

Review

LINE dancing in the human genome: transposable elements and disease

Victoria P Belancio[†], Prescott L Deininger^{*} and Astrid M Roy-Engel^{*}

Addresses: ^{*}Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane Cancer Center, Tulane University, SL-49 1430 Tulane Ave, New Orleans, LA 70112, USA. [†]Department of Structural and Cellular Biology, School of Medicine, Tulane Cancer Center and Tulane Center for Aging, Tulane University, SL-49 1430 Tulane Ave, New Orleans, LA 70112, USA.

Correspondence: Prescott L Deininger. Email: pdeinin@tulane.edu

Abstract

Transposable elements (TEs) have been consistently underestimated in their contribution to genetic instability and human disease. TEs can cause human disease by creating insertional mutations in genes, and also contributing to genetic instability through non-allelic homologous recombination and introduction of sequences that evolve into various *cis*-acting signals that alter gene expression. Other outcomes of TE activity, such as their potential to cause DNA double-strand breaks or to modulate the epigenetic state of chromosomes, are less fully characterized. The currently active human transposable elements are members of the non-LTR retroelement families, LINE-1, Alu (SINE), and SVA. The impact of germline insertional mutagenesis by TEs is well established, whereas the rate of post-insertional TE-mediated germline mutations and all forms of somatic mutations remain less well quantified. The number of human diseases discovered to be associated with non-allelic homologous recombination between TEs, and particularly between Alu elements, is growing at an unprecedented rate. Improvement in the technology for detection of such events, as well as the mounting interest in the research and medical communities in resolving the underlying causes of the human diseases with unknown etiology, explain this increase. Here, we focus on the most recent advances in understanding of the impact of the active human TEs on the stability of the human genome and its relevance to human disease.

Introduction to mammalian transposable elements

Transposable elements (TEs) occupy almost half, 46%, of the human genome, making the TE content of our genome one of the highest among mammals, second only to the opossum genome with a reported TE content of 52% [1,2]. The total representation of TE-related sequences in the human genome is probably even higher, as many of the sequences of the most ancient TEs have deteriorated beyond recognition [3]. The human genome contains two major classes of TEs, DNA and RNA transposons, defined by the type of molecule used as an intermediate in their mobilization.

DNA TEs encode a transposase that re-enters the nucleus to specifically recognize transposon sequences in chromosomal DNA. The transposase excises these sequences from their genomic location and inserts them into a new genomic site (reviewed in [4]); this is also referred to as 'cut and paste' transposition. Human DNA TE activity subsided over 37 million years ago [5]; as a result, DNA TEs no longer contribute significantly to the ongoing mutagenesis in humans.

Retrotransposons or retroelements make use of an RNA-mediated transposition process. Retroelements are subdivided into two major groups: those containing long-terminal repeats, LTR retroelements, and all others, lumped into the category of non-LTR retroelements. Although inactive in humans for millions of years, the best known LTR retrotransposons, the endogenous retroviruses, make up approximately 8% of the human genome [1]. This contrasts with rodent genomes, in which LTR elements continue to contribute a high proportion of the germline TE-associated mutations (reviewed in [6]).

Non-LTR retrotransposons include autonomous and non-autonomous members. The autonomous long interspersed element-1 (LINE-1 or L1), and its non-autonomous partners, such as 'SINE-R, VNTR, and Alu' (SVA) and the short interspersed element (SINE) Alu, are the only mobile elements with clear evidence of current retrotranspositional activity in the human genome [7] and will therefore be the primary focus of this article.

The human L1 is about 6 kb long and encodes two open reading frames, ORF1 and ORF2, which are both required for L1 retrotransposition (Figure 1a) [8]. ORF2 encodes endonuclease and reverse transcriptase activities that are crucial for the insertion mechanism [8,9]. SINEs and SVA elements do not encode any proteins [10], instead they

AML, acute myelogenous leukemia; *BRCA1*, breast cancer-1 gene; CGH, comparative genomic hybridization; DSB, double-strand break; LINE-1, L1, long interspersed element-1; LTR, long terminal repeat; NAHR, non-allelic homologous recombination; NHEJ, non-homologous end joining; ORF, open reading frame; SINE, short interspersed element; SVA, SINE-R, VNTR, and Alu element; TE, transposable element.

