Regulation of alternative pre-mRNA splicing

Louise Woodley and Juan Valcárcel

Date received (in revised form): 15th June 2002

Abstract

The first wave of bioinformatic studies that followed genome and complementary DNA sequencing projects revealed that alternative splicing of messenger RNA precursors (pre-mRNAs) contributes substantially to transcriptome complexity in higher eukaryotes. Together with the realisation of the impact of the process on cell differentiation, development and disease, these studies portray alternative splicing as a fundamental component of gene regulation. Both detailed mechanistic studies and genome-wide analyses will be necessary to unravel the molecular basis for cell type-specific splice site selection. This paper will highlight some recent progress and future challenges for functional genomics and bioinformatics in this rapidly developing area.

PREVALENCE

Most genes in higher eukaryotes contain introns: non-coding sequences that need to be eliminated by splicing together the coding sequences (exons) in order to generate translatable messenger (mRNAs). The splicing process can be regulated such that specific introns can be retained, exons skipped or alternative splice sites used in different cell types or at different stages during development.

Comparison between expressed sequence tag (EST) sequences and either genomic or complementary DNA (cDNA) sequences has revealed that 35–60 per cent of human genes produce multiple transcripts that can be explained by alternative choice of splice sites.

These figures represent a ten-fold increase in the prevalence of alternative splicing compared with previous estimates. Furthermore, as persuasively argued in a recent review, the current estimates are likely to represent minimal values due to the limited coverage of EST libraries and other biases associated with cDNA synthesis. It is therefore reasonable to assume that the majority of human genes are alternatively spliced, an assumption that may soon be independently tested by high throughput proteomic analyses.

It has been argued that alternative splicing could help to explain the discrepancy between the number of human genes estimated by EST clustering (or suspected from proteomic analyses and considerations of organismal complexity) and the significantly lower number of genes predicted from the human genome draft sequence. Could the prevalence of alternative splicing be a more realistic measure of the complexity of an organism than the number of genes? A recent study comparing an equal, arbitrary number of cDNAs and ESTs from six animal species, including Caenorhabditis elegans, Drosophila and humans, did not find a significant difference in the prevalence of alternative splicing when the number of alternatively spliced genes or the number of isoforms per gene were considered. Alternative splicing therefore appears to have made significant contributions to transcriptome and proteome complexity for long evolutionary periods.

This is not completely unexpected, given that a Drosophila gene encoding the cell adhesion molecule Dscam has the potential to generate more than 38,000 different mRNA isoforms — more than twice the number of genes in the fruitfly. Although only a limited fraction of all possible Dscam isoforms has been reported, extensive heterogeneity has been detected in this subset, correlating with expression in specific neurones and...
in response to certain regulatory factors.\textsuperscript{15,16} As the gene encodes a protein important for neural connectivity, and the different isoforms are likely to have distinct protein–protein interaction properties, it is conceivable that alternative splicing of Dscam contributes to the establishment of neural networks.

Given the combinatorial potential achievable, it is not surprising that alternative splicing is particularly prevalent in genes encoding regulatory factors, including cell adhesion molecules, receptors, signalling molecules and transcription factors.\textsuperscript{11,17} It will be interesting for future bioinformatic studies to determine whether the prevalence of alternative splicing within these families — and whether the number and combinatorial potential for isoforms in their orthologue genes — varies along with the complexity of the organisms.

For similar reasons, the incidence of alternative splicing is high in ‘adaptive’ tissues like the nervous or immune systems.\textsuperscript{18} Exploring the genes subject to extensive alternative processing in these tissues may also provide clues about the plasticity of splice site and regulatory signals.

In summary, it has become apparent that to obtain a truly global view of an organism’s coding capacity, massive efforts will be required to identify and characterise all splice variants. Similarly, understanding gene expression regulation will require a more complete knowledge of the mechanisms that control splice site choice. These problems are particularly suited to the fields of bioinformatics, functional genomics and proteomics, and encouragingly in the past few years these approaches have identified an order of magnitude more alternatively spliced genes than had been identified by 20 years of more traditional molecular biology approaches. In this paper, some of the achievements to date and the challenging issues ahead will be outlined. Additional information and discussions on this topic can be found in three excellent recent reviews (refs. 13, 19 and 20).

