# The impact of retrotransposons on human genome evolution

*Richard Cordaux* and Mark A. Batzer‡*

Abstract | Their ability to move within genomes gives transposable elements an intrinsic propensity to affect genome evolution. Non-long terminal repeat (LTR) retrotransposons — including LINE-1, *Alu* and SVA elements — have proliferated over the past 80 million years of primate evolution and now account for approximately one-third of the human genome. In this Review, we focus on this major class of elements and discuss the many ways that they affect the human genome: from generating insertion mutations and genomic instability to altering gene expression and contributing to genetic innovation. Increasingly detailed analyses of human and other primate genomes are revealing the scale and complexity of the past and current contributions of non-LTR retrotransposons to genomic change in the human lineage.

**Long terminal repeats**
Sequences of 300–1,000 bp that are directly repeated at the 5′ and 3′ ends of long terminal repeat retrotransposons and retroviruses.

**SVA element**
An element that is made up of a short interspersed element (SINE) region, a variable number of tandem repeats (VNTR) region and an *Alu*-like region.

**Université de Poitiers, CNRS UMR 6556 Ecologie, Evolution, Symbiose, 40 Avenue du Recteur Pineau, 86022 Poitiers, France.*
*‡Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA.*
*Correspondence to M.A.B.*
*e-mail: mbatzer@lsu.edu*
doi:10.1038/nrg2640

Transposable elements (TEs; also known as 'jumping genes') are discrete pieces of DNA that can move within (and sometimes between) genomes. Although they were discovered in the 1940s[1], it was approximately half a century before scientists began to understand how TEs interact with their genomic environment. Crucially, the completion of the first human genome sequence revealed that nearly half of our genome is derived from TEs[2,3] (FIG. 1a). This is likely to be an underestimate, as many ancient TEs in the human genome have probably diverged beyond recognition[3]. The scale of the contribution of TEs to the human genome is all the more remarkable when one considers that protein-coding regions account for just 1.5% of the human genome[3].

TEs can be separated into two major classes: DNA transposons and retrotransposons. DNA transposons, which constitute ~3% of the human genome (FIG. 1a), can excise themselves from the genome, move as DNA and insert themselves into new genomic sites[4]. Although DNA transposons are currently not mobile in the human genome, they were active during early primate evolution until ~37 million years (Myr) ago[5]. Retrotransposons duplicate through RNA intermediates that are reverse transcribed and inserted at new genomic locations[4]. Retrotransposons can be subdivided into two groups distinguished by the presence or absence of long terminal repeats (LTRs). Human LTR elements are endogenous retroviruses (HERVs), which along with related elements account for ~8% of the genome (FIG. 1a). Most HERVs inserted into the human genome >25 Myr ago, and their activity is presently very limited in humans, if it occurs at all[3,6]. By contrast, the majority of human TEs result from the present and past activity of non-LTR retrotransposons, including the LINE-1 (long interspersed element 1, abbreviated here to L1), *Alu* and SVA elements, which collectively account for approximately one-third of the human genome[3] (FIG. 1a). L1, *Alu* and SVA non-LTR retrotransposons are the only TEs that have unequivocally been shown to be currently active in humans, as indicated by the more than 60 reported cases of *de novo* insertions that are responsible for genetic disorders[7–11].

The extremely high density of TEs in our genome poses the question: what impact have they had on human evolution? The development of innovative molecular methodologies — such as retrotransposition assays in cultured cells[12,13] and computational techniques for comparative genomics — in conjunction with the sequencing of multiple primate genomes (such as the human[3], chimpanzee[14] and rhesus macaque[15] genomes) has resulted in a progressive shift in the focus of TE research towards investigating the extent of the impact of TE activity on genomic evolution. Important discoveries regarding how TEs affect human genome evolution have recently made it possible to quantify the overall impact that TE activity has had on our genome. For example, although it has long been recognized that recombination between TEs can trigger genomic deletions in humans (these deletions have caused several genetic disorders[8]), only recently have genome-wide comparisons of human and other primate genomes allowed us to determine the



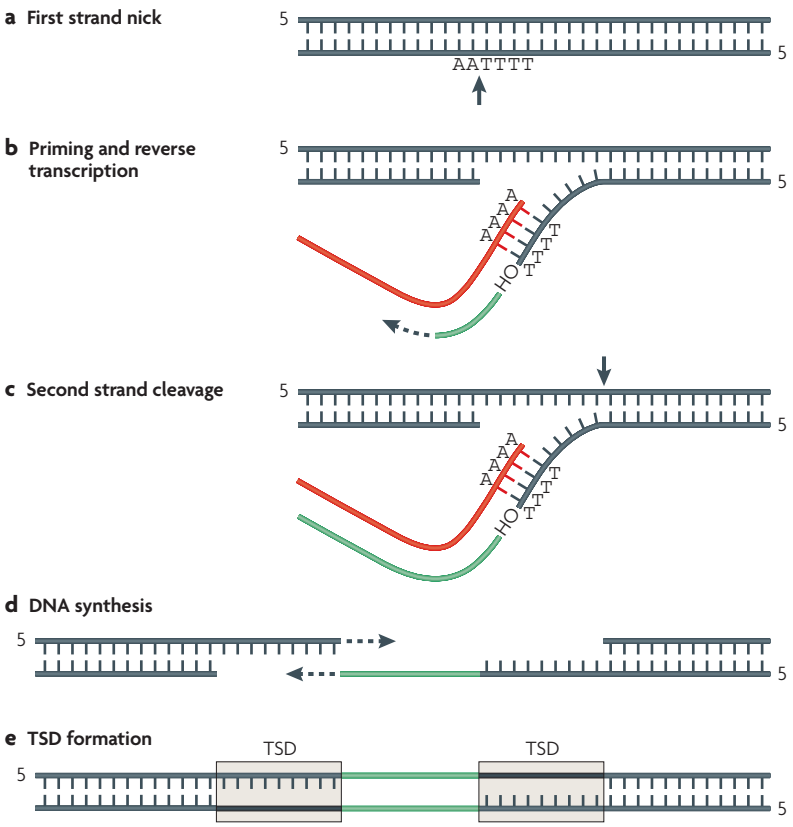
© 2009 Macmillan Publishers Limited. All rights reserved

Box 1 | **The retrotransposition cycle**

The increase in copy numbers of non-long terminal repeat (LTR) retrotransposons occurs through an RNA-based duplication process termed retrotransposition. The first step in LINE-1 (L1) retrotransposition involves RNA polymerase II-mediated transcription of a genomic L1 locus from an internal promoter that directs transcription initiation at the 5′ boundary of the L1 element[19,129]; an internal promoter allows a retrotransposon to generate autonomous duplicate copies at multiple locations in the genome. The L1 RNA is exported to the cytoplasm, in which ORF1 (which encodes an RNA-binding protein) and ORF2 (which encodes a protein with endonuclease and reverse-transcriptase activities) are translated. Both proteins show a strong *cis*-preference[28]; consequently, they preferentially associate with the L1 RNA transcript that encoded them to produce a ribonucleoprotein (RNP) particle. The RNP is then transported back into the nucleus by a mechanism that is poorly understood.

The integration of the L1 element into the genome is likely to occur through a process termed target-primed reverse transcription (TPRT)[13,130,131], which was originally described for the R2 non-LTR retrotransposon of the silkworm *Bombyx mori*[132]. During TPRT, it is thought that the L1 endonuclease cleaves the first strand of target DNA, generally between T and A at 5′-TTTTAA-3′ consensus sites[133] (see the figure, part **a**). The free 3′ hydroxyl (OH) generated by the nick is then used to prime reverse transcription of L1 RNA (red) by the L1 reverse transcriptase (**b**). The second strand of the target DNA is cleaved (**c**) and used to prime second-strand synthesis (**d**) through poorly understood mechanisms. Hallmarks of the integration process include frequent 5′ truncations, the presence of an oligo(dA)-rich tail at the 3′ end and target site duplications (TSDs) of between 2 and 20 base pairs in length[3,21] (**e**).

*Alu* and SVA retrotransposition is also likely to occur through TPRT using the L1 retrotransposition machinery[12,29,30]. The mechanism of *Alu* and SVA *trans*-mobilization by L1 proteins remains elusive. RNA polymerase III-mediated *Alu* transcripts are exported to the cytoplasm and bound to signal recognition particle 9 kDa protein (SRP9) or SRP14 to form stable RNPs[134,135]. It has been suggested that *Alu* RNPs interact with ribosomes, thereby positioning *Alu* transcripts in close proximity to nascent L1 ORF2 proteins[12,42] (the ORF1 protein enhances, but is not strictly required for, *Alu* retrotransposition[12,136]). However, it remains unclear whether *Alu* RNPs gain access to the L1 retrotransposition machinery in the cytoplasm or in the nucleus, as *Alu* RNPs might recruit L1 ORF2 proteins in the nucleus and immediately proceed with TPRT[137].




***SVA elements.*** SVA elements have been active throughout the ~25 Myr of hominoid evolution, and there are now ~3,000 copies in the human genome[29,30]. A typical full-length SVA element is ~2 kb long and is composed of a hexamer repeat region, an *Alu*-like region, a region consisting of a variable number of tandem repeats, a HERV-K10-like region and a polyadenylation signal ending with an oligo(dA)-rich tail of variable length[29,30] (FIG. 1b). Several lines of evidence suggest that SVA elements are transcribed by RNAPII[29,30]. However, they apparently contain no internal promoter and might rely, at least in part, on promoter activity in flanking regions[29,30]. Like *Alu* elements, SVA elements are non-autonomous TEs that are presumably *trans*-mobilized by the L1 retrotransposition machinery[29,30] (BOX 1).

