Which transposable elements are active in the human genome?

Ryan E. Mills[1,2], E. Andrew Bennett[1,3], Rebecca C. Iskow[1,3] and Scott E. Devine[1,2,3]

[1] Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA
[2] Center for Bioinformatics, Emory University School of Medicine, Atlanta, GA 30322, USA
[3] Genetics and Molecular Biology Graduate Program, Emory University School of Medicine, Atlanta, GA 30322, USA

Although a large proportion (44%) of the human genome is occupied by transposons and transposon-like repetitive elements, only a small proportion (<0.05%) of these elements remain active today. Recent evidence indicates that ~35–40 subfamilies of *Alu*, L1 and SVA elements (and possibly HERV-K elements) remain actively mobile in the human genome. These active transposons are of great interest because they continue to produce genetic diversity in human populations and also cause human diseases by integrating into genes. In this review, we examine these active human transposons and explore mechanistic factors that influence their mobilization.

Introduction

Transposable genetic elements (TEs) are ubiquitous in both prokaryotes and eukaryotes [1]. TEs can mutate the genomes of their hosts either by 'jumping' to new locations or by facilitating chromosomal rearrangements through homologous recombination [1]. The human genome is no exception, and transposons have been documented to cause mutations that lead to human diseases, including cancers, through such mechanisms [2,3]. Several dozen disease-causing transposon insertions have been identified [4], and the full extent of transposon mutagenesis in humans is likely to extend far beyond these initial cases. Active human transposons have been estimated to generate about one new insertion per 10–100 live births [5–7]. Thus, humans harbor a large genetic load of recent transposon insertions along with several million fixed insertions [8]. 'Private' *de novo* insertions that occur only once in the human population (in just a single individual) are expected to represent a particularly abundant class of insertions. The full impact of these private insertions on human diversity and disease is only just beginning to be studied, and these insertions are likely to influence a range of human phenotypes. Therefore, it is of crucial importance to determine which endogenous human transposons remain active and continue to produce new insertions. Here, we examine these active transposons (and subfamilies) and explore several mechanistic factors that influence their mobilization.

The human genome harbors many transposon families and subfamilies

The human genome browser hosted by the University of California, Santa Cruz [9], currently contains over 4 million annotated transposon copies belonging to at least 848 families and subfamilies of elements (http://genome.ucsc.edu). These transposons collectively occupy almost half (44%) of the human genome and, thus, are major components of human genes and chromosomes (http://genome.ucsc.edu). Most of the largest transposon families in humans initially were identified as dispersed repetitive sequences that contain hallmark features of transposons (e.g. target site duplications, terminal repeats and transposases [10,11]). Subfamilies are defined by specific sets of sequence changes that can be useful for tracking the evolution and activity of elements (Box 1). Repbase [10] can be consulted for additional information on human transposons and their subfamilies (http://www.girinst.org/repbase/index.html).

Active transposons are dispersed among a vast genomic graveyard of dead copies

Alu and LINE-1 (L1) elements were first identified in the human genome during the late 1960s [12], but active *Alu* and L1 copies were not recognized until the 1980s, when they transposed into genes and caused human diseases [13,14]. Why did it take so long to recognize these active *Alu* and L1 transposons? Active copies of L1 were obscured by an overwhelming majority of inactive copies in the genome. Almost all of the several hundred thousand copies of L1 in the human genome are truncated at their 5′ ends [3]. Most L1s also have stop codons within the regions that encode the two proteins that drive L1 retrotransposition, ORF1p and ORF2p [15]. The coding capacity of L1 was not fully recognized until an active copy with intact ORFs 'jumped' into the factor VIII gene and caused hemophilia [14,16]. Thus, active L1 copies are rare in the human genome and were not easily recognized among the large collection of inactive L1 copies that have accumulated in humans. Active copies of *Alu* were even more difficult to recognize because *Alu* elements are nonautonomous elements that rely upon L1-encoded proteins for their own mobilization (see below).

Strategies for identifying recently mobilized transposon copies

Because most (or all) copies of a transposon class can be inactive, it has been necessary to develop targeted strategies to identify active and potentially active transposons in the human genome [17–21]. Similar to the example cited above in which a *de novo* L1 insertion in a gene signaled

Corresponding author: Devine, S.E. (sedevin@emory.edu).
Available online 27 February 2007.

0168-9525/$ – see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2007.02.006

Box 1. How are transposon subfamilies defined?

Transposon subfamilies are classified using unique sequence changes that serve as markers to indicate a common phylogeny among subfamily members. These diagnostic changes occur in a progenitor element and are copied to all new transposon insertions that are derived from this founder or its progeny. For the *Alu* Y family, new subfamilies are named using the number of nucleotide changes relative to the *Alu* Y consensus sequence. *Alu* Ya4, for example, contains four specific changes from the Y consensus. *Alu* Ya5 contains those same four changes plus one additional change. Letters following Y specify a particular cluster of additive changes that track the evolution of the active elements; both Ya4 and Yj4 contain four changes from the Y consensus sequence but they occur at different positions.

Similar methods have been used to define L1, SVA and HERV-K elements. The youngest L1 elements in humans are designated L1-Ta (Transcribed subset a [37]). This group is further divided into Ta-0, Ta-1nd and Ta-1d subfamilies on the basis of additional diagnostic nucleotide changes in these elements [37]. Six SVA subfamily consensus sequences, A–F, have been proposed [38]. SVA-A, -B, -C and -D predate the divergence of humans, chimps and gorillas. SVA-E and -F elements are only found in humans (these are the youngest SVA subfamilies). SVA-D is the largest subfamily, representing nearly half of all human SVAs. HERV-K subfamilies also have been identified.

transposon activity [14], several of these strategies have been directed at identifying recently mobilized transposon copies [17–21]. Active transposons often generate many recent insertions in genomes and such insertions can serve as surrogate indicators of activity. Therefore, the most active transposons in a genome are identified by assembling large collections of new transposon insertions. Recent activity does not necessarily indicate that a particular element remains active, however, and additional studies are necessary to determine whether a given candidate truly remains functional (see below).

