

Neighborhood-Inference: Predicting Splicing Regulatory Elements Using Functionally Antagonistic Groups

Michael B. Stadler¹, Gene W. Yeo^{1,2}, Noam Shomron¹,
Christopher B. Burge¹

Keywords: sequence analysis, sequence motif, alternative splicing, splicing regulatory elements

Introduction

The precise location of exons in higher eukaryotes is partially determined by the presence of canonical splice signals, such as the 5' splice site, the 3' splice site and the branch signal. The degeneracy of the canonical signals necessitates accompanying auxiliary splicing cis-regulatory elements, such as exonic splicing enhancers and silencers (ESEs/ESSs) [5], which act as binding sites for protein trans-factors, to promote or repress exon inclusion. For example, serine-arginine rich proteins bind to ESEs [2], and recruit the spliceosomal machinery to promote splice site selection. In contrast, ESSs presumably interact with another class of trans-factors to suppress the recognition of splice sites, thereby 'masking' decoy exons in non-coding regions [3]. In addition to their roles in constitutive splicing, both splicing enhancers and silencers contribute significantly to the regulation of alternative splicing in a variety of biological scenarios [1], and their importance is underscored by the fact that genomic mutations that fall in splicing cis-elements have led to numerous splicing-associated genetic diseases [6].

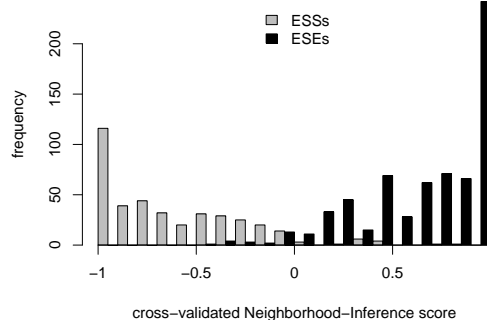


Figure 1: Histogram of Neighborhood-Inference scores obtained by tenfold crossvalidation of trusted ESE (666) and ESS (386) hexanucleotides.

Method and Results

Because of their importance in both constitutive and alternative splicing, as well as splicing-associated diseases, recent efforts have been undertaken to systematically identify mammalian ESEs [4, 8] and ESSs [8, 7]. Here, we developed a new algorithm, *Neighborhood-Inference*, that utilizes sets of trusted ESE and functionally antagonistic ESS sequences to predict new ESEs and ESSs, based on the premise that ESE and ESS motifs are compact in sequence space. The Neighborhood-Inference score for a test sequence is mainly dependent on the ratio between trusted ESEs and ESSs in its neighborhood, as defined by Hamming

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, MA. E-mail: MBS stadler@mit.edu, NS nshomron@mit.edu, CBB cburge@mit.edu

²Current address: Laboratory of Genetics, The Salk Institute, La Jolla, CA. E-mail: geneyeo@salk.edu

distance, with higher weights applied to near neighbors compared to distant neighbors. In effect, our method predicts that oligonucleotides whose sequence neighbors are predominantly of one class (e.g., ESEs) are likely to belong to the same class. Cross-validation tests using known elements support the validity of this method (Figure 1). Most true ESSs obtain Neighborhood-Inference scores below zero, and most true ESEs scores above zero, with only minor fractions of the two sets obtaining overlapping scores. A cohort of newly predicted cis-elements is being experimentally verified in a splicing reporter mini-gene construct in vivo.

Outlook

Our approach allows expansion and refinement of the class boundaries of known splicing cis-elements and could potentially be generalized and applied to other classes of antagonistic regulatory elements.

References

1. D.L. Black. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, 72:291–336, 2003.
2. B.J. Blencowe. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci*, 25(3):106–10, 2000.
3. F. Del Gatto-Konczak, M. Olive, M.C. Gesnel, and R. Breathnach. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol Cell Biol*, 19(1):251–60, 1999.
4. W.G. Fairbrother, R.F. Yeh, P.A. Sharp, and C.B. Burge. Predictive identification of exonic splicing enhancers in human genes. *Science*, 297(5583):1007–13, 2002.
5. L.P. Lim and C.B. Burge. A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A*, 98(20):11193–8, 2001.
6. H.X. Liu, L. Cartegni, M.Q. Zhang, and A.R. Krainer. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet*, 27(1):55–8, 2001.
7. Z. Wang, M.E. Rolish, G. Yeo, V. Tung, M. Mawson, and C.B. Burge. Systematic identification and analysis of exonic splicing silencers. *Cell*, 119(6):831–45, 2004.
8. X.H. Zhang and L.A. Chasin. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*, 18(11):1241–50, 2004.