Comprehensive Study Illuminates Regulation and Significance of NAGNAG Splicing

February 20, 2012

ME Arnegard

The diversity of proteins encoded by a genome is expanded by alternative splicing, in which introns of variable size are removed from the primary transcript and exons are stitched back together in assorted ways. A common mode of alternative splicing is exon skipping, which results in entire chunks of a protein being excluded or included in the final product. Use of alternate 3' or 5' splice sites (separated by more than three nucleotides) causes longer or shorter versions of a particular exon to be included in the protein. The adaptive significance and cis-regulation of these textbook splicing modes are well understood compared to other forms of splicing. Examples abound of how these processes allow protein isoform expression to be tuned to different developmental stages or tissues. Moreover, mutations affecting these forms of splicing play known roles in many human diseases, including cancer.

A less well understood mode of splicing involves introns ending in NAGNAG, where 'N' stands for any nucleotide. Either NAG can act as a 3' splice site, resulting in proteins that differ by only a single amino acid. Genome-wide analyses of NAGNAG splicing have been hindered by insufficient depths of EST databases and the difficulty of distinguishing such similar cDNAs using hybridization arrays. Despite these limitations, NAGNAG splicing is known to affect thousands of human genes, yet little is known about its regulation and adaptive significance. Some scientists have even suggested that differential NAGNAG splicing occurs randomly and has no relevance to an organism's fitness.

Dr. Robert Bradley (Public Health Sciences and Basic Sciences Divisions) and several colleagues recently employed deep RNA-Seq data collected from 24 distinct human or mouse tissues to analyze the regulation and evolution of NAGNAG splicing. They show that a quarter of all NAGNAGs undergo tissue-specific regulation, and that the evolutionary conservation of NAGNAG splicing is positively related to the strength of differential regulation across tissues. These results strongly suggest that NAGNAG splicing is physiologically important to survival and, therefore, selectively maintained by evolution. Based on their proximity, it has been difficult to imagine how NAGNAG splice sites could be regulated by a cell's standard splicing machinery. Nevertheless, Bradley *et al*. pinpoint features of the upstream intron that shift splicing toward either the proximal or distal NAG

(see fig.), and they experimentally confirm these findings in cell culture. Finally, the authors show that NAGNAG splicing is associated with accelerated protein evolution at exon boundaries, with a bias towards some amino acids over others. This impressively comprehensive computational study makes a clear case for the importance of NAGNAG splicing as a generator of proteome diversity, at both developmental and evolutionary timescales.

Bradley RK, Merkin J, Lambert NJ, Burge CB. 2012. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS Biology* 10:e1001229.

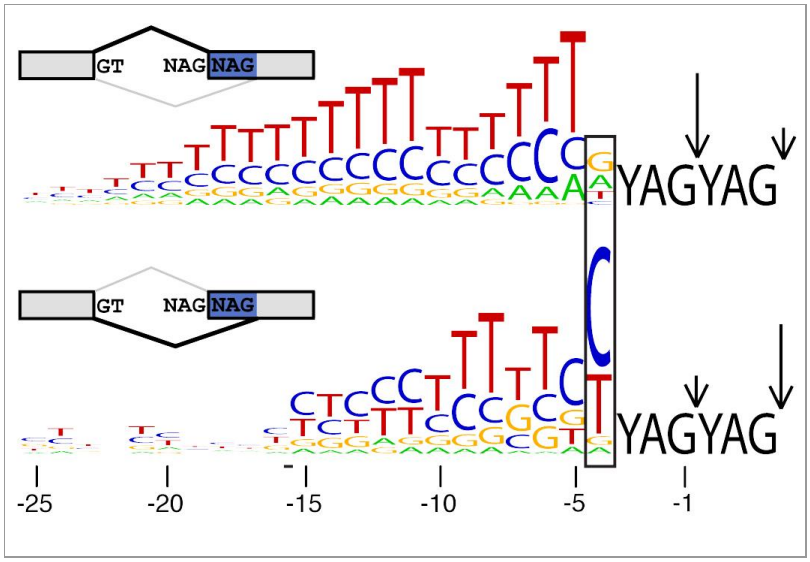

Image courtesy of Robert Bradley

Example of one of many novel insights provided by Bradley et al. (2012). To help illuminate upstream sequence features that affect NAGNAG isoform choice, the authors consider here a large subclass of NAGNAGs with favored pyrimidines (either C or T) at both -3 base positions. Such NAGNAGs are therefore called YAGYAGs. As depicted schematically by the top and bottom splicing diagrams inset at the left (and also by the size of arrows at the right), either the proximal or distal splice site can be favored. Frequency diagrams of the upstream nucleotide sequence, with nucleotide frequency proportional to text size, clearly show that distinct sequence features favor the alternate YAGYAG splice sites.