A Guide to One of the Genome's Best-Kept Secrets

Alternative Splicing in the Postgenomic Era

Benjamin J. Blencowe and Brenton R. Graveley, Eds.
Austin, TX: Landes Bioscience (2007)
228 pp., ISBN 9780387773735.

When Jean François Champollion visited Egypt in 1828, after having spent many years in France cracking the ancient Egyptian alphabet, he could read hieroglyphic messages portraying a splendid long-lost civilization, a feat which no other person had accomplished for millennia. Although the central dogma of molecular biology provides a universal principle to decode genomes, important aspects of the syntax of complex genomes remain fraught with mystery, like Egyptian hieroglyphs before Champollion. One of these is the generation of multiple mRNAs through alternative splicing of mRNA precursors. This process is intimately linked to the counterintuitive organization of eukaryotic genes in coding modules (exons), separated by longer stretches of sequence (introns). Introns must be eliminated from primary transcripts to generate translatable mRNA species. The distinction between exonic and intronic sequences is sufficiently flexible to allow the generation of mRNA variants harboring different combinations of exonic sequences, sometimes adding up to thousands of potential mRNAs generated from a single primary transcript (reviewed by Blencowe, 2006). Future molecular biologists might look back in disbelief to a time when it was not possible to directly interpret genomic sequences to predict whether transcriptional activation of a gene in a particular cell will generate one type of mRNA or another, frequently with significant—sometimes dramatic—consequences for the functional output of the gene. Ben Blencowe and Brent Graveley have gathered timely contributions from many experts in this area to produce the first comprehensive book on alternative splicing, capturing both the current excitement and the challenges within this field: the prevalence and biological relevance of the process, its molecular mechanisms of regulation, and the impact on disease.

Intron removal requires one of the most complex molecular machineries in the cell, the spliceosome, comprising five small nuclear ribonucleoprotein particles (snRNPs) and more than 300 proteins harboring various domains and enzymatic activities. Three chapters contributed by Moore, Fu, Chabot, and their colleagues provide a parts list as well as an overview of the properties and dynamics of this complex and of two families of factors (SR and hnRNP proteins) that modulate its assembly. The overarching concept is that an intricate network of RNA-protein, protein-protein, pre-mRNA/snRNA, and snRNA/snRNA interactions directs and proofreads spliceosome assembly and ultimately provides a scaffold that brings together the chemical moieties that participate in the two catalytic steps of the splicing reaction. Regulation of splice site selection can occur at early steps of splice site recognition but also at subsequent steps in complex assembly, thus positioning the intricacies of spliceosome function as potential targets for modulation (House and Lynch, 2008). These chapters provide an updated view of established facts and models, and they also rightly emphasize our limited understanding of basic principles in splice site selection, including (1) the extent to which our view of spliceosome assembly, based upon detailed studies on an embarrassingly low number of pre-mRNAs, will be useful to understand regulation in general; (2) what general properties of SR proteins and their associated arginine and serine-rich (RS) domains provide them with the capacity to display a wide variety of effects on splicing control; (3) why, despite these widespread effects, individual SR proteins appear to have a limited set of targets and functions in vivo; (4) how these factors and hnRNP proteins achieve specificity to influence the structure of nascent RNPs (e.g., looping out of sequences to approximate splice sites or cause exon skipping) and the accessibility of core splicing factors to the splice sites.

A conundrum in the splicing field, with resonances on genome structure and genetic disease, is the nature of the sequences that determine efficient and precise intron excision. The very limited conservation at splice sites in complex eukaryotes suggests that additional sequences contribute, positively or negatively, to splice site recognition. A chapter by Chasin exhaustively reviews the nature of these regulatory sequences. The collective knowledge accumulated experimentally or through computational analyses puts forward the remarkable conclusion that a large fraction of all possible sequence motifs do indeed influence the splicing process. This "embarrassment of riches" implies that exonic sequences often weave two information codes into one sequence, one for protein synthesis, another for proper pre-mRNA processing. Another question posed by these findings is whether the function of this large variety of sequence elements and associated factors can be understood through some general and simple principles, or whether the outcome of splice site selection will be dictated by the unique configuration of the RNP in each exon or pre-mRNA, i.e., understandable only upon structural determination of each ribonucleoprotein assembly. In other words, will we ever be able to predict alternative splicing outcomes on the basis of sequence alone? A chapter by Ule and Darnell dedicated to the neuron-specific Nova regulators provides hope for building predictive rules. Will these RNA maps be extensible to more ubiquitous regulators, and can their cooperative and antagonistic effects be factored into sequence analysis algorithms? Indeed, owing to the difficulty in confidently predicting and determining complex RNA structures in the context of RNPs, the field largely ignores their role in splice site selection, but as argued by Park and Graveley in another chapter, the availability of sequences from multiple, related genomes can greatly help to fill this gap.

