Genome-wide Analysis of Alternative Pre-mRNA Splicing*

Claudia Ben-Dov1, Britta Hartmann1, Josefín Lundgren1, and Juan Valcárcel*

From the 1Centre de Regulació Genòmica, the 2Institució Catalana de Recerca i Estudis Avançats, and the 3Universitat Pompeu Fabra, 08003 Barcelona, Spain.

Alternative splicing of pre-mRNAs allows the synthesis of multiple mRNAs from a single primary transcript, significantly expanding the information content and regulatory possibilities of higher eukaryotic genomes. High-throughput enabling technologies, particularly large-scale sequencing and splicing-sensitive microarrays, are providing unprecedented opportunities to address key questions in this field. The picture emerging from these pioneering studies is that alternative splicing affects most human genes and a significant fraction of the genes in other multicellular organisms, with the potential to greatly influence the evolution of complex genomes. A combinatorial code of regulatory signals and factors can deploy physiologically coherent programs of alternative splicing that are distinct from those regulated at other steps of gene expression. Pre-mRNA splicing and its regulation play important roles in human pathologies, and genome-wide analyses in this area are paving the way for improved diagnostic tools and for the identification of novel and more specific pharmaceutical targets.

Removal of introns from pre-mRNAs is an essential step in eukaryotic gene expression (see Fig. 1A). Alternative patterns of intron removal allow the synthesis of multiple mRNAs from a single gene encoding different proteins (see Fig. 1B). This mini-review focuses on how the recent application of high-throughput technologies, particularly large-scale sequencing and splicing-sensitive microarrays, are providing unprecedented opportunities to address key questions in this field. The picture emerging from these pioneering studies is that alternative splicing affects most human genes and a significant fraction of the genes in other multicellular organisms, with the potential to greatly influence the evolution of complex genomes. A combinatorial code of regulatory signals and factors can deploy physiologically coherent programs of alternative splicing that are distinct from those regulated at other steps of gene expression. Pre-mRNA splicing and its regulation play important roles in human pathologies, and genome-wide analyses in this area are paving the way for improved diagnostic tools and for the identification of novel and more specific pharmaceutical targets.

Biological Relevance

It is unquestionable that AS can generate mRNAs with important and distinct biological functions (10–13). The more difficult question is what fraction of the extensive transcript variation generated by AS is truly biologically relevant and what fraction may be due to stochastic noise in the splicing process. Biological Incidence—Consistent with the idea that AS plays important roles in cellular function, bioinformatic and array data indicate that the process is more prominent in tissues with diverse cell types and among genes playing regulatory functions (4, 12, 14). An important insight from global studies in a variety of biological situations is that the overlap between genes that show changes in AS and those regulated through changes in transcript levels is relatively limited (but also see Ref. 18). This observation suggests the existence of dedicated regulatory programs that coordinately control multiple AS events. Such programs are evident, for example, in Drosophila sex determination (19) and mammalian synaptic transmission (20, 21) and strongly argue for the biological relevance of AS.

Predicted Functional Effects—Bioinformatic analyses indicate that 75% of AS events affect coding regions, with predicted effects ranging from subtle amino acid substitutions to removal of protein motifs or protein truncations (10–12). Although compelling instances of biologically relevant changes due to a

Supplemental Material can be found at:
http://www.jbc.org/content/suppl/2007/11/19/R700033200.DC1.html
untranslated regions that are subject to translation. Thus, 40% of predicted microRNA targets are not the transcripts to factors that regulate their stability or also lead to functional differences, for example, to expose (or regulators (28–32).

of this mechanism in the control of expression of splicing regulators (27).

light the remarkable evolutionary conservation and importance genetically conserved (16). Nonetheless, recent reports highlight the remarkable evolutionary conservation and importance of this mechanism in the control of expression of splicing regulators (28–32).

AS events occurring in untranslated regions of mRNAs can also lead to functional differences, for example, to expose (or not) the transcripts to factors that regulate their stability or translation. Thus, 40% of predicted microRNA targets are located in regions of 3’-untranslated regions that are subject to AS and/or alternative polyadenylation (33).

