

Meeting report

Splicing bioinformatics to biology

Douglas L Black* and Brenton R Graveley†

Addresses: *Department of Microbiology, Immunology and Molecular Genetics, Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095-1662, USA. †Department of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030-3301, USA.

Correspondence: Douglas L Black. Email: dougb@microbio.ucla.edu

Published: 26 May 2006

Genome Biology 2006, **7**:317 (doi:10.1186/gb-2006-7-5-317)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2006/7/5/317>

© 2006 BioMed Central Ltd

A report on the 2nd Symposium on Alternative Transcript Diversity, Heidelberg, Germany, 21-23 March 2006.

Alternative splicing affects many aspects of eukaryotic biology and is studied by groups with diverse interests. Geneticists and biochemists have long been interested in understanding the molecular mechanisms that underlie changes in splice-site choice, and the role of splicing regulation in particular biological systems. More recently, computational biologists have entered the field with the goals of defining the products of genomes and understanding the role of alternative splicing in genome evolution. Although their interests broadly overlap, these fields often utilize distinct languages, and there have been relatively few meetings dedicated to bringing the two groups together. Exceptions have been the symposia on alternative transcript diversity organized by the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI); the second symposium was held in March in Heidelberg. This meeting made clear that the interests of these two groups coincide more than ever, and that combining genomic approaches with mechanistic analyses is leading to significant new understanding of splicing regulation.

The combined approach was apparent in the opening talk given by one of us (B.G.) describing the use of comparative genomics in analysis of the splicing of the *Dscam* locus in *Drosophila*. This gene is the most complex system of alternative splicing yet described. *Dscam* contains several large arrays of alternative exons that are used in a mutually exclusive manner where only one exon in each array is spliced into the *Dscam* mRNA. The mechanisms that enforce the mutually exclusive choice in such a large array are obscure. For one array (exon 6), the comparative sequence analysis

identified conserved features that predict base pairing between a docking site in the intron upstream of the array and selector sequences adjacent to each alternative exon. This finding leads to a unique model for the regulation of exon 6 splicing, in which mutually exclusive pairing between the docking sequence and one of the selector sequences ensures that only one exon 6 variant is included.

Splice sites and control elements in RNA

Comparative genomics, specifically identifying conserved splicing patterns and regulatory elements, was a recurring theme. Chris Lee (University of California, Los Angeles, USA) described how major-form alternative exons, those that are included in more than two-thirds of a gene's transcripts, are more highly conserved than minor-form exons, which are included less frequently. Following the evolution of exons through the mammalian lineage, Lee estimates that it takes roughly 40 million years for a newly evolved exon to become functionalized and fixed in a genome. This suggests that a low level of inclusion allows newly evolved exons to persist even if deleterious. In this way, the exon can continue to evolve and, if it gains advantageous features, can become a major-form exon. This evolutionary pathway seems particularly common in exons that show tissue-specific inclusion. It has been known for some time that exons that make functionally significant changes to an mRNA or protein are generally highly conserved across related species, such as within the mammalian or wider vertebrate lineages. Extending this idea, Peer Bork (EMBL, Heidelberg, Germany) described searches for exons whose regulation is conserved across all metazoans. Starting with a set of defined orthologous genes, his group defined exons whose variable inclusion is conserved across multiple species. Interestingly, few exons are regulated in all species, but larger numbers show apparent conserved regulation

between humans and at least one insect. This is an interesting strategy for identifying splicing events of particular biological importance, given the conservation of their regulation over such a large evolutionary distance.

The high degree of conservation of regulated exons can be used to identify new splicing-regulatory sequences. Interestingly, Lee pointed out that the selection against synonymous codon changes in alternative exons appears too great to be explained by the set of known exonic regulatory elements, indicating that there are potentially many more elements to be identified. To address this issue, Gil Ast (Tel Aviv University, Tel Aviv, Israel) described a novel strategy for identifying new exonic splicing-regulatory sequences by searching for dicodons (two consecutive codons) whose synonymous positions are unusually conserved in alternative exons. Some of these elements were functionally validated in heterologous reporter genes, where their effects on splicing were surprisingly variable and depended on their exact location within an exon. Such analyses, which are being conducted by several different research groups, will ultimately help to identify the full spectrum of *cis*-regulatory elements that control alternative splicing. This is important both for understanding mechanisms of regulation and as a predictive tool in defining alternative exons within genomic sequence.

