



**HAL**  
open science

# 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.

# 1 Mobile Elements in the Human Genome: Implications for Disease

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

3

## 4 **Abstract**

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

17

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

21

## 22 **Key concepts**

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

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

31

## 32 **Introduction: Mobile elements in the human genome**

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

48 [10.1002/9780470015902.a0001899.pub3](#); [DOI: 10.1002/9780470015902.a0006126.pub3](#)

49 By far the most successful type of TEs in the human genome is class-I elements, which is  
50 subdivided into Long Terminal Repeats (LTR) and non-LTR elements. The latter are very  
51 predominant and 76% of human TEs are non-LTR elements, which represent at least 33.6%  
52 of the whole human genome (Lander *et al.*, 2001) (**Fig. 1**). On its own, the long interspersed  
53 element (LINE)-1 accounts for 38% of the human TEs and 16.9% of the entire genome (**Fig.**  
54 **1**), followed by the non-autonomous SINEs, particularly the *Alu* repeats, which account for  
55 24% of TEs and 10.6% of the genome (de Koning *et al.*, 2011, Lander *et al.*, 2001). Other  
56 SINEs found in humans are MIR and MIR3 elements. The LTR-elements are less present in  
57 the human genome and mainly represented by endogenous retroviruses (HERVs, including  
58 *HERV-I*, *HERV-K*, *HERV-L*) that resemble retroviruses in both their structure and mobility

59 mechanism, but lacking a functional envelope gene. See also: [DOI:](https://doi.org/10.1002/9780470015902.a0026264)  
60 [10.1002/9780470015902.a0026264](https://doi.org/10.1002/9780470015902.a0026264); [DOI:](https://doi.org/10.1002/9780470015902.a0026711) [10.1002/9780470015902.a0026711](https://doi.org/10.1002/9780470015902.a0026711)

61 Class-II elements, or DNA transposons, are present in small amounts in the human genome,  
62 accounting for only 6% of TEs and 2.8% of the entire genome (**Fig. 1**). They mainly belong to  
63 the TC-1/*mariner* superfamily (*i.e. mariner*, MER2-Tigger, Tc2), *hAT* superfamily (*i.e. MER-1-*  
64 *Charlie*, *Zaphod*), and some *PiggyBac*-like elements (Lander *et al.*, 2001). See also: [DOI:](https://doi.org/10.1002/9780470015902.a0026264)  
65 [10.1002/9780470015902.a0026264](https://doi.org/10.1002/9780470015902.a0026264); [DOI:](https://doi.org/10.1002/9780470015902.a0020996.pub2) [10.1002/9780470015902.a0020996.pub2](https://doi.org/10.1002/9780470015902.a0020996.pub2)

66 Long been considered as junk DNA or parasite DNA, TEs are now accepted as an evolutionary  
67 force that contributes to the genetic diversity of organisms. TE could be recruited on behalf  
68 of the genome to become a new gene and such TE exaptation or domestication is a source of  
69 genetic innovation, but the transposition activity may also contribute to human disorders  
70 through chromosome rearrangements, insertional mutagenesis, deregulation of gene  
71 expression, and epigenetic changes (Chénais, 2015). See also: [DOI:](https://doi.org/10.1002/9780470015902.a0026265)  
72 [10.1002/9780470015902.a0026265](https://doi.org/10.1002/9780470015902.a0026265); [DOI:](https://doi.org/10.1002/9780470015902.a0020996.pub2) [10.1002/9780470015902.a0020996.pub2](https://doi.org/10.1002/9780470015902.a0020996.pub2)

