The necessary junk: new functions for transposable elements

Alysson R. Muotri, Maria C.N. Marchetto, Nicole G. Coufal and Fred H. Gage[*]

1-Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

Transposable elements have been shaping the genome throughout evolution, contributing to the creation of new genes and sophisticated regulatory network systems. Today, most of genomes (animals and plants) allow the expression and accommodate transposition of a few transposon families. The potential genetic impact of this small fraction of mobile elements should not be underestimated. Although new insertions that happen in germ cells are likely to be passed to the next generation, mobilization in pluripotent embryonic stem cells or in somatic cells may contribute to the differences observed in genetic makeup and epigenetic gene regulation during development at the cellular level. The fact that these elements are still active, generating innovative ways to alter gene expression and genomic structure, suggests that the cellular genome is not static or deterministic but rather dynamic. In this short review, we collect a set of recent observations that point to a new appreciation of transposable elements as a source of genetic variation.

INTRODUCTION

Long dismissed as selfish or 'junk' DNA, retroelements are frequently thought to be mere intracellular parasites from our distant evolutionary past. Transposons and transposon-derived repetitive elements are responsible for more than 40% of the human genome. However, only a small proportion of these retroelements, <0.05%, remains able to transpose (Fig. 1). Recent evidence indicates that among the non-long terminal repeat (LTR) retrotransposonies, only some long interspersed nucleotide elements-1 (LINE-1, or L1) and short interspersed element (SINE) (*Alu* and SVA) subfamilies continue to be mobile in mammals today (1,2).

LINE-1 comprises ~20% of mammalian genomes (3–5). On the basis of reverse transcriptase (RT) phylogeny, L1 elements are most closely related to the group II introns of mitochondria and eubacteria (6,7). These studies revealed that the RT enzyme is extremely old and that retroelements can be viewed as relics or molecular fossils of the first primitive replication systems in the progenote. The origin of retroelements possibly traces back to the conversion of RNA-based systems, the RNA World, to modern DNA-based systems. Current models suggest that mobile introns of eubacteria were transmitted to eukaryotes during the initial fusion of the eubacterial and archaebacterial genomes or during the symbiosis that gave rise to the mitochondria, generating the modern-day spliceosomal introns (8). Further acquisition of an endonuclease enzyme and a promoter sequence certainly represented important steps in the evolution of L1 retrotransposons, providing autonomy for L1s to insert into many locations throughout the genome. Nevertheless, the L1 machinery is frequently hijacked by other non-autonomous elements.

SINEs are known to occur throughout the genomes of eukaryotes. They originate from retrotransposition events of small RNAs such as 7SL RNA, tRNA or their derivatives featuring internal RNA polymerase III promoters (9–12). Specific types of SINE, called *Alu* elements, are the most abundant repetitive elements in the human genome; they emerged 65 million years ago and amplified throughout the human genome by retrotransposition to reach the present number of more than one million copies. Several pieces of evidence have demonstrated that these elements modulate gene expression at the post-transcriptional level in at least three independent ways. They have been shown to be involved in alternative splicing, RNA editing and translation regulation (10). In normal growth conditions, *Alu* RNAs are present at very low levels in the cytosol (10^3–10^4 molecules per cell), but numerous stressful conditions, such as viral infection, cycloheximide exposure or heat shock, can transiently increase their level of expression, which rapidly decreases upon recovery (13,14). This precisely controlled regulation

*To whom correspondence should be addressed. Tel: +1 8584534100; Fax: +1 8585970824; Email: gage@salk.edu.
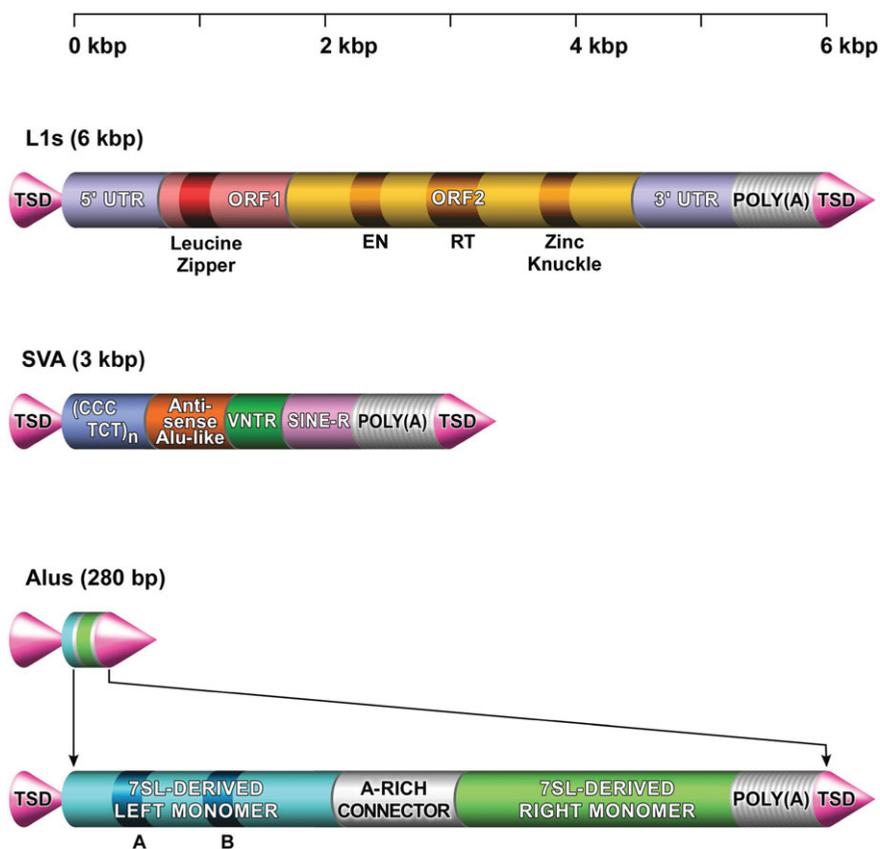
Figure 1. Active transposable elements in the human genome. Although full-length active L1 elements can transcribe all the necessary open-reading frames (ORF1 and ORF2) for autonomous mobilization, *Alu*s and SVAs only can transpose using the L1 machinery *in trans*. TSD, target site duplication; UTR, untranslated region; EN, endonuclease domain; RT, reverse transcriptase domain; VNTR, variable number of tandem repeats.

