

LINEs, SINEs and other retroelements: do birds of a feather flock together?

Astrid M. Roy-Engel¹

¹*Tulane University, Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane Cancer Center, SL-66
1430 Tulane Ave., New Orleans, LA 70112*

TABLE OF CONTENTS

1. Abstract
2. Introduction to the active non-LTR retroelements
3. Expression of non-LTR retroelements
 - 3.1. Expression of L1: the driving force
 - 3.2. Expression of Alu and other non-autonomous elements
4. Insertion mechanism
 - 4.1. Retrotransposition process and Target Primed Reverse Transcription (TPRT)
 - 4.2. Structural requirements
 - 4.3. Cellular regulatory mechanisms of retrotransposition
5. Activity and distribution non-LTR retroelements
 - 5.1. Activity profiles during evolution
 - 5.2. Evidence of current activity
 - 5.3. Insertional distribution
6. Summary and perspective
7. Acknowledgements
8. References

1. ABSTRACT

Mobile elements account for almost half of the mass of the human genome. Only the retroelements from the non-LTR (long terminal repeat) retrotransposon family, which include the LINE-1 (L1) and its non-autonomous partners, are currently active and contributing to new insertions. Although these elements seem to share the same basic amplification mechanism, the activity and success of the different types of retroelements varies. For example, Alu-induced mutagenesis is responsible for the majority of the documented instances of human disease induced by insertion of retroelements. Using copy number in mammals as an indicator, some SINEs have been vastly more successful than other retroelements, such as the retropseudogenes and even L1, likely due to differences in post-insertion selection and ability to overcome cellular controls. SINE and LINE integration can be differentially influenced by cellular factors, indicating some differences between in their amplification mechanisms. We focus on the known aspects of this group of retroelements and highlight their similarities and differences that may significantly influence their biological impact.

2. INTRODUCTION TO THE ACTIVE NON-LTR RETROELEMENTS

Completion of the human genome sequence revealed mobile element numbers to be higher than expected, making up nearly half the mass of the genome (1). Most of these sequences represent fossils of previously active elements. Mobile elements are classified into RNA (Class I) or DNA (Class II) transposons based on the type of intermediate used in their amplification mechanism. Class I RNA transposons, also commonly referred to as retroelements or retrotransposons, are further subdivided into two families based on the presence or absence of long terminal repeats (LTRs). These elements amplify through a process termed retrotransposition (2), during which an RNA intermediate is reverse transcribed to generate a new copy. Although LTR- and non-LTR retrotransposons both use an RNA intermediate, their mechanism of insertion varies greatly (reviewed in 3). Evidence of current mobile element activity in humans is mostly limited to the subgroup of non-LTR retrotransposons: Long Interspersed Element -1 (LINE-1 or L1) and its non-autonomous partners, such as SVA and Alu. Recent publications (4-6)

Differences and similarities of active retroelements

and data from the 1000 human genome project (7) demonstrate that the occurrence of new insertions is much higher than previously estimated dramatically contributing to diversity. Elements or inserts generated by an L1-driven retrotransposition process are identified by the characteristic hallmark features which include flanking target site duplications (TSD) of varying lengths, the usual presence of a 3' poly adenosine stretch or A-tail and its integration into an A-rich sequence (Figure 1).

L1 belongs to the family of non-LTR retrotransposons. There are about half a million L1 copies in the human genome. However, most L1 copies are 5' truncated (1, 8) or contain inactivating mutations so that only about 100 are thought to be retrotranspositionally active (9). The human L1 is approximately 6000 bp in length with two open reading frames, ORF1 and ORF2 (Figure 1). ORF1 codes for a 40 kDa RNA binding protein thought to function as a nucleic acid chaperone (10). The ORF2p is a 149 kDa multifunctional protein with endonuclease and reverse transcriptase capabilities. Both proteins are essential for L1 retrotransposition (11). The L1 encoded proteins are thought to bind to the RNA to form a cytoplasmic RNP (12-14). Because L1 codes for the proteins required for its amplification, it is referred to as an autonomous element. The L1 5'UTR contains an atypical internal RNA polymerase II (pol II) promoter that initiates transcription upstream of the promoter (15). This helps L1 preserve the ability to self-mobilize since the promoter sequence is present in the transcript and can be passed on to new inserts. Another unusual feature of the L1 5'UTR is the presence of an antisense promoter (ASP), which transcribes the upstream flanking sequence (16, 17).

Alu elements are classified as SINEs (Short Interspersed Elements), due to their short length (~300 bp) and their transcription by RNA polymerase III (pol III). Because Alu has no coding capacity, it depends on L1 to obtain the factors needed for retrotransposition and thus considered a non-autonomous element. With over 1 million copies in the human genome (1), Alu is one of the most successful SINE elements known. However, the vast majority of the Alu copies in the human genome are not retrotranspositionally active. Only a few Alu copies are likely to be active Alu source elements (18). Alu elements are ancestrally derived from the 7SL RNA gene (19), the nucleic acid component of the signal recognition particle (SRP) involved in protein secretion (20). The region of the 7SL gene sharing sequence homology to Alu is defined as the "Alu domain". An Alu is composed of two different GC-rich monomers separated by an adenine rich region and flanked at the 3' end by another polyA stretch usually referred to as an "A-tail" (Figure 1). Only the 5' monomer or "left monomer" of an Alu maintains the internal bipartite RNA polymerase III promoter composed of an A and B box (21). However, the Alu sequence does not contain an RNA pol III terminator (usually four thymidine residues). Thus, transcription of a typical Alu element will continue past the A-tail and into the genomic flank, resulting in Alu transcripts from various loci containing unique 3' sequences of different lengths (Figure 1).

SVA (SINE/VNTR/Alu) elements are a composite of several sequences (Figure 1). About 2700 SVA elements have been identified in the human genome (22). A "full-length" SVA consists of 5' tandem array of a hexameric sequence (CCCTCT), a segment with two partial Alu sequences in reverse orientation, followed by variable number tandem repeat (VNTR) segment, and a human endogenous retrovirus-K (HERV-K) like sequences, also referred to as SINE-R, at its 3' end. A promoter has yet to be identified within the SVA sequence. Because SVA displays the characteristic hallmarks of L1-mediated retrotransposition and predicted to have no coding capacity, this element is considered a non-autonomous member of the non-LTR retrotransposon family. However, due to the limited knowledge of its amplification mechanism, further classification of the SVA element as a SINE is still controversial. Not only is SVA uncommonly long (over 4kb) for a SINE which is typically 100-300 bp in length, this element is unlikely to be transcribed by RNA polymerase III due to the presence of multiple internal RNA pol III termination signals.

Retropseudogenes or processed pseudogenes arise from the reverse transcription of mRNA from a large variety of genes. Estimates for the number of retropseudogenes present in the human genome vary from about 8000 (23) up to 33,000 (24) depending on the parameters used for identification. Because they derive from processed transcripts, retropseudogenes typically do not contain introns and many contain a recognizable 3' polyA tail and flanking TSD (Figure 1). Some genes are more prone to retropseudogene formation, thus the majority of retropseudogene copies being derived from a few genes. Some of the retropseudogenes with higher copy numbers include the highly expressed housekeeping genes GAPDH, cyclophilin, cytochrome c and the ribosomal proteins, which contribute to 20% of the total retropseudogene copies (23). One study suggests a strong association between short length genes and higher numbers of retropseudogene copies (25).

Other retrotransposed sequences from non-coding genes can be found in the human genome attesting to the promiscuity of the L1 retrotransposition machinery. These retrotransposed sequences will be flanked by the hallmark TSD. There are multiple examples of RNA polymerase III-generated transcripts that have been retrotransposed. For example, the human genome contains several hundred retropseudogene copies of 7SL (26) and multiple examples of U6 copies. Interestingly, many of the retrotransposed U6 sequences are found as part of a chimera with a 5' truncated L1 sequence. This chimeric element is postulated to form *in vivo* by a "template switch" mechanism, where the ORF2p switches between the L1 RNA to the U6 transcript during reverse transcription (27). Another relatively successful family of RNA pol-III retrotransposition events is derived from the Y (hY) RNA genes associated with the Ro60 autoantigen with almost 1000 copies (28). Among the shorter retrotransposed sequences found in mammals are the "tailless" inserts derived from portions of tRNA or pre-tRNA sequences (29). Interestingly, retrotransposed copies from another mobile element, the endogenous retrovirus

