Review

Transposable elements and human cancer: A causal relationship?

Benoît Chénais *

Université du Maine, EA2160 Mer Molécules Santé, UFR Sciences et Techniques, Avenue Olivier Messiaen, 72085 Le Mans, France

Abstract

Transposable elements are present in almost all genomes including that of humans. These mobile DNA sequences are capable of invading genomes and their impact on genome evolution is substantial as they contribute to the genetic diversity of organisms. The mobility of transposable elements can cause deleterious mutations, gene disruption and chromosome rearrangements that may lead to several pathologies including cancer. This mini-review aims to give a brief overview of the relationship that transposons and retrotransposons may have in the genetic cause of human cancer onset, or conversely creating protection against cancer. Finally, the cause of TE mobility may also be the cancer cell environment itself.

Contents

1. Introduction ............................................................... 28
2. Non-LTR retroelement mobility: a genetic cause of cancer ......................................... 29
3. Alu-mediated chromosomal recombination at the origin of cancer................................. 30
4. Transposable element domestication: an evolutionary gain sometimes related to cancer or cancer suppression................. 31
5. But what may cause or promote transposable element mobility? ...................................... 31
6. Concluding remarks ........................................................... 32
Acknowledgements .............................................................. 32
References .................................................................. 32

1. Introduction

Transposable elements (TEs) are repeated and mobile DNA sequences, with the ability to move in and invade genomes. They are present in most genomes from bacteria to eukaryotes and generally represent a substantial fraction of the genome depending on the species [1–5]. Two classes of TE are roughly distinguished (Fig. 1). Class-I elements, or retrotransposons, use reverse transcriptase to copy an RNA genome into the host DNA; they are divided into Long Terminal Repeat (LTR) and non-LTR elements [6,7]. Classed as LTRs, the human endogenous retroviruses (HERVs) resemble retroviruses in both their structure and mobility mechanism. Most HERVs contain a nonfunctional envelope gene, which relegates them to an intracellular existence [6]. Class-I elements present in the human genome are mainly non-LTR long interspersed elements (LINEs), i.e., LINE-1 (L1), LINE-2 and LINE-3, and LTR-element HERVs (HERV-I, HERV-K, HERV-L) [8]. With Class-II elements, or DNA transposons, the DNA genome of the element itself serves as the template for transposition either by a “cut and paste” mechanism involving the excision and reinsertion of the DNA sequence of the element, or by using a rolling circle process or a virus-like process [7,9]. DNA transposons present in the human genome belong to TC-1/mariner super-family (i.e. mariner, MER2-Tigger, Tc2), hAT super-family (i.e. MER-1-charlie, Zaphod), and some PiggyBac like elements [8]. The two classes have been subdivided into super-families and then into families on the basis of the transposition mechanism, sequence similarities and structural relationships [10].
A growing number of TEs are non-autonomous and appear dependent on another helper element. They are unable to remove and insert themselves elsewhere in the genome and need the enzyme(s) encoded by autonomous elements to move. Nevertheless, non-autonomous elements such as miniature inverted-repeat transposable elements (MITEs) and short interspersed nuclear elements (SINEs) have greatly invaded genomes [4], notably MITEs in rice [11] and SINEs in mammals [3]. The human genome contains three main SINE families, i.e. Alu repeats, MIR and MIR3 [8].

Transposable element activity and its impact on eukaryotic genome evolution are substantial and TEs are currently accepted as an evolutionary force [2,4,5]. Transposition activity may contribute to the genetic diversity of organisms acting as gene regulatory elements by providing their own promoters or by altering the state of chromatin [7,12–14]. Another way for a TE to be a source of genetic innovation is its recruitment on behalf of the genome to become a new gene, a phenomenon known as TE domestication [7,13,14]. But the TE mobility can cause mutations, some of which can result in several diseases, i.e., about 11% [9,10]. The Alu element is specific to primates and has colonized primate genomes for 65 million years [21]. Another group of non-autonomous retroelements active in humans is the SVA (for SINE-VNTR-Alu; with VNTR for variable number of tandem repeats) group of TEs (Fig. 1B) [22]. Like SINEs, SVAs use the machinery of L1 elements for transposition and are specific to hominids [23]. The colonization of the human genome by SVAs is relatively recent, i.e., less than 25 million years, and SVAs account for nearly 3000 copies [16].