raised the attractive possibility that *Alu* RNAs might serve a specific function in cell metabolism, which is required during stress conditions. This hypothesis was supported by two independent studies showing that an overexpressed *Alu* RNA was able to stimulate translation of a co-transfected reporter gene in mammalian cells (13,14). These data suggested for the first time that transcribed *Alu* elements might have a particular function in translation regulation.

SVA is a hominid-specific element that was derived from three other repeats (SINE-R, VNTR and *Alu*). It is the youngest retrotransposon family in primates and is also mobilized by L1 (15). Recent studies have demonstrated that this element is capable of transducing genomic sequences upon its mobilization and may also act as a promoter (16). The relationship between *Alu* and SVA remains unclear, but it has been proposed that *Alu*s activate the SVA transcript for entry into the transmechanism of L1 via an SVA homologous portion to *Alu* (2).

OPOSSUM GENOME SEQUENCE SURPRISES

The recent publication of the first marsupial genome to be sequenced (short-tailed opossum or *Monodelphis domestica*) called attention not so much to the protein coding regions of its genome as to the non-coding portion. Interspersed elements cover ∼52% of the opossum genome, compared with ∼44%

for human and ∼38% for mouse (17,18). That increased amount of transposon-derived sequences were probably generated by the combined activity of three different non-LTR transposons (L1; CR1, a more common repetitive element in birds and reptiles; RTE, first described in *C. elegans*), along with a suggested reduced rate of genome recombination, promoting the fixation of the transposed locus (17).

A comparison of opossum and eutherian (placental mammals) genomes revealed a sharp difference in evolutionary innovation between protein-coding and non-coding functional elements. Approximately 20% of eutherian conserved non-coding elements are recent inventions that postdate the divergence of eutheria and metatheria (marsupials), most of which arose from sequences inserted by transposable elements. Such data point to transposons as a major creative force in the evolution of mammalian gene regulation (17).

THE IMPACT OF MOBILE ELEMENTS IN THE GENOME

The proliferation and evolution of retrotransposons have had multiple impacts on the genome, which are only gradually being discovered. Classically, numerous reports have shown that retrotransposition can destabilize the genome, shaping genomic landscapes by insertional mutagenesis, deletions

and gene rearrangements. L1s and *Alu*s may also introduce intragenic polyadenylation signals, influencing evolutionary pressures (19). In addition, transposable elements in high copy number can provide a substrate for illegitimate homologous recombination (also called ectopic recombination), causing rearrangements that may be deleterious, advantageous or null (20,21). The genomic deletions generated by recombination between *Alu* elements in the human lineage since the human–chimpanzee divergence was recently estimated to have occurred approximately 6 million years ago (22). The authors identified 492 human-specific deletions (∼400 kb) attributable to *Alu*-derived ectopic recombination. The majority of the deletions (295 of 492) coincided with known or predicted genes, suggesting that a substantial portion of the genomic differences between humans and chimpanzees was caused by *Alu* recombination (22). This work indicates that this profound recombination process has had a much higher impact in the genome than was inferred from previously identified isolated events.

Transposons have also been notably involved in the emergence of new genes during mammalian evolution, mainly by generating pseudogenes. The estimated number of pseudogenes in the human genome is approximately 8000 and the vast majority has been generated by L1 retrotransposition activity (23–25). Also known as processed pseudogenes or retrogenes (usually lacking intronic regions), they typically contain a 3' poly-A tract and short direct repeats of varying lengths flanking the two ends. These hallmarks can be used to distinguish L1-derived pseudogenes from other types of pseudogenes generated by tandem duplication or unequal crossing-over (26,27).

Extensive retrotransposition activity has been proposed to cause the presence in the genome of a vast number of pseudogenes derived from highly expressed embryonic stem (ES) cells and also germline-specific genes (25,28,29). Two distinct observations could explain such bias. First, L1 intermediates (mRNA and protein) have been shown to be increased in those cells (30). Secondly, retrotransposons that emerge in either ES or germline cells have the advantage of becoming *bona fide* structures to be transmitted to the next generation (29).