Differences and similarities of active retroelements

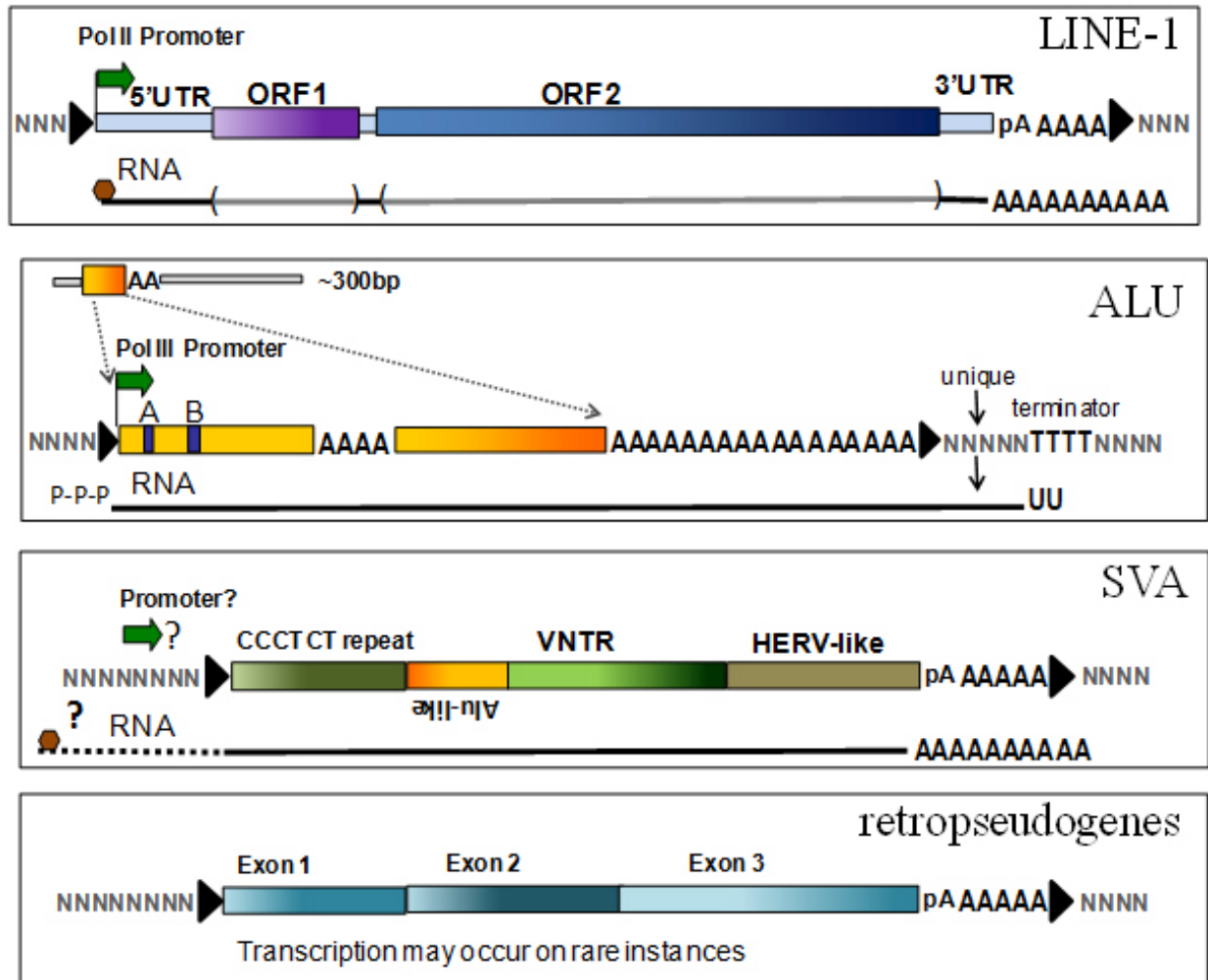


Figure 1. Schematic representation of the structural organization of active human retroelements and their transcripts. Alu, SVA and LINE-1 belong to the non-LTR group of currently active human retroelements. Each panel shows the schematic of the element with the expected transcript (represented as a line and labeled RNA) shown below it. The boundaries of a retroelement are defined by the presence of tandem site duplications (TSDs) (shown as black arrowheads) generated during the retrotransposition process and which are a hallmark of retrotransposed sequences. Ns represent the flanking genomic sequence. The transcription start sites are represented by a green arrow. LINE-1 (L1) elements are composed of a 5'UTR that contains an internal RNA polymerase II promoter, followed by two open reading frames (ORF1 and ORF2) and a 3'UTR containing a polyadenylation signal (pA) and an A-tail. Full-length L1 transcripts are polyadenylated and capped (represented as a brown hexagon). Alu elements are about 300 bp in length (shown at the top of the panel, with a larger version that includes structural details shown below). Alu elements are composed of two non-identical sequences (monomers shown in yellow) separated by an adenine-rich region and flanked at the 3' end by a longer A-rich region commonly referred to as an "A-tail". The Alu left monomer contains the internal bipartite RNA pol III promoter composed of an A and B box (shown in blue). Alu transcripts will usually contain the 3' flanking unique genomic sequence (shown in gray) past the A-tail due to the lack of a RNA pol III terminator sequence (TTTT) within the Alu sequence. Alu transcripts are thought to have a cap-like structure composed of a gamma-monomethyl phosphate (represented as P-P-P) that has been shown to be present in other pol III transcripts (135, 136). SVA elements are composite elements that contain a 5' region with a variable number of CCCTCT repeats, an antisense Alu-like sequence, a variable number tandem repeat (VNTR) region, and a HERVK-like region previously referred to as SINE-R. A promoter has yet to be identified (represented as "?"), but it is presumed that RNA polymerase II likely drives transcription to generate a capped (brown hexagon) and polyadenylated transcript. Retropseudogenes or processed pseudogenes can originate from the retrotransposition of either RNA pol II or pol III transcripts. Retropseudogenes also are flanked by the hallmark TSD. Although the retrotransposed copies are not expected to contain a promoter sequence with them when they mobilize, sometimes the new flanking genomic sequence into which they insert provides promoter function, and thus some retropseudogenes generate transcripts.

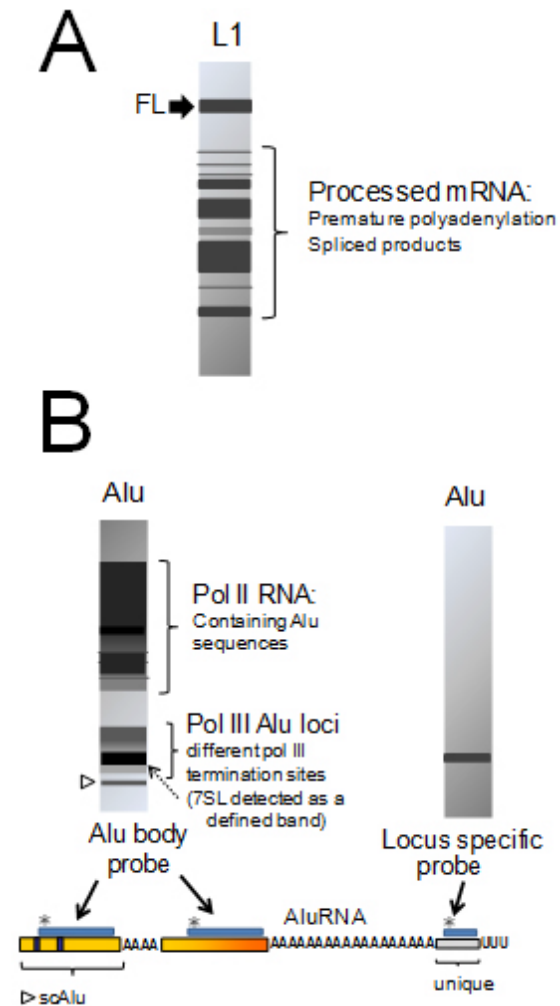


Figure 2. RNA profiles of L1 and Alu elements. Schematic representations of northern blot analysis of L1 (A) and Alu (B) are shown. L1 transcripts are extensively processed both by premature polyadenylation and splicing (42, 44). An L1 RNA profile may show full-length (FL) transcripts (indicated by a black arrow), in addition to a variety of processed products depending on the human tissue analyzed (45). Alu RNA profiles will vary depending on the probe utilized. The structure of a typical Alu transcript is shown at the bottom. Probes (shown as blue boxes marked with asterisks *) that hybridize with the Alu body sequence will detect both large (over 1 kb) and small (100- 600 bp) RNA products (schematic shown on the left). The large products observed include RNA pol II transcripts that contain Alu sequences, while the smaller products observed include the RNA pol III transcribed Alu RNAs of various lengths, the left monomer, a degradation intermediate known as small cytoplasmic Alu (scAlu) (54), indicated by an open arrowhead) and the 7SL transcript (56), indicated by a dotted arrow) due to shared sequence homology. The use of a probe that hybridizes to the unique sequence (derived from the genomic flank) will only detect those RNA pol III Alu transcripts generated from that specific locus (schematic shown on the right).

HERV-W, an LTR-retrotransposon, have also been reported (30).

3. EXPRESSION OF NON-LTR RETROELEMENTS

3.1. Expression of L1: the driving force

Expression of RNA is a requisite for amplification of retroelements. The vertical transmission of retroelements provides evidence of the expression somewhere in the germline. Because the non-autonomous elements depend on L1 products, understanding the distribution and extent of L1 expression is of great importance. L1 activity requires the L1 transcript as a template for the new copy as well as the expression of both ORF1 and ORF2 proteins (11). ORF1 protein (ORF1p) appears to be more abundant and easier to detect, making the evaluation of its endogenous expression more common in the literature. Endogenous expression of ORF1 has been reported in several human cell lines including teratocarcinoma and choriocarcinoma cells (31). Different studies using a variety of tumor samples detected ORF1p in breast and testicular cancers, pediatric germ cell tumors, ileal carcinoids, bladder and pancreatic neuroendocrine tumors including some samples of prostate and colorectal tumors (32-37). Although most examples detected ORF1p expression in the cytoplasm, some cancers displayed a nuclear localization of ORF1p. In these cases, nuclear detection of ORF1p correlated with poor prognosis (32). A detailed analysis of ORF1p expression in mice demonstrated its temporal regulation in germ line and steroidogenic tissue (38). ORF1p has also been detected in somatic cells (syncytiotrophoblasts from placenta) of adult mice (39) and in different regions of the brain of L1-transgenic mice (40). At the time of publication, data on endogenous ORF2p expression in human tissues are scarce. One study detects ORF2p in a variety of tissues including male gonads, prespermatogonia of fetal testis and germ cells of adult testis, Leydig, Sertoli and microvascular endothelial cells (41). As expected, ORF1p expression was also observed in the same cell types.

