). Even though currently there are no direct examples of SVA interference with human gene expression leading to specific phenotypic changes, SVA's ability to contribute splice sites present within their genomes raises a possibility of postinsertional interference with normal gene expression.

4.4. HERVs

Although human endogenous retroviruses (HERVs) are no longer actively undergoing retrotransposition (51,58), their long terminal repeats (LTR) often still contain functional regulatory sequences which may serve a biological role, influencing gene expression in their host (reviewed in (76)). Research into the HERV-K family of human-specific endogenous retroviruses revealed that at least 50% of these LTRs retained promoter activity in normal and malignant germ line tissues (18). 5'-proviral LTRs displayed a two to five-fold higher promoter activity when compared to solitary and 3'-proviral LTRs, demonstrating the importance of the LTR status to promoter activity (18). Further study demonstrated that the LTR distance from genes was a key factor affecting promoter activity, as the relative content of promoter-active LTRs were higher in gene-rich regions compared to gene-poor loci (18).

HERV LTRs predominantly act as alternative promoters with a similar expression pattern to the native promoters, although there are exceptions in which the LTR promoter leads to new expression patterns (25,33–35,101). LTRs from the HERV-E family are used as alternative promoters for the endothelin B receptor, apolipoprotein CI and Opitz syndrome midline 1 genes (60,75). These are a few representative examples in which the LTRs only have a subtle effect on gene expression, because the native promoters are also driving gene expression, thus, transcription initiated at the alternative LTR promoter only makes a minor contribution to the overall mRNA pool (25).

In addition to providing alternative promoters, LTRs can confer tissue-specific expression of genes, such as in the examples of gasdermin B (*GSDMB*) and nitric oxide synthase 3 genes (*NOS3*) (25). An LTR element from the HERV-H family operates not only as the predominantly used promoter for *GSDMB*, but diversifies the tissue specificity from the stomach to a more widespread expression in various human tissues (100). Conversely, LTR-driven expression of the *NOS3* gene is restricted to the placenta whereas non-LTR expression of *NOS3* is widespread due to its presence in endothelial cells (25,50). Recent data have also demonstrated human-specific antisense regulation of gene expression due to HERVs. LTRs integrated into the introns of the sodium bicarbonate cotransporter and intraflagellar transport protein 172 generate antisense transcripts, leading to a down-regulation in the mRNA levels of the corresponding genes (41).

In contrast to most examples of gene expression regulated by HERVs, sperm adhesion molecule 1 (*SPAM1*) transcripts initiate within the polymerase coding region of an HERV, rather than within the LTR (33). Expression of *SPAM1* from this HERV-derived promoter is gender-specific and largely tissue-specific as well. It represents an "evolution of promoter from protein coding sequence". In addition to being a source of alternative promoters, HERVs are also known to provide primary polyadenylation signals (67) and, when included in the 5'UTRs of human transcripts, to alter mRNA translation efficiency (5,56).

HERV's effect on the host transcriptome is not limited to the direct interference with gene expression. Changes in cellular gene expression may also occur in response to transcription of the HERV envelope (ENV) gene. Recent studies have revealed a regulatory link between the expression of the HERV-W family ENV gene and the expression of schizophrenia-associated genes (48). ENV transcripts were detected in plasma collected from a significant

number of individuals with recent-onset schizophrenia, but not in those obtained from samples of unaffected individuals (48). Furthermore, overexpression of ENV in human glioma cells resulted in an up-regulation of several schizophrenia-associated genes including brain-derived neurotrophic factor, neurotrophic tyrosine kinase receptor type 2 and dopamine receptor 3 (48).

HERV expression has also been associated with a number of autoimmune disorders such as multiple sclerosis (92), psoriasis (13) and systemic lupus erythematosus (91). The immunopathogenic response to the envelope protein from the HERV-W family can activate a pro-inflammatory reaction and autoimmune cascade (92), which may be associated with multiple sclerosis (MS). In addition, expression of epitopes from the HERV-H and HERV-W family envelope proteins was significantly higher on B cells and monocytes from patients with active MS disease compared to stable MS or control individuals (16).

Expression of viral-like particles from the HERV-K family has recently been implicated in the progression of melanoma (98). Human melanoma cell lines transition from an adherent to a non-adherent growth phenotype representative of increased malignancy, which is accompanied by the activation of HERV-K expression and massive viral-like particle production (98). Gene knockdown of the HERV-K *pol* and *env* prevented the conversion of the cells from the adherent to the non-adherent phenotype, suggesting that HERV-K expression is critical in regulating this transition in melanoma progression (98). HERV loci, particularly the envelope genes, have been shown to be abundantly expressed in a variety of other cancers including ovarian (118), colon (62), and testicular (40) when compared to normal tissues. In a quantitative evaluation of the expression of different HERV family envelope genes in several tumor and adjacent normal tissues, high levels of ENV transcripts were detected in testis tumor tissues for HERV-K, liver and lung tumor tissues for HERV-H, and in colon and liver tumor tissues for HERV-P (1). Although the functional role of HERVs in these cancers have not been identified, the upregulation of HERVs associated with some cancers may be useful as a diagnostic tool, and suggests that HERV expression may contribute to tumorigenesis by altering host gene expression in cancer genes (1).

