Genome Biology

REVIEW

Alu elements: know the SINEs

Prescott Deininger*

Abstract

Alu elements are primate-specific repeats and comprise 11% of the human genome. They have wide-ranging influences on gene expression. Their contribution to genome evolution, gene regulation and disease is reviewed.

Keywords *Alu*, diversity, evolution, gene expression, genetic instability, mutagenesis, retroelements, transposition.

Alu elements represent one of the most successful of all mobile elements, having a copy number well in excess of 1 million copies in the human genome [1] (contributing almost 11% of the human genome). They belong to a class of retroelements termed SINEs (short interspersed elements) and are primate specific. These elements are non-autonomous, in that they acquire *trans*-acting factors for their amplification from the only active family of autonomous human retroelements: LINE-1 [2].

Although active at higher levels earlier in primate evolution, *Alu* elements continue to insert in modern humans, including somatic insertion events, creating genetic diversity and contributing to disease through insertional mutagenesis. They are also a major factor contributing to non-allelic homologous recombination events causing copy number variation and disease. *Alu* elements code for low levels of RNA polymerase III transcribed RNAs that contribute to retrotransposition. However, the ubiquitous presence of *Alu* elements throughout the human genome has led to their presence in a large number of genes and their transcripts. Many individual *Alu* elements have wide-ranging influences on gene expression, including influences on polyadenylation [3,4], splicing [5-7] and ADAR (adenosine deaminase that acts on RNA) editing [8-10].

This review focuses heavily on studies generated as a result of the advent of high-throughput genomics providing huge datasets of genome sequences, and data on gene expression and epigenetics. These data provide tremendous insight into the role of *Alu* elements in genetic instability and genome evolution, as well as their many impacts on expression of the genes in their vicinity. These roles then influence normal cellular health and function, as well as having a broad array of impacts on human health.

Alu structure and amplification mechanism

The general structure of an *Alu* element is presented in Figure 1a. The body of the *Alu* element is about 280 bases in length, formed from two diverged dimers, ancestrally derived from the 7SL RNA gene, separated by a short A-rich region (reviewed in [11]). The 3' end of an *Alu* element has a longer A-rich region that plays a critical role in its amplification mechanism [12]. The entire *Alu* element is flanked by direct repeats of variable length that are formed by duplication of the sequences at the insertion site. *Alu* elements have an internal RNA polymerase III promoter that potentially initiates transcription at the beginning of the *Alu* and produces RNAs that are responsible for their amplification. However, *Alu* elements have no terminator for transcription and the transcripts terminate at nearby genomic locations using a TTTT terminator sequence.

Each RNA polymerase III generated *Alu* RNA is unique in terms of: (i) accumulated mutations in the *Alu* element itself; (ii) the length and accumulated sequence heterogeneity in the encoded A-rich region at its 3' end; and (iii) the unique 3' end on each RNA transcribed from the adjacent genomic site. Those RNAs are then thought to assemble into ribonucleoprotein particles (Figure 1b) that involve the SRP9/14 heterodimer [13], polyA-binding protein (PABP) [14,15] and at least one other unidentified protein that binds to the RNA structure [14,15]. The SRP9/14 proteins and PABP are thought to help the *Alu* RNA associate with a ribosome, where it might become associated with ORF2 protein (ORF2p) being translated from L1 elements [2,16,17]. *Alu* RNAs then utilize the purloined ORF2p to copy themselves at a new genomic site using a process termed target-primed reverse transcription (Figure 1c; reviewed in [18,19]).

Although *Alu* is dependent on the L1 ORF2p protein, *Alu* retrotransposition is not simply an extension of the L1 retrotransposition process. For instance, L1 depends

*Correspondence: pdeinin@tulane.edu
Tulane Cancer Center, SL66, 1430 Tulane Avenue, New Orleans, LA 70112, USA