An important question in defining splice sites is the fidelity of the splicing reaction. Error rates in splicing have been difficult to measure *in vivo* because many mis-spliced transcripts are degraded by the nonsense-mediated mRNA decay pathway. Mihaela Zavolan (Biozentrum, University of Basel, Switzerland) and Michael Hiller (University of Freiburg, Germany) have analyzed a special case in which spliced products differ by three nucleotides through the use of tandemly duplicated 3' splice sites, also known as NAGNAG acceptors. Zavolan described the finding that in most cases, the upstream AG is preferentially used. In approximately 25% of these sequences, however, either the downstream AG or both are used. Hiller described the identification of single-nucleotide polymorphisms (SNPs) that affect the relative use of the two AGs in NAGNAG sequences. These are being used to predict which NAGNAG sequences in the genome will behave as typical sites to splice only at the upstream AG, and which will produce splicing at both positions.

Proteins regulating splicing

Several talks examined the interdependence of the sequence elements controlling the inclusion of a particular exon. Bertrand Séraphin (Centre de Génétique Moléculaire, Gif-sur-Yvette, France) described how the human homolog of the yeast snu30 protein, a component of the U1 small nuclear ribonucleoprotein (snRNP), can bind to the pre-mRNA and affect 5' splice-site choice. Interestingly, human snu30 is not stoichiometrically associated with the U1

snRNP and is not apparently required for all splicing events, making it a possible point of regulation for splice-site choice. Séraphin also described the coordination of the U5 and U6 snRNPs in determining the site of 5' splice-site cleavage and the role of another splicing factor, the Res protein, in this selection. Looking at the other end of the intron, Angela Krämer (University of Geneva, Switzerland) described extensive studies of the protein SF1, which recognizes the branchpoint and is required for splicing in yeast. Interestingly, in mammals this protein is not required for all splicing events, but is needed for certain alternatively spliced exons. In this case, the requirement for SF1 may be affected by how well the splicing factor U2AF binds to the 3' splice site. An RNA interference (RNAi) screen in *Drosophila* described by one of us (B.G.) has also identified a number of spliceosome components as effectors of alternative splicing. Thus, the role of ostensibly constitutive splicing factors in splicing regulation was another recurring theme of the meeting.

Goran Akusjarvi (University of Uppsala, Sweden) described his group's recent studies of splicing regulation during adenovirus infection, which have uncovered a highly specific regulatory protein. In the viral *IIIa* gene they discovered a 3' splice site that is active late in viral infection and is not dependent on the standard spliceosome component U2AF. In biochemical experiments, they identified the viral protein L4-33K as the activator of this splice site. L4-33K has an interesting domain structure that includes the arginine-serine repeats required for splicing activation. Studies of this protein should yield important information on how 3' splice sites are chosen by the spliceosome.

Also apparent is a wave of interest in understanding alternative splicing on a genome-wide level. Krämer and Javier Caceres (MRC Human Genetics Unit, Edinburgh, UK) both used a crosslinking and immunoprecipitation (CLIP) procedure developed in Robert Darnell's laboratory to identify large sets of *in vivo* binding sites for the splicing factors SF1 and SF2/ASF, respectively. DNA microarrays are also becoming more widely used to characterize alternative splicing throughout the genome, as reported by several groups. A powerful approach is to examine splicing changes after RNAi knockdown of particular splicing factors. For example, Donald Rio (University of California, Berkeley, USA) described his laboratory's use of genome-wide splice-junction arrays to identify exons that are regulated by four *Drosophila* heterogeneous nuclear ribonucleoprotein (hnRNP) family proteins. This was coupled with standard DNA array analysis of the RNA composition of the pre-messenger RNPs containing these factors. Using this combined approach, Rio showed that these related *Drosophila* proteins each bind distinct, but partially overlapping, sets of transcripts and regulate the splicing of different sets of exons. In addition to their importance for understanding these specific regulators, these results constitute very exciting progress in global splicing analysis.

Some regulatory targets of splicing factors are proving to be other splicing factors. Two groups presented work showing that homologs within a family of splicing factors can regulate the expression of one another. Albrecht Bindereif (University of Giessen, Germany) described the properties of hnRNP L and its homolog the L-like protein. HnRNP L targets CA-rich elements that can act as splicing enhancers or silencers depending on their location in introns or exons, respectively. Bindereif and colleagues have identified the L-like protein as a target of L and *vice versa*, where RNAi knockdown of one protein leads to an increase in the other. Similarly, one of us (D.B.) presented analyses of the regulation of the neuronal polypyrimidine tract binding protein (nPTB) by its more widely expressed homolog PTB. This regulation is not simply due to the regulation of nPTB splicing, but also to regulation of the translation of nPTB mRNA. This mRNA is present in most cells, but the protein is only found in certain cell types, most notably neurons. The tissue specificity of protein expression apparently results from the repression of nPTB mRNA translation by PTB in many cell types and, in muscle cells, by microRNAs. It is likely that these systems of cross regulation are just the initial observations of a large network of genetic interactions between splicing factors.

