Chapter title: Alternative splicing in humans

Authors: John G. Conboy[1] and Inna Dubchak[2]

Life Sciences Division and Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 95720
Email: jgconboy@lbl.gov, ildubchak@lbl.gov.

Key words: alternative splicing, regulation, splicing enhancer, silence, disease

Length: 1985 words (from abstract to end of references)

Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics
Part 4. Bioinformatics
4.2. Gene Finding and Gene Structure

DOI: 10.1002/047001153X.g402305
Article Online Posting Date: April 15, 2005

Abstract

Alternative pre-mRNA splicing is the mechanism by which a cell fine-tunes gene expression, by deciding what discrete segments of a gene's total coding information should be selected for expression in a specific cell type. The phenomenon of multiple alternative splicing pathways in the human genome, and the opportunity to regulate these pathways during development, allows a relatively small number of genes to encode a complex proteome in tissue-specific patterns. Individual splicing decisions are controlled by interplay among splicing regulatory proteins, acting at specific enhancer and silencer sequences in the pre-mRNA, and the catalytic machinery known as the spliceosome. Proper splicing regulation is a critical aspect of gene expression during normal human development, and derangements in this process are responsible for many human diseases. Recent advances in genome sequencing, gene expression profiling, bioinformatics, and biochemical analysis of the splicing machinery give hope that the rules by which splicing regulatory networks govern alternative splicing in humans will be deduced.

Introduction

In the human genome, the protein coding sequence of most genes is organized into discrete blocks of sequence, the exons, which are interrupted by noncoding sequences called introns. Exons and introns are transcribed together as long nuclear pre-mRNAs that must be spliced so as to excise the introns and ligate the exons, forming mature mRNAs that are exported to the cytoplasm for translation. For small genes this process of RNA splicing is often hard-wired so that the same array of exons is always joined to create a single unique mRNA. Many large eukaryotic genes, in contrast, exhibit alternative competing pathways for splicing, generating multiple mRNA isoforms that differ with respect to inclusion or exclusion of selected exons or parts of exons. Moreover, because alternative exons can encode unique peptides or introduce new translation initiation or termination signals, any aspect of protein structure and function may be affected, having important physiological impacts on virtually every major cellular process. In recognition of the critical role of splicing as a key mediator of gene expression, its discovery was celebrated with the 1993 Nobel Prize award in Physiology or Medicine to Richard Roberts and Phillip Sharp.

Others have reviewed the structure and function of the eukaryotic spliceosomal machinery and the biochemical mechanisms involved in catalysis of the splicing reactions (Black 2003). Here, we will focus briefly on the importance of alternative splicing in the human genome. It is now known that a majority of human genes are alternatively spliced with respect to one or more exons (Modrek, 2002), and that this

phenomenon allows a relatively small genome to encode an enormously complex proteome. A number of efforts are under way to catalog and classify human alternative splicing events on a genome-wide basis, and to compare alternative splicing among genomes of other species (e.g., (Lee, 2003;Nurtdinov, 2003;Pospisil, 2004;Thanaraj, 2004). Some splicing events are poorly conserved phylogenetically and may represent experiments of nature, while others are very highly conserved through evolution and likely regulate important cellular functions. Many conserved splicing events are tightly regulated in cell type- and differentiation stage-specific patterns, so that individual exons or sets of exons can be switched on or off in response to appropriate environmental signals. Moreover, regulation of alternative splicing is functionally coupled to other RNA processing events including transcription initiation and polyadenylation (Hirose, 2000;Maniatis, 2002). A major challenge in the splicing field remains identification and characterization of the regulatory networks that control alternative splicing switches and thereby play important roles in defining the specialized properties of differentiated cells.

Main text

Cis-acting regulatory sequences in the genome. The decision to activate splicing of an alternative exon requires mechanism(s) for recruiting spliceosomal machinery to the appropriate splice sites, a daunting task given the observed degeneracy of splice site consensus sequences and the abundance of cryptic splice sites in the genome. The cellular splicing machinery meets this challenge by employing not only the splice sites themselves, but also accessory regulatory elements in the exon and/or its flanking introns, for the process of "exon definition" (Robberson, 1990). Analogous to regulatory elements that stimulate or repress transcription, positively acting enhancer elements and negatively acting silencer elements can exert dramatic effects on the control of pre-mRNA splicing by promoting or inhibiting spliceosome recruitment. While exons are generally enriched in enhancers and introns contain abundant silencers (Cartegni, 2002;Fairbrother, 2000;Fairbrother, 2002), tissue-specific alternative exons often contain both types of elements as an integral part of the regulatory mechanism. As a result, exon sequences are evolutionarily constrained by the need to ensure their own splicing via retention of appropriate enhancers and silencers.

Trans-acting splicing factors. A major paradigm in pre-mRNA splicing regulation is the concept that splicing is controlled by mutually antagonistic splicing factor proteins that bind to enhancer and silencer elements in pre-mRNA, in order to control recruitment of the spliceosomes. Typically these proteins fall into two classes, the heterogeneous nuclear ribonucleoprotein (hnRNP) proteins, and the SR proteins characterized by domains rich in arginine and serine. Splicing activator proteins typically act as spliceosome recruiting proteins by binding to

regulatory site(s) in or near an alternative exon (via a consensus RNA binding motif), and to a component(s) of the spliceosome (via a protein interaction domain). Interestingly, recruitment itself can involve any one of several different adaptors interacting with distinct spliceosomal proteins, a flexibility that may enable a cell to switch on specific sets of alternative exons in response to specific environmental signals. Equally important to splicing regulation are splicing silencer proteins that can interfere with spliceosome recruitment by blocking activity of the positive regulators. This interplay between splicing activators and inhibitors, termed "dynamic antagonism" (Charlet, 2002), plays a critical role in regulating cell type-specific splicing.