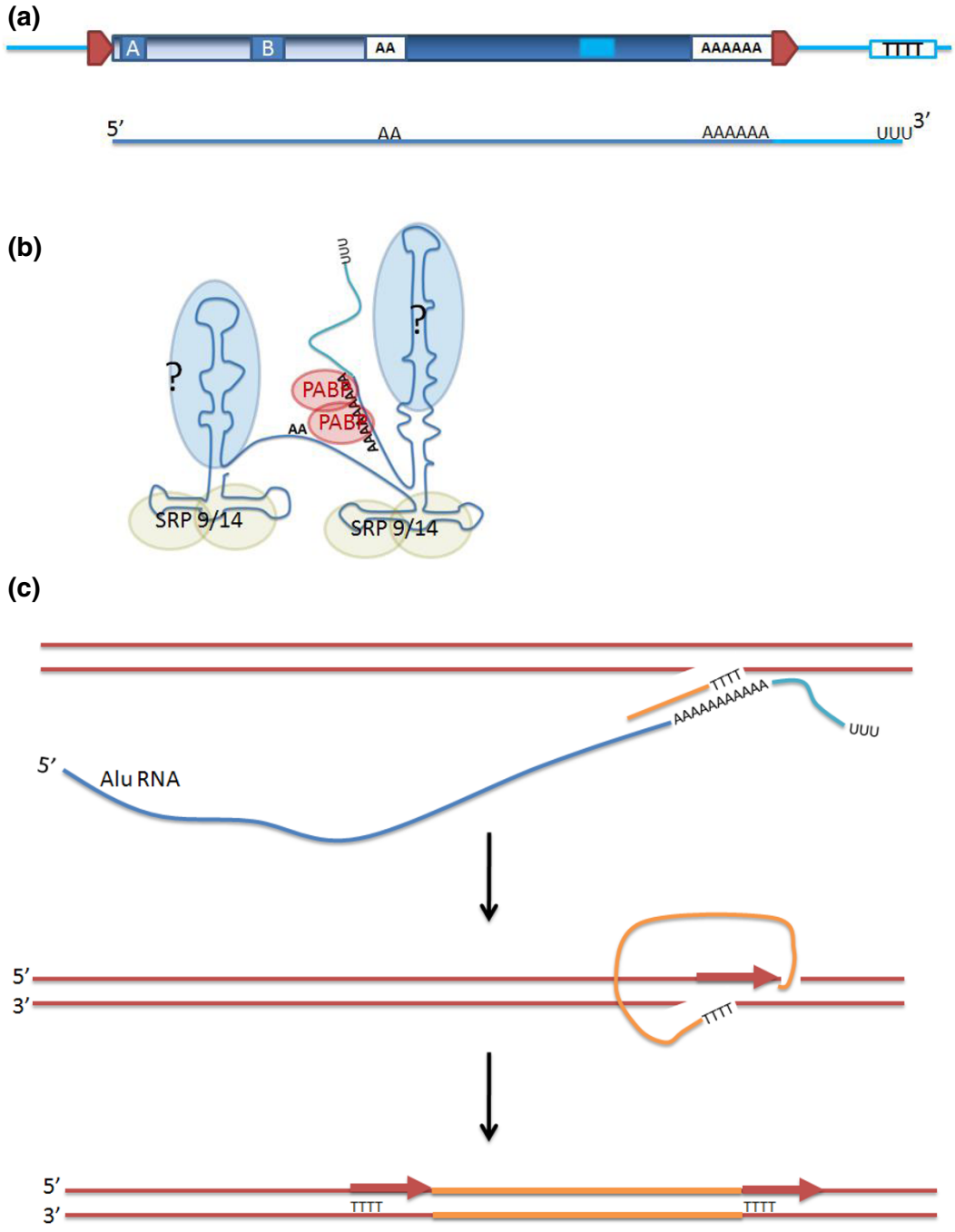

Figure 1. The structure of an *Alu* element. (a) The top portion shows a genomic *Alu* element between two direct repeats formed at the site of insertion (red arrowheads). The *Alu* ends with a long A-run, often referred to as the A-tail, and it also has a smaller A-rich region (indicated by AA) separating the two halves of a diverged dimer structure. *Alu* elements have the internal components of a RNA polymerase III promoter (boxes A and B), but they do not encode a terminator for RNA polymerase III. They utilize whatever stretch of T nucleotides is found at various distances downstream of the *Alu* element to terminate transcription. A typical *Alu* transcript is shown below the genomic *Alu*, showing that it encompasses the entire *Alu*, including the A-tail, and has a 3′ region that is unique for each locus. **(b)** The *Alu* RNA is thought to fold into separate structures for each monomer unit. The RNA has been shown to bind the 7SL RNA SRP9 and 14 heterodimer, as well as polyA-binding protein (PABP). It is thought that at least one other protein binds the duplex portion of the RNA structure. **(c)** In the target-primed reverse transcription mechanism, the *Alu* RNA (blue) brings the ORF2p to the genome where its endonuclease activity cleaves at a T-rich consensus sequence. The T-rich region primes reverse transcription by ORF2p on the 3′ A-tail region of the *Alu* element. This creates a cDNA copy of the body of the *Alu* element. A nick occurs by an unknown mechanism on the second strand and second-strand synthesis is primed. The new *Alu* element is then flanked by short direct repeats that are duplicates of the DNA sequence between the first and second nicks.

on ORF1p and ORF2p, while *Alu* requires ORF2p only [2,20,21]. This may be one of the reasons why *Alu* causes several times as many diseases as L1 through insertion [22,23] and has twice the copy number of L1 [1]. Because L1 elements have been shown to have a splice variant that makes only ORF2p [24], or that may express ORF2p from elements with a mutated ORF1, *Alu* might be able to amplify in cells that do not effectively amplify L1. In fact, although L1 transcription is high in the testis, almost all of the RNA is not full-length, mostly due to splicing [24]. This means that *Alu* may retrotranspose well in the testis, even though L1 retrotransposes poorly. *Alu* and L1s have several other differences. Following expression, *Alu* RNAs can retrotranspose rapidly, whereas L1 RNAs take almost 24 h longer [25]. Retrotransposition of *Alu* and L1 elements is also differentially influenced by different APOBEC3 proteins [26-28]. *Alu* elements encode the A-tail separately at each locus rather than through post-transcriptional polyadenylation, as with L1. Thus, *Alu* A-tails are prone to shrinkage and accumulation of mutations that can affect the amplification process from each particular locus (discussed below) [16].

Only a handful of the greater than 1 million genomic *Alu* elements can amplify [29,30]. It seems highly likely that relatively few polymorphic elements in the population have high amplification capability that maintains *Alu* amplification within the population. There are many factors that contribute to the relative amplification activity of an *Alu* locus (Figure 2) [29,31]. These include: (i) the influence of the primary genomic sequence on transcription; (ii) epigenetic influences on transcription; (iii) the length, and possibly the specific nature, of the 3' unique region of the *Alu* RNAs; (iv) the length and heterogeneity of the A-tail of the *Alu*; and (v) divergence of the body of the *Alu* element, which seems likely to influence RNA structure and probably relevant protein binding (Figure 1b).

These mechanistic features all contribute to the observed paucity of actively amplifying 'master' or 'source' *Alu* elements in the human genome. The internal RNA polymerase III promoter is not strong unless it fortuitously lands near appropriate flanking sequences [32]. Furthermore, epigenetics seems to silence the majority of *Alu* transcripts. Thus, there are generally very low levels of RNA polymerase III transcribed *Alu* RNAs in a cell and it is transcribed by a number of dispersed loci, including many loci that are incapable of active retrotransposition [33]. Because the A-tail grows during the insertion process [2,34], most new inserts have a sufficiently long A-tail for effective amplification. However, because each new insert lands in a different genomic environment, the new loci will vary tremendously in their transcription potential owing to the influences of flanking sequences [32] and epigenetics. In addition, the 3' flanking sequence will provide the RNA polymerase III terminator, and those with longer 3' unique regions will be poor at retrotransposition [29]. Following insertion, those elements that are initially capable of retrotransposition will gradually lose that capability by a series of sequence changes. The most rapid change will be that the long, relatively unstable A-tails will shrink rapidly [16], resulting in lower retrotransposition capability [12,29]. In addition, the A-tails will rapidly accumulate mutations and often form variant microsatellite-like sequences at their ends that will also impair the activity [29]. Over the long run, the body of the *Alu* element will accumulate mutations [31], first CpG mutations, and then other random mutations, which will alter the promoter, RNA folding, and/or interactions with cellular proteins, leaving relatively few of the older *Alu* elements capable of retrotransposition. The sum of all of these factors contributes to the lack of activity of most *Alu* elements.

Alu elements and genome evolution

Alu elements are ancestrally derived from the 7SL RNA gene [35,36]. Although the details of the origin are not known, it seems likely that a relatively inefficient retrotransposon was formed by a deleted version of the 7SL RNA gene sometime before the primate/rodent evolutionary divergence. This precursor then evolved into B1 repeats in rodents, and into FLAM (free left *Alu* monomer) and FRAM (free right *Alu* monomer) sequences in the primate lineage [36,37]. A dimer of FLAM and FRAM eventually took on the highly efficient amplification characteristics of the *Alu* elements.

Large-scale sequencing studies of primate genomes have provided a great deal of detail on the evolution of *Alu* elements. Because there is no specific mechanism for removal of *Alu* insertions, *Alu* evolution is dominated by the accumulation of new *Alu* inserts. These new *Alu* inserts accumulate sequence variation over time and are rarely removed by non-specific deletion processes. Different periods of evolutionary history have given rise to different subfamilies of *Alu* elements with a very limited and homogeneous group of subfamilies active in any given species because of a very limited number of source, or master, *Alu* loci (Figure 3) [38,39]. The earliest *Alu* elements were the J subfamily, followed by a very active series of S subfamilies. The dominant S subfamilies included Sx, Sq, Sp and Sc (*Alu* subfamilies and their nomenclature are defined in [40]). More recently, most of the *Alu* amplification in old world monkey and ape lineages has been from a series of Y subfamilies, with Ya5 and Yb8 dominating in humans. The *Alu* amplification rate peaked with the S subfamilies [38]. Comparisons between chimpanzee and human genomes have shown that, since their divergence about 6 million years ago, there have been about 2,400 and 5,000 lineage-specific

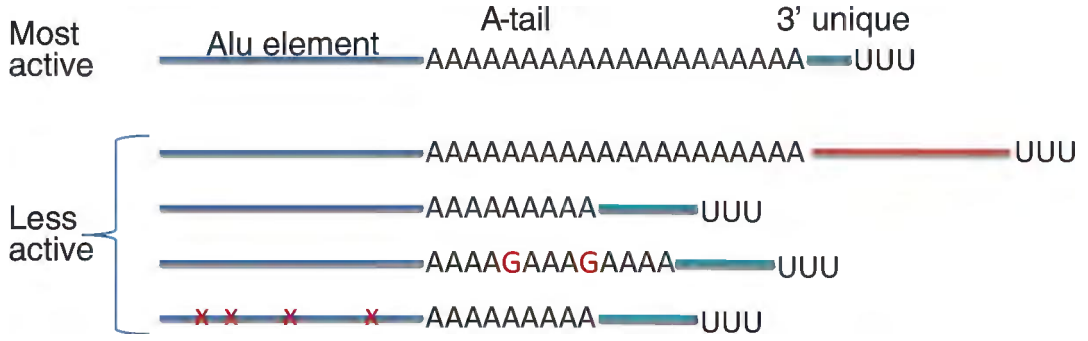

Figure 2. Why so few *Alu* elements are active. Out of the more than 1 million *Alu* elements in the human genome, very few are capable of making copies, although many make transcripts. Upon insertion in a new locus, the factors that make a very active *Alu* element are the flanking sequences influencing the promoter, creating a short unique region. Active elements match the consensus *Alu* element fairly closely and they have a long and fairly perfect A-tail. Active elements degrade rapidly on an evolutionary time scale by A-tail shortening, heterogeneous base interruptions accumulating in the A-tail, and eventually by the accumulation of random mutations in the *Alu* element. At least some of these changes alter *Alu* activity through disruption of the various proteins binding to the RNA in the ribonucleoprotein (Figure 1b).