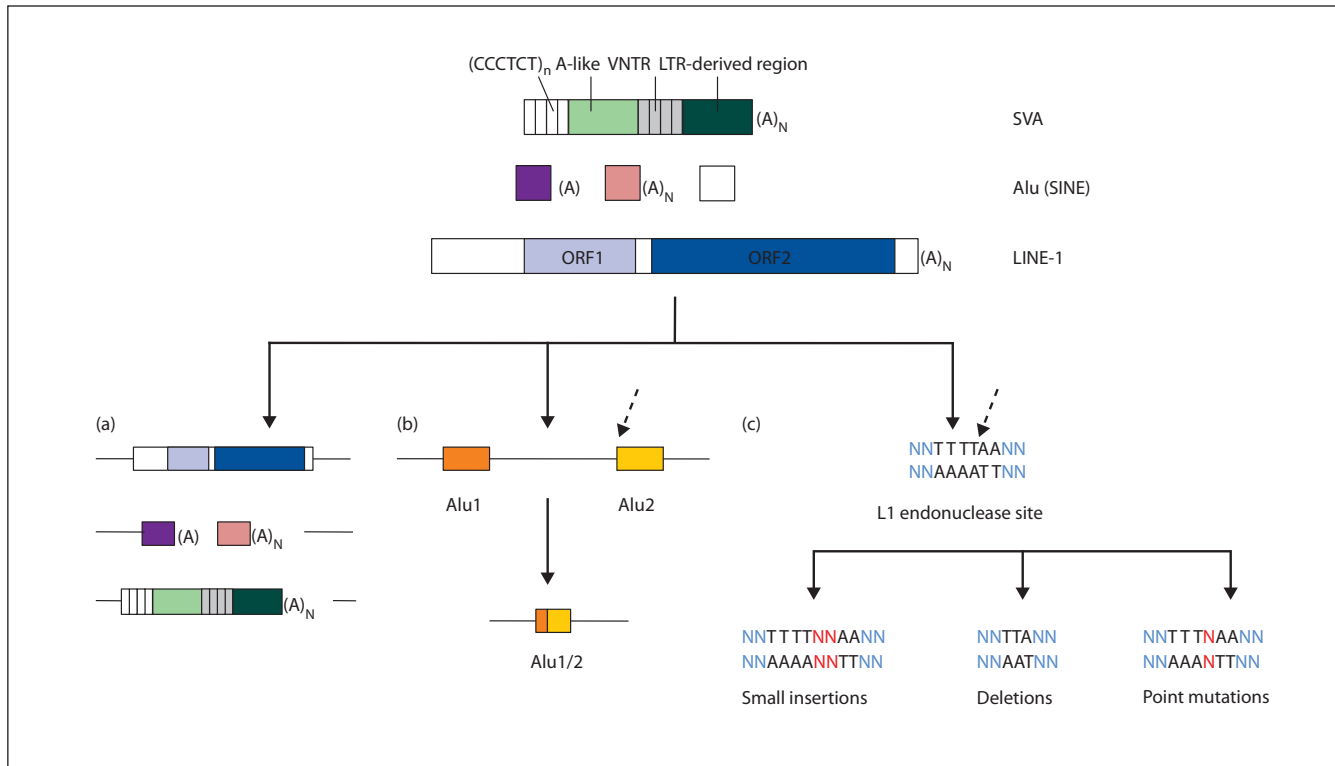


Figure 1

L1 expression leads to different types of DNA damage. Schematic structures of an SVA element (labeled SVA), showing the CCCTCT repeat, the Alu-derived (A-like) region, the variable number tandem repeat (VNTR) region, and the long terminal repeat (LTR)-derived region; an Alu element (labeled Alu (SINE)), showing left (purple) and right (pink) halves separated by the A-rich region (A) and the variable length A-tail ((A)_n) followed by the 3' region (white), which has a variable length and sequence; and an L1 element (labeled LINE-1), showing open reading frame (ORF)1 (light blue) and ORF2 (dark blue) and the 5' untranslated region, inter-ORF region and 3' untranslated region (white). **(a)** The typical insertion of these elements into the genome, which can lead to insertional mutagenesis. **(b)** Dispersed repetitive elements such as Alu elements can undergo non-allelic homologous recombination, which can cause a deletion (shown) or duplication (not shown). The dashed arrow indicates the potential site of DNA damage by an L1 endonuclease that may help initiate these recombination events. **(c)** Potential outcomes of the repair of the L1-induced double-strand breaks (DSBs). The L1 recognition site is in black; surrounding sequence is in blue; inserted nucleotides are in red. The associated changes are typical of what might be seen with repair of the DSB by non-homologous end joining. It is also possible that the sites are simply re-ligated with no mutation occurring, or alternatively, these sites may cause recombination, as shown in (b).

depend on the presence of the functional L1s, and they are therefore often referred to as L1 parasites [11]. In contrast to L1, Alu elements require only ORF2 of L1 for their mobilization [11,12]. Alu elements are transcribed by RNA polymerase III and encode a variable length adenosine-rich region at their 3' end, a critical feature for retrotransposition [10]. SVA is a composite element containing a complex sequence composed of a (CCCTCT)_n hexamer repeat region, an Alu-derived region, a variable number tandem repeat (VNTR) region and a retroviral-derived sequence (Figure 1a) [13]. The requirements for SVA mobilization are still poorly understood [13,14].

TE activity has often been assumed to be confined to the germline, early embryogenesis, and potentially cancer cells [15–18]. The most recent reports indicate that expression of L1 RNA (VP Belancio, A Roy-Engel, R Pochampally and

P Deininger, personal communication) and L1 protein [16] occurs in human somatic tissues and that somatic L1 retrotransposition takes place in transgenic mouse models [19,20]. Interestingly, L1 transgenic mice show higher L1 mobilization in somatic tissues than in the germline [19,21]. Other evidence of somatic L1 mobilization comes from a somatic L1 insertion that inactivates the adenomatous polyposis coli (APC) gene, leading to colon cancer [22]. There are currently very limited data on the somatic expression of Alu and SVA elements, and we do not have a true appreciation of the level of somatic insertion that is occurring from endogenous elements.