One approach for identifying recently mobilized transposons in humans has been to compare the human and chimpanzee genomes to detect species-specific transposon insertions [17–19,22]. Element copies that are found in only one of the two species generally were mobilized during the past ~6 million years (the time since the last common ancestor of humans and chimps). More than 10 000 of these species-specific transposon insertions have been identified and most of these insertions (>95%) belong to *Alu*, L1 and SVA element subfamilies (Table 1 and Figure 1; SVA is an unusual composite element that was derived from three other repeats: SINE-R, VNTR and *Alu*). A relatively small number of species-specific human endogenous retrovirus K (HERV-K) copies also were identified (Table 1 and Figure 1). The remaining elements in the human genome generally lack species-specific copies and, therefore, have not been significantly active since the divergence of humans and chimps [17–19,22].

Comparative genomics also has been used to identify recently mobilized transposons in genetically diverse humans. For example, over 600 recent transposon insertions were identified by examining DNA resequencing traces from 36 genetically diverse humans [20]. An additional 800 *Alu* insertions were identified by comparing the full Celera genome sequence to the reference human genome sequence [21]. Likewise, polymerase chain reaction (PCR)-based assays (including transposon display assays) have been used extensively to screen for dimorphic transposon copies that are differentially present in genetically diverse humans (Table 1) [2,3,23–27]. These human–human comparisons have identified most of the same element families and subfamilies that were identified in the human–chimp comparisons (Table 1). A possible exception is SVA, for which only SVA-D, SVA-E and SVA-F elements were identified in both types of studies. Together, these studies indicate that 37 subfamilies of *Alu*, L1, SVA and HERV-K elements have been active in recent human history (Table 1).

Alu, L1 and SVA elements cause human diseases by jumping into genes

Several dozen transposon insertions have been shown to cause diseases by integrating into human genes (for a comprehensive review see Ref. [4]). We have examined the sequences of all element copies that have caused human diseases (including additional insertions that were not listed in Ref. [4]) and have reannotated them using custom Repbase libraries that include all available subfamilies of *Alu*, L1, SVA and HERV-K (Table 1). These insertions belong to the same *Alu*, L1 and SVA element subfamilies that were identified in the human–chimp and human–human comparisons described above. Thus, this third line of evidence indicates that *Alu*, L1 and SVA elements and their subfamilies have been the most active transposons in recent human history. To date, HERV-K insertions have not been documented to cause human diseases, suggesting that such insertions are rare or nonexistent.

L1 elements are active *in vitro*

The final test of whether an element remains actively mobile is to demonstrate functional transposition *in vivo* or in cell culture. In 1996 the Kazazian laboratory reported a cell-culture assay for L1 retrotransposition in human HeLa cells [28]. Two disease-causing L1 elements were engineered to carry a special intron-containing neomycin marker that served as a reporter gene for retrotransposition events. These marked L1 elements were placed on the extrachromosomal pCEP4 plasmid and tested for their ability to retrotranspose into the genome of HeLa cells. New chromosomal L1 insertions were identified that had the hallmark features of *bone fide* L1 retrotransposition events, including L1-like target site duplications (TSDs).

This assay was used systematically to test 89 'intact' full-length human L1 elements [29]. Almost half of these copies were found to be retrotransposition competent, but most L1 copies supported relatively low levels of retrotransposition. In fact, only six 'hot' L1 copies were shown to be responsible for the majority of L1 retrotransposition activity in humans [29]. These studies demonstrated that several L1 subfamilies (Ta-0, Ta-1d, Ta-1nd and pre-Ta elements) remain actively mobile in the human genome (Tables 1 and 2) [29]. Most of the same L1 subfamilies were identified in the comparative genomics studies described above [17–27] (Table 1). One possible exception is L1-PA2, which was greatly abundant in the chimp–human and human–human comparisons, but was not active in the

Table 1. Summary of recently mobilized transposons

Transposon family	Subfamily	Differentially present in humans and chimps	Dimorphic among humans	Disease-causing	Active in cell culture[a]
Alu	Sc	31 [19]	5 [20,44]	None found	NT
	Sg	56 [19]	8 [20,44]	None found	NT
	Sp	25 [19]	5 [20,21,44]	None found	NT
	Sq	46 [19]	3 [20,21,44,45]	1 [45]	NT
	Sx	46 [19]	7 [20,21,44]	None found	NT
	Sz	58 [19]	No [20,21,44]	1 [46]	NT
	Y	475 [19]	66 [20,21,44]	None found	NT
	Ya1	67 [19]	12 [20,21,44]	1 [47]	NT
	Ya4	170 [19]	54 [20,21,44]	None found	NT
	Ya5	1676 [19]	587 [13,20,21,23,24,44,48–53]	11 [3,13,45,54–61]	Yes [32]
	Ya5a2	38 [19]	9 [20,21,44]	None found	NT
	Ya8	36 [19]	9 [20,21,24,44,49,51,62]	None found	NT
	Yb3a1	17 [19]	10 [20,44]	1 [6]	NT
	Yb3a2	87 [19]	8 [20,44]	None found	NT
	Yb8	1290 [19]	409 [20,21,23,44,62–64]	4 [58,65–67]	NT
	Yb9	137 [19,22]	24 [20,21,44,52,63]	4 [68–71]	NT
	Yc1	356 [19]	113 [20,21,44,72]	4 [73–76]	NT
	Yc2	68 [19]	13 [20,21,44,52]	None found	NT
	Yd2	35 [19]	5 [44]	None found	NT
	Yd3	40 [77]	3 [21,44,77]	None found	NT
	Yd8	102 [19]	12 [20,21,44,77]	None found	NT
	Ye2	31 [19]	2 [20,44]	None found	NT
	Ye5	144 [19]	35 [20,44]	None found	NT
	Yf1	19 [19]	4 [20,21,44]	None found	NT
	Yg6	261 [19]	42 [20,21,25,44]	None found	NT
	Yh9	10 [19]	4 [21,44]	None found	NT
	Yi6	116 [19]	17 [20,21,25,44]	None found	NT
	Yj	10 [78]	6 [21,44]	None found	NT
LINE	L1-PA2	490 [17,19,22]	21 [20,29]	1 [79]	No [29]
	Pre-Ta	252 [19]	4 [20,29]	1 [14]	Yes [29]
	Ta[b]	270 [19]	101 [20,29]	9 [6,80–87]	Yes [29]
	Ta-0	43 [19]	2 [20,29]	1 [88]	Yes [29]
	Ta-1d	91 [19]	7 [20,29]	3 [14,89,90]	Yes [28,29]
	Ta-1nd	20 [19]	2 [20,29]	None found	Yes [29]
SVA	A	No [19]	No [20]	None found	NT
	B	5 [19]	No [20]	None found	NT
	C	15 [19]	No [20]	None found	NT
	D	259 [19,20]	5 [20]	None found	NT
	E	55 [19,20]	32 [20,21]	3 [91–93,95]	NT
	F	23 [19,20]	26 [20,21]	1 [94]	NT
HERV-K		64 [19]	8 [96]	None found	NT

[a] NT, not tested.

