Heredity (2009) 103, 279–280

www.nature.com/hdy

NEWS AND COMMENTARY

Human specific exons

When new exons are born

R Sorek

Heredity (2009) **103,** 279–280; doi:10.1038/hdy.2009.62; published online 3 June 2009

In contrast to the common belief before the genomic era, we now know that new exons are frequently 'born' in the genomes of individual species. Thousands of these new exons are alternatively spliced into human transcripts, but as the vast majority of these exons are expressed at very low levels, it was hypothesized that almost none of them were functional. Using a microarray-based interrogation of newly born, *Alu*-derived exons, Lin *et al.* (2008) now show that some of these exons are constitutively expressed in multiple human tissues, indicating that they have 'adapted' to become integral, functional parts of protein-coding human genes.

Alu elements are primate-specific retrotransposons, sized ~300 bp, which constitute more than 10% of the human genome, appearing in it over one million times (Lander *et al*, 2001). They are abundant in introns and are generally considered 'selfish' DNA elements. Intriguingly, it has been shown that some *Alu*s can be 'exonized', that is, turned into new exons. Such exonizations occur when random mutations activate 'hidden' splice sites within the *Alu* sequences, so that the splicing machinery recognizes part of the *Alu* element as a *bone fide* exon (Sorek, 2007).

Historically, the identification of *Alu* exons was mostly achieved using alignments of expressed sequence tags (ESTs) to the human genome. Such alignments pinpoint pieces of transcripts that are derived from intronic *Alu*s, thus identifying *Alu*s that give rise to new exons. When multiple ESTs are aligned to the same genomic locus, it is possible to use their alignments to identify alternative splicing, that is, exons that appear in some ESTs but are absent from others. Indeed, EST-based analyses have revealed that nearly all exonized *Alu* elements in the human genome are alternatively spliced (Sorek *et al*, 2002). Moreover, these newly created *Alu* exons were found to be spliced into the transcript at low frequencies with only a small fraction of the transcripts containing the new exon. The original transcript therefore remains intact, making the effect of the exon insertion nearly neutral. This exon creation mechanism, where new *Alu* exons are alternatively spliced in low frequencies, was suggested to allow exonized sequences to increase the coding and regulatory versatility of the human transcriptome but at the same time keep the original transcripts intact (Xing and Lee, 2006).

It is of course interesting to look for cases of *Alu* exonizations that were fixed in the human transcriptome as constitutively spliced exons, or that have evolved to be tightly regulated in a tissue-specific manner. Such unique events are most probably cases of 'functional' new exons that contribute to primate-specific or human-specific traits. However, until the study by Lin *et al.* (2008), no systematic expression analysis had been conducted on *Alu* exons, because the resolution of differential expression analysis based on ESTs is too low. Microarray-based analyses were also challenging, as most expression arrays contain probes that cover only the minority of exons, and hence are usually blind to alternative splicing.

Lin *et al.* (2008) have elegantly addressed this challenge by using public data from Affymetrix human exon array experiments on 11 tissues. This array is a high-density exon-tiling microarray platform that contains 6 million probes targeting all human exons at a density of 4 probes per exon (Clark *et al*, 2007). As *Alu* elements are repetitive in the genome, the authors only tested the 330 *Alu* exons for which at least three probes could uniquely report on the expression of the individual exon, therefore minimizing the possibility of detection based on false cross-hybridization between probes.

Using the tiling array data, Lin *et al.* (2008) uncovered eight human *Alu*-derived exons that were constitutively spliced in all 11 tissues tested, and an additional six exons that were included in the majority of transcripts ('major form' exons) in at least one tissue. These *Alu* exons, therefore, represent a subset of sequences that were recently adapted as functional elements in the primate lineage. Until this study, constitutively spliced *Alu* exons were thought to be almost exclusively associated with genetic disorders, as their insertion within a protein coding region is expected to have strongly deleterious effects. The results of Lin *et al.* (2008) show that, in rare cases, constitutively spliced *Alu*s can occur without causing adverse effects. Moreover, the authors also detected four exons that are alternatively spliced in a regulated manner, that is, are included or skipped in a specific tissue, highlighting the possibility that these exons may have gained a new tissue-specific function.

As *Alu* elements are primate-specific, the exons examined in this study could represent genomic innovations shared by all members of the primate lineage. It is even more intriguing to find exons showing splicing patterns unique to the human genome only. Using comparative human/chimp/macaque qPCR experiments on the SEPN1 gene (encoding selenoprotein N1, linked to a form of congenital muscular dystrophy), Lin *et al.* (2008) showed that the transcript containing the *Alu* exon is the major form in muscle tissues in the human lineage only. Therefore, the muscle specificity of the human SEPN1 *Alu* exon was acquired after the divergence of humans and chimpanzees.

Although the Lin *et al.* (2008) study is the first to tackle the challenge of testing *Alu* exon expression on a large scale, its results fit well to expectations. The vast majority (97%) of tested exons were either expressed below the detection threshold, or represented as minor spliced forms, in agreement with previous observations (Xing and Lee, 2006). Furthermore, most (seven out of eight) *Alu* exons that were found to be constitutively spliced were inserted in the untranslated regions (UTR) of genes, not affecting the protein coding portion of the gene they were inserted into. These results therefore strongly indicate that almost all *Alu*-derived exons represent non-functional evolutionary intermediates that are rarely incorporated into transcripts.

Even so the Lin *et al.* (2008) study was able to pinpoint the very small subset of *Alu* exons that might have gained 'true' functions. On the basis of their results, the next challenge will be to pin down how these new exons affect the function of the genes in which they reside. As these exons are by their nature primate-specific, their adaptation could affect primate-specific traits that were acquired after the separation of the primate lineage. Future studies on this specific set of exons could therefore

shed more light on the uniqueness of the human species.

R Sorek is at the Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: rotem.sorek@weizmann.ac.il

Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H *et al*. (2007). Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol* **8**: R64.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al*. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL *et al*. (2008). Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet* **4**: e1000225.

Sorek R (2007). The birth of new exons: mechanisms and evolutionary consequences. *RNA* **13**: 1603–1608.

Sorek R, Ast G, Graur D (2002). Alu-containing exons are alternatively spliced. *Genome Res* **12**: 1060–1067.

Xing Y, Lee C (2006). Alternative splicing and RNA selection pressure–evolutionary consequences for eukaryotic genomes. *Nat Rev Genet* **7**: 499–509.

Editor's suggested reading:

Loreto ELS, Carareto CMA, Capy P (2008). Revisiting horizontal transfer of transposable elements in Drosophila. *Heredity* **100**.

Sabot F, Schulman AH (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**.