***Other non-LTR retrotransposons.*** In addition to the L1, *Alu* and SVA elements, which are currently active, there are families of old, inactive non-LTR retrotransposons that comprise ~6% of the human genome (FIG. 1a). Although they are far less numerous than L1 and *Alu* elements, these elements provide a rich molecular 'fossil record' that testifies to the long relationship between TEs and the human genome[3]. This record indicates, for example, that before the autonomous L1 element and its *Alu* parasite expansions, the genome experienced retrotransposition of the autonomous LINE-2 element and its mammalian-wide interspersed repeat (MIR) parasite[3]. These old elements might have substantially affected human genome evolution[31–34].

## Evolutionary dynamics

The impact of non-LTR retrotransposons on human genome evolution largely results from their extremely high copy numbers (for example, there is one *Alu* insertion every 3 kb on average[3]) and their continued activity over tens of millions of years. These two features are particularly striking when considering the various cellular processes that control retrotransposon activity (BOX 2). At an evolutionary scale, the vertical persistence of non-LTR retrotransposons, not only in primates but also in mammals in general, sets them apart from most other TEs in mammals and other eukaryotes[3,5,35]. In this section, we discuss the evolutionary dynamics that have made non-LTR retrotransposons so prolific during primate genome evolution.

***Subfamily structure and source elements.*** A key concept that is relevant to the evolutionary dynamics of L1, *Alu* and SVA sequences is that they can all be divided into subfamilies or 'clades' of related elements based on diagnostic nucleotide substitutions and insertions or deletions that are exclusively shared by all subfamily members. For example, more than 200 *Alu* subfamilies are currently recognized in the human genome[36] but only 6 subfamilies of the younger SVA family exist[30]. Not only are subfamilies different in age, but the diagnostic sequence mutations or changes that define subfamilies tend to accumulate hierarchically[23,37]; that is, instead of two subfamilies being independently derived from an ancestral subfamily, most subfamilies represent an



© 2009 Macmillan Publishers Limited. All rights reserved

Box 2 | **Cellular factors that influence retrotransposition**

Transposable elements (TEs) can be seen as selfish genetic entities, the spread of which can be deleterious to the host cell due to the genomic instability that is induced by a massive increase in copy number. As a result of the conflicting interests of TEs and the host genome, the cell has developed various processes to control retrotransposon activity, as predicted by the Red Queen hypothesis[138]. Below we provide examples of how LINE-1 (L1) and *Alu* retrotransposition activity is regulated in host cells (for more detailed discussions, see REFS 11,139).

Regulation of L1 retrotransposition can occur at the transcriptional level — for example, new regulatory regions have frequently been recruited during the evolution of L1 (REF. 38). The current L1 5′UTR contains several transcription factor-binding sites that are important for transcription activation or initiation[140–142]. In addition, DNA methylation at the promoter is known to repress L1 expression[124,143]. L1 elements are also subject to post-transcriptional regulation. For example, RNA-induced silencing through RNA interference has been suggested to reduce L1 retrotransposition in cultured cells[144,145]. The A-rich coding strand of the full-length human L1 contains 19 potential canonical and non-canonical polyadenylation signals that lead to truncation of full-length L1 transcripts by premature polyadenylation, therefore contributing to the attenuation of L1 activity[110]. Furthermore, cells produce proteins, such as those of the apolipoprotein B mRNA-editing complex 3 (APOBEC3) family, that can inhibit L1 and *Alu* retrotransposition[146].

*Alu* activity is influenced by its primary sequence in that the accumulation of mutations can alter important motifs, such as the internal RNA polymerase III promoter or signal recognition particle 9 kDa protein (SRP9)- or SRP14-binding motifs[26,42]. The accumulation of mutations is facilitated by the high density of CpG dinucleotides, which are prone to mutation as a result of the deamination of 5-methylcytosine residues[125]. Overall, it has been estimated that when an *Alu* copy reaches ~10% divergence from its subfamily consensus sequence, the likelihood that it will continue to be active is remote[42]. The length and homogeneity of the oligo(dA)-rich tail also seem to be important for activity[147,148]. The genomic environment into which *Alu* copies are inserted is crucial for retrotranspositional activity[149–151], and the distance between the oligo(dA)-rich tail at the 3′ end of the *Alu* sequence and the RNA polymerase III terminator, which is located in the downstream sequence and determines the overall length of *Alu* transcripts[26], is also important.

ongoing linear sequential evolution pattern in which one subfamily is derived successively from another. For example, it has been shown that all L1 subfamilies in the human genome were derived sequentially from a single lineage over the past ~40 Myr[38]. Similar patterns of subfamily evolution have been reported for *Alu*[23] and SVA[30] elements. These observations can be explained if one assumes that only a few elements (so-called 'source' or 'master' elements) are involved in the retrotransposition process and are responsible for the formation of all other subfamily members[37].

The original 'master gene' model of retrotransposon amplification[37] has been refined, in particular by quantifying the number of retrotransposition-competent elements in the human genome. Analysis of the >200 *Alu* subfamilies in the human genome suggested the existence of at least 143 *Alu* source elements[36], and it has been estimated that an average human genome carries 80–100 retrotransposition-competent L1 copies, 6 of which — known as 'hot L1s' — are probably responsible for the majority of L1 retrotransposition[22,39,40]. These results further indicate that several source elements may exist within a subfamily, because all 6 hot L1 elements belong to the L1-Ta subfamily[22]. A network-based analysis also revealed that human-specific *Alu* subfamilies typically contain secondary source elements in addition to a master element. Secondary source elements comprise ~15% of subfamily members and give rise to approximately 30% of all subfamily copies[41]. Therefore, there may be hundreds of active *Alu* 'core' sequences in the human genome[42]. Although they only represent a tiny fraction of all human non-LTR retrotransposons, source elements can be considered as the ultimate drivers of evolutionary change in the human genome because they are responsible for most of the L1, *Alu* and SVA elements that have been inserted into our genome.

***Stealth drivers and long-term evolution.*** Another distinguishing feature of human non-LTR retrotransposons is their persistent activity over tens of millions of years of evolution. How have active retrotransposons been maintained over this time? Reconstruction of the evolutionary history of the *Alu* Yb lineage showed that it originated during early hominoid evolution, 18–25 Myr ago[43]. Strikingly, the *Alu* Yb lineage has dramatically expanded to ~2,000 copies over the past few million years specifically in the human genome, as shown by the finding that non-human hominoid primates carry only a few *Alu* Yb elements[43–45]. Therefore, the *Alu* Yb lineage remained in the genome with little or no retrotransposition for 15–20 Myr while preserving the ability to generate a high number of new copies in a species-specific manner. These results suggest that long-lived, low-activity source elements may act as 'stealth drivers' that occasionally produce copies, some of which can become highly active. Whereas highly active master elements might be deleterious and negatively selected, low-activity stealth drivers might allow the *Alu* lineage to persist in the long term[43]. Attenuation of mobilization activity might be a common evolutionary strategy of various retrotransposons[46,47]; therefore, the ability to maintain low to moderate levels of retrotransposition activity might be an important feature that allowed human non-LTR retrotransposons to maintain long-term activity.

## Impact on genome evolution

***Amplification rates.*** Because of their continued activity and accumulation in the genome over tens of millions of years, L1, *Alu* and SVA elements have had a tremendous impact on the evolution of primate genomes in terms of both structure and function. To assess this impact, one can consider how frequently retrotransposition occurs in the germ line. The current rate of *Alu* retrotransposition has been estimated as approximately 1 insertion for every 20 births in humans, which is based both on the frequency of disease-causing *de novo* insertions compared with nucleotide substitutions[48] and on comparisons between the human and chimpanzee genomes[48] and between multiple human genome sequences[49]. The current rate of L1 retrotransposition has also been estimated as approximately 1 insertion for every 20 births in humans based on disease-causing *de novo* insertions[50], but as approximately 1 insertion for every 200 births based on genome comparisons[49]. The difference between the two estimates might lie in the underlying assumptions of the methods used, but no such bias is observed for *Alu* element insertion estimates based on

**Hominoids**
The group of primates comprised of humans and apes. Hominoids diverged from Old World monkeys approximately 25 million years ago.

***Trans*-mobilization**
The process by which non-autonomous retrotransposons, such as *Alu* and SVA elements, borrow the LINE-1 retrotransposition machinery to perform their own retrotransposition.

**Red Queen hypothesis**
Proposed by Van Valen in 1973, this hypothesis states that, for an evolutionary system, continuing development is needed to maintain its fitness relative to the systems it is co-evolving with.



© 2009 Macmillan Publishers Limited. All rights reserved

Box 3 | **Retrotransposon insertion polymorphisms as genetic markers**

As revealed by pioneering studies on humans[152–154], primates[155] and non-primate groups[156,157], retrotransposons provide several advantages that make them very powerful tools as genetic markers for studying human and non-human primate evolutionary history[23,157,158]. They are essentially homoplasy-free markers, as individuals that share retrotransposon copies at orthologous sites are almost certain to have inherited them from a common ancestor (the precise excision of retrotransposons is extremely rare)[158,159]. When comparing genomes, the absence of an element at a locus indicates that the individual carries an ancestral version of that locus, and this makes it possible to include hypothetical ancestors to root phylogenetic trees[153]. As there are only two possible character states for each locus — the presence or absence of the element — genotyping of individuals for retrotransposon insertions is technically straightforward. Furthermore, the vast majority of retrotransposon insertions are neutral residents of the genome[160], and the gradual accumulation of elements over time makes it possible to determine the loci that are most suitable for investigating evolutionary relationships at a range of time points in primate history. As a result, retrotransposon insertion polymorphisms (most notably *Alu* elements) have been used to decipher the phylogenetic relationships of various primate groups[161,162], including the resolution of the human–chimpanzee–gorilla trichotomy that demonstrated the close relationship between humans and chimpanzees[163].