[b] These L1 Ta elements are 5′ truncated and lack the necessary 5′ diagnostic sequences to classify them further into Ta subfamilies.

cell-culture assays. Because only a single L1-PA2 element was tested in these assays, it is possible that some L1-PA2 copies remain active.

Brouha *et al.* identified the 89 full-length L1 elements described above from an early working draft of the human genome sequence [29]. We examined a more recent version of the genome sequence (build hg17) and identified 58 additional full-length L1 elements with intact open reading frames (ORFs) that belong to the same L1 subfamilies identified by Brouha *et al.* (bringing the total to 147 'intact' L1 copies in the reference sequence; Table 2 [19]). As in the Brouha *et al.* study, about half of these elements would be expected to be active. In addition to these full-length L1 elements, we also identified 80 solo ORF2 sequences in the human genome (i.e. L1 elements that have intact ORF2 sequences but lack intact ORF1 sequences). These solo ORF2 elements could potentially serve as drivers for *Alu* and perhaps other human transposons such as SVA, assuming that they could be expressed and translated (Table 2; see below). This hypothesis is somewhat speculative and requires additional validation.

Table 2. Summary of intact L1 transposons in the human genome

Intact L1s			
Subclass	**Brouha *et al.* [29]**	**Mills *et al.* [19][a]**	**Total**
Ta-1d	22	11	33
Ta-1nd	12	3	15
Ta-0	21	8	29
Pre-Ta	17	8	25
Pre-Ta (ACG/A)	2	1	3
Ta	13	21	34
L1-PA2	2	6	8
Total	**89**	**58**	**147**
Intact ORF2 only[b]			
Total	**0**	**80**	**80**

[a] Larger than 5500 nt and having an ORF1 and ORF2 within ±3 nt of hot L1 ORF sizes, respectively.

[b] Larger than 3800 nt and having an ORF2 within ±3 nt of hot L1 ORF2 size.

Sequence changes that influence L1 activity

A consensus sequence for active L1 elements has been developed using eight strongly active L1 copies [29]. Brouha *et al.* concluded that active L1 elements generally

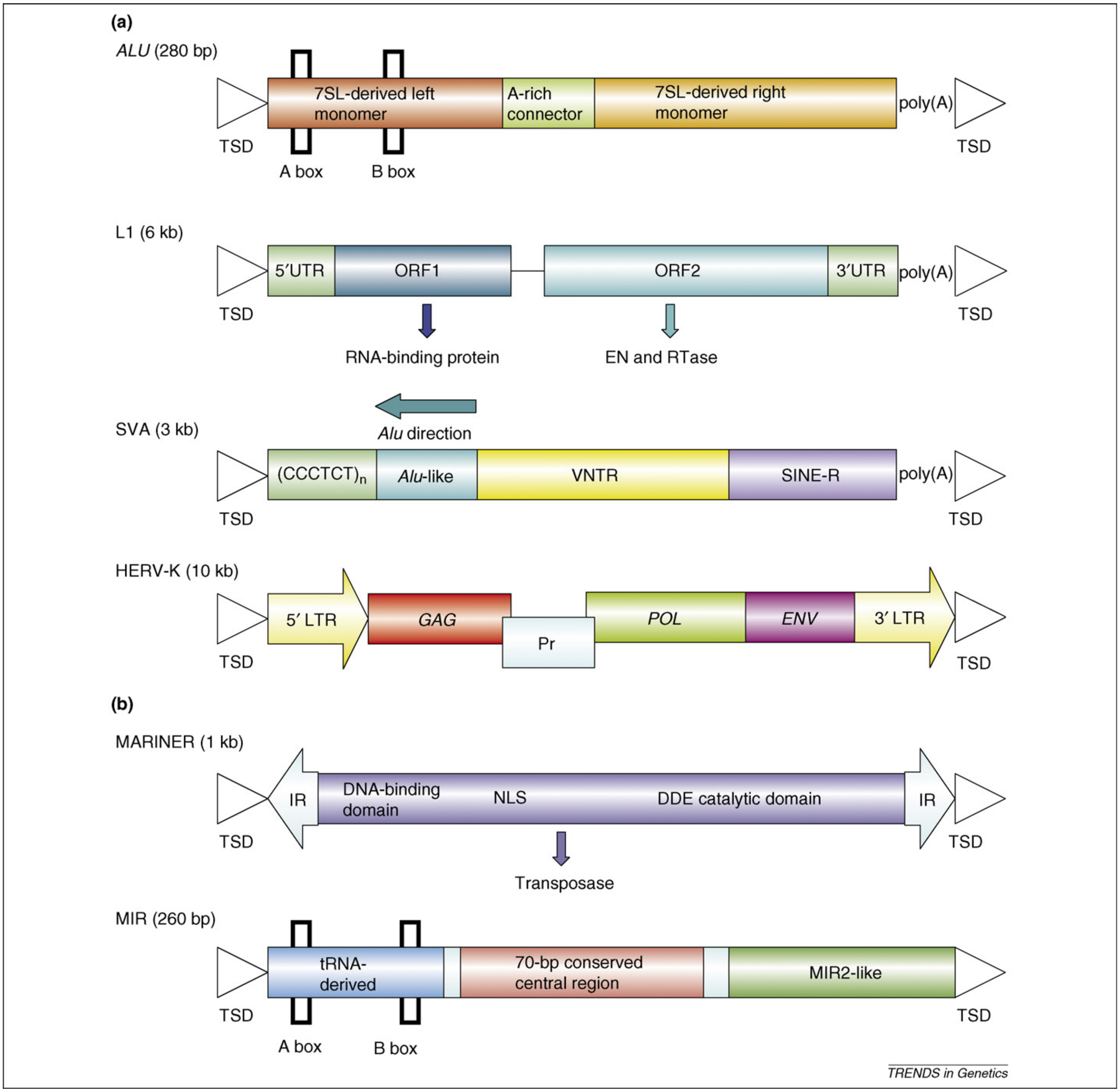

Figure 1. Structures of active and inactive human transposable elements. The structures of the four recently active human transposon classes are depicted **(a)**, along with two examples of inactive transposons **(b)**. For our purposes, 'recently active' means actively mobile sometime during the past ~6 million years. DDE, the conserved DDE sequence of the mariner transposase; IR; inverted repeat; LTR, long terminal repeat; MIR, mammalian-wide interspersed repeat; NLS, nuclear localization signal; ORF, open reading frame; Pr, protease; SINE-R, short interspersed repetitive element-R; TSD, target site duplication; UTR, untranslated region; VNTR, variable number of tandem repeats.

had sequences that closely resembled this consensus [29]. However, none of the active L1 copies in the human genome was identical to the consensus sequence. All active copies varied by at least two nucleotide positions and some active copies varied by as many as 25 nucleotide positions from this consensus [29]. Weak or inactive L1 elements generally had even more changes [29].