Evolutionary Conservation—Conserved AS events are more likely to be functionally important because they show a higher tendency to preserve the reading frame, to modify protein-coding sequences, and to conserve regulatory sequences and be tissue-specific (12). However, computational predictions indicate that only 10–20% of AS events appear to be conserved between human and mouse, a figure substantially lower than the overall gene conservation between these organisms (12, 34). Although this figure may increase as more extensive data sets become available, does it argue against the functional relevance of the majority of AS events? An alternative view is that AS can play important roles in the evolution of individual genes. A compelling example is the exonization of transposable elements present in intronic regions of some vertebrate genes, which evolved mechanisms of limited exon inclusion that facilitate exploration of novel functions without compromising expression of the host gene (35). There is substantial evidence for the creation or loss of tens of thousands of exons across vertebrate genomes, which can function as evolutionary “hot spots” (36).

In summary, neither simple structure/function predictions nor extensive phylogenetic comparisons provide straightforward evidence of biological relevance for the majority of AS events. However, evidence accumulated through classical gene-by-gene studies and persuasive examples of coordinated regulation of biological processes through AS readily justify future systematic functional analyses of AS isoforms.

Regulatory Mechanisms

As elaborated in other minireviews in this series (37–39), splice site selection depends not only on the identity of the individual splice sites but also on numerous regulatory sequence motifs in the neighboring exons and introns, which are recognized by proteins of the SR protein and hnRNP families as well as other splicing factors that associate with the RNA as it emerges from transcriptional complexes (Fig. 1B) (10–12, 40).

Bioinformatics—Splicing regulatory sequences have been identified through a variety of approaches (reviewed in Ref.40), including (i) comparisons between exons in genes and in pseudogenes or intronless genes, (ii) phylogenetic nucleotide conservation not explained by conservation of amino acid sequence, and (iii) correlations between sequence motifs and the strength of the neighboring splice sites. Additional regulatory sequences were identified as characteristic signatures of tissue-specific splicing or of activation of specific signaling pathways (10, 12, 40). When added to motifs identified as preferred binding sites for known regulatory factors, the picture coming from these studies is that a substantial fraction of exonic and flanking intronic sequences, both constitutively and alternatively spliced, play roles in modulating the splicing process. Furthermore, the extent and nature of these effects some-
times depend on their position relative to splicing signals (41) or to other regulatory motifs (42).

Combinatorial assemblies of regulatory motifs and factors act on individual transcripts to facilitate or preclude splice site recognition by the spliceosome and thus establish patterns of AS. Cell type-specific splicing can be achieved either by expression of cell type-specific regulators or through cell type-specific variations in the levels or activity of more ubiquitous factors (10–12).

Splicing-sensitive Microarrays—Various platforms for high-throughput detection of splice variants and changes in AS are contributing to decipher these cellular codes (Fig. 2B and supplemental Table 1). First, results from these approaches have increased substantially the number of known tissue-specific AS events (4, 51–53) or of events changing under specific conditions like T cell activation (17). They have even revealed an unexpected flexibility in the splicing machinery of the yeast *Saccharomyces cerevisiae* (previously thought to be mostly devoid of regulated splicing) to adapt to stress conditions (43).

Splicing microarrays have also been used to identify AS events regulated upon ablation or overexpression of specific factors (20, 44–48), significantly expanding the known targets for tissue-specific (20, 45) or ubiquitous (45–47) regulators. In addition, results from yeast arrays and systematic RNA interference screens in *Drosophila* revealed substrate-specific splicing defects and changes in AS caused by depletion of core components of the spliceosome, uncovering the unexpected regulatory potential of these factors (44, 48, 49).

Genome-wide studies have helped to evaluate the generality of mechanistic insights gained from gene-by-gene studies. For example, previous work indicated that SR protein splicing factors have redundant biochemical activities and that some members of the SR protein and hnRNP families display general antagonistic activities in AS regulation (10). Results from splicing microarrays did show changes in AS consistent with these conclusions, but also provided extensive evidence for specific functions of individual splicing regulators and for a complex network of genetic interactions between regulators and AS events (45). For example, the targets of a soma-specific splicing regulator (PSI) were included within a subset of the targets of a ubiquitous hnRNP protein (Hrp48), suggesting that PSI requires Hrp48 as a cofactor, whereas Hrp48 has additional, PSI-independent functions.