**BIOLOGICAL RELEVANCE**

Dramatic differences in the functions of the proteins encoded by alternatively spliced transcripts have been documented.\textsuperscript{2} These include genes in which alternative splicing efficiently acts as an on/off switch for expression by introducing premature stop codons in some of the mRNA isoforms. Frameshifts can also be used to produce totally different proteins from the same gene. A common theme is the production of protein pairs with antagonistic activities which, for example, generate either a transcription activator or a repressor, an apoptotic activator or an inhibitor, a membrane-bound receptor or a soluble factor that ‘squelches’ ligands in extracellular space.

The functional consequences of alternative splicing are sometimes more subtle, including enzymes with different allosteric regulation, muscle proteins with different mechanical properties, ion channels with distinct conductivities, etc.\textsuperscript{21} There are genes for which the functional consequences of alternative splicing are essentially unknown.

Although 70–88 per cent of alternative splicing events affect coding regions,\textsuperscript{11} and often the ratio between isoforms is tightly regulated in a cell-type or stage-specific manner and/or altered in disease states,\textsuperscript{5,22} the question of what fraction of alternative splicing is biologically relevant can still be raised. Bioinformatic approaches using large enough datasets and suitable analysis packages, combined with structural information, have the potential to make predictions regarding the function of different isoforms, and therefore they may be able to provide informed estimates about the fraction of relevant alternative splicing events. This issue is particularly difficult to address on a large scale because it may require gene-specific assays to probe subtle differences in activity between protein isoforms.

Three approaches have been used to assess functional differences between alternatively spliced mRNAs. First, isoform-specific knockouts have proved...
useful in establishing isoform-specific functions in vivo. The Wilms’ tumour gene (WT1), for example, is involved in sex determination and kidney development and produces two isoforms that differ by the presence or absence of a KTS tripeptide.\textsuperscript{23} The –KTS isoform is involved in transcription regulation, while the +KTS isoform associates with splicing factors.\textsuperscript{24} Children with Frasier’s syndrome have reduced levels of WT1(+KTS) due to a mutation in the alternative splice site that normally results in KTS inclusion. Selective knockouts of each isoform in mice have provided evidence for both overlapping and distinct functions at the later stages of genitourinary development, as well as a useful model system for Frasier’s syndrome.\textsuperscript{25}

A second method to study the function of different protein isoforms is the use of anti-sense strategies to block one of the alternative splice sites.\textsuperscript{26} This has been used, for example, to block the 5’ splice site of the apoptosis inhibitor Bcl-x\textsubscript{L} so that production of the pro-apoptotic splice variant Bcl-x\textsubscript{S} is increased in prostate and breast cancer cell lines.\textsuperscript{27} This strategy was also used in tissue culture cells to revert an aberrant alternative splicing event in the β-globin pre-mRNA that results in β-thalassaemia.\textsuperscript{28}

Third, RNA interference (RNAi) induced by double-stranded (ds) RNA is becoming a powerful tool for gene inactivation,\textsuperscript{29} and its feasibility for large-scale functional genomics has been recently demonstrated in C. elegans.\textsuperscript{30} The use of short dsRNA molecules to target short sequences within an mRNA specifically\textsuperscript{31} and the apparently catalytic effects observed can make RNAi a particularly useful technique for the assessment of isoform-specific function.\textsuperscript{32,33}

A major motivation to understanding alternative splicing is the aim of characterising and developing therapies for the 15 per cent of human genetic diseases caused by aberrant splicing.\textsuperscript{5,22,34,35} Aberrant splicing can be a direct consequence of splice site inactivation or of the generation of cryptic sites.

When genetic defects — or mutations generated, for example, in the course of tumour progression — are analysed at the level of cDNAs, the incidence of mutations that alter splicing can be higher than 40 per cent. This has been observed in the case of the neurofibromatosis-1, BRCA1 and ATM genes, where nonsense, missense or silent mutations affect the regulatory sequences required for exon definition or cryptic exon silencing.\textsuperscript{33–37} Such effects have been observed in more than 30 disease-related genes.\textsuperscript{35}

Another category of mutations affects alternative splicing indirectly. CTG trinucleotide expansion in the 5’ untranslated region of the DM1 kinase gene is associated with myotonic dystrophy type 1 (DM1). CTG expansion affects the activity of CUG-binding proteins (CUG-BP) as regulators of alternative splicing of the cardiac troponin T (cTNT) and insulin receptor (IR) genes.\textsuperscript{39,40} These alterations can explain, at least in part, aspects of the pathology associated with DM1, including insulin resistance in skeletal muscle.