Other studies have revealed functionally important genetic variation in human L1 elements. Lutz *et al.* identified two alleles of the disease-causing L1.2 element (L1.2A and L1.2B) that differ by three nucleotide positions and consequently support dramatically different levels of retrotransposition [30]. Two of these changes caused amino acid substitutions within the ORF2 region [30]. Seleme *et al.* re-examined three hot L1s by resequencing these copies in diverse humans and found functionally important genetic variation [31]. In some cases, only a few changes occurred in the L1 element and the functionally important change(s) could be identified unambiguously. In many cases, however, individual copies of L1 encoded several nucleotide and/or amino acid changes making it difficult to sort out the specific changes that produced altered L1 activity. Thus, additional studies are needed to identify the specific changes in these elements that influence activity. Taken together, these studies indicate that natural genetic variation within L1 copies can have a major impact on L1 activity.

Alu elements are mobilized *in trans* by L1 elements

Because *Alu* elements are flanked by L1-like TSDs and lack substantial protein-coding capability, it seemed that *Alu* somehow was hijacking the L1 machinery to drive its own retrotransposition. Dewannieux *et al.* tested this hypothesis and showed that *Alu* is indeed mobilized *in trans* by the L1 machinery [32]. A young, disease-causing *Alu* Ya5 insertion was engineered to contain both a special RNA polymerase III promoter (derived from the 7SL gene) and an intron-containing neomycin marker. L1 proteins were then expressed from a separate, unmarked plasmid to drive the retrotransposition of the *Alu* Ya5 copy. L1-dependent *Alu* retrotransposition was shown in the presence of a driver plasmid that contained an active L1 element (a source of L1 proteins) but not with a plasmid that lacked L1 sequences. In HeLa chromosomes, new *Alu* retrotransposition events were observed that had the hallmark features of true *Alu* retrotransposition events, including L1-like TSDs. Interestingly, *Alu* retrotransposition required only the ORF2p of L1, and did not require ORF1p [32]. By contrast, L1 requires both ORF1p and ORF2p for its own retrotransposition using the *cis* mechanism [28]. Thus, the *trans* mechanism of retrotransposition uses the L1 machinery in a fundamentally different way than the *cis* mechanism that drives L1 itself.

Which *Alu* elements are active in humans?

Unlike L1 elements, which have been tested extensively, *Alu* elements have not been tested systematically for activity in cell-culture assays. In fact, only a single *Alu* Ya5 copy has been tested thus far [32], and additional studies are needed to determine whether other *Alu* subfamilies listed in Table 1 remain active. Current data indicate that 22 *Alu* Y and six Alu S subfamilies have been the most active *Alu* elements in recent human history (Table 1).

Sequence changes are commonly found in *Alu* Y element copies and we have exploited these changes to search for additional *Alu* Y elements that have been mobilized recently in the human genome. We reasoned that clusters of identical *Alu* Y copies with unique sets of sequence changes must have been mobilized recently (sufficiently recent that they have not had time to acquire new mutations that distinguish them from the other members of the clusters). We identified 49 clusters of at least five *Alu* Y elements that had identical sets of sequence changes. Although nine of these clusters were equivalent to named elements that had been described, 40 clusters had not been identified previously. Thus, in addition to the recently mobilized *Alu* copies in Table 1, these 40 additional *Alu* Y elements seem to have been amplified by recent bursts of retrotransposition.

The sequence changes in these clusters define positions within *Alu* Y that can be altered without abolishing retrotransposition. The 40 clusters mentioned above harbor a total of 99 changes that are dispersed throughout the *Alu* Y sequence. An additional 57 sequence changes are harbored by previously named *Alu* Y subfamilies (these changes also must be compatible with activity). Thus a total of 156 sequence changes (equivalent to 56% of the *Alu* Y sequence) have been identified at positions throughout the *Alu* Y sequence that are compatible with transposition (even within the A and B boxes that mediate *Alu* expression; Figure 2b,c). Therefore, it seems that *Alu* Y can sustain small changes throughout its sequence without losing the ability to transpose. Two small conserved regions were noted, however, on the left monomeric arm of *Alu* Y (Figure 2c), suggesting that some short sequences within *Alu* Y are under selective pressure. As a control experiment, we identified 316 *Alu* Y insertions that only occur once in the genome and have one additional (non-CpG) change from a known *Alu* Y subfamily. Some of these elements might have sustained mutations that are not compatible with retrotransposition, whereas others did not amplify for other reasons. These *Alu* Y ‘singletons’ had mutations throughout the *Alu* Y sequence (including the small areas of conservation observed above; Figure 2a). Taken together, these data indicate that selective pressure is acting to maintain only small regions within active *Alu* Y elements (stretches of only a few to several bases at most).

How does *Alu* hijack the L1 machinery?

Jef Boeke previously proposed an elegant model to explain why *Alu* RNAs have been amplified extensively by the L1 machinery, whereas most cellular RNAs have not [33] (Dewannieux *et al.* later presented a similar model [32]; Figure 3). In this model, *Alu* is docked on ribosomes and captures the L1 ORF2 protein as it is translated from an active L1 element mRNA. By capturing ORF2p at the ribosome, *Alu* can efficiently substitute its RNA for the normal L1 mRNA during the process of target primed reverse transcription (TPRT) that occurs at sites of integration on chromosomes. Importantly, this ribosomal docking is proposed to be mediated by two proteins, SRP9p and SRP14p (components of the Signal Recognition Particle), which bind to specific sequences on *Alu* and escort it to a docking site on the ribosome (Figure 3).

