Seminars in Cancer Biology 20 (2010) 197–199

Contents lists available at ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Editorial

Repetitive elements and genome instability

Transposable elements are discrete segments of DNA that have the distinctive ability to move and replicate within genomes. Transposons were discovered in the 1940s by Barbara McClintock (who later was awarded with the Nobel Prize) in the maize genome, and have since been found ubiquitous in essentially all living organisms. The process of element movement is generally called transposition, and can contribute to insertional mutagenesis, altered gene expression and recombination. Transposons make up significant fractions of genomes; for example, about 45% of the human genome is composed of sequences of a variety of different elements [1]. Despite their parasitic nature, there is increasing evidence that transposable elements are a powerful force in gene evolution. Indeed, about 50 human genes are derived from transposable elements [1], among them genes that are responsible for immunoglobulin gene recombination in vertebrates.

Transposable elements fall into two major classes. The first is retroelements that transpose through an RNA intermediate and include long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeat (LTR) retrotransposons. Retroviruses are specialized LTR retrotransposons that have the capacity to move from one cell to another by infection, whereas endogenous retroviruses (ERVs) are retroviruses that lost their capacity to infect cells, but nevertheless can retrotranspose in the genome. Retroelements transpose through a replicative (copy-and-paste) mode of transposition, in which the transposon does not get excised from its donor locus but instead a copy of it is produced which inserts elsewhere in the genome. Thus, replicative transposition leads to an increase in the copy number of the transposon within a genome.

The major non-LTR retrotransposon in mammals is LINE-1 (or L1). L1 has about 5×10^5 copies in the human genome, thereby making up about 17% of human genomic DNA. LINEs contain two open reading frames (ORFs) that encode proteins required for transposition (Fig. 1A). The element-encoded endonuclease generates a single-stranded nick in the target DNA, and the reverse transcriptase uses the nicked DNA to prime reverse transcription from the 3′ end of the LINE RNA. The LINE retrotransposition machinery is believed to be responsible for most reverse transcription activity in the genome, including retrotransposition of the non-autonomous SINEs and generation of processed pseudogenes.

SINEs are short (about 100–400 bp) retrotransposable elements that contain a polymerase III promoter and encode no proteins (Fig. 1B). These non-autonomous transposons are thought to use the LINE machinery for transposition [2]. The vast majority of known SINEs are derived from tRNA sequences, with the exception of the human *Alu* element, which is derived from the 7SL component of the signal recognition particle. *Alu* elements were originally identified as repetitive DNA elements in human DNA renaturation curves, and contain a recognition site for the restriction enzyme *Alu*I. *Alu* elements are represented in the human genome with $>1 \times 10^6$ copies which make up about 11% of the total genome. *Alu* is the only active SINE in humans. Another non-autonomous retrotransposon that likely uses the L1 machinery for retrotransposition is the hominid-specific SVA (SINE-VNTR-*Alu*) family. The composite SVA elements are currently active in humans, and are present in about 2700 copies in the human genome [3].

LTR retrotransposons are similar to retroviruses in that they are flanked by long terminal repeats that contain transcriptional regulatory elements that drive the expression of at least two ORFs similar to the *gag* and *pol* genes (Fig. 1C). LTR retrotransposons are widely distributed in eukaryotes, and make up about 8% of the human genome. Transposition occurs through reverse transcription of the retrotransposon RNA, and integration of the resultant

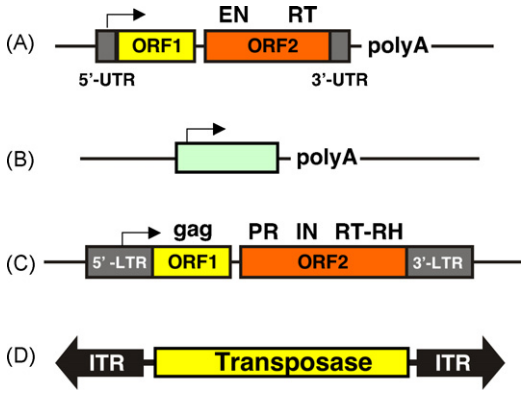

Fig. 1. Structures and organization of the main types of transposable elements. (A) Non-LTR retrotransposon. The element consists of a 5′ untranslated region that has promoter activity (arrow pointing towards the downstream genes), which is required to drive transcription of the element-encoded genes. ORF1 encodes a nucleic acid binding protein. ORF2 encodes an endonuclease (EN) and a reverse transcriptase (RT). The element has a polyA tail. (B) A typical SINE. The element is a small, RNA-derived pseudogene, which is transcribed from an RNA polymerase III promoter within the element (arrow). The element has a polyA tail. (C) LTR retrotransposon. The element consists of long terminal repeats (LTRs) similar to those of retroviruses. The LTRs flank two open reading frames. ORF1 encodes the group specific antigen (gag), ORF2 encodes a protease (PR), an integrase (IN), and a reverse transcriptase-RNaseH (RT-RH) function. (D) DNA transposon. The central transposase gene (yellow box) is flanked by inverted terminal repeats (ITRs, shown as black arrows). The ITRs contain the binding sites for the transposase and sequences that are required for transposase-mediated cleavage.

