

Mobile Elements in the Human Genome: Implications for Disease

Benoît Chénais

▶ **To cite this version:**

Benoît Chénais. Mobile Elements in the Human Genome: Implications for Disease. eLS, John Wiley & Sons, Ltd., 2017, 10.1002/9780470015902.a0024377 . hal-01901944

HAL Id: hal-01901944
https://univ-lemans.hal.science/hal-01901944

Submitted on 23 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mobile Elements in the Human Genome: Implications for Disease

Benoît Chénais, *Le Mans Université, Le Mans, France*

Abstract

Transposable elements (TEs) are mobile DNA sequences that represent a great portion of the human genome. TEs are divided into retrotransposons, the more abundant, and DNA transposon, which differ by their structure and transposition mechanisms. The major *LINE-1* and *Alu* elements represent about one third of the human genome and are responsible for several chromosomal rearrangements, gene deletion, insertional mutagenesis as well as RNA splicing alteration and epigenetic regulation. DNA transposons and human endogenous retroviruses are also a source of genetic alterations. Consequently, TE insertion and mobility lead to human diseases including (but not limited to) cancer, hematologic disease, metabolic disease, neurodegenerescence, neurologic and psychiatric diseases. The herein presented examples of TE-related diseases aim to highlight the interest of increasing our knowledge of TEs through extensive genomic approaches, which may lead to new potential diagnostic markers of diseases.

Key words: Transposable elements, transposon, retrotransposon, human disease, chromosome rearrangement, gene recombination, cancer, neurologic disease, metabolic disease, genetic disease

Key concepts

- Transposable element are numerous in the human genome and the most abundant *Alu* and *LINE-1* retro-elements represent about one third of the human genome
- Transposable elements are source of genomic rearrangements that may lead to several human diseases
- Gene expression may be affected by transposable elements through the modification of regulatory elements and epigenetic control

- Examples of human disease related to transposable element include (but are not limited to) cancer, hematologic, metabolic and neurologic diseases

Introduction: Mobile elements in the human genome

Mobile elements or transposable elements (TEs) represent a substantial fraction of the human genome, from 45% according to the most common data (Lander *et al.*, 2001) and even up to 66% according to further analysis (de Koning *et al.*, 2011). The two major characteristics of these DNA sequences are their mobility in a given genome and across different genomes and their ability to increase in number in the genome, both leading to an important role in genome plasticity and evolution. TEs are diverse in nature and two main classes are identified on the basis of their mechanism of transposition: class-I (retro-) elements, or retrotransposons, use a reverse transcriptase enzyme to copy an RNA transcript into the host DNA, whereas class-II elements, or DNA transposons, mainly move through a "cut and paste" mechanism involving the excision and reinsertion of the DNA sequence of the element. In addition, a large number of repeated sequences are related to TEs but non-autonomous, i.e. dependent of the enzyme(s) encoded by a helper element, which is an autonomous TE, in order to ensure their mobility and amplification in the genome. The short interspersed nuclear elements (SINEs) are the leader class of such non-autonomous TEs that have greatly invaded the human genome (de Koning *et al.*, 2011). See also: DOI: 10.1002/9780470015902.a0001899.pub3; DOI: 10.1002/9780470015902.a0006126.pub3

By far the most successful type of TEs in the human genome is class-I elements, which is subdivided into Long Terminal Repeats (LTR) and non-LTR elements. The latter are very predominant and 76% of human TEs are non-LTR elements, which represent at least 33.6% of the whole human genome (Lander *et al.*, 2001) (**Fig. 1**). On its own, the long interspersed element (LINE)-1 accounts for 38% of the human TEs and 16.9% of the entire genome (**Fig. 1**), followed by the non-autonomous SINEs, particularly the *Alu* repeats, which account for 24% of TEs and 10.6% of the genome (de Koning *et al.*, 2011, Lander *et al.*, 2001). Other SINEs found in humans are MIR and MIR3 elements. The LTR-elements are less present in the human genome and mainly represented by endogenous retroviruses (HERVs, including *HERV-I*, *HERV-K*, *HERV-L*) that resemble retroviruses in both their structure and mobility

mechanism, but lacking a functional envelope gene. See also: DOI: 10.1002/9780470015902.a0026264; DOI: 10.1002/9780470015902.a0026711

Class-II elements, or DNA transposons, are present in small amounts in the human genome, accounting for only 6% of TEs and 2.8% of the entire genome (**Fig. 1**). They mainly belong to the TC-1/*mariner* superfamily (*i.e. mariner*, MER2-Tigger, Tc2), *hAT*superfamily (*i.e.* MER-1-*Charlie*, *Zaphod*), and some *PiggyBac*-like elements (Lander *et al.*, 2001). See also: DOI: 10.1002/9780470015902.a0026264; DOI: 10.1002/9780470015902.a0020996.pub2

Long been considered as junk DNA or parasite DNA, TEs are now accepted as an evolutionary force that contributes to the genetic diversity of organisms. TE could be recruited on behalf of the genome to become a new gene and such TE exaptation or domestication is a source of genetic innovation, but the transposition activity may also contribute to human disorders through chromosome rearrangements, insertional mutagenesis, deregulation of gene expression, and epigenetic changes (Chénais, 2015). See also: DOI: 10.1002/9780470015902.a0026265; DOI: 10.1002/9780470015902.a0020996.pub2

Mechanisms of TE's impact on human genome

Several mechanisms are involved to explain the impact of TE on the genome and gene expression (Chénais, 2015). Genome rearrangements are favoured by the repetition of TE in the genome and the abundance of *LINE-1* and *Alu* repeats (i.e. about 17% and 11% of the human genome, respectively) promotes ectopic recombination, leading to gene deletions, duplications and chromosome translocations or inversions (**Table 1**). In addition, *LINE-1* and *Alu* elements are frequently observed within or near the breakage points of chromosomal rearrangement and *LINE-1* alone account for 19% of the 2081 breakpoints junction analysed in a dataset of 17 whole human genomes (Kidd *et al.*, 2010). Although less present in the human genome, DNA transposons may also be a source of chromosome rearrangement as shown by a study conducted on ten types of hematolymphoid neoplasms (Rodic *et al.*, 2013). This work showed that most translocations have no propensity to involve genomic repeats, but 47% of the translocations at the *TCF3* locus on chromosome 19 involve a *MER20* DNA transposon (Rodic *et al.*, 2013). See also: DOI: 10.1002/9780470015902.a0005493.pub2; DOI: 10.1002/9780470015902.a0020996.pub2