Splicing and human disease

Another area where progress is particularly evident is in understanding the role of splicing in human disease and in applying this understanding to new therapeutic approaches. Tito Baralle (International Center for Genetic Engineering and Biotechnology, Trieste, Italy) described how the effect of mutations in splicing regulatory elements is dependent on genetic background. His team has found that because alternative exons are frequently controlled by multiple elements, mutations in one regulatory sequence may be silent on their own, but can make an exon more dependent on other elements. The complexity of this interplay was further highlighted by Cyril Bourgeois (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France) in regard to splicing of dystrophin exon 31 and by Joerg Gromoll (University of Münster, Germany) for luteinizing hormone receptor (LHR) exon 10. In each case, exonic mutations destroy or create binding sites for splicing regulators that alter the splicing of the exon and cause human disease.

This theme was extended further by several speakers who discussed the link between alternative splicing and tumorigenesis. Mariano Garcia-Blanco (Duke University, Durham, USA) described a system that allows the splicing of particular alternative exons to be visualized in mice, and his group's use of this system to explore changes in the splicing of the fibroblast growth factor receptor FGFR2 during prostate cancer progression. Adrian Krainer (Cold Spring Harbor Laboratory, Cold Spring Harbor, USA) described a wide array of molecular and genomic approaches to show how overexpression of the splicing regulator SF2/ASF leads to

cell transformation by activating both the Ras/MAP kinase and mTOR intracellular signaling pathways. Of particular interest was the identification of SF2/ASF-induced splice variants of S6 kinase and other components in the mTOR pathway, some of which have oncogenic activity on their own. In a complementary talk, Claudia Ghigna (University of Pavia, Italy) described how skipping of exon 11 of the gene for the tyrosine kinase receptor Ron in breast and colon cancer leads to its constitutive activation and to increased cell mobility and invasiveness. In a satisfying counterpart to Krainer's results, Ghigna showed that exon 11 skipping in these cells resulted from increased expression of SF2/ASF.

Given the role of alternative splicing in a multitude of human diseases, there is great interest in the possibility of therapeutic alteration of splicing. Jamal Tazi (University of Montpellier II, Montpellier, France) presented the results of high-throughput screens to identify small molecules that can alter splicing. Molecules identified were shown to alter both spliceosome assembly and the modification of specific splicing factors. Ryszard Kole (University of North Carolina, Chapel Hill, USA) discussed his most recent results using antisense oligonucleotides to alter splicing patterns. He presented new chemical modifications that improve the targeting of oligonucleotides to specific tissues and, importantly, efficiently change the splicing of a variety of therapeutic targets, including beta-globin genes carrying thalassemia mutations, the tumor necrosis factor (TNF) receptor, and dystrophin, the protein that is defective in muscular dystrophy. Dystrophin splicing is a particularly appealing target, because simply inducing exon skipping clearly yields therapeutic benefit. Annemieke Aartsma-Rus (Leiden University, Leiden, The Netherlands) presented work on this system that has identified oligonucleotides that strongly alter dystrophin splicing in model systems; trials are now under way in humans.

Another intensively studied disease where the ability to alter splicing would clearly have therapeutic benefit is spinal muscular atrophy (SMA). This disease is caused by mutations in the gene *SMN* that result in a loss of SMN protein, a ubiquitously expressed protein involved in a process - the assembly of snRNPs - that is common to all cells. In addition to describing new results on the mechanism of snRNP assembly by the SMN protein, Utz Fischer (University of Würzburg, Germany) presented a zebrafish model of SMA. One unanswered question posed by SMA is why the loss of SMN leads specifically to motor neuron degeneration. It was not known whether this specific defect was due to a function of SMN specific to motor neurons, or whether motor neurons are simply more dependent than other cells on SMN for its normal role in snRNP assembly. Fischer described how knocking down SMN expression in zebrafish did indeed lead to a motor neuron defect, which could be rescued by the co-injection of assembled snRNPs. This argues that the degeneration of the motor neurons is due to their special

need for efficient snRNP assembly. Many approaches to altering splicing in disease genes are aimed at inducing exon skipping. This is not the case for SMA, where therapies are needed that will increase the inclusion of exon 7. Julien Marquis (University of Bern, Switzerland) described several strategies for accomplishing this using modified U7 snRNPs targeted to specific sites in the *SMN* transcript. These include attaching a splicing enhancer to exon 7 *in trans*, weakening exon 8 splicing by masking the branchpoint, and improving binding to the nonoptimal exon 7 5' splice site by a mutant U1 snRNA. All of these generated increased splicing in cell culture and are now being tested *in vivo*.

The range of questions being addressed and the variety of techniques described at the meeting made clear that the field of alternative splicing studies is robust and growing (and daunting to review). Important progress has been made in a multitude of directions, but the combination of genome-wide analyses with focused genetic or biochemical assays is proving to be particularly powerful. The meeting spurred very useful dialog between these global and more focused views, and we are looking forward to the next such gathering.