4.5. DNA transposons

The human genome contains about seven major classes of DNA transposons that appear to have become completely inactive for transposition (58). While DNA transposons no longer contribute to human disease through insertional mutagenesis, their transpositionally incompetent loci continue to influence human gene expression. It appears that some DNA transposons may have been domesticated shortly after their integration as long as 45 million years ago to provide an initially beneficial function to the host that later became a latent liability to the normal gene function. The insertion of the *PiggyBac* transposable element *PGBD3* into intron 5 of the Cockayne syndrome Group B (*CSB*) gene (85) is an interesting example of the potential for both outcomes. This fusion protein is as equally abundant as the original full-length CSB protein and evolutionarily highly conserved, suggesting the fusion protein has been selected for some beneficial function. It has been speculated that the CSB-transposase fusion protein assists in host genome defense by repressing transposition of *PGBD3* and non-autonomous MER85 elements (85), i.e. it is advantageous in the presence of functional *CSB*. However in cells lacking functional CSB, the fusion protein may be detrimental and contribute to Cockayne syndrome (CS) (85). Mutations in the CSB chromatin remodeling protein are linked to CS (23), a devastating form of progeria.

In addition to insertional mutagenesis, DNA transposons may also influence host gene expression or function by creating novel genes. Recent studies identified a repressor derived from the hobo-*Ac*-Tam3 transposase family involved in the regulation of muscle growth in pigs (17,69). Disruption of the binding site of this transposon-derived repressor by a single

nucleotide substitution in intron 3 of the insulin-like growth factor 2 (*IGF2*) gene results in increased muscle growth and reduced fat deposition in pigs (113). Despite the fact that this observation has not been confirmed in humans, data show this repressor is specific to placental mammals and it is implicated as an important transcription factor in other biological functions affecting cell development, proliferation and growth (69). Another example is *SETMAR*, a new primate chimeric gene that originated from the fusion of a SET histone methyltransferase gene to the transposase gene of an *Hsmar1* transposon via *de novo* exonization (27).

The above described summary of TE influence on normal gene expression should be viewed as a complex process of TE interplay with the regulatory sequences present within genes as well as signals brought by other TEs. Introns of human genes usually harbor a mixture of sequences of different origins, thus, the potential effects of individual TE insertions should not be considered without taking into account the composition of the immediate genomic landscape.

5. PERSPECTIVE

To the detriment or benefit of the host, continuing accumulation of TE sequences within genomes provides an array of opportunities for their influence on the normal biological functions of host genes. While the mutagenic burden from these elements associated with their insertional mutagenesis has been long recognized, the considerable influence imposed by TE buildup on gene expression has only recently become a subject of genome-wide studies. A growing number of individual examples reflecting the spectrum of TE influences on mammalian transcriptomes combined with accumulation of large-scale studies strongly support the notion that these elements are not simple bystanders of the evolutionary course, but rather, by their mere presence, are involuntary participants in the processes that mount biological consequences. Even though TE influence on gene expression can be rather dramatic, these elements may also provide regulatory sequences conducive to fine-tuning of cellular expression profiles ensuring genome plasticity potentially important for adaptation or survival.

Acknowledgments

We thank Dr. Prescott Deininger, Dr. Cecily Bennett, and Vincent Strevi for critical comments. This publication was made possible by Grants Number P20RR020152, NIA 5K01AG030074-02 from the National Institutes of Health (NIH) and The Ellison Medical Foundation New Scholar in Aging award 547305G1 to VPB.

References

1. Ahn K, Kim HS. Structural and quantitative expression analyses of HERV gene family in human tissues. *Mol Cells*. 2009; 28:99–103. [PubMed: 19669627]
2. Allen E, Horvath S, Tong F, Kraft P, Spiteri E, Riggs AD, Marahrens Y. High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:9940–9945. [PubMed: 12909712]
3. An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD. Active retrotransposition by a synthetic L1 element in mice. *Proc Natl Acad Sci USA*. 2006; 103:18662–18667. [PubMed: 17124176]
4. Awano H, Malueka RG, Yagi M, Okizuka Y, Takeshima Y, Matsuo M. Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA. *J Hum Genet*. 2010; 55:785–790. [PubMed: 20827276]
5. Baban S, Freeman JD, Mager DL. Transcripts from a novel human KRAB zinc finger gene contain spliced Alu and endogenous retroviral segments. *Genomics*. 1996; 33:463–472. [PubMed: 8661005]

6. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr. L1 integration in a transgenic mouse model. *Genome Res.* 2006; 16:240–250. [PubMed: 16365384]
7. Beier F, Lammi MJ, Bertling W, von der MK. Transcriptional regulation of the human type X collagen gene expression. *Ann NY Acad Sci.* 1996; 785:209–211. [PubMed: 8702131]
8. Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.* 2006; 34:1512–1521. [PubMed: 16554555]
9. Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 2008; 18:343–358. [PubMed: 18256243]
10. Belancio VP, Roy-Engel AM, Deininger P. The impact of multiple splice sites in human L1 elements. *Gene.* 2008; 411:38–45. [PubMed: 18261861]
11. Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* 2010; 38:3909–3922. [PubMed: 20215437]
12. Belancio VP, Whelton M, Deininger P. Requirements for polyadenylation at the 3' end of LINE-1 elements. *Gene.* 2007; 390:98–107. [PubMed: 17023124]
13. Bessis D, Moles JP, Basset-Seguín N, Tesniere A, Arpin C, Guilhou JJ. Differential expression of a human endogenous retrovirus E transmembrane envelope glycoprotein in normal, psoriatic and atopic dermatitis human skin. *Br J Dermatol.* 2004; 151:737–745. [PubMed: 15491412]
14. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci USA.* 2006; 103:9590–9594. [PubMed: 16766655]
15. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology.* 2006; 13:1097–1101.
16. Brudek T, Christensen T, Aagaard L, Petersen T, Hansen HJ, Moller-Larsen A. B cells and monocytes from patients with active multiple sclerosis exhibit increased surface expression of both HERV-H Env and HERV-W Env, accompanied by increased seroreactivity. *Retrovirology.* 2009; 6:104. [PubMed: 19917105]
17. Butter F, Kappei D, Buchholz F, Vermeulen M, Mann M. A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. *EMBO Rep.* 2010
18. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E. At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. *J Virol.* 2006; 80:10752–10762. [PubMed: 17041225]
19. Chen C, Ara T, Gautheret D. Using Alu elements as polyadenylation sites: A case of retroposon exaptation. *Mol Biol Evol.* 2009; 26:327–334. [PubMed: 18984903]
20. Chen J, Rattner A, Nathans J. Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements. *Hum Mol Genet.* 2006; 15:2146–2156. [PubMed: 16723373]
21. Chen LL, Carmichael GG. Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle.* 2008; 7:3294–3301. [PubMed: 18948735]
22. Chen LL, DeCervo JN, Carmichael GG. Alu element-mediated gene silencing. *EMBO J.* 2008; 27:1694–1705. [PubMed: 18497743]
23. Christiansen M, Stevnsner T, Modin C, Martensen PM, Brosh RM Jr, Bohr VA. Functional consequences of mutations in the conserved SF2 motifs and post-translational phosphorylation of the CSB protein. *Nucleic Acids Res.* 2003; 31:963–973. [PubMed: 12560492]
24. Claverie-Martin F, Flores C, Anton-Gamero M, Gonzalez-Acosta H, Garcia-Nieto V. The Alu insertion in the CLCN5 gene of a patient with Dent's disease leads to exon 11 skipping. *J Hum Genet.* 2005; 50:370–374. [PubMed: 16041495]
25. Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene.* 2009; 448:105–114. [PubMed: 19577618]
26. Cordaux R, Hedges DJ, Herke SW, Batzer MA. Estimating the retrotransposition rate of human Alu elements. *Gene.* 2006; 373:134–137. [PubMed: 16522357]
27. Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci USA.* 2006; 103:8101–8106. [PubMed: 16672366]

28. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 2002; 21:5899–5910. [PubMed: 12411507]
29. Cruickshanks HA, Tufarelli C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. *Genomics.* 2009; 94:397–406. [PubMed: 19720139]
30. Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, Batzer MA, Lower R, Schumann GG. 5′-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res.* 2009; 19:1992–2008. [PubMed: 19652014]
31. Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J Mol Biol.* 1981; 151:17–33. [PubMed: 6276559]
32. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet.* 2003; 35:41–48. [PubMed: 12897783]
33. Dunn CA, Mager DL. Transcription of the human and rodent SPAM1/PH-20 genes initiates within an ancient endogenous retrovirus. *BMC Genomics.* 2005; 6:47. [PubMed: 15804358]
34. Dunn CA, Romanish MT, Gutierrez LE, van de Lagemaat LN, Mager DL. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene.* 2006; 366:335–342. [PubMed: 16288839]
35. Dunn CA, van de Lagemaat LN, Baillie GJ, Mager DL. Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3GAL-T5. *Gene.* 2005; 364:2–12. [PubMed: 16112824]
36. Fanning TG, Singer MF. LINE-1: a mammalian transposable element. *Biochim Biophys Acta.* 1987; 910:203–212. [PubMed: 2445384]
37. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell.* 1996; 87:905–916. [PubMed: 8945517]
38. Gal-Mark N, Schwartz S, Ast G. Alternative splicing of Alu exons--two arms are better than one. *Nucleic Acids Res.* 2008; 36:2012–2023. [PubMed: 18276646]
39. Gilbert N, Lutz-Prigge S, Moran JV. Genomic deletions created upon LINE-1 retrotransposition. *Cell.* 2002; 110:315–325. [PubMed: 12176319]
40. Gimenez J, Montgiraud C, Pichon JP, Bonnaud B, Arzac M, Ruel K, Bouton O, Mallet F. Custom human endogenous retroviruses dedicated microarray identifies self-induced HERV-W family elements reactivated in testicular cancer upon methylation control. *Nucleic Acids Res.* 2010
41. Gogvadze E, Stukacheva E, Buzdin A, Sverdlov E. Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions. *J Virol.* 2009; 83:6098–6105. [PubMed: 19339349]
42. Graham T, Boissinot S. The genomic distribution of L1 elements: the role of insertion bias and natural selection. *J Biomed Biotechnol.* 2006; 2006:75327–75332. [PubMed: 16877820]
43. Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature.* 2004; 429:268–274. [PubMed: 15152245]
44. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res.* 2009; 19:1983–1991. [PubMed: 19635844]
45. Harris CR, Dewan A, Zupnick A, Normart R, Gabriel A, Prives C, Levine AJ, Hoh J. p53 responsive elements in human retrotransposons. *Oncogene.* 2009; 28:3857–3865. [PubMed: 19718052]
46. Hata K, Sakaki Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene.* 1997; 189:227–234. [PubMed: 9168132]
47. Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A, Jonasdóttir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthíasdóttir S, Jónsdóttir T, Palsson S, Einarsson H, Gunnarsdóttir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdóttir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science.* 2007; 316:1491–1493. [PubMed: 17478679]