73

#### 74 **Mechanisms of TE's impact on human genome**

75 Several mechanisms are involved to explain the impact of TE on the genome and gene  
76 expression (Chénais, 2015). Genome rearrangements are favoured by the repetition of TE in  
77 the genome and the abundance of *LINE-1* and *Alu* repeats (*i.e.* about 17% and 11% of the  
78 human genome, respectively) promotes ectopic recombination, leading to gene deletions,  
79 duplications and chromosome translocations or inversions (**Table 1**). In addition, *LINE-1* and  
80 *Alu* elements are frequently observed within or near the breakage points of chromosomal  
81 rearrangement and *LINE-1* alone account for 19% of the 2081 breakpoints junction analysed  
82 in a dataset of 17 whole human genomes (Kidd *et al.*, 2010). Although less present in the  
83 human genome, DNA transposons may also be a source of chromosome rearrangement as  
84 shown by a study conducted on ten types of hematolymphoid neoplasms (Rodic *et al.*,  
85 2013). This work showed that most translocations have no propensity to involve genomic  
86 repeats, but 47% of the translocations at the *TCF3* locus on chromosome 19 involve a *MER20*  
87 DNA transposon (Rodic *et al.*, 2013). See also: [DOI:](https://doi.org/10.1002/9780470015902.a0005493.pub2)  
88 [10.1002/9780470015902.a0005493.pub2](https://doi.org/10.1002/9780470015902.a0005493.pub2); [DOI:](https://doi.org/10.1002/9780470015902.a0020996.pub2) [10.1002/9780470015902.a0020996.pub2](https://doi.org/10.1002/9780470015902.a0020996.pub2)

89 The second way in which TEs may impact the human genome and induce diseases is through  
90 the modification of gene structure and regulatory regions (**Table 2**). This include the creation  
91 of new polyadenylation sites (**Table 3**), new exons (also called “exonisation”) and new  
92 introns (or “intronisation”) as well as alternative splicing modifications, e.g. exon skipping  
93 and intron retention (**Table 2, Fig. 2**) (Chénais, 2015). The creation of a new exon causes a  
94 frameshift or creates premature stop codons in about 79% of cases, resulting in non-  
95 functional sequences, which are often alternatively spliced. However, exonisation events  
96 following *LINE-1* or *Alu* insertion may lead to human diseases (**Table 2**) (reviewed in Chénais,  
97 2015). Using the LTR of HERV as promoter/enhancer is also a way to modify gene expression.  
98 This was observed, for example, in the case of the *MID1* gene, associated with the X-linked  
99 form of Opitz syndrome, which transcript is fused to *HERV-E* LTR acting as a tissue-specific  
100 promoter/enhancer (Landry *et al.*, 2002). See also: [DOI: 10.1002/9780470015902.a0005493.pub2](https://doi.org/10.1002/9780470015902.a0005493.pub2);  
101 [DOI: 10.1002/9780470015902.a0020996.pub2](https://doi.org/10.1002/9780470015902.a0020996.pub2);  
102 [DOI: 10.1002/9780470015902.a0022886](https://doi.org/10.1002/9780470015902.a0022886); [DOI: 10.1002/9780470015902.a0026711](https://doi.org/10.1002/9780470015902.a0026711)

103 The third way to modify the expression of the genome concerns the epigenetic mechanisms  
104 induced or modified by TEs. A correlation between the activity of TEs and chromatin  
105 regulation, mainly DNA methylation, has been observed and hypomethylation of DNA is a  
106 cause of TEs activation (Ross *et al.*, 2010). The relationship between pathological markers  
107 and the methylation status of whole genome CpG islands and *LINE-1* elements has been  
108 particularly studied in the case of cancers (Chénais 2015). In addition, the repression of TEs  
109 through RNA interference involved essentially endogenous small interfering RNA (endo-  
110 siRNA) in somatic cells and PIWI-interacting RNA (piRNA) in the germline (Saito & Siomi,  
111 2010). Both endo-siRNA and piRNA are transcribed from TEs themselves and *a priori* do not  
112 disturb the expression of protein-coding genes (Saito & Siomi, 2010). Cancer cells are known  
113 to contain significantly lower quantities of micro-RNAs, but currently, in the absence of  
114 endo-siRNA and piRNA expression pattern in pathologic cells, it is impossible to answer the  
115 question of the putative impact of TE-related regulatory RNA on cancer or other diseases.  
116 See also: [DOI: 10.1002/9780470015902.a0023593](https://doi.org/10.1002/9780470015902.a0023593); [DOI: 10.1002/9780470015902.a0025688](https://doi.org/10.1002/9780470015902.a0025688);  
117 [DOI: 10.1002/9780470015902.a0026316](https://doi.org/10.1002/9780470015902.a0026316)