Detection of L1 proteins in a cell is not a reliable indicator of L1 retrotransposition activity as both ORF1p and ORF2p may derive from defective L1 copies and thus be non-functional. Analysis of L1 RNA expression is complex due to extensive processing by splicing (42, 43) and/or premature polyadenylation (44) of L1 transcripts. Northern blot analysis of L1 transcripts presents the advantage allowing distinction between full-length and other L1 products (Figure 2A), but may be limited in sensitivity. In contrast, RT-PCR approaches can detect very small amounts of L1 RNA. However, it is difficult to envision an RT-PCR approach that distinguishes between full-length and processed L1 products making RT-PCR data unreliable as an indicator of L1 activity. The application of some of the newer technologies such as paired end RNAseq may prove valuable for evaluating L1 transcripts. However, due to the large L1 copy number, even small amounts of DNA contamination will skew the data by particularly enriching for sequence reads matching the 3' regions of L1, as the 5' truncated inserts are more abundant than full length L1 elements. Another limitation

Differences and similarities of active retroelements

of using methods based on short sequence reads derives from the inability to distinguish between reads that derive from L1 fragments present within other non-L1 mRNAs and *bona-fide* L1 transcripts. In addition, most of these methodologies lack information on the orientation of the obtained sequence (sense vs. antisense), making it difficult to discern those reads derived from RNA products generated by the antisense activity of the L1 promoter or other flanking promoters. Published data demonstrate that expression of full-length as well as processed L1 transcripts is widespread in human somatic tissues, including human adult stem cells, and that the amount of L1 RNA and proportion of processed products varies extensively between tissues (45).

The subcellular localization of L1 components is of importance for two reasons: 1) after transcription, the L1 RNA needs to reach the cytoplasm for translation and 2) because of the cis-preference exerted by L1 transcripts (46), the L1 RNA undergoing translation and its two proteins, ORF1p and ORF2p [likely as an RNP (12, 47)], must reach the nucleus to generate a new insert. Analyses using living and fixed cells showed the colocalization of the L1 RNA and both the ORF1 and ORF2 proteins in the cytoplasm (47, 48). The L1 components are often observed associated with markers of cytoplasmic stress granules (49). Previous studies by the same group, using a GFP tagged L1 ORF2p showed that the deletion of its carboxy-terminus caused the protein to also localize to nucleoli (50). However, the importance of the association of L1 components with stress granules and the nucleolus is unclear, as it is unknown if they represent functional intermediate steps of the retrotransposition cycle or the accumulation of byproducts.

3.2. Expression of Alu and other non-autonomous elements

Evaluation of Alu expression is also complex, but for different reasons. First, being particularly enriched in genic regions, Alu sequences are commonly found within RNA pol II transcribed mRNA (51, 52). However, only the RNA pol III Alu transcripts are efficiently retrotransposed (53) making it critical to distinguish between the two types of transcripts. Secondly, Alu elements share sequence homology to the ubiquitously and highly expressed 7SL RNA. Third, Alu transcripts vary in length. Because Alu elements do not contain an RNA pol III terminator within their sequence, transcription will continue past the A-tail into the genomic flanking region (unique sequence) until it reaches an RNA pol III terminator, usually four thymidine residues, somewhere within the downstream genomic DNA (Figure 1). Alu transcripts from different loci will contain a unique 3' sequence characteristic of the individual element generating the RNA. Due to the diverse location of Alu elements, pol-III generated Alu transcripts can vary in length from 300 bp to over 600 bp. Finally, the pol-III generated Alu transcript can be processed to a smaller product (left monomer) also known as small cytoplasmic Alu (scAlu) (54), which is prevalent due to its higher stability than the Alu dimer RNA (55). Thus, a hybridization analysis which uses a probe that anneals to the Alu sequence (Alu body) will likely detect all of these

transcripts generating a variety of smears rather than a distinct band in a Northern blot (see Figure 2B for a schematic example). Due to shared sequence homology, usually the 7SL RNA will appear as an intense discrete band within the smear containing the faster migrating products. Previous work on Alu expression in HeLa cells estimates that 7SL RNA is at least 500 fold more abundant than RNA pol III Alu transcripts (56), making it vastly easier to detect 7SL RNA than Alu RNA. In contrast, a probe to the unique region of the Alu RNA (Figure 2B) should only detect transcripts generated from that specific locus. Size fractionation of transcripts (57) and the use of poly-A selected RNA (58) will enrich for RNA pol III Alu transcripts and reduce the presence of 7SL RNA in the samples. Alu sequences are commonly found in mRNAs; it has been estimated that about 4% of human mRNAs harbor Alu elements (59). One of the better approaches for evaluating Alu expression, is the use of primer extension with reverse transcriptase which gives the advantage of identifying the RNA pol III Alu transcripts by size (60). In contrast, the use of the standard RT-PCR, 3' RACE using an oligo dT primer and microarray approaches will also encounter the same problem of trying to distinguish between *bona fide* RNA pol III Alu transcripts and other RNAs containing Alu and Alu-like sequences. Evaluations using these types of approaches, not only will not adequately reflect Alu expression, but in addition, due to favorable oligo annealing, PCR based assays will selectively amplify the younger less mutated Alu elements introducing a bias in the distribution of Alu subfamily expression.

Due to the complexity of Alu expression analysis, reliable data are scarce. Except for the predicted germ line expression, currently there are minimal data available on tissue or developmental expression of Alu elements. Most of the data are derived from a few studies on cell lines. One of these studies demonstrates that transcripts from different Alu subfamilies can be detected in NTera2D1 cells (61). Recently, several laboratories performed genome wide chromatin immunoprecipitation (ChIP) analyses which identified multiple Alu loci (mostly older elements) bound by RNA polymerase III factors in both cancerous and normal cell lines (62-64). Stringent analyses of their data support the active transcription of several of the identified Alu elements of which different loci express in different cell lines. Detection of transcripts from the older non-active Alu subfamilies suggests that expression alone is not sufficient for retrotransposition of Alu elements, but that other factors determine the retrotranspositional capability of individual Alu elements (discussed in section 4.2). Table 1 illustrates the discrepancy between the distribution of Alu subfamily transcripts in relation to their current activity as estimated by reports of diseases caused by *de novo* insertion of an Alu.

Evidence indicates that Alu expression appears to be limited. Because the internal Alu RNA pol III promoter alone is insufficient to drive expression (65, 66), it is likely that even Alu elements with intact A and B boxes are transcriptionally inactive (56). In addition, Alu expression

Table 1. Alu subfamilies genomic distribution, activity and relative expression

Alu subfamily	% total genomic Alu	# diseases caused by <i>de novo</i> insertion (%) (107)	% Alu transcripts (61) ^{\$}
S + J	82	0 (0%)	66
Y	17	11 (27%)	33
Ya5*	0.3	18 (44%)	0.8
Yb8*	0.2	12 (29%)	0.5
Total	100	41 (100%)	100

* Includes all subfamily variants, \$ Only includes data from validated *bona fide* RNA pol III transcribed Alu elements

s thought to be downregulated through methylation (58, 67, 68). However, RNA pol III driven Alu expression levels can be transiently increased under various stress conditions that alter chromatin structure (69, 70), such as heat shock (69) and viral infection (71-73). Studies indicate that a few loci are the main contributors to this transient increase in Alu expression in response to stress (74), rather than a global response from all Alu elements in the genome. Considering that each of the one million plus Alu elements is flanked upstream by different sequences, it is not surprising that any individual Alu locus may respond differently to changes in chromatin structure or external stimuli such as viral infection. Early studies detected both full length Alu transcripts and the processed scAlu form primarily in the cytoplasm (54, 58, 60). A recent study using *in situ* hybridization and Alu-MS2 tagged RNA detected Alu transcripts mostly in the nucleus (47). However, it is unknown if either the nuclear or cytoplasmic localization of Alu transcripts favors its retrotransposition. It is possible that only those Alu transcripts that reach the cytoplasm are capable of sequestering the L1 ORF2p required to retrotranspose.

Data on expression of other retrotransposed sequences vary considerably. Details on SVA expression remain elusive, largely due to the lack of identification of a well-characterized internal promoter. However, a recent discovery shows that some of these elements have acquired promoter sequences through 5' transduction events (75). This occurs when transcription from an upstream genomic site continues into the SVA sequence and, through splicing, a promoter sequence is juxtaposed with the SVA sequence. Retrotransposition of this transcript generates an insert of an SVA with its own promoter that can subsequently generate new copies creating new SVA subfamilies. Bioinformatic analyses have led to the identification and study of two SVA families that were created through this mechanism: the CpG-SVA family (75) and the SVA_F1 family (76). Evaluations using EST data and RT-PCR demonstrated the expression of the SVA_F1 family in germline, stem cells, tumor cells and different types of somatic cells (76). Interestingly, although L1, Alu and SVA use the L1 retrotransposition machinery, the subcellular localization of their transcripts differs, which possibly contributes to the observed variations in their amplification mechanism (47). Expression of retropseudogenes is unexpected due to the lack of mobilization of the promoter sequence. However, if the copies insert downstream of another promoter, transcription of retropseudogenes is possible. An analysis of pseudogenes identified in ENCODE regions demonstrated that several are transcribed in cell lines or tissues (77). It is still unknown, whether any of the

expressed retropseudogenes are capable of furthering their own amplification.