Some retrotransposons were inserted so recently that they are polymorphic for presence or absence among human populations and individuals[23,49,164]. In particular, *Alu* elements have proved highly informative for the study of human origins by providing strong evidence for an African origin of anatomically modern humans[153,154]. More recently, *Alu* element insertion polymorphisms have been used to investigate human population structure and demography[154,165,166]. Retrotransposon insertion polymorphisms are also being used as forensic tools — for example, for species-specific DNA detection and quantification, for the analysis of complex biomaterials, for human gender determination and for the inference of geographic origin of human samples[167].

the same approaches. Alternatively, the difference might reflect recent variation in the L1 retrotransposition rate or intense negative selection against L1 insertions. The current SVA retrotransposition rate has tentatively been estimated as approximately 1 insertion for every 900 births based on genome comparisons[49]; this rate is more uncertain owing to the smaller data sets available for analysis. Although new heritable retrotransposition events take place in the germ line, retrotransposition also occurs in somatic tissues and has been implicated in processes ranging from cancer to brain development[8,51,52]. Retrotransposon-induced somatic variation is a fascinating area of investigation that is likely to provide new insights into the biology of TEs and their impact on humans.

The amplification rates of TEs have not been uniform over time. For example, the majority of *Alu* elements were inserted ~40 Myr ago following a peak of amplification during which there was approximately one new *Alu* insertion in every birth[53]. Similarly, over the past ~70 Myr of evolution, variation in the L1 amplification rate has been observed, and the most prolific L1 subfamilies were amplified 12–40 Myr ago[38]. Genome-wide comparisons of the human and chimpanzee genomes have provided additional evidence for recent variation in L1, *Alu* and SVA retrotransposition rates, as judged by the different numbers of species-specific elements that have been inserted since the divergence of the two species ~6 Myr ago[14,54,55]. Such fluctuation in amplification rates over a short timescale suggests influences at the host population level[40,54].

Homoplasy
Similarity due to independent evolutionary change — that is, not inherited from a common ancestor.

X inactivation
The process by which, in female mammals, one of the two copies of the X chromosome is inactivated during early embryogenesis. The inactive X chromosome is silenced by being packaged into transcriptionally inactive heterochromatin.

***Changes in copy number.*** Perhaps one of the most intuitive consequences of TE accumulation is their contribution to increases in genome size[56]: L1 and *Alu* elements alone have contributed ~750 Mb to the human genome[3] (FIG. 1a). This increase in genome size is an ongoing process, as the human genome has accumulated ~2,000 L1, ~7,000 *Alu* and ~1,000 SVA copies over the past ~6 Myr of human evolution, which is a combined addition of >8 Mb[14]. Equally importantly, the ongoing expansion of non-LTR retrotransposons has also created significant inter-individual variation in retrotransposon content; several hundred new mobile element insertions have been detected in multiple human genomic sequences[49,57–59]. These human-specific retrotransposon insertions are often polymorphic (present or absent) at orthologous loci among human individuals, and they constitute highly informative genetic markers that are being used to investigate human evolutionary history, population structure and demography (BOX 3).

## Local genomic instability

Retrotransposons can generate genomic instability in many ways. In this section, we consider the effects of retrotransposons at a local genomic scale.

***Insertion mutagenesis.*** The most straightforward way a retrotransposon can alter genome function, and thereby potentially influence genome evolution, is by inserting into protein-coding or regulatory regions (FIG. 2a). Owing to the immediate phenotypic impact of such insertions, they were the first to be detected[7]. Examples of human genetic disorders caused by *de novo* L1, *Alu* and SVA insertions continue to accumulate, and 65 cases have been shown to cause heritable diseases, such as haemophilia, cystic fibrosis, Apert syndrome, neurofibromatosis, Duchenne muscular dystrophy, β-thalassaemia, hypercholesterolaemia and breast and colon cancers[8,9,11]. Overall, it has been estimated that ~0.3% of all human mutations are attributable to *de novo* L1, *Alu* and SVA insertions[10]. Interestingly, L1 (and to a lesser extent *Alu* and SVA) disease-causing insertions seem to be enriched on the X chromosome[8,9,11]. This might partly be attributable to ascertainment bias, as X-linked genetic disorders are often dominant in males and are thus more easily detected. Alternatively, L1 elements might preferentially insert into the X chromosome. One possible explanation is that the preference might be linked to a proposed involvement of L1 elements in X inactivation[60,61], in which they might help to spread silencing signals.

***DNA double-strand breaks.*** It has recently been shown in mammalian cell lines that the number of DNA double-strand breaks (DSBs) generated by L1 ORF2 proteins, which have endonuclease activity, is much higher than the number associated with actual L1 insertions[62] (FIG. 2b). The extent to which these DSBs contribute to human genomic instability remains unknown because levels of L1 expression under these experimental conditions were much higher than those expected under normal cellular conditions. However, the repair of L1-mediated DSB lesions would leave no signature of L1



© 2009 Macmillan Publishers Limited. All rights reserved

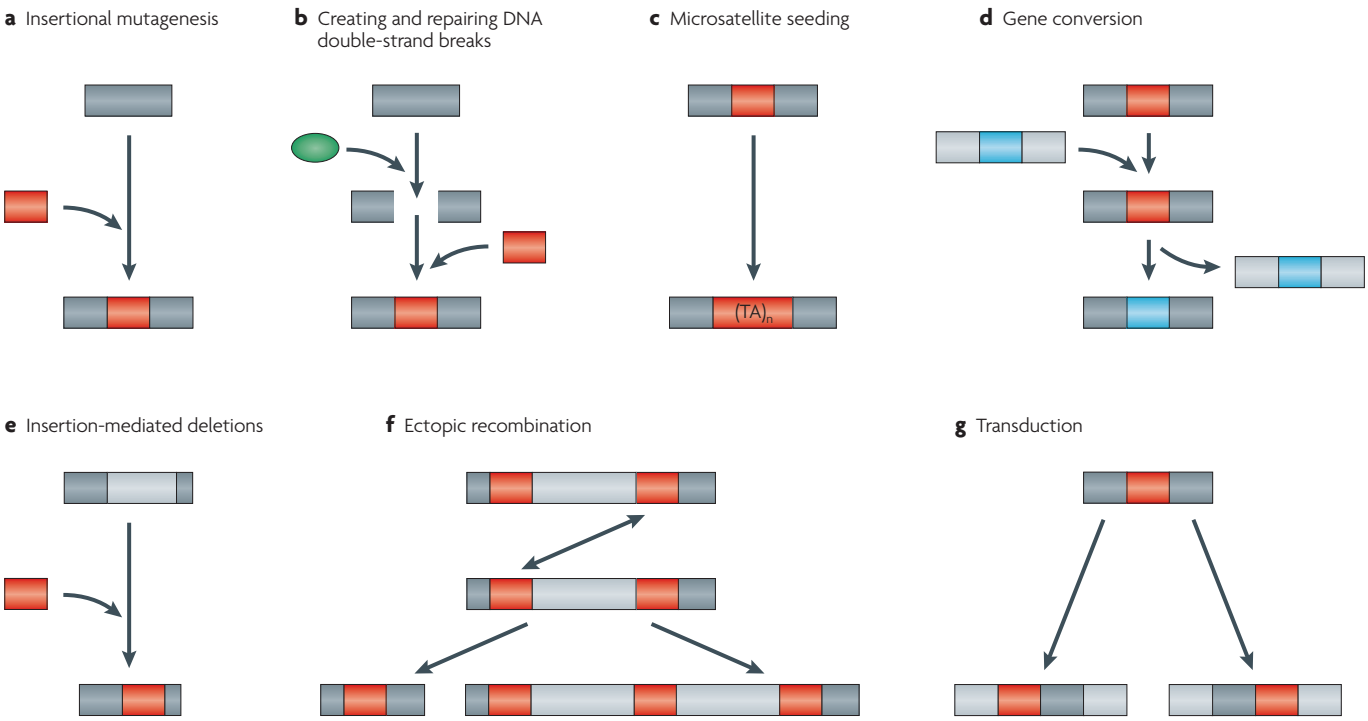



Figure 2 | **Impact of retrotransposons on human genome structure. a** | Typical insertion of a LINE-1 (L1), *Alu* or SVA retrotransposon (red box) at a new genomic site (dark grey). If the new genomic site is a genic region, the retrotransposon may cause insertional mutagenesis. **b** | The protein product (green oval) of an L1 element may create DNA double-strand breaks (broken dark grey area). Alternatively, an existing double-strand break may be repaired by non-classical endonuclease-independent insertion of a retrotransposon. **c** | Microsatellites (for example, $(TA)_n$) may arise from the homopolymeric tracts that are endogenous to retrotransposons. **d** | Gene conversion may alter the sequence compositions of homologous retrotransposon copies (red and blue boxes). **e** | The insertion of a retrotransposon is sometimes associated with the concomitant deletion of a target genomic sequence (light grey box). **f** | Ectopic recombination (double arrowhead) between non-allelic homologous retrotransposons may result in genomic rearrangements, such as deletions (left) or duplications (right) of intervening genomic sequences. **g** | During the duplication of a retrotransposon, the downstream 3′ flanking sequence or the upstream 5′ flanking sequence (dark grey boxes) may also be duplicated (known as 3′ or 5′ transduction, respectively). This results in the retrotransposition of the 3′ flanking sequence (left) or the 5′ flanking sequence (right) along with the retrotransposon.

involvement, so it is possible that a substantial fraction of the genomic instability associated with DSBs, which are highly mutagenic and prone to recombination, is attributable to L1 activity.