118

119 **Involvement of TEs in human cancer**

120       • ***Alu/LINE-1 insertion***

121 Several examples of cancers have been reported in the literature as related to LINE or  
122 SINE/*Alu* insertion (**Table 1**) (reviewed in Chénais, 2015). More recently an increased  
123 insertional activity of *LINE-1* and an elevated level of *LINE-1* transcript were found in primary  
124 tumour samples of breast invasive carcinoma, head and neck squamous cell carcinoma and  
125 lung adenocarcinoma (Clayton *et al.*, 2016). The insertion of the *LINE-1* retroelement into  
126 the human proto-oncogene *c-MYC* was reported as early as in 1988 in human breast  
127 carcinoma cells (Morse *et al.*, 1988). This study compared the structure of the *c-MYC* gene  
128 from breast ductal adenocarcinoma and from normal breast tissue of the same patient and  
129 revealed a tumour-specific rearrangement of one *MYC* locus and the amplification of the  
130 other *MYC* locus. A non-*MYC* sequence with near complete homology to a *LINE-1* sequence  
131 was found within the second intron of the rearranged locus. In this case, the *LINE-1*  
132 sequence functioned as a mobile genetic element to produce a somatic mutation (Morse *et*  
133 *al.*, 1988). Another example of a *LINE-1* element somatic insertion was observed in the  
134 tumor-suppressing gene *APC* (adenomatous polyposis coli) from several patients with a  
135 colon cancer and also in association with desmoid tumors (reviewed in Chénais, 2015). The  
136 *BRCA1* and *BRCA2* genes, which are associated with breast and ovarian cancer, are also the  
137 sites of *Alu* insertions. For example, one of the first evidenced germline mutations was the  
138 insertion of a transcriptionally active *Alu* element into exon 22 of *BRCA2*, which resulted in  
139 alternative splicing that skipped exon 22 (reviewed in Chénais, 2015). Finally, *Alu* insertions  
140 in the von Hippel-Lindau (*VHL*) gene cause large alteration at the origin of the von Hippel-  
141 Lindau disease, a familial autosomal-dominant syndrome, predisposing to the development  
142 of benign and malignant tumors (Franke *et al.*, 2009). See also: [DOI:  
143 10.1002/9780470015902.a0005493.pub2](https://doi.org/10.1002/9780470015902.a0005493.pub2)

144

145       • ***Alu recombination leading to deletion, duplication or translocation***

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

152 cadherin-1 (*CDH1*) gene occur in 4% of hereditary diffuse gastric cancer families (Oliveira *et*  
153 *al.*, 2009). These deletions involve mainly non-allelic homologous recombination in *Alu*  
154 repeat sequences, as supported by the statistically significant over-representation of *Alu*  
155 elements around breakpoints (Oliveira *et al.*, 2009). Homologous recombination between  
156 *Alu* sequences and subsequent deletion have been detected at high frequency within the  
157 caspase-activated DNase (*CAD*) gene in 13 out of the 20 human hepatoma tissues and in  
158 eight of the para-cancerous cirrhotic livers (Hsieh *et al.*, 2005) (**Table 1**). The recombination  
159 caused exon-3 deletion, which in turn led to exon-3 skipping or replacement with a partial  
160 *Alu* sequence, and consequential C-terminal truncation of the *CAD* protein (Hsieh *et al.*,  
161 2005). These data highlight the role of *Alu* repeats in mediating genome instability in human  
162 cancers. See also: DOI: 10.1002/9780470015902.a0005493.pub2

