

SpliceAid-F: a database of human splicing factors and their RNA binding sites

M. Giulietti¹, F. Piva², M. D'Antonio³, P. D'Onorio De Meo³, T. Castrignanò³, G. Pavesi⁴, G. Pesole⁵✉

¹Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Bari, Italy

²Department of Specialized Clinical Sciences and Odontostomatology, Polytechnic University of Marche, Ancona, Italy

³Consorzio per le Applicazioni di Supercalcolo per Università e Ricerca, Rome, Italy

⁴Department of Biomolecular Sciences and Biotechnology, University of Milan, Milan, Italy

⁵CNR - Institute of Biomembrane and Bioenergetics, Bari, Italy

Motivations

The pre-mRNA sequences of eukaryotes harbour multiple information layers in addition to the one specifying the amino acid sequence. Indeed, specific signals regulating the splicing process, RNA folding, capping, polyadenylation, stability and nuclear export, overlap each other and with the coding information. Mutagenesis experiments have highlighted that splicing regulatory elements are scattered in the entire pre-mRNA sequence and that all nucleotide positions are potentially involved in the generation of the specific splicing pattern through the specific interaction with trans-acting RNA-binding proteins. In particular, the identification of cis-regulatory motifs in exonic regions, such as exonic splicing enhancers (ESEs) or silencers (ESSs), superimposed on the coding sequence of the gene can be driven by the observation of purifying selection occurring at synonymous codons [5]. The coordinated binding of combinations of regulatory proteins to their binding sites modulate the expression of specific transcript isoforms in a cell/tissue type-, development stage-, disease- and/or other condition-specific manners and may also promote or repress the formation of the spliceosome, the large (~60S) RNA-protein machinery that catalyzes intron removal. A growing list of mammalian protein factors involved in splicing regulation and their target sites in the pre-mRNA has been reported in the literature [2]. In order to establish a curated and retrievable repository of splicing regulatory factors and target sites for human genes we have recently created SpliceAid [4] that can also be used to find putative regulatory motifs in user submitted sequences. As a further evolution of SpliceAid, we present here SpliceAid-F, a database compilation of splicing regulatory factors and their experimentally validated target RNAs extracted from an exhaustive hand-curated literature search.

Methods

For each known splicing factor, cross-links to gene (NCBI Entrez) and protein (Uniprot) IDs, as well as information on the structure of the RNA binding domain, the protein-defect associated disease (MIM), and the interacting proteins (from STRING and IntAct) have been collected. Moreover, we have extracted from the literature the relevant information on the genome coordinates of RNA binding sites (or experimentally validated non-binding sites), type of binding assay, gene information and the context-specific splicing effects of splicing factor binding in terms of exonization or intronization. Furthermore, binding site information have been also reported in the transcript view of ASPicDB [1,3].

Results

SpliceAid-F, currently collects 68 records for splicing regulatory factors and 2489 records for their related binding sites and can be retrieved through a web interface. SpliceAid-F collects in a unique resource heterogeneous information about splicing regulatory proteins, related RNA binding sites and context-specific activity of their interaction. Our database may be a useful resource to retrieve and visualize experimentally known splicing factor binding sites in a gene and to investigate their context where additional binding sites may establish a potential competition or co-regulation. All these information may help to explain the observed splicing pattern as well as the effect of mutations, which if located in functional regulatory elements may generate an aberrant and possibly pathological splicing pattern. This resource can also be useful to develop a new generation of prediction software taking into account all the splicing regulatory element and so allowing to attain a more accurate prediction of splicing patterns.

Availability

<http://www.caspur.it/SpliceAidF>

References

1. Castrignano T, D'Antonio M, Anselmo A, Carrabino D, D'Onorio De Meo A et al. (2008) ASPicDB: a database resource for alternative splicing analysis. *Bioinformatics* 24(10):1300-1304.
2. Gabut M, Chaudhry S, Blencowe BJ (2008) SnapShot: The splicing regulatory machinery. *Cell* 133(1): 192e191.
3. Martelli PL, D'Antonio M, Bonizzoni P, Castrignano T, D'Erchia AM, D'Onorio De Meo P et al. (2011) ASPicDB: a database of annotated transcript and protein variants generated by alternative splicing. *Nucleic Acids Res.* 39(Database issue):D80-85.
4. Piva F, Giulietti M, Nocchi L, Principato G (2009) SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics* 25(9):1211-1213.
5. Schattner P, Diekhans M (2006) Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res* 34(6):1700-1710