

Codon bias uncovers splicing regulatory sequences

Noam Shomron¹, Michael B. Stadler², Christopher B. Burge³

Keywords: codon bias, pre-mRNA splicing, alternative splicing, cis-regulatory sequences

1 Introduction

During pre-mRNA splicing non-coding introns are removed and coding exons are ligated to form a mature mRNA. In a given pre-mRNA transcript, exon choice, either constitutive or alternative, is determined by the concentration of different nuclear proteins that bind cis-regulatory sequences and these can be altered in the course of different cellular events [1]. Canonical pre-mRNA sequences that are essential for splicing are the 5' and 3' splice sites, the polypyrimidine tract and the branch-site. In addition, exonic and intronic splicing enhancers (ESE/ISE) and silencers (ESS/ISS) play a vital role in this process [2] which has led to the efforts in identifying such elements. Methodical screens employing both experimental and computational methods have been undertaken for large-scale detection of these sequences [4,8,10]. However, the complete list of splicing regulatory sequences is estimated to be far from completion.

Codon bias, the preference for some of the synonymous codons encoding the same amino acid, is correlated with many genomic factors such as gene expression level, G/C content and gene length [7]. In humans, codon preference was shown to vary with distance from the splice sites suggesting that selection at the pre-mRNA level influences codon usage in humans [9].

2 Methods and Results

The activity of some regulatory elements is increased with proximity to splice sites [3,5]. Our working hypothesis is based on the assumption that synonymous codon usage might reflect the presence of splicing regulatory sequences in the proximity of splice sites. On the contrary, this bias would be less pronounced with increasing distance from the splice sites. We thus recorded synonymous codon frequency for each of the 61 sense codons as a function of distance from the nearest splice site in human coding exons. Using these frequencies we identified the codons showing a strong positional bias and searched for candidate regulatory motifs flanking these codons. The resulting motifs were clustered and aligned, and further investigated and confirmed using a splicing minigene reporter system.

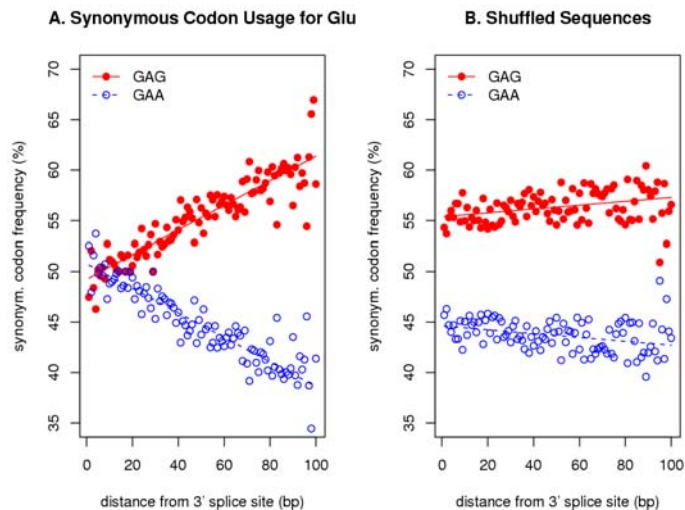
¹⁻³ Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

¹ E-mail: nshomron@mit.edu

² E-mail: stadler@mit.edu

³ E-mail: cburge@mit.edu

Synonymous codon usage for Glutamic acid (GAG; filled circles, GAA; opened circles) relative to its position from the 3' splice site. Synonymous codon usage was calculated for real exons (left panel) and for shuffled exons (right panel). Shuffled exons were generated using 'CodonShuffle' which permutes synonymous codons while preserving the encoded amino acid sequence and nucleotide frequencies [6].



3 Outlook

To understand the splicing mechanism it is necessary to identify all of the features that govern its regulation. Our findings indicate that natural selection acts most strongly on codon bias to preserve splicing regulatory sequences in the vicinity of splice sites. We demonstrate that codon bias can be used to infer splicing regulatory motifs and we provide evidence that selection at the RNA transcript level influences codon usage in the context of splicing regulation.

4 References

- [1] Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem.* 72:291-336.
- [2] Blencowe, B.J. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci.* 25:106-10.
- [3] Fairbrother, W.G., Holste, D., Burge, C.B. and Sharp, P.A. 2004. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* 2:E268.
- [4] Fairbrother, W.G., Yeh, R.F., Sharp, P.A. and Burge, C.B. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science.* 297:1007-13.
- [5] Graveley, B.R., Hertel, K.J. and Maniatis, T. 1998. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.* 17:6747-6756.
- [6] Katz, L. and Burge, C.B. 2003. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.* 13:2042-2051.
- [7] Lavner, Y. and Kotlar, D. 2005. Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene.* Dec 24
- [8] Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M. and Burge, C.B. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell.* 119:831-45.
- [9] Willie, E. and Majewski, J. 2004. Evidence for codon bias selection at the pre-mRNA level in eukaryotes. *Trends Genet.* 20:534-8.
- [10] Zhang, X.H. and Chasin, L.A. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* 18:1241-50.