Multiple changes in alternative splicing have been detected in complex diseases like cancer. These often contribute to pathology (eg by generating autocrine proliferative loops or by promoting cellular metastasis), and often have diagnostic value.\textsuperscript{22,41}

Heterogeneity in the relative levels of alternatively spliced isoforms, due to the effects of single nucleotide polymorphisms, for example, has been associated both with disease states and also with variability in disease phenotypes.\textsuperscript{22,35} Alternative splicing can therefore be a genetic modifier for disease severity.

**GENOMIC TOOLS FOR THE ANALYSIS OF ALTERNATIVE SPlicing**

**Annotation and archiving of alternative splicing**

Alignments of ESTs to genomic or full-length cDNA sequences provide readily
available information about transcript structure and its variations. Frequently used servers include Gene2EST (http://woody.embl-heidelberg.de/gene2est/; Ref. 42) and the Transcript Assembly Program (TAP) (http://sapiens.wustl.edu/~zkan/TAP/; Ref. 43). When applied to large collections of sequences, these alignments provide vast amounts of data on putative alternative splicing events. This creates the problem of how to archive such information meaningfully, and how to assess its validity and functional relevance. Modrek and Lee\textsuperscript{13} have discussed available databases, and additional databases have been released earlier this year. The Putative Alternative Splicing database (http://palsdb.ym.edu.tw/; Ref. 44) aligns the longest mRNA sequence in each Unigene cluster to related Unigene sequences and EST data, and provides tools to assess the quality of alternative splicing predictions, to cluster related ESTs and to find information about their sources and other gene features. The SpliceNest (http://splicenest.molgen.mpg.de; Ref. 45) database aligns consensus sequences of EST clusters to human genomic DNA using an interactive graphical representation.

Alternative splicing databases can be divided into two broad categories: (1) curated, bibliography-based databases, and (2) databases generated by sequence alignments. Curated databases have great descriptive value but are necessarily of limited scope. Databases generated by sequence alignments have the potential of providing a complete picture of the transcriptome in various organisms, but fail to register experimental information and lack further verification.

Ideally, databases should combine the best features of each type of approach, providing information with various degrees of verification, rapid access to EST annotations (eg statistical coverage, tissue of origin, developmental stage) and links to programs for protein and RNA structure prediction (some of these being already implemented in recent database releases). Other interesting additions would be alignments to orthologous genes, identification of known regulatory sequences and the capacity to integrate new high throughput sources of information (eg microarray data, proteomic analysis).

The authors endorse the recent proposal\textsuperscript{13} for integrative databases that will have a strong input from annotation contributed by the scientific community, as has been the case for highly successful protein databases such as SWISSPROT. Although the specific values of individual databases (alignment methods, sources and scope) should remain, a central depository with links to other resources may be essential to catalogue efficiently alternatively spliced variants. This would enable the wider scientific community to attempt to decode information such as tissue-specific splicing patterns, disease isoforms and even the regulatory elements present.

**Microarrays**

Another promising approach to identifying alternative splicing events and quantifying their relative occurrence is the use of microarrays. These are nucleic acids (cDNAs or oligonucleotides) arrayed at high density onto solid surfaces that can hybridise to fluorescently labelled cRNAs or cDNAs such that the fluorescent signal can serve as a measure of the abundance of transcripts corresponding to each of the genes represented in the array.\textsuperscript{46}

Microarrays can be designed to analyse pre-mRNA splicing and its regulation. Clark \textit{et al.}\textsuperscript{47} have recently used microarrays able to distinguish between spliced and unspliced RNA to analyse, at the genomic level in yeast, the impact of mutation of splicing factors. The results identified groups of genes with different dependencies on processing factors as well as novel functional relationships between the processing factors themselves. Alternatively spliced transcripts could be detected using probes that hybridise differentially to each of the isoforms.
Given that alternatively spliced transcripts often differ by relatively short sequences, only oligonucleotide-based microarrays can provide the degree of specificity required to assess alternative splicing. 