Finally, important cellular regulatory circuits have been uncovered using splicing microarrays. These include, for instance, changes in AS relevant for cell transformation induced by the SR protein SF2/ASF (47) (see below) and regulation of a neuron-specific form of PTB (nPTB) by the more ubiquitous isoform of this splicing regulator, a feedback mechanism that controls an extensive alternative splicing program during neural development (46).

Combining Technologies—The combined use of experimental and computational approaches has been very fruitful for identifying regulatory sequences (50, 51), including motifs involved in tissue-specific AS (52, 53). Interestingly, although muscle-specific sequence motifs were largely overlapping with
MINIREVIEW: Genome-wide Analysis of Alternative Splicing

known binding sites for muscle-specific regulators (52), numerous potential novel regulatory sequences for nervous system-specific splicing were found (53), consistent with the high prevalence of AS in this tissue (4, 14).

Combinations of various high-throughput, biochemical, and computational methods have allowed the identification of numerous target genes of the Nova neuron-specific splicing regulators and also the drawing of a regulatory map that correctly predicts the effect of these factors on AS regulation (20, 21, 54). High-throughput technologies include ultraviolet cross-linking and immunoprecipitation of RNAs bound to Nova in mouse brain, which identified >30 novel RNAs from genes involved in synapse inhibition (54). Splicing-sensitive microarrays identified 40 neuron-specific exons from a largely distinct set of synapse-related genes that showed splicing defects in the neocortex of Nova-2 knock-out mice (20). Interestingly, a significant fraction of the protein products of these genes were known to form complexes, suggesting that AS modulates the function of these complexes to shape the synapse.

Impact on Disease

As illustrated by the examples below, a mounting body of evidence implicates splicing defects and altered splicing regulation as causes or modifiers of numerous pathologies. Systematic screening for mRNA defects indicates that 50% of the mutations identified in patients with neurofibromatosis or ataxia-telangiectasia affect splicing of the NF1 or ATM genes, respectively (55). Noncoding nucleotide expansion disorders affect the function of splicing factors: CTG expansions characteristic of myotonic dystrophy patients, transcribed into RNA, alter the activity of CUG-binding proteins, resulting in changes in AS of other genes that can explain some associated symptoms (56). Overexpression of the SR protein SF2/ASF leads to efficient cell transformation and tumor formation by altering the ratios between isoforms of key regulators of cell growth (47).

High-throughput technologies for detecting AS hold the promise of improved diagnostic and prognostic tools (57). Both computational predictions and microarray experiments have identified hundreds of AS and aberrant splicing events associated with disease states, particularly various cancers (18, 58–67). These add to a growing list of AS changes and aberrant splicing events known to affect cellular features very relevant for tumor growth, including cell transformation, motility, invasiveness, and angiogenesis (reviewed in Ref. 68).

Consistent with the notion that changes in AS can be relevant for cellular phenotypes and lymphoma grade, proper classification of histologically distinct glial brain tumors, and improved diagnosis of prostate cancer based upon differential expression of exons have been reported using splicing microarrays (63, 66, 67). Other medically relevant insights provided by these studies include a potential autocrine mechanism for the development of chorionicarcinomas, ectopic functional expression of neuron-specific splicing regulators in lymphomas, and the influence of growth substrate on the expression profile of breast cancer cell lines (61, 63, 69).

Detailed knowledge about isoform expression provides the possibility to identify novel, more specific, and safer targets for drug design (57). For example, voltage-independent inhibition of N-type calcium channels in the pain pathway depends critically on an alternatively spliced exon that makes these channels more sensitive to neurotransmitters and drugs (22). In this regard, individual variation in splicing patterns related to population haplotypes (5, 6) may add yet another dimension to personalized medicine.

Future

Likely additions to the arsenal of high-throughput technologies to study AS include proteomic technologies able to distinguish and quantify protein isoforms. The potential of two-dimensional difference gel electrophoresis, which compares protein samples labeled with different fluorescent dyes, has recently been illustrated by the discovery of cross-regulation and functional redundancy between paralogs of the splicing regulator PTB (31). High-throughput comparative mass spectrometry, e.g. through differential isoform detection techniques (70, 71), and the identification of novel AS events, which can be helped by the construction of cDNA libraries enriched in alternatively spliced transcripts (72).

