ASPIC: a Novel Method to Predict Alternative Splicing

Paola Bonizzoni [(1)] , Riccardo Dondi [(1)] , Raffaella Rizzi [(1)], and Graziano Pesole [(2)]

[(1)] Dipartimento di Informatica Sistemistica e Comunicazione, University of Milan Bicocca, via Bicocca degli Arcimboldi 8, 20135 Milan - Italy
{bonizzoni,dondi,rizzi}@disco.unimib.it
[(2)] Dipartimento di Scienze Biomolecolari e Biotecnologie, University of Milan, via Celoria 26, 20133 Milan - Italy
graziano.pesole@unimi.it

Keywords. Alternative splicing, ESTs alignments, Computational methods

Introduction

In this paper a new method for detecting splicing sites is proposed. It is based on a combined analysis of all available transcript data in order to produce all transcript alignments to the genomic sequence. The algorithm requires that all transcript-genome alignments are fully compatible with a plausible common exon-intron structure within the genomic sequence. The algorithm was implemented in the ASPIC (Alternative Splicing PredICtion) software.

Some Details

1. Alternative Splicing Prediction: the Method

The problem of identifying the correct exon-intron structure, is complicated by the fact that a genomic sequence may be alignable to the same EST sequence in a number of equally probable ways. For this reason, we designed a method, for obtaining the exon-intron structure of a genomic sequence, by aligning a collection of transcript data (mostly ESTs) so that common regions of different ESTs are aligned to the same region of the gene. The criterion used minimizes the number of exons in the exon factorization and EST alignments, compatible with a gene structure under given parameters and optimization criteria, are produced. The parameters describe properties of the exon-intron structure to which all ESTs alignments must agree and they are: minimum exon length, maximum number of exons allowed along the genomic sequence, error parameters for evaluating the quality of the EST alignment. The optimization criteria were mainly designed in order to increase the quality of the alignments near the intron boundaries. The intron regions are validated, where it is possible, by the *gt-ag* rule. When an EST alignment produces an incompatible exon, then the current EST factorization (or previous ESTs factorizations) is recomputed, by backtracking, in order to obtain a compatible EST alignment. For more details we refer the reader to [1].

2. ASPIC Software

The method described above has been implemented in C and we call it ASPIC (Alternative Splicing PredICtion). It can be accessed at http://aspic.bio.disco.unimib.it/. A basic searching form permits to input an official gene name for the genomic sequence (e.g. ABCB10) and a Unigene cluster identifier for the ESTs collection (e.g. Hs.1710) and ASPIC runs under standard parameters. An advanced searching form permits to specify also the running parameters. The genomic sequences are retrieved by Ensembl database. The program provides a table view and a graphical view describing respectively the intron list (with confirming ESTs and links to the alignments near the intron boundaries) and the exon-intron structure of the input gene. ASPIC method has been tested on 80 Unigene clusters on Chromosome 1 and its accuracy has been analyzed by comparison with the ASAP database [4] accessible at http://www.bioinformatics.ucla.edu/ASAP. In almost all cases ASPIC

confirmed ASAP data, while providing additional novel splice sites. A total of 161 splices were uniquely found by ASPIC whereas 25 splices were detected by ASAP but not by ASPIC. ASPIC and ASAP detected 921 and 785 splices respectively.

3. Developments

We plan to integrate ASPIC representation, in more detail, with all isoforms and all patterns of alternative splicing that may occur in the gene: exon skipping, mutually exclusive exon, competing 5'-3' ends and intron retention.

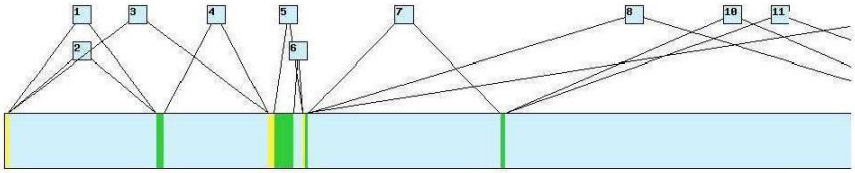

Fig. 1 Snapshot of a portion of the exon-intron strucure of gene HNRPR (introns are in blue, exons are in yellow, alternative spliced regions are in green) provided by ASPIC

Intron 2 [5101]-[8293]

```
CCCCGCTCTGCCCTGGTGAGTCTCGCGCCGGCCCG----TCCTGAATTTTTTCCAGCAGCATAATAAAATGGCT
CCCCGCTCTGCCCTG       75          76       CATAATAAAATGGCT      gnl|UG| Hs#S14790655
CCCCGCTCTGCCCTG      108         109       CATAATAAAATGGCT      gnl|UG| Hs#S15980077
CCCCGCTCTGCCCTG      111         112       CATAATAAAATGGCT      gnl|UG| Hs#S16104631
CCCCGCTCTGCCCTG       53          54       CATAATAAAATGGCT      gnl|UG| Hs#S1711540
CCCCGCTCTGCCCTG       81          82       CATAATAAAATGGCT      gnl|UG| Hs#S1729818
CCCCGCTCTGCCCTG       74          75       CATAATAAAATGGCT      gnl|UG| Hs#S2077579
CCCCGCTCTGCCCTG      134         135       CATAATAAAATGGCT      gnl|UG| Hs#S2182336
CCCCGCTCTGCCCTG      100         101       CATAATAAAATGGCT      gnl|UG| Hs#S3333678
CCCCGCTCTGCCCTG      138         139       CATAATAAAATGGCT      gnl|UG| Hs#S3605033
CCCCGCTCTGCCCTG       55          56       CATAATAAAATGGCT      gnl|UG| Hs#S3905706
CCCCGCTCTTCCTTG       77          78       TTTAATAAAATGGCT      gnl|UG| Hs#S4005136
CCCCGCTCTGCCCTG      115         116       CATAATAAAATGGCT      gnl|UG| Hs#S4229059
CCCCGCTCTGCCCTG       52          53       CATAATAAAATGGCT      gnl|UG| Hs#S4838983
CCCCGCTCTGCCCTG       83          84       CATAATAAAATGGCT      gnl|UG| Hs#S5959336
CCCCGCTCTGCCCTG       52          53       CATAATAAAATGGCT      gnl|UG| Hs#S952995
```

Fig. 2 Example of ESTs alignments near intron boundaries (intron 2 of gene HNRPR) provided by ASPIC

Acknowledgements

This work was supported by FIRB project "Bioinformatica per la Genomica e la Proteomica" (Ministero dell'Istruzione e Ricerca Scientifica, Italy) and by Telethon.

References

[1] P. Bonizzoni, G. Pesole and R. Rizzi, A method to detect gene structure and alternative splice sites by agreeing ESTs to a genomic sequence, in G. Benson and R. Page (eds.), *Lecture Notes in Bioinformatics*, *Proc. WABI 2003,* 2812:63-77, 2003

[2] J.F. Caceres and A.R. Kornblihtt, Alternative splicing: multiple control mechanism and involvment in human disease. *Trends Genet*, 18(4):186-193, 2002

[3] B.R. Graveley, Alternative splicing: increasing diversity in the proteomic world. *Trends Genet*, 17(2):100-107, 2001

[4] C. Lee, L Atanelov, B. Modrek and Y. Xing, ASAP: the alternative splicing annotation project. *Nucleic Acids Research*, 31(1):101-105, 2003