Hu et al. showed that current oligonucleotide-based microarray technology can be used, at least to a certain extent, to analyse alternative splicing. Gene expression was monitored in ten different rat tissues by 20 pairs of 25-mer oligonucleotide probes per gene analysed, each consisting of a perfect match and a mismatch control. Algorithms were developed to compare chip hybridisation signals between individual probes in the different tissues. Normalisation of the data revealed variation among individual probes that could be assigned to alternative splicing in 17% of the genes analysed in the ten different tissues. A significant fraction of these putative alternative splicing events was supported by EST alignments and further verified by reverse transcriptase polymerase chain reaction. A limitation of this source of data is that probe design usually focuses on the 3' region of the transcripts, which is better represented when oligo-dT is used to prime cDNA synthesis.

Shoemaker et al. used 60-nucleotide long probes corresponding to 8,183 exons annotated on human chromosome 22 to study systematically their expression in cell lines from normal and disease tissues. Comparison of 69 pairs of conditions using algorithms that detect groups of co-regulated exons established the boundaries of 'expression-verified genes', provided verification for ab initio gene prediction and also provided data consistent with alternative splicing events. One limitation of this approach is that it depends on exon annotation by gene prediction algorithms. As alternative splice sites often have non-consensus splice sites (see below), regulated exons may be more difficult to predict.

A modification of the previous approach has been used to generate a high-resolution picture of the transcripts corresponding to a particular genomic sequence without a priori knowledge of the possible exon content. Tiling arrays consist of 60-mer overlapping oligonucleotides spaced at ten-nucleotide intervals that cover a particular genomic region or, potentially, a complete chromosome. Hybridisation of labelled mRNA revealed boundaries of novel exons with ten-nucleotide resolution, the precise limits being easily assigned thereafter by search for consensus splice sites within the ten-nucleotide window. Although no systematic assessment of alternative splicing was reported, the system has the potential to identify all transcribed regions in a chromosome and their variation by post-transcriptional regulation.

Recently, Yeskley et al. reported microarrays specifically designed to detect alternatively spliced mRNAs. This technology employs two oligonucleotides to detect each alternative splicing event: one oligonucleotide hybridises to sequences upstream of the splice junction, while the other hybridises to downstream sequences. When the two oligos are annealed on the splice junction, they are ligated and the splice junction sequence amplified using additional sequence tags incorporated in the oligos. Detection is accomplished on beads coated with sequences able to identify the products of amplification, loaded onto the tip of optical fibres. The system proved to be able to detect splice variants from 10 pg of total cellular RNA or from as few as ten cells without prior RNA purification, and provided biological insights into the development of choriocarcinomas. Its major current limitation is the requirement for sequence information about each of the splicing events analysed.

UNDERSTANDING ALTERNATIVE SPICING REGULATION

Our ability to predict the pattern of alternative splicing of a gene in a given tissue or physiological situation is very limited at present. Currently, seven
categories of elements have been shown to influence splice site choice. These include sequences in the pre-mRNA, trans-acting (mainly protein) factors and molecular events that integrate splicing with other steps in gene expression. The interplay between these elements can define cell-specific patterns of splicing by affecting the behaviour of the splicing machinery. Splicing complexes (spliceosomes) are composed of five small nuclear ribonucleoprotein particles (snRNPs) and 50–100 proteins not directly associated with the snRNPs. RNA:RNA interactions involving the RNA components of snRNPs (snRNAs) and the pre-mRNA are essential for spliceosome assembly and most likely also for catalysis.

Let us now discuss the factors that influence splice site selection and how genomic tools can be used for their investigation.

**Intrinsic strength of the splice sites**
A measure of the strength of a splice site can be made by comparing its sequence to the consensus of the sequences that surround exon/intron junctions in a particular organism. This can be quantified by comparing the sequence at each position with the frequency with which that nucleotide appears in splice sites. The score at each position then contributes to determine an overall score. Using this criterion, regulated splice sites are often weaker than constitutive sites. A rationale for these observations is that weaker sites are more likely to be dependent on (or sensitive to) regulatory factors recognising other less conserved sequence elements nearby. It is important to point out that several classes of consensus splice sites can be found in the same organism, including those defining the introns removed by a less abundant class of snRNPs which constitutes the minor spliceosome. Statistical analyses can reveal significant correlations between specific constellations of splice site sequences and different modes of splicing regulation, intron length, G+C richness of the genomic locus, etc. These correlations are likely to reveal novel mechanistic insights into splice site recognition.

