Minireview

TEs or not TEs? That is the evolutionary question

Keren Vaknin, Amir Goren and Gil Ast

Address: Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69987, Israel.

Correspondence: Keren Vaknin. Email: uakninke@post.tau.ac.il

Abstract

Transposable elements (TEs) have contributed a wide range of functional sequences to their host genomes. A recent paper in *BMC Molecular Biology* discusses the creation of new transcripts by transposable element insertion upstream of retrocopies and the involvement of such insertions in tissue-specific post-transcriptional regulation.

Among the many factors that contribute to the diversity of genome structure and organization in different eukaryotes are transposable elements, which comprise a large fraction of many eukaryotic genomes. It is now well established that the activities of these elements represent a major evolutionary force that has shaped the genes and genomes of many species, contributing a wide range of functional sequences. Some transposable elements encode the enzyme reverse transcriptase, which as well as being involved in the proliferation and movement of the element within the genome, occasionally reverse transcribes a mature spliced cellular mRNA and inserts the DNA copies (cDNAs) into new locations within the genome by retrotransposition [1] (Figure 1). Because they have been generated from a mature mRNA, these DNA sequences lack introns, promoter sequences and upstream regulatory elements and are known as 'retrocopies'. This mini-review addresses work published recently in *BMC Molecular Biology* by Chiu-Jung Huang and colleagues [2] in which they demonstrated that, over the course of evolution, some retrocopies can acquire a new promoter, often by the insertion of a transposable element upstream of the retrocopies, and are transcribed into a functional gene product. Functional genes derived from retrocopies are known as 'retrogenes'.

Transposable element sequences provide new exons for host genes

The generation of new exons and new genes is a major force that advances genomic complexity. Three mechanisms are thought to be responsible for the origin of new exons. Two of these yield new exons within existing genes. The first is known as exon shuffling (or exon duplication); in this process, a new exon is inserted into an existing gene by recombination or is duplicated within the same gene, and by alternative splicing some of the mature transcript contains this exon. In the second mechanism, alternative exon cassettes are derived from constitutively spliced ones by mutations at splicing signal sites that weaken the selection of particular exons by the splicing machinery [3]. The third mechanism is the exonization of transposable element sequences. In this process transposable element sequences are first inserted into introns, and then gain mutations that allow the RNA splicing machinery to recruit part of the inserted transposable element into the mature mRNA [4].

The proliferation of transposable elements within the genome provides repeated sequences that promote recombination and can also provide sites that regulate transcription, polyadenylation sites, splicing signals and protein-coding sequence [5]. Most exonizations of transposable elements generate internal exons that are alternatively spliced [4]. Two mRNAs are thus produced from these genes: one is the original mRNA that skips the new exon, while the other includes it by alternative splicing. The latter mRNA is a minor product, and its function can be 'tested' by natural selection without losing the original function of the gene. Exonization can also lead to the extension of existing exons by the activation of alternative donor or acceptor splice sites; or splicing may even be abolished by the mutation, which leads to retention of the intron in the mRNA.

In mammalian genomes, the process of exonization just described is restricted to transposable elements inserted into introns or exons that are part of untranslated regions (UTRs). However, there is no indication that transposable element sequence has become incorporated into existing protein-coding exons. It was shown that insertion of a transposable element into UTR exons sometimes leads to a phenomenon called 'intronization' [5]. In this case, the insertion generates a new intron within an existing exon, which can alter gene expression and create, for example, a new binding site for a regulatory microRNA [6].

Thus, the incorporation of transposable element sequence into a genome is one means of generating diversity among transcriptomes. A functional exonized transposable element usually does not disrupt the coding integrity of the
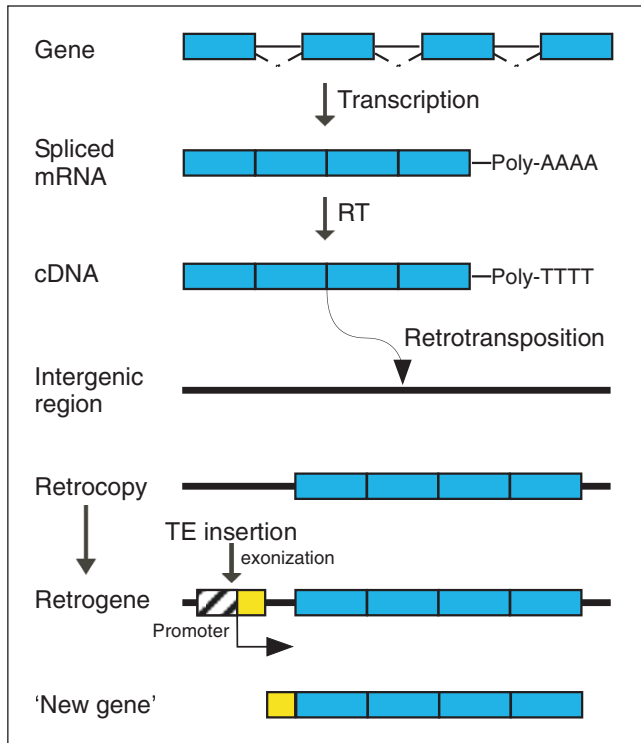
Figure 1