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

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

177

#### 178 • ***TE-related epigenetic regulation and cancer***

179 The relation between *LINE-1* activation and the methylation status of DNA has been  
180 particularly investigated in the case of cancers. Indeed the frequent hypomethylation of  
181 chromatin in tumour cells is regarded as a facilitator of TE mobility, and especially  
182 concerning *LINE-1* (Ross *et al.*, 2010). Whole genome studies revealed the preferential *LINE-*  
183 *1* insertion in hypomethylated DNA regions of cancer cells of different origins (Ferreira *et al.*,

184 2014, Tubio *et al.*, 2014). Besides, several studies highlighted the hypomethylation of the  
185 *LINE-1* element itself in several types of cancers (Chénais 2015) and the transcriptome  
186 analysis of chronic lymphocytic leukaemia showed that TEs are globally derepressed  
187 (Ferreira *et al.*, 2014). However Lee *et al.* have shown that *LINE-1* and *Alu* insertions  
188 occurred preferentially in epithelial cancer but not in blood and brain cancer (Lee *et al.*,  
189 2012). In this study the authors identified almost two hundred somatic insertions of *LINE-1*  
190 (94.5%), *Alu* (5.0%) and HERV (0.5%) in 43 high-coverage whole genome sequencing datasets  
191 from five cancer types (i.e. colorectal, glioblastoma, multiple myeloma, ovarian and prostate  
192 cancer) (Lee *et al.*, 2012). Finally, in the case of melanoma, the *LINE-1* hypomethylation is  
193 accompanied by the shortened relapse-free survival of patients, and a global under-  
194 methylation of DNA impacts the metastatic formation of primary melanomas (Ecsedi *et al.*,  
195 2013).

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

201

## 202 **TEs in hematologic, metabolic and/or genetic diseases**

### 203 • ***Alu-mediated deletion causing human diseases***

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

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

217 The reported TE-mediated rearrangements in the  $\beta$ -globin gene cluster include  $\gamma$ -,  $\delta$ - and  $\beta$ -  
218 globin genes (Harteveld *et al.*, 1997, Yu *et al.*, 2014). For example, in an individual with  
219 hereditary persistence of foetal haemoglobin, the  $\beta$ -globin cluster displays a large deletion of  
220 48.5 kb downstream of the  $\beta$ -globin gene and the 5' deletion breakpoint occurs in an *Alu*  
221 element 3' to the  $A\gamma$ -globin gene. However, in this particular case, the recombination  
222 mechanism involve non-homologous recombination since the 3' breakpoint was located  
223 within a region that contains various repeats including a portion of a *LINE-1* repeat, a perfect  
224 160-bp palindrome, and a set of 41-bp direct repeats also present elsewhere in the human  
225 genome (Henthorn *et al.*, 1986).

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

246

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

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

265

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

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

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

279 intron 11 was observed in individuals with X-linked dilated cardiomyopathy (Ferlini *et al.*,  
280 1998). The insertion induces a rearrangement that activates a cryptic splice site in intron 11  
281 and produces an alternative transcript containing the *Alu*-like sequence and part of the  
282 adjacent intron 11, spliced between exons 11 and 12, but not translated due to the presence  
283 of numerous stop codons (Ferlini *et al.*, 1998). The *LINE-1* element is responsible for a large  
284 deletion in the X-linked pyruvate dehydrogenase E1 (*PDHX*) gene, but the *LINE-1* insertion  
285 may also lead, at last in one patient, to an aberrantly spliced isoform resulting from the use  
286 of two cryptic splice sites (Mine *et al.*, 2007).

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

292

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

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

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

306

307 • ***LINE-1 and chromatin regulation***

308 As described above for cancer, the chromatin structure and/or epigenetic regulation  
309 associated with TE insertion may lead to non-cancer human diseases but examples are rare  
310 to date. The  $\beta$ -globin<sub>L1</sub> allele results from the insertion of a full length *LINE-1* element in the

311 intron 2 of the  $\beta$ -globin gene. The decreased rate of transcription of this  $\beta$ -globin<sub>L1</sub> allele is  
312 due to the hypermethylated profile of its promoter and enhancer sequences, which results  
313 in a transcriptional down-regulation and thus the  $\beta^+$ -thalassemia phenotype (Lanikova *et al.*,  
314 2013).