A link between the transposition of mobile retroelements (mainly LINE or SINE/Alu) in the human genome and pathologies was noted many years ago. Several examples of LINE or SINE/Alu related cancers are listed in Table 1. For instance, in the late 1980s L1 retroelement insertion into the human proto-oncogene c-myc was found in human breast carcinoma cells (Table 1) [24]. In the study by Morse et al., the structure comparison of myc from a breast ducal adenocarcinoma and from normal breast tissue of the same patient revealed a tumor-specific rearrangement of one myc locus and amplification of the other myc locus. Within the second intron of the rearranged locus was a non-myc sequence with near complete homology to a LINE-1 sequence. In this case, the L1 sequence functioned as a mobile genetic element to produce a somatic mutation [24]. Another example of an L1 element somatic insertion is its integration into the tumor-suppressing gene Apc (adenomatous polyposis coli), found in colon cancer patients (Table 1) [25–27]. The insertion of an Alu element in Apc was also observed in association with desmoid tumors [28].

2. Non-LTR retroelement mobility: a genetic cause of cancer

Most non-LTR retroelement families in the human genome are currently inactive, except for three. L1 elements make up nearly 17% of human genomic DNA (Fig. 2), with a total of about 300,000 copies [8]. Full-length L1 elements, stretching for some 6 kb, have two open reading frames encoding proteins required for the transposition and relocation of non-autonomous elements of the SINE family. L1 elements have been active in the human genome for near 160 million years [8]. SINEs are short elements (100–400 bp) that contain a promoter for polymerase III and do not encode proteins (Fig. 1B). Most known SINE elements are tRNA derivatives, except the Alu element that is present in over a million copies in the human genome, i.e., about 11% [9,10]. The Alu element is specific to primates and has colonized primate genomes for 65 million years [21].
Insertions of the non-autonomous SINE Alu element into the intron of the NF-1 (neurofibromatosis type I) gene lead to a deletion and a reading frame shift in the downstream exon during splicing, which might be associated with neurofibromatosis [29]. The BRCA1 and BRCA2 breast/ovarian cancer associated genes are also the sites of Alu insertions [30,31]. So, one of the germline mutations evidenced by Miki et al. was an insertion of an Alu element into exon 22, which resulted in alternative splicing that skipped exon 22. This event was caused by the retrotransposonal insertion of a transcriptionally active Alu element [30].

3. Alu-mediated chromosomal recombination at the origin of cancer

As noted above Alu repeats account for 11% of the human genome (Fig. 2). These repeats are thought to induce ectopic recombination (also known as non-allelic homologous recombination) at the origin of several cancer diseases. The unusually high density of Alu repeats in the BRCA1 gene favor ectopic recombination and numerous Alu-mediated deletion events have been reported at this locus [32,33]. At least four cases of BRCA1 gene duplication were reported involving the ectopic recombination of Alu elements [33,34]. Although less frequent, Alu-mediated deletions were also described for BRCA2 gene [33,35].

In addition, Alu elements may play a role in chronic myeloid leukemia. This leukemia develops as a result of a translocation between the human chromosomes 9 and 22. The chromosome breakpoints producing this chromosome aberration contained nucleotide sequences of Alu elements [36]. Similarly, an internal tandem duplication of part of the Mll (myeloid/lymphoid or mixed-lineage leukemia) gene, resulting from ectopic recombination between those very Alu elements, may trigger a cascade of events frequently associated with acute myeloid leukemia [37,38]. Duplication of the MYB locus, which encodes an essential transcription factor, is frequent in human cancers [39,40]. The human MYB locus is flanked by Alu repeats and O’Neil et al. showed that its duplication is mediated somatically by homologous recombination.

### Table 1

Examples of TE insertion and TE-mediated chromosomal rearrangements associated with cancer.