insertions fixed, respectively [41,42]. There are 110,000 lineage-specific insertions in the Rhesus macaque genome [43]. However, this estimate was measured over a longer period of time than the estimates for human and chimpanzee insertion rates. Thus, we are unable to compare rates over the same period of time. The orangutan has only acquired approximately 250 lineage-specific insertions in the last 12 million years [44], demonstrating a marked decrease in amplification rate in that lineage. L1 elements do not show a significant difference in their lineage-specific insertions between human, chimp and orangutan, and it therefore appears that changes in *Alu* source elements or other *Alu*-specific amplification changes have occurred to cause the slow rate in orangutan. Further studies from incomplete, large-scale analyses of other primate genomes [45] show that the overall rates of *Alu* insertion in the marmoset lineage were generally lower than towards the human lineage, supporting the idea that *Alu* amplification rates vary in a species-specific or lineage-specific manner. Subfamily analysis and these rate studies suggest that the bottleneck events that occur during speciation can result in altered levels of *Alu* activity, probably through fixation of different numbers or levels of activity of source elements.

Alu elements have an even larger impact than that provided by their insertional mutagenesis through their influence on genome instability by providing the most common source of homology for non-allelic homologous recombination events leading to disease [23,46]. The bioinformatics required to analyze these types of rearrangements from comparative genomic data is technically more difficult than characterizing insertions. However, studies of the human and chimpanzee genomes show that approximately 500 deletion events have occurred in both genomes (Figure 3) [47,48]. It has not been possible to assess the duplication events that are also caused by this type of recombination, but it is likely that there is approximately the same number of events, and these events have also been suggested to contribute to genomic inversions [49] and segmental duplications [50]. The lower number of apparent non-allelic *Alu*/*Alu* recombination events between human and chimpanzee relative to the number of *Alu* insertion events (Figure 3) suggests that the recombination events cause a stronger negative selection because there are many more *Alu* recombination events than insertions causing disease [23]. Thus, they contribute more to disease, but are less well fixed in the population. This is consistent with the relatively short length of the fixed deletions relative to the longer deletions commonly found associated with disease [46].

Alu elements are preferentially enriched in regions that are generally gene rich, whereas L1 elements are enriched in the gene-poor regions [1]. This also correlates with *Alu* elements being enriched in reverse G bands [51], as well as in G+C-rich genomic isochores [52]. However, younger *Alu* and L1 elements do not show much disparity in their locations, making it most likely that the differences in location are the result of losses of L1 and *Alu* elements in different genomic regions. It is easy to understand why the much larger L1 elements might have more negative selection when located in genes, making *Alu* elements much more stably maintained within the genes. It is more difficult to understand why *Alu* elements seem to be preferentially lost between genes over evolutionary time compared with L1. It is most likely that the tendency of *Alu* elements to participate in non-allelic homologous recombination events might allow loss of these elements when not under selection [53,54].

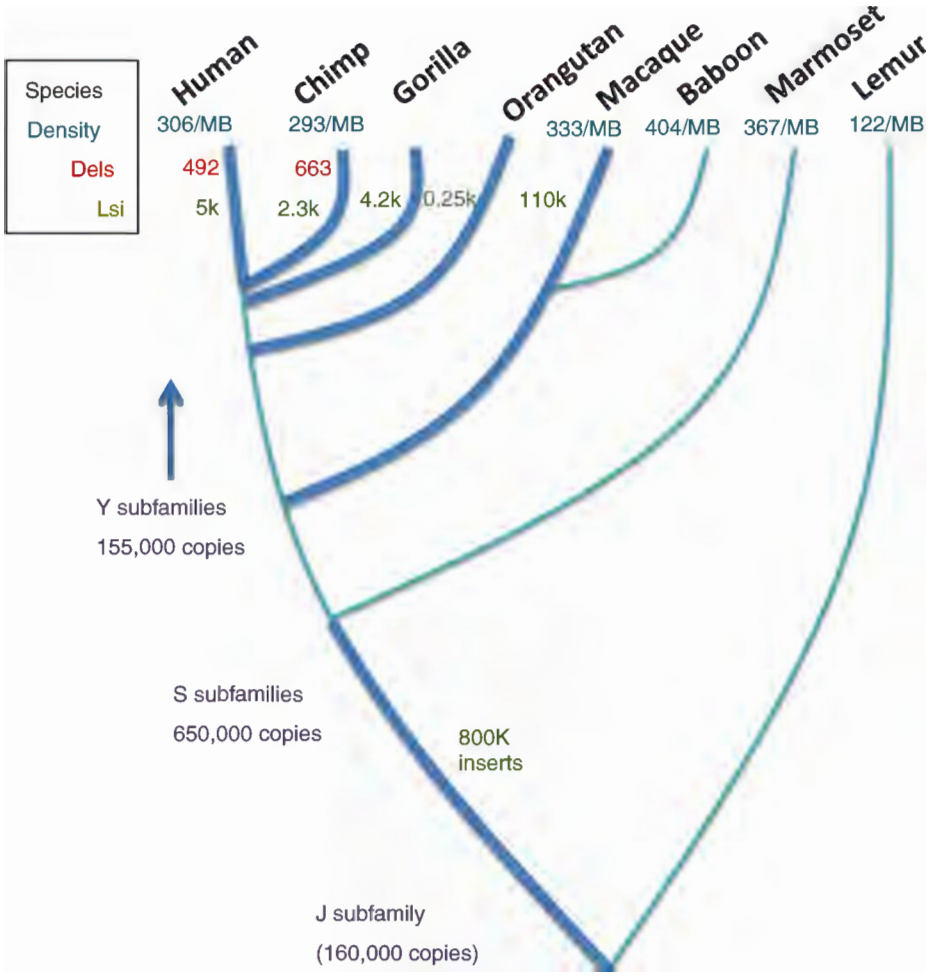

Figure 3. Evolutionary impact of *Alu* elements in primates. An approximate evolutionary tree is shown for various primate species. The approximate density of *Alu* elements in the genomes of those species is shown as the number of *Alu* elements per megabase (MB). For specific evolutionary time periods, marked by thicker lines, the number of lineage-specific *Alu* insertions (Lsi) is marked. Data of *Alu*/*Alu* recombination causing deletions (Dels) between the human and chimp genomes are also shown. Note that the rate of *Alu* insertion, as well as recombination, seems to vary with different lineages and different evolutionary time periods.

Human diversity

Alu elements have continued to insert in the modern human lineage as evidenced by their continued contribution to human genetic disease. It is estimated that there is about one new *Alu* insert per 20 human births [55], leading to about one in every 1,000 new human genetic diseases [23]. Comparison between two completed human genomes showed that there were approximately 800 polymorphic *Alu* elements between those two individuals [55].

Alu insertions contribute to disease by either disrupting a coding region or a splice signal [23,56] (Table 1). Although *Alu* element insertions causing disease are broadly spread throughout the genome, some genes seem more prone to disease-causing insertions of this type, particularly on the X chromosome. Fourteen new Alu insertions inactivating the *NF1* gene have been reported [57], representing 0.4% of known mutations in this gene. Similarly, many diseases caused by non-allelic homologous recombination between *Alu* elements have been discussed previously [23,57]. Although these events are also broadly spread throughout the genome, some regions, such as the *MSH2*, *VHL* and *BRCA1* genes, are much more subject to this instability than others [58]. Most *Alu*-related genomic instability events will either have no major functional consequence, and over many generations simply be lost from the human population gene pool through random fixation, or be deleterious and therefore lost through negative selection. Thus, the events described above represent only a tiny proportion of the overall genetic instability in the human population caused by such elements.