Human diseases caused by TE-mediated insertional mutagenesis

The most obvious form of mutagenesis common to all TEs is the disruption of gene function or regulation resulting

from the insertion of new element copies (Figure 1b). The fortuitous discovery of the first known active human L1 was the result of its retrotransposition into the factor VIII gene, causing a *de novo* case of hemophilia [23]. L1, Alu, and SVA are reported to cause a broad range of human diseases (reviewed in [7,10,24]). Examples include a diverse collection of diseases, such as neurofibromatosis, choroideremia, cholinesterase deficiency, Apert syndrome, Dent's disease, β -thalassemia, and Walker-Warburg syndrome. Because of the relatively random insertion process, there is great diversity in the type of genetic diseases associated with TE insertions. However, there is a very strong overrepresentation of X-chromosome-linked diseases caused by TEs that could be a result of ascertainment bias (that is, X-linked genetic defects are more easily detected because of the single X-chromosome in males or could also reflect the higher density of L1 elements on the X chromosome). Compilations of the known human diseases attributed to TE insertions (33 Alu, 11 L1, and 4 SVA) are provided in recent reviews [7,10]. Most of these diseases are due to germline insertions and have been detected as rare recessive diseases. However, some cases of cancers have been identified that are probably somatic mutations in which a TE insertion has disrupted a critical gene, such as *BRCA1* and *BRCA2* in breast cancer or *APC* in colon cancer.

Interference with gene expression

Almost all of the reported retroelement insertions that cause human diseases have either interrupted the ORF or inserted in close proximity to a splice site, leading to a major disruption of gene function [7,10]. However, many insertions that do not cause disease may still influence the expression of the genes in which they insert, thus predisposing cells or individuals to disease by slightly changing gene expression. For example, insertion of TE elements might introduce functional splice and polyadenylation sites [25-29], resulting in aberrant processing of some of the transcripts produced by a gene. In addition, they might introduce regulatory regions that would influence the strength of its promoter, or even add promoter sequences [30,31]. It has been suggested that L1 elements inserted in the intron of a gene could cause 'gene breaking' [25,28] that could create proteins truncated from either end, possibly leading to altered functions or dominant-negative effects. In contrast to L1, Alu elements need to accumulate a critical mutation(s) that creates an appropriate functional *cis*-acting sequence (both splicing and polyadenylation) to have this effect [32-34].

Human disease caused by post-insertional TE mutagenesis

Recombination

TEs continue to contribute to genetic instability after insertion through non-allelic homologous recombination (NAHR). The presence of multiple closely related sequences throughout the genome facilitates misalignment

of repeated sequences, allowing uneven genetic exchange between alleles that contribute to deletions and duplications (Figure 1c; reviewed in [35,36]). Comparisons of the human and chimpanzee genomes have shown that L1 and Alu recombination deletions caused over a megabase of difference in more than 100 individual deletions [37-39]. Alu elements not only cause deletions, but also seem to contribute to the formation of segmental duplications. A genome-wide set of 2,366 duplication alignments demonstrated the enrichment of Alu elements near the junction between the two duplicated sequences in all cases, suggesting Alu involvement in these rearrangements [40]. These segmental duplications lead to altered expression of the genes located in these regions and result in further instability by promoting non-allelic recombination between duplicated segments, leading to recurrent genetic disease.

TE-mediated NAHR (in particular, recombination between Alu elements) contributes directly to a large variety of genetic diseases. The frequency of this type of genetic rearrangement varies depending on the affected gene (reviewed in [35,41]). Genes such as *MLL-1* (which is involved in acute myelogenous leukemia (AML)) [42], *VHL* (von Hippel-Lindau syndrome) [43], and *BRCA1* (familial breast cancer) [44] seem to be hotspots for Alu-Alu recombination, with a series of independent recombination events occurring with different Alu elements in the region. *BRCA1* has 137 Alus in its introns, making up over 40% of its gene sequence. Studies of *BRCA1* mutations have shown that, in 23 different individuals, 44 of these Alu elements were involved in duplication/deletion events in this gene. *VHL* is also subject to extensive Alu-Alu recombination, with almost a third of its germline mutations resulting from large deletions, and 90% of the mapped events involving Alu-Alu NAHR [45]. Of 30 Alu-Alu recombination events mapped, seven involve one particular Y-subfamily Alu element recombining with other Alus in the gene. The Y-subfamily is young and therefore shows lower than average divergence relative to other genomic Alus, which might explain its high recombination rate. Similar observations implicating a particular 'hotspot' Alu element were reported for multiple cases of rearrangements in the LDL receptor gene causing familial hypercholesterolemia [46], and also for the *SLC7A7* gene, where one Alu accounted for 38% of all rearranged chromosomes in patients with lysinuric protein intolerance [47].

The *MLL1* gene, which is associated with AML, is often involved in chromosomal translocations causing expression of an oncogenic fusion gene. Of the cases of AML without a visible translocation, seven out of nine cases studied involved a duplication caused by Alu-Alu recombination events in intron 1 and 6, which resulted in a duplication of exons 2 to 6 of the gene [42]. Similar duplications have consistently been found in the blood of healthy individuals [48], suggesting that these types of

recombination events occur spontaneously and regularly throughout the lifespan of an individual. The cellular environment can potentially increase Alu-Alu NAHR. Mutants in *TP53* (which encodes the tumor-suppressor p53) increase these Alu-Alu recombination events, possibly contributing to malignancy. Although it is clear that active human TEs contribute to spontaneous genetic diseases, the exact extent of their involvement in this process remains elusive. This uncertainty makes the contribution of TEs to human diseases difficult to assess, for the most part due to the absence of uniform and reliable diagnostic methods.