1044-579X/$ – see front matter
doi:10.1016/j.semcancer.2010.08.002

cDNA into a new location by the integrase protein. ERVs appear to have been recently active in mammalian genomes [4].

Members of the second major class of transposable elements are called DNA transposons because they move directly as DNA, by a conservative (cut-and-paste) mechanism of transposition, in which the element gets excised from the donor locus and is subsequently reinserted elsewhere. The simplest DNA transposons are composed of short terminal inverted repeats flanking a transposase gene (Fig. 1D). The transposase binds to the inverted repeats, and catalyzes 'cut-and-paste' transposition. Although DNA transposons are widespread in prokaryotes, they are less abundant in eukaryotes. For example, only 2–3% of the human genome is composed of DNA transposons, and there is no indication for their activity in the past 50 million years [1,5]. DNA transposons carry an attractive and elaborate enzymatic machinery as well as DNA components that have been exapted by the host genome *via* an evolutionary process referred to as molecular domestication. Indeed, approximately 50 human genes evolved from transposable elements, mostly from DNA transposons [1]. These include the RAG1 and RAG2 immunoglobulin gene recombinases that mediate V(D)J recombination that is central in the adaptive immune response common in all jawed vertebrates [6]. Thus, although transposable elements have not been selected for conferring selective advantage to the host, they can contribute useful functions to genomes.

A general feature of transposable elements is that they can replicate independently of the cellular replication cycle, and new copies can emerge in new locations in the genome, a process that is inherently carries a mutagenic potential. The power of transposition to shuffle DNA around, to generate novel gene expression patterns and thereby drive evolution was first hypothesized by McClintock, and is today a widely accepted paradigm in the scientific community. However, a process that might have beneficial consequences on an evolutionary time-scale at the population level, can easily turn out to be deleterious at the level of the individual, in case it leads to disease. This issue of *Seminars in Cancer Biology* highlights the potential of transposable elements (genomic repeats in general) to contribute to genomic instability and cancer. There are three major mechanisms through which transposable elements might trigger genomic rearrangements and deregulation of gene expression: (i) *de novo* insertion events that can cause insertional mutagenesis if they land within a gene (disruption of gene function), (ii) transcriptional deregulation of gene function, and (iii) secondary DNA rearrangements including deletions, duplications, inversions and translocations promoted by dispersed copies of transposable elements scattered around the genome. As highlighted by Belancio et al. [7], and by Konkel and Batzer [8], non-LTR retotransposons are especially prone to contribute to the genesis of structural variations in the human genome due to their inherent mobility, abundance, and high sequence identity. The potential of dispersed transposon copies to promote non-allelic homologous recombination can be even more damaging to the genome than a *de novo* insertion, contributing to the development of disease, including cancer. This effect has nothing to do with mobility of genetic elements in the genome; rather, the sheer presence of large numbers of repeats (a large fraction of them having been distributed across the genome by transposition) can trigger recombinational mechanisms. Chen et al. [9] describe the various mechanisms responsible for such rearrangements, including non-allelic homologous recombination, gene conversion, single strand annealing and microhomology-mediated replication-dependent recombination.

Transposon movement also leaves its mark in the genome by aberrant transposition events that induce genomic rearrangements including deletions, translocations and duplications of chromosomal DNA. For example, L1 elements can carry non-transposon sequences into new places, a process that can contribute to "exon shuffling" and thus to gene evolution [10]. This is because L1 transcription can read through the native transcription termination site of the element into flanking genomic sequences. It is estimated that about 0.5–1% of the human genome may have been generated by L1-mediated transduction of 3′-flanking sequences. Similarly, the composite non-LTR retrotransposon SVA occasionally carries over 5′-flanking genomic sequences to new chromosomal locations [11]. As discussed by Hancks and Kazazian [12], this is presumably due to the requirement of external promoters to drive transcription of the elements that produce transcripts containing the entire SVA element plus upstream sequences. These 5′-transduced SVA elements may give rise to entire subfamilies as a result of repeated rounds of retrotransposition events. Another mechanism by which SVA elements might mutagenize gene expression is interruption of the normal splicing pattern of genes by the splice signals that can be incorporated in genes by SVA retrotransposition.

DNA hypomethylation is a hallmark of cancer, and deregulation of the transcriptional activity of the LTRs of human ERVs (HERVs) can contribute to further destabilization of the cancer genome via transcriptional effects. For example, as outlined by Romanish et al. [13], deregulated transcription of HERV sequences could facilitate oncogene activation by donating their LTRs as promoters and/or enhancers.