The second way in which TEs may impact the human genome and induce diseases is through the modification of gene structure and regulatory regions (**Table 2**). This include the creation of new polyadenylation sites (**Table 3**), new exons (also called "exonisation") and new introns (or "intronisation") as well as alternative splicing modifications, e.g. exon skipping and intron retention (**Table 2, Fig. 2**) (Chénais, 2015). The creation of a new exon causes a frameshift or creates premature stop codons in about 79% of cases, resulting in non-functional sequences, which are often alternatively spliced. However, exonisation events following *LINE-1* or *Alu* insertion may lead to human diseases (**Table 2**) (reviewed in Chénais, 2015). Using the LTR of HERV as promoter/enhancer is also a way to modify gene expression. This was observed, for example, in the case of the *MID1* gene, associated with the X-linked form of Opitz syndrome, which transcript is fused to *HERV-E* LTR acting as a tissue-specific promoter/enhancer (Landry *et al.*, 2002). See also: DOI: 10.1002/9780470015902.a0005493.pub2; DOI: 10.1002/9780470015902.a0020996.pub2; DOI: 10.1002/9780470015902.a0022886; DOI: 10.1002/9780470015902.a0026711

The third way to modify the expression of the genome concerns the epigenetic mechanisms induced or modified by TEs. A correlation between the activity of TEs and chromatin regulation, mainly DNA methylation, has been observed and hypomethylation of DNA is a cause of TEs activation (Ross *et al.*, 2010). The relationship between pathological markers and the methylation status of whole genome CpG islands and LINE-1 elements has been particularly studied in the case of cancers (Chénais 2015). In addition, the repression of TEs through RNA interference involved essentially endogenous small interfering RNA (endo-siRNA) in somatic cells and PIWI-interacting RNA (piRNA) in the germline (Saito & Siomi, 2010). Both endo-siRNA and piRNA are transcribed from TEs themselves and *a priori* do not disturb the expression of protein-coding genes (Saito & Siomi, 2010). Cancer cells are known to contain significantly lower quantities of micro-RNAs, but currently, in the absence of endo-siRNA and piRNA expression pattern in pathologic cells, it is impossible to answer the question of the putative impact of TE-related regulatory RNA on cancer or other diseases. See also: DOI: 10.1002/9780470015902.a0023593; DOI: 10.1002/9780470015902.a0025688; DOI: 10.1002/9780470015902.a0026316

Involvement of TEs in human cancer

- ***Alu/LINE-1 insertion***

Several examples of cancers have been reported in the literature as related to LINE or SINE/*Alu* insertion (**Table 1**) (reviewed in Chénais, 2015). More recently an increased insertional activity of *LINE-1* and an elevated level of *LINE-1* transcript were found in primary tumour samples of breast invasive carcinoma, head and neck squamous cell carcinoma and lung adenocarcinoma (Clayton *et al.*, 2016). The insertion of the *LINE-1* retroelement into the human proto-oncogene *c-MYC* was reported as early as in 1988 in human breast carcinoma cells (Morse *et al.*, 1988). This study compared the structure of the *c-MYC* gene from breast ductal adenocarcinoma and from normal breast tissue of the same patient and revealed a tumour-specific rearrangement of one *MYC* locus and the amplification of the other *MYC* locus. A non-*MYC* sequence with near complete homology to a *LINE-1* sequence was found within the second intron of the rearranged locus. In this case, the *LINE-1* sequence functioned as a mobile genetic element to produce a somatic mutation (Morse *et al.*, 1988). Another example of a *LINE-1* element somatic insertion was observed in the tumor-suppressing gene *APC* (adenomatous polyposis coli) from several patients with a colon cancer and also in association with desmoid tumors (reviewed in Chénais, 2015). The *BRCA1* and *BRCA2* genes, which are associated with breast and ovarian cancer, are also the sites of *Alu* insertions. For example, one of the first evidenced germline mutations was the insertion of a transcriptionally active *Alu* element into exon 22 of *BRCA2*, which resulted in alternative splicing that skipped exon 22 (reviewed in Chénais, 2015). Finally, *Alu* insertions in the von Hippel-Lindau (*VHL*) gene cause large alteration at the origin of the von Hippel-Lindau disease, a familial autosomal-dominant syndrome, predisposing to the development of benign and malignant tumors (Franke *et al.*, 2009). See also: DOI: 10.1002/9780470015902.a0005493.pub2

- ***Alu recombination leading to deletion, duplication or translocation***

As noted above *Alu* repeats account for 11% of the human genome (**Fig. 1**). These repeats are thought to induce non-allelic homologous recombination, or ectopic recombination, at the origin of several cancer diseases. Numerous deletion events have been reported at the *BRCA1* locus and in a lesser extent for the *BRCA2* gene, related to the unusually high density of *Alu* repeats (**Table 1**). The insertion of *Alu* elements in the *VHL* gene is also involved in *Alu*-mediated deletions associated with cancer (Franke *et al.*, 2009). Large deletions in the

cadherin-1 (*CDH1*) gene occur in 4% of hereditary diffuse gastric cancer families (Oliveira *et al.*, 2009). These deletions involve mainly non-allelic homologous recombination in *Alu* repeat sequences, as supported by the statistically significant over-representation of *Alu* elements around breakpoints (Oliveira *et al.*, 2009). Homologous recombination between *Alu* sequences and subsequent deletion have been detected at high frequency within the caspase-activated DNase (*CAD*) gene in 13 out of the 20 human hepatoma tissues and in eight of the para-cancerous cirrhotic livers (Hsieh *et al.*, 2005) (**Table 1**). The recombination caused exon-3 deletion, which in turn led to exon-3 skipping or replacement with a partial *Alu* sequence, and consequential C-terminal truncation of the CAD protein (Hsieh *et al.*, 2005). These data highlight the role of *Alu* repeats in mediating genome instability in human cancers. See also: DOI: 10.1002/9780470015902.a0005493.pub2

Homologous recombination between *Alu* elements results sometimes in gene duplication as reported for the *BRCA1* gene, the *MLL-1* (myeloid/lymphoid or mixed-lineage leukemia) gene involved in acute myeloid leukemia, and the *MYB* proto-oncogene the duplication of which leads to T-cell acute lymphoblastic leukemia (reviewed in Chénais, 2015). Duplication of the *MYB* locus, which encodes an essential transcription factor, is frequent in human cancers. The human *MYB* locus is flanked by *Alu* repeats that are responsible for its duplication through somatic homologous recombination between the flanking *Alu* elements on sister chromatids as observed in T cell acute lymphoblastic leukemia (O'Neil *et al.*, 2007).