48. Huang W, Li S, Hu Y, Yu H, Luo F, Zhang Q, Zhu F. Implication of the env Gene of the Human Endogenous Retrovirus W Family in the Expression of BDNF and DRD3 and Development of Recent-Onset Schizophrenia. *Schizophr Bull.* 2010
49. Huda A, Jordan IK. Epigenetic regulation of Mammalian genomes by transposable elements. *Ann NY Acad Sci.* 2009; 1178:276–284. [PubMed: 19845643]
50. Huh JW, Ha HS, Kim DS, Kim HS. Placenta-restricted expression of LTR-derived NOS3. *Placenta.* 2008; 29:602–608. [PubMed: 18474398]
51. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet.* 2008; 42:709–732. [PubMed: 18694346]
52. Jern P, Sperber GO, Ahlsen G, Blomberg J. Sequence variability, gene structure, and expression of full-length human endogenous retrovirus H. *J Virol.* 2005; 79:6325–6337. [PubMed: 15858016]
53. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proceedings of the National Academy of Sciences of the United States of America.* 1997; 94:1872–1877. [PubMed: 9050872]
54. Kazazian HH. An estimated frequency of endogenous insertional mutations in humans. *Nature Genetics.* 1999; 22:130. [PubMed: 10369250]
55. Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Hemophilia-A Resulting from Denovo Insertion of L1 Sequences Represents A Novel Mechanism for Mutation in Man. *Nature.* 1988; 332:164–166. [PubMed: 2831458]
56. Kowalski PE, Mager DL. A human endogenous retrovirus suppresses translation of an associated fusion transcript, PLA2L. *J Virol.* 1998; 72:6164–6168. [PubMed: 9621083]
57. Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM. The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet.* 2009; 5:e1000458. [PubMed: 19390602]
58. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409:860–921. [PubMed: 11237011]
59. Landry JR, Medstrand P, Mager DL. Repetitive elements in the 5' untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. *Genomics.* 2001; 76:110–116. [PubMed: 11549323]
60. Landry JR, Rouhi A, Medstrand P, Mager DL. The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter. *Mol Biol Evol.* 2002; 19:1934–1942. [PubMed: 12411602]
61. Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G. Intronic Alus influence alternative splicing. *PLoS Genet.* 2008; 4:e1000204. [PubMed: 18818740]
62. Liang Q, Ding J, Xu R, Xu Z, Zheng S. Identification of a novel human endogenous retrovirus and promoter activity of its 5' U3. *Biochem Biophys Res Commun.* 2009; 382:468–472. [PubMed: 19289097]
63. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet.* 2010; 11:75–87. [PubMed: 20019688]
64. Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL, Xing Y. Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet.* 2008; 4:e1000225. [PubMed: 18841251]

65. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ III, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshtsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolun O, Elong R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005; 438:803–819. [PubMed: 16341006]
66. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*. 1993; 72:595–605. [PubMed: 7679954]
67. Mager DL, Hunter DG, Schertzer M, Freeman JD. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). *Genomics*. 1999; 59:255–263. [PubMed: 10444326]
68. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet*. 2006; 2:e2. [PubMed: 16440055]
69. Markljung E, Jiang L, Jaffe JD, Mikkelsen TS, Wallerman O, Larhammar M, Zhang X, Wang L, Saenz-Vash V, Gnirke A, Lindroth AM, Barres R, Yan J, Stromberg S, De S, Ponten F, Lander ES, Carr SA, Zierath JR, Kullander K, Wadelius C, Lindblad-Toh K, Andersson G, Hjalmar G, Andersson L. ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol*. 2009; 7:e1000256. [PubMed: 20016685]
70. Martin SL. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol*. 1991; 11:4804–4807. [PubMed: 1715025]
71. Martin SL, Li J, Epperson LE, Lieberman B. Functional reverse transcriptases encoded by A-type mouse LINE-1: defining the minimal domain by deletion analysis. *Gene*. 1998; 215:69–75. [PubMed: 9666081]
72. Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science*. 1991; 254:1808–1810. [PubMed: 1722352]
73. Matlik K, Redik K, Speck M. L1 antisense promoter drives tissue-specific transcription of human genes. *J Biomed Biotechnol*. 2006; 2006:1–16.
74. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007; 316:1488–1491. [PubMed: 17478681]
75. Medstrand P, Landry JR, Mager DL. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J Biol Chem*. 2001; 276:1896–1903. [PubMed: 11054415]
76. Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL. Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res*. 2005; 110:342–352. [PubMed: 16093686]
77. Medstrand P, van de Lagemaat LN, Mager DL. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res*. 2002; 12:1483–1495. [PubMed: 12368240]
78. Mersch B, Sela N, Ast G, Suhai S, Hotz-Wagenblatt A. SERpredict: detection of tissue- or tumor-specific isoforms generated through exonization of transposable elements. *BMC Genet*. 2007; 8:78. [PubMed: 17986331]

79. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SM, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Grealley JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Graves JA, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature*. 2007; 447:167–177. [PubMed: 17495919]
80. Montoya-Durango DE, Liu Y, Teneng I, Kalbfleisch T, Lacy ME, Steffen MC, Ramos KS. Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. *Mutat Res*. 2009; 665:20–28. [PubMed: 19427507]
81. Morales JF, Snow ET, Murnane JP. Environmental factors affecting transcription of the human L1 retrotransposon. I. Steroid hormone-like agents. *Mutagenesis*. 2002; 17:193–200. [PubMed: 11971989]
82. Morales JF, Snow ET, Murnane JP. Environmental factors affecting transcription of the human L1 retrotransposon. II. Stressors. *Mutagenesis*. 2003; 18:151–158. [PubMed: 12621071]
83. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. High frequency retrotransposition in cultured mammalian cells. *Cell*. 1996; 87:917–927. [PubMed: 8945518]
84. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. 2005; 435:903–910. [PubMed: 15959507]
85. Newman JC, Bailey AD, Fan HY, Pavelitz T, Weiner AM. An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. *PLoS Genet*. 2008; 4:e1000031. [PubMed: 18369450]
86. Nigumann P, Redik K, Matlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics*. 2002; 79:628–634. [PubMed: 11991712]
87. Oei SL, Babich VS, Kazakov VI, Usmanova NM, Kropotov AV, Tomilin NV. Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics*. 2004; 83:873–882. [PubMed: 15081116]
88. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet*. 2003; 73:1444–1451. [PubMed: 14628287]
89. Ovchinnikov I, Troxel AB, Swergold GD. Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. *Genome Res*. 2001; 11:2050–2058. [PubMed: 11731495]
90. Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet*. 2003; 35:363–366. [PubMed: 14625551]
91. Perl A, Nagy G, Koncz A, Gergely P, Fernandez D, Doherty E, Telarico T, Bonilla E, Phillips PE. Molecular mimicry and immunomodulation by the HRES-1 endogenous retrovirus in SLE. *Autoimmunity*. 2008; 41:287–297. [PubMed: 18432409]
92. Perron H, Lang A. The Human Endogenous Retrovirus Link between Genes and Environment in Multiple Sclerosis and in Multifactorial Diseases Associating Neuroinflammation. *Clin Rev Allergy Immunol*. 2009
93. Piriyaopongsa J, Marino-Ramirez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. *Genetics*. 2007; 176:1323–1337. [PubMed: 17435244]
94. Pontius JU, Mullikin JC, Smith DR, Lindblad-Toh K, Gnerre S, Clamp M, Chang J, Stephens R, Neelam B, Volfovsky N, Schaffer AA, Agarwala R, Narfstrom K, Murphy WJ, Giger U, Roca AL, Antunes A, Menotti-Raymond M, Yuhki N, Pecon-Slattery J, Johnson WE, Bourque G, Tesler G, O'Brien SJ. Initial sequence and comparative analysis of the cat genome. *Genome Res*. 2007; 17:1675–1689. [PubMed: 17975172]
95. Robertson HM, Zumpano KL. Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. *Gene*. 1997; 205:203–217. [PubMed: 9461395]

96. Schmid CW, Deininger PL. Sequence organization of the human genome. *Cell*. 1975; 6:345–358. [PubMed: 1052772]
97. Schmid CW, Jelinek WR. The Alu family of dispersed repetitive sequences. *Science*. 1982; 216:1065–1070. [PubMed: 6281889]
98. Serafino A, Balestrieri E, Pierimarchi P, Matteucci C, Moroni G, Oricchio E, Rasi G, Mastino A, Spadafora C, Garaci E, Vallebona PS. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Exp Cell Res*. 2009; 315:849–862. [PubMed: 19167380]
99. Shephard EA, Chandan P, Stevanovic-Walker M, Edwards M, Phillips IR. Alternative promoters and repetitive DNA elements define the species-dependent tissue-specific expression of the FMO1 genes of human and mouse. *Biochem J*. 2007; 406:491–499. [PubMed: 17547558]
100. Sin HS, Huh JW, Kim DS, Kang DW, Min DS, Kim TH, Ha HS, Kim HH, Lee SY, Kim HS. Transcriptional control of the HERV-H LTR element of the GSDML gene in human tissues and cancer cells. *Arch Virol*. 2006; 151:1985–1994. [PubMed: 16625320]
101. Sin HS, Huh JW, Kim DS, Kim TH, Ha HS, Kim WY, Park HK, Kim CM, Kim HS. Endogenous retrovirus-related sequences provide an alternative transcript of MCJ genes in human tissues and cancer cells. *Genes Genet Syst*. 2006; 81:333–339. [PubMed: 17159294]
102. Slagel V, Flemington E, Traina-Dorge V, Bradshaw H, Deininger P. Clustering and subfamily relationships of the Alu family in the human genome. *Mol Biol Evol*. 1987; 4:19–29. [PubMed: 3128713]
103. Smit AFA, Riggs AD. Tiggers and other DNA transposon fossils in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93:1443–1448. [PubMed: 8643651]
104. Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, King MC. Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. *Genome Res*. 1996; 6:1029–1049. [PubMed: 8938427]
105. Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. *Genome Res*. 2002; 12:1060–1067. [PubMed: 12097342]
106. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Molecular and Cellular Biology*. 2001; 21:1973–1985. [PubMed: 11238933]
107. Steel G, Lutz EM. Characterisation of the mouse vasoactive intestinal peptide receptor type 2 gene, *Vipr2*, and identification of a polymorphic LINE-1-like sequence that confers altered promoter activity. *J Neuroendocrinol*. 2007; 19:14–25. [PubMed: 17184482]
108. Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol*. 1990; 10:6718–6729. [PubMed: 1701022]
109. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell*. 2002; 110:327–338. [PubMed: 12176320]
110. Tang W, Gunn TM, McLaughlin DF, Barsh GS, Schlossman SF, Duke-Cohan JS. Secreted and membrane attractin result from alternative splicing of the human ATRN gene. *Proc Natl Acad Sci USA*. 2000; 97:6025–6030. [PubMed: 10811918]
111. Tsigos A, Rigoutsos I. Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput Biol*. 2009; 5:e1000610. [PubMed: 20019790]
112. Ustyugova SV, Lebedev YB, Sverdlov ED. Long L1 insertions in human gene introns specifically reduce the content of corresponding primary transcripts. *Genetica*. 2006; 128:261–272. [PubMed: 17028956]
113. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*. 2003; 425:832–836. [PubMed: 14574411]
114. Vervoort R, Gitzelmann R, Lissens W, Liebaers I. A mutation (IVS8+0.6kbpdelTC) creating a new donor splice site activates a cryptic exon in an Alu-element in intron 8 of the human beta-glucuronidase gene. *Hum Genet*. 1998; 103:686–693. [PubMed: 9921904]