Detection and diagnosis of diseases caused by TE insertion

The introduction of PCR technology for diagnostics revolutionized the field of human genetic testing. Within a clinical setting, PCR across the exons of a gene and sequence analysis is commonplace. This approach is great for identifying point mutations and small insertion/deletion events, but will detect only small TE insertions very near the exons. Most PCR-based tests are inadequate for the detection of the large deletions, rearrangements, and duplications often associated with TE-induced mutagenesis. Awareness of this bias led to the use of alternative methods that can detect copy number variations (CNVs), such as long-range PCR, targeted array comparative genome hybridization (array-CGH) analysis, and multiplex ligation-dependent probe assays (MLPAs). These approaches are better at detection of duplicated or deleted exons that occur from Alu-Alu recombination events. New tests using either MLPA or array-CGH are becoming more commonplace, particularly for diagnostics in cancer, and can detect most genomic duplications and deletions but not larger TE insertions. Traditional Southern blot analysis is still one of the few robust methods for detecting large TE insertions but it is rarely used in diagnostic tests today. The fact that the majority of sporadic human diseases have a subset of cases of unknown etiology leaves a possibility that TE-induced DNA damage may be responsible for at least some of them. In fact, genomic analysis by methods specifically targeting potential involvement of TEs in the sporadic human diseases revealed that a significant proportion of them are, indeed, caused by TEs [43]. One of the most promising technologies for characterizing all of the TE-based variation with minimal ascertainment bias is the potential usage of some of the upcoming next-generation DNA sequencing approaches for random sequencing of the entire genome of an individual. However, this approach is still some years from clinical usefulness.

L1-associated DNA double-strand breaks

Recent publications from several laboratories have reported the formation of DNA double-strand breaks (DSBs) associated with L1 expression [49-51]. These DSBs depend on the enzymatic activity of the L1 ORF2 endonuclease domain [51], and their formation triggers

various cellular responses [51,52], including apoptosis, cellular senescence, cell-cycle checkpoints, and DNA repair responses. DSBs are highly mutagenic and can lead to small deletions or insertions if repaired by the non-homologous end-joining (NHEJ) repair machinery (Figure 1d). L1-induced DSBs may also cause recombination events when repaired by homology-driven repair, potentially leading to large genomic rearrangements (reviewed in [41]). Homologous recombination (HR) repairs damaged genomic sequence either by using the unaltered counterpart as a template in a gene conversion event or through non-allelic homologous interactions that lead to deletions or duplications between the homologous sequences, as described above for Alu element-mediated NAHR. Given that all L1, Alu, and SVA copies in the human genome are generated with target site duplications that contain an L1 endonuclease recognition site, there are roughly 3 million potential cleavage sites adjacent to these elements that may help them contribute to NAHR-mediated events. Because many L1-endonuclease-mediated events may lack the typical hallmarks of L1 involvement (such as the target site duplications that normally flank mobilized sequences and a run of adenosines), we cannot currently assess the relative contribution of this process to genetic instability.

Modulators of TE activity

The TE amplification cycle involves complex interactions with various cellular factors and compartments, any of which can be positively or negatively regulated by intrinsic or extracellular environmental factors. The L1 lifecycle and some of its known modulators are depicted in Figure 2.

Modulations by the genomic environment

Levels of TE activity can vary both because of the polymorphism of these elements between different individuals, as well as variations in epigenetic regulation of TE loci. Even though each human genome averages 500,000 L1 copies, of which about 3,000 are full-length and roughly 200 are potentially functional [1,53], only a handful of elements have high levels of activity in each genome [53]. The rest have mutated sufficiently to lose most or all retrotransposition potential. All of the highly active elements found to date are polymorphic in the population, with each individual probably having a different assortment of active elements [53]. Because these loci consist of the youngest L1 integration events, they have had the least time to accumulate inactivating mutations and are more likely to remain active. In addition to the presence/absence polymorphism of these 'hot' elements, the same L1 loci accumulate distinct point mutations in various individuals that contribute to the diversity in their potential activity [54,55]. Thus, there may be as much as several hundred-fold variability in L1 activity in different individuals [55]. Recent advances in understanding of the sequence components controlling Alu activity [56-58]