The recombination between *Alu* repeats is also involved in the generation of the Philadelphia chromosome (*i.e.* translocation between chromosomes 9q34 and 22q11) leading to *BCR/ABL* gene fusion in chronic myeloid leukemia (Jeffs *et al.*, 1998). Finally, the translocation between chromosomes 5q23-31 and 18q12, which involves the *TRE-2* oncogene (TRE-USP6, ubiquitin-specific protease-6), is due to recombination between *Alu* elements and play an important role in Ewing sarcoma development (Onno *et al.*, 1992).

- ***TE-related epigenetic regulation and cancer***

The relation between *LINE-1* activation and the methylation status of DNA has been particularly investigated in the case of cancers. Indeed the frequent hypomethylation of chromatin in tumour cells is regarded as a facilitator of TE mobility, and especially concerning *LINE-1* (Ross *et al.*, 2010). Whole genome studies revealed the preferential *LINE-1* insertion in hypomethylated DNA regions of cancer cells of different origins (Ferreira *et al.*,

2014, Tubio *et al.*, 2014). Besides, several studies highlighted the hypomethylation of the *LINE-1* element itself in several types of cancers (Chénais 2015) and the transcriptome analysis of chronic lymphocytic leukaemia showed that TEs are globally derepressed (Ferreira *et al.*, 2014). However Lee et al. have shown that *LINE-1* and *Alu* insertions occurred preferentially in epithelial cancer but not in blood and brain cancer (Lee *et al.*, 2012). In this study the authors identified almost two hundred somatic insertions of *LINE-1* (94.5%), *Alu* (5.0%) and HERV (0.5%) in 43 high-coverage whole genome sequencing datasets from five cancer types (i.e. colorectal, glioblastoma, multiple myeloma, ovarian and prostate cancer) (Lee *et al.*, 2012). Finally, in the case of melanoma, the *LINE-1* hypomethylation is accompanied by the shortened relapse-free survival of patients, and a global under-methylation of DNA impacts the metastatic formation of primary melanomas (Ecsedi *et al.*, 2013).

Epigenetic modification of TE's regulatory sequences can impact the expression of neighbouring genes. When comparing the histone enrichment, TE neighbourhood and expression divergence of human genes between normal and cancer cells, a correlation was observed between the presence of TEs near genes and greater changes in histone enrichment (Grégoire *et al.*, 2016).

TEs in hematologic, metabolic and/or genetic diseases

- ***Alu*-mediated deletion causing human diseases**

In addition to cancer, several human diseases can be attributed to *Alu*-mediated rearrangements (**Table 1**). Deletions involving *Alu* repeats account for numerous mutations in several other genes, among which the apolipoprotein B (*APOB*) gene, the LDL receptor (*LDLR*) gene, and the α-globin and β-globin gene clusters. In a patient with homozygous hypobetalipoproteinemia, the deletion of exon 21 of *APOB* is the consequence of a crossover event between the *Alu* sequences in introns 20 and 21 resulting in nonreciprocal exchange between two chromosomes (Huang *et al.*, 1989). The *LDLR* gene is also a place of large deletions (or duplications) through homologous recombination between *Alu* repetitive elements, which are numerous within almost all introns and in the 3′-UTR (reviewed in (Hobbs *et al.*, 1992). For example, the analysis of autosomal dominant hypercholesterolemia in two Tunisian families showed the presence of a large deletion in the *LDLR* gene due to

intrachromatid non-allelic homologous recombination between two homologous *Alu* sequences located apart from splice sites in intronic sequences (Jelassi *et al.*, 2012).

The reported TE-mediated rearrangements in the β-globin gene cluster include γ-, δ- and β-globin genes (Harteveld *et al.*, 1997, Yu *et al.*, 2014). For example, in an individual with hereditary persistence of foetal haemoglobin, the β-globin cluster displays a large deletion of 48.5 kb downstream of the β-globin gene and the 5' deletion breakpoint occurs in an *Alu* element 3' to the Aγ-globin gene. However, in this particular case, the recombination mechanism involve non-homologous recombination since the 3' breakpoint was located within a region that contains various repeats including a portion of a *LINE-1* repeat, a perfect 160-bp palindrome, and a set of 41-bp direct repeats also present elsewhere in the human genome (Henthorn *et al.*, 1986).

Concerning the α-globin gene cluster, recombination events are numerous and globally lead to diverse forms of thalassemia. Following are a few examples of *Alu*- and/or *LINE-1*-mediated recombination that lead to α-thalassemia, an inherited hemoglobin disorder characterized by a microcytic hypochromic anaemia caused by a quantitative reduction of the α-globin chain. In a Chinese family with α-thalassemia patients, a 28.5 kb deletion involving *Alu* repeats has been described that eliminated one of the duplicated α-globin genes (Yu *et al.*, 2014). A 33-kb α^0-thalassaemia deletion encompassing the α- and ζ-globin genes and pseudogenes has been observed in a Dutch family and this deletion appears similar to a previously described one (Harteveld *et al.*, 1997). The molecular analysis of deletion breakpoints indicate that the mechanism leading to these α^0-thalassaemia deletions involves misalignment between the highly homologous *Alu* tandem repeats at both parental sides, which are normally 33 kb apart. Moreover, comparison of breakpoint positions along the *Alu* consensus sequence indicate the involvement of a 26-bp core sequence in two out of five α^0-thalassaemia deletions, suggesting that this sequence is a possible hotspot of recombination (Harteveld *et al.*, 1997). This suggests that *Alu* repeats could stimulate recombination events not only by homologous pairing, but also by providing binding sites for recombinogenic proteins (Harteveld *et al.*, 1997). Finally a 8.2 kb deletion in the α-globin gene cluster was found in a Dutch family. This deletion involves both α-globin genes in *cis* and is caused by a non-homologous recombination event between an *Alu* and a *LINE-1* element (Phylipsen *et al.*, 2010). See also: DOI: 10.1002/9780470015902.a0005493.pub2

