Computational detection of cancer-specific splice sites

Anselmo A (1), Iacono M (1), Felice B (2), Guffanti A (2), Pesole G (3)

(1) Department of Biomolecular Sciences and Biotechnologies, University of Milan, Italy
(2) IFOM - The FIRC Institute of Molecular Oncology Foundation, Milan, Italy
(3) Department of Biochemistry and Molecular Biology, University of Bari, Italy

Motivation

Alternative splicing is a mechanism allowing the generation of multiple transcripts from a single gene. This can lead to the expression of structurally and functionally different proteins. Recent studies have shown that alternative splicing can also play an important role in modulating gene expression in different tissues or developmental stages. Moreover, some alternatively spliced isoforms are associated with diseases (different isoforms of some genes such as MLH1, APC and hSNF5 are expressed in normal and neoplastic tissues). For example, mutations in the splice sites of the MLH1 gene induce a double exon skipping that ultimately leads to nonpolyposis colorectal cancer (Venables 2004). Here we present a systematic search of tumour-specific splice sites in order to find new tumor markers.

Methods

We previously developed an algorithm and a web based tool (ASPIC) for the prediction of alternative splicing (Bonizzoni, Rizzi et al. 2005). The algorithm is based on the alignment of multiple transcript sequences against their correspondening genomic sequence. We are now further developing the ASPIC tool by introducing a new module that infers the library source of the ESTs supporting each predicted intron. Knowing the tissue source of each spliced EST, a suitable statistic (with an associated P value) indicates the possible library specificity of splice events (e.g. splice sites or introns). This approach can be used to identify splicing events which are specific to different tissue types or, for example, neoplastic vs "normal" alternative splicing patterns.

Results

We have tested our method on a gene set for which different isoforms are known to be expressed in normal and neoplastic conditions. For these genes, our approach provides good correspondence with experimental results. The systematic application of such method to a larger set of cancerrelated genes may lead to the identification of novel "cancer introns", that ultimately can be used to define novel cancer biomarkers.

Availability: The ASPIC software tool can be found in the website http://www.caspur.it/ASPIC/

Contact email: mailto:graziano.pesole@unimi.it