Poster H-44

"Co-transcriptional" prediction of splice sites and analysis of spliceosomal introns in human pre-mRNAs

Authors:
Kai Wang (*Center for Biological Sequence Analysis, Biocentrum, Technical University of Denmark*)
Rasmus Wernersson (*Center for Biological Sequence Analysis, Biocentrum, Technical University of Denmark*)
Søren Brunak (*Center for Biological Sequence Analysis, Biocentrum, Technical University of Denmark*)

Short Abstract: We developed a new prediction program to detect splice site, including alternative splice sites, by combining neural networks. We presented a detailed review of various anticipated and novel statistical features of exon-intron structures, splicing enhancers and silencer, ultra-conserved elements, and pseudogenes in human to enhance our understanding of the many intriguing questions in different splicing scenarios. We also investigated the spliceosomal intron evolution on co-transcriptional splicing processes.

Long Abstract:
"Co-transcriptional" prediction of splice sites and analysis of spliceosomal introns in human pre-mRNAs Kai Wang, Rasmus Wernersson and Søren Brunak Center for Biological Sequence Analysis, Biocentrum, Building 208, Technical University of Denmark, DK-2800 Lyngby, Denmark Alternative splicing has been considered as a major actor to widely yield rearranged genes with altered functions based on given sequences. How to detect the incredible abundance of alternatively spliced RNAs, and how to characterize splicing enhancers and silencers remain challenges. We developed a new prediction program to detect splice site, including alternative splice sites, by combining neural networks. We presented a detailed review of various anticipated and novel statistical features of exon-intron structures, splicing enhancers and silencer, ultra-conserved elements, and pseudogenes in human to enhance our understanding of the many intriguing questions in different splicing scenarios. By comparing with NetAspGene, a specific splice site predictor on Aspergillus, we also investigated the spliceosomal intron evolution on co-transcriptional splicing processes. The key points to identify protein coding genes in genomics DNA de novo are to detect splice site signals and locate potential coding regions. Artificial neural networks (ANNs) have been used successfully on prediction for a long time. ANNs are able to exhibit generalization beyond the training gene sequences and detect splicing enhancers and silencers even without any prior knowledge about them. Performance of neural network models largely relies on their appropriate architecture and robust postprocessing of the data. In our model, several neural network predictors were selected and combined into one single ensemble program to detect coding sequences, splice sites, alternative splice sites, and reading frames respectively. Then, the prediction of coding sequences regulated a cutoff level for splice site prediction of local neural networks in a certain region around predicted splice sites. Finally, by postprocessing methods after predictions, we discarded wrong predicted splice sites in uniformly predicted regions and selected between nearby

predictions. We significantly increased the overall performance of the prediction by combination. To distinguish bona fide splice sites from the numerous putative false splice sites is critical for spliceosomal intron recognition during splicing. Furthermore, gene expression is affected at all stages of the processes on the transcriptional, post-transcriptional and post-translational levels. We used an algorithm to analyze the snRNA:pre-mRNA interactions to illustrate the fidelity of splicing, and utilized information on co-transcription and translatability to integrate the prediction, including the post-transcriptional regulation process, such as nonsense-mediated mRNA decay (NMD), nonsense-associated altered splicing (NAS) and suppression of splicing (SOS).