- ***Alu- and LINE-1-mediated exonisation***

The creation of a new exon, so called exonisation (**Fig. 2A, Table 2**), is involved in the ornithine delta-aminotransferase (OAT) deficiency, which causes the gyrate atrophy of the choroid and retina, a progressive loss of vision. In this case, a G-to-C mutation in an *Alu* element present in the intron 3 of the *OAT* gene results in the creation of a new donor splice site and potentially to a new exon (Mitchell *et al.*, 1991). This *Alu* insertion also causes premature termination, which again results in OAT deficiency (Mitchell *et al.*, 1991). The *Alu*-mediated exonisation in the gene of the collagen type IV alpha-3 chain (*COL4A3*) may be at the origin of the Alport syndrome, a nephropathy (glomerulonephritis) due to a collagen IV defect (Knebelmann *et al.*, 1995). In the same way, *Alu*-mediated exonisation has been reported in the glucuronidase beta (*GUSB*) gene that causes the Sly syndrome or type VII mucopolysaccharidosis (Vervoort *et al.*, 1998). The *LINE-1* element is also a potential source of exonisation as observed for the X-linked *CYBB* gene, encoding the beta chain of cytochrome B245 and involved in at least one case of chronic granulomatous disease (Meischl *et al.*, 2000). The *LINE-1* insertion into intron 5 of the *CYBB* gene cause internal rearrangements and new splice sites that result in a highly heterogeneous splicing pattern with introduction of two *LINE-1* fragments as new exons into the transcripts and concomitant skipping of exonic coding sequence (Meischl *et al.*, 2000).

- ***Alu- and LINE-1 mediated alternative splicing and intron retention***

Alternative splicing is often affected by TE insertion (**Table 2**). For example, insertion of the non-autonomous SINE *Alu* element into an intron of the human Factor VIII (*F8*) gene results in exon skipping (**Fig. 2B**) and consequent onset of haemophilia A (Ganguly *et al.*, 2003). Another example is the insertion of an *Alu* element within the exon 9, or just upstream of the start of exon 9, of the fibroblast growth factor receptor-2 (*FGFR2*) gene, which was reported as a cause of Apert syndrome (Oldridge *et al.*, 1999). The latter mutation affects the utilisation of the 3' splice site leading to the generation of different, and non-functional, splice forms of *FGFR2* transcripts (Oldridge *et al.*, 1999).

The dystrophin (*DMD*) gene is a giant gene, sizing about 2.5 Mb of the X chromosome with more than 60 exons, where numerous TEs have inserted to create deletions or exon skipping. For example the insertion of *LINE-1* induces skipping of exon 44, causing Duchenne muscular dystrophy (Narita *et al.*, 1993). Insertion of an *Alu*-like sequence downstream of

intron 11 was observed in individuals with X-linked dilated cardiomyopathy (Ferlini *et al.*, 1998). The insertion induces a rearrangement that activates a cryptic splice site in intron 11 and produces an alternative transcript containing the *Alu*-like sequence and part of the adjacent intron 11, spliced between exons 11 and 12, but not translated due to the presence of numerous stop codons (Ferlini *et al.*, 1998). The *LINE-1* element is responsible for a large deletion in the X-linked pyruvate dehydrogenase E1 (*PDHX*) gene, but the *LINE-1* insertion may also lead, at last in one patient, to an aberrantly spliced isoform resulting from the use of two cryptic splice sites (Mine *et al.*, 2007).

Finally, intron retention (**Fig. 2C**) may happen when an error occurs in normal splicing leading to retention of a transcript-confirmed intron within a transcript-confirmed exon and thus in the mature mRNA. Such events may be affected by TEs, as demonstrated in the case of *LINE-1* insertion in the *CYBB* gene that causes chronic granulomatous disease (Meischl *et al.*, 2000).

- ***Alu and LINE-1 derived new polyadenylation sites***

LINE-1 elements include a polyadenylation signal within their own sequence, and the polyadenylation signal (i.e. AATAAA sequence) is frequently created in the A-rich tails of both SINEs and LINEs. Therefore human retroelements, such as *LINE-1* and *Alu*, may introduce intragenic polyadenylation signals creating new mRNA isoforms. Several diseases have been linked to the introduction of a new polyadenylation site through *LINE-1* or *Alu* element insertion in human genes, such as haemophilia A and B, X-linked dilated cardiomyopathy, autoimmune lymphoproliferative syndrome, hypo-calciuric hypercalcemia and neonatal severe hyperparathyroidism as reviewed in Chénais (2015) (**Table 3**).

Besides, the LTR sequence of HERV is also a source of new polyadenylation signal as observed with *HERV-K* in the breast cancer cell line T47D (Baust *et al.*, 2000) and with *HERV-H* in the *HHLA2* and *HHLA3* genes (Mager *et al.*, 1999); but no obvious relation with a pathology was described to date.

- ***LINE-1 and chromatin regulation***

As described above for cancer, the chromatin structure and/or epigenetic regulation associated with TE insertion may lead to non-cancer human diseases but examples are rare to date. The β-globin$_{L1}$ allele results from the insertion of a full length *LINE-1* element in the

intron 2 of the β-globin gene. The decreased rate of transcription of this β-globin$_{L1}$ allele is due to the hypermethylated profile of its promoter and enhancer sequences, which results in a transcriptional down-regulation and thus the β^{+}-thalassemia phenotype (Lanikova *et al.*, 2013).

TE involvement in neurological diseases

As for cancer, metabolic and hematologic diseases, TE insertion may be the cause of neurological diseases or syndromes. The insertion of a *LINE-1* element in the ribosomal S6 kinase 2 (*RSK2*) gene was reported in a patient with Coffin-Lowry syndrome, a disease characterised by psychomotor and growth retardation, facial dysmorphy and squeletic abnormalities. This *LINE-1* insertion at position -8 of intron 3 leads to a skipping of exon 4 and then to a reading-frame shift and the appearance of a premature stop codon (**Table 2**) (Martinez-Garay *et al.*, 2003).

The Fabry disease, characterized by a neurodegenerescence of the central nervous system, is an X-linked recessive disorder of glycosphingolipid catabolism, which results from the deficient activity of the lysosomal hydrolase α-galactosidase A encoded by an *Alu*-rich gene, that displays 12 *Alu* elements within its 12 kb sequence (Kornreich *et al.*, 1990). Several mutations and gene rearrangements have been described to be at the origin of the disease (**Table 1**), among which only one deletion was originally identified as an *Alu-Alu* recombination, and ten other as *Alu-Alu* rearangements (Kornreich *et al.*, 1990). Thereafter, a large 3.1 kb deletion due to *Alu-Alu* recombination and including the whole exon 2 was identified among 50 new mutations at the origin of Fabry disease (Shabbeer *et al.*, 2006).