Transcription factors that are essential for the maintenance of ES pluripotency, such as Oct4, Nanog and Stella, have been said to generate more pseudogenes than non-ES cell-specific genes (29). Likewise, the expression of at least 10 retroposons was detected only in testis-derived tissue, including primary spermatocytes. The fact that those pseudogenes are conserved in mouse and human genomes and the evidence that some of them could actually be transcribed (31) suggest a functional and evolutionary relevance. Moreover, a potential 'pseudogene-signature' of retrotransposition activity for ES and germline cells can be envisioned.

Using a database developed for human-processed pseudogenes (www.pseudogene.org), combined with the decay rates for human mRNAs, Pavlicek *et al.* (26) proposed that not only germline expression but also RNA stability, translational competence and localization (in free ribosomes outside the endoplasmic reticulum) likely contributed to retrotransposition of processed pseudogenes. Moreover, similar to embryos and oocytes, ES cells may also lack interferon response to long, double-stranded RNA, allowing the expression of long

reverse-transcribed mRNAs (32,33). The combination of these features may explain why few genes are actually capable of giving rise to many pseudogenes and ∼90% of the transcribed genome is not (23,24).

SINEs can also generate novel insertions when hijacking the L1 RT enzyme. It was recently shown that the entire *AMAC* gene was duplicated three times in the human genome through an SVA-mediated transduction event, creating a hybrid SINE-pseudogene (16). Surprisingly, two copies of the retroposed *AMAC* gene can be actively transcribed in different human tissues. The authors proposed that the LTR region from the SVA element could be acting as a promoter for the newly integrated gene duplicate, proposing an explanation for how processed pseudogenes are expressed (16).

Another intriguing role attributed to transposable elements is a process termed 'exaptation', in which relics of transposable elements have acquired a regulatory function, providing the source of some of the most conserved vertebrate-specific genomic sequences (Fig. 2). Ultra conserved SINE sequences have recently been proposed to act as distal *cis*-regulatory elements (12). Also, there is strong evidence that a SINE-derived conserved region situated 0.5 million bases from the neuro-developmental gene *ISL1* can act as an *in vivo* enhancer for a reporter gene. Moreover, the reporter expression recapitulates the expression pattern of islet1 (34).

MECHANISMS FOR CONTROLLING RETROTRANSPOSITION ACTIVITY

Given the deleterious nature of many retrotransposon insertions and rearrangements, evolution has generated multiple pathways to inhibit retrotransposition; these pathways are now being elucidated. Recent evidence has linked miRNA and siRNA pathways with retrotransposon inhibition. In *Drosophila*, a novel, repeat-associated small interfering RNAs (rasiRNAs) homologous to the retrotransposons' antisense strand have been shown to prevent the detrimental retrotransposition of the *gypsy* endogenous retrovirus (35,36). This pathway is Dicer-independent and functions through *Piwi*, a member of the Argonaute family, which acts as splicer for the rasiRNAs, thereby inhibiting retrotransposition in germline cells (37). In addition, recent reports have indicated that AGO3, another member of the Argonaute family, associates and processes sense strand rasiRNAs, leading to a model for rasiRNA 5′ end formation that involves formation of a 5′ terminus that is guided by rasiRNA transcripts from the opposite strand and the activity of PIWI (38). Mammalian PIWI proteins (MIWI and MILI) have also been identified and demonstrated to associate with PIWI-interacting RNAs (piRNAs) to regulate spermatogenesis (39–41). Moreover, Mili-null mice express lower amounts of L1 and intracisternal A particle (IAP) mRNA and lose the DNA methylation of L1 elements. On the basis of this evidence, the authors (41) propose that Mili is regulating L1 and IAP elements in mouse.

In plants, DNA methylation can be induced by double-stranded RNA through the RNAi pathway, a response known as RNA-directed DNA methylation. This process requires a specialized set of RNAi components, including AGO4. This member of the Argonaute family binds to small
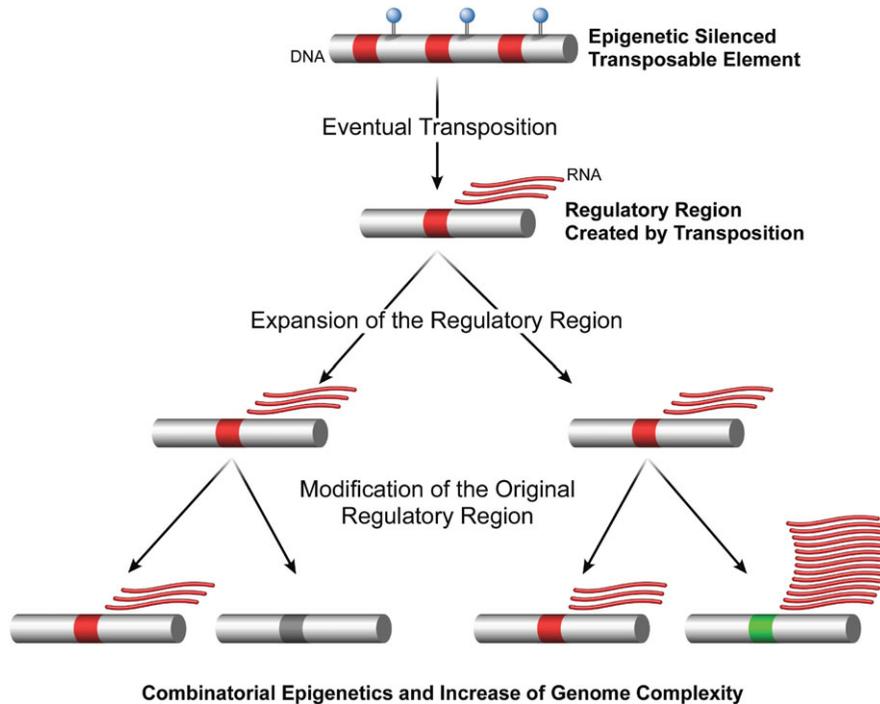
Figure 2. Generation of regulatory sequences by transposition. Epigenetically silenced transposable elements can escape the genome defense and generate a new regulatory region in a different region of the genome. Such insertion can work as an enhancer and drive the expression of genes nearby. After expansion, the transposable regulatory region can mutate or even fall under the control of epigenetic mechanisms. Under these circumstances, genes that were previously regulated by the ancient insertion will now be under the control of the new insertional version, causing gene silencing or upregulation and thus increasing genome complexity. Such processes can take several generations to develop, but the detection of somatic *de novo* insertions may help to elucidate the genetic molecular mechanisms that shape new regulatory regions during evolution.