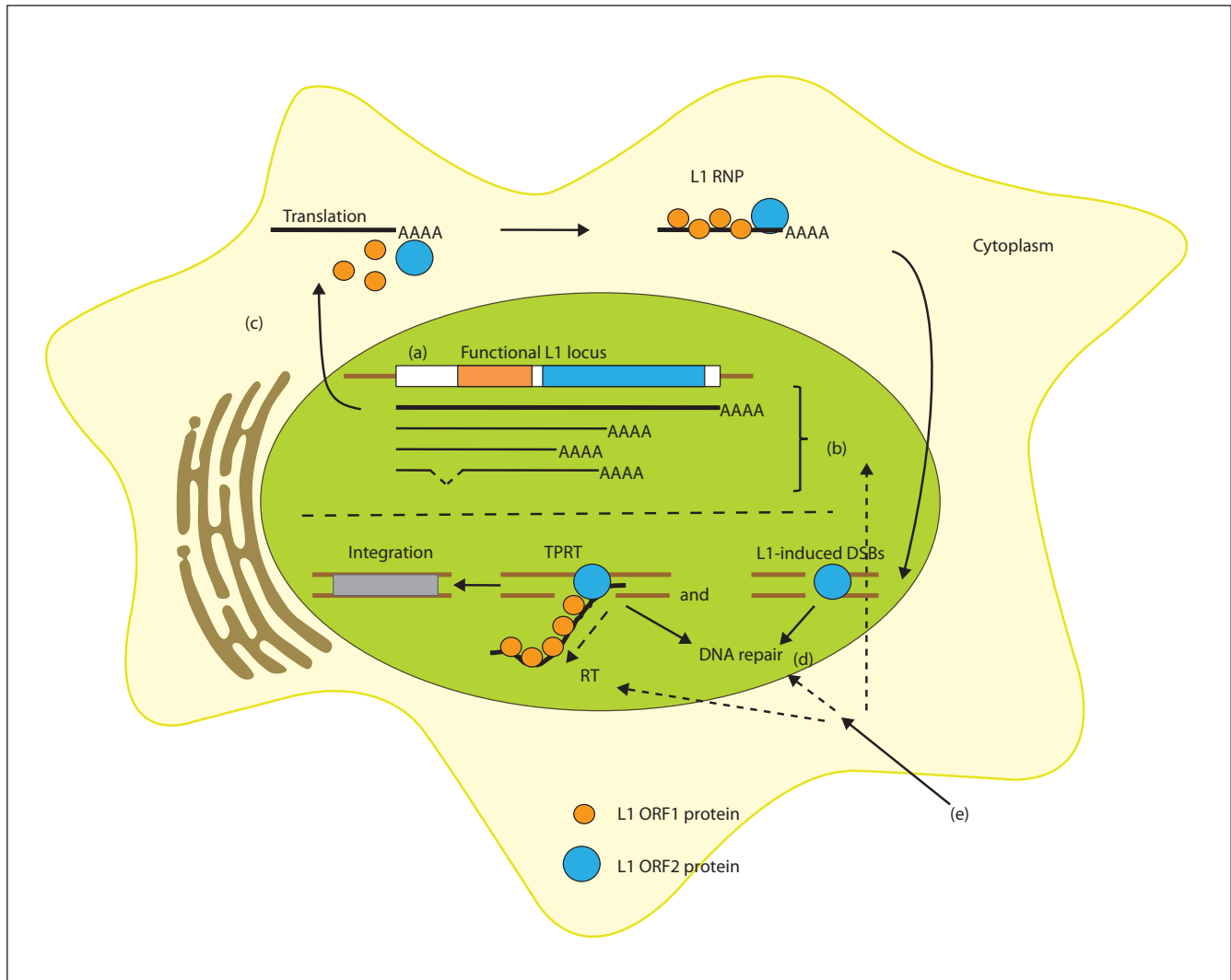


Figure 2

Modulators of the L1 lifecycle. The L1 amplification cycle can be divided into several steps. **(a)** Transcription. L1 amplification initiates with transcription, and regulation of L1 at this step can be modified by epigenetic modifications, DNA methylation, and recruitment of transcription factors. **(b)** Before leaving the nucleus, the number of retrocompetent full-length L1 transcripts can be reduced by RNA processing through premature polyadenylation and splicing. **(c)** Translation. Full-length L1 enters the cytoplasm to be translated, producing ORF1 and ORF2 proteins for retrotransposition. The two proteins interact with the L1 transcript to form an L1 ribonucleoprotein particle (RNP). RNA interference can affect this step. **(d)** Insertion of a new L1 copy. The L1 RNP reaches the nucleus, where the DNA is cleaved by the L1 ORF2 endonuclease activity. It is proposed that reverse transcription occurs through a process referred to as target primed reverse transcription (TPRT) [71]. The L1 ORF2 reverse transcriptase activity generates the first strand of DNA. DNA repair proteins are likely to be involved in inhibiting the L1 integration step. **(e)** Effects of external stimuli. Ionizing radiation or heavy metals can affect L1 at multiple steps, such as transcriptional activation or altering DNA repair pathways.

indicate that its retrotransposition is also likely to vary in individual genomes.

DNA methylation of the CpG island in the 5' region of L1 [59] is one of the powerful mechanisms controlling L1 promoter activity that minimizes the exposure of genomic DNA to L1-associated damage. The genome-wide hypomethylation of repetitive sequences observed during

malignant transformation unleashes L1 expression that is usually tightly regulated in untransformed cells [60]. Methylation of genomic DNA often triggers specific histone modifications, resulting in chromatin remodeling. The role of epigenetic control in L1 expression has recently attracted significant interest, particularly because little is known about the effects of the intronic or near-genic full-length L1 insertions on the epigenetic state of the affected human genes.

