Predictive identification of alternative exons conserved in human and mouse

Gene Yeo, Christopher Burge and Tomaso Poggio, CBCL

The Problem: Alternative pre-messenger RNA splicing affects a majority of human genes and plays important roles in development and disease. Alternative splicing (AS) events conserved since the divergence of human and mouse are likely of primary biological importance, but relatively few such events are known.

Previous work: In our previous work, we described sequence features that distinguish exons subject to evolutionarily conserved AS, which we call ‘alternative conserved exons’ (ACEs) from other orthologous human/mouse exons, and integrated these features into an exon classification algorithm, ACEScan. Genome-wide analysis of annotated orthologous human-mouse exon pairs identified 2,000 predicted ACEs. Alternative splicing was verified in both human and mouse tissues using an RT-PCR-sequencing protocol for 21 of 30 (70%) predicted ACEs tested, supporting the validity of a majority of ACEScan predictions. By contrast, AS was observed in mouse tissues for only 2 of 15 (13%) tested exons which had EST or cDNA evidence of AS in human but were not predicted ACEs, and was never observed for eleven negative control exons in human or mouse tissues. Predicted ACEs were much more likely to preserve reading frame, and less likely to disrupt protein domains than other AS events, and were enriched in genes expressed in the brain and in genes involved in transcriptional regulation, RNA processing and development. Our results also imply that the vast majority of AS events represented in the human EST databases are not conserved in mouse, and therefore may represent aberrant, disease- or allele-specific, or highly lineage-restricted splicing events.

Motivation: Our current motivation is in trying to understand which trans-factors will interact with these predicted ACEs.

Approach: We are building motif searches based on comparative sequence analyses within our predicted ACEs to predict co-regulation of subsets of ACEs by different trans-factors.

Research Support: This report describes research done at the Center for Biological & Computational Learning, which is in the McGovern Institute for Brain Research at MIT, as well as in the Dept. of Brain & Cognitive Sciences, and which is affiliated with the Computer Sciences & Artificial Intelligence Laboratory (CSAIL).