Table 1. *Alu* insertions in human disease

Locus	Chromosome	Subfamily	Disease	Reference[a]
3 × *HEMB* (IX)	X	Ya5, Ya5, Yb8	Hemophilia B	[23,56]
2 × *HEMA* (VIII)	X	Yb8, Yb9	Hemophilia A	[23,56]
2 × *CLCN5*	X	Ya6, Ya5	Dent's disease	[23,56]
2 × *BTK*	X	Y, Y	X-linked agammaglobulinemia	[23,56]
IL2RG	X	Ya5	X-linked severe combined immunodeficiency disease	[23,56]
GK	X	Yc1	Glycerol kinase deficiency	[23,56]
CD40LG	X	Yb8	Hyper IgM syndrome	[23,56]
ATP7A	X	Ya5a2	Menkes disease	[23,56]
CRB1	1	Y	Retinitis pigmentosa	[23,56]
ZFHX1B	2	Ya5	Mowat-Wilson syndrome	[23,56]
BCHE	3	Yb8	Cholinesterase deficiency	[23,56]
OPA1	3	Yb8	Autosomal dominant optic atrophy	[89]
CASR	3	Ya4	Hypocalciuric hypercalcemia and hyperparathyroidism	[23,56]
MLVI2	5	Ya5	Associated with leukemia	[23,56]
APC	5	Yb8	Hereditary desmoid disease	[23,56]
P5N1	7	Ya5	Chronic hemolytic anemia	[23,56]
EYA1	8	Y	Branchio-oto-renal syndrome	[23,56]
LPL	8	Yb9	Lipoprotein lipase deficiency	[23,56]
POMT1	9	Ya5	Walker Warburg syndrome	[23,56]
3 × *FGFR2*	10	Ya5, Yb8, Yc1	Apert's syndrome	[23,56]
TNFRSF6	10	Yb8	Autoimmune lymphoproliferative syndrome	[23,56]
C1NH	11	Yc1	Complement deficiency	[23,56]
AIP	11	Ya5	Acute intermittent porphyria	[23,56]
GNPTAB	12	Y	Mucolipidosis	[90]
3 × *BRCA2*	13	Ya5, Yc1, Y	Breast cancer	[91]
PMM2	16	Yb8	Congenital disorder of glycosylation type I	[92]
BRCA1	17	Ya5	Breast cancer	[23,56]
15 x *NF1*	17	Y subfamilies	Neurofibromatosis	[23,56,57]

Genomic studies are now beginning to delve into the diversity of *Alu* elements in the human population. Several studies involve the resequencing of multiple independent human genomes, resulting in the discovery of many new polymorphic *Alu* elements [59-61]. These studies largely confirm earlier work on the tremendous amount of diversity contributed to individual genomes by *Alu* insertions, as well as *Alu* subfamily types and distribution. These studies have utilized multiple available human genome sequences, primarily those available with low-to-moderate sequence coverage from the first 185 genomes from the 1000 Genomes Project. New, focused, next-generation sequencing (NGS) approaches seem very promising for looking at more specific questions about *Alu* activity. Among these approaches is a PCR method to isolate sequences flanking L1 or *Alu* sequences [62,63]. This approach isolated an additional 403 polymorphic *Alu* inserts from a number of individuals (also see a second method in the section Somatic insertions of *Alu* elements). The added sensitivity of these directed NGS approaches will aid in studies for detecting rare insertions in germline tissues, as well as for detecting somatic insertions present in only a few cells within an organ or tumor.

Somatic insertions of *Alu* elements

Almost all studies on *Alu* element activity have focused on germ line or tissue culture cell inserts [2,12,29,31]. However, there is reason to believe that *Alu* elements are also active in somatic tissues and may continue to contribute to genetic instability throughout the life of an individual, possibly leading to cancer or other age-related degenerations. The high levels of *Alu* insertion in tissue culture cells from transfected tagged constructs demonstrate that *Alu* is capable of retrotransposing in cells that are at least somewhat differentiated [2,29]. However, the

only way to demonstrate endogenous activity of *Alu* elements in tissues is by utilizing the power of high-throughput NGS technologies.

One NGS approach has claimed detection of somatic *Alu* elements. This approach uses hybrid selection with probes to *Alu* elements to enrich *Alu*-containing regions prior to NGS. DNA was sequenced from several brain regions, particularly the hippocampus, which has been reported to have higher levels of somatic L1 retrotransposition [64]. Using very deep sequencing, this study found evidence of thousands of individual *Alu* insertions. These studies were unable to quantify the relative insertion rate per cell. Each insertion is also extremely low in sequence coverage in these studies as if each one is specific to only a small proportion of cells within the tissue, consistent with insertion very late in the differentiation process. However, with so many of these rare insertions, these data suggest that there is a significant amount of genetic mosaicism created by the activity of mobile elements. A feature of note for the somatic *Alu* insertions was that there were apparently a large number of insertions of the older S subfamilies. This group of subfamilies is almost completely inactive in the human germ line, implying that the rules of *Alu* amplification [29,31] may differ between the somatic cells and the germ line. However, this study needs to be further substantiated, as the NGS reads are short and may have led to some misassignments or misinterpretations.

Alu elements in RNA molecules

Alu elements are extremely prevalent within RNA molecules, owing to their preference for gene-rich regions (Figure 4) [1]. The abundance of *Alu* elements within introns means that most primary nuclear transcripts (hnRNAs) will have *Alu* sequences located in one or both orientations. These will be found almost exclusively in the nucleus, but might represent a significant proportion of whole-cell RNA preparations and are likely to significantly contaminate cytoplasmic RNA preparations. *Alu* elements are also commonly found in the non-coding portion of the 3' exon of mRNAs: 5% to 10% of all mRNAs have *Alu* elements in their 3' ends.

The hnRNA and mRNA molecules described above are transcribed by RNA polymerase II and are not involved in the *Alu* amplification process. What is often not appreciated is that RNA polymerase III generated *Alu* transcripts are generally expressed at very low levels. It has been estimated that HeLa cells express about 100 molecules of *Alu* RNA (defined as RNA polymerase III generated) [65], although this could increase under various cellular stresses, including heat shock and viral infection [66]. By contrast, there are hundreds of thousands of mRNA molecules in each cell, and therefore tens of thousands of RNA polymerase II transcribed RNAs that contain *Alu* sequences. Thus, only a tiny proportion of *Alu*-containing RNAs in the cell are transcribed by RNA polymerase III. This makes it extremely difficult to measure and characterize the authentic *Alu* transcripts that might be involved in the amplification process relative to those that are just 'passengers' in other RNAs.

Given the technical challenges involved, it is not surprising that very few studies have looked properly at *Alu* RNA polymerase III transcripts. These studies have used either a primer extension approach to define the 5' end of the *Alu* transcript to prove that they were generated from RNA polymerase III rather than read-throughs of *Alu* elements in RNA polymerase II transcripts [67], or size fractionation combined with a 3' RACE (rapid amplification of cDNA ends) technique after *in vitro* tailing of the RNA to define the 3' end of the *Alu* RNA [33]. Any other traditional method of RNA characterization, such as northern blots, RT-PCR or cDNA cloning, is more likely to study either the closely related 7SL RNA (300 bp band in northern blots) or *Alu* elements included in RNA polymerase II transcripts, rather than those that might be transcribed by RNA polymerase III.

