Poster H-23

Are competitive tandem splice donors utilized to shift out-of-frame and trigger nonsense-mediated decay?

Authors:
Ralf H. Bortfeldt (*Friedrich Schiller University of Jena*)
Stefan Schuster (*Friedrich Schiller University of Jena*)
Dirk Holste (*Institute of Molecular Pathology Vienna*)

Short Abstract: Alternative splicing (AS) is a central mode of genetic regulation, and the classification and subsequent characterization of patterns is the first step to obtain a global picture of AS exons. Here, we investigate whether alternative 5'ss exons might be produced by competitive tandem donor splice sites in the human genome.

Long Abstract:
The splicing of precursor to messenger RNAs (pre-mRNAs) is a central mode of genetic regulation, and a significant number of human genes encode proteins by alternative splicing (AS) pathways. Using the transcript-to-genome alignment system GENOA [1], AS events are detected and distinguished in terms of whether mRNA isoforms differ by inclusion or exclusion of an exon, or whether isoforms differ in the usage of a 3' splice site (3'ss) or 5'ss, producing alternative 3'ss exons (A3Es) or alternative 5'ss exons (A5Es), respectively. A3Es and A5Es give rise to two types of exon segments – the 'core' segment common to both (short and long) splice forms and the 'extension' that is present in only the longer isoform. These descriptions are not necessarily exclusive, but an exon can make several alternative splice site choices. In additional to biochemical studies, computational identification and analysis of AS events have been conducted, enabled by the abundance of millions of different transcripts compiled in sequence databases. Here, two challenges for bioinformatics of AS are the recognition of authentic AS patterns from an average of 200-300 expressed sequence tags (ESTs) for each annotated human gene, and the generation of hypotheses for regulatory mechanisms of alternative splice site choice. In the set of human alternative exons, the length distribution of occurrences of A3E and A5E extensions ranges for moderately or higher frequent AS events from about 3 to 100 nucleotides (nt), and shows several peaks at smaller length scales, including a clear trimer extension for A3Es as well as a tetramer extension of A5Es. Typically, such small extensions (shorter than 6nt) were excluded from analysis because of the possibility to result from EST sequencing or alignment errors. Recently, the authenticity and functional impact of observed trimer competitive acceptor sites of A3E has been demonstrated, which arise from tandem acceptors that utilize NAG/NAG/ [2], and about 5% of human genes have been estimated to encode protein isoforms by using this subtle AS type. Here, we study numerically observed tetramer competitive donor sites of A5Es, which arise from /GTRA/GT, and possible implications of this yet uncharacterized AS event.
For 5'ss recognition, it is necessary to have both U1 snRNP-binding to positions -3 to 6 (U1 binds also to splicing-enhancing G-triplets, if present, at positions -3 to –1), and subsequently U6 snRNP-binding to positions 2 to 6. For each AS event with tetramer donor sites (~1,500 binary 5'-distinct alternative splice site choices made by GENOA), the

alternative exon was labelled as a ‘major’ or ‘minor’ splice form, by the number of aligned transcripts (see Figure 1). In order to assess possible alignment errors, we applied stringent filtering criteria and compared/contrasted genomic 5'ss (and the corresponding 3'ss) position in three different ways with/against (i) mRNA-to-genome alignments extracted from the UCSC Genome Browser [3]; (ii) spliced-alignments detected by the program EXALIN [4]; and (iii) to spliced-alignments made by the program SIM4 with modified parameters (used by GENOA only for EST-to-genomic alignments). We found that modest 12 % (179/1,493) of initially detected tetramer donors were exactly redetected by (i); within this overlap, about 50% (89/179) were also contained in (ii); while about 14% (202/1,493) initially detected by (iii). Further manual inspection of the above 89 and 202 tandem donors left 21% (19/89) and 53% (106/202) ambiguous alignments in total, highlighting the abovementioned difficulties of aligning short splice variations to genomic DNA. Under manual inspection of alignments, one can estimate a lower specificity of Sp = 6% (96 TP/1,493 AP) for the SIM4-aligned set, and Sp = 48% (96 TP /202 AP) for EXALIN-aligned A5E set. While this yielded improvement of Sn, neither computational method tested outputs sufficiently accurate AS patterns of competitive donor sites, possibly generally affecting alignments of this type positioned around splice sites. The analysis was indicative several possible sources of misaligned sequences, pointing to the occurrence of nearly repetitive oligonucleotides (4-10nt long) at donor and acceptor sites, and to single-nucleotide mismatches as well as indels within splice sites. For further clarification, the remaining set of tandem donor-spliced exons is subjected to ongoing investigations, involving RT-PCR validation, since these genes and their translational products were rich in biological-regulatory functions, e.g., breast cancer, cell-cycle control and signaling, or transcriptional machinery.

It is tempting, albeit with caution, to speculate about implications that arise from the utilization of competitive donor sites. Firstly, one immediate consequence is an out-of-frame shift that could subject extended messages to the nonsense-mediated RNA decay pathway (NMD). The fact that tandem donors that are frequently supported by a small number of longer splice forms, in turn, might be explained as a consequence downstream of NMD. Secondly, for coding exons that escape NMD, these subtle variations can severely affect the whole message of a gene. Thirdly, it is known that the initial contact of spliceosomal components is mediated by RNA-RNA interactions of U1 snRNP with the pre-mRNA at donor splice sites. This was further supported by 5'ss in vitro selection [5], where /GTRAGT-motifs were highly enriched, while it also been observed, e.g. in human chromosome 6, that only a small proportion of all naturally occurring 5'ss match perfectly to all 11nt of the 5'-end of U1 snRNA. Usage of the proximal-GT includes the second (distal) GT within the consensus donor splice site, thus allowing for more base-pairing between snRNA and pre-mRNA at the 5'ss, in contrast to the utilization of the distal-GT. Our results support the observation that less complementarity between the snRNA and 5'ss could be advantageous at low RNA concentrations [5], since splice forms that use the distal-GT are less frequently transcript-supported than shorter splice forms.

1. Holste, D., et al.,Nucleic Acids Res, 2006. 34: p. D56-62.
2. Hiller, M., et al., Nat Genet, 2004. 36(12): p. 1255-57.
3. Kent, J., UCSC. 2003.
4. Zhang, M. and W. Gish, Bioinformatics, 2006. 22(1): p. 13-20.
5. Lund, M. and J. Kjems, 8(2): p. 166-79.