RNAs originated from transposable and repetitive elements, cleaving target RNA transcripts (42). The model suggests that AGO4 can function at target loci through two distinct and separable mechanisms. First, AGO4 can recruit components that signal DNA methylation in a manner independent of its catalytic activity. Secondly, AGO4 catalytic activity can be crucial for the generation of secondary siRNAs that reinforce its repressive effects. A similar mechanism has not yet been found for silencing mammalian transposable elements.

In humans, evidence that the RNAi pathway can silence L1 retrotransposon activity was recently found in cultured cell lines (43,44). The double-stranded RNA generated by the sense and antisense activity of the L1 5′-UTR promoter region is subject to processing, generating siRNAs that may induce neighbor L1 elements to epigenetic regulation (44). Recent research has also indicated that miRNAs can act to silence retrotransposon elements. Unlike the rasiRNAs, which are processed from long, double-stranded RNA precursors, these miRNAs arise from conventional hairpins. Four miRNAs have been shown to target MIR/LINE-2 elements (45), and 30 miRNAs, a remarkably large number, have exhibited seed matches to *Alu* elements (46).

Other cellular inhibitors of retrotransposon action include the APOBEC family of proteins. Originally shown to have intrinsic antiretroviral ability against HIV and other retroviruses, several members of the APOBEC family have since been shown to inhibit retrotransposon action (47–50). In humans, APOBEC 3A inhibits L1 retrotransposition as much as 85% (47), whereas APOBEC 3B, 3C and 3F all decrease the rate of L1 retrotransposition by ∼75% (47,51); in contrast, APOBEC 3D, 3G and 3H have little effect on L1 activity (47,51). The effect on L1 seems to be independent of APOBEC deaminase activity and may occur through a novel but still unknown mechanism (51,52). APOBEC 3G has been shown to inhibit the activity of *Alu* elements (53) by sequestering *Alu* RNAs in cytoplasmic high-molecular-weight complexes away from L1 elements.

HANDY JUNK

Rather than viewing retrotransposons as only a destabilizing genomic entity, recent reports indicate that the co-evolution of retrotransposons with their host genomes has led to the incorporation and integration of retrotransposons into complex genomic processes. For instance, given that *Alu* is the most prominent human repeat and is found in >5% of all human 3′-UTRs (54), new evidence for miRNA targeting of *Alu* sequences suggests the intriguing prospect (or idea) of Alu participation in gene expression regulation on a global scale. Similarly, the relationship between miRNA and retrotransposon is complex, as recent genomic analysis provides evidence that, in addition to being miRNA targets, *Alu*

elements can drive miRNA expression in the chromosome 19 miRNA cluster through RNA polymerase III (55).

Recent evidence from Morrish *et al*. (56) suggests that L1 element retrotransposons are ancestral mechanisms for telomere repair. The authors utilized cell lines that were deficient in non-homologous end-joining and had telomere capping defects induced through either the loss of DNA protein kinase catalytic subunit activity or through overexpression of dominant negative TRF2. They found that L1s containing a disabled endonuclease domain (ENi) could use the 3′OH overhang present at a dysfunctional telomere as a substrate to initiate retrotransposition, and they demonstrated orientation-specific retrotransposition events at the terminal telomere repeat in ∼30% of ENi events. This seemingly convoluted appropriation could represent a mechanism for RNA-mediated DNA repair associated with non-LTR retrotransposons before the development of the endonuclease domain, illustrating distinct similarities between ENi retrotransposition and the action of telomerase.