115. Walker WH, Delfino FJ, Habener JF. RNA processing and the control of spermatogenesis. *Front Horm Res.* 1999; 25:34–58. [PubMed: 10941401]
116. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM. LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene.* 2008; 419:1–6. [PubMed: 18534786]
117. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA elements: a hominid-specific retroposon family. *J Mol Biol.* 2005; 354:994–1007. [PubMed: 16288912]
118. Wang-Johanning F, Liu J, Rycaj K, Huang M, Tsai K, Rosen DG, Chen DT, Lu DW, Barnhart KF, Johanning GL. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. *Int J Cancer.* 2007; 120:81–90. [PubMed: 17013901]
119. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermizakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyas E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002; 420:520–562. [PubMed: 12466850]
120. Weber B, Kimhi S, Howard G, Eden A, Lyko F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. *Oncogene.* 2010; 29:5775–5784. [PubMed: 20562909]
121. Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* 2005; 15:1073–1078. [PubMed: 16024818]
122. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, Jorde LB. Mobile elements create structural variation. Analysis of a complete human genome. *Genome Res.* 2009
123. Yang N, Kazazian HH Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol.* 2006; 13:763–771. [PubMed: 16936727]
124. Yang N, Zhang L, Zhang Y, Kazazian HH. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Research.* 2003; 31:4929–4940. [PubMed: 12907736]
125. Zemojtel T, Kielbasa SM, Arndt PF, Chung HR, Vingron M. Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. *Trends Genet.* 2009; 25:63–66. [PubMed: 19101055]
126. Zhao F, Qi J, Schuster SC. Tracking the past. Interspersed repeats in an extinct Afrotherian mammal, *Mammuthus primigenius*. *Genome Res.* 2009; 19:1384–1392. [PubMed: 19508981]