4. INSERTION MECHANISM

4.1. Retrotransposition process and target primed reverse transcription (TPRT)

The L1 retrotransposition amplification cycle initiates with transcription followed by the localization of the L1 transcript to the cytoplasm for translation. The retrocompetent L1 mRNA is bicistronic, generating both ORF1 and ORF2 proteins required for L1 retrotransposition (11). As mentioned earlier, ORF1p possesses nucleic acid chaperone activity (78, 79), an essential property for L1 retrotransposition (79, 80). L1 RNA shows a strong cis-preference for its own translated proteins during the retrotransposition process (46, 81). The cis-preference of L1 makes it likely that RNA/protein interactions near the ribosome during translation are important in forming the appropriate RNP (Figure 3A) for retrotransposition of L1 to occur (14). In addition, the cis-preference may decrease the ability of other mRNAs to access the L1 ORF1p and ORF2p and explaining the lower retropseudogene copy numbers in the genome. Because multiple ORF1p molecules interact with the L1 RNA (80), and are present in higher numbers than ORF2p (50), it is reasonable to speculate that ORF1p outnumbers ORF2p in the L1 RNP. It is possible that these properties of ORF1p may be important to help "dislodge" the L1 transcript from the ribosomes (Figure 3A) and to protect the RNA from being diverted to RNA degradation pathways. Also, the ORF1p coating of the L1 RNA may prevent new reassembly of ribosomes and allow for the transcript to proceed to the nucleus. This may also be the reason why L1 and retropseudogenes that require ORF1p (82) for retrotransposition are kinetically slower in culture, while the non-translated RNAs, such as Alu, are both ORF1p independent (83, 84) and kinetically faster (53). After L1 RNP assembly, there are limited data on how L1, or any other transcripts such as SVA or processed transcripts (retropseudogenes), reach the nucleus for the insertion process to ensue.

Although all retroelements use an RNA template in their process of amplification, there are fundamental differences in the insertion mechanism between LTR-retrotransposons and non-LTR retrotransposons (3). In general, LTR-retrotransposons reverse transcribe the RNA into a ds-cDNA molecule that is subsequently integrated into the genome. In contrast, L1 and its non-autonomous partners use a unique process to integrate new copies into the genome based on target primed reverse transcription (TPRT) of its RNA. The TPRT model is based on the *in vitro* studies of the amplification cycle of the R2 element in



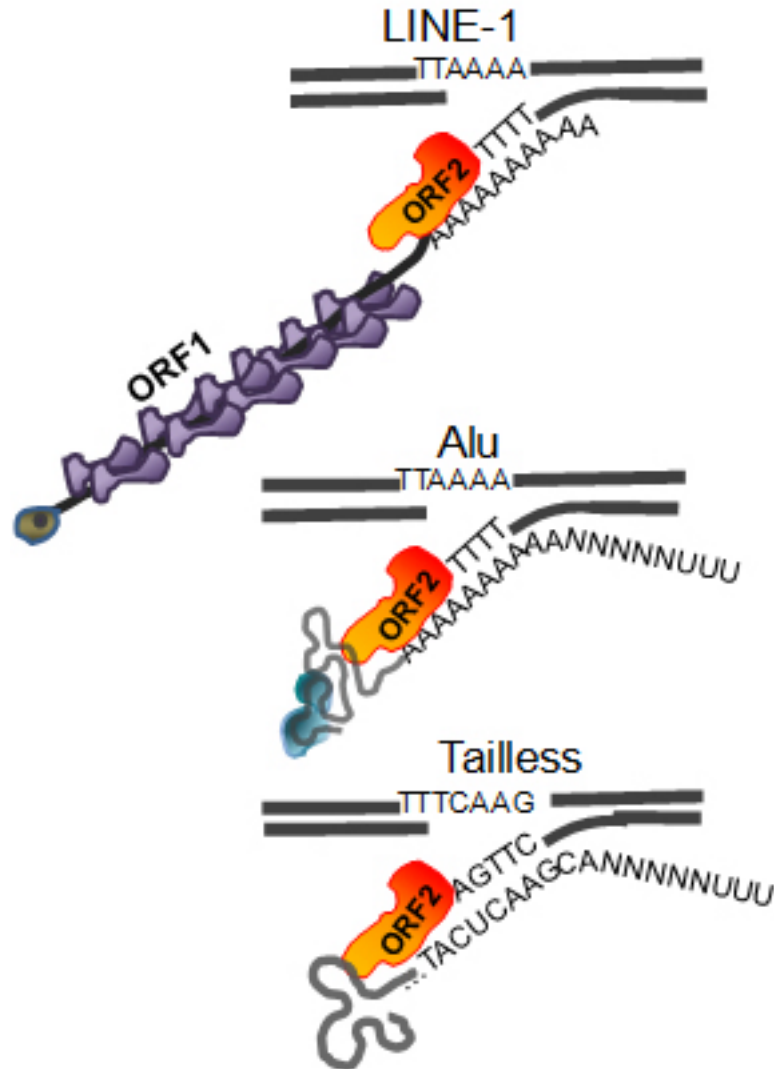


Figure 4. Target primed reverse transcription (TPRT). A schematic representation of the TPRT step of the retrotransposition process is shown for L1, Alu, and tailless tRNA-derived transcripts. The genomic DNA is depicted as dark lines. The consensus 5'-TTAAAA-3' is shown on the top strand of the insertion sites for the L1 and Alu, while a representative sequence of the 5'-TTWX1-17-3' consensus is shown for the tailless retropseudogene. The bottom strand shows the nicked DNA with the exposed nucleotides (e.g. TTTT) thought to anneal with the RNA and provide the priming site for reverse transcription by ORF2p (depicted in orange). The Alu and tailless retropseudogene transcripts are shown with extended 3' sequences (shown as Ns) which represent the unique region of Alu or part of the tRNA-related sequences followed by the RNA pol III terminator (shown as Us).

Bombyx mori (85) and has been adapted into a model for other elements that use the non-LTR insertion mechanism. Under this model, reverse transcription of the template RNA is primed by a nick at the chromosomal integration site. In the case of L1 and its parasites (such as Alu), the L1 ORF2 protein generates the initial nick (86). The ORF2p endonuclease preferentially cleaves the sequence 5'-TTTT/AA-3', which resembles *in vivo* L1 integration sites (86, 87). The poly-A stretch at the 3' end of LINES, SINEs, and processed pseudogenes plays an important role for retrotransposition (88), likely by providing the annealing site for the genomic thymidine residues that

prime reverse transcription (89) (Figure 4). Although very inefficiently, the retrotransposition of Alu transcripts without 3' poly-A stretches can occur (83). In addition, the presence of a few "tailless" 3' truncated Alu inserts in the human genome indicates the likely occurrence of internal priming during the TPRT step (90) (Figure 4). Also, a large number of retrotransposed copies of RNA pol III transcripts (tRNA, tRNA-related, and 4.5S RNA) that lack A-tails and referred to as tailless retropseudogenes have been described (29). Evaluation of the insertion sites demonstrated that the cleaved DNA strand which functions as a primer for reverse transcription by ORF2p presents

perfect matched complementarity to the RNA template sequence (Figure 4). The cleavage consensus sequence for the tailless retropseudogenes varies from the one observed for Alu and L1 inserts, with TT/WX1-17 (29) as being more prevalent (Figure 4). On the other hand, *in vitro* studies show that structural parameters of the target DNA appear more important for recognition and determination of the nicking profiles of the L1 endonuclease (91). These data suggest that the consensus sequences derived from the computational studies of retrotransposed inserts are not exclusively determined by the cleavage preference of the ORF2p endonuclease, but also reflect the ability of the RNA to interact with the cleaved DNA to prime reverse transcription. ORF1p has been suggested to mediate nucleic acid strand transfer steps during L1 reverse transcription (78, 92). Because ORF1p is not strictly required for Alu retrotransposition, as the process can occur in chicken cells which are devoid of endogenous L1 ORF1p (84), it is difficult to define the role of ORF1p in Alu retrotransposition.

Finally, little is known about the retrotransposition process after ORF2p completes reverse transcription of the cDNA. How the second nick is created or how the second DNA strand is generated and ligation of the DNA occurs is still unknown. However, the hallmark tandem site duplications (TSDs) flanking the new retrotransposed copy are well-characterized.