Many recent studies attempting to measure *Alu* RNA transcripts do not seem to be aware of the difficulties described above. Some groups using northern blots to look at *Alu* transcripts [67] have detected a band that is more likely to be 7SL rather than the expected smear of heterogeneous *Alu* transcripts. Similarly, investigators often do not realize that typical cDNA cloning approaches [68,69] or RT-PCR of *Alu* elements [70] are also unable to distinguish RNA polymerase III transcripts from those that are contained within RNA polymerase II transcripts (Figure 4). Thus, many claims regarding *Alu* non-coding RNAs probably reflect the inclusion of *Alu* elements in mRNAs.

Alu elements and gene regulation

Every time an *Alu* element inserts in or near a gene, it has the potential to influence expression of that gene in several ways. It is very likely that the majority of such influences would be under negative selection. Thus, only rarely would an *Alu* element insert and evolve in conjunction with a specific gene to truly become a regulator of that gene.

Alu elements are relatively rich in CpG residues, which appear to be widely subject to methylation and therefore are responsible for approximately 25% of all of the methylation in the genome [71]. Because methylated CpGs readily mutate to TpG, the higher density of methylation occurs in the younger elements. Methylation of *Alu* elements does vary in different tissues and appears to decrease in many tumors. It is likely that demethylation of an *Alu* increases expression from that *Alu* locus. It has

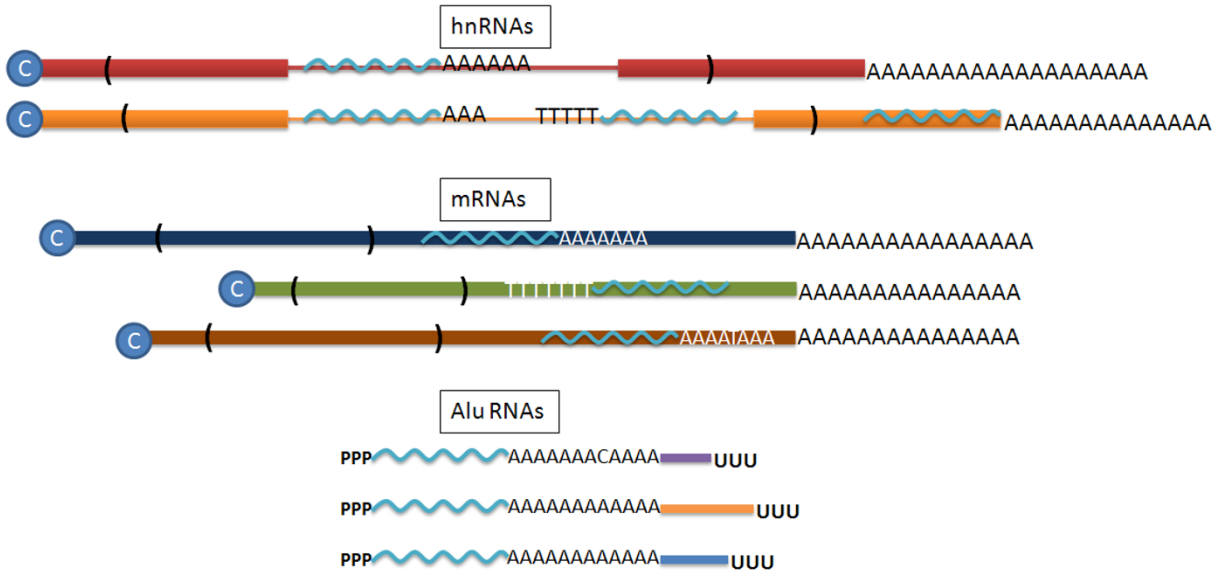

Figure 4. RNAs containing *Alu* sequences. Because of the high density of *Alu* elements found in introns, there are many *Alu* elements located (in both sense and antisense orientations) within the introns of the primary nuclear transcripts (hnRNAs) found in the nucleus. Even after splicing, a high proportion of mature cytoplasmic mRNAs also contain *Alu* elements in their 3' non-coding regions. Both of the above types of transcript involve capped (blue circle with C), polyadenylated RNAs where one or more *Alu* elements is included in either orientation within the transcript. The transcripts made from the *Alu* RNA polymerase III promoter are not capped, include a genomically encoded A-rich region, and terminate at a typical RNA pol III terminator. The transcripts are similar, but not identical, in the *Alu* region (wavy light blue line), with an A-tail encoded from the genome that is variable in length and often includes non-A bases. The 3' ends of each *Alu* RNA are unique, and arise from the flanking sequences of the different genomic *Alu* elements (Figure 1).

also been proposed that *Alu* elements might be a source of new CpG islands that could influence the regulation of nearby genes. However, studies to date do not make a clear case for *Alu* methylation being the driving force for nearby gene expression changes rather than the alternative, that *Alu* methylation is influenced by other nearby genome features.

Alu elements have also been found to host a number of transcription-factor-binding sites. Some of these binding sites are specific to certain *Alu* subfamilies, and some are also enhanced by changes that occur in *Alu* elements post-insertion. Dozens of different transcription-factor-binding sites have been predicted within subsets of *Alu* elements [72]. Although most of these are not validated, it does illustrate the opportunity for such sites to evolve at specific loci into regulatory elements. Sites that have used transcription-factor binding to demonstrate the association with *Alu* include several families of nuclear receptors [73-75], NF-kappaB [76] and p53 [77]. Thus, *Alu* elements have, at the least, a tremendous capacity to serve as a sink of bound transcription factors, and in limited specific cases have been found to influence expression of nearby genes.

The data are even more compelling for *Alu* elements to contribute to an array of post-transcriptional processes. These include providing polyadenylation sites [3,4], sites for alternative splicing [5-7] and sites for RNA editing [8-10] that then influences the fate of the RNA. *Alu* elements have two runs of A in their consensus sequence that can be readily mutated to the AATAAA consensus polyadenylation site. An analysis suggested that the modest bias for *Alu* elements in the reverse orientation to the gene in which they insert might be because of negative selection against the introduction of potential polyA sites [4]. This was further confirmed by a bioinformatic analysis demonstrating that a number of human genes utilize *Alu* sites to provide polyadenylation [3,78], including some that caused differences in human gene transcripts relative to chimpanzee [79].

Alternative splicing involving *Alu* elements is referred to as *Alu* 'exonization' [80] (Figure 5). This phenomenon is widespread, certainly affecting hundreds, if not thousands, of human genes. In some cases the exonized *Alu* RNA may make up a relatively minor portion of the transcripts from a gene, although in a study of human brain transcripts, hundreds of genes were found to have *Alu* exonization in the majority of their transcripts [5]. In general, the use of *Alu* sequences to generate alternative splicing seems to cause only decreased expression of the appropriate transcript. However, it appears that those alternative splices that survive over evolutionarily long periods of time become dominant and are more likely to

