ACADEMIC TESTS
AT THE FACULTY OF MEDICINE OF THE UNIVERSITY OF LISBON
INSTITUTE OF ADVANCED TRAINING

PhD:
Biomedical Sciences

Name of Student:
Inês Maria de Stoop Camões Guerra Mollet

Subject of Thesis:
Genome Wide Mining of Alternative Splicing in Metazoan Model Organisms

Field:
Biomedical Sciences

Specialty:
Morphological Sciences

Date of Defence:
14th of April 2009

Classification:
Unanimously approved with Distinction and Praise

Jury:
The president of the jury was Professor Maria do Carmo Fonseca from the FMUL, and others present were: Professors Didier Auboeuf, from the Institut Universitaire d'Hématologie, Hôpital Saint Louis, Paris, França, Gil Ast, Tel Aviv University, Israel, Mário Ramirez, Margarida Gama-Carvalho, Luis Ferreira Moita and João Eurico da Fonseca from the University of Lisbon.

ABSTRACT

Background

Mining current mRNA and EST databases for novel alternatively spliced isoforms is of paramount importance for shedding light on the way in which the maturation of RNA is used to regulate gene expression. Preliminary observations revealed a tendency for greater amounts of potentially non-protein-coding alternative transcripts in human genes than in orthologous genes from other organisms. However, many of these isoforms did not appear in recently published alternative splicing databases on account of constraints imposed in the selection of transcripts. This prompted us to develop a less constrained database with the aim of contributing to the identification of the full repertoire of splice variants in the transcriptome of different organisms. Given that mechanisms of control of gene expression involving non-protein-coding splice variants have been described in a variety of genes, this information may be crucial to deciphering more intricate layers of gene regulation in complex organisms brought about by alternative splicing.

Description

An algorithm was developed to cluster mRNA and EST BLAT alignments to annotated gene regions. Consensus splice sites were the main requirement imposed on the selection of transcripts. The method was applied to thirteen model organisms. The alternative splicing information generated has been incorporated into a database with clear graphical displays representing the splicing patterns and is available from the ExonMine website (http://www.imm.fm.ul.pt/exonmine). It incorporates information on constitutive exons, poly-A signals, open reading frames and translation, expression specificity of any exon or splicing pattern relative to biological source of mRNA/EST, alternative splicing events and respective exon and junction sequences for microarray probe design. The ExonMine interface also provides several tools to support laboratory validation of splicing patterns.

Conclusions

ExonMine detects a higher percentage of spliced genes and isoforms than currently available alternative splicing databases. The analysis reveals a marked increase, in complex organisms, of splice variants with either retained introns or incorporating novel exons with no apparent protein-coding potential. About 18% of unannotated exons detected in ExonMine were found expressed in primary human cells using tiling arrays. Validation of some of these results for the U2AF family of splicing factors was successfully performed in collaboration with members of the lab revealing primate specific transcripts and an alternatively spliced transcript carrying a microRNA. The database was also successfully used for genome wide analysis of sequence elements involved in the regulation of alternative splicing and for custom alternative splicing microarray design. Matching of ExonMine data to a commercial exon microarray platform covering the majority of human exons was also performed and will assist in large-scale analysis of alternative splicing data. The algorithm developed also provides for easy updatability, taking only 48 hours to generate data for the whole human genome and far less time for less complex organisms. In conclusion, ExonMine represents a new useful resource for future research on alternative splicing and gene regulation.