***L1 and* Alu *elements have been linked to DSB repair.*** Evidence from L1 retrotransposition assays in cultured cells showed that L1 insertions can occur independently of endonuclease in mammalian cell lines that cannot perform non-homologous end joining, which is a major mechanism of DSB repair[63] (FIG. 2b). Endonuclease-independent (ENi) L1 insertions lack the hallmarks of TPRT (BOX 1), which suggests that L1 elements can integrate into and repair DSBs[63]. In addition, dysfunctional telomeres can serve as substrates for ENi L1 retrotransposition, and endonuclease-deficient LINE-like (*Penelope*) elements are present at the telomeres of several eukaryotes, which suggests that ENi retrotransposition might be an ancestral mechanism of RNA-mediated DNA repair that was used before non-LTR retrotransposons acquired an endonuclease domain[64,65]. Recent analyses of the human genome have shown that 0.5–0.7% of all L1 and *Alu* insertions have non-canonical structures and might have resulted from ENi retrotransposition[66,67], which suggests that non-LTR retrotransposons in general, not just L1 elements, might provide an additional mechanism for maintaining human genome integrity.

***Sources of microsatellites.*** Because of their abundance in the genome and because they contain homopolymeric tracts, non-LTR retrotransposons can generate microsatellites at many loci in the genome (FIG. 2c). In particular, this has been studied for *Alu* elements[68,69], each new copy of which provides two potential sources of microsatellites: the linker region in the middle of the element and the 3`oligo(dA)-rich tail (FIG. 1b). These homopolymeric repeats can be subject to mutations, such as nucleotide substitutions and replication slippage, which can produce microsatellites of varying length and complexity. Consequently, it is not surprising that ~20% of all microsatellites (including ~50% of mononucleotide

**Homopolymeric tract**
A DNA sequence made of the same nucleotide repeated in tandem.

**Microsatellite**
A class of repetitive DNA made up of tandem repeats that are 1–8 bp in length.



© 2009 Macmillan Publishers Limited. All rights reserved

microsatellites) shared by the human and chimpanzee genomes lie within *Alu* elements[70]. In addition, there are at least two examples of genetic disorders that are caused by the expansion of microsatellites that arose from A-rich regions of *Alu* elements[71,72].

***Gene conversion.*** Several studies indicate that *Alu* elements undergo gene conversion[73,74] (FIG. 2d), which is a type of recombination that is defined as the non-reciprocal transfer of information between homologous sequences. Gene conversion might play a part in the evolution of *Alu* elements by inactivating active copies or reactivating inactivated copies[23]. For example, it has recently been shown that the master element of the *Alu* Yh3a3 subfamily has been inactivated by gene conversion in humans, therefore preventing further amplification of this subfamily[75]. In addition, because *Alu* elements make up >10% of the human genome, *Alu*-mediated gene conversion might have a substantial impact on the overall nucleotide diversity of our genome. Also, it might impair the use of SNPs located within *Alu* sequences as genetic markers, as gene conversion would make these SNPs identical by state rather than identical by descent[23]. However, the significance of this phenomenon has not been tested formally; next-generation sequencing and personal genomics will open new avenues for resolving this question.

## Genomic rearrangements

In addition to generating local genomic instability, retrotransposons can generate genomic rearrangements, such as deletions, duplications and inversions. In this section, we discuss three ways in which retrotransposons can create structural variation in the genome.

***Insertion-mediated deletions.*** The insertion of L1 and *Alu* elements at new genomic sites sometimes results in the concomitant deletion of an adjacent genomic sequence (FIG. 2e). This phenomenon was first observed through the analysis of L1 integrations in cultured human cells: ~20% of L1 insertions were associated with structural rearrangements, including concomitant deletions at the insertion site that ranged in size from 1 bp to possibly >130 kb[76–78]. These deletions can arise by endonuclease-dependent and ENi mechanisms[78]. L1 and *Alu* insertion-mediated deletions have subsequently been shown to occur naturally in the human and chimpanzee genomes, although the deletions are usually shorter (<800 bp on average) and occur at a much lower frequency than in cultured cells (a frequency of ~2% and ~0.3% for L1 and *Alu* insertion events, respectively)[79,80]. This might reflect, at least partly, negative selection against large, disruptive, insertion-mediated deletions. Consistent with these observations, a 46 kb L1 insertion-mediated deletion event in the pyruvate dehydrogenase complex, component X (*PDHX*) gene has recently been implicated in pyruvate dehydrogenase complex deficiency[81], and human–chimpanzee genome comparisons have identified a single insertion-mediated deletion event that caused loss of a functional gene in the past ~6 Myr[79].

It has also been noted that ~90% of non-classical ENi L1 and *Alu* insertions are associated with deletions of flanking sequences that range in size from 1 bp to 14 kb, including one deletion that removed an olfactory receptor gene from the human and chimpanzee genomes[66,67]. Altogether, it has been estimated that during primate evolution, as many as 45,000 insertion-mediated deletions might have removed >30 Mb of genomic sequence[18].

***Ectopic recombination.*** Due to their extremely high copy numbers, L1 and *Alu* elements can also create structural genomic variation at the post-insertion stage through recombination between non-allelic homologous elements (FIG. 2f), including between elements that have been present in the genome for a long time. Ectopic recombination can result in various types of genomic rearrangements, such as deletions, duplications and inversions.

It has long been recognized that *Alu* recombination-mediated deletions (RMDs) occur in the human genome: there are >70 reported cases of *Alu* RMDs being responsible for various cancers and genetic disorders[8,10]. By contrast, only three disease-causing L1 RMD events have been reported[17]. Genome-wide comparisons have identified 492 *Alu* RMD events and 73 L1 RMD events that have taken place in the human genome since the human–chimpanzee divergence[16,17]. L1 RMDs are larger on average than *Alu* RMDs and occur more frequently in gene-poor regions of the genome. These results suggest that there might be negative selection against long, deleterious L1 RMDs in gene-rich regions of the genome[18,82,83]. Therefore, *Alu* and L1 RMD events that are detectable by comparative genomics approaches are likely to represent the fraction of RMDs that have escaped negative selection. However, based on human and chimpanzee genome comparisons, these events have collectively removed nearly 1 Mb of genomic sequence from the human genome over the past few million years[16–18], thereby underscoring their important evolutionary impact.

The human genome contains many large (>10 kb) and highly similar (>90% sequence identity) duplicated genomic regions, which are termed segmental duplications. Interestingly, the boundaries of human segmental duplications are significantly enriched in *Alu* elements — that is, they comprise ~24% of boundary sequences but only ~11% of the total human genome[84]. Considering that ~5% of the human genome has been duplicated in the past ~40 Myr, recombination between *Alu* elements might represent an important mechanism for the origin and expansion of human segmental duplications[84].

The contribution of L1 and *Alu* elements to chromosomal inversions has also been investigated recently by comparative genomics. Nearly half of the inversions that have taken place in the human and chimpanzee genomes since their divergence have involved L1 and *Alu* elements, and ~20% of all inversions can clearly be identified as products of L1–L1 or *Alu*–*Alu* recombination events[85]. Although this type of rearrangement does not result in gain or loss of genomic

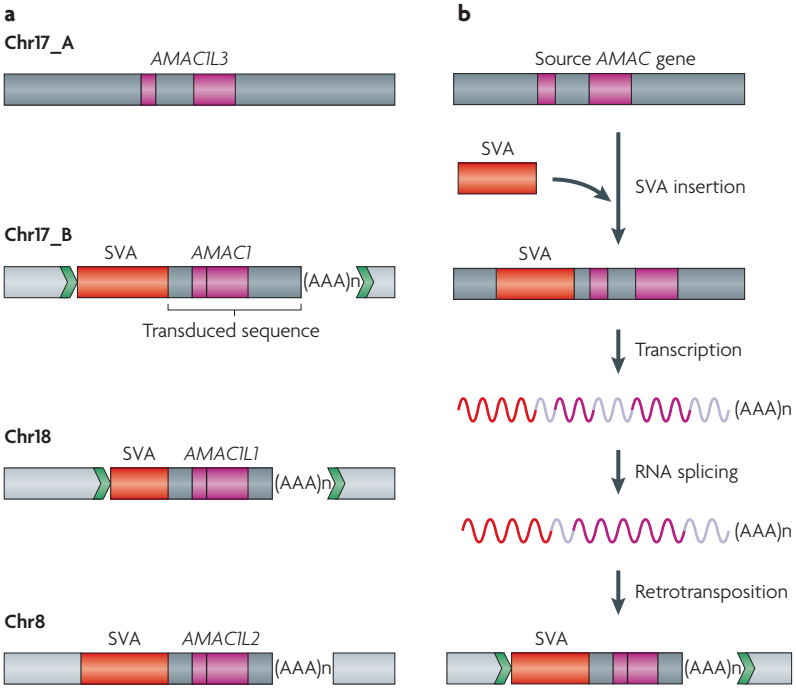
**Identical by state**
Alleles that have the same character state as a result of independent evolutionary changes (that is, the alleles were not inherited from a common ancestor).

**Identical by descent**
Alleles that have the same character state as a result of being directly inherited from a common ancestor.



© 2009 Macmillan Publishers Limited. All rights reserved

Box 4 | **Birth of a gene family by retrotransposon-mediated transduction**




It has been experimentally shown using cell culture assays that LINE-1 (L1) retrotransposons can mediate exon shuffling by 3⃗transduction[88]. Subsequent analyses of the human genome have confirmed that L1-mediated transduction took place during human genome evolution and that it may account for 0.6–1% of human DNA[3,90,91]. However, whether it contributes to the evolution of new gene functions remains an open question. A recent analysis of SVA retrotransposons has demonstrated the evolutionary significance of retrotransposon-mediated 3⃗ transduction by showing that SVA-mediated transduction is responsible for the creation of the acyl-malonyl condensing enzyme 1 (*AMAC1*) gene family, which has four members in the human genome[89].