There is now abundant evidence for the role of retrotransposons in modulating post-transcriptional gene expression through influences on alternative splicing, RNA editing and translation regulation. Retrotransposons contain an internal polymerase II promoter, allowing them to influence transcription of neighboring genes and acting as alternative promoters (9). In addition, L1 and *Alu* contain numerous internal splice donor and splice acceptor sites that act to influence splicing patterns and expression of many genomic ESTs (57). In a computational comparison of eight vertebrate species, the majority of new exons were found to contain repetitive sequences, especially *Alu*s (58), suggesting that only very few mutations are necessary to convert a repeat to a novel exon. There is also evidence to indicate that repeat elements, especially *Alu*s, are abundant targets of adenosine-to-inosine RNA editing (59). Although the consequences of this editing are as yet unknown, it may act as a mechanism for introducing new intronic splice sites and thereby affecting new exon formation. Lev-Maor *et al*. (60) have shown that the primate-specific *Alu*-exon formation depends on RNA editing for exonization in a tissue-specific fashion, and only a very few mutations are needed to exonize an *Alu*. Retrotransposons have also recently been shown to contain transcription factor-binding sites, and most binding sites are for factors known to be early markers of development (61). Lastly, new research implicates retrotransposons in protein translation, specifically *Alu* elements acting to increase translation initiation, and in complex with SRP9/14 protein, they can inhibit translation (10).

UNCOVERING SOMATIC RETROTRANSPOSITION

The magnitude of endogenous retrotransposition in somatic cells is still unknown, and retrotransposons are often assumed to be active only in germline or early ES cells. This assumption is mainly based on the silencing of transposon expression by DNA methylation in somatic cells. However, methylation can be regulated, especially during the differentiation process in several tissues. Also, the fidelity of DNA methylation is not perfectly reproduced after each cell division, allowing a window of opportunity for retrotransposition to happen. Moreover, non-phenotypic integrations are likely to be overlooked, causing a detection bias, where only visible mutants (usually leading to diseases) are scored.

The notion that somatic cells are free of new retrotransposition events is slowly changing. The development of better detection tools, such as synthetic L1 elements carrying reporter genes, is revealing the likelihood of retrotransposition in several somatic tissues (62,63). If the behavior of artificial L1s used in these studies is similar to that of endogenous elements, each individual can be considered to be a living library of *de novo* insertions, allowing stochastic gene expression in every single cell of the body. Of interest, we recently showed that L1 retrotransposition might have an insertional preference for neuronal genes when active in progenitor cells during the neuronal differentiation process (64). Newly generated L1 insertions can affect gene expression and influence neuronal fate. Such an observation indicates that neurons are a genetic mosaic for L1 content and that phenomenon may contribute to neuronal diversity (65). It remains to be determined the actual amount of endogenous mobilization, if such genetic mosaicism can influence neuronal networks and, as a result, have a functional consequence on behavior (66).

RETROTRANSPOSON POLYMORPHISMS

Following a detailed characterization of chimpanzee-specific L1 subfamily diversity and a comparison with their human-specific counterparts, it has been reported that L1 elements have experienced different evolutionary fates in humans and chimpanzees within the past approximately 6 million years (67). Although the species-specific L1 copy numbers are on the same order in both species (1200–2000 copies), the number of retrotransposition-competent elements appears to be much higher in the human genome than in the chimpanzee genome (67).

Mobile elements are highly polymorphic in human population (68–70). Taking advantage of that characteristic, Witherspoon *et al*. (71) used L1 and *Alu* insertion polymorphisms to analyze human population structure. After genotyping 75 recent, polymorphic L1 insertions in 317 individuals from 21 populations in sub-Saharan Africa, East Asia, Europe and the Indian subcontinent, they obtained congruent results that support the recent African origin model of human ancestry. Collectively, L1 insertion polymorphisms represent useful markers for population diversity and evolutionary studies. Also, as shown for *Alu* elements, L1 insertion polymorphisms have potential use in forensics (72).