The variety of sequences and factors influencing splicing patterns provide a multitude of Achilles' heels that allow mutations to promote disease development. Orengo and Cooper document several compelling examples of pathologies caused by alterations in RNA regulatory elements (e.g., tauophaties), regulatory factors (e.g., various muscular dystrophies), or combinations of both that influence disease progression (e.g., cancer).

The growing number of RNA-based pathologies should spur renewed interest in alternative splicing and begs for the involvement of the medical community. Another emerging area is the impact that natural sequence variation can have on splicing patterns and its possible role as disease modifier (Hull et al., 2007; Kwan et al., 2008).

Most of our detailed knowledge of splicing regulatory mechanisms stems from biochemical analysis in nuclear extracts. Although essential in dissecting molecular processes, these approaches have intrinsic limitations, including the absence of a physiological cellular context and the difficulty of recapitulating splicing reactions in the context of transcriptionally active chromatin. Carmo-Fonseca and Carvalho review recent advances in live-cell imaging that hold the promise of realistically describing the localization and dynamics of splicing factors within the nucleus, their association with other factors and with nascent transcripts, as well as the identification of limiting steps in RNP production, processing, quality control, and transport to the cytoplasm. Although the impact of these features on splicing regulation remains incompletely understood, the use of these technologies can provide unique insight into critical steps in splicing control. Splicing can occur cotranscriptionally, and a chapter by Kornblihtt reviews evidence for effects of promoter structure, recruitment of processing factors through the RNA polymerase II Carboxy-Terminal-Domain (CTD), and changes in transcription elongation rates on alternative splicing. This view is reinforced by recent evidence that classical splicing regulatory factors can influence transcription elongation rates of certain genes (Lin et al., 2008). Although the generality of coupling and its—sometimes long-range—effects on alternative RNA processing requires genome-wide monitoring, it seems likely that understanding physiological changes in alternative splicing will require the stereo vision of transcription and splicing in the context of chromatin structure and, possibly, chromatin modifications.

Another topic whose exploration requires the cellular context is the convergence between signaling pathways and alternative splicing, a relatively slim area of activity in RNA processing that is bound to expand considerably judging by the impact that signaling has in other key steps of gene regulation and by the interesting examples provided in a chapter by Lynch. Posttranslational modifications of splicing regulatory factors, which alter their RNA and protein binding properties, likely influence in vivo splicing. Indeed, examples of physiological changes in splicing and the associated changes in activity of specific regulatory factors linked to particular biological processes remain relatively rare (e.g., Hou et al., 2002; Ule et al., 2005; Xu et al., 2005; Karni et al., 2007; Ohno et al., 2008). In this regard, an additional chapter focusing on the potential manipulation of splicing factor genes or alternative splicing reporters in animal models would have been useful.

It is clear that alternative splicing can diversify the genetic repertoire of multicellular organisms and has the potential to serve as a key contributor to the complexity of their architecture and adaptability. Yet it is also important to realize that we ignore both the functional consequences of the vast majority of the known or predicted alternative splicing events as well as the extent to which the observed transcript heterogeneity might reflect noise in splice site selection. The book offers a balanced view of this field's *soul search*. A chapter by Park and Graveley provides examples of genes undergoing spectacularly complex patterns of alternative splicing with deep functional implications on key biological functions, where a handful of *Drosophila* genes can collectively generate more than one million protein variants. The chapter by Ule and Darnell provides a compelling example of how neuron-specific regulatory factors influence alternative splicing of multiple genes to orchestrate synaptic function. It is difficult to ignore the tremendous regulatory potential and the elegant expansion of genomic information afforded by alternative splicing through these examples and not be tempted to extrapolate to whole genomes. The time is ripe for this wishful thinking because rapidly expanding data sets of cDNA and EST sequences, as well as data generated through the use of high-throughput technologies, including splicing-sensitive microarrays and proteomics, indicate the widespread prevalence of alternative splicing in multicellular organisms (Ben-Dov et al., 2008). The detailed discussion of these technologies by Blencowe and colleagues is an important asset of the book, which also predicts the impact of emerging high-throughput sequencing technologies (Cloonan et al., 2008; Mortazavi et al., 2008; Sultan et al., 2008). It is also clear, as stated by the editors, that new high-throughput functional assays are needed to directly assess the biological relevance of alternatively spliced variants from a genome-wide perspective, perhaps the greatest challenge for the field today.