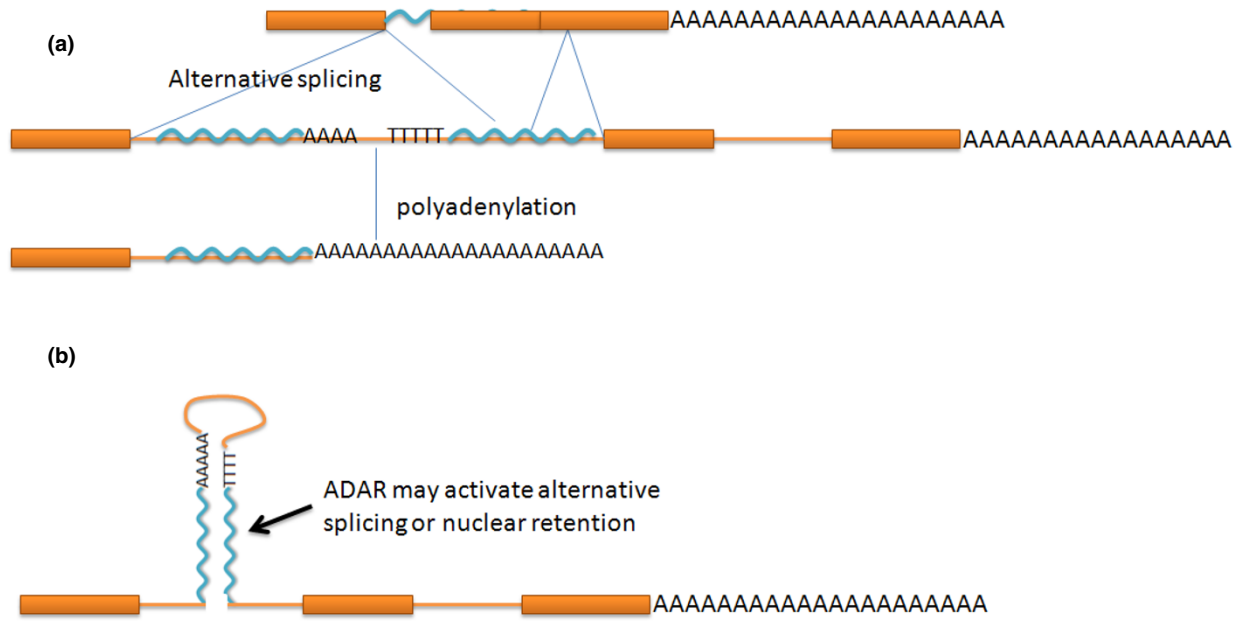

Figure 5. *Alu* elements and post-transcriptional processing of transcripts. (a) The majority of primary transcripts from genes contain *Alu* elements, both sense and antisense, within their introns. These *Alu* elements gradually accumulate mutations that can activate cryptic splice sites, or polyadenylation sites, within the *Alu*. This can lead to alternative splicing of RNAs that can either include a portion of an *Alu* in the coding region or result in premature termination of translation. Similarly, *Alu* elements may cause premature termination and polyadenylation resulting in truncated genes. **(b)** *Alu* elements in introns located in opposite orientations can fold into secondary structures that are then a major substrate for ADAR (adenosine deaminase that acts on RNA) activity. The edited RNAs may then have cryptic splice sites activated or may also result in retention of the RNA in the nucleus.

represent those transcripts that serve functionally [81]. *Alu* elements have only relatively weak, cryptic splice sites upon insertion. However, as elements accumulate more mutations, these sites can be further activated. There are also a number of cases where the evolution of a cryptic *Alu* splice site to a more functional form disrupts gene expression sufficiently to lead to disease [7]. A wide range of diseases are caused by this mechanism, and they include Alport syndrome, Leigh syndrome, chorioretinal degeneration and mucopolysaccharidosis VII. There are also two cases of Duchenne muscular dystrophy, probably because the *DMD* gene is so large and requires many splicing events. There are also examples of Alu exonization, where the *Alu* sequences require ADAR editing to become functional [6]. These are particularly prevalent in the brain, where ADAR activity is particularly high.

Alu elements appear to contribute to a relatively unique form of gene regulation involving ADARs [82]. These enzymes recognize RNAs with double-strand character and deaminate some adenosines to form inosines in those duplex regions. Most ADAR editing in cells occurs on primary transcripts in the nucleus in which two *Alu* elements in opposite orientations form a hairpin (Figure 5b). One of the major consequences of this editing process is the retention of transcripts in the nucleus [9]. Because ADAR is most prevalent in the brain, but also present in other tissues and tumors, it seems likely that this results in a tissue-specific alteration in RNA retention in the nucleus [8,82].

There have also been suggested associations between miRNAs and *Alu* elements. It has been suggested that the *Alu* promoter drives expression of sequences that can be processed into miRNAs [83]. However, at least in one case this has been suggested to be due to the co-presence of *Alu* and the miRNA in the intron of an hnRNA molecule, rather than a RNA polymerase III generated *Alu* RNA [84]. Additionally, some miRNAs appear to recognize *Alu* elements in other transcripts and may lead to regulation of the large number of transcripts with *Alu* elements in their 3' ends [5,85]. This regulation can be altered by RNA editing of the *Alu* elements, influencing the specificity of the regulation [86].

There are several cases where the RNA polymerase III transcribed *Alu* RNAs have been suggested to play roles in gene expression and function (that is, in response to stress [81]). It has similarly been suggested that the interaction of *Alu* RNAs with the RNA polymerase II molecule can attenuate transcription [87]. More recently it was reported that alterations in Dicer expression in

age-related macular degeneration would lead to increased accumulation of *Alu* RNAs that were responsible for the pathogenesis [88]. All of these studies are supported either by transient overexpression of *Alu* RNAs or *in vitro* studies. However, given the relatively low levels of endogenous *Alu* transcripts, even upon stress stimulation, it is not completely clear that the necessary levels of RNA to achieve these influences are made in cells.

Concluding remarks

The abundance of *Alu* elements in the human genome demonstrates that they have had a tremendous impact on insertional mutagenesis and evolution of the primate genome. Their distribution throughout the genome has acerbated that impact, supplying the primary sequences for non-allelic homologous recombination events throughout the genome. Extensive genomic sequencing efforts demonstrate that these forms of instability have not only resulted in major evolutionary changes in genomes, but continue to cause human diversity and contribute to human diseases. The ubiquity of *Alu* elements throughout the genome, and their enrichment in genes, has also led them to be inextricably mixed with a number of types of influence on gene expression and regulation. Many high-throughput studies have ignored *Alu* elements because of the technical difficulties in analyzing such high-copy-number elements. New NGS approaches are beginning to address the intricate relationships between *Alu* elements and other genomic features.

Abbreviations

ADAR, adenosine deaminase that acts on RNA; bp, base pair; FLAM, free left *Alu* monomer; FRAM, free right *Alu* monomer; hnRNA, primary nuclear transcript; miRNA, microRNA; NF, nuclear factor; NGS, next-generation sequencing; ORF, open reading frame; PABP, polyA-binding protein; RT-PCR, reverse transcriptase PCR.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

Thanks to Drs Astrid Engel and Victoria Belancio for helpful discussions and comments on the manuscript, and to Melanie Cross for editorial help. Dr Deininger's research on *Alu* elements is supported by a grant from the NIH (R01GM45668).