This model predicts that SRP binding sites should be conserved in active *Alu* elements, and our analysis supports this prediction (Figure 2c). The largest regions of conservation identified in our analysis of *Alu* Y clusters correspond precisely to sites in the left monomeric half of *Alu* that bind SRP9p and SRP14p, suggesting that these SRP-binding sites must be preserved to maintain *Alu* Y retrotransposition (Figure 2c). However, unlike the original model (which proposes two SRP sites), our data suggest that only the left monomer SRP-binding site is necessary for *Alu* Y retrotransposition (Figure 2c). This agrees with a previous observation that SRP9p–SRP14p binds strongly to the left monomer of *Alu* Y, but only weakly to the right monomer [34]. Otherwise, only short stretches of a few bases are conserved (Figure 2c). Thus, *Alu* is unlikely to contain a substrate sequence that is necessary for direct recognition by the L1 reverse transcriptase (other than the poly(A)+ tail). However, sequence changes at conserved positions might have an impact on *Alu* Y retrotransposition through structural changes [34]. Such structural changes could be envisioned to influence

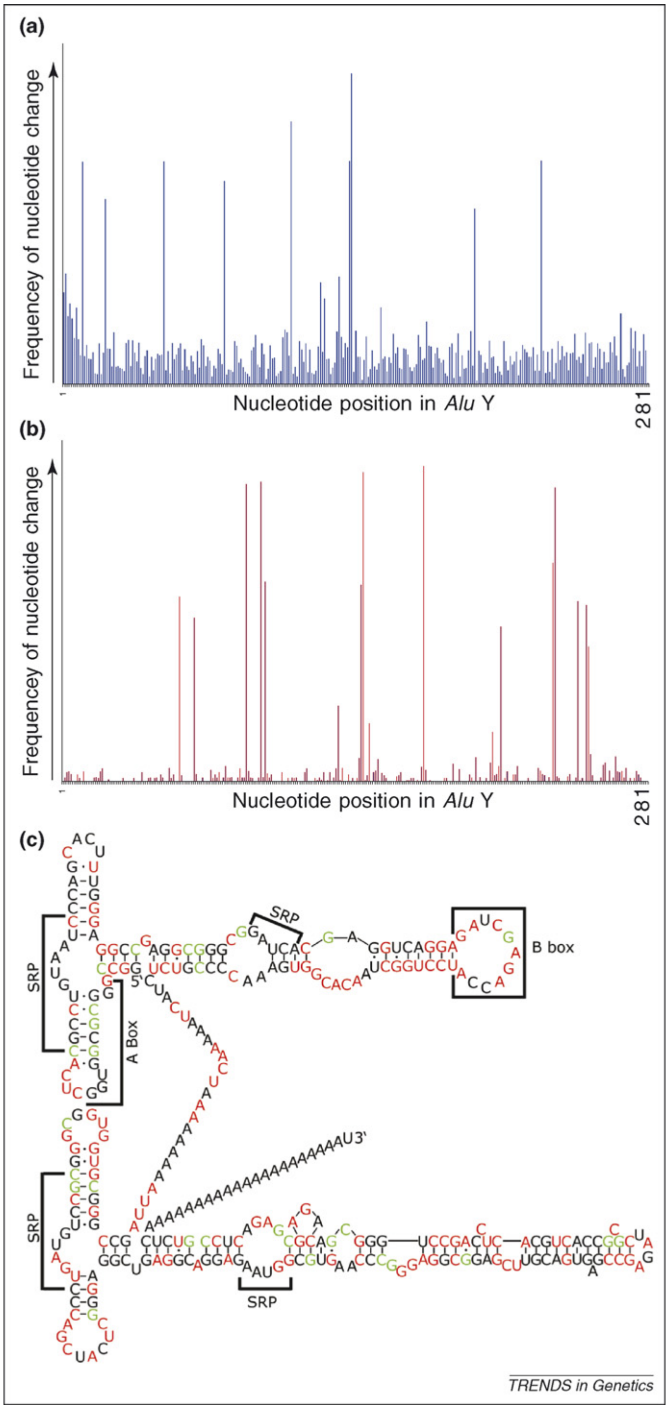

Figure 2. Variation in *Alu* Y sequences. Within the *Alu* Y sequence 156 nucleotide positions were identified that can be altered without losing *Alu* activity. These changes are dispersed throughout the *Alu* sequence, suggesting that most of the *Alu* Y sequence is not under strong selective pressure. **(a)** A total of 316 *Alu* Y insertions were identified that occur once in the genome and have one additional (non-CpG) change from a known *Alu* Y subfamily sequence. Some of these elements might have acquired mutations that are not compatible with retrotransposition, whereas others do not amplify for other reasons. As depicted in (a), these singleton *Alu* Y copies have changes throughout the Alu Y sequence. **(b)** Clusters (>5 copies) of *Alu* Y element copies in the human genome that had identical sequence changes. The nucleotide changes in these elements must be compatible with activity, because multiple copies with identical changes were generated in the genome. Thus, the nucleotide positions that are altered in these clusters define sequences within *Alu* that are not strictly necessary for function. Clusters defined by single CpG changes were omitted in (a) and (b) because CpG changes do not necessarily have to be caused by *Alu* Y amplification (they might have occurred independently in separate copies). **(c)** *Alu* Y secondary folding structure [43] showing permissible positional changes (red) from (b) and CpG changes (green) found in the clusters. The black positions represent sites that were not mutated in any of the active clusters or subfamilies in (b), and these positions could be under selective pressure. Thus, if mutated, such sites might be expected to inactivate *Alu* Y. The SRP binding sites that are conserved in 7SL are indicated along with the A and B Boxes of the *Alu* promoter.