The generation of a retrogene. Infrequently, a spliced, capped and polyadenylated cellular mRNA molecule is reverse transcribed (RT) into cDNA and integrated by retrotransposition into the genome in an intergenic region, creating an intronless copy of the gene, a retrocopy (blue), lacking its own promoter and regulatory elements. Over time, the insertion of a transposable element (TE) upstream of the retrocopy can provide both a promoter and, by the process of exonization, a new 5' UTR exon (yellow), such that, after splicing, the transcript yields a functional mRNA. The new functional gene is termed a retrogene and if useful to the organism, will be maintained in the genome.

gene of which it has become a part - the length of the exonized region is divisible by three, avoiding the generation of stop codons - and has a relatively high probability of inclusion by alternative splicing compared with non-functional exonized transposable elements [5]. Exonization can occur in any gene that undergoes RNA splicing - it is not restricted to protein-coding genes but to all spliced genes.

Formation of new genes by retrotransposition of transposable elements within retrogenes

Mammalian genomes contain intronless DNA copies of more than 1,000 different spliced mRNAs, and some of these retrocopies have been converted into functional retrogenes by the processes outlined above [7]. In their recent paper, Huang *et al*. [2] provide insight into the creation of the retrogenes *Rtdpoz-T1* and *Rtdpoz-T2* (which will be referred to as *T1* and *T2*) in the rat genome.

The 5' UTRs of these two genes have been the sites of multiple transposable element insertions, resulting in the generation of 11 different transcripts (isoforms). The *RTdpoz* family of elements are distributed over seven different chromosomes of the rat genome but the bulk of them map over an approximately 700 kb segment on chromosome 2 (including *T1* and *T2*). *T1* and *T2* exons are derivatives of mostly repetitive sequences of L1 and ERV transposable elements, particularly in the *T1* transcripts. The first exon of both genes is the result of exonization of the same transposable element, and both *T1* and *T2* are transcribed from a common promoter associated with this leader exon, which is located upstream of the retrogene. Thus, the exonization of a transposable element has resulted in transcriptional activation of the intronless *T1* and *T2* retrocopies.

Interestingly, most mammalian retrogenes are expressed mainly in the testes, where their transcripts participate in spermatogenesis and other unique male germline functions. Transcription in testes appears to be less regulated than in other somatic tissues [8], which might lead to a higher level of exonization of transposable elements in this organ. In support of this hypothesis, Huang *et al*. [2] show that *T1* and *T2* are expressed exclusively in the testis and during early stages of embryonic development.

The authors also show that exonization within a retrogene can add new regulatory motifs and new protein-coding sequences. They find that some of the alternatively spliced transposable-element-derived exons located upstream of the original ATG translation start site of the retrocopy can provide a new open reading frame (ORF) and a new start codon. These insertions have both an influence on gene expression at the level of transcription, and in the *T1* gene, the new ORF and ATG triplet also repress translation of the RNA transcript.

The study by Huang *et al*. [2] adds a new twist to exonization: transposable elements not only provide functional sequences within genes, but they can also provide promoter sequences located upstream of retrocopies of intronless mRNA. Transcription from such sites results in mRNA precursors containing 5' UTR exon and intron sequence from the transposable element and the exon from the retrocopy gene. Splicing results in mRNAs that are 'live on arrival' as they maintain the coding capacity of the original gene. The fate of such new genes is determined by selective pressures during evolution.

References

1. Pace JK 2nd, Feschotte C: **The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage.** *Genome Res* 2007, **17:**422-432.
2. Huang C-J, Lin W-Y, Chang C-M, Choo K-B: **Transcription of the rat testis-specific *Rtdpoz*-T1 and - T2 retrogenes during embryo development: co-transcription and frequent**

exonisation of transposable element sequences. *BMC Mol Biol* 2009, **10:**74.

3. Lev Maor G, Goren A, Sela N, Kim E, Keren H, Doron-Faigenboim A, Leibman-Barak S, Pupko T, Ast G: **The "alternative" choice of constitutive exons throughout evolution.** *PLoS Genet* 2007, **3:**e203.

4. Makalowski W, Mitchell GA, Labuda D: **Alu sequences in the coding regions of mRNA: a source of protein variability.** *Trends Genet* 1994, **10:**188-193.

5. Sela N, Mersch B, Gal Mark N, Lev Maor G, Hotz-Wagenblatt A, Ast G: **Comparative analysis of transposed elements' insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome.** *Genome Biol* 2007, **8:**R127.

6. Babushok DV, Ostertag EM, Kazazian HH: **Current topics in genome evolution: Molecular mechanism of new gene formation.** *Cell Mol Life Sci* 2007, **64:**542-554.

7. Kedde M, Agami R: **Interplay between microRNAs and RNA-binding proteins determine developmental processes.** *Cell Cycle* 2008, **7:**899-903.

8. Schmidt EE: **Transcriptional promiscuity in testes.** *Curr Biol* 1996, **6:**768-769.