The Sandhoff disease, a recessive lysosomal storage disease, is another example of TE-related neurodegenerative disease (**Table 1**). The deficience of β-hexosaminidase (HEXB) activity is due to deletion alleles of the *HEXB* gene in 27% of the examined dataset (Neote *et al.*, 1990). *HEXB* contains two *Alu* repeats that are involved in the deletion event, which removes about 16kb including the *HEXB* promoter, exons 1 to 5, and part of intron 5. The deletion also resulted in a reconstituted *Alu* sequence, the left half of which comes from the 5' *Alu* sequence and the right half of which from the 3' *Alu* sequence (Neote *et al.*, 1990).

Neurofibromatosis type I is a common genetic disease linked to the *NF-1* tumor suppressor gene and characterized by changes in skin coloring and the growth of benign tumors along nerves (i.e. neurofibromas). The insertion of an *Alu* sequence in the intron of the *NF-1* gene leads to a deletion and reading frame shift in the downstream exon during splicing, which could be a cause of the disease (**Table 2**) (Wallace *et al.*, 1991). See also: DOI: 10.1002/9780470015902.a0005493.pub2

Elevated expression of TEs has been observed in several human neurodegenerative disorders including sporadic Creutzfeld-Jakob disease, age-related macular degeneration and amyotrophic lateral sclerosis (ALS). For example, the RNA-binding protein TDP-43, which is involved in ALS and frontotemporal lobar degeneration (FTLD), was shown to interact with many TE transcripts and moreover this association of TDP-43 with many of its TE targets is reduced in FTLD patients (Li *et al.*, 2012).

Finally, a link between TEs and psychiatric disorders was also highlighted (Guffanti *et al.*, 2014). Particularly, whole-genome sequencing revealed brain specific *LINE-1* insertion in patients localized preferentially to synapse- and schizophrenia-related genes, suggesting a causal role of *LINE-1* in the onset of the disease (Bundo *et al.*, 2014).

Conclusion

First considered as junk DNA, TEs are now considered as major actors of genome evolution and their role in genome plasticity is now obviously demonstrated. This involves several molecular mechanisms besides TEs insertion *per se*, including DNA recombination and chromosome rearrangements (especially deletion), modification of gene expression through the introduction or alteration of poly-A sites, splicing sites, exons or introns and regulatory elements, and finally the domestication of TEs leading to new functionalities.

The high abundance of TEs in the human genome, especially *Alu* and *LINE-1* repeats, make them responsible for the molecular origin of several diseases as presented above. Therefore increasing our knowledge of TEs through extensive genomic approaches may lead to new potential diagnostic markers of diseases and prenatal markers of genetic disorders. In this way the recent development of combined molecular and bioinformatics tools is of great interest, such as the work by Tang *et al.*, which have detected somatic *LINE-1* insertions in ovarian cancer (Tang *et al.*, 2017).

References

Baust C, Seifarth W, Germaier H, Hehlmann R, Leib-Mosch C (2000) HERV-K-T47D-related long terminal repeats mediate polyadenylation of cellular transcripts. *Genomics* **66**(1): 98-103.

Bundo M, Toyoshima M, Okada Y *et al.* (2014) Increased l1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* **81**(2): 306-313.

Chénais B (2015) Transposable elements in cancer and other human diseases. *Current Cancer Drug Targets* **15**(3): 227-242.

Clayton EA, Wang L, Rishishwar L *et al.* (2016) Patterns of transposable element expression and insertion in cancer. *Frontiers in Molecular Biosciences* **3**: 76.

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics* **7**(12): e1002384.

Ecsedi SI, Hernandez-Vargas H, Lima SC *et al.* (2013) Transposable hypomethylation is associated with metastatic capacity of primary melanomas. *International Journal of Clinical and Experimental Pathology* **6**(12): 2943-2948.

Ferlini A, Galie N, Merlini L *et al.* (1998) A novel *Alu*-like element rearranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy. *American Journal of Human Genetics* **63**(2): 436-446.

Ferreira PG, Jares P, Rico D *et al.* (2014) Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Research* **24**(2): 212-226.

Franke G, Bausch B, Hoffmann MM *et al.* (2009) *Alu-Alu* recombination underlies the vast majority of large *VHL* germline deletions: Molecular characterization and genotype-phenotype correlations in VHL patients. *Human Mutation* **30**(5): 776-786.

Ganguly A, Dunbar T, Chen P, Godmilow L, Ganguly T (2003) Exon skipping caused by an intronic insertion of a young *Alu YB9* element leads to severe hemophilia A. *Human Genetics* **113**(4): 348-352.

Grégoire L, Haudry A, Lerat E (2016) The transposable element environment of human genes is associated with histone and expression changes in cancer. *BMC Genomics* **17**: 588.

401 Guffanti G, Gaudi S, Fallon JH *et al.* (2014) Transposable elements and psychiatric disorders.
402 *American Journal of Medical Genetics, part B: Neuropsychiatric Genetics* **165B**(3): 201-
403 216.
404 Harteveld KL, Losekoot M, Fodde R, Giordano PC, Bernini LF (1997) The involvement of *Alu*
405 repeats in recombination events at the alpha-globin gene cluster: Characterization of
406 two alphazero-thalassaemia deletion breakpoints. *Human Genetics* **99**(4): 528-534.
407 Henthorn PS, Mager DL, Huisman TH, Smithies O (1986) A gene deletion ending within a
408 complex array of repeated sequences 3' to the human beta-globin gene cluster.
409 *Proceedings of the National Academy of Sciences of the United States of America* **83**(14):
410 5194-5198.
411 Hobbs HH, Brown MS, Goldstein JL (1992) Molecular genetics of the *LDL* receptor gene in
412 familial hypercholesterolemia. *Human Mutation* **1**(6): 445-466.
413 Hsieh SY, Chen WY, Yeh TS, Sheen IS, Huang SF (2005) High-frequency *Alu*-mediated genomic
414 recombination/deletion within the caspase-activated DNase gene in human hepatoma.
415 *Oncogene* **24**(43): 6584-6589.
416 Huang LS, Ripps ME, Korman SH, Deckelbaum RJ, Breslow JL (1989)
417 Hypobetalipoproteinemia due to an apolipoprotein B gene exon 21 deletion derived by
418 *Alu-Alu* recombination. *The Journal of Biological Chemistry* **264**(19): 11394-11400.
419 Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM (1998) The *BCR* gene recombines
420 preferentially with *Alu* elements in complex *BCR-ABL* translocations of chronic myeloid
421 leukaemia. *Human Molecular Genetics* **7**(5): 767-776.
422 Jelassi A, Slimani A, Rabes JP *et al.* (2012) Genomic characterization of two deletions in the
423 *LDLR* gene in tunisian patients with familial hypercholesterolemia. *Clinica Chimica Acta*
424 **414**: 146-151.
425 Kidd JM, Graves T, Newman TL *et al.* (2010) A human genome structural variation
426 sequencing resource reveals insights into mutational mechanisms. *Cell* **143**(5): 837-847.
427 Knebelmann B, Forestier L, Drouot L *et al.* (1995) Splice-mediated insertion of an *Alu*
428 sequence in the *COL4A3* mRNA causing autosomal recessive Alport syndrome. *Human*
429 *Molecular Genetics* **4**(4): 675-679.
430 Kornreich R, Bishop DF, Desnick RJ (1990) Alpha-galactosidase A gene rearrangements
431 causing Fabry disease. Identification of short direct repeats at breakpoints in an *Alu*-rich
432 gene. *The Journal of Biological Chemistry* **265**(16): 9319-9326.

