Role of alternative splicing in modulating protein-protein interactions

V. Bianchi✉, A. Colantoni, M. Helmer-Citterich, F. Ferrè

Department of Biology, University of Tor Vergata, Rome, Italy

Motivations

Alternative splicing (AS) permits the synthesis of multiple transcript variants from a single gene increasing the diversity of proteins encoded by a genome. Through the use of recent high-throughput sequencing technologies, it has been demonstrated that approximately 95% of multi-exon genes undergo AS in panels of human tissues [5], shaping the expressed transcriptome in various ways, from effectively turning off gene expression by the inclusion of early stop codons in the sequence, to subtle changes in protein function [1]. In addition, new data suggest that aberrant mRNAs generated through the AS machinery and their protein products have unique characteristics that confer new properties to cancer cells [2,3]. In this context it becomes crucial to link together heterogeneous data from different sources such as domain composition, protein structures, gene-interaction networks, in order to better understand AS mechanism and regulation, and its effects on protein products. We aim at a detailed analysis of how AS can modulate protein interactions by the differential expression of isoforms encoding or not for the interaction interfaces.

Methods

We analyzed RNA-seq data from two experiments, a panel of 9 human tissues [6] and a panel of 16 human tissues (Illumina BodyMap 2.0), the former downloaded from the Gene Expression Omnibus repository (GEO identifier: GSE12946), the latter from the ArrayExpress Archive database (ID: E-MTAB-513). We used the Tuxedo suite (bowtie, tophat, cufflinks) to map RNA-seq reads on the reference human genome (the hg19 assembly) up to two mismatches [4] and evaluate tissue-specific expression level for each isoform annotated in the Ensembl database (release 65). In a recent work of our group, we identified all human hetero-dimeric interactions solved by X-ray crystallography present in the Protein Data Bank, and whose residues are involved in the formation of the protein-protein interface using distance and energy criteria. Such residues were mapped to the hg19 human genome assembly to establish how many interface residues are part of each splicing isoform of a gene.

Results

We calculated that a considerable amount of human genes, about 24%, for which an interaction is known at the molecular level in the PDB, encode for at least one isoform where all interface residues are lost due to an alternative splicing event. We computed splice isoform expression levels in all tissues under analysis, and draw tissue-level interaction maps based on the expression of splice variants encoding or losing the interface residues, detecting that in many cases (and in fractions that differ in different tissues), even if two binding partner genes are actively expressed, the usage of splicing isoforms not encoding for the interface residues prevents the interaction. The distribution of the number of tissues that lose the interface highlights a large number of ubiquitous interfaces (214 pairs out of 620) and a smaller number of tissue-specific interfaces, which are lost in all except one tissue (32 pairs). The functional characterization made by enrichment in GO terms confirmed the characteristics of the ubiquitous and tissue-specific interactions, the former being mostly involved in general metabolic pathways, the latter involved in tissue-specific pathways. Our results indicate that AS is a powerful modulator of protein interactions, and that splicing isoform usage is finely tuned to allow or prevent specific interactions.

References

1. David CJ, Manley JL: The search for alternative splicing regulators: new approaches offer a path to a splicing code. Genes and development 2008, 22:279.
2. Fackenthal JD, Godley LA: Aberrant RNA splicing and its functional consequences in cancer cells. Disease models and mechanisms , 1:37-42.
3. Kim E, Goren A, Ast G: Insights into the connection between cancer and alternative splicing. Trends in genetics : TIG 2008, 24:7-10.
4. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 2009, 10:R25.

Genomics

5. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: Deep sur-veying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics 2008, 40:1413-5.

6. Wang ET, Sandberg R, Luo S, et al.: Alternative Isoform Regulation in Human Tissue Transcriptomes. Nature 2008, 456:470.

Genomics