4.2. Requirements for retrotransposition of non autonomous elements

Retrotransposition of the non autonomous elements requires two basic components: expression of the RNA template and accessibility to the L1 proteins. Although ORF1p and ORF2p are essential for L1 retrotransposition, this is not a universal requirement for the non-autonomous elements. As mentioned above, RNA pol II generated transcripts (L1 and retropseudogenes) require ORF1p (82), while RNA pol III transcripts (SINEs) only require ORF2p (53, 83, 84). Regardless, all non-autonomous retroelements require the interaction with the at least the L1 ORF2p to undergo retrotransposition. The physical interaction with ribosomes during translation probably limits other mRNA transcripts (retropseudogene source) from achieving the sufficient proximity to those ribosomes translating L1 to be able to sequester the required ORF1p and ORF2p for retrotransposition (Figure 3A). This limitation may explain the lower abundance of retropseudogenes, particularly when compared to the SINE or LINE copy numbers in the human genome. In contrast, Alu transcripts are not encumbered by the translation machinery. Still, Alu RNA must at least come in contact with ORF2p to facilitate trans-mobilization. However it is uncertain of how or where Alu transcripts interact with ORF2p.

Previously, the ability of the Alu transcript to come in close proximity to the translating L1, has been proposed as a model to explain how Alu efficiently competes for access to ORF2p (89). This proximity model is based on the ability of Alu RNA to target ribosomes through its interaction with two of the signal recognition

particle (SRP) proteins. The Alu RNA retains the folding pattern of its ancestral 7SL transcript, which provides binding sites for SRP9p and SRP14p (93, 94) (Figure 3B). Because SRP14p normally targets the 7SL RNP to the ribosomes and causes the arrest of nascent chain elongation through direct contacts of the “Alu domain” of 7SL RNA and the ribosome (95), it is proposed that SRP9/14 proteins bound Alu RNA will have the same fate. Mutations altering the interaction of Alu transcripts with SRP9/14p diminish the retrotransposition capability, suggesting a potential role for these proteins in the process (96, 97). Further support for a critical role of SRP9/14p comes from a study using another 7SL-derived SINE, the rodent B1 element. A single nucleotide substitution in a key region for SRP9/14p-RNA interaction resulted in a significant increase in the retrotransposition rate of B1 (98). In addition, retrotransposition of Alu transcripts is about 10 fold more efficient than its monomeric form (83), possibly a reflection of having two sets of SRP9/14p proteins per Alu RNP (96) (Figure 3B). Another protein thought to help target the Alu RNP to the ribosomes is the cytoplasmic polyA binding protein (PABP). Because PABP binds SINEs that contain A-tails (99, 100), it is thought to form part of the Alu RNP complex. Due to the PABP interaction with eIF4G in the CAP complex (reviewed in 101), the PABP in the Alu complex may bring the RNP into proximity with newly synthesized ORF2p by binding the CAP complex of the translating L1 RNA(102) (Figure 3B). However, the role of PABP in SINE retrotransposition is unclear.

Alu sequence variations also affect activity of an individual element, as not all Alu transcripts are able to undergo retrotransposition. In addition to mutations that limit SRP9 and 14 binding, studies demonstrate that disruptions in the homopolymeric A-tail and increases in the unique 3' sequence length of the Alu transcript (Figure 1) reduce retrotransposition rates (103). These sequence variations at the Alu 3' end are likely to influence the annealing dynamics of the RNA and the cleaved DNA during the TPRT process (Figure 4), where mismatches and lengthy 3' ends may destabilize the interaction.

Most SINE elements are not derived from 7SL, but instead share sequence homology to tRNAs. It is unlikely that SRP9p and SRP14p interact with tRNA derived SINEs or contribute to their retrotransposition. However, it is possible that alternate cellular proteins substitute for the role of SRP9/14p in the tRNA derived SINEs. Interestingly, the transcripts of these SINEs do not retain the folded cloverleaf structure of the tRNA, but instead form a structure that shares similar folding patterns between different SINEs (104). The conservation of this RNA structural similarity may reflect a binding motif for critical proteins that aid retrotransposition of these SINEs.

Currently, little is known about what other cellular factors may be required for either L1 or the non-autonomous elements. Some DNA repair proteins may play important roles in L1 retrotransposition. Components of the non-homologous end joining (NHEJ) pathway (105) and the damage sensor, ATM (106), appear necessary for

Differences and similarities of active retroelements

promoting L1 retrotransposition. We expect that the role of these proteins may be elucidated when the remainder of the insertion process is better understood.

4.3. Cellular regulatory mechanisms of retrotransposition

Unregulated activity of mobile elements can cause significant genetic damage, thus leading cells to evolve mechanisms to minimize mobile element activity. Active elements cause disease predominantly through insertional mutagenesis and non-allelic homologous recombination (NAHR), resulting in loss of gene function (reviewed in 107). For example, L1 activity contributes to several types of genetic instability by causing inversions, insertions, deletions and chromosomal rearrangements. Recent studies also show that the expression of L1 or L1 ORF2p in human cells induces cell cycle arrest (106), apoptosis (108), and senescence (45, 109). Due to the highly damaging effects of mobile elements, cells have developed surveillance mechanisms to limit their impact.

The first line of defense against the damaging effects of retroelement activity is the prevention of their expression. Several observations demonstrate that methylation plays a major role in silencing retrotransposons. Deficiencies in the DNA methyltransferase 3-like gene (Dnmt3L) prevented *de novo* methylation of retrotransposons causing meiotic failure in the male germ line of mice (110). During malignant transformation, genome-wide hypomethylation of repetitive sequences occurs, altering the normal regulation of expression of these elements (111). Additionally, the piRNA pathway has been implicated in transposon control through methylation. Studies in mice showed that piRNAs and the Piwi proteins MILI and MIWI2 determine *de novo* methylation of retrotransposons in germ cells in mice (112, 113). siRNA activity appears to also regulate L1 expression (114).

Suppression of L1 retrotransposition is further controlled by mRNA processing through extensive splicing and premature polyadenylation (42-44). However, L1 RNA processing varies dramatically across tissues (45), with splicing being the potentially dominant mechanism of L1 RNA regulation in some tissues. However, processing of L1 mRNA does not fully mitigate L1-mediated damage because some L1 splice products are capable of generating active L1 ORF2p (45). Because SINEs only require ORF2p, tissues favoring L1 processing leading to the generation of ORF2p containing splice products could support Alu retrotransposition. Thus, Alu retrotransposition could be actively occurring even if no L1 activity is observed in those tissues.

Other cellular processes function as control mechanisms of retroelements. The members of the apolipoprotein B mRNA editing complex 3 (APOBEC3) have been implicated in the post-transcriptional regulation of L1 and Alu retrotransposition (115, 116). However, not all the proteins in the APOBEC3 gene family universally inhibit retrotransposons; instead, some show specificity to individual elements. For example, APOBEC3G has been

shown to specifically inhibit Alu but not L1 activity (117). Interestingly, the ability of the different APOBEC3 proteins to inhibit retrotransposition is independent of their RNA editing function. Currently, the mechanism by which APOBEC3 inhibits retrotransposition remains unclear (118). However, recent advances on APOBEC 3A proteins indicate that they may function through the inhibition of the elongation of cDNA by reverse transcriptase (reviewed in 119). Another example, the Nucleotide Excision Repair endonuclease, ERCC1/XPF, has been demonstrated to limit L1 activity (120). The authors propose that L1 retrotransposition triggers a DNA damage signal which causes the recruitment of ERCC1/XPF and removal of the L1 insertion intermediate, preventing completion of the process. Although yet to be discovered, it is highly likely that other cellular factors participate in the regulation of retrotransposons.

5. ACTIVITY AND DISTRIBUTION OF NON-LTR RETROELEMENTS

5.1. Activity profiles during evolution

Throughout evolution, different mobile elements have dominated the genomic landscape. L1 elements have been active for over 100 million years (1, 121). Not surprisingly, the non-autonomous elements such as Alu which depend on L1, appeared much later (122). The peak Alu amplification rate occurred between 55 and 35 million years ago, generating approximately 80% of the Alu elements currently present in the human genome (18). During this same time period, peak activity was also observed for both the RNA pol II derived retropseudogenes (123) and the RNA pol III derived tailless retropseudogenes (29) possibly reflecting a time period favorable for non-autonomous element expansion. However, SVA does not follow the same pattern of expansion as its origins date back only 18 to 25 million years ago (22).

5.2. Evidence of current activity

The presence of polymorphic elements between different individuals attests to the current activity of non-LTR retroelements in humans. The extent of genomic diversity contributed by these elements was highlighted by research from several laboratories (4-6) and corroborated further with the analyses from the 1000 human genome project (7). However, the calculated retrotransposition rates for the different elements vary, with an estimated one insertion per every 21, 212, and 916 births for Alu, L1, and SVA, respectively (124). Even more compelling are the multiple *de novo* cases of human retroelement induced diseases where Alu contributes for the majority of documented cases (107).

Although the impact of the insertional mutagenesis by mobile elements is well documented, data on the activity at somatic level are scarce. Transgenic mouse models have shown that somatic activity occurs (40, 125-127). There are a few reports of cancers describing LINE-1 insertions disrupting tumor suppressors (128, 129) or present at a translocation breakpoint (130). However, only one of these examples appears to be an LINE-1 that inserted through retrotransposition. More recently, data

from high throughput DNA sequence analyses of lung tumors detected nine somatic L1 insertions specific to the tumor DNA that were absent from the adjacent normal samples (4). As the field and methodology have advanced, the amount of evidence that these elements may have a significant role at the somatic level has increased. Further studies will provide a better understanding of the somatic role of these elements.