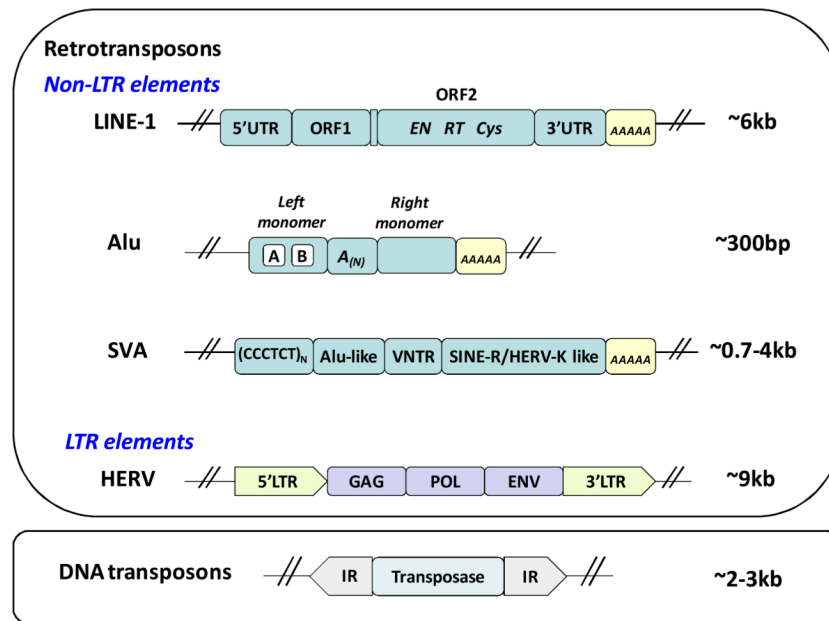


Figure 1. Schematic representation of TE organization

Transposable elements are divided into retro- and DNA transposons. Retroelements are further subdivided into Non-Long Terminal Repeat (Non-LTR) and LTR elements. Long Interspersed Element-1 (**LINE-1**) is an autonomous element that is about 6 kb long. It is composed of a 5' UTR (untranslated region) that contains sense- and antisense promoters, two open reading frames (ORF1 and 2) separated by an intergenic region, and a 3' UTR that ends in a polyadenylation signal followed by a stretch of adenosine residues, A-tail, (AAAAA yellow box) of variable length (36,106,108). ORF2 encodes three domains critical for L1 retrotransposition: endonuclease (EN), reverse transcriptase (RT), and cystein-rich domain (Cys) (37,72,83). Alu and SVA elements do not encode any ORFs and belong to a group of non-autonomous retroelements that parasitize on L1 retrotranspositional machinery. **Alus** are primate-specific elements that originated from the 7SL gene. Alu elements are about 300 bp long. They are composed of left and right monomers separated by an A-rich region (A(N)) (31,97). Like L1 elements, Alus end in an A-tail of variable length. **SVAs** identified in the human genome vary in length from 0.7 to 4 kb. SVAs are composed of a combination of sequences of different origin. They begin with a variable copy number of CCCTCT repeats followed by an Alu-like sequence (two antisense Alu fragments), a variable nucleotide repeat (VNTR), and a reverse SINE/HERV-K-like sequence (88,117). SVA elements contain an A-tail at their 3' end. Human endogenous retroviruses (**HERVs**) contain 5' and 3' LTRs that flank genes coding for the structural protein (GAG), polymerase (POL), and envelope protein (ENV) necessary for their mobilization (52). Most all of the HERV loci identified in the human genome harbor mutations incompatible with mobilization. Full-length HERVs are about 9 kb long. DNA transposons mobilize through a “cut-and-paste” mechanism. A functional DNA transposon locus encodes an enzyme transposase that is flanked by inverted repeats (IR) (95,103). DNA transposons vary from 2 to 3 kb in length.

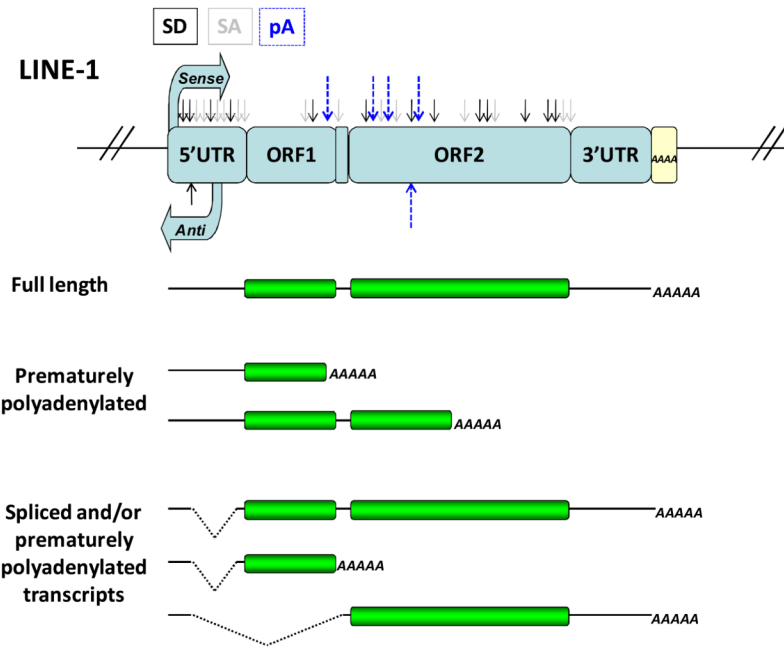


Figure 2. Cis-acting elements within the L1 sequence and their interference with gene expression
 Schematic representation of the major polyadenylation (pA), splice donor (SD) and acceptor (SA) sites within the L1 sequence and L1 mRNAs resulting from their usage (8,43,90). Solid black arrows depict SD sites, solid gray arrows depict SA sites and dashed blue arrows depict pA sites.

