Improved prediction of conserved exon skipping using Bayesian Networks

Rileen Sinha[1], Ulrike Gausmann[1], Michael Hiller[2,] Rainer Pudimat[2,]
Stefan Schuster[3], Matthias Platzer[1], Rolf Backofen[2]

[1] Leibniz Institute for Age Research - Fritz Lipmann Institute, Genome Analysis,.
Beutenbergstrasse 11, 07745 Jena, Germany
[2]Albert-Ludwigs-University, Institute of Computer Science, Bioinformatics Group,
Georges-Koehler-Allee 106, 79110 Freiburg, Germany
[3]Friedrich-Schiller-University, Faculty of Biology and Pharmacy, Department of Bioinformatics
Ernst-Abbe-Platz 2, 07743 Jena, Germany

Alternative splicing is now well established as a widespread phenomenon in higher eukaryotes, and a major contributor to proteome diversity. Over half of the multiexonic human genes are believed to have splice variants Large-scale detection of alternative splicing usually involves expressed sequence tags (ESTs) or microarray analysis. However, due to various sampling biases, not all alternative splicing events can be detected by these methods. Moreover, nowadays genomic sequence data is being churned out at a much faster rate than transcript data, that is, several genomes do not have a very high amount of transcript data. This situation is likely to continue for the foreseeable future. Thus, there is a need for independent methods of detecting alternative splicing. Previous studies have shown that discriminative features can be used to distinguish alternatively splice exons from constitutively spliced ones. We used Bayesian Networks, a state of the art machine learning tool, to accurately distinguish conserved alternative exons from conserved constitutive ones. Using a combination of previously described features and novel ones, we were able to achieve a classification performance competitive with the state of the art from the literature (Dror et al, Bioinformatics. 2005 Apr 1;21(7):897-90). Future plans include prediction without using conservation based features, prediction of species-specific alternative splicing, and prediction of alternative splicing in human-specific exons.