Technologies enabling genome-wide analyses are likely to become standard tools for addressing virtually every functional, mechanistic, medical, or evolutionary question in gene function and AS. The possibility to combine various technologies (e.g. cross-linking and immunoprecipitation with microarray analyses), particularly merging large-scale data gathering with bioinformatic methods, will be of great value for modeling biological impact and regulatory mechanisms and networks (21).

Acknowledgments—We thank Ben Blencowe and Chris Smith for comments on the manuscript.

REFERENCES

1. Zavolan, M., and van Nimwegen, E. (2006) Curr. Opin. Struct. Biol. 16, 362–367
2. Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R., and Shoemaker, D. D. (2003) Science 302, 2141–2144
3. Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., Tammana, H., and Gingeras, T. R. (2004) Genome Res. 14, 331–342
4. Clark, T. A., Schweitzer, A. C., Chen, T. X., Staples, M. K., Lu, G., Wang, H., Williams, A., and Blume, J. E. (2007) Genome Biol. 8, R64
5. Kwan, T., Benovoy, D., Dias, C., Gurd, S., Serre, D., Zuzan, H., Clark, T. A., Schweitzer, A., Staples, M. K., Wang, H., Blume, J. E., Hudson, T. J., Sladek, R., and Majewski, J. (2007) Genome Res. 17, 1210–1218
6. Hull, J., Campino, S., Rowlands, K., Chan, M. S., Copley, R. R., Taylor, M. S., Rockett, K., Elvidge, G., Keating, B., Knight, J., and Kwiatkowski, D. (2007) PLoS Genet. 3, e99
MINIREVIEW: Genome-wide Analysis of Alternative Splicing