As part of a genome-wide analysis of SVA-mediated transduction, Xing *et al.*[89] identified 143 events that transduced sequences ranging in size from a few dozen base pairs to almost two kilobase pairs. Interestingly, 3 transduced sequences located on chromosomes 8, 17 and 18 were found to originate from the same source locus that was located elsewhere on chromosome 17 (see the figure, part **a**). The flanking sequences of the original locus are shown as dark grey boxes and the sequences flanking the transduced loci are shown as light grey boxes. Target site duplications are shown as green arrows. SVA elements are shown as red bars, and the coding regions are shown as purple bars. SVA element oligo(dA)-rich tails are shown as '(AAA)n'. Analysis of the four paralogous sequences identified four copies of the *AMAC1* gene. The ancestral *AMAC1L3* gene copy at the source locus consisted of two exons separated by an intron. By contrast, the three transduced copies of *AMAC1L3* (*AMAC1*, *AMAC1L1* and *AMAC1L2*) were intronless as a result of the splicing of the intron during the retrotransposition process (**b**). Evolutionary analyses indicated that the three transduction events all took place ~7–14 Myr ago, as humans and African great apes share all four *AMAC1* copies, whereas orangutans and other primate and non-primate species that have been analysed only possess the ancestral *AMAC1L3* gene. Experimental studies indicated that, in addition to *AMAC1L3*, at least two of the three transduced *AMAC1* genes are expressed in human tissues. RNA transcript sequence analyses of the expressed *AMAC1* duplicates further revealed that the promoter sequence had been duplicated along with the *AMAC1* coding sequence as part of the 3⃗transduction process. This indicates that retrotransposon-mediated gene transduction can duplicate not only coding regions of genes but also their regulatory regions; therefore, genes retain their functional potential after duplication, and retrotransposon-mediated duplication can lead to the rapid generation of functional gene families. Figure modified, with permission, from REF. 89 © 2006 National Academy of Sciences, USA.

sequence, it contributes to genomic variation and can be of functional importance — for example, by causing the inversion of exons[85].

***Transduction of flanking sequences.*** In addition to duplicating themselves, L1 and SVA elements sometimes carry upstream or downstream flanking genomic sequences with them (termed 5` and 3` transduction, respectively) (FIG. 2g). In 3` transduction, the RNA transcription machinery skips the weak retrotransposon polyadenylation signal and terminates transcription by using an alternative polyadenylation signal located downstream in the 3` flanking sequence. Similarly, 5` transduction occurs when a promoter located upstream of the retrotransposon is used to transcribe the sequence down to the retrotransposon[86,87]. The transcript containing the retrotransposon, along with the extra genomic sequence, is integrated into the genome through retrotransposition. Initially characterized using cell culture-based methods[88], 3` transduction has subsequently been shown to occur frequently in the human genome: ~10% of L1 and SVA insertions are associated with 3` transduction events[30,89–91].

## Genetic innovation

Variation in the number of genes among species indicates that new genes are continuously generated over evolutionary time. Comparative genomic studies have confirmed the notion of 'evolutionary tinkering'[92], according to which new genes most commonly arise by rearrangements between pre-existing genetic structures. In this section, we explore mechanisms by which retrotransposons have fostered genetic innovations in the human lineage.

***Transduction-mediated gene formation.*** The process of retrotransposon-mediated transduction (discussed above) can lead to the duplication of coding sequences that are located in the transduced flanking genomic sequence. The potential of L1 retrotransposons to mediate exon shuffling through 3` transduction has been experimentally confirmed using cell culture assays[88]. This mechanism has subsequently been shown to have mediated the formation of a new gene family during recent human evolution through multiple SVA-mediated transduction events of the acyl-malonyl condensing enzyme 1 (*AMAC1*) gene[89] (BOX 4).

***Gene retrotransposition.*** In contrast to transduction, gene retrotransposition only duplicates gene sequences and no retrotransposon sequence is co-duplicated in the process. This is because gene retrotransposition is based on the hijacking of the L1 retrotransposition machinery by host mRNA transcripts[93] in a similar way to *Alu* and SVA retrotransposition. As a result, gene retrotransposition generally does not duplicate upstream regulatory regions; therefore, duplicated genes must fortuitously acquire new regulatory regions to become functional. Gene retrotransposition was long thought to generate non-functional duplicate gene copies termed retropseudogenes. However, genome-wide searches



© 2009 Macmillan Publishers Limited. All rights reserved

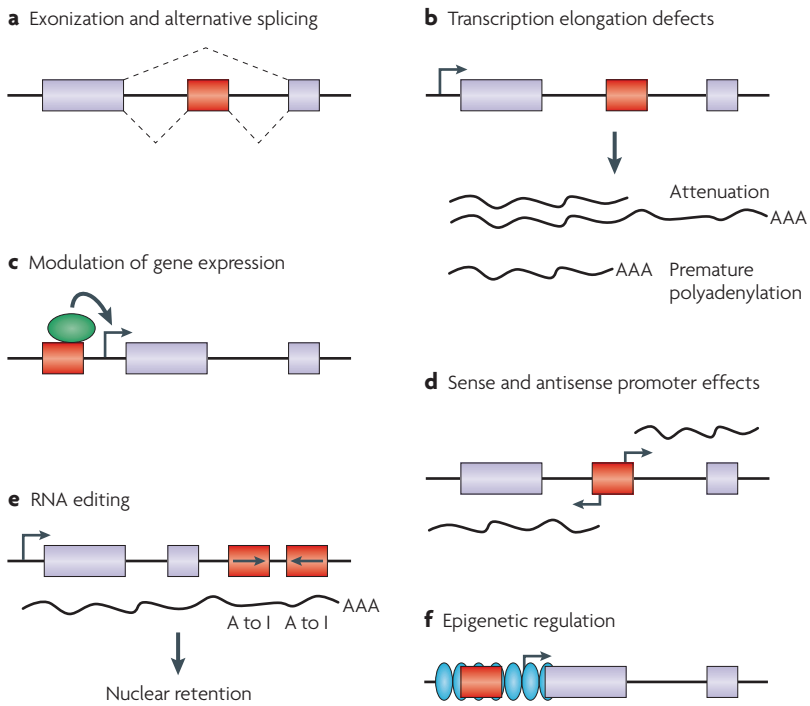



Figure 3 | **Impact of retrotransposons on human gene expression.** **a** | A retrotransposon sequence (red box) can be recruited as a coding sequence and be integrated into a gene (made up here of two exons, grey boxes). This is often associated with alternative splicing (dashed lines). **b** | The presence of a retrotransposon in the intron of a gene (the sequence between the two grey boxes, which represent exons) can result in transcription elongation defects, such as attenuation or premature polyadenylation. **c** | Retrotransposons carry transcription factor-binding sites. A transcription factor (green oval) carried by a retrotransposon can upregulate or downregulate (curved arrow) the expression (horizontal arrow) of neighbouring genes (grey boxes). **d** | A retrotransposon carries sense and antisense promoters (horizontal arrows) that can initiate downstream and upstream transcription. **e** | The presence of two *Alu* elements in the opposite orientation in gene transcripts can lead to adenosine to inosine (A to I) editing, which can result in suppression of expression through nuclear retention of edited RNA transcripts. **f** | A retrotransposon sequence can be methylated, which may initiate and spread the formation of heterochromatin (blue ovals), thereby altering the expression (horizontal arrow) of neighbouring genes (grey boxes).

have confirmed the importance of gene retrotransposition in the emergence of new primate genes[94–96], and it has been estimated that at least one new retrogene has emerged every million years in the human lineage over the past ~65 Myr[97] (for a more detailed discussion, see REF. 96).

***Exonization.*** Alternative splicing is a widespread mechanism that occurs in 40–60% of human genes[3,98]. By producing more than one type of mRNA from a single gene, alternative splicing substantially contributes to human proteome variation[98]. Interestingly, retrotransposon sequences are sometimes recruited as exons that become integrated into genes in a process termed exonization (FIG. 3a). It was initially estimated, based on transcript sequence data, that ~4% of human protein-coding sequences contained TEs (mostly *Alu* and L1)[99]. However, a recent analysis at the protein level suggested that this proportion is closer to ~0.1%[100].

Exonization is thought to be facilitated by the fact that many TEs carry cryptic donor and acceptor splice sites. For example, a typical *Alu* sequence contains 9 GT dinucleotides and 14 AG dinucleotides that represent the same numbers of cryptic donor and acceptor splice sites, respectively[101,102]. *Alu* exonization has occurred repeatedly and consistently during primate evolution[103]. It has been estimated that ~5% of alternatively spliced exons are derived from *Alu* elements in humans and that most — if not all — *Alu* exons are alternatively spliced, presumably because constitutively expressed *Alu* exons are deleterious and negatively selected[101]. Consistent with this assumption, the three reported cases of exonized *Alu* elements becoming constitutively expressed are all associated with genetic disorders[98].

Non-LTR retrotransposons have also been involved in facilitating the molecular domestication of other TEs. This is exemplified by the SET domain and mariner transposase fusion gene (*SETMAR*), a chimeric primate gene that resulted from the fusion of a SET histone methyltransferase gene to the transposase gene of an *Hsmar1* DNA transposon[104]. The birth of *SETMAR* might never have occurred without the contribution of an *Alu* element that inserted into and partially deleted the 5`-terminal inverted repeat of the *Hsmar1* element[104]. Because both of the terminal inverted repeats of DNA transposons are necessary for transposition, the *Alu* insertion might have contributed to the recruitment of the *Hsmar1* transposon as part of *SETMAR* by immobilizing it at a period when the *Hsmar1* family was experiencing high levels of transposition in primate genomes[5]. Overall, it is striking that non-LTR retrotransposons seem to directly contribute a disproportionately small number of domesticated genes to genomes compared with other TEs (such as DNA transposons), despite the fact that they are the most numerous TEs in the human genome[105,106].

## Impact on gene expression

As described above, retrotransposons have dramatically affected human evolution at the DNA level. Evidence is also accumulating that retrotransposons have substantially shaped human evolution at the RNA level through various mechanisms, which we discuss in this section.