<table>
<thead>
<tr>
<th>Locus and/or gene</th>
<th>Associated cancer</th>
<th>TE</th>
<th>Distribution</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APC, adenomatous polyposis coli gene</td>
<td>Desmoids tumors</td>
<td>Alu</td>
<td>Germine</td>
<td>[28]</td>
</tr>
<tr>
<td>APC</td>
<td>Colon cancer</td>
<td>L1</td>
<td>Germine</td>
<td>[26]</td>
</tr>
<tr>
<td>APC</td>
<td></td>
<td>L1</td>
<td>Somatic</td>
<td>[27]</td>
</tr>
<tr>
<td>BRCA1, breast cancer 1 gene</td>
<td></td>
<td>Alu</td>
<td>Germine</td>
<td>[31]</td>
</tr>
<tr>
<td>BRCA2, breast cancer 2 gene</td>
<td></td>
<td>Alu</td>
<td>Germine</td>
<td>[30,31]</td>
</tr>
<tr>
<td>MYC, c-myc proto-oncogene</td>
<td></td>
<td>L1</td>
<td>Somatic</td>
<td>[24]</td>
</tr>
<tr>
<td>NF1, neurofibromatosis 1 gene</td>
<td>Neurofibroma</td>
<td>Alu</td>
<td>Germine</td>
<td>[29]</td>
</tr>
<tr>
<td>Chromosomal deletions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VHL, von Hippel Lindau gene</td>
<td>von Hippel Lindau disease</td>
<td>Alu</td>
<td>Germine</td>
<td>[43,44]</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Breast/ovarian cancers</td>
<td>Alu</td>
<td>Germine</td>
<td>[32,33]</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Breast/ovarian cancers</td>
<td>Alu</td>
<td>Germine</td>
<td>[33,35]</td>
</tr>
<tr>
<td>CDH1, cadherin 1 gene</td>
<td>Hereditary diffuse gastric cancer</td>
<td>Alu</td>
<td>Germine</td>
<td>[45]</td>
</tr>
<tr>
<td>CAD, caspase activated DNase gene</td>
<td>Hepatoma</td>
<td>Alu</td>
<td>Somatic</td>
<td>[46]</td>
</tr>
<tr>
<td>Chromosomal duplication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML1L, myeloid/lymphoid mixed lineage leukemia gene</td>
<td>Acute myeloid leukemia</td>
<td>Alu</td>
<td>Somatic</td>
<td>[37,38]</td>
</tr>
<tr>
<td>MYB, myb transcription factor gene</td>
<td></td>
<td>T-acute lymphoblastic lymphoma</td>
<td>Alu</td>
<td>Somatic</td>
</tr>
<tr>
<td>BRCA1</td>
<td></td>
<td>Breast/ovarian cancers</td>
<td>Alu</td>
<td>Somatic</td>
</tr>
<tr>
<td>Chromosomal translocation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EWSR1-ETV, t(5q23q31)(18q12)</td>
<td>Ewing sarcoma</td>
<td>Alu</td>
<td>Somatic</td>
<td>[42]</td>
</tr>
<tr>
<td>BCR-ABL, t(9;22)(q34;q11)</td>
<td>Chronic myeloid leukemia</td>
<td>Alu</td>
<td>Somatic</td>
<td>[36]</td>
</tr>
</tbody>
</table>
between the flanking Alu elements on sister chromatids in T cell acute lymphoblastic leukemia [41].

Finally, recombination between Alu elements, which causes a translocation involving the Tre-2 oncogene (TRE-USP6, ubiquitin-specific protease-6), has been shown to play an important role in Ewing sarcoma development [42]. Several other examples exist such as Alu-mediated deletions in the von Hippel Lindau gene [43,44], the cadherin-1 gene [45], or the caspase-activated DNase gene [46] (Table 1).

4. Transposable element domestication: an evolutionary gain sometimes related to cancer or cancer suppression

Transposable element domestication corresponds to the diversion of these sequences and the reuse of original functions of recombinases (transposases, integrases) for another role in the cell. At least 47 human genes are known to originate from TEs [8]. In mammals, the Synctin genes that play crucial roles in placenta formation are derived from HERV LTR-retroelements [47,48]. In D. melanogaster, the non-LTR retroelements Heta and Tart assume the function of telomeres and telomerase [49]. TEs are also involved in the formation and epigenetic regulation of heterochromatin as described in Drosophila [50,51] and several other vertebrate and invertebrate species [52–55].