One approach to gauge the biological relevance of alternative splicing events is to assess their predicted impact on protein function and their evolutionary conservation. Chapters contributed by Lee, Brenner, and their colleagues address these issues. Once again, compelling examples of conserved alternative splicing events with clear functional implications can be identified. Ultraconserved sequence elements are found, for instance, in exons containing stop codons within genes encoding splicing factors. Alternative splicing of these exons establishes feedback loops and other regulatory networks that control the expression of key splicing regulators across long evolutionary distances. Yet the majority of alternative splicing events do not appear to be conserved, prolonging the debate on whether they provide a test ground for species-specific tinkering with new protein modules or whether they reflect mostly gene expression noise.

Collectively, the contributions presented in this book make a clear, rigorous, and solidly documented case for continued and enhanced efforts in alternative splicing studies. Additional chapters on alternative splicing in plants, the potential of *C. elegans* to combine genetics, genomics, and fluorescent reporters to crack splicing codes (Barberan-Soller and Zahler, 2008; Ohno et al., 2008), and also the potential of even simpler models organisms to explore the flexibility of the core splicing apparatus (Pleiss et al., 2007) would have enhanced the allure of the book even further. Although we cannot yet claim to have cracked the splicing code or fully understand its biological impact, the volume edited by Blencowe and Graveley provides both a guide to current knowledge, models, and paradigms and a road map for future Champollions.

REFERENCES

Barberan-Soller, S., and Zahler, A.M. (2008). PLoS Genet. *4*, e1000001. 10.1371/journal.pgen.1000001.

Ben-Dov, C., Hartmann, B., Lundgren, J., and Valcárcel, J. (2008). J. Biol. Chem. *283*, 1229–1233.

Blencowe, B.J. (2006). Cell *126*, 37–47.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., et al. (2008). Nat. Methods *5*, 613–619.

Hou, V.C., Lersch, R., Gee, S.L., Ponthier, J.L., Lo, A.J., Wu, M., Turck, C.W., Koury, M., Krainer, A.R., Mayeda, A., and Conboy, J.G. (2002). EMBO J. *21*, 6195–6204.

House, A.E., and Lynch, K.W. (2008). J. Biol. Chem. *283*, 1217–1221.

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. 10.1371/journal.pgen.0030099.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). Nat. Struct. Mol. Biol. *14*, 185–193.

Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R., and Majewski, J. (2008). Nat. Genet. *40*, 225–231.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). Nat. Struct. Mol. Biol. *15*, 819–826.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Nat. Methods *5*, 621–628.

Ohno, G., Hagiwara, M., and Kuroyanagi, H. (2008). Genes Dev. *22*, 360–374.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Mol. Cell *27*, 928–937.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). Science *321*, 956–960.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. (2005). Nat. Genet. *37*, 844–852.

Xu, X., Yang, D., Ding, J.H., Wang, W., Chu, P.H., Dalton, N.D., Wang, H.Y., Bermingham, J.R., Jr., Ye, Z., Liu, F., et al. (2005). Cell *120*, 59–72.

Mafalda Martins-Araujo[1,2] and Juan Valcárcel[1,2,3,*]
[1]Centre de Regulació Genòmica
[2]Universitat Pompeu Fabra
[3]Institució Catalana de Recerca i Estudis Avançats
Dr. Aiguader 88, 08003 Barcelona, Spain
*Correspondence: juan.valcarcel@crg.es
DOI 10.1016/j.molcel.2008.09.004