***Modulation of gene expression.*** Non-LTR retrotransposons affect the expression of nearby genes through a range of mechanisms. Similar to *Alu* elements, L1 sequences can provide new splice sites that might promote exonization and alternative splicing[107,108] (FIG. 3a). In addition, intronic L1 elements can interfere with transcriptional elongation of the host gene owing to RNAPII having a reduced ability to read through L1 sequences[109] (FIG. 3b). Furthermore, retrotransposon sequences can provide polyadenylation signals that induce the termination of gene transcripts[110–112] (FIG. 3b). It has also been shown that *Alu* elements carry transcription factor-binding sites that might modulate gene expression[113,114] (FIG. 3c). The functional promoter sequences of L1 and *Alu* elements can also initiate sense or antisense transcription through other genes[115–117] (FIG. 3d).

**Retrogene**
An expressed and functional gene that is generated by retrotransposition and that usually has an intact ORF that is consistent with that of the parental gene.

**Molecular domestication**
The recruitment of a transposable element-derived sequence into a new functional role by the genome.



© 2009 Macmillan Publishers Limited. All rights reserved

The potential of L1 endogenous promoter and polyadenylation signals to create transcriptome diversity in humans is shown by 15 human genes that were apparently split by L1 elements, which were inserted into intronic sequences in antisense orientations[118]. In each of these genes, a transcript containing exons upstream of the insertion site terminates at the L1 3`antisense polyadenylation signal; a second transcript derived from the L1 5`antisense promoter drives the expression of a transcript that includes the downstream exons of the gene. These observations provide a mechanistic basis for the emergence of new gene structures by gene fission.

***RNA editing.*** RNA editing is a process by which RNA nucleotide sequences are co- or post-transcriptionally modified, such as by the conversion of adenosine to inosine (A to I) in dsRNA (FIG. 3e). A to I editing is widespread in humans, and >90% of all A to I substitutions occur in *Alu* sequences embedded in mRNA transcripts[119–122]. Editing in *Alu* elements might be favoured by the dimeric structure of these elements and the occasional occurrence of pairs of *Alu* elements in inverted orientations. A to I editing can eliminate splice sites and therefore might affect the alternative splicing of exonized *Alu* sequences. Furthermore, it has recently been shown that A to I editing of pairs of inverted *Alu* elements in 3` UTRs can suppress expression through the nuclear retention of mRNA transcripts[123].

***Epigenetic regulation.*** The epigenetic silencing of retrotransposon activity through DNA methylation is an important defence mechanism for the cell (BOX 2). The L1 promoter CpG island is typically highly methylated[124], and *Alu* and SVA elements have a high density of CpG sites[30,125], to the extent that one-third of all human CpG sites are contained within *Alu* sequences[126]. Because L1, *Alu* and SVA elements are frequently found in or near genes, heterochromatin formed at retrotransposons could spread and repress the transcription of nearby genes (FIG. 3f). Consistent with this is the observation that *Alu* elements might be excluded from human imprinted regions owing to their potential negative effect on methylation, which is associated with imprinting[127]. The proposed involvement of L1 elements in X inactivation (discussed above) is also linked to methylation. However, the formal demonstration of retrotransposon-mediated epigenetic control of neighbouring genes in humans and the evaluation of the extent of this phenomenon at a genome-wide scale are active topics of investigation in the field.

***Conserved non-coding elements.*** Recent genome comparisons have revealed the occurrence of numerous conserved non-coding elements (CNEs) in the human genome. Strikingly, many CNEs seem to be derived from ancient TE sequences and in particular from a class of non-LTR retrotransposons known as short interspersed elements (SINEs), to which *Alu* elements belong[31–33]. These ancient SINE-derived sequences are currently evolving under strong negative selection and have apparently taken on regulatory functions[31–33]. It remains unclear whether the frequent recruitment of SINEs as CNEs indicates an endogenous functional property of these elements, is a by-product of their high copy numbers in mammalian genomes or results from their distinctive sequence architecture, which makes them more readily identifiable as old retrotransposons[106]. In any event, the genome-wide contribution of this phenomenon to human evolution remains to be determined but is likely to be important.

## Conclusions and future directions

For tens or even hundreds of millions of years, TEs have shaped the evolution of the genomes in which they reside[128]. The maintenance of activity over extended periods of time is a distinguishing feature of non-LTR retrotransposons that has been instrumental in their evolutionary success in the human lineage. Our understanding of the factors underlying this evolutionary success is still incomplete, and new insights into this intriguing topic are likely to emerge over the next few years. The intricate relationship of non-LTR retrotransposons with the human genome does not mean that they have been maintained because they confer evolutionary advantages. On the contrary, we believe that the profound impact of retrotransposons on genome evolution is a by-product of, not the reason for, the evolutionary success of these selfish genetic elements.

This view is supported by the notion that retrotransposons often pose a threat to human health. Although it has long been established that retrotransposons can cause genetic diseases through insertional mutagenesis as a result of their sustained mobilization activity, other mechanisms are less well understood. For example, investigating the contribution of L1 endonuclease to the generation of DSBs in germline and somatic tissues might provide insights into the L1 integration process and the interaction of L1 with DNA repair mechanisms, as well as into chromosomal damage and human health more generally. Although the contribution of retrotransposons to genomic deletions, such as insertion-mediated deletions and RMDs, is well established, other types of genomic rearrangements, such as retrotransposon recombination-mediated duplications, are less well understood, partly because they are more difficult to characterize through computational comparisons of genome sequences. Given that duplications are a key contributor to genetic innovation, the extent to which retrotransposons have contributed to the formation of new genes in the human genome might still be underestimated. This is also true for many aspects of the impact of retrotransposons on gene expression. For example, there is growing evidence that TEs in general, not just non-LTR retrotransposons, have been a rich source of material for the assembly and evolution of regulatory networks[106]. The increased genomic and transcriptomic sequence data provided by next-generation sequencing are likely to shed new light on the dynamic roles that TEs have in shaping within- and inter-individual variation and will allow researchers to dissect retrotransposon-induced variation at an ever-increasing resolution. Such information is crucial if we are to better understand the overall impact of TEs on human health and genome evolution and their contribution to the unique traits that make us human.