Such polymorphisms are probably generated in germ cells or during early embryonic stages. In support of this assertion, we recently showed that human ES cells (HESC) express endogenous L1 elements and can accommodate L1 retrotransposition *in vitro*. The resultant retrotransposition events can occur into genes and can result in the concomitant deletion of genomic DNA at the target site. These data suggest that L1 retrotransposition events are likely to occur during early stages of human development, contributing to the genome fluidity and variability of HESC (30). In fact, an endogenous L1 retrotransposition in the *CHM* gene of a patient with
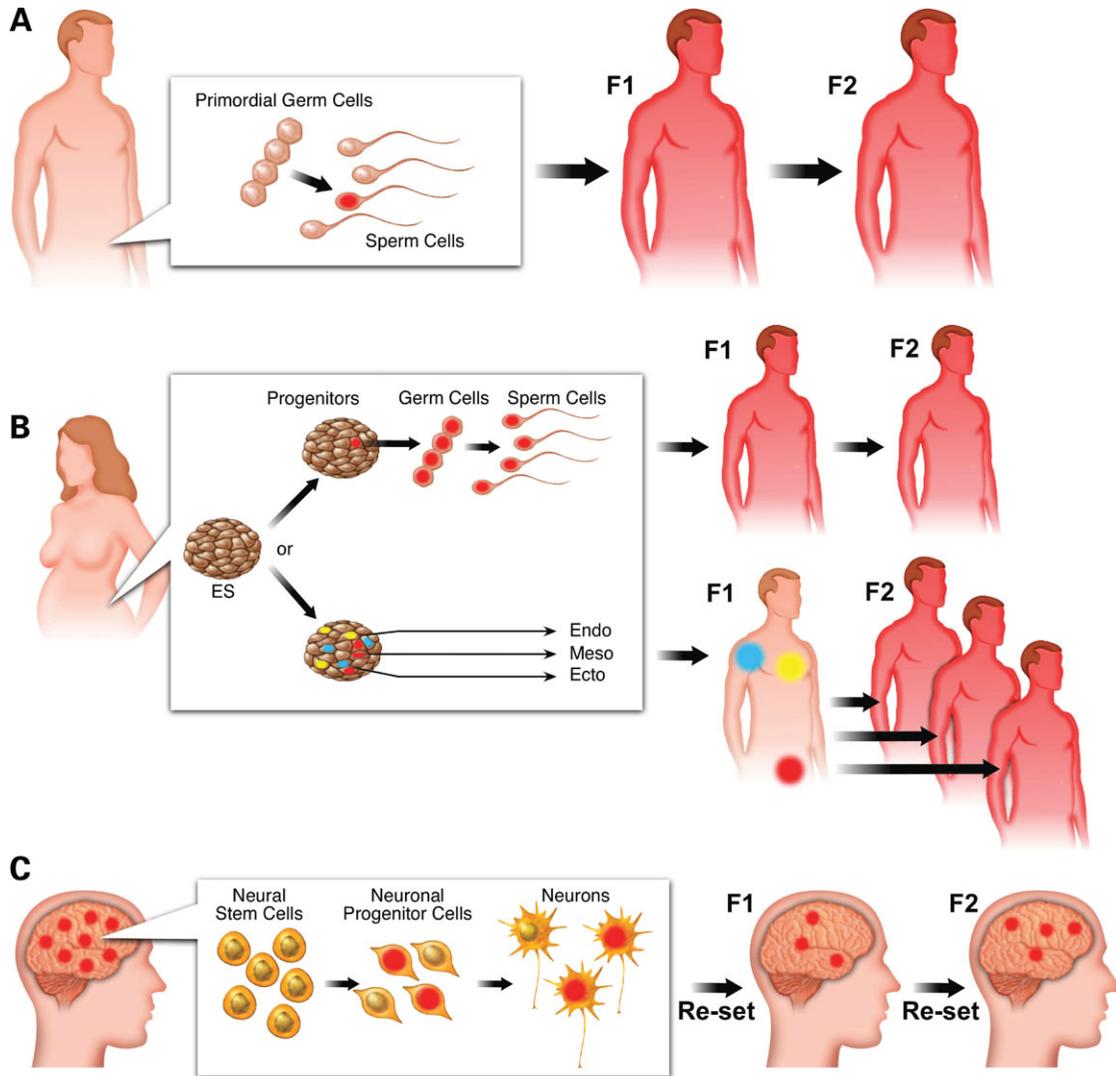
Figure 3. Evolutionary consequences of retrotransposition in different cell types. (**A**) Retrotransposition in primordial germ cells will give rise to germ cells carrying a new insertion (in red). If the germ cells carrying the new insertion are used to produce an egg, the resultant individual will pass the new insertion on throughout the generations. (**B**) New insertions that occurred in early development will only be carried through a new generation if the insertion is in germ progenitor cells (red). Alternatively, new insertions may not be present in germ cells, but in specific cells derived from ES cells (blue, yellow and red). In this case, the resultant individual will be a mosaic for new transposable insertions and only insertions in the germ cells (red) will be carried to the next generation. (**C**) Somatic new insertions will never be passed on to the next generation. As an example, insertions that occurred in neuronal progenitor cells during the development of an individual represent a genetic mosaic signature of the brain. The F1 progeny and subsequently derived generations will not have the same signature because the brain development will be re-set in each generation.

choroideremia, an X-linked progressive eye disease, revealed that the insertion occurred very early during the embryonic development of the patient's mother (73).

It is estimated that only 100 or so L1 copies in a given genome are capable of retrotransposition; roughly 10% of these active elements are classified as 'hot' or highly active in an artificial culture system (74), probably due to allele-specific differences in retrotransposition capability (75). Interestingly, L1 activity seems to vary depending on the population analyzed, indicating that there is substantial individual variation in retrotransposition capability (75). Of course, the ability of individual L1 elements to retrotranspose is related to the chromatin context in which the element is located in the genome, independent of being a 'hot' element in culture. The different levels of retrotransposition could be dependent not only on the population, but also on the epigenetic regulation that the full-length element is subjected to at a given time. We predict that such variation may be directly or indirectly able to influence individual personalities and predisposition to certain diseases.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

It has been more than 20 years since autonomous transposable elements were suggested to be selfish sequences, parasites in

the genome that increased the amount of 'junk' DNA in the different living organisms (76,77). In the interim, considerable evidence suggesting how transposable elements can shape the genome through evolution has challenged this idea. As more and more genomes are being sequenced, our understanding of the impact of transposable elements in the genome increases. The literature holds several examples that illustrate how transposons can serve as a reservoir of genetic innovation. Such innovation includes the creation of new coding or non-coding genes, with beneficial functions to the cells and/or modifying regulatory *cis* elements that may affect the balance of bound and free regulatory factors. Moreover, it is becoming clear, due to mechanistic similarities, that the epigenetic regulation of certain genes is derived from the defense mechanism against the activity of ancestral transposable elements.