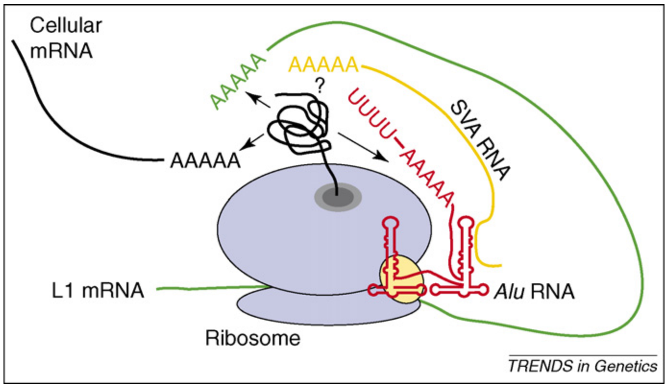

Figure 3. Model for L1, *Alu* and SVA retrotransposition. The model depicts possible scenarios for the *cis* and *trans* mechanisms of L1 retrotransposition. In the original model for *cis* preference, the L1 ORF proteins normally bind to the L1 mRNA (green) that makes them, as they are translated. The *trans* mechanism proposes that docking of *Alu* (red) on ribosomes enables it to hijack ORF2p (black line) as it is translated. This docking is proposed to be mediated by interactions with SRP9p and SRP14p in a manner that is analogous to SRP9p–SRP14p binding on 7SL (these binding sites are similar because *Alu* was derived from 7SL). It seems that only a single site for SRP9p–SRP14p (yellow) binding on the left monomer of *Alu* Y is required for this mechanism. SVA (orange) might use a similar mechanism, perhaps by first hybridizing to an active *Alu* RNA. The black mRNA represents all other cellular mRNAs that are not thought to serve as good substrates for the L1 machinery because they are not localized to ribosomes, whereas ORF2p is translated. The gray arrows indicate possible fates of ORF2p. Adapted with permission from Ref. [32].

SRP binding, ribosome docking, and/or the manner in which *Alu* RNA is presented to ORF2p on the ribosome. Thus, sequence changes at conserved nucleotide positions might inactivate *Alu* Y copies.

How does SVA hijack the L1 machinery?

Because SVA insertions have many of the hallmark features of L1-mediated retrotransposition events, SVA also is likely to be driven *in trans* by the L1 machinery [19,20,35]. A range of cellular RNAs can participate in the TPRT mechanism that normally drives L1 retrotransposition, including *Alu* and other mRNAs [32,36]. However, *Alu* serves as a much better substrate than most cellular RNAs. Thus, *Alu* has been amplified extensively by the L1 *trans* mechanism whereas most of the remaining cellular mRNAs have not. Like *Alu*, SVA RNA seems to hijack the L1 machinery more efficiently than most RNAs. But how might this happen? The ‘A’ of SVA actually stands for ‘Alu’, and the composite SVA element includes an unusual, rearranged *Alu* element at its 5′ end. So perhaps this *Alu* activates the SVA transcript for entry into the *trans* mechanism of L1 retrotransposition. However, because this 5′ *Alu* is positioned in reverse orientation, it would seem incapable of using the SRP-mediated mechanism described above for *Alu*. One possible model for SVA retrotransposition that is consistent with the SRP-mediated mechanism would be that these antisense *Alu* sequences in SVA serve as hybridization sites for active *Alu* RNAs. These stably hybridized *Alu* RNAs then, in turn, would escort SVA RNAs to ribosomes where they could efficiently participate in the *trans* mechanism of L1 retrotransposition (Figure 3).

Concluding remarks and future directions

The Human Genome Project has provided new resources to identify transposons that are moving around in our genomes. Recent studies indicate that ~35–40 subfamilies of *Alu*, L1, SVA and HERV-K elements have been actively mobile in recent human history. Most or all of these elements are likely to remain active today. Experimental systems that have been established to confirm the activity of these elements will continue to be useful for studying both the full scope of transposon activity in humans and the mobilization mechanisms of these elements.

Several questions remain.

(i) What is the full extent of *Alu* activity in the human genome? Clearly, it will be important to test the activities of additional *Alu* Y subfamilies along with representatives of *Alu* S and *Alu* J subfamilies (*Alu* J would be expected to be inactive). Likewise, it will be important to test the impact of *Alu* sequence variation on activity.

(ii) What is the full extent of L1 activity in the human genome? Although L1 has been studied extensively, the full scope of L1 activity remains unclear. The additional 58 copies of L1 that are described in Table 2 could be tested to determine which of these elements remain active. Human populations also are expected to harbor additional private 'hot' L1 alleles [31].

(iii) Does the human genome harbor active copies of SVA? It should be possible to test the hypothesis that SVA is driven in *trans* by L1.

(iv) Is HERV-K truly extinct? HERV-K has the lowest number of recent insertions, and full-length copies are variable in structure, suggesting that this family of elements became extinct recently [19,20] (Box 2).

Box 2. Phylogenetic reconstruction of active transposons

Most of the transposons in the human genome (those not listed in Table 1, main text) have become extinct (i.e. are no longer capable of transposition). Despite the apparent extinction of these elements, it might be possible to bring them back to life through phylogenetic reconstruction [39]. Successful transposon reconstruction projects were first accomplished for two extinct Tc1/mariner transposons, Sleeping Beauty [39] and Frog Prince [40]. Sleeping Beauty was 'awakened' from extinction using a combination of phylogenetic analysis to develop a model for an active element, followed by DNA engineering and testing to reconstruct an active copy [39]. A similar process was used to reconstruct Frog Prince, an active representative of an otherwise extinct transposon in the frog *Rana pipiens* [40].

Researchers have begun to apply this approach to the human genome. For example, an active human HERV-K retrotransposon was recently reconstructed from inactive copies using this approach [41]. HERV-K is dimorphic in humans (copies are differentially present in diverse humans) suggesting that it has been active recently (Table 1). HERV-K virus-like particles containing HERV-K mRNA also have been observed in cancer cell lines, further suggesting that active HERV-K copies reside in the human genome [42]. Several full-length HERV-K elements with significant ORFs (larger than several hundred base pairs in length) have been identified. Thus, HERV-K might either have become inactive relatively recently or be active today (active copies might be present in some humans). A phylogenetic model for an active HERV-K element was developed from known copies and an active synthetic HERV-K copy was produced [41]. This process, in principle, could now be applied to other extinct (or nearly extinct) elements in the human genome to study the basic transposition mechanisms of these elements and to learn how they have helped to shape the human genome.

One of the biggest challenges is to develop efficient methods to identify private transposon insertions in humans and to study the impact of these insertions on human biology. With a rate of one new insertion per ~10–100 live births [5–7], humans could have up to 60–600 million private transposon insertions (equivalent to one insertion per 5–50 bp of the human genome). This represents an impressive mutagenesis of the human genome, and these mutations are expected to influence a range of human phenotypes and diseases.

Finally, because all of the active and potentially active transposons that have been identified in humans are retrotransposons, these elements must be transcribed to generate new retrotransposition events. Even the hottest L1 element could never generate new L1 insertions if it was located in an unfavorable genomic location and was not expressed. Thus, the level of retrotransposition that is achieved by a given transposon copy reflects the sum of the transcription levels of the copy and the intrinsic activity as dictated by the sequence of the copy. It should be possible to examine these factors for each active copy in the genome.