7. Kim, E., Magen, A., and Ast, G. (2007) Nucleic Acids Res. 35, 125–131
8. Stoic, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M. F., Rifkin, S. A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P. E., Bussemaker, H. J., and White, K. P. (2004) Science 306, 655–660
9. Chen, B. E., Kondo, M., Garnier, A., Watson, F. L., Puettmann-Holgado, R., Lamar, D. R., and Schmucker, D. (2006) Cell 125, 607–620
10. Black, D. L. (2003) Annu. Rev. Biochem. 72, 291–326
11. Matlin, A. J., Clark, F., and Smith, C. W. (2005) Nat. Rev. Mol. Cell Biol. 6, 386–398
12. Blencowe, B. J. (2006) Cell 126, 37–47
13. Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A., and Soreq, H. (2005) Gene (Annot) 344, 1–20
14. Pan, Q., Shi, O., Misquitta, C., Zhang, W., Saltzman, A. L., Mohammad, N., Babak, T., Shi, H., Hughes, T. R., Morris, Q. D., Frey, B. J., and Blencowe, B. J. (2004) Mol. Cell 16, 929–941
15. Le, K., Mitsouras, K., Roy, M., Wang, Q., Xu, Q., Nelson, S. F., and Lee, C. J. (2007) Nucleic Acids Res. 35, 773–787
16. Li, H. R., Wang-Rodriguez, J., Nair, T. M., Yeakley, J. M., Kwon, Y. S., Liu, L., Rozenhek, A., Mohammad, N., Williams, A., Kato, M., Shiue, L., Ares, M., Jr., and Black, D. L. (2007) Genes Dev. 21, 1636–1652
17. Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., and Krainer, A. R. (2007) Nat. Struct. Mol. Biol. 14, 185–193
18. Pleiss, J. A., Whitworth, G. B., Bergkessel, M., and Guthrie, C. (2007) Mol. Cell 27, 928–937
19. Clark, T. A., Sugnet, C. W., and Ares, M., Jr. (2002) Science 296, 907–910
20. Blancho, M., Green, R. E., Brenner, S. E., and Rio, D. C. (2005) Genes Dev. 19, 1306–1314
21. Bouz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., Jr., and Black, D. L. (2007) Genes Dev. 21, 1636–1652
22. Raingo, J., Castiglioni, A. J., and Lipscombe, D. (2007) Cell 130, 202–212
23. Han, K., Yeo, G., An, P., Burge, C. B., and Grabowski, P. J. (2005) PLoS Biol. 3, e158
24. Pleiss, J. A., Whitworth, G. B., Bergkessel, M., and Guthrie, C. (2007) PLoS Biol. 5, e90
25. Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A., and Graveley, B. R. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 15974–15979
26. Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., and Burge, C. B. (2004) Cell 119, 831–845
27. Sugnet, C. W., Srinivasan, K., Clark, T. A., O’Brien, G., Cline, M. S., Wang, H., Williams, A., Kulp, D., Blume, J. E., Haussler, D., and Ares, M., Jr. (2006) PLoS Comput. Biol. 2, e4
28. Das, D., Clark, T. A., Schweitzer, A., Yamamoto, M., Harr, H., Arribere, J., Minovitsky, S., Pollakover, A., Dubchak, I., Blume, J. E., and Conboy, J. G. (2007) Nucleic Acids Res. 35, 4845–4857
29. Fagnani, M., Barash, Y., Ip, J., Misquitta, C., Pan, Q., Saltzman, A. L., Shi, O., Lee, L., Rozenhek, A., Mohammad, N., Willaime-Morawek, S., Babak, T., Zhang, W., Hughes, T. R., van der Kooy, D., Frey, B. J., and Blencowe, B. J. (2007) Genome Biol. 8, R108
30. Ule, J., Jensen, K. B., Ruggini, M., Mele, A., Ule, A., and Darnell, R. B. (2003) Science 302, 1212–1215
31. Pagani, F., and Baralle, F. E. (2004) Nature 430, 389–396
32. Ranum, L. P., and Cooper, T. A. (2006) Annu. Rev. Neurosci. 29, 257–277
33. Brinkman, B. M. (2004) Clin. Biochem. 37, 584–594
34. Zhu, H., Ye, G., Wang, S., Hu, Y., Buetow, K. H., and Lee, M. P. (2003) Cancer Res. 63, 655–667
35. Xu, Q., and Lee, C. (2003) Nucleic Acids Res. 31, 5635–5643
36. Roy, M., Xu, Q., and Lee, C. (2005) Nucleic Acids Res. 33, 5026–5033
37. Yeakley, J. M., Fan, J. B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M. S., and Fu, X. D. (2002) Nat. Biotechnol. 20, 353–358
38. Nagao, K., Toyoda, M., Takeuchi-Inoue, K., Fujii, K., Yamada, M., and Miyashita, T. (2005) Genomics 85, 462–471
39. Relogio, A., Ben-Dov, C., Baum, M., Ruggini, M., Gemund, C., Benes, V., Darnell, R. B., and Valcarcel, J. (2005) J. Biol. Chem. 280, 4779–4784
40. Gardina, P. J., Clark, T. A., Shimada, B., Staples, M. K., Yang, Q., Veitch, J., Schweitzer, A., Awad, T., Sugnet, C., Dee, S., Davies, C., Williams, A., and Turpur, Y. (2006) BMC Genomics 7, 325
41. Milani, L., FREDRIKSSON, M., and Syvanen, A. C. (2006) Clin. Chem. 52, 202–211
42. Zhang, C., Li, H. R., Fan, J. B., Wang-Rodriguez, J., Downs, T., Fu, X. D., and Zhang, M. Q. (2006) BMC Bioinformatics 7, 202
43. French, P. J., Peeters, H., Horsman, S., Duijum, E., Siccama, I., van den Bent, M. J., Luider, T. M., Kros, J. M., van der Spek, P., and Sillevis Smitt, P. A. (2007) Cancer Res. 67, 5635–5642
44. Venables, I. P. (2006) BioEssays 28, 378–386
45. Li, C., Kato, M., Shiozawa, K., Eres, J. A., Eres, M. J., and Lin, R. I. (2006) Cancer Res. 66, 1990–1999
46. Zhu, J., Shendure, J., Mitra, R. D., and Church, G. M. (2003) Science 301, 838–838
47. Calarco, J. A., Saltzman, A. L., Ip, J. Y., and Blencowe, B. J. (2007) in Alternative Splicing in the Postgenomic Era (Blencowe, B. J., and Gravelle, B. R., eds) pp. 85–106, Landes Bioscience, Austin, TX, in press
48. Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., and Burge, C. B. (2004) Cell 119, 831–845
49. Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., and Burge, C. B. (2004) Cell 119, 831–845
50. Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., and Burge, C. B. (2004) Cell 119, 831–845