433 Lander ES, Linton LM, Birren B *et al.* (2001) Initial sequencing and analysis of the human
434 genome. *Nature* **409**(6822): 860-921.
435 Landry JR, Rouhi A, Medstrand P, Mager DL (2002) The Opitz syndrome gene *MID1* is
436 transcribed from a human endogenous retroviral promoter. *Molecular Biology and*
437 *Evolution* **19**(11): 1934-1942.
438 Lanikova L, Kucerova J, Indrak K *et al.* (2013) Beta-thalassemia due to intronic LINE-1
439 insertion in the beta-globin gene (HBB): Molecular mechanisms underlying reduced
440 transcript levels of the beta-globin(L1) allele. *Human Mutation* **34**(10): 1361-1365.
441 Lee E, Iskow R, Yang L *et al.* (2012) Landscape of somatic retrotransposition in human
442 cancers. *Science* **337**(6097): 967-971.
443 Li W, Jin Y, Prazak L, Hammell M, Dubnau J (2012) Transposable elements in TDP-43-
444 mediated neurodegenerative disorders. *PLoS One* **7**(9): e44099.
445 Mager DL, Hunter DG, Schertzer M, Freeman JD (1999) Endogenous retroviruses provide the
446 primary polyadenylation signal for two new human genes (*HHLA2* and *HHLA3*). *Genomics*
447 **59**(3): 255-263.
448 Martinez-Garay I, Ballesta MJ, Oltra S *et al.* (2003) Intronic L1 insertion and F268S, novel
449 mutations in RPS6KA3 (RSK2) causing Coffin-Lowry syndrome. *Clinical Genetics* **64**(6):
450 491-496.
451 Meischl C, Boer M, Ahlin A, Roos D (2000) A new exon created by intronic insertion of a
452 rearranged LINE-1 element as the cause of chronic granulomatous disease. *European*
453 *Journal of Human Genetics* **8**(9): 697-703.
454 Mine M, Chen JM, Brivet M *et al.* (2007) A large genomic deletion in the *PDHX* gene caused
455 by the retrotranspositional insertion of a full-length LINE-1 element. *Human Mutation*
456 **28**(2): 137-142.
457 Mitchell GA, Labuda D, Fontaine G *et al.* (1991) Splice-mediated insertion of an *Alu* sequence
458 inactivates ornithine delta-aminotransferase: A role for *Alu* elements in human mutation.
459 *Proceedings of the National Academy of Sciences of the United States of America* **88**(3):
460 815-819.
461 Morse B, Rotherg PG, South VJ, Spandorfer JM, Astrin SM (1988) Insertional mutagenesis of
462 the MYC locus by a LINE-1 sequence in a human breast carcinoma. *Nature* **333**(6168): 87-
463 90.

Narita N, Nishio H, Kitoh Y *et al.* (1993) Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *The Journal of Clinical Investigation* **91**(5): 1862-1867.

Neote K, McInnes B, Mahuran DJ, Gravel RA (1990) Structure and distribution of an *Alu*-type deletion mutation in Sandhoff disease. *The Journal of Clinical Investigation* **86**(5): 1524-1531.

O'Neil J, Tchinda J, Gutierrez A *et al.* (2007) *Alu* elements mediate MYB gene tandem duplication in human T-ALL. *The Journal of Experimental Medicine* **204**(13): 3059-3066.

Oldridge M, Zackai EH, McDonald-McGinn DM *et al.* (1999) De novo *Alu*-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. *American Journal of Human Genetics* **64**(2): 446-461.

Oliveira C, Senz J, Kaurah P *et al.* (2009) Germline CDH1 deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics* **18**(9): 1545-1555.

Onno M, Nakamura T, Hillova J, Hill M (1992) Rearrangement of the human tre oncogene by homologous recombination between *Alu* repeats of nucleotide sequences from two different chromosomes. *Oncogene* **7**(12): 2519-2523.

Phylipsen M, Vogelaar IP, Schaap RA *et al.* (2010) A new alpha(0)-thalassemia deletion found in a Dutch family (--(aw)). *Blood Cells Molecular Disease* **45**(2): 133-135.

Rodic N, Zampella JG, Cornish TC, Wheelan SJ, Burns KH (2013) Translocation junctions in TCF3-PBX1 acute lymphoblastic leukemia/lymphoma cluster near transposable elements. *Mobile DNA* **4**(1): 22.

Ross JP, Rand KN, Molloy PL (2010) Hypomethylation of repeated DNA sequences in cancer. *Epigenomics* **2**(2): 245-269.

Saito K, Siomi MC (2010) Small RNA-mediated quiescence of transposable elements in animals. *Developmental Cell* **19**(5): 687-697.

Shabbeer J, Yasuda M, Benson SD, Desnick RJ (2006) Fabry disease: Identification of 50 novel alpha-galactosidase a mutations causing the classic phenotype and three-dimensional structural analysis of 29 missense mutations. *Human Genomics* **2**(5): 297-309.

Tang Z, Steranka JP, Ma S *et al.* (2017) Human transposon insertion profiling: Analysis, visualization and identification of somatic line-1 insertions in ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America* **114**(5): E733-E740.

Tubio JM, Li Y, Ju YS *et al.* (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**(6196): 1251343.

