Effect of alternative splicing on structure and function of mouse transcription factors.

Bahar Taneri[1], Ben Snyder[2], Terry Gaasterland[3].

Keywords: alternative splicing, transcription factor, proteome

Abstract

Analyzing proteins in the context of all available genome and transcript sequence data has the potential to reveal functional properties not accessible through protein sequence analysis alone. We are studying the effect of alternative splicing on mouse proteins and specifically on transcription factors. We hypothesize that by creating a protein structure alteration, alternative splicing could be a critical determining factor for the specific DNA binding sites and cofactors that interact with a given transcription factor. We further hypothesize that expression data will indicate tissue specific control of changes in protein structure and function due to alternative splicing. We use MouSDB3, a database of splice variants in the mouse transcriptome (http://genomes.rockefeller.edu/MouSDB3) to study these phenomena.[1] Protein domain organizations for different splice forms are determined by using SMART (Simple Modular Architecture Research Tool- http://smart.embl-heidelberg.de).[2] Initial analyses revealed that 62% of the transcription factor loci in MouSDB3 have variant exons, compared to 29% of all loci. These variant transcription factor loci contain a total of 325 facultative exons, which are excluded in some transcripts and included in others. 24% of these facultative exons are in-frame, i.e their nucleotide number is in multiples of three, they do not introduce a stop codon when skipped and the exon starts at the first base of a codon. When excluded, in-frame facultative exons alter the domain architecture of the protein 81% of the time, as computed by SMART. 67% of these in-frame facultative exons are either fully or partially within the coding regions for motifs that are important in transcription factor function. These include helix-loop-helix, leucine zipper and homeobox domains. Our integrated genomic and proteomic approach addresses the general question of how alternative splicing affects the proteome, and gives insight into control of transcription.

[1] Laboratory of Computational Genomics, The Rockefeller University, 1230 York Avenue, New Yok, New York, 10021. E-mail: bahar@genomes.rockefeller.edu

[2] Laboratory of Computational Genomics, The Rockefeller University, 1230 York Avenue, New Yok, New York, 10021. E-mail: ben@genomes.rockefeller.edu

[3] Laboratory of Computational Genomics, The Rockefeller University, 1230 York Avenue. New Yok, New York 10021. E-mail: gaasterl@genomes.rockefeller.edu

References

[1] Zavolan, M., van Nimwegen, E. and Gaasterland, T. 2002. Splice variation in mouse full-length cDNAs identified by mapping to the mouse genome. *Genome Research* 9:1377-1385.

[2] Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P. and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Research* 30(1):242-4.