Acknowledgements

This work was supported by NIH training grant 2T32GM00849 (R.C.I.) and NIH grant 5R01HG002898 (S.E.D.). We thank Shari Corin for helpful discussions and critical review of the manuscript, and Chris Luttig, Jackie Griffith and Steve Pittard for helpful discussions.

References

1 Craig, N.L. *et al.*, eds (2002) *Mobile DNA II*, ASM Press

2 Batzer, M.A. and Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379

3 Ostertag, E.M. and Kazazian, H.H., Jr (2001) Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35, 501–538

4 Chen, J.M. *et al.* (2005) Meta-analysis of gross insertions causing human genetic disease, novel mutational mechanisms and the role of replication slippage. *Hum. Mutat.* 25, 207–221

5 Cordaux, R. *et al.* (2006) Estimating the retrotransposition rate of human Alu elements. *Gene* 373, 134–137

6 Li, X. *et al.* (2001) Frequency of recent retrotransposition events in the human factor IX gene. *Hum. Mutat.* 17, 511–519

7 Kazazian, H.H. (1999) An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* 22, 130

8 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

9 Kent, W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.* 12, 996–1006

10 Jurka, J. (2000) Repbase update a database and an electronic journal of repetitive elements. *Trends Genet.* 16, 418–420

11 Smit, A.F.A. and Riggs, A.D. (1996) Tiggers and other DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1443–1448

12 Britten, R.J. and Kohn, D.E. (1968) Repeated sequences in DNA. *Science* 161, 529–540

13 Wallace, M.R. *et al.* (1991) A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature* 353, 864–866

14 Kazazian, H.H., Jr *et al.* (1988) Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164–166

15 Skowronski, J. *et al.* (1988) Unit length Line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* 8, 1384–1397

16 Mathias, S.L. *et al.* (1991) Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810

17 Hedges, D.J. *et al.* (2004) Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* 14, 1068–1075

18 The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87

19 Mills, R.E. *et al.* (2006) Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* 78, 671–679

20 Bennett, E.A. *et al.* (2004) Natural genetic variation caused by transposable elements in humans. *Genetics* 168, 933–951

21 Wang, J. *et al.* (2006) Whole genome computational comparative genomics, a fruitful approach for ascertaining Alu insertion polymorphisms. *Gene* 365, 11–20

22 Watanabe, H. *et al.* (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* 429, 382–388

23 Carroll, M.L. *et al.* (2001) Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* 311, 17–40

24 Batzer, M.A. *et al.* (1996) Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.* 42, 22–29

25 Salem, A.H. *et al.* (2003) Recently integrated Alu elements and human genomic diversity. *Mol. Biol. Evol.* 20, 1349–1361

26 Badge, R.M. *et al.* (2003) ATLAS: A system to selectively identify human-specific L1 insertions. *Am. J. Hum. Genet.* 72, 823–838

27 Myers, J.S. *et al.* (2002) A comprehensive analysis of recently integrated human Ta L1 elements. *Am. J. Hum. Genet.* 71, 312–326

28 Moran, J.V. *et al.* (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917–927

29 Brouha, B. *et al.* (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5280–5285

30 Lutz, S.M. *et al.* (2003) Allelic heterogeneity in LINE-1 retrotransposition activity. *Am. J. Hum. Genet.* 73, 1431–1437

31 Seleme Mdel, C. *et al.* (2006) Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6611–6616

32 Dewannieux, M. *et al.* (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48

33 Boeke, J.D. (1997) LINEs and Alus – the polyA connection. *Nat. Genet.* 16, 6–7

34 Sarrowa, J. *et al.* (1997) 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

35 Ostertag, E.M. *et al.* (2003) SVA is a nonautonomous retrotransposon that causes diseases in humans. *Am. J. Hum. Genet.* 73, 1444–1451

36 Wei, W. *et al.* (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439

37 Boissinot, S. *et al.* (2000) L1 (LINE1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* 17, 915–928

38 Wang, H. *et al.* (2005) SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* 354, 994–1007

39 Ivics, Z. *et al.* (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91, 501–510

40 Miskey, C. *et al.* (2003) The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic Acids Res.* 31, 6873–6881

41 Dewannieux, M. *et al.* (2006) Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.* 16, 1548–1556

42 Brodsky, I. *et al.* (1993) Expression of HERV-K proviruses in human leukocytes. *Blood* 81, 2369–2374

43 Hasler, J. and Strub, K. (2006) Alu elements as regulators of gene expression. *Nucleic Acids Res.* 34, 5491–5497

44 Wang, J. *et al.* (2006) dbRIP: A highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* 27, 323–329

45 Teugels, E. *et al.* (2005) *De novo* Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Hum. Mutat.* 26, 284

46 Kloor, M. *et al.* (2004) A large MSH2 Alu insertion mutation causes HNPCC in a German kindred. *Hum. Genet.* 115, 432–438

47 Janicic, N. *et al.* (1995) Insertion of an Alu sequence in the Ca(2+)-sensing receptor gene in familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Am. J. Hum. Genet.* 56, 880–886

48 Arcot, S.S. *et al.* (1997) Identification and characterization of two polymorphic Ya5 Alu repeats. *Mutat. Res.* 382, 5–11

49 Watkins, W.S. *et al.* (2001) Patterns of ancestral human diversity, an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* 68, 738–752

50 Otieno, A.C. *et al.* (2004) Analysis of the human Alu Ya-lineage. *J. Mol. Biol.* 342, 109–118

51 Mamedov, I.Z. *et al.* (2005) Whole-genome experimental identification of insertion/deletion polymorphisms of interspersed repeats by a new general approach. *Nucleic Acids Res.* 33, e16

52 Roy, A.M. *et al.* (2000) Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* 10, 1485–1495

53 Economou-Pachnis, A. and Tsichlis, P.N. (1985) Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. *Nucleic Acids Res.* 13, 8379–8387

54 Abdelhak, S. *et al.* (1997) Clustering of mutations responsible for branchio-oto-renal BOR syndrome in the eyes absent homologous region eyaHR of EYA1. *Hum. Mol. Genet.* 16, 2247–2255

55 Claverie-Martin, F. *et al.* (2003) *De novo* insertion of an Alu sequence in the coding region of the CLCN5 gene results in Dent's disease. *Hum. Genet.* 113, 480–485