315

### 316 **TE involvement in neurological diseases**

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

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

333 The Sandhoff disease, a recessive lysosomal storage disease, is another example of TE-  
334 related neurodegenerative disease (**Table 1**). The deficiency of  $\beta$ -hexosaminidase (*HEXB*)  
335 activity is due to deletion alleles of the *HEXB* gene in 27% of the examined dataset (Neote *et*  
336 *al.*, 1990). *HEXB* contains two *Alu* repeats that are involved in the deletion event, which  
337 removes about 16kb including the *HEXB* promoter, exons 1 to 5, and part of intron 5. The  
338 deletion also resulted in a reconstituted *Alu* sequence, the left half of which comes from the  
339 5' *Alu* sequence and the right half of which from the 3' *Alu* sequence (Neote *et al.*, 1990).

340 Neurofibromatosis type I is a common genetic disease linked to the *NF-1* tumor suppressor  
341 gene and characterized by changes in skin coloring and the growth of benign tumors along  
342 nerves (i.e. neurofibromas). The insertion of an *Alu* sequence in the intron of the *NF-1* gene  
343 leads to a deletion and reading frame shift in the downstream exon during splicing, which  
344 could be a cause of the disease (**Table 2**) (Wallace *et al.*, 1991). See also: DOI:  
345 10.1002/9780470015902.a0005493.pub2

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

352 Finally, a link between TEs and psychiatric disorders was also highlighted (Guffanti *et al.*,  
353 2014). Particularly, whole-genome sequencing revealed brain specific *LINE-1* insertion in  
354 patients localized preferentially to synapse- and schizophrenia-related genes, suggesting a  
355 causal role of *LINE-1* in the onset of the disease (Bundo *et al.*, 2014).

356

## 357 **Conclusion**

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

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

371

372 **References**

- 373 Baust C, Seifarth W, Germaier H, Hehlmann R, Leib-Mosch C (2000) HERV-K-T47D-related  
374 long terminal repeats mediate polyadenylation of cellular transcripts. *Genomics* **66**(1):  
375 98-103.
- 376 Bundo M, Toyoshima M, Okada Y *et al.* (2014) Increased I1 retrotransposition in the  
377 neuronal genome in schizophrenia. *Neuron* **81**(2): 306-313.
- 378 Chénais B (2015) Transposable elements in cancer and other human diseases. *Current*  
379 *Cancer Drug Targets* **15**(3): 227-242.
- 380 Clayton EA, Wang L, Rishishwar L *et al.* (2016) Patterns of transposable element expression  
381 and insertion in cancer. *Frontiers in Molecular Biosciences* **3**: 76.
- 382 de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may  
383 comprise over two-thirds of the human genome. *PLoS Genetics* **7**(12): e1002384.
- 384 Ecsedi SI, Hernandez-Vargas H, Lima SC *et al.* (2013) Transposable hypomethylation is  
385 associated with metastatic capacity of primary melanomas. *International Journal of*  
386 *Clinical and Experimental Pathology* **6**(12): 2943-2948.
- 387 Ferlini A, Galie N, Merlini L *et al.* (1998) A novel *Alu*-like element rearranged in the  
388 dystrophin gene causes a splicing mutation in a family with X-linked dilated  
389 cardiomyopathy. *American Journal of Human Genetics* **63**(2): 436-446.
- 390 Ferreira PG, Jares P, Rico D *et al.* (2014) Transcriptome characterization by RNA sequencing  
391 identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia.  
392 *Genome Research* **24**(2): 212-226.
- 393 Franke G, Bausch B, Hoffmann MM *et al.* (2009) *Alu-Alu* recombination underlies the vast  
394 majority of large *VHL* germline deletions: Molecular characterization and genotype-  
395 phenotype correlations in *VHL* patients. *Human Mutation* **30**(5): 776-786.
- 396 Ganguly A, Dunbar T, Chen P, Godmilow L, Ganguly T (2003) Exon skipping caused by an  
397 intronic insertion of a young *Alu* *YB9* element leads to severe hemophilia A. *Human*  
398 *Genetics* **113**(4): 348-352.
- 399 Grégoire L, Haudry A, Lerat E (2016) The transposable element environment of human genes  
400 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.