**Imprinting**
An epigenetic phenomenon in which certain genes are expressed in a parent-of-origin-specific manner.



© 2009 Macmillan Publishers Limited. All rights reserved

1. McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216 (1956).
2. Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
3. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
**A landmark study that analysed the entire human genome sequence and revealed that TEs make up nearly half of our genome.**
4. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. Mobile DNA II (American Society for Microbiology, Washington, 2002).
5. Pace, J. K. & Feschotte, C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* **17**, 422–432 (2007).
6. Mills, R. E., Bennett, E. A., Iskow, R. C. & Devine, S. E. Which transposable elements are active in the human genome? *Trends Genet.* **23**, 183–191 (2007).
7. Kazazian, H. H. Jr *et al.* Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166 (1988).
8. Deininger, P. L. & Batzer, M. A. *Alu* repeats and human disease. *Mol. Genet. Metab.* **67**, 183–193 (1999).
9. Chen, J. M., Stenson, P. D., Cooper, D. N. & Ferec, C. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum. Genet.* **117**, 411–427 (2005).
10. Callinan, P. A. & Batzer, M. A. Retrotransposable elements and human disease. *Genome Dyn.* **1**, 104–115 (2006).
11. Belancio, V. P., Hedges, D. J. & Deininger, P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* **18**, 343–358 (2008).
12. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked *Alu* sequences. *Nature Genet.* **35**, 41–48 (2003).
**In this paper, the authors establish an experimental test of *Alu* retrotransposition in cultured cells and show that the L1 ORF2 protein is required for *Alu* retrotransposition.**
13. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
**A landmark study that presents the development and characterization of an *in vitro* assay to measure L1 retrotransposition in cultured cells.**
14. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
15. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
16. Sen, S. K. *et al.* Human genomic deletions mediated by recombination between *Alu* elements. *Am. J. Hum. Genet.* **79**, 41–53 (2006).
17. Han, K. *et al.* L1 recombination-associated deletions generate human genomic variation. *Proc. Natl Acad. Sci. USA* **105**, 19366–19371 (2008).
**References 16 and 17 report genome-wide analyses showing that L1 and *Alu* recombination-mediated deletions have greatly affected human genome evolution.**
18. Cordaux, R. The human genome in the LINE of fire. *Proc. Natl Acad. Sci. USA* **105**, 19033–19034 (2008).
19. Swergold, G. D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* **10**, 6718–6729 (1990).
20. Babushok, D. V. & Kazazian, H. H. Jr. Progress in understanding the biology of the human mutagen LINE-1. *Hum. Mutat.* **28**, 527–539 (2007).
21. Szak, S. T. *et al.* Molecular archeology of L1 insertions in the human genome. *Genome Biol.* **3**, research0052 (2002).
22. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA* **100**, 5280–5285 (2003).
**This study indicates that most L1 retrotransposition events in humans may result from the activity of just six highly active L1 elements.**
23. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nature Rev. Genet.* **3**, 370–379 (2002).
24. Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J. & Schmitz, J. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.* **23**, 158–161 (2007).
25. Shaikh, T. H., Roy, A. M., Kim, J., Batzer, M. A. & Deininger, P. L. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* **271**, 222–234 (1997).
26. Comeaux, M. S., Roy-Engel, A. M., Hedges, D. J. & Deininger, P. L. Diverse *cis* factors controlling *Alu* retrotransposition: what causes *Alu* elements to die? *Genome Res.* **19**, 545–555 (2009).
27. Weiner, A. M. SINEs and LINEs: the art of biting the hand that feeds you. *Curr. Opin. Cell Biol.* **14**, 343–350 (2002).
28. Wei, W. *et al.* Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell. Biol.* **21**, 1429–1439 (2001).
29. Ostertag, E. M., Goodier, J. L., Zhang, Y. & Kazazian, H. H. Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **73**, 1444–1451 (2003).
30. Wang, H. *et al.* SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* **354**, 994–1007 (2005).
31. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**, 87–90 (2006).
32. Nishihara, H., Smit, A. F. & Okada, N. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **16**, 864–874 (2006).
33. Kamal, M., Xie, X. & Lander, E. S. A large family of ancient repeat elements in the human genome is under strong selection. *Proc. Natl Acad. Sci. USA* **103**, 2740–2745 (2006).
34. Krull, M., Petrusma, M., Makalowski, W., Brosius, J. & Schmitz, J. Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res.* **17**, 1139–1145 (2007).
35. Cantrell, M. A., Scott, L., Brown, C. J., Martinez, A. R. & Wichman, H. A. Loss of LINE-1 activity in the megabats. *Genetics* **178**, 393–404 (2008).
36. Price, A. L., Eskin, E. & Pevzner, P. A. Whole-genome analysis of *Alu* repeat elements reveals complex evolutionary history. *Genome Res.* **14**, 2245–2252 (2004).
37. Deininger, P. L., Batzer, M. A., Hutchison, C. A. & Edgell, M. H. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**, 307–311 (1992).
38. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).
39. Lutz, S. M., Vincent, B. J., Kazazian, H. H. Jr, Batzer, M. A. & Moran, J. V. Allelic heterogeneity in LINE-1 retrotransposition activity. *Am. J. Hum. Genet.* **73**, 1431–1437 (2003).
40. Seleme Mdel, C. *et al.* Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc. Natl Acad. Sci. USA* **103**, 6611–6616 (2006).
41. Cordaux, R., Hedges, D. J. & Batzer, M. A. Retrotransposition of *Alu* elements: how many sources? *Trends Genet.* **20**, 464–467 (2004).
**The authors refine the master gene model of *Alu* amplification by showing that in a typical human-specific *Alu* subfamily, 10–20% of the *Alu* copies are retrotransposition competent.**
42. Bennett, E. A. *et al.* Active *Alu* retrotransposons in the human genome. *Genome Res.* **18**, 1875–1883 (2008).
43. Han, K. *et al.* Under the genomic radar: the stealth model of *Alu* amplification. *Genome Res.* **15**, 655–664 (2005).
44. Carter, A. B. *et al.* Genome-wide analysis of the human *Alu* Yb-lineage. *Hum. Genomics* **1**, 167–178 (2004).
45. Gibbons, R. *et al.* Distinguishing humans from great apes with AluYb8 repeats. *J. Mol. Biol.* **339**, 721–729 (2004).
46. Li, T. H. & Schmid, C. W. Alu's dimeric consensus sequence destabilizes its transcripts. *Gene* **324**, 191–200 (2004).
47. Han, J. S. & Boeke, J. D. A highly active synthetic mammalian retrotransposon. *Nature* **429**, 314–318 (2004).
48. Cordaux, R., Hedges, D. J., Herke, S. W. & Batzer, M. A. Estimating the retrotransposition rate of human *Alu* elements. *Gene* **373**, 134–137 (2006).
49. Xing, J. *et al.* Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* 13 May 2009 (doi:10.1101/gr.091827.109).
**The first genome-wide comparison of inter-individual structural variation due to TEs in humans.**
50. Kazazian, H. H. Jr. An estimated frequency of endogenous insertional mutations in humans. *Nature Genet.* **22**, 130 (1999).
51. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
52. Kano, H. *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23**, 1303–1312 (2009).
53. Shen, M. R., Batzer, M. A. & Deininger, P. L. Evolution of the master Alu gene(s). *J. Mol. Evol.* **33**, 311–320 (1991).
54. Hedges, D. J. *et al.* Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* **14**, 1068–1075 (2004).
55. Mills, R. E. *et al.* Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* **78**, 671–679 (2006).
56. Liu, G. *et al.* Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358–368 (2003).
57. Badge, R. M., Alisch, R. S. & Moran, J. V. ATLAS: a system to selectively identify human-specific L1 insertions. *Am. J. Hum. Genet.* **72**, 823–838 (2003).
58. Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S. & Devine, S. E. Natural genetic variation caused by transposable elements in humans. *Genetics* **168**, 933–951 (2004).
59. Boissinot, S., Entezam, A., Young, L., Munson, P. J. & Furano, A. V. The insertional history of an active family of L1 retrotransposons in humans. *Genome Res.* **14**, 1221–1231 (2004).
60. Lyon, M. F. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* **80**, 133–137 (1998).
61. Bailey, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl Acad. Sci. USA* **97**, 6634–6639 (2000).
62. Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **357**, 1383–1393 (2006).
63. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31**, 159–165 (2002).
64. Morrish, T. A. *et al.* Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446**, 208–212 (2007).
65. Gladyshev, E. A. & Arkhipova, I. R. Telomere-associated endonuclease-deficient *Penelope*-like retroelements in diverse eukaryotes. *Proc. Natl Acad. Sci. USA* **104**, 9352–9357 (2007).
66. Sen, S. K., Huang, C. T., Han, K. & Batzer, M. A. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res.* **35**, 3741–3751 (2007).
67. Srikanta, D. *et al.* An alternative pathway for *Alu* retrotransposition suggests a role in DNA double-strand break repair. *Genomics* **93**, 205–215 (2009).
68. Arcot, S. S., Wang, Z., Weber, J. L., Deininger, P. L. & Batzer, M. A. *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* **29**, 136–144 (1995).
69. Jurka, J. & Pethiyagoda, C. Simple repetitive DNA sequences from primates: compilation and analysis. *J. Mol. Evol.* **40**, 120–126 (1995).
70. Kelkar, Y. D., Tyekucheva, S., Chiaromonte, F. & Makova, K. D. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res.* **18**, 30–38 (2008).
71. Justice, C. M. *et al.* Phylogenetic analysis of the Friedreich ataxia GAA trinucleotide repeat. *J. Mol. Evol.* **52**, 232–238 (2001).
72. Kurosaki, T., Ninokata, A., Wang, L. & Ueda, S. Evolutionary scenario for acquisition of CAG repeats in human *SCA1* gene. *Gene* **373**, 23–27 (2006).
73. Kass, D. H., Batzer, M. A. & Deininger, P. L. Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Mol. Cell. Biol.* **15**, 19–25 (1995).