The most famous example of TE domestication in mammals is the derivation of an ancient transib-type element that causes the V(DJ) system at the base of the vertebrate immune system [56–59]. The RAG-1 and RAG-2 genes that promote recombination, retain their ability to relocate their nucleotide sequences during V(DJ) recombination in the lymphocyte genome [60]. Furthermore, V(DJ) recombination events are biochemically similar to the transposition of the Hermes transposon family, such as hobo, Activator, and Tom3, which relocate via the “cut-and-paste” mechanism [59]. In fact, a nucleotide sequence fragment cut out during V(DJ) recombination resembled transposon DNA, except that the former is circularized [58,59]. Fragments cut out by RAG proteins are usually degraded. However, sometimes the proteins can reinsert these fragments into other genome sites with a frequency of about 1 per 13,000–50,000 recombination events in cultured human cells [61–63]. Although this rate of illegitimate insertion cannot be directly extrapolated from cell culture to an organism, a link between these events and B- and T-cell cancerization in the human organism can be presumed [64]. Furthermore, a link between V(DJ) recombination and the onset of cancer associated with recombination-induced chromosome rearrangements has been shown [65,66]. RAG proteins may induce double-strand breaks in sites similar in their structure to signal sequences for V(DJ) recombination. Such double-strand breaks in DNA are potential players in recombination of the genes or receptors of mature T and B cells. This, in turn, entails expression deviations of such proto-oncogenes as LMO2 and BCL2 [65,66].

As an example of Class-II transposon domestication in the human genome, the fusion of the mariner transposon hsmr1 with a methylease coding sequence (SET) is of particular interest with respect to cancer [67,68]. The resulting chimeric protein, referred to as SETMAR, or Metnase, shows both methylease and nuclease activities [68–70] and promotes non-homologous end joining (NHEJ) repair in humans [71–74]. The fusion of the histone methyle SET domain and the transposase domain in the anthropoid lineage to form primate Metnase promotes accurate intrachromosomal NHEJ and thereby suppresses interchromosomal translocations. Metnase may have been selected because it has the function of opposing transposases and thus may play a key role in suppressing translocations that underlie oncogenicity [75–77]. Metnase transposase has been remarkably conserved through evolution; however, there is a clustering of substitutions located in alpha helices 4 and 5 within the putative DNA-binding site, consistent with loss of transposition-specific DNA cleavage activity and acquisition of repair-specific DNA cleavage activity [78]. Both methylease (SET) and transposase domains are necessary for Metnase function in double-strand break repair, and Metnase possesses a unique endonuclease activity that preferentially acts on ssDNA and ssDNA-overhang of a partial duplex DNA [71]. This result of Beck et al. suggests a role for Metnase’s endonuclease activity in promoting the joining of non-compatible ends [71]. Metnase is widely expressed, especially in proliferating tissues, and its involvement in cancer cell resistance to Topoisomerase II-α inhibitors has been suggested. Indeed acute leukemia cells as well as breast cancer cells have an attenuated mitotic arrest due to decatenation inhibition by Metnase [76,77]. Therefore Metnase inhibition will restore chemotherapy sensitivity in these cells.

5. But what may cause or promote transposable element mobility?

Activation of TE mobility may occur as a result of different mechanisms. For example, L1 elements are known to actively transpose during early embryogenesis, which is believed to be triggered by total genome demethylation, or epigenetic reprogramming as shown in murine primordial cells between the E11.5 and E13.5 early embryo stages [79]. As DNA methylation is known to repress various nucleotide sequences, including L1 elements, demethylation may cause TE activation and transposition events [80,81]. Kano et al. have demonstrated that mRNA from L1 elements transcribed in the parental organism can be passed on through oocytes or sperm cells to progeny where reverse transcription ensures further insertions of the element’s copies into the genome of the developing organism during the pre-implantation stage, leading to somatic mosaicism [82]. It seems, therefore, that at least L1 activation paths exist during early stages of mammal development [83]. TE activation is rather common during embryogenesis, so retroelement insertions, including those associated with cancer development may be considered as the cost of phenotypic diversity formation.