464 Narita N, Nishio H, Kitoh Y *et al.* (1993) Insertion of a 5' truncated L1 element into the 3' end  
465 of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a  
466 case of Duchenne muscular dystrophy. *The Journal of Clinical Investigation* **91**(5): 1862-  
467 1867.

468 Neote K, McInnes B, Mahuran DJ, Gravel RA (1990) Structure and distribution of an *Alu*-type  
469 deletion mutation in Sandhoff disease. *The Journal of Clinical Investigation* **86**(5): 1524-  
470 1531.

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

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

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

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

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

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

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

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

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

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

496 Tubio JM, Li Y, Ju YS *et al.* (2014) Mobile DNA in cancer. Extensive transduction of  
497 nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science*  
498 **345**(6196): 1251343.

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

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

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

506

507 **Further readings**

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

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

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

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

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

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

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

524 Reilly MT, Faulkner GJ, Dubnau J, Ponomarev I, Gage FH (2013) The role of transposable  
525 elements in health and diseases of the central nervous system. *Journal of Neuroscience*  
526 **33**(45): 17577-17586.

527 Roy-Engel AM, El-Sawy M, Farooq L *et al.* (2005) Human retroelements may introduce  
528 intragenic polyadenylation signals. *Cytogenetics and Genome Research* **110**(1-4): 365-  
529 371.

530 Sinzelle L, Izsvak Z, Ivics Z (2009) Molecular domestication of transposable elements: From  
531 detrimental parasites to useful host genes. *Cellular and Molecular Life Sciences* **66**(6):  
532 1073-1093.

533

## 534 **Glossary**

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

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

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

543 Intronsation: creation of new intron following the insertion of a transposable element  
544 sequence.

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

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

550

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

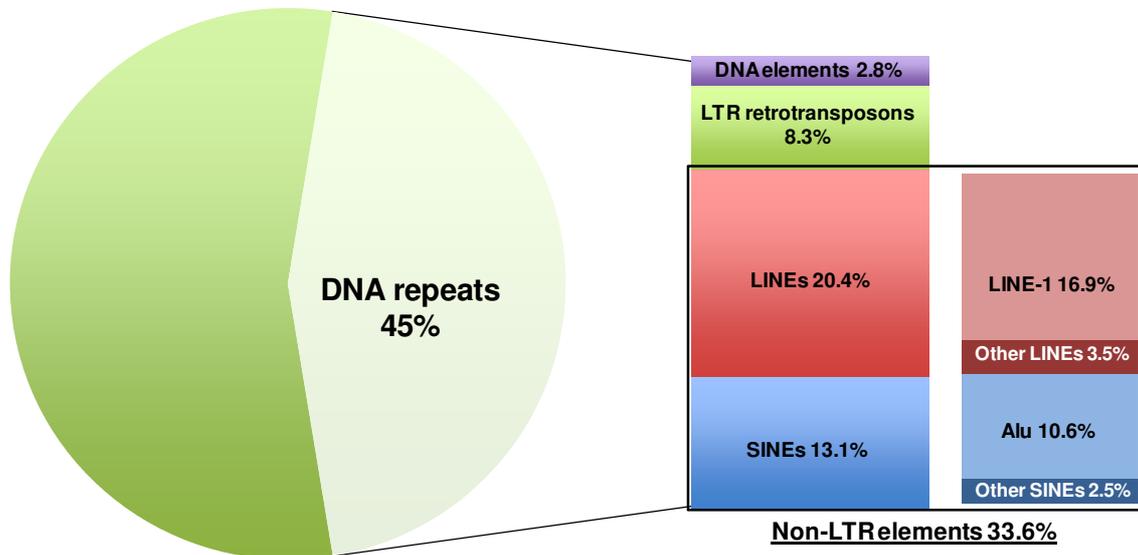
<b>Gene</b>	<b>Type of TE</b>	<b>Type of event</b>	<b>Disease or disorder</b>
<i>APC</i>	<i>LINE-1</i>	insertion	Colon cancer, desmoid tumours
<i>APOB</i>	<i>Alu</i>	deletion	Hypo beta lipoproteinemia
<i>BCR-ABL</i>	<i>Alu</i>	translocation t(9;22)(q34;q11)	Chronic myeloid leukemia
<i>BRCA1</i>	<i>Alu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