© 2009 Macmillan Publishers Limited. All rights reserved

74. Roy, A. M. *et al.* Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* **10**, 1485–1495 (2000).
75. Styles, P. & Brookfield, J. F. Y. Source gene composition and gene conversion of the AluYh and AluYi lineages of retrotransposons. *BMC Evol. Biol.* **9**, 102 (2009).
76. Gilbert, N., Lutz-Prigge, S. & Moran, J. V. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**, 315–325 (2002).
77. Symer, D. E. *et al.* Human l1 retrotransposition is associated with genetic instability *in vivo*. *Cell* **110**, 327–338 (2002).
**References 76 and 77 show that L1 retrotransposition can be associated with various forms of genomic instability in cultured cells.**
78. Gilbert, N., Lutz, S., Morrish, T. A. & Moran, J. V. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* **25**, 7780–7795 (2005).
79. Callinan, P. A. *et al.* *Alu* retrotransposition-mediated deletion. *J. Mol. Biol.* **348**, 791–800 (2005).
80. Han, K. *et al.* Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.* **33**, 4040–4052 (2005).
81. Miné, M. *et al.* A large genomic deletion in the *PDHX* gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Hum. Mutat.* **28**, 137–142 (2007).
82. Boissinot, S., Davis, J., Entezam, A., Petrov, D. & Furano, A. V. Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl Acad. Sci. USA* **103**, 9590–9594 (2006).
83. Song, M. & Boissinot, S. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **390**, 206–213 (2007).
84. Bailey, J. A., Liu, G. & Eichler, E. E. An *Alu* transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**, 823–834 (2003).
85. Lee, J., Han, K., Meyer, T. J., Kim, H. S. & Batzer, M. A. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE* **3**, e4047 (2008).
86. Hancks, D., Ewing, A., Chen, J. E., Tokunaga, K. & Kazazian, H. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res.* 27 Jul 2009 (doi:10.1101/gr.093153.109).
87. Dament, A. *et al.* 5`-transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res.* 3 Aug 2009 (doi:10.1101/gr.093435.109).
88. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H. Jr. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530–1534 (1999).
**This paper shows that retrotransposon-mediated transduction can create new genes in cultured cells.**
89. Xing, J. *et al.* Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl Acad. Sci. USA* **103**, 17608–17613 (2006).
**This study shows the evolutionary significance of retrotransposon-mediated transduction during human evolution.**
90. Goodier, J. L., Ostertag, E. M. & Kazazian, H. H. Jr. Transduction of 3`-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9**, 653–657 (2000).
91. Pickeral, O. K., Makalowski, W., Boguski, M. S. & Boeke, J. D. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* **10**, 411–415 (2000).
92. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
93. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**, 363–367 (2000).
94. Sayah, D. M., Sokolskaja, E., Berthoux, L. & Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569–573 (2004).
95. Babushok, D. V. *et al.* A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Res.* **17**, 1129–1138 (2007).
96. Kaessmann, H., Vinckenbosch, N. & Long, M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Rev. Genet.* **10**, 19–31 (2009).
97. Marques, A. C., Dupanloup, I., Vinckenbosch, N., Reymond, A. & Kaessmann, H. Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.* **3**, e357 (2005).
98. Kreahling, J. & Graveley, B. R. The origins and implications of *Alu*ternative splicing. *Trends Genet.* **20**, 1–4 (2004).
99. Nekrutenko, A. & Li, W. H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621 (2001).
100. Gotea, V. & Makalowski, W. Do transposable elements really contribute to proteomes? *Trends Genet.* **22**, 260–267 (2006).
101. Sorek, R., Ast, G. & Graur, D. *Alu*-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067 (2002).
102. Lev-Maor, G., Sorek, R., Shomron, N. & Ast, G. The birth of an alternatively spliced exon: 3`splice-site selection in *Alu* exons. *Science* **300**, 1288–1291 (2003).
103. Krull, M., Brosius, J. & Schmitz, J. Alu-SINE exonization: en route to protein coding function. *Mol. Biol. Evol.* **22**, 1702–1711 (2005).
104. Cordaux, R., Udit, S., Batzer, M. A. & Feschotte, C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl Acad. Sci. USA* **103**, 8101–8106 (2006).
105. Volff, J. N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**, 913–922 (2006).
106. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9**, 397–405 (2008).
107. Belancio, V. P., Hedges, D. J. & Deininger, P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.* **34**, 1512–1521 (2006).
108. Belancio, V. P., Roy-Engel, A. M. & Deininger, P. The impact of multiple splice sites in human L1 elements. *Gene* **411**, 38–45 (2008).
109. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274 (2004).
110. Perepelitsa-Belancio, V. & Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genet.* **35**, 363–366 (2003).
**This study shows that the L1 element contains many polyadenylation signals, which result in truncated transcripts and attenuated L1 activity.**
111. Lee, J. Y., 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.* **36**, 5581–5590 (2008).
112. Chen, C., Ara, T. & Gautheret, D. Using *Alu* elements as polyadenylation sites: a case of retroposon exaptation. *Mol. Biol. Evol.* **26**, 327–334 (2009).
113. Shankar, R., Grover, D., Brahmachari, S. K. & Mukerji, M. Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements. *BMC Evol. Biol.* **4**, 37 (2004).
114. 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* **7**, 133 (2006).
115. Speek, M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* **21**, 1973–1985 (2001).
116. Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nature Struct. Mol. Biol.* **13**, 1097–1101 (2006).
117. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41**, 563–571 (2009).
118. Wheelan, S. J., Aizawa, Y., Han, J. S. & Boeke, J. D. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* **15**, 1073–1078 (2005).
**This paper shows how the antisense promoter and polyadenylation signals of L1 elements can lead to the formation of new genes by fission of pre-existing genes.**
119. Kim, D. D. *et al.* Widespread RNA editing of embedded *Alu* elements in the human transcriptome. *Genome Res.* **14**, 1719–1725 (2004).
120. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).
121. Levanon, E. Y. *et al.* Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nature Biotechnol.* **22**, 1001–1005 (2004).
122. Li, J. B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
123. Chen, L. L., DeCerbo, J. N. & Carmichael, G. G. *Alu* element-mediated gene silencing. *EMBO J.* **27**, 1694–1705 (2008).
124. Hata, K. & Sakaki, Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* **189**, 227–234 (1997).
125. Xing, J. *et al.* *Alu* element mutation spectra: molecular clocks and the effect of DNA methylation. *J. Mol. Biol.* **344**, 675–682 (2004).
126. Rubin, C. M., VandeVoort, C. A., Teplitz, R. L. & Schmid, C. W. Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res.* **22**, 5121–5127 (1994).
127. Greally, J. M. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc. Natl Acad. Sci. USA* **99**, 327–332 (2002).
128. Jurka, J., Kapitonov, V. V., Kohany, O. & Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8**, 241–259 (2007).
129. Lavie, L., Maldener, E., Brouha, B., Meese, E. U. & Mayer, J. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.* **14**, 2253–2260 (2004).
130. Feng, Q., Moran, J. V., Kazazian, H. H. Jr & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
131. Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.* **21**, 5899–5910 (2002).
132. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
133. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl Acad. Sci. USA* **94**, 1872–1877 (1997).
134. Sinnett, D., Richer, C., Deragon, J. M. & Labuda, D. *Alu* RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J. Biol. Chem.* **266**, 8675–8678 (1991).
135. Chang, D. Y., Hsu, K. & Maraia, R. J. Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res.* **24**, 4165–4170 (1996).
136. Wallace, N., Wagstaff, B. J., Deininger, P. L. & Roy-Engel, A. M. LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene* **419**, 1–6 (2008).
137. Kroutter, E. N., Belancio, V. P., Wagstaff, B. J. & Roy-Engel, A. M. The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet.* **5**, e1000458 (2009).
138. Van Valen, L. A new evolutionary law. *Evol. Theor.* **1**, 1–30 (1973).
139. Goodier, J. L. & Kazazian, H. H. Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135**, 23–35 (2008).
140. Tchenio, T., Casella, J. F. & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28**, 411–415 (2000).
141. Yang, N., Zhang, L., Zhang, Y. & Kazazian, H. H. Jr. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* **31**, 4929–4940 (2003).
142. Athanikar, J. N., Badge, R. M. & Moran, J. V. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **32**, 3846–3855 (2004).
143. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
144. Soifer, H. S., Zaragoza, A., Peyvan, M., Behlke, M. A. & Rossi, J. J. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res.* **33**, 846–856 (2005).
145. Yang, N. & Kazazian, H. H. Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Struct. Mol. Biol.* **13**, 763–771 (2006).




© 2009 Macmillan Publishers Limited. All rights reserved

146. Schumann, G. G. APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem. Soc. Trans.* **35**, 637–642 (2007).
147. Roy-Engel, A. M. *et al.* Active Alu element 'A-tails': size does matter. *Genome Res.* **12**, 1333–1344 (2002).
148. Dewannieux, M. & Heidmann, T. Role of poly(A) tail length in Alu retrotransposition. *Genomics* **86**, 378–381 (2005).
149. Chesnokov, I. & Schmid, C. W. Flanking sequences of an Alu source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors. *J. Mol. Evol.* **42**, 30–36 (1996).
150. Aleman, C., Roy-Engel, A. M., Shaikh, T. H. & Deininger, P. L. *Cis*-acting influences on Alu RNA levels. *Nucleic Acids Res.* **28**, 4755–4761 (2000).
151. Roy, A. M. *et al.* Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* **302**, 17–25 (2000).
152. Perna, N. T., Batzer, M. A., Deininger, P. L. & Stoneking, M. *Alu* insertion polymorphism: a new type of marker for human population studies. *Hum. Biol.* **64**, 641–648 (1992).
153. Batzer, M. A. *et al.* African origin of human-specific polymorphic Alu insertions. *Proc. Natl Acad. Sci. USA* **91**, 12288–12292 (1994).
154. Stoneking, M. *et al.* *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**, 1061–1071 (1997).
**References 153 and 154 show how *Alu* insertion polymorphisms can be used to study human evolutionary history and demography.**
155. Ryan, S. C. & Dugaiczyk, A. Newly arisen DNA repeats in primate phylogeny. *Proc. Natl Acad. Sci. USA* **86**, 9360–9364 (1989).
156. Shimamura, M. *et al.* Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388**, 666–670 (1997).
157. Shedlock, A. M. & Okada, N. SINE insertions: powerful tools for molecular systematics. *Bioessays* **22**, 148–160 (2000).
158. Ray, D. A., Xing, J., Salem, A. H. & Batzer, M. A. SINEs of a nearly perfect character. *Syst. Biol.* **55**, 928–935 (2006).
159. van de Lagemaat, L. N., Gagnier, L., Medstrand, P. & Mager, D. L. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res.* **15**, 1243–1249 (2005).
160. Cordaux, R., Lee, J., Dinoso, L. & Batzer, M. A. Recently integrated *Alu* retrotransposons are essentially neutral residents of the human genome. *Gene* **373**, 138–144 (2006).
161. Roos, C., Schmitz, J. & Zischler, H. Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl Acad. Sci. USA* **101**, 10650–10654 (2004).
162. Ray, D. A. *et al.* *Alu* insertion loci and platyrrhine primate phylogeny. *Mol. Phylogenet. Evol.* **35**, 117–126 (2005).
163. Salem, A. H. *et al.* Alu elements and hominid phylogenetics. *Proc. Natl Acad. Sci. USA* **100**, 12787–12791 (2003).
164. Cordaux, R., Srikanta, D., Lee, J., Stoneking, M. & Batzer, M. A. In search of polymorphic *Alu* insertions with restricted geographic distributions. *Genomics* **90**, 154–158 (2007).
165. Watkins, W. S. *et al.* Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Res.* **13**, 1607–1618 (2003).
166. Bamshad, M. J. *et al.* Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **72**, 578–589 (2003).
167. Ray, D. A., Walker, J. A. & Batzer, M. A. Mobile element-based forensic genomics. *Mutat. Res.* **616**, 24–33 (2007).

## Acknowledgements
We apologize to colleagues whose work could not be discussed or cited owing to space constraints. Our research on various aspects of mobile elements is supported by a Young Investigator ATIP Award from the Centre National de la Recherche Scientifique to R.C. and by grants from the Louisiana Board of Regents Governor's Biotechnology Initiative (GBI 2002-005), the National Science Foundation (BCS-0218338) and the National Institutes of Health (PO1 AG022064 and RO1 GM59290) to M.A.B.

## DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
*AMAC1* | *PDHX* | *RN7SL1* | *SETMAR*
UniProtKB: http://www.uniprot.org
SRP9 | SRP14

## FURTHER INFORMATION
Richard Cordaux's homepage: http://site.voila.fr/rcordaux
Mark A. Batzer's homepage: http://batzerlab.lsu.edu
dbRIP, a database of retrotransposon insertion polymorphisms: http://dbrip.brocku.ca
Dolan DNA Learning Center, *Alu* insertion polymorphism module: http://www.geneticorigins.org/pv92/aluframeset.htm
Repbase, a database of eukaryotic TEs: http://www.girinst.org/repbase/index.html

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**



© 2009 Macmillan Publishers Limited. All rights reserved