Alternative splicing and normal development. Implementation of alternative splicing switches during differentiation and development requires signaling mechanisms to transduce extracellular stimuli into changes in expression or activity of splicing activators and inhibitors. Alternative splicing switches can be signaled by insulin, stress hormones, or antigen treatment (Stamm 2002), and may ultimately be mediated by tissue-restricted expression of a splicing regulatory protein, or by physiological changes in expression of a widely expressed factor (e.g., (Hou, 2002;Jensen, 2000;Ladd, 2001). However, at present these mechanisms are poorly understood.

Alternative splicing and human disease. This subject is clearly presented in a recent review (Faustino, 2003). Mutations that alter normal splice sites or create aberrant splice sites have long been known to underlie many human diseases. More recently it has been appreciated that subtle exon mutations and distant intron mutations can also cause disease by disrupting enhancer or silencer elements. Alternatively, disturbances in the splicing machinery itself, whether engineered in mouse models or occurring naturally in human disease states, have been shown to affect splicing patterns of many genes and cause global disturbances of cellular metabolism. Quantitative shifts in the relative concentrations of enhancer and silencer proteins could alter splice site selection of key growth regulatory genes, DNA repair functions, or apoptosis genes, and contribute to cancer progression. Many individual reports of altered splicing in cancer, and two broad bioinformatic surveys of cancer-associated splicing defects, have been published recently (Wang, 2003;Xu, 2003).

Future perspectives. This is an exciting time in the alternative splicing field, as new approaches offer great promise for a better understanding of the extent of alternative splicing in the human genome and better knowledge of the mechanisms by which alternative splicing is regulated. Advances in genome-wide sequence analysis and microarray technologies offer opportunities to deduce global patterns of alternative splicing in normal tissues (Johnson, 2003). This will provide a more detailed and perhaps more accurate view of splicing than what is available today from studies that are often derived from transformed cell lines or tumors that may exhibit aberrant splicing compared to normal differentiated cells. Although computational analysis of splicing

regulatory elements has already begun (Brudno, 2001;Cartegni, 2003;Fairbrother, 2002), improved methods should increase our ability to mine this sequence data for splicing regulatory signals. Together with new biochemical and genetic techniques methods for identifying the targets of splicing regulator proteins, these advances should ultimately lead to a map of the splicing regulatory networks that control alternative splicing. Finally, the next several years should see significant progress in understanding splicing aberrations in human disease and, more importantly, in development of exciting new potential therapies for correction of such splicing defects (Cartegni, 2003;Sazani, 2003).

References

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, 72, 291-336.

Brudno M, Gelfand MS, Spengler S, Zorn M, Dubchak I and Conboy JG (2001) Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res*, 29, 2338-48.

Cartegni L, Chew SL and Krainer AR (2002) Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.*, 3, 285-298.

Cartegni L and Krainer AR (2003) Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat Struct Biol*, 10, 120-5.

Cartegni L, Wang J, Zhu Z, Zhang MQ and Krainer AR (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*, 31, 3568-71.

Charlet BN, Logan P, Singh G and Cooper TA (2002) Dynamic Antagonism between ETR-3 and PTB Regulates Cell Type-Specific Alternative Splicing. *Mol Cell*, 9, 649-58.

Fairbrother WG and Chasin LA (2000) Human genomic sequences that inhibit splicing. *Mol Cell Biol*, 20, 6816-25.

Fairbrother WG, Yeh RF, Sharp PA and Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. *Science*, 297, 1007-13.

Faustino NA and Cooper TA (2003) Pre-mRNA splicing and human disease. *Genes Dev.*, 17, 419-437.

Hirose Y and Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev*, 14, 1415-29.

Hou VC, Lersch R, Gee SL, Ponthier JL, Lo AJ, Wu M, Turck CW, Koury M, Krainer AR, Mayeda A and Conboy JG (2002) Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. *EMBO J*, 21, 6195-204.

Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, Okano HJ, Yang YY and Darnell RB (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, 25, 359-71.

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R and Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, 302, 2141-4.

Ladd AN, Charlet N and Cooper TA (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol*, 21, 1285-96.

Lee C, Atanelov L, Modrek B and Xing Y (2003) ASAP: the Alternative Splicing Annotation Project. *Nucleic Acids Res*, 31, 101-5.

Maniatis T and Reed R (2002) An extensive network of coupling among gene expression machines. *Nature*, 416, 499-506.

Modrek B and Lee C (2002) A genomic view of alternative splicing. *Nat Genet*, 30, 13-9.

Nurtdinov RN, Artamonova, II, Mironov AA and Gelfand MS (2003) Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum Mol Genet*, 12, 1313-20.

Pospisil H, Herrmann A, Bortfeldt RH and Reich JG (2004) EASED: Extended Alternatively Spliced EST Database. *Nucleic Acids Res*, 32 Database issue, D70-4.

Robberson BL, Cote GJ and Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol*, 10, 84-94.

Sazani P and Kole R (2003) Modulation of alternative splicing by antisense oligonucleotides. *Prog Mol Subcell Biol*, 31, 217-39.

Stamm S (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum Mol Genet*, 11, 2409-16.

Thanaraj TA, Stamm S, Clark F, Riethoven JJ, Le Texier V and Muilu J (2004) ASD: the Alternative Splicing Database. *Nucleic Acids Res*, 32 Database issue, D64-9.

Wang H, Lo S, Yang H, Gere S, Hu Y, Buetow KH and Lee MP (2003) Computational Analysis and Experimental Validation of Tumor-associated Alternative RNA Splicing in Human Cancer. *Cancer Research*, 63, 655–657.

Xu Q and Lee C (2003) Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic Acids Res*, 31, 5635-43.