Modulations by the cellular environment

Among the multiple cellular pathways influencing L1 expression and activity are DNA methylation, tissue-specific transcription factors (Figure 2a), RNA processing (Figure 2b), and RNA interference (Figure 2c) [25-27,29, 61-63]. In addition, some cellular proteins greatly influence integration (Figure 2d) of L1 and Alu elements; these include DNA repair proteins, such as the ataxia telangiectasia mutated kinase (ATM) and the endonuclease dimer composed of excision repair complementing protein 1 (ERCC1) and xeroderma pigmentosum complement group (XPF) [51,64,65], and also viral defense proteins, such as the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C (APOBEC3) family of proteins [66,67] (Figure 2). L1 mobilization in NHEJ-negative hamster cells causes the element to lose the endonuclease dependence that it shows in a wild-type genetic background, and it then requires only functional L1 reverse transcriptase to achieve wild-type retrotransposition levels [68]. Because of the diversity of the L1-associated mutagenesis, it will not be surprising if additional DNA repair pathways are reported to modulate L1 activity.

Given the multitude of cellular factors influencing L1 activity, it is easy to imagine that polymorphisms or mutations in any of the genes whose function is important for suppressing L1 activity may have an impact on its contribution to genetic instability. One of the most profound examples is the mouse knockout of DNA-methyltransferase-3-like protein (Dnmt3L), a modulator of *de novo* DNA methylation in the germline, which results in upregulation of the expression of endogenous L1 and LTR elements that coincides with meiotic catastrophe during spermatogenesis [69,70].

Modulations by the extracellular environment

TE activity is influenced not only by the intrinsic cellular environment, but also by external stimuli (Figure 2e). Ionizing radiation, heavy metals (present in cigarette smoke and workplace exposures), anti-cancer therapies, air pollutants, and DNA demethylation agents can locally or systemically cause increases in endogenous TE activity (reviewed in [50,70]), potentially leading to new health problems (such as sporadic cancers) or exacerbating preexisting conditions (such as the rise of a more aggressive cancer phenotype). The mechanisms of the environmental influences on human TE activity are only just beginning to emerge as we are learning more about their interaction with various cellular pathways. Some of the environmental factors enhance TE expression by changing the epigenetic state of the genome; others, such as heavy metals, probably exert their effect by influencing cellular enzymes that are important for keeping TE activity at bay. Because of the early stage of this area of investigation, no diseases have yet been directly associated with increased activity of TEs due to exposure to

environmental toxicants. However, with the new advances in whole-genome studies, some of these crucial questions are likely to be answered in the near future.

Conclusions

TE activity can generate a wide-spectrum of genomic mutations, ranging from point mutations to gross rearrangements with gain of genomic information, as well as interference with normal gene processing and expression after insertion. These mutations contribute to idiopathic human disease. Because of the intimate relationship between L1 activity and multiple cellular processes, it is likely that people with genetic backgrounds that produce defects in any of the pathways influencing the L1 lifecycle are more vulnerable to insult from TEs. Thus, to evaluate the impact of these elements on the stability of the human genome and human disease, it is crucial to take into account their cumulative activity in a specific genetic background as well as the potential modulating effects of the extracellular environment.

The increasing ease of sequencing genomes is likely to help clarify the extent of the contribution of mobile elements to genetic instability in many human diseases. This information is critical in determining the full spectrum of mutations contributing to human disease. However, the full impact of these ubiquitous, high-copy-number elements on the biology of the cell may remain elusive for some time.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors participated equally in the conception and writing of this article.

Acknowledgements

This article was made possible by grants P20RR020152 (PLD, VPB, and AMR-E), R01GM45668 (PLD), and R01GM079709A (AMR-E) from the National Institutes of Health (NIH) and an EPSCOR grant from the National Science Foundation (PLD). VPB is supported by NIH/NIA grant 5K01AG030074 and an Ellison Medical Foundation New Scholar in Aging award (AG-NS-0447-08). The contents of the article are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Research Resources or the NIH. Competitive Advantage Funds (2006) from the Louisiana Cancer Research Consortium (LCRC) were also awarded to AMR-E.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
2. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, Lopez-Otin C, Ordonez GR, Eichler EE, Chen L, Cheng Z, Deakin JE,