5.3. Insertional distribution

Retrotransposons are distributed in a non-random manner in the human genome. Although L1 and its non-autonomous parasites use ORF2p cleavage sites for insertion, they do not share the same distribution throughout the genome (1). Alu elements are enriched in GC-rich regions while L1s are found at higher densities in AT-rich regions. Data indicate that although the initial insertion distribution of these elements is comparable, selection through time contributes to differential fixation of Alu and L1 elements (1). Changes in the distribution of retrotransposons in the human genome through time indicate that recombination between chromosomes is likely an important mechanism for their redistribution (131). In addition, the distribution between the sex chromosomes and autosomal chromosomes also varies. Genomic analyses demonstrate that young Alu elements are more dense in chromosome Y relative to chromosome X (3 fold) and to autosomes (2 fold) (132, 133). However, the density of Alu elements increase over time in chromosome X relative to autosomal densities (133). Retrotransposons are also significantly higher on the human X chromosome (134) while underrepresented on the Y chromosome. Human diseases caused by insertion of these elements, which represent recent events with a limited exposure to selection processes, also show a skewed distribution, with Alu and L1 insertions highly overrepresented on the X chromosome (107). However, it is difficult to determine if there is a biological component driving the enrichment of these elements on the X chromosome or if ascertainment bias of disease detection contributes to the observed numbers.

6. SUMMARY AND PERSPECTIVE

The impact of mobile element activity on the human genome became clearly evident after determining that these elements have contributed to almost half of its mass. As the field progresses in understanding the biology of the currently active elements, further data demonstrate their significant role in human disease. Among the currently active human transposable elements, LINE-1 takes a central role by not only providing factors for its own retrotransposition, but also enabling the trans-mobilization of non-autonomous elements such as Alu and SVA. Because non-autonomous elements depend on LINE-1 for the active components/protein(s) for their mechanism, it is reasonable to presume that the parasitic element would benefit from a constant and close proximity to an active LINE-1 element *i.e.* a requirement for “flocking together”. However, the amplification behavior and impact of these parasitic non-autonomous elements do not parallel that of LINE-1. For example, diseases caused by *de novo* Alu inserts outnumber those caused by LINE-1

inserts by about 2 to 1. Although it is still unclear why a non-autonomous element would outnumber autonomous elements, ongoing studies demonstrate that these elements exhibit different expression patterns, cellular localization, retrotransposition requirements and possibly insertional preferences that may help explain differences in amplification success rates. The recent discovery that some tissues can express transcripts that generate ORF2p only further demonstrates the unique genetic impact that each family of non-LTR retrotransposons may have and underscores the importance of evaluating these elements individually. To be able to comprehend the collective impact of the active mobile elements on the human genome, the particularities regarding each type of element must first be elucidated.

7. ACKNOWLEDGEMENTS

This publication was made possible by Grants Number R01GM079709A and P20RR020152 to AMR-E from the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

8. REFERENCES

1. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissole, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier: Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001)
2. J. Boeke, D. J. Garfinkel, C. A. Styles, and C. R. Fink: Ty elements transpose through an RNA intermediate. *Cell* 40, 491-500 (1985)
3. T. H. Eickbush and V. K. Jamburuthugoda: The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* 134, 221-234 (2008)
4. R. C. Iskow, M. T. McCabe, R. E. Mills, S. Torene, W. S. Pittard, A. F. Neuwald, E. G. Van Meir, P. M. Vertino, and S. E. Devine: Natural mutagenesis of human genomes

Differences and similarities of active retroelements

- by endogenous retrotransposons. *Cell* 141, 1253-1261 (2010)
5. C. R. Beck, P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge, and J. V. Moran: LINE-1 Retrotransposition Activity in Human Genomes. *Cell* 141, 1159-1170 (2010)
6. C. R. Huang, A. M. Schneider, Y. Lu, T. Niranjana, P. Shen, M. A. Robinson, J. P. Steranka, D. Valle, C. I. Civin, T. Wang, S. J. Wheelan, H. Ji, J. D. Boeke, and K. H. Burns: Mobile interspersed repeats are major structural variants in the human genome. *Cell* 141, 1171-1182 (2010)
7. The 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. *Nature* 467, 1061-1073 (2010)
8. G. Grimaldi, J. Skowronski, and M. F. Singer: Defining the beginning and end of KpnI family segments. *Embo J.* 3, 1753-1759 (1984)
9. B. Brouha, J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran, and H. H. Kazazian, Jr.: Hot L1s account for the bulk of retrotransposition in the human population. *Proc.Natl.Acad.Sci.U.S.A* 100, 5280-5285 (2003)
10. S. L. Martin: Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA.Biol* 7, 67-72 (2010)
11. J. V. Moran, S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, and J. r. Kazazian HH: High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917-927 (1996)
12. H. Hohjoh and M. F. Singer: Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *Embo J.* 15, 630-639 (1996)
13. H. Hohjoh and M. F. Singer: Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *Embo J.* 16, 6034-6043 (1997)
14. D. A. Kulpa and J. V. Moran: Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human Molecular Genetics* 14, 3237-3248 (2005)
15. G. D. Swergold: Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol.Cell Biol.* 10, 6718-6729 (1990)
16. M. Speak: Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol.Cell Biol.* 21, 1973-1985 (2001)
17. P. Nigumann, K. Redik, K. Matlik, and M. Speak: Many Human Genes Are Transcribed from the Antisense Promoter of L1 Retrotransposon. *Genomics* 79, 628-634 (2002)
18. M. R. Shen, M. A. Batzer, and P. L. Deininger: Evolution of the master Alu gene(s) *J Mol.Evol.* 33, 311-320 (1991)
19. A. Weiner, P. Deininger, and A. Efstratiadis: The Reverse Flow of Genetic Information: pseudogenes and transposable elements derived from nonviral cellular RNA. *Annual Reviews of Biochemistry* 55, 631-661 (1986)
20. P. Walter and G. Blobel: Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* 299, 691-698 (1982)
21. S. Fuhrman, P. L. Deininger, P. LaPorte, T. Friedmann, and P. Geiduschek: Analysis of Transcription of the Human Alu Family of Ubiquitous Repeating Element by Eukaryotic RNA Polymerase III. *Nucleic Acid.Res* 9, 6439-6455 (1981)
22. H. Wang, J. Xing, D. Grover, D. J. Hedges, K. Han, J. A. Walker, and M. A. Batzer: SVA elements: a hominid-specific retroposon family. *J Mol.Biol.* 354, 994-1007 (2005)
23. Z. Zhang, P. M. Harrison, Y. Liu, and M. Gerstein: Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13, 2541-2558 (2003)
24. I. Goncalves, L. Duret, and D. Mouchiroud: Nature and structure of human genes that generate retropseudogenes. *Genome Res.* 10, 672-678 (2000)
25. A. N. Khachane and P. M. Harrison: Strong association between pseudogenization mechanisms and gene sequence length. *Biol Direct.* 4, 38 (2009)
26. E. Ullu and A. M. Weiner: Human genes and pseudogenes for the 7SL RNA component of signal recognition particle. *EMBO J* 3, 3303-3310 (1984)
27. A. Buzdin, E. Gogvadze, E. Kovalskaya, P. Volchkov, S. Ustyugova, A. Illarionova, A. Fushan, T. Vinogradova, and E. Sverdlov: The human genome contains many types of chimeric retrogenes generated through *in vivo* RNA recombination. *Nucleic Acids Res* 31, 4385-4390 (2003)
28. J. Perreault, J. F. Noel, F. Briere, B. Cousineau, J. F. Lucier, J. P. Perreault, and G. Boire: Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. *Nucleic Acids Res* 33, 2032-2041 (2005)
29. J. Schmitz, G. Churakov, H. Zischler, and J. Brosius: A novel class of Mammalian-specific tailless retropseudogenes. *Genome Res.* 14, 1911-1915 (2004)
30. A. Pavlicek, J. Paces, D. Elleder, and J. Hejnar: Processed Pseudogenes of Human Endogenous Retroviruses Generated by LINEs: Their Integration, Stability, and Distribution. *Genome Res.* 12, 391-399 (2002)