Environmental factors may also cause TE mobility. Several chemicals containing mercury (HgS), cadmium (CdS), and nickel (NiO) have been found to increase 3-fold L1 element activity in human cell culture [84]. Meanwhile nickel chloride, which increases L1 activity 2.5-fold, has no direct effect on the sequence of the element or its proteins, but instead inhibits DNA repair systems, leading to L1 transpositions [85]. In general, active TE transposition in living organisms can be induced by a number of environmental factors, like heat shock, viral infection, various chemicals, γ irradiation, etc. [86–90]. Although not clearly established, environmentally-induced TE activation could, albeit indirectly, contribute to human carcinogenesis. This external factor effect on TE-mediated carcinogenesis is further supported by the geographic occurrence patterns of such recombination events. For example, some studies link the expression rates of breast cancer associated gene BRCA2, specific to the Portuguese population, to Alu activity [31,91].

The tumor cell environment itself is favorable to TE mobility. In fact the frequent hypomethylation of chromatin in tumor cells will facilitate TE activation [80,81,92,93]. L1 hypomethylation has been observed in several cancers such as breast carcinoma [94], prostate cancer [95–97], hepatocellular carcinomas [98] or colorectal cancer [99]. Moreover recent findings using second-generation DNA sequencing revealed a total of 9 de novo L1 retrotransposition events in 6 out of 20 examined non-small cell lung tumors [100]. In this study the new L1 insertions were accompanied by a specific genome-wide hypomethylation signature. This is consistent with the idea that epigenome alterations create a permissive environment for TE expression and/or retrotransposition [7,100].

Recently Lee et al., identified almost two hundred somatic TE insertions in 43 high-coverage whole genome sequencing datasets from five cancer types (i.e. colorectal, glioblastoma, multiple myeloma, ovarian and prostate) [101]. Most of the identified TE insertions are L1s (94.5%), then Alu (5.0%) and ERV (0.5%). All the somatic L1 and Alu insertions were observed in epithelial cancer whereas no L1 and Alu insertions were detected in blood and brain cancers. Interestingly, the somatic L1 insertions are biased away from transcriptionally active
regions and conversely overrepresented in common cancer-specific hypomethylation domains [101]. This supports the view that loss of DNA methylation promotes TE integration. Lee et al. also showed that somatic L1 insertions tend to disrupt the expression of tumor suppressor genes commonly mutated in cancer, suggesting that some TE insertions provide a selective advantage during tumorigenesis [101].

Furthermore, tumor cells are known to contain significantly lower quantities of micro RNAs [102], so the repression of TEs through RNA interference, the mechanisms of which are detailed elsewhere [103,104], may be bypassed in cancerous cells. Therefore, while TEs may be linked to cancer development, they themselves can become activated by the cell malignization process that promotes increased mutation and recombination rates in the genome of the transformed cells.

6. Concluding remarks

The list of cancer-linked chromosome rearrangements extends far beyond those caused by the function defects of V(D)J recombination genes. Indeed oncogenic chromosomal rearrangements can be formed at fragile chromosome sites due to imperfect functioning of the NHEJ and homologous recombination reparation systems. The breakpoints of such recombination events are often localized in or near the nucleotide sequences of Alu elements [105]. However, the presence of an Alu sequence itself is not sufficient. It is more the nature of this sequence that induces double-strand breaks and provides an initiation point for recombination [106,107]. Although rare in somatic cells, the existence of ectopic recombination between Alu sequences leads to DNA deletions in germ cells (Table 1) [107]. As noted above, such events can cause tumors such as acute myeloid leukemia and Ewing sarcoma (Table 1) [37,42].