**Secondary structures in the RNA**
Higher order RNA structure can modulate the access of splicing factors to splicing signals or affect the effective distance between splice sites. Despite the limited value of RNA folding programs in the context of the ribonucleoprotein complexes where splice sites are actually recognized, progress has been reported in predicting 3’ splice sites.

**Regulatory signals**
Sequences present within introns or exons, distinct from the splice sites themselves, can enhance or silence the use of specific splice sites. A classic example is of the purine-rich exonic enhancers, which are bound by members of the arginine/serine-rich (RS) domain-containing splicing factors that promote the use of weak 3’ splice sites. Bioinformatic analysis can be critical to unravelling these elements in four ways. First, once an element is identified and a functional consensus derived, the use of scoring matrices may allow the identification of similar sequences in other genes. A recent example demonstrating the power of this approach was the identification of a single nucleotide that disrupts an exonic enhancer recognised by the splicing factor SF2/ASF. This enhancer is functional in the survival of motor neurone 1 gene (SMN1), but the nucleotide substitution in the related locus SMN2 prevents exon 7 inclusion and the production of functional SMN protein. Absence of SMN by inactivation of SMN1 causes spinal muscular atrophy (SMA), a leading cause of mortality in infants. It is conceivable that SMA pathology could be alleviated if SF2/ASF could be forced to recognise the sub-optimal enhancer in SMN2.

Secondly, novel regulatory elements...
can be identified through statistical analysis of specific sequence contexts. For example, based on the conjecture that exons with non-consensus splice sites are under a stronger selective pressure to retain splicing enhancer elements, novel exonic enhancers were identified by analysing large datasets of human gene sequences enriched in exons (as compared to intronic regions near splice sites) in the neighbourhood of weak splice sites.59

Thirdly, tissue-specific regulatory signals can be identified by comparing genes subject to alternative splicing in those tissues. Thus, signals characteristic of exons specifically included in the brain have been identified using this approach,60 Comparing 25 brain-specific exons and neighbouring introns with a control set of constitutive exons and using a word contrast algorithm, the hexanucleotide UGCAUG was found to be over-represented in the proximal downstream intron of brain-specific exons. It will be important to determine whether brain-specific factors recognise this sequence.

Fourthly, phylogenetic comparisons between orthologous genes in different species can help to identify conserved regulatory elements. This approach, which has provided important insights for the dissection of promoters in transcriptional regulation, can be particularly fruitful for intronic signals, given that the majority of intron sequences are apparently not subject to strong evolutionary pressure. Limitations of these studies include the variability in regulatory mechanisms across species, and the fact that relatively short regulatory signals may be difficult to discover if they are located at non-equivalent positions, often in non-identical exon/intron configurations.

**Tissue-specific trans-acting factors**

Extensive genetic and molecular analysis has established the existence of sex- and tissue-specific splicing regulatory factors in *Drosophila*.2 The protein Sex-lethal (SXL), for example, is only present in female flies and regulates alternative splicing of target genes to induce the production of female-specific isoforms. By doing so, SXL coordinates all aspects of sexual differentiation, X-chromosome dosage compensation and sexual behaviour in fruitflies.61 Tissue-specific factors have been also documented in mammalian cells. The proteins NOVA-1 and a neurone-specific isoform of the polypyrimidine tract-binding protein (nPTB) have been shown to trigger neurone-specific patterns of alternative splicing.18,62–64

Most of the tissue-specific factors identified so far have characteristic protein motifs (RNA-binding domains of the RNA Recognition Motif (RRM) or hnRNP K-Homology (KH) type, RS domains, etc.). Searches for these motifs63,64 in transcriptome and proteome databases can be useful in order to identify new regulators, particularly if combined with information about their patterns of expression.

**Changes in ratios between general splicing factors and hnRNP proteins**

Variations in the concentration of proteins with constitutive functions in pre-mRNA splicing can induce changes in alternative splicing.5,65 An early paradigm of the field is that antagonism between members of two families of
proteins, hnRNP and SR proteins, often serves to modulate the relative use of competing splice sites. Microarrays and other forms of expression analysis, including, ultimately, proteomic analysis, can help to generate a global view of the relative levels of expression of known regulators in particular cell types. Combined with analyses to identify the occurrence of binding sites for these factors within a particular pre-mRNA, these tools may become useful in predicting patterns of alternative splicing in different cell types or physiological situations.

