ASPIC: a novel bioinformatics tool to investigate the impact of alternative splicing in the expansion of human transcriptome and proteome

ID - 196

Pesole Graziano[1], Mangiulli Marina[1], Picardi Ernesto[1], D'erchia Annamaria[1], Castrignano' Tiziana[2], Rizzi Raffaella[3], Anselmo Anna[4], Bonizzoni Paola[3], Mignone Flavio[4]

[1]University of Bari, Dipartimento di Biochimica e Biologia Molecolare 'E. Quagliariello', via Orabona, 4, Bari 70126, Italy

[2]Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, CASPUR, Rome, Italy

[3]DISCo, University of Milan Bicocca, via Bicocca degli Arcimboldi, 8, Milan, 20135, Italy

[4]University of Milan, Dipartimento di Scienze Biomolecolari e Biotecnologie, via Celoria 26, Milan 20133, Italy

Motivation

Alternative splicing is now emerging as a major mechanism contributing to the expansion of the transcriptome and proteome complexity of multicellular organisms.

The fact that a single gene locus may give rise to multiple mRNAs and protein isoforms, showing both major and subtle structural variations, is an exceptionally versatile tool in the optimization of the coding capacity of the eukaryotic genome.

The huge and continuously increasing number of genome and transcript sequences provides an essential information source for the computational detection of genes alternative splicing pattern. However, much of this information is not optimally or comprehensively used in gene annotation by current genome annotation pipelines.

Methods

We recently developed ASPIC, a novel algorithm for the investigation of alternative splicing of user submitted genes and a related web resource, based on comparative analysis of available transcript and genome data from a variety of species.

The ASPIC web resource provides graphical and tabular views of the splicing patterns of all full-length mRNA isoforms compatible with the detected splice sites of genes under investigation as well as relevant structural and functional annotation. The ASPIC web resource - available at http://www.caspur.it/ASPIC/ - is dynamically interconnected with the Ensembl and Unigene databases and also implements an upload facility.

Results

The ASPIC software has been applied to a set of 300 human genes included in the 44 regions included in the ENCODE project and accounting for 1% of the entire human genome. The results we obtained confirm the effectiveness and reliability of the ASPIC tool and show that about 90% of human genes undergo alternative splicing and that on average more than ten different transcript are generated by each gene. This implies that the human transcriptome and proteome complexity is expanded of at least one order of magnitude with respect to its gene complement.

Availability: http://www.caspur.it/ASPIC

Email: graziano.pesole@biologia.uniba.it