Besides inducing chromosome rearrangements, TEs may also affect gene expression [5,108]. Gene disruption is perhaps the most evident effect of TE insertion, and the inclusion of a TE often results in genetic disorders caused by deleterious premature termination of the peptide sequence. However, gene expression may be affected by TE insertion in different ways, such as the introduction of new polyadenylation sites [109,110], the creation of new exons [111–113], or the insertion of a cis-element at promoter and enhancer regions [114–116]. Such dysregulation of gene expression may be the cause of a number of human diseases including cancers [15–18].

Studies in animal models also support a causative role of TEs in cancer onset. The observation of TEs in mouse cancer cells [117–119] and the use of engineered TEs, such as Sleeping Beauty and PiggyBac as mutagen agents, demonstrate that TE insertion can lead to cancer [120–123]. In the first studies by Dupuy et al. [124] and Collier et al. [125], the reconstructed Tc-1 element Sleeping Beauty transposase was expressed in mice that also carried concatemers of gene trap transposons. In the work of Dupuy et al., crosses of offspring frequently died in utero, but mice which survived until birth succumbed to cancer rapidly [124]. In the second study by Collier et al., the mice did not develop tumors spontaneously but mobilization of the transposase appeared to accelerate the onset of sarcomas in p19 ARF null mice [125]. In further studies a conditional Sleeping Beauty transposon system was used in order to target specific tissues such as gastrointestinal tract and liver [120]. Such a TE-based insertional mutagenesis provides great help in identifying cancer candidate genes in mice [120,121] and fish [126].

Several transcription factors, such as Foxa1, GATA, Lyf-1, Sp1, YY1; retinoic acid receptors (RARs) and p53 have binding sites in Alu elements [127–133]. In particular, putative p53 binding sites have been discovered in several LTR- and non-LTR elements of the human genome [127]. The p53 tumor suppressor gene is a transcription factor that induces activation and repression of more than a thousand human genes and is involved in about 50% of human cancers [134,135]. Previous studies have shown that p53 can repress RNA polymerase III transcription [136]. Therefore, mutations in p53 leading to DNA binding defects would result in RNA polymerase III hyperactivity that characterizes many tumors [136]. Moreover, an increased level of Alu RNA has been evidenced in hepatocellular carcinoma [137]. Cui et al., hypothesized that p53 binding to Alu elements, plus the transcriptional repression can change the expression of the host genes through steric interaction between RNA polymerase II, which is responsible for host gene expression, and RNA polymerase III that ensures Alu transcription [127]. This hypothesis has to be demonstrated as well as its consequences with respect to cancer onset and/or its development.

Finally, similarities between the reverse transcriptase of human L1 elements and telomerase were described [138,139]. The telomerase function and the telomere sequence are provided by domesticated TEs in several organisms, such as Drosophila, rotifers, the diatom microalga Phaeodactylum tricornutum, the silkworm Bombyx mori, etc. [49,140–142]. Due to the crucial involvement of telomerase in cancer development and the complex regulation of telomere lengthening [143,144], this suggests that retrotransposons, by providing an alternative mechanism to maintain telomere structure, may play a yet unelucidated role in cancer biology.

Although a direct relationship between TE insertion and cancer onset is difficult to evidence, the above examples illustrate that TE-linked rearrangements in the human genome may be associated with cancers of various etiology. To demonstrate that TE insertion or TE-mediated rearrangement causes human cancer is difficult because 1) sequencing repetitive elements remain a challenge and 2) experimental demonstration is impossible in humans. However, recent animal model studies with refined bioinformatic analyses of insertional mutagenesis screens [145–148], and improved human genome analysis such as the recent work of Lee et al. [101], offer solid arguments in favor of the causative role of TEs in cancer. Conversely, the SETMAR example illustrates the case of a positive antitumor effect due to TE domestication.

According to the available data, TEs are clearly a source of mutations and are therefore able to lead to cancer development. Cancer is not a single mutation event, but a result of the accumulation of several mutations affecting the signaling of cell cycle, apoptosis, angiogenesis and DNA repair. Thus TEs can produce mutations causing cancer susceptibility, but are not the sole cause of cancer.

Acknowledgements

The author is grateful to Marilyn Twell for revising the English text and to Dr. Albane de Gentile for her continuous support during this project.

References


N. Lane, W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, W. Reik, Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse, Genesis 35 (2003) 85–93.