Vervoort R, Gitzelmann R, Lissens W, Liebaers I (1998) A mutation (IVS8+0.6kbdeltc) creating a new donor splice site activates a cryptic exon in an alu-element in intron 8 of the human beta-glucuronidase gene. *Human Genetics* **103**(6): 686-693.

Wallace MR, Andersen LB, Saulino AM *et al.* (1991) A de novo *Alu* insertion results in neurofibromatosis type 1. *Nature* **353**(6347): 864-866.

Yu J, Xie J, Luo L, Li Z (2014) An *Alu* element-mediated 28.5 kb alpha-thalassemia deletion found in a chinese family. *Hemoglobin* **38**(6): 427-430.

Further readings

Chénais B, Caruso A, Hiard S, Casse N (2012) The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* **509**(1): 7-15.

Chénais B (2013) Transposable elements and human cancer: A causal relationship? *Biochimica et Biophysica Acta* **1835**(1): 28-35.

Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* **10**(10): 691-703.

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41**: 331-368.

Kulis M, Queiros AC, Beekman R, Martin-Subero JI (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochimica et Biophysica Acta* **1829**(11): 1161-1174.

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics* **12**(9): 615-627.

Mazoyer S (2005) Genomic rearrangements in the BRCA1 and BRCA2 genes. *Human Mutation* **25**(5): 415-422.

Reilly MT, Faulkner GJ, Dubnau J, Ponomarev I, Gage FH (2013) The role of transposable elements in health and diseases of the central nervous system. *Journal of Neuroscience* **33**(45): 17577-17586.

Roy-Engel AM, El-Sawy M, Farooq L *et al.* (2005) Human retroelements may introduce intragenic polyadenylation signals. *Cytogenetics and Genome Research* **110**(1-4): 365-371.

Sinzelle L, Izsvak Z, Ivics Z (2009) Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cellular and Molecular Life Sciences* **66**(6): 1073-1093.

Glossary

Ectopic recombination: aberrant form of DNA recombination between non-homologous loci, rather than along homologous loci. Such recombination is often driven by transposable element and results in dramatic chromosomal rearrangement, which is generally deleterious.

Exonisation: recruiting of a new exon from non-protein-coding, such as intronic DNA sequences or transposable elements.

Exon skipping: splicing modification following the insertion of a transposable element that result in the skipping of one or more exon during mRNA maturation.

Intronisation: creation of new intron following the insertion of a transposable element sequence.

Retrotransposon or retroelement: transposable element that needs to be retro-transcribed from RNA before reinsertion in the genomic DNA. The structure of retrotransposons may be more or less complex and could include, or not, long terminal repeat (LTR) sequences.

Transposon: in eukaryotes, sequence of repeated DNA that encodes the transposase enzyme sufficient to assume the mobility of the transposable element.

Table 1. Examples of human diseases linked to TE insertion, TE-mediated deletions or TE-mediated chromosomal rearrangements.

Gene	Type of TE	Type of event	Disease or disorder
APC	*LINE-1*	insertion	Colon cancer, desmoid tumours
APOB	*Alu*	deletion	Hypo beta lipoproteinemia
BCR-ABL	*Alu*	translocation t(9;22)(q34;q11)	Chronic myeloid leukemia
BRCA1	*Alu*	insertion, deletion and duplication	Breast/ovarian cancer
BRCA2	*Alu*	insertion, duplication	Breast/ovarian cancer
CAD	*Alu*	deletion	Hepatoma
CDH1	*Alu*	deletion	Hereditary diffuse gastric cancer
EWSR1-ETV	*Alu*	translocation t(5q23q31)(18q12)	Ewing sarcoma
α-GALA	*Alu*	deletion	Fabry disease
HBB	*Alu*	deletion	Thalassemia
HEXB	*Alu*	deletion	Sandhoff disease
LDLR	*Alu*	deletion	Familial hyper-cholesterolemia
NF1	*Alu*	insertion	Neurofibroma
MLL1	*Alu*	duplication	Acute myeloid leukemia
MYB	*Alu*	duplication	T-cell lymphoblastic leukemia
MYC	*LINE-1*	insertion	Breast carcinoma
TCF3	*MER20*	translocation	hematolymphoid neoplasm
VHL	*Alu*	deletion	Von Hippel Lindau disease

Table 2. Examples of TE-mediated alterations of mRNA splicing at the origin of human diseases, including exonisation, exon skipping, intron retention and aberrant splicing.

Gene	Type of TE	Type of event[a]	Disease or disorder
BRCA2	*Alu*	aberrant splicing	Breast cancer
COL4A3	*Alu*	exonisation	Alport syndrome
CYBB	*LINE-1*	exonisation, intron retention	Chronic granulomatous disease
DMD	*LINE-1*	exon skipping	Duchenne Muscular Dystrophy
DMD	*Alu*	aberrant splicing	X-linked dilated cardiomyopathy
F8	*Alu*	exon skipping	Hemophilia A
FAS	*Alu*	exon skipping	Autoimmune lymphoproliferative syndrome
FGFR2	*Alu*	exon skipping, aberrant splicing	Apert syndrome
FKTN	*LINE-1*	exon skipping	Fukuyama-type congenital muscular dystrophy
GUSB	*Alu*	exonisation	Sly syndrome
IDS	*Alu*	exon skipping	Hunter syndrome
MID1	*HERV-E*	new regulatory region	Opitz syndrome
NF1	*Alu*	aberant splicing, reading frame shift	Neurofibromatosis type 1
OAT	*Alu*	exonisation	ornithine delta-aminotransferase deficiency
OPA1	*Alu*	exon skipping	Autosomal dominant optic atrophy
PDHX	*LINE-1*	aberrant splicing	Pyruvate dehydrogenase complex deficiency
RP2	*LINE-1*	aberrant splicing, reading frame shift	X-linked retinitis pigmentosa
RSK2	*LINE-1*	exon skipping	Coffin-Lowry syndrome

Table 3. Examples of TE-mediated creation of new polyadenyation site involved in the onset of human diseases.