- Alsop A, Thompson K, Kirby P, *et al.*: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453**:175-183.
3. Gu W, Castoe TA, Hedges DJ, Batzer MA, Pollock DD: **Identification of repeat structure in large genomes using repeat probability clouds.** *Anal Biochem* 2008, **380**:77-83.
 4. Sinzelle L, Izsvak Z, Ivics Z: **Molecular domestication of transposable elements: from detrimental parasites to useful host genes.** *Cell Mol Life Sci* 2009, **66**:1073-1093.
 5. Pace JK, Feschotte C: **The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage.** *Genome Res* 2007, **17**:422-432.
 6. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL: **Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line.** *PLoS Genet* 2006, **2**:e2-
 7. Chen JM, Stenson PD, Cooper DN, Ferec C: **A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease.** *Hum Genet* 2005, **117**:411-427.
 8. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr: **High frequency retrotransposition in cultured mammalian cells.** *Cell* 1996, **87**:917-927.
 9. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD: **Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.** *Cell* 1996, **87**:905-916.
 10. Belancio VP, Hedges DJ, Deininger P: **Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health.** *Genome Res* 2008, **18**:343-358.
 11. Dewannieux M, Esnault C, Heidmann T: **LINE-mediated retrotransposition of marked Alu sequences.** *Nat Genet* 2003, **35**:41-48.
 12. Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM: **LINE-1 ORF1 protein enhances Alu SINE retrotransposition.** *Gene* 2008, **419**:1-6.
 13. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr: **SVA elements are nonautonomous retrotransposons that cause disease in humans.** *Am J Hum Genet* 2003, **73**:1444-1451.
 14. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA: **SVA elements: a hominid-specific retroposon family.** *J Mol Biol* 2005, **354**:994-1007.
 15. Branciforte D, Martin SL: **Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition.** *Mol Cell Biol* 1994, **14**:2584-2592.
 16. Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Stratling WH, Schumann GG: **Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues.** *J Biol Chem* 2004, **279**:27753-27763.
 17. Martin SL: **Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells.** *Mol Cell Biol* 1991, **11**:4804-4807.
 18. Martin SL, Branciforte D: **Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells.** *Mol Cell Biol* 1993, **13**:5383-5392.
 19. An W, Han JS, Schrum CM, Maitra A, Koentgen F, Boeke JD: **Conditional activation of a single-copy L1 transgene in mice by Cre.** *Genesis* 2008, **46**:373-383.
 20. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH Jr: **L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism.** *Genes Dev* 2009, **23**:1303-1312.
 21. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr: **L1 integration in a transgenic mouse model.** *Genome Res* 2006, **16**:240-250.
 22. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y: **Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer.** *Cancer Res* 1992, **52**:643-645.
 23. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr: **Isolation of an active human transposable element.** *Science* 1991, **254**:1805-1808.
 24. Babushok DV, Kazazian HH Jr: **Progress in understanding the biology of the human mutagen LINE-1.** *Hum Mutat* 2007, **28**:527-539.
 25. Belancio VP, Hedges DJ, Deininger P: **LINE-1 RNA splicing and influences on mammalian gene expression.** *Nucleic Acids Res* 2006, **34**:1512-1521.
 26. Belancio VP, Roy-Engel AM, Deininger P: **The impact of multiple splice sites in human L1 elements.** *Gene* 2008, **411**:38-45.
 27. Perepelitsa-Belancio V, Deininger P: **RNA truncation by premature polyadenylation attenuates human mobile element activity.** *Nat Genet* 2003, **35**:363-366.
 28. Wheelan SJ, Aizawa Y, Han JS, Boeke JD: **Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution.** *Genome Res* 2005, **15**:1073-1078.
 29. Han JS, Szak ST, Boeke JD: **Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes.** *Nature* 2004, **429**:268-274.
 30. Matlik K, Redik K, Speek M: **L1 antisense promoter drives tissue-specific transcription of human genes.** *J Biomed Biotechnol* 2006, **2006**:1-16.
 31. Speek M: **Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes.** *Mol Cell Biol* 2001, **21**:1973-1985.
 32. Chen C, Ara T, Gautheret D: **Using Alu elements as polyadenylation sites: a case of retroposon exaptation.** *Mol Biol Evol* 2009, **26**:327-334.
 33. Lee YK, Chew A, Phan H, Greenhalgh DG, Cho K: **Genome-wide expression profiles of endogenous retroviruses in lymphoid tissues and their biological properties.** *Virology* 2008, **373**:263-273.
 34. Sorek R, Ast G, Graur D: **Alu-containing exons are alternatively spliced.** *Genome Res* 2002, **12**:1060-1067.
 35. Deininger PL, Batzer MA: **Alu repeats and human disease.** *Mol Genet Metab* 1999, **67**:183-193.
 36. Deininger PL, Batzer MA: **Mammalian retroelements.** *Genome Res* 2002, **12**:1455-1465.
 37. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA: **Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages.** *Nucleic Acids Res* 2005, **33**:4040-4052.
 38. Han K, Lee J, Meyer TJ, Wang J, Sen SK, Srikanta D, Liang P, Batzer MA: **Alu recombination-mediated structural deletions in the chimpanzee genome.** *PLoS Genet* 2007, **3**:1939-1949.
 39. Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA: **L1 recombination-associated deletions generate human genomic variation.** *Proc Natl Acad Sci USA* 2008, **105**:19366-19371.
 40. Bailey JA, Liu G, Eichler EE: **An Alu transposition model for the origin and expansion of human segmental duplications.** *Am J Hum Genet* 2003, **73**:823-834.
 41. Hedges DJ, Deininger PL: **Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity.** *Mutat Res* 2006, **616**:46-59.
 42. Strout MP, Marcucci G, Bloomfield CD, Caligiuri MA: **The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia.** *Proc Natl Acad Sci USA* 1998, **95**:2390-2395.
 43. Franke G, Bausch B, Hoffmann MM, Cybulla M, Wilhelm C, Kohlhasse J, Scherer G, Neumann HP: **Alu-Alu recombination underlies the vast majority of large VHL germline deletions: molecular characterization and genotype-phenotype correlations in VHL patients.** *Hum Mutat* 2009, **30**:776-786.
 44. Mazoyer S: **Genomic rearrangements in the BRCA1 and BRCA2 genes.** *Hum Mutat* 2005, **25**:415-422.
 45. Casarin A, Martella M, Polli R, Leonardi E, Anesi L, Murgia A: **Molecular characterization of large deletions in the von Hippel-Lindau (VHL) gene by quantitative real-time PCR: the hypothesis of an Alu-mediated mechanism underlying**