Most of these modifications must have happened in germ cells or progenitor germ cells to be passed to the next generation, and only then be shaped by evolution. However, a completely new paradigm may be anticipated. ES cells are the perfect incubator for the transposable genetic toolbox to act, contributing to differences in gene regulation in a tissue-specific manner (Fig. 3). Furthermore, the realization that active transposons are mobile in somatic tissues brings a new dimension to the understanding of individual development and challenges the idea of a single and static genome per individual. Organisms are likely genetic mosaics influenced by environmental conditions to elicit *de novo* insertions from transposable elements. It remains to be seen how such genetic mosaicism contributes to individual differences, from cognition to disease predisposition.

ACKNOWLEDGEMENTS

REFERENCES

1. Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S. and Devine, S.E. (2006) Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.*, **78**, 671–679.
2. Mills, R.E., Bennett, E.A., Iskow, R.C. and Devine, S.E. (2007) Which transposable elements are active in the human genome? *Trends Genet.*, **23**, 183–191.
3. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
4. Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
5. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
6. Xiong, Y. and Eickbush, T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
7. Cavalier-Smith, T. (1991) Intron phylogeny: a new hypothesis. *Trends Genet.*, **7**, 145–148.
8. Zimmerly, S., Guo, H., Perlman, P.S. and Lambowitz, A.M. (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, **82**, 545–554.
9. Hasler, J. and Strub, K. (2006) Alu elements as regulators of gene expression. *Nucleic Acids Res.*, **34**, 5491–5497.
10. Hasler, J. and Strub, K. (2006) Alu RNP and Alu RNA regulate translation initiation *in vitro*. *Nucleic Acids Res.*, **34**, 2374–2385.
11. Batzer, M.A. and Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.*, **3**, 370–379.
12. Nishihara, H., Smit, A.F. and Okada, N. (2006) Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.*, **16**, 864–874.
13. Rubin, C.M., Kimura, R.H. and Schmid, C.W. (2002) Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res.*, **30**, 3253–3261.
14. Liu, W.M., Chu, W.M., Choudary, P.V. and Schmid, C.W. (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.*, **23**, 1758–1765.
15. Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A. and Batzer, M.A. (2005) SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.*, **354**, 994–1007.
16. Xing, J., Wang, H., Belancio, V.P., Cordaux, R., Deininger, P.L. and Batzer, M.A. (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl Acad. Sci. USA*, **103**, 17608–17613.
17. Gentles, A.J., Wakefield, M.J., Kohany, O., Gu, W., Batzer, M.A., Pollock, D.D. and Jurka, J. (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.*, **17**, 992–1004.
18. Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A. *et al.* (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature*, **447**, 167–177.
19. Roy-Engel, A.M., El-Sawy, M., Farooq, L., Odom, G.L., Perepelitsa-Belancio, V., Bruch, H., Oyeniran, O.O. and Deininger, P.L. (2005) Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.*, **110**, 365–371.
20. Boissinot, S., Davis, J., Entezam, A., Petrov, D. and Furano, A.V. (2006) Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl Acad. Sci. USA*, **103**, 9590–9594.
21. Song, M. and Boissinot, S. (2007) Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene*, **390**, 206–213.
22. Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A., Dyer, M., Cordaux, R., Liang, P. and Batzer, M.A. (2006) Human genomic deletions mediated by recombination between Alu elements. *Am. J. Hum. Genet.*, **79**, 41–53.
23. Zhang, Z., Harrison, P.M., Liu, Y. and Gerstein, M. (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.*, **13**, 2541–2558.
24. Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y. and Okada, N. (2003) Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol.*, **4**, R74.
25. Karro, J.E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrrison, P. and Gerstein, M. (2007) Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res.*, **35**, D55–D60.
26. Pavlicek, A., Gentles, A.J., Paces, J., Paces, V. and Jurka, J. (2006) Retroposition of processed pseudogenes: the impact of RNA stability and translational control. *Trends Genet.*, **22**, 69–73.

27. Dewannieux, M. and Heidmann, T. (2005) LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. *Cytogenet. Genome Res.*, **110**, 35–48.

28. Pain, D., Chirn, G.W., Strassel, C. and Kemp, D.M. (2005) Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. *J. Biol. Chem.*, **280**, 6265–6268.

29. Kleene, K.C., Mulligan, E., Steiger, D., Donohue, K. and Mastrangelo, M.A. (1998) The mouse gene encoding the testis-specific isoform of poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes. *J. Mol. Evol.*, **47**, 275–281.

30. Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N., Gage, F.H., Sue O'shea, K. and Moran, J.V. (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.*, **16**, 1569–1577.

31. Lin, H., Shabbir, A., Molnar, M. and Lee, T. (2007) Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. *Biochem. Biophys. Res. Commun.*, **355**, 111–116.

32. Svoboda, P., Stein, P., Hayashi, H. and Schultz, R.M. (2000) Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. *Development*, **127**, 4147–4156.

33. Wianny, F. and Zernicka-Goetz, M. (2000) Specific interference with gene function by double-stranded RNA in early mouse development. *Nat. Cell Biol.*, **2**, 70–75.

34. Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J. and Haussler, D. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, **441**, 87–90.

35. Pelisson, A., Sarot, E., Payen-Groschene, G. and Bucheton, A. (2007) A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J. Virol.*, **81**, 1951–1960.

36. Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. and Zamore, P.D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, **313**, 320–324.

37. Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H. and Siomi, M.C. (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.*, **20**, 2214–2222.

38. Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H. and Siomi, M.C. (2007) A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science*, **315**, 1587–1590.

39. Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P. and Kingston, R.E. (2006) Characterization of the piRNA complex from rat testes. *Science*, **313**, 363–367.

40. Grivna, S.T., Pyhtila, B. and Lin, H. (2006) MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc. Natl Acad. Sci. USA*, **103**, 13415–13420.

41. Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. and Hannon, G.J. (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, **316**, 744–747.

42. Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J. and Hannon, G.J. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, **443**, 1008–1012.

43. Soifer, H.S., Zaragoza, A., Peyvan, M., Behlke, M.A. and Rossi, J.J. (2005) A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res.*, **33**, 846–856.

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

45. Smalheiser, N.R. and Torvik, V.I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends Genet.*, **21**, 322–326.

46. Smalheiser, N.R. and Torvik, V.I. (2006) Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.*, **22**, 532–536.

47. Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G. and Munk, C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.*, **281**, 22161–22172.

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

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

50. Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R. and Weitzman, M.D. (2006) APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.*, **16**, 480–485.

51. Stenglein, M.D. and Harris, R.S. (2006) APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.*, **281**, 16837–16841.

52. Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T. and Tokunaga, K. (2007) All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.*, **35**, 2955–2964.

53. Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T. and Greene, W.C. (2006) High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc. Natl Acad. Sci. USA*, **103**, 15588–15593.

54. Yulug, I.G., Yulug, A. and Fisher, E.M. (1995) The frequency and position of Alu repeats in cDNAs, as determined by database searching. *Genomics*, **27**, 544–548.

55. Borchert, G.M., Lanier, W. and Davidson, B.L. (2006) RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.*, **13**, 1097–1101.

56. Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., Taccioli, G.E., Sekiguchi, J. and Moran, J.V. (2007) Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature*, **446**, 208–212.

57. Belancio, V.P., Hedges, D.J. and Deininger, P. (2006) LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.*, **34**, 1512–1521.

58. Zhang, X.H. and Chasin, L.A. (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc. Natl Acad. Sci. USA*, **103**, 13427–13432.

59. Neeman, Y., Levanon, E.Y., Jantsch, M.F. and Eisenberg, E. (2006) RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA*, **12**, 1802–1809.

60. Lev-Maor, G., Sorek, R., Levanon, E.Y., Paz, N., Eisenberg, E. and Ast, G. (2007) RNA-editing-mediated exon evolution. *Genome Biol.*, **8**, R29.

61. Polak, P. and Domany, E. (2006) Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics*, **7**, 133.

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

63. Kubo, S., Seleme Mdel, C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian, H.H., Jr. and Kasahara, N. (2006) L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl Acad. Sci. USA*, **103**, 8036–8041.

64. Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V. and Gage, F.H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, **435**, 903–910.

65. Muotri, A.R. and Gage, F.H. (2006) Generation of neuronal variability and complexity. *Nature*, **441**, 1087–1093.

66. Cao, X., Yeo, G., Muotri, A.R., Kuwabara, T. and Gage, F.H. (2006) Noncoding RNAs in the mammalian central nervous system. *Annu. Rev. Neurosci.*, **29**, 77–103.

67. Lee, J., Cordaux, R., Han, K., Wang, J., Hedges, D.J., Liang, P. and Batzer, M.A. (2007) Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene*, **390**, 18–27.

68. Batzer, M.A., Arcot, S.S., Phinney, J.W., Alegria-Hartman, M., Kass, D.H., Milligan, S.M., Kimpton, C., Gill, P., Hochmeister, M., Ioannou, P.A. *et al.* (1996) Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.*, **42**, 22–29.

69. Bennett, E.A., Coleman, L.E., Tsui, C., Pittard, W.S. and Devine, S.E. (2004) Natural genetic variation caused by transposable elements in humans. *Genetics*, **168**, 933–951.

70. Boissinot, S., Entezam, A., Young, L., Munson, P.J. and Furano, A.V. (2004) The insertional history of an active family of L1 retrotransposons in humans. *Genome Res.*, **14**, 1221–1231.

71. Witherspoon, D.J., Marchani, E.E., Watkins, W.S., Ostler, C.T., Wooding, S.P., Anders, B.A., Fowlkes, J.D., Boissinot, S., Furano, A.V., Ray, D.A. *et al.*

(2006) Human population genetic structure and diversity inferred from polymorphic L1 (LINE-1) and Alu insertions. *Hum. Hered.*, **62**, 30–46.

72. Ray, D.A., Walker, J.A. and Batzer, M.A. (2007) Mobile element-based forensic genomics. *Mutat. Res.*, **616**, 24–33.

73. van den Hurk, J.A., Meij, I.C., Del Carmen Seleme, M., Hoefsloot, L.H., Sistermans, E.A., de Wijs, I.J., Plomp, A.S., de Jong, P.T., Kazazian, H.H. and Cremers, F.P. (2007) L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.*, **16**, 1587–1592.

74. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian, H.H., Jr. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA*, **100**, 5280–5285.

75. Seleme Mdel, C., Vetter, M.R., Cordaux, R., Bastone, L., Batzer, M.A. and Kazazian, H.H., Jr. (2006) Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc. Natl Acad. Sci. USA*, **103**, 6611–6616.

76. Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604–607.

77. Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601–603.