Just like retroelements, DNA transposons are also inherently mutagenic because they can insert into many places of the genome. Thus, a cellular DNA recombination system that evolved from an ancient transposon, like V(D)J recombination, is better to be downregulated in its transpositional activity in order to avoid mutational damage. Indeed, the integration step of this specialized transposon appears to be efficiently blocked in cells. However, as Ramsden et al. [14] argue, V(D)J recombination is not completely risk-free, because the excised DNA fragments may be involved in "hit-and-run" transposition events (an insertion event followed by re-excision) as well as in intermolecular recombination.

Finally, genomic trouble can not only be triggered by transposable elements that are endogenous in the genome, but also by transposon-based vectors that we exogenously insert into the genome for the purpose of genetic screens or for transgenesis including gene therapy. Indeed, the capacity of DNA insertion across the genome can be exploited to deliberately disrupt gene function. For example, the *Sleeping Beauty* transposon [15] has become a useful tool for somatic mutagenesis in mice. Dupuy [16] discusses the various components (strong transcriptional signals as well as splice sites) that have been incorporated in the transposon to either upregulate the expression of oncogenes or disrupt the expression of tumor tuppressor genes, and highlights the power of this approach to devise tissue-specific screens to uncover the genetic basis of various cancers. Just the opposite: transcriptional interference with endogenous gene function is a side effect that gene therapists want to avoid at all costs. As discussed by Kustikova et al. [17], selection of gene-modified cells in the patient may support the preferential survival of clones with insertional deregulation of genes that are involved in the control of engraftment, proliferation or differentiation, in the worst case initiating oncogenic progression. Nevertheless, several lines of evidence show that the genomic risk of gene therapy can be potentially reduced by targeting cells that lack sustained replicative potential, targeting vector integration away from genes, reducing the number of vector copies per cell, and designing gene expression cassettes that avoid long-distance enhancer interactions or fusion transcripts.

The reader finds in this issue of *Seminars in Cancer Biology* selected examples of mutagenicity and genomic instability mediated by repetitive DNA, most of which has been amplified and distributed across the genome by transposition. These examples cover the whole spectrum of the various types of transposable elements and the consequences of their presence and activities in the genome. Continued efforts of large-scale sequencing of tumor

samples will shed further light on the role these repeats play in tumorigenesis, whereas transposon-based genetic screens will continue to enhance our understanding of other genetic determinants of cancer.

Acknowledgements

We thank Eva Klein and George Klein for the invigorating discussions on transposable elements and the link between their activities and cancer. We are grateful to all contributing authors for their commitment and spirited participation in covering their respective fields of research. Finally, we thank Ben Rowe for his assistance and patience in editing and producing this issue.

References

[1] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
[2] Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 2003;35:41–8.
[3] Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, et al. SVA elements: a hominid-specific retroposon family. J Mol Biol 2005;354:994–1007.
[4] de Parseval N, Heidmann T. Human endogenous retroviruses: from infectious elements to human genes. Cytogenet Genome Res 2005;110:318–32.
[5] Pace 2nd JK, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 2007;17:422–32.
[6] Jones JM, Gellert M. The taming of a transposon: V(D)J recombination and the immune system. Immunol Rev 2004;200:233–48.
[7] Belancio VP, Roy-Engel AM, Deininger PL. All y'all need to know bout retroelements in cancer. Semin Cancer Biol 2010;20:200–10.
[8] Konkel MK, Batzer MA. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol 2010;20:211–21.
[9] Chen JM, Cooper DN, Ferec C, Kehrer-Sawatzki H, Patrinos GP. Genomic rearrangements in inherited disease and cancer. Semin Cancer Biol 2010;20:222–33.
[10] Moran JV, DeBerardinis RJ, Kazazian Jr HH. Exon shuffling by L1 retrotransposition. Science 1999;283:1530–4.
[11] Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, et al. 5′-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res 2009;19:1992–2008.
[12] Hancks DC, Kazazian Jr HH. SVA retrotransposons: evolution and genetic instability. Semin Cancer Biol 2010;20:234–45.
[13] Romanish MT, Cohen CJ, Mager DL. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. Semin Cancer Biol 2010;20:246–53.
[14] Ramsden DA, Weed BD, Reddy YV. V(D)J recombination: born to be wild. Semin Cancer Biol 2010;20:254–60.
[15] Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 1997;91:501–10.
[16] Dupuy AJ. Transposon-based screens for cancer gene discovery in mouse models. Semin Cancer Biol 2010;20:261–8.
[17] Kustikova O, Brugman M, Baum C. The genomic risk of somatic gene therapy. Semin Cancer Biol 2010;20:269–78.

Zoltán Ivics*
Zsuzsanna Izsvák
Max Delbrück Center for Molecular Medicine, Robert Rössle Str. 10, D-13092 Berlin, Germany

* Corresponding author. Tel.: +49 30 9406 2546; fax: +49 30 9406 2547.
E-mail address: zivics@mdc-berlin.de (Z. Ivics)