- VHL gene rearrangements.** *Mol Diagn Ther* 2006, **10**:243-249.
46. Kim SH, Bae JH, Chae JJ, Kim UK, Choe SJ, Namkoong Y, Kim HS, Park YB, Lee CC: **Long-distance PCR-based screening for large rearrangements of the LDL receptor gene in Korean patients with familial hypercholesterolemia.** *Clin Chem* 1999, **45**:1424-1430.
 47. Font-Llitjos M, Rodriguez-Santiago B, Espino M, Sillue R, Manas S, Gomez L, Perez-Jurado LA, Palacin M, Nunes V: **Novel SLC7A7 large rearrangements in lysinuric protein intolerance patients involving the same AluY repeat.** *Eur J Hum Genet* 2009, **17**:71-79.
 48. Schnittger S, Wormann B, Hiddemann W, Griesinger F: **Partial tandem duplications of the MLL gene are detectable in peripheral blood and bone marrow of nearly all healthy donors.** *Blood* 1998, **92**:1728-1734.
 49. Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A: **Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells.** *Cancer Cell Int* 2006, **6**:13.
 50. Farkash EA, Kao GD, Horman SR, Prak ET: **Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay.** *Nucleic Acids Res* 2006, **34**:1196-1204.
 51. Gasior SL, Wakeman TP, Xu B, Deininger PL: **The human LINE-1 retrotransposon creates DNA double-strand breaks.** *J Mol Biol* 2006, **357**:1383-1393.
 52. Wallace NA, Belancio VP, Deininger PL: **L1 mobile element expression causes multiple types of toxicity.** *Gene* 2008, **419**:75-81.
 53. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr: **Hot L1s account for the bulk of retrotransposition in the human population.** *Proc Natl Acad Sci USA* 2003, **100**:5280-5285.
 54. Lutz SM, Vincent BJ, Kazazian HH Jr, Batzer MA, Moran JV: **Allelic heterogeneity in LINE-1 retrotransposition activity.** *Am J Hum Genet* 2003, **73**:1431-1437.
 55. Seleme MC, Vetter MR, Cordaux R, Bastone L, Batzer MA, Kazazian HH Jr: **Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity.** *Proc Natl Acad Sci USA* 2006, **103**:6611-6616.
 56. Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL: **Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die?** *Genome Res* 2009, **19**:545-555.
 57. Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL: **Active Alu element "A-tails": size does matter.** *Genome Res* 2002, **12**:1333-1344.
 58. Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE: **Active Alu retrotransposons in the human genome.** *Genome Res* 2008, **18**:1875-1883.
 59. Hata K, Sakaki Y: **Identification of critical CpG sites for repression of L1 transcription by DNA methylation.** *Gene* 1997, **189**:227-234.
 60. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, Ehrlich M, Laird PW: **Analysis of repetitive element DNA methylation by MethyLight.** *Nucleic Acids Res* 2005, **33**:6823-6836.
 61. Yang N, Zhang L, Zhang Y, Kazazian HH: **An important role for RUNX3 in human L1 transcription and retrotransposition.** *Nucleic Acids Res* 2003, **31**:4929-4940.
 62. Yang N, Kazazian HH Jr: **L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells.** *Nat Struct Mol Biol* 2006, **13**:763-771.
 63. Yu F, Zingler N, Schumann G, Stratling WH: **Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription.** *Nucleic Acids Res* 2001, **29**:4493-4501.
 64. Gasior SL, Roy-Engel AM, Deininger PL: **ERCC1/XPF limits L1 retrotransposition.** *DNA Repair (Amst)* 2008, **7**:983-989.
 65. Suzuki J, Yamaguchi K, Kajikawa M, Ichiyangi K, Adachi N, Koyama H, Takeda S, Okada N: **Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition.** *PLoS Genet* 2009, **5**:e1000461.
 66. Hulme AE, Bogerd HP, Cullen BR, Moran JV: **Selective inhibition of Alu retrotransposition by APOBEC3G.** *Gene* 2007, **390**:199-205.
 67. Stenglein MD, Harris RS: **APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.** *J Biol Chem* 2006, **281**:16837-16841.
 68. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV: **DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.** *Nat Genet* 2002, **31**:159-165.
 69. Bourc'his D, Bestor TH: **Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.** *Nature* 2004, **431**:96-99.
 70. Belancio V, Roy-Engel AM: **Xenobiotics-modulation of human mobile elements and genetic instability.** In *Encyclopedia of Environmental Health*. Edited by Nriagu JO. Amsterdam: Elsevier; 2009:1-10.
 71. Cost GJ, Feng Q, Jacquier A, Boeke JD: **Human L1 element target-primed reverse transcription in vitro.** *EMBO J* 2002, **21**:5899-5910.

Published: 27 October 2009
doi:10.1186/gm97
© 2009 BioMed Central Ltd