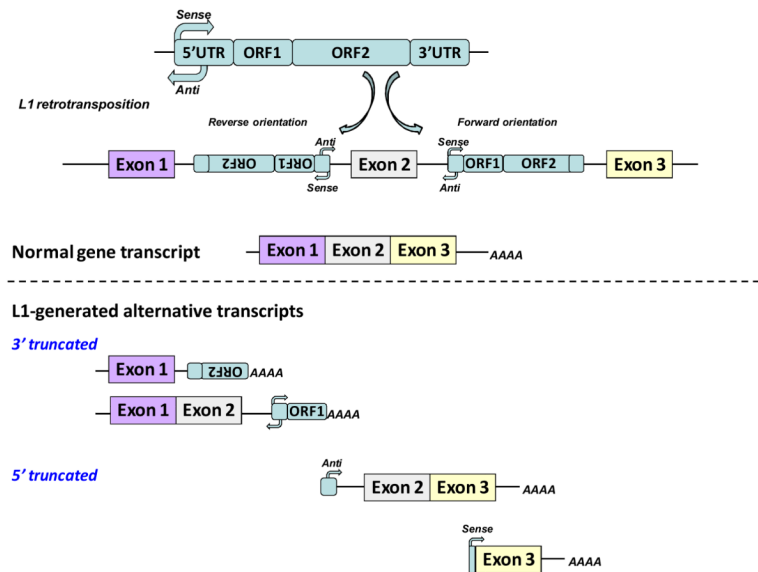


Figure 3. Examples of L1 interference with gene expression

L1 retrotransposition can land within the intronic regions of human genes. *De novo* L1 integration events can be positioned in the forward or reverse orientation relative to the transcription of the gene. Transcription of genes harboring L1 sequences within their introns can produce 3' and 5' truncated transcripts in addition to normal mRNA. 3' truncated mRNAs can arise from premature transcriptional termination at the L1-encoded polyadenylation sites present in both positive and negative strands of L1. 5' truncated transcripts can be initiated by the antisense promoter (anti) of the L1 in the reverse orientation or by the sense promoter (sense) of the L1 in the forward orientation. 5' truncated mRNAs generated by the antisense promoter can contain the entire sequence present between the L1 promoter and the intron as shown in this figure (top transcript of the two depicted 5' truncated mRNAs) (86,106). Alternatively, splice donor sites within sense or antisense L1 sequences can be used with the splice acceptor site of the adjacent exon generating a spliced hybrid L1 mRNA represented by the transcript depicted at the bottom of the figure.

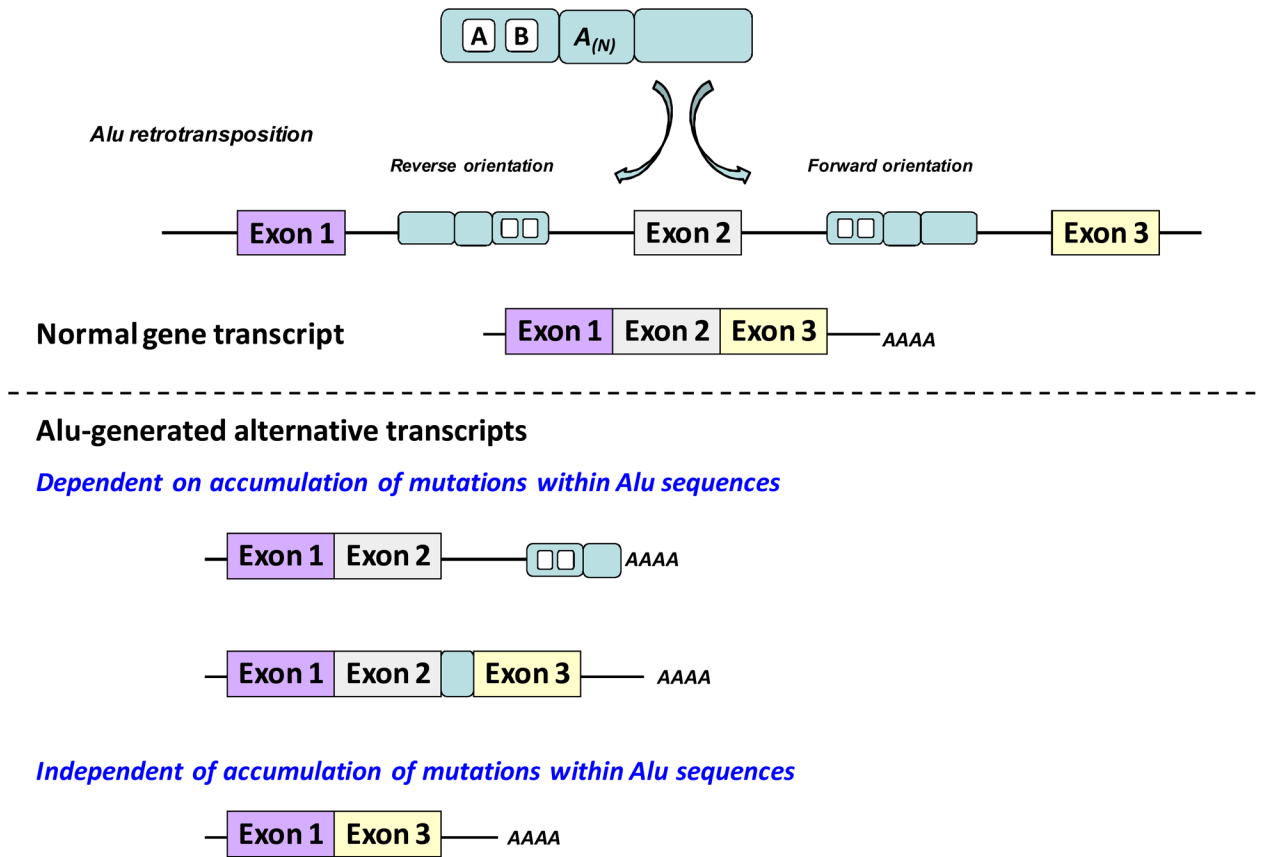


Figure 4. Schematic representation of Alu inserts interference with gene expression
 Alu sequences that are enriched within the intronic regions of human genes can be positioned in the forward or reverse orientation relative to the transcription of the gene. Alu elements can interfere with normal gene expression in a manner dependent or independent of the accumulation of mutations after integration. As Alu elements accumulate mutations they accrue functional polyadenylation usage of which generates 3' truncated mRNAs (top transcript in lower half of the figure) and splice sites that result in Alu exonization (middle transcript in lower half of the figure). Inverted Alus flanking an exon (shown in the top half of the figure) can lead to exon skipping (a transcript depicted in the bottom of the lower half of the figure).