Published: 19 December 2011

References

1. 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, *et al.*: **Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium.** *Nature* 2001, **409:**860-921.
2. Dewannieux M, Esnault C, Heidmann T: **LINE-mediated retrotransposition of marked Alu sequences.** *Nat Genet* 2003, **35:**41-48.
3. Chen C, Ara T, Gautheret D: **Using *Alu* elements as polyadenylation sites: A case of retroposon exaptation.** *Mol Biol Evol* 2009, **26:**327-334.
4. Roy-Engel AM, El-Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H, Oyeniran OO, Deininger PL: **Human retroelements may introduce intragenic polyadenylation sites.** *Cytogenet Genome Res* 2005, **110:**365-371.
5. Shen S, Lin L, Cai JJ, Jiang P, Kenkel EJ, Stroik MR, Sato S, Davidson BL, Xing Y: **Widespread establishment and regulatory impact of *Alu* exons in human genes.** *Proc Natl Acad Sci U S A* 2011, **108:**2837-2842.
6. Sela N, Mersch B, Hotz-Wagenblatt A, Ast G: **Characteristics of transposable element exonization within human and mouse.** *PLoS One* 2010, **5:**e10907.
7. Vorechovsky I: **Transposable elements in disease-associated cryptic exons.** *Hum Genet* 2010, **127:**135-154.
8. Dominissini D, Moshitch-Moshkovitz S, Amariglio N, Rechavi G: **Adenosine-to-inosine RNA editing meets cancer.** *Carcinogenesis* 2011, **32:**1569-1577.
9. Chen LL, DeCerbo JN, Carmichael GG: ***Alu* element-mediated gene silencing.** *EMBO J* 2008, **27:**1694-1705.
10. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, Olshansky M, Rechavi G, Jantsch MF: **Systematic identification of abundant A-to-I editing sites in the human transcriptome.** *Nat Biotechnol* 2004, **22:**1001-1005.
11. Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr: **Mobile elements and genome evolution.** *Curr Opin Genet Dev* 2003, **136:**651-658.
12. Dewannieux M, Heidmann T: **Role of poly(A) tail length in *Alu* retrotransposition.** *Genomics* 2005, **86:**378-381.
13. Hsu K, Chang DY, Maraia RJ: **Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9.** *J Biol Chem* 1995, **270:**10179-10186.
14. West N, Roy-Engel A, Imataka H, Sonenberg N, Deininger P: **Shared protein components of SINE RNPs.** *J Mol Biol* 2002, **321:**423.
15. Muddashetty RS, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskothen J, Duning K, Barnekow A, Huttenhofer A, Tiedge H, Brosius J: **Poly(A) binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles.** *J Mol Biol* 2002, **321:**433-445.
16. Roy-Engel AM, Salem AH, Oyeniran OO, Deininger LA, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL: **Active Alu element "A-tails"; size does matter.** *Genome Res* 2002, **12:**1333-1344.
17. Boeke JD: **LINEs and Alus - the polyA connection.** *Nat Genet* 1997, **16:**6-7.
18. Batzer MA, Deininger PL: **Alu repeats and human genomic diversity.** *Nat Rev Genet* 2002, **3:**370-379.
19. Deininger P: *Alu Elements*. Totowa, NJ: Humana Press; 2006.
20. 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.
21. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM: **LINE-1 ORF1 protein enhances *Alu* SINE retrotransposition.** *Gene* 2008, **419:**1-6.
22. Belancio VP, Deininger PL, Roy-Engel AM: **LINE dancing in the human genome: transposable elements and disease.** *Genome Med* 2009, **1:**97.
23. Deininger PL, Batzer MA: ***Alu* repeats and human disease.** *Mol Genet Metab* 1999, **67:**183-193.
24. 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.
25. 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.
26. Hulme AE, Bogerd HP, Cullen BR, Moran JV: **Selective inhibition of *Alu* retrotransposition by APOBEC3G.** *Gene* 2007, **390:**199-205.
27. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR: **Cellular inhibitors of long interspersed element 1 and *Alu* retrotransposition.** *Proc Natl Acad Sci U S A* 2006, **103:**8780-8785.
28. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C: **APOBEC3 proteins inhibit human LINE-1 retrotransposition.** *J Biol Chem* 2006, **281:**22161-22172.
29. Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL: **Diverse *cis* factors controlling *Alu* retrotransposition: what causes *Alu* elements to die?** *Genome Res* 2009, **19:**545-555.
30. Deininger P, Batzer M, Maraia R: **SINE master genes and population biology.** In *The Impact of Short, Interspersed Elements (SINEs) on the Host Genome.* Georgetown, TX: R G Landes; 1995:43-60.
31. Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE: **Active *Alu* retrotransposons in the human genome.** *Genome Res* 2008, **18:**1875-1883.
32. Roy AM, West NC, Rao A, Adhikari P, Aleman C, Barnes AP, Deininger PL: **Upstream flanking sequences and transcription of SINEs.** *J Mol Biol* 2000, **302:**17-25.
33. Shaikh TH, Roy AM, Kim J, Batzer MA, Deininger PL: **cDNAs derived from primary and small cytoplasmic *Alu* (scAlu) transcripts.** *J Mol Biol* 1997,