56 Ishihara, N. *et al.* (2004) Clinical and molecular analysis of Mowat-Wilson syndrome associated with ZFHX1B mutations and deletions at 2q22-q24.1. *J. Med. Genet.* 41, 387–393

57 Mustajoki, S. *et al.* (1999) Insertion of Alu element responsible for acute intermittent porphyria. *Hum. Mutat.* 13, 431–438

58 Oldridge, M. *et al.* (1999) *De novo* alu-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. *Am. J. Hum. Genet.* 64, 446–461

59 Tighe, P.J. *et al.* (2002) Inactivation of the Fas gene by Alu insertion, retrotransposition in an intron causing splicing variation and autoimmune lymphoproliferative syndrome. *Genes Immun.* 3 (Suppl. 1), S66–S70

60 Vidaud, D. *et al.* (1993) Haemophilia B due to a *de novo* insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* 1, 30–36

61 Wulff, K. *et al.* (2000) Identification of a novel large F9 gene mutation-an insertion of an Alu repeated DNA element in exon e of the factor 9 gene. *Hum. Mutat.* 15, 299

62 Batzer, M.A. *et al.* (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *J. Mol. Biol.* 247, 418–427

63 Carter, A.B. *et al.* (2004) Genome-wide analysis of the human Alu Yb-lineage. *Hum. Genomics* 1, 167–178

64 Callinan, P.A. *et al.* (2003) Comprehensive analysis of Alu-associated diversity on the human sex chromosomes. *Gene* 317, 103–110

65 Halling, K.C. *et al.* (1999) Hereditary desmoid disease in a family with a germline Alu I repeat mutation of the APC gene. *Hum. Hered.* 49, 97–102

66 Sukarova, E. *et al.* (2001) An Alu insert as the cause of a severe form of hemophilia A. *Acta Haematol.* 106, 126–129

67 Sobrier, M.L. *et al.* (2005) Alu-element insertion in the homeodomain of HESX1 and aplasia of the anterior pituitary. *Hum. Mutat.* 25, 503

68 Ganguly, A. *et al.* (2003) Exon skipping caused by an intronic insertion of a young Alu Yb9 element leads to severe hemophilia *A. Hum. Genet.* 113, 348–352

69 Muratani, K. *et al.* (1991) Inactivation of the cholinesterase gene by Alu insertion, possible mechanism for human gene transposition. *Proc. Natl. Acad. Sci. U. S. A.* 88, 11315–11319

70 Kutsche, K. *et al.* (2002) Characterization of breakpoint sequences of five rearrangements in L1CAM and ABCD1 (ALD) genes. *Hum. Mutat.* 19, 526–535

71 Su, L.K. *et al.* (2000) Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis. *Hum. Genet.* 106, 101–107

72 Garber, R.K. *et al.* (2005) The Alu Yc1 subfamily: sorting the wheat from the chaff. *Cytogenet. Genome Res.* 110, 537–542

73 Conley, M.E. *et al.* (2005) Two independent retrotransposon insertions at the same site within the coding region of BTK. *Hum. Mutat.* 25, 324–325

74 Miki, Y. *et al.* (1996) Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* 13, 245–247

75 Stoppa-Lyonnet, D. *et al.* (1990) Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. *Proc. Natl. Acad. Sci. U. S. A.* 87, 1551–1555

76 Zhang, Y. *et al.* (2000) AluY insertion (IVS4-52ins316alu) in the glycerol kinase gene from an individual with benign glycerol kinase deficiency. *Hum. Mutat.* 15, 316–323

77 Xing, J. *et al.* (2003) Comprehensive analysis of two Alu Yd subfamilies. *J. Mol. Evol.* 57 (Suppl. 1), S76–S89

78 Park, E.S. *et al.* (2005) Analysis of newly identified low copy AluYj subfamily. *Genes Genet. Syst.* 80, 415–422

79 Narita, N. *et al.* (1993) Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* 91, 1862–1867

80 Martinez-Garay, I. *et al.* (2003) Intronic L1 insertion and F268S, novel mutations in RPS6KA3 (RSK2) causing Coffin-Lowry syndrome. *Clin. Genet.* 64, 491–496

81 Holmes, S.E. *et al.* (1994) A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.* 7, 143–148

82 Mukherjee, S. *et al.* (2004) Molecular pathology of haemophilia B, identification of five novel mutations including a LINE 1 insertion in Indian patients. *Haemophilia* 10, 259–263

83 van den Hurk, J.A. *et al.* (2003) Novel types of mutation in the choroideremia (CHM) gene, a full-length L1 insertion and an intronic mutation activating a cryptic exon. *Hum. Genet.* 113, 268–275

84 Kondo-Iida, E. *et al.* (1999) Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). *Hum. Mol. Genet.* 8, 2303–2309

85 Miki, Y. *et al.* (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645

86 Meischl, C. *et al.* (2000) A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Eur. J. Hum. Genet.* 8, 697–703

87 Yoshida, K. *et al.* (1998) Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Hum. Mol. Genet.* 7, 1129–1132

88 Divoky, V. *et al.* (1996) A novel mechanism of beta-thalassemia. The insertion of L1 retrotransposable element into beta globin IVSII. *Blood* 88, 148a

89 Brouha, B. *et al.* (2002) Evidence consistent with human L1 retrotransposition in maternal meiosis I. *Am. J. Hum. Genet.* 71, 327–336

90 Schwahn, U. *et al.* (1998) Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat. Genet.* 19, 327–332

91 Hassoun, H. *et al.* (1994) A novel mobile element inserted in the alpha spectrin gene, spectrin Dayton. A truncated alpha spectrin associated with hereditary elliptocytosis. *J. Clin. Invest.* 94, 643–648

92 Kobayashi, K. *et al.* (1998) Founder-haplotype analysis in Fukuyama-type congenital muscular dystrophy (FCMD). *Hum. Genet.* 103, 323–327

93 Wilund, K.R. *et al.* (2002) Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum. Mol. Genet.* 11, 3019–3030

94 Rohrer, J. *et al.* (1999) Unusual mutations in Btk, an insertion, a duplication, an inversion, and four large deletions. *Clin. Immunol.* 90, 28–37

95 Kobayashi, K. *et al.* (1998) An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 394, 388–392

96 Belshaw, R. *et al.* (2005) Genomewide screening reveals high levels of insertional polymorphisms in the Human Endogenous Retrovirus Family HERV-K (HML-2): implications for present-day activity. *J. Virol.* 79, 12507–12514