Gene	Type of TE	Disease or disorder
APC	*LINE-1*	Colon cancer
CASR	*Alu*	Hypocalciuric hypercalcemia and neonatal severe hyper-parathyroidism
DMD	*LINE-1*	X-linked dilated cardiomyopathy
F9	*Alu*	Hemophilia B
F9	*LINE-1*	Hemophilia B
FAS	*Alu*	Autoimmune lympho-proliferative syndrome
HHLA2, 3	*HERV-H*	unkwown

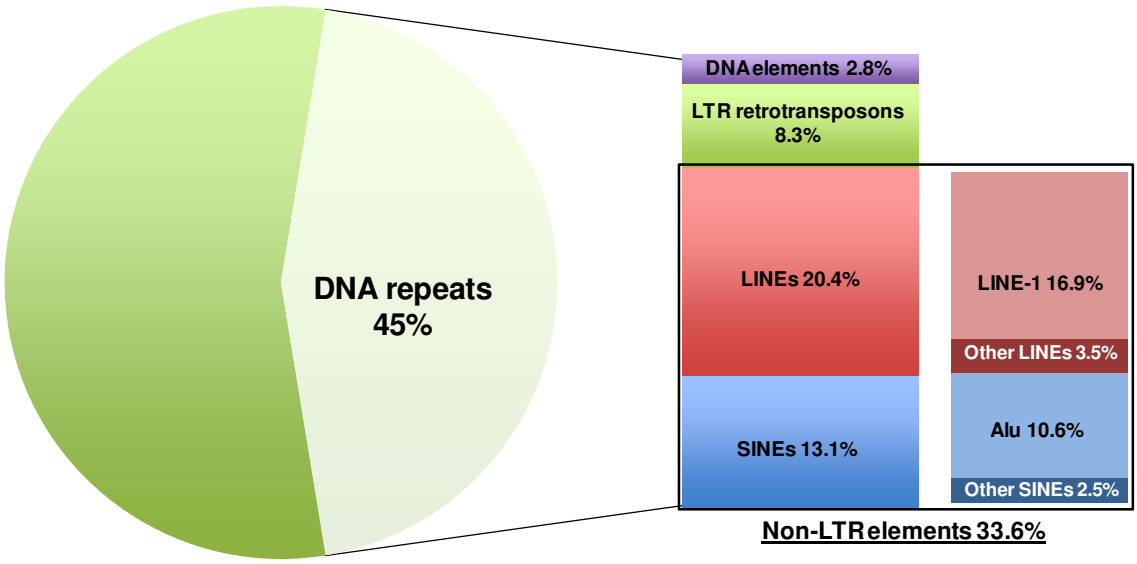


Figure 1. Proportion of TEs in the human genome (Chénais, Current Cancer Drug Targets 2015). The percentage of each class or sub-class of TE is indicated with respect to the whole genome according to data from Lander et al. (2001).

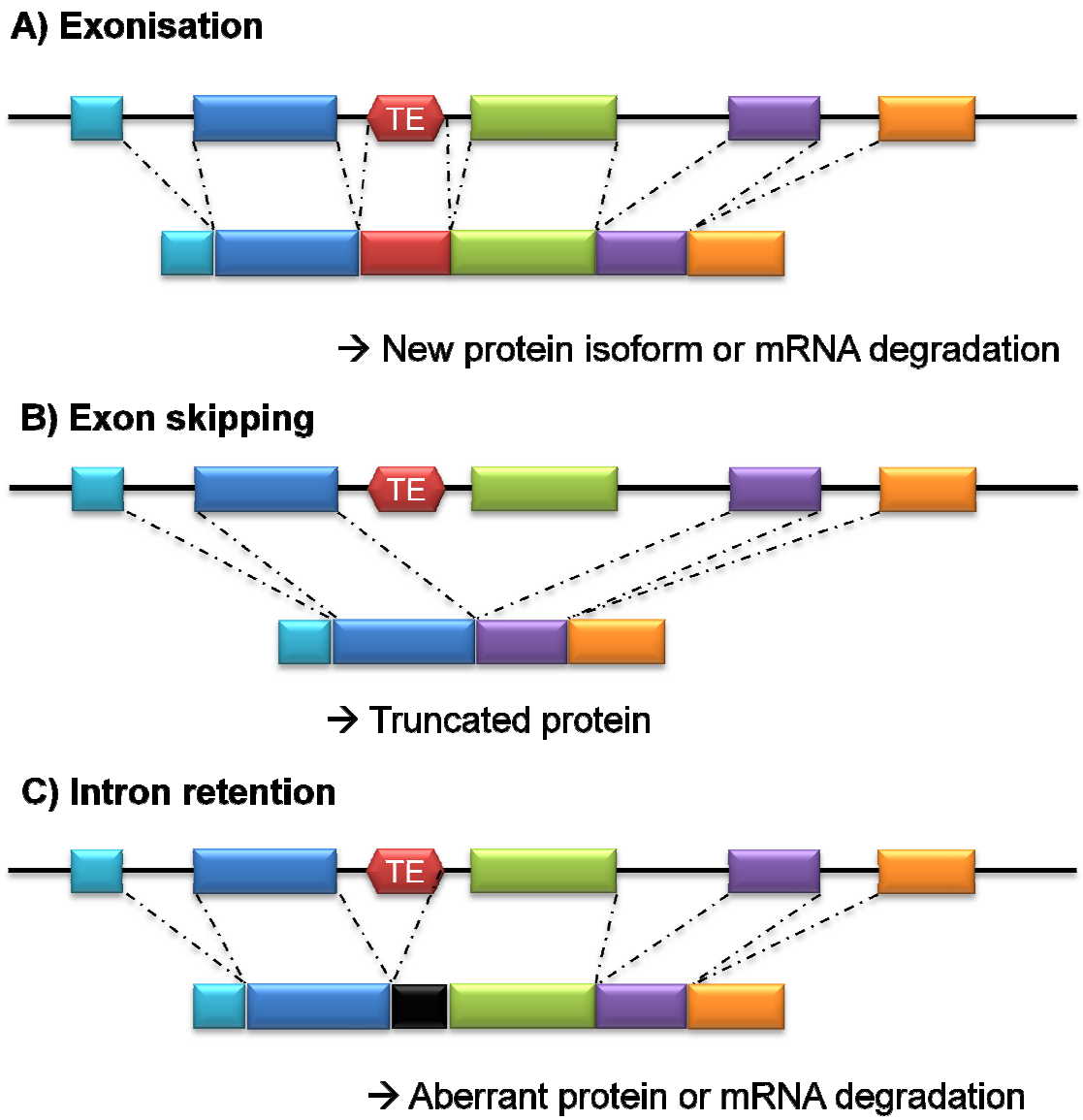


Figure 2. Simplified view of the main consequences of TE insertion on mRNA splicing. A) exonisation; B) exon skipping; C) intron retention. Red box figure the TE and colored boxes correspond to exons; introns are represented as black lines; dashed lines indicate the splicing events. (Chénais, Current Cancer Drug Targets 2015).