Differences and similarities of active retroelements

31. D. M. Leibold, G. D. Swergold, M. F. Singer, R. E. Thayer, B. A. Dombroski, and T. G. Fanning: Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc.Natl.Acad.Sci.U.S.A* 87, 6990-6994 (1990)
32. C. R. Harris, R. Normart, Q. Yang, E. Stevenson, B. G. Haffty, S. Ganesan, C. Cordon-Cardo, A. J. Levine, and L. H. Tang: Association of Nuclear Localization of a Long Interspersed Nuclear Element-1 Protein in Breast Tumors with Poor Prognostic Outcomes. *Genes Cancer* 1, 115-124 (2010)
33. G. L. Bratthauer, R. D. Cardiff, and T. G. Fanning: Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* 73, 2333-2336 (1994)
34. H. L. Asch, E. Eliacin, T. G. Fanning, J. L. Connolly, G. Bratthauer, and B. B. Asch: Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol.Res.* 8, 239-247 (1996)
35. G. L. Bratthauer and T. G. Fanning: Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* 7, 507-510 (1992)
36. G. L. Bratthauer and T. G. Fanning: LINE-1 retrotransposon expression in pediatric germ cell tumors. *Cancer* 71, 2383-2386 (1993)
37. Y. Su, S. Davies, M. Davis, H. Lu, R. Giller, M. Krailo, Q. Cai, L. Robison, and X. O. Shu: Expression of LINE-1 p40 protein in pediatric malignant germ cell tumors and its association with clinicopathological parameters: A report from the Children's Oncology Group. *Cancer Lett.* (2006)
38. S. A. Trelogan and S. L. Martin: Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl Acad Sci U S A* 92, 1520-1524 (1995)
39. D. Branciforte and S. L. Martin: Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol* 14, 2584-2592 (1994)
40. A. R. Muotri, V. T. Chu, M. C. Marchetto, W. Deng, J. V. Moran, and F. H. Gage: Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903-910 (2005)
41. S. Ergun, C. Buschmann, J. Heukeshoven, K. Dammann, F. Schnieders, H. Lauke, F. Chalajour, N. Kilic, W. H. Stratling, and G. G. Schumann: Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J.Biol.Chem.* 279, 27753-27763 (2004)
42. V. P. Belancio, D. J. Hedges, and P. Deininger: LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Research* 34, 1512-1521 (2006)
43. V. P. Belancio, A. M. Roy-Engel, and P. Deininger: The impact of multiple splice sites in human L1 elements. *Gene* 411, 38-45 (2008)
44. V. Perepelitsa-Belancio and P. L. Deininger: RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet* 35, 363-366 (2003)
45. V. P. Belancio, A. M. Roy-Engel, R. R. Pochampally, and P. Deininger: Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* (2010)
46. W. Wei, N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke, and J. V. Moran: Human L1 retrotransposition: cis preference versus trans complementation. *Mol.Cell Biol.* 21, 1429-1439 (2001)
47. J. L. Goodier, P. K. Mandal, L. Zhang, and H. H. Kazazian, Jr.: Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum.Mol Genet* (2010)
48. A. J. Doucet, A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran, and N. Gilbert: Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet* 6, (2010)
49. J. L. Goodier, L. Zhang, M. R. Vetter, and H. H. Kazazian, Jr.: LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex. *Mol.Cell Biol.* 27, 6469-6483 (2007)
50. J. L. Goodier, E. M. Ostertag, K. A. Engleka, M. C. Seleme, and H. H. Kazazian, Jr.: A potential role for the nucleolus in L1 retrotransposition. *Human Molecular Genetics* 13, 1041-1048 (2004)
51. D. Grover, M. Mukerji, P. Bhatnagar, K. Kannan, and S. K. Brahmachari: Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* 20, 813-817 (2004)
52. R. Versteeg, B. D. C. van Schaik, M. F. van Batenburg, M. Roos, R. Monajemi, H. Caron, H. J. Bussemaker, and A. H. C. van Kampen: The Human Transcriptome Map Reveals Extremes in Gene Density, Intron Length, GC Content, and Repeat Pattern for Domains of Highly and Weakly Expressed Genes. *Genome Research* 13, 1998-2004 (2003)
53. E. N. Krouter, V. P. Belancio, B. J. Wagstaff, and A. M. Roy-Engel: The RNA Polymerase Dictates ORF1 Requirement and Timing of LINE and SINE Retrotransposition. *PLoS Genet* 5, e1000458 (2009)
54. R. J. Maraia, C. T. Driscoll, T. Bilyeu, K. Hsu, and G. J. Darlington: Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol.Cell Biol.* 13, 4233-4241 (1993)

Differences and similarities of active retroelements

55. T. H. Li and C. W. Schmid: Alu's dimeric consensus sequence destabilizes its transcripts. *Gene* 324, 191-200 (2004)
56. K. E. Paulson and C. W. Schmid: Transcriptional inactivity of Alu repeats in HeLa cells. *Nucleic.Acids.Res.* 14, 6145-6158 (1986)
57. D. Sinnott, C. Richer, J. M. Deragon, and D. Labuda: Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J.Biol.Chem.* 266, 8675-8678 (1991)
58. W. M. Liu, R. J. Maraia, C. M. Rubin, and C. W. Schmid: Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic.Acids.Res.* 22, 1087-1095 (1994)
59. A. Jasinska and W. J. Krzyzosiak: Repetitive sequences that shape the human transcriptome. *FEBS Letters* 567, 136-141 (2004)
60. T. H. Li, C. Kim, C. M. Rubin, and C. W. Schmid: K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res.* 28, 3031-3039 (2000)
61. T. H. Shaikh, A. M. Roy, J. Kim, M. A. Batzer, and P. L. Deininger: cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J Mol.Biol.* 271, 222-234 (1997)
62. D. Canella, V. Praz, J. H. Reina, P. Cousin, and N. Hernandez: Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Res.* 20, 710-721 (2010)
63. A. J. Oler, R. K. Alla, D. N. Roberts, A. Wong, P. C. Hollenhorst, K. J. Chandler, P. A. Cassiday, C. A. Nelson, C. H. Hagedorn, B. J. Graves, and B. R. Cairns: Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat Struct Mol Biol* 17, 620-628 (2010)
64. Z. Moqtaderi, J. Wang, D. Raha, R. J. White, M. Snyder, Z. Weng, and K. Struhl: Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nat Struct Mol Biol* 17, 635-640 (2010)
65. I. Chesnokov and C. W. Schmid: Flanking sequences of an Alu source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors. *J.Mol.Evol.* 42, 30-36 (1996)
66. A. M. Roy, N. C. West, A. Rao, P. Adhikari, C. Alemán, A. P. Barnes, and P. L. Deininger: Upstream flanking sequences and transcription of SINEs. *J Mol.Biol.* 302, 17-25 (2000)
67. W. M. Liu and C. W. Schmid: Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. *Nucleic.Acids.Res.* 21, 1351-1359 (1993)
68. C. M. Rubin, C. A. VandeVoort, R. L. Teplitz, and C. W. Schmid: Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic.Acids.Res.* 22, 5121-5127 (1994)
69. C. Kim, C. M. Rubin, and C. W. Schmid: Genome-wide chromatin remodeling modulates the Alu heat shock response. *Gene* 276, 127-133 (2001)
70. W. M. Liu, W. M. Chu, P. V. Choudary, and C. W. Schmid: Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic.Acids.Res.* 23, 1758-1765 (1995)
71. B. Panning and J. R. Smiley: Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. *Mol.Cell Biol.* 13, 3231-3244 (1993)
72. B. Panning and J. R. Smiley: Activation of expression of multiple subfamilies of human Alu elements by adenovirus type 5 and herpes simplex virus type 1. *J.Mol.Biol.* 248, 513-524 (1995)
73. V. R. Russanova, C. T. Driscoll, and B. H. Howard: Adenovirus type 2 preferentially stimulates polymerase III transcription of Alu elements by relieving repression: a potential role for chromatin. *Mol.Cell Biol.* 15, 4282-4290 (1995)
74. T. H. Li and C. W. Schmid: Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* 276, 135-141 (2001)
75. O. B. Bantysh and A. A. Buzdin: Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. *Biochemistry (Mosc.)* 74, 1393-1399 (2009)
76. A. Damert, J. Raiz, A. V. Horn, J. Lower, H. Wang, J. Xing, M. A. Batzer, R. Lower, and G. G. Schumann: 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Research* 19, 1992-2008 (2009)
77. D. Zheng, A. Frankish, R. Baertsch, P. Kapranov, A. Reymond, S. W. Choo, Y. Lu, F. Denoeud, S. E. Antonarakis, M. Snyder, Y. Ruan, C. L. Wei, T. R. Gingeras, R. Guigo, J. Harrow, and M. B. Gerstein: Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res.* 17, 839-851 (2007)
78. S. L. Martin and F. D. Bushman: Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol.Cell Biol.* 21, 467-475 (2001)
79. S. L. Martin, M. Cruceanu, D. Branciforte, L. P. Wai-Lun, S. C. Kwok, R. S. Hodges, and M. C. Williams: LINE-1 retrotransposition requires the nucleic acid