271:222-234.
34. Hagan CR, Sheffield RF, Rudin CM: **Human *Alu* element retrotransposition induced by genotoxic stress.** *Nat Genet* 2003, **35:**219-220.
35. Ullu E, Tschudi C: ***Alu* sequences are processed 7SL RNA genes.** *Nature* 1984, **312:**171-172.
36. Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J: **Evolutionary history of 7SL RNA-derived SINEs in supraprimates.** *Trends Genet* 2007, **23:**158-161.
37. Quentin Y: **Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes.** *Nucleic Acids Res* 1992, **20:**487-493.
38. Shen M, Batzer M, Deininger P: **Evolution of the master *Alu* gene(s).** *J Mol Evol* 1991, **33:**311-320.
39. Deininger PL, Batzer MA, Hutchison CA, 3rd, Edgell MH: **Master genes in mammalian repetitive DNA amplification.** *Trends Genet* 1992, **8:**307-312.
40. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E: **Standardized nomenclature for *Alu* repeats.** *J Mol Evol* 1996, **42:**3-6.
41. Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, Devine SE: **Recently mobilized transposons in the human and chimpanzee genomes.** *Am J Hum Genet* 2006, **78:**671-679.
42. Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA: **Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages.** *Genome Res* 2004, **14:**1068-1075.
43. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, Huang CT, Sandifer E, Hebert K, Barnes EW, Hubley R, Miller W, Smit AF, Ullmer B, Batzer MA: **Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome.** *Science* 2007, **316:**238-240.
44. Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, Mitreva M, Cook L, Delehaunty KD, Fronick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Cree A, Dinh HH, Hume J, Kovar CL, Fowler GR, Lunter G, Meader S, Heger A, *et al.*: **Comparative and demographic analysis of orang-utan genomes.** *Nature* 2011, **469:**529-533.
45. Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE: **Comparative analysis of *Alu* repeats in primate genomes.** *Genome Res* 2009, **19:**876-885.
46. Hedges DJ, Deininger PL: **Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity.** *Mutat Res* 2007, **616:**46-59.
47. Han K, Lee J, Meyer TJ, Wang J, Sen SK, Srikanta D, Liang P, Batzer MA: ***Alu* recombination-mediated structural deletions in the chimpanzee genome.** *PLoS Genet* 2007, **3:**1939-1949.
48. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA: **Human genomic deletions mediated by recombination between *Alu* elements.** *Am J Hum Genet* 2006, **79:**41-53.
49. Lee J, Han K, Meyer TJ, Kim HS, Batzer MA: **Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons.** *PLoS One* 2008, **3:**e4047.
50. Bailey JA, Liu G, Eichler EE: **An *Alu* transposition model for the origin and expansion of human segmental duplications.** *Am J Hum Genet* 2003, **73:**823-834.
51. Korenberg JR, Rykowski MC: **Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands.** *Cell* 1988, **53:**391-400.
52. Costantini M, Bernardi G: **Mapping insertions, deletions and SNPs on Venter's chromosomes.** *PLoS One* 2009, **4:**e5972.
53. Abrusan G, Krambeck HJ: **The distribution of L1 and Alu retroelements in relation to GC content on human sex chromosomes is consistent with the ectopic recombination model.** *J Mol Evol* 2006, **63:**484-492.
54. Gasior SL, Preston G, Hedges DJ, Gilbert N, Moran JV, Deininger PL: **Characterization of pre-insertion loci of de novo L1 insertions.** *Gene* 2007, **390:**190-198.
55. 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, **19:**1516-1526.
56. 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.
57. Wimmer K, Callens T, Wernstedt A, Messiaen L: **The *NF1* gene contains hotspots for L1 endonuclease-dependent *de novo* insertion.** *PLoS Genet* 2011, 7:1-11.
58. Belancio VP, Roy-Engel AM, Deininger PL: **All y'all need to know 'bout retroelements in cancer.** *Semin Cancer Biol* 2010, **20:**200-210.
59. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, *et al.*: **Mapping copy number variation by population-scale genome sequencing.** *Nature* 2011, **470:**59-65.
60. Hormozdiari F, Alkan C, Ventura M, Hajirasouliha I, Malig M, Hach F, Yorukoglu D, Dao P, Bakhshi M, Sahinalp SC, Eichler EE: ***Alu* repeat discovery and characterization within human genomes.** *Genome Res* 2011, **21:**840-849.
61. Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, Urban AE, Grubert F, Lam HY, Lee WP, Busby M, Indap AR, Garrison E, Huff C, Xing J, Snyder MP, Jorde LB, Batzer MA, Korbel JO, Marth GT; 1000 Genomes Project: **A comprehensive map of mobile element insertion polymorphisms in humans.** *PLoS Genet* 2011, **7:**e1002236.
62. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE: **Natural mutagenesis of human genomes by endogenous retrotransposons.** *Cell* 2010, **141:**1253-1261.
63. Witherspoon DJ, Xing J, Zhang Y, Watkins WS, Batzer MA, Jorde LB: **Mobile element scanning (ME-Scan) by targeted high-throughput sequencing.** *BMC Genomics* 2010, **11:**410.
64. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH: **L1 retrotransposition in human neural progenitor cells.** *Nature* 2009, **460:**1127-1131.
65. Paulson KE, Schmid CW: **Transcriptional inactivity of *Alu* repeats in HeLa cells.** *Nucleic Acids Res* 1986, **14:**6145-6158.
66. Liu WM, Chu WM, Choudary PV, Schmid CW: **Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts.** *Nucleic Acids Res* 1995, **23:**1758-1765.
67. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA: **Human *Alu* RNA is a modular transacting repressor of mRNA transcription during heat shock.** *Mol Cell* 2008, **29:**499-509.
68. Macia A, Munoz-Lopez M, Cortes JL, Hastings RK, Morell S, Lucena-Aguilar G, Marchal JA, Badge RM, Garcia-Perez JL: **Epigenetic control of retrotransposon expression in human embryonic stem cells.** *Mol Cell Biol* 2011, **31:**300-316.
69. Kiesel P, Gibson TJ, Ciesielczyk B, Bodemer M, Kaup FJ, Bodemer W, Zischler H, Zerr I: **Possible editing of *Alu* transcripts in blood cells of sporadic Creutzfeldt-Jakob disease (sCJD).** *J Toxicol Environ Health A* 2011, **74:**88-95.
70. Rana T, Misra S, Mittal MK, Farrow AL, Wilson KT, Linton MF, Fazio S, Willis IM, Chaudhuri G: **Mechanism of down-regulation of RNA polymerase III-transcribed non-coding RNA genes in macrophages by *Leishmania*.** *J Biol Chem* 2011, **286:**6614-6626.
71. Xie H, Wang M, Bonaldo Mde F, Smith C, Rajaram V, Goldman S, Tomita T, Soares MB: **High-throughput sequence-based epigenomic analysis of *Alu* repeats in human cerebellum.** *Nucleic Acids Res* 2009, **37:**4331-4340.
72. Polak P, Domany E: ***Alu* elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes.** *BMC Genomics* 2006, **7:**133.
73. Norris J, Fan D, Aleman C, Marks JR, Futreal PA, Wiseman RW, Iglehart JD, Deininger PL, McDonnell DP: **Identification of a new subclass of *Alu* DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers.** *J Biol Chem* 1995, **270:**22777-22782.
74. Vansant G, Reynolds WF: **The consensus sequence of a major *Alu* subfamily contains a functional retinoic acid response element.** *Proc Natl Acad Sci U S A* 1995, **92:**8229-8233.
75. Cotnoir-White D, Laperriere D, Mader S: **Evolution of the repertoire of nuclear receptor binding sites in genomes.** *Mol Cell Endocrinol* 2011, **334:**76-82.
76. Antonaki A, Demetriades C, Polyzos A, Banos A, Vatsellas G, Lavigne M, Apostolou E, Mantouvalou E, Papadopoulou D, Mosialos G, Thanos D: **Genomic analysis reveals a novel NF-kappaB binding site in *Alu* repetitive elements.** *J Biol Chem* 2011, **286:**38768-82.
77. 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.
78. Lee JY, Ji Z, Tian B: **Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes.** *Nucleic Acids Res* 2008, **36:**5581-5590.
79. Kim DS, Hahn Y: **Identification of human-specific transcript variants induced by DNA insertions in the human genome.** *Bioinformatics* 2011, **27:**14-21.

80. Sorek R, Ast G, Graur D: ***Alu*-containing exons are alternatively spliced.** *Genome Res* 2002, **12:**1060-1067.
81. Lin L, Jiang P, Shen S, Sato S, Davidson BL, Xing Y: **Large-scale analysis of exonized mammalian-wide interspersed repeats in primate genomes.** *Hum Mol Genet* 2009, **18:**2204-2214.
82. Hogg M, Paro S, Keegan LP, O'Connell MA: **RNA editing by mammalian ADARs.** *Adv Genet* 2011, **73:**87-120.
83. Borchert GM, Lanier W, Davidson BL: **RNA polymerase III transcribes human microRNAs.** *Nat Struct Mol Biol* 2006, **13:**1097-1101.
84. Bortolin-Cavaille ML, Dance M, Weber M, Cavaille J: **C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts.** *Nucleic Acids Res* 2009, **37:**3464-3473.
85. Smalheiser NR, Torvik VI: ***Alu* elements within human mRNAs are probable microRNA targets.** *Trends Genet* 2006, **22:**532-536.
86. Borchert GM, Gilmore BL, Spengler RM, Xing Y, Lanier W, Bhattacharya D, Davidson BL: **Adenosine deamination in human transcripts generates novel microRNA binding sites.** *Hum Mol Genet* 2009, **18:**4801-4807.
87. Ponicsan SL, Kugel JF, Goodrich JA: **Genomic gems: SINE RNAs regulate mRNA production.** *Curr Opin Genet Dev* 2010, **20:**149-155.
88. Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, Karikó K, Yoo JW, Lee DK, Hadziahmetovic M, Song Y, Misra S, Chaudhuri G, Buaas FW, Braun RE, Hinton DR, Zhang Q, Grossniklaus HE, Provis JM, Madigan MC, Milam AH, Justice NL, Albuquerque RJ, Blandford AD, Bogdanovich S, Hirano Y, *et al.*: **DICER1 deficit induces *Alu* RNA toxicity in age-related macular degeneration.** *Nature* 2011, **471:**325-330.
89. Gallus GN, Cardaioli E, Rufa A, Da Pozzo P, Bianchi S, D'Eramo C, Collura M, Tumino M, Pavone L, Federico A: ***Alu*-element insertion in an OPA1 intron sequence associated with autosomal dominant optic atrophy.** *Mol Vis* 2010, **16:**178-183.
90. Tappino B, Regis S, Corsolini F, Filocamo M: **An *Alu* insertion in compound heterozygosity with a microduplication in *GNPTAB* gene underlies Mucolipidosis II.** *Mol Genet Metab* 2008, **93:**129-133.
91. Machado PM, Brandao RD, Cavaco BM, Eugenio J, Bento S, Nave M, Rodrigues P, Fernandes A, Vaz F: **Screening for a BRCA2 rearrangement in high-risk breast/ovarian cancer families: evidence for a founder effect and analysis of the associated phenotypes.** *J Clin Oncol* 2007, **25:**2027-2034.
92. Schollen E, Keldermans L, Foulquier F, Briones P, Chabas A, Sanchez-Valverde F, Adamowicz M, Pronicka E, Wevers R, Matthijs G: **Characterization of two unusual truncating PMM2 mutations in two CDG-Ia patients.** *Mol Genet Metab* 2007, **90:**408-413.

doi:10.1186/gb-2011-12-12-236
Cite this article as: Deininger P: ***Alu* elements: know the SINEs.** *Genome Biology* 2011, **12:**236.