An important concept emerging from detailed studies of mechanisms of splicing control is that regulation is achieved by protein complexes rather than individual factors. These complexes are often assembled on signals that contain adjacent, usually short sequence elements bound by individual components of the complex. This design is likely to be functionally important: as complex assembly depends on multiple, cooperative recognition steps, each step can be the target of regulation, thus providing combinatorial potential to the system.

The architecture of regulatory complexes has two consequences for in silico identification of regulatory sequences. First, prediction of binding sites for regulatory factors gains further credibility when several signals cluster within a relatively small region of the pre-mRNA. Second, as protein–RNA interactions are strongly influenced by cooperative effects mediated by protein–protein interactions within the complexes, individual binding sites do not need to be of high affinity/specificity (as are those resulting from SELEX experiments using purified factors). Bioinformatic tools able to gauge the relevance of a putative, sub-optimal binding site on the basis of the context of other nearby sequence elements may become instrumental in identifying splicing enhancers and silencers.

Post-translational modifications/signalling pathways
Phosphorylation/dephosphorylation of splicing factors has been shown to modulate their activity and to be important for splice site selection. Activation of signal transduction pathways can affect alternative splicing, by, for example, affecting the nucleo-cytoplasmic distribution of hnRNP and therefore their effective concentration in the nucleus.

Cis-acting regulatory sequences that connect alternative splicing regulation with specific signalling pathways have been identified. Alternative splicing of the attachment molecule CD44 plays an important role in tumour progression, and a composite signal has been found that couples ras activation with particular splicing events. Similarly, a sequence element able to induce exon skipping in response to Ca2+/calmodulin-dependent protein kinase IV activation has been described. As this pathway is activated upon membrane depolarisation and affects ion channel pre-mRNAs, it may implement changes in gene expression that contribute to complex processes such as learning and memory.

It is likely that response elements characteristic of other signalling pathways will be found in the near future. The next challenge will be to decode the chain of molecular events leading to their recognition or activation upon the presence of the relevant stimulus.

Integration with other steps of gene expression
It is becoming increasingly apparent that many of the individual steps of gene expression are coordinated and influence each other. At least some of these effects are due to physical association between the enzymatic complexes that direct transcription, capping, splicing, polyadenylation, editing and export of the mature mRNA to the cytoplasm. Recent work indicates that alternative splicing can be influenced by the structure
of the promoter directing the synthesis of the primary transcript.\textsuperscript{57} These effects could be based on the assembly of factors with dual function in splicing and transcription\textsuperscript{79,80} or be related to the processivity of transcript synthesis.\textsuperscript{80,81} Combination of promoter sequence analyses and alternative splicing could eventually establish correlations with predictive value. Further insights into the mechanisms of coupling between alternative splicing and transcription (or other steps in gene expression) are likely to be required to facilitate these predictions.

**PERSPECTIVES**

The authors have argued for the need for comprehensive efforts to catalogue splice variants in ways that will combine curated databases with purely bioinformatic predictions. Microarray and proteomic analyses should become instrumental to quantify and provide functional information for known alternative splicing events, and should also become essential for validation of predicted products. Recent/imminent developments in bioinformatics may provide useful tools to define, categorise and help understand the function of splicing regulatory signals.

The functional consequences of alternative splicing cannot always be inferred from the differences between the predicted protein products. The design of high throughput methods to knock out specific isoforms of a gene should be in high demand in the near future. Although obviously limited by the availability of relevant functional assays, RNA interference and anti-sense strategies, together with genomic engineering, will provide essential tools for studies of isoform function.

Finally, a combination of genomic-wide research tools and detailed mechanistic analyses will be required to understand the molecular logic underlying the regulated selection of splice sites in different cells. These studies will be essential to draw a sensible picture of eukaryotic gene regulation and, ultimately, to be able to predict the occurrence and tissue-specificity of alternative splicing events.

**Acknowledgments**

We thank Richard Copley, Rodric Guigó, Nuno Morais, colleagues at the EMBL Gene Expression Programme and an anonymous reviewer for useful comments on the manuscript. Work in our laboratory is supported in part by a grant from the Human Frontier Science Program.

**References**


Regulation of alternative pre-mRNA splicing