Differences and similarities of active retroelements

- chaperone activity of the ORF1 protein. *J Mol.Biol.* 348, 549-561 (2005)
80. S. Basame, L. P. Wai-Lun, G. Howard, D. Branciforte, D. Keller, and S. L. Martin: Spatial Assembly and RNA Binding Stoichiometry of a LINE-1 Protein Essential for Retrotransposition. *J Mol.Biol.* 357, 351-357 (2006)
81. D. A. Kulpa and J. V. Moran: Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat.Struct.Mol.Biol.* 13, 655-660 (2006)
82. C. Esnault, J. Maestre, and T. Heidmann: Human LINE retrotransposons generate processed pseudogenes. *Nat.Genet.* 24, 363-367 (2000)
83. M. Dewannieux, C. Esnault, and T. Heidmann: LINE-mediated retrotransposition of marked Alu sequences. *Nat.Genet.* 35, 41-48 (2003)
84. N. Wallace, B. J. Wagstaff, P. L. Deininger, and A. M. Roy-Engel: LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene* 419, 1-6 (2008)
85. D. D. Luan, M. H. Korman, J. L. Jakubczak, and T. H. Eickbush: 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 (1993)
86. Q. Feng, J. V. Moran, J. r. Kazazian HH, and J. D. Boeke: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905-916 (1996)
87. J. Jurka: Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc.Natl.Acad.Sci.U.S.A.* 94, 1872-1877 (1997)
88. M. Dewannieux and T. Heidmann: Role of poly(A) tail length in Alu retrotransposition. *Genomics* 86, 378-381 (2005)
89. J. D. Boeke: LINEs and Alus--the polyA connection. *Nat.Genet.* 16, 6-7 (1997)
90. D. Srikanta, S. K. Sen, E. M. Conlin, and M. A. Batzer: Internal priming: an opportunistic pathway for L1 and Alu retrotransposition in hominins. *Gene* 448, 233-241 (2009)
91. K. Repanas, N. Zingler, L. E. Layer, G. G. Schumann, A. Perrakis, and O. Weichenrieder: Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. *Nucleic Acids Res.* 35, 4914-4926 (2007)
92. S. L. Martin: The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition. *J Biomed.Biotechnol.* 2006, 45621 (2006)
93. K. Hsu, D. Y. Chang, and R. J. Maraia: Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9. *J Biol Chem* 270, 10179-10186 (1995)
94. D. Y. Chang, K. Hsu, and R. J. Maraia: Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res* 24, 4165-4170 (1996)
95. Y. Thomas, N. Bui, and K. Strub: A truncation in the 14 kDa protein of the signal recognition particle leads to tertiary structure changes in the RNA and abolishes the elongation arrest activity of the particle. *Nucleic Acids Res.* 25, 1920-1929 (1997)
96. J. Sarrowa, D. Y. Chang, and R. J. Maraia: The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA. *Mol.Cell Biol.* 17, 1144-1151 (1997)
97. E. A. Bennett, H. Keller, R. E. Mills, S. Schmidt, J. V. Moran, O. Weichenrieder, and S. E. Devine: Active Alu retrotransposons in the human genome. *Genome Res* 18, 1875-1883 (2008)
98. M. Dewannieux and T. Heidmann: L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J Mol.Biol.* 349, 241-247 (2005)
99. N. West, A. Roy-Engel, H. Imataka, N. Sonenberg, and P. Deininger: Shared Protein Components of SINE RNPs. *J.Mol.Biol.* 321, 423-432 (2002)
100. R. Muddashetty, T. Khanam, A. Kondrashov, M. Bundman, A. Iacoangeli, J. Kremerskothen, K. Duning, A. Barnekow, A. Huttenhofer, H. Tiedge, and J. Brosius: Poly(A)-binding Protein is Associated with Neuronal BC1 and BC200 Ribonucleoprotein Particles. *J.Mol.Biol.* 321, 433-445 (2002)
101. A. C. Gingras, B. Raught, and N. Sonenberg: eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu.Rev.Biochem.* 68, 913-963 (1999)
102. A. M. Roy-Engel, A. H. Salem, O. O. Oyeniran, L. Deininger, D. J. Hedges, G. E. Kilroy, M. A. Batzer, and P. L. Deininger: Active alu element "A-Tails": size does matter. *Genome Res.* 12, 1333-1344 (2002)
103. M. S. Comeaux, A. M. Roy-Engel, D. J. Hedges, and P. L. Deininger: Diverse cis factors controlling Alu retrotransposition: What causes Alu elements to die? *Genome Res* 19, 545-555 (2009)
104. F. J. Sun, S. Fleurdepine, C. Bousquet-Antonelli, G. Caetano-Anolles, and J. M. Deragon: Common evolutionary trends for SINE RNA structures. *Trends in Genetics* 23, 26-33 (2007)
105. J. Suzuki, K. Yamaguchi, M. Kajikawa, K. Ichianagi, N. Adachi, H. Koyama, S. Takeda, and N. Okada: Genetic

Differences and similarities of active retroelements

evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet* 5, e1000461 (2009)

106. S. L. Gasior, T. P. Wakeman, B. Xu, and P. L. Deininger: The human LINE-1 retrotransposon creates DNA double-strand breaks. *J.Mol.Biol.* 357, 1383-1393 (2006)

107. V. P. Belancio, D. J. Hedges, and P. Deininger: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 18, 343-358 (2008)

108. S. M. Belgnaoui, R. G. Gosden, O. J. Semmes, and A. Haoudi: Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.* 6, 13 (2006)

109. N. A. Wallace, V. P. Belancio, and P. L. Deininger: L1 mobile element expression causes multiple types of toxicity. *Gene* 419, 75-81 (2008)

110. D. Bourc'his and T. H. Bestor: Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96-99 (2004)

111. D. J. Weisenberger, M. Campan, T. I. Long, M. Kim, C. Woods, E. Fiala, M. Ehrlich, and P. W. Laird: Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* 33, 6823-6836 (2005)

112. A. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, and G. J. Hannon: A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell* 31, 785-799 (2008)

113. S. Kuramochi-Miyagawa, T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, and T. Nakano: DNA methylation of retrotransposon genes is regulated by Piwi family members ML1 and MIW12 in murine fetal testes. *Genes & Development* 22, 908-917 (2008)

114. N. Yang and H. H. Kazazian, Jr.: L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat.Struct.Mol.Biol.* 13, 763-771 (2006)

115. H. P. Bogerd, H. L. Wiegand, A. E. Hulme, J. L. Garcia-Perez, K. S. O'shea, J. V. Moran, and B. R. Cullen: Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc.Natl.Acad.Sci.U.S.A* 103, 8780-8785 (2006)

116. A. E. Hulme, H. P. Bogerd, B. R. Cullen, and J. V. Moran: Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* 390, 199-205 (2007)

117. A. E. Hulme, H. P. Bogerd, B. R. Cullen, and J. V. Moran: Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* 390, 199-205 (2007)

118. A. M. Niewiadomska, C. Tian, L. Tan, T. Wang, P. T. N. Sarkis, and X. F. Yu: Differential Inhibition of Long Interspersed Element 1 by APOBEC3 Does Not Correlate with High-Molecular-Mass-Complex Formation or P-Body Association. *The Journal of Virology* 81, 9577-9583 (2007)

119. R. S. Aguiar and B. M. Peterlin: APOBEC3 proteins and reverse transcription. *Virus Research* 134, 74-85 (2008)

120. S. L. Gasior, A. M. Roy-Engel, and P. L. Deininger: ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst)* 7, 983-989 (2008)

121. H. Khan, A. Smit, and S. Boissinot: Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* 16, 78-87 (2006)

122. Y. Quentin: Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic.Acids.Res.* 20, 3397-3401 (1992)

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

124. J. Xing, Y. Zhang, K. Han, A. H. Salem, S. K. Sen, C. D. Huff, Q. Zhou, E. F. Kirkness, S. Levy, M. A. Batzer, and L. B. Jorde: Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* 19, 1516-1526 (2009)

125. H. Kano, I. Godoy, C. Courtney, M. R. Vetter, G. L. Gerton, E. M. Ostertag, and H. H. Kazazian, Jr.: L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303-1312 (2009)

126. W. An, J. S. Han, C. M. Schrum, A. Maitra, F. Koentgen, and J. D. Boeke: Conditional activation of a single-copy L1 transgene in mice by Cre. *Genesis.* 46, 373-383 (2008)

127. W. An, J. S. Han, S. J. Wheelan, E. S. Davis, C. E. Coombes, P. Ye, C. Triplett, and J. D. Boeke: Active retrotransposition by a synthetic L1 element in mice. *Proc.Natl.Acad.Sci.U.S.A* 103, 18662-18667 (2006)

128. B. Morse, P. G. Rotherg, V. J. South, J. M. Spandorfer, and S. M. Astrin: Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333, 87-90 (1988)

129. Y. Miki, I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K. W. Kinzler, B. Vogelstein, and Y. Nakamura: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643-645 (1992)

Differences and similarities of active retroelements

130. J. Liu, M. M. Nau, J. Zucman-Rossi, J. I. Powell, C. J. Allegra, and J. J. Wright: LINE-I element insertion at the t(11,22) translocation breakpoint of a desmoplastic small round cell tumor. *Genes Chromosomes.Cancer* 18, 232-239 (1997)

131. P. Medstrand, L. N. van de Lagemaat, and D. L. Mager: Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* 12, 1483-1495 (2002)

132. J. Jurka, M. Krnjajic, V. V. Kapitonov, J. E. Stenger, and O. Kokhanyy: Active Alu elements are passed primarily through paternal germlines. *Theor.Popul.Biol* 61, 519-530 (2002)

133. J. Jurka, O. Kohany, A. Pavlicek, V. V. Kapitonov, and M. V. Jurka: Duplication, coclustering, and selection of human Alu retrotransposons. *Proc.Natl.Acad.Sci.U.S.A* 101, 1268-1272 (2004)

134. G. Drouin: Processed pseudogenes are more abundant in human and mouse X chromosomes than in autosomes. *Mol Biol Evol.* 23, 1652-1655 (2006)

135. G. P. Shumyatsky, S. V. Tillib, and D. A. Kramerov: B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. *Nucleic Acids Res.* 18, 6347-6351 (1990)

136. G. Shumyatsky, S. Shimba, and R. Reddy: Capping signals correspond to the 5' end in four eukaryotic small RNAs containing gamma-monomethylphosphate cap structure. *Gene Expr.* 4, 29-41 (1994)

Key Words: LINE, SINE, Alu, Retropseudogene, Non-LTR retrotransposon, SVA, TPRT, Review

Send correspondence to: Astrid M. Roy-Engel, Tulane University, Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane Cancer Center, SL-66 1430 Tulane Ave., New Orleans, LA 70112, Tel: 504-988-6316, Fax: 504-988-5516, E-mail: aengel@tulane.edu

<http://www.bioscience.org/current/vol17.htm>