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

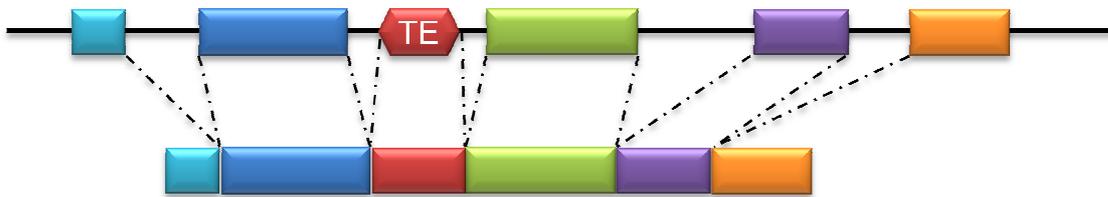
**Table 3.** Examples of TE-mediated creation of new polyadenylation site involved in the onset of human diseases.

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



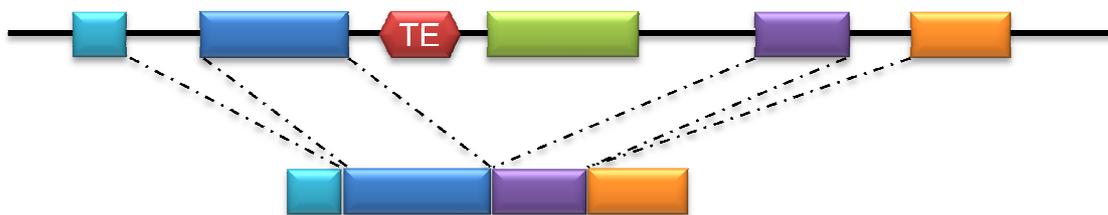
**Figure 1.** Proportion of TEs in the human genome (Chénaïs, Current Cancer Drug Targets 2015). The percentage of each class or sub-class of TE is indicated with respect to the whole genome according to data from Lander et al. (2001).

### A) Exonisation



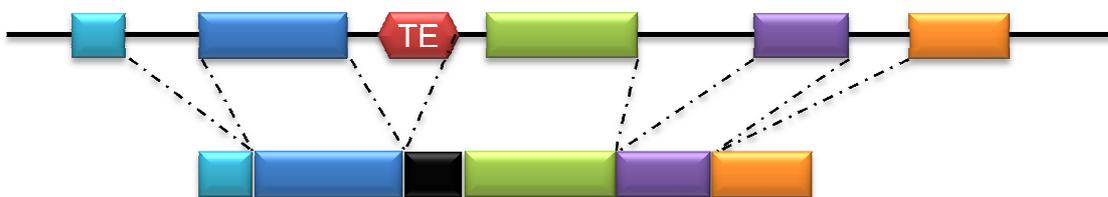
→ New protein isoform or mRNA degradation

### B) Exon skipping



→ Truncated protein

### C) Intron retention



→ Aberrant protein or mRNA degradation

**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. (Chénaïs, Current Cancer Drug Targets 2015).