Evolutionarily Conserved Intronic Splicing Regulatory Elements in the Human Genome

Eric L Van Nostrand, *Stanford University, Stanford, California, USA*

Gene W Yeo, *Salk Institute, La Jolla, California, USA*

Advanced article

Article Contents

Online posting date: 15th July 2008

Splicing is the process by which introns are removed from premature messenger ribonucleic acid (pre-mRNA), and exons are ligated to form mature mRNA before translation into protein products. Aside from the consensus splice signals such as the branch point, acceptor and donor splice sites, additional *cis*-regulatory elements embedded within the exons and introns control the recognition of exonic (approximately 150 bases) from intronic sequence (approximately 100–100 000 bases). Intronic splicing elements that are evolutionarily conserved across multiple species play important roles in the regulation of constitutive splicing, where a single mRNA is produced, and alternative splicing, where multiple mRNA isoforms are generated.

Introduction

Pre-mRNA (premature messenger ribonucleic acid) splicing is a critical step by which the genetic information of higher Eukaryotes encodes for the full spectrum of expressed proteins. Once a pre-mRNA is transcribed it undergoes the highly regulated process of splicing, consisting of the removal of intronic regions and sequential joining of exonic regions, which results in the production of mature mRNA transcript that will specify the synthesis of a particular protein isoform. Although most exons are constitutively spliced, meaning that the exon is always recognized and spliced in an identical manner, it has recently become clear that up to 70% of genes contain alternatively spliced exons, in which there is variable recognition of exon–intron boundaries in a tissue-, condition-, developmental stage- or disease-specific manner. **See also**: mRNA Splicing: Regulated and Differential; Splicing of pre-mRNA

Recognition of the boundaries of an intron is often depicted as the identification of a 5′ (donor) and 3′ (acceptor) splice site, which is equivalent to describing the regulation of transcription as being dictated solely by the transcription start site. However, much like transcription is controlled in a context-specifc manner by the combinatorial interplay of a wide array of transcription factors and other regulatory proteins, exon-recognition is also tightly regulated by the orchestrated interactions of numerous RNA-binding proteins (RBPs) that recognize *cis*-regulatory sequence elements in both the exon and flanking introns. The primary demarcation of exon–intron boundaries is achieved by the recognition of the 5′ donor and 3′ acceptor splice sites, with additional specificity conferred by the presence of the branch point and poly pyrimidine tract upstream of the 3′ splice site itself. The 5′ splice site, with consensus sequence GURAGU is typically bound by the U1 small nuclear ribonucleoprotein (snRNA); the branch point by the U2 snRNA; the poly pyrimidine tract (a 18–23 nucleotide C/U-rich region) and 3′ splice site (with consensus sequence UUUCAG) by the U2AF splicing factors. In the case of some exons, these four sequence elements are sufficient to recruit the splicing machinery and accurately splice the desired introns. However, computational analysis of the frequency of these canonical splice site motifs indicates that a significant number of introns contain 'decoy' splice sites with weak matches to the consensus sequence. Thus, proper splice site recognition must require auxiliary splicing factors that can recognize specific sequence motifs (elements) and act to repress or enhance splicing at specific loci. Although the importance of such auxiliary elements was recognized early on, it remained difficult to characterize the full set of regulatory elements for many years due to the complexity of identifying

ELS subject area: Evolution and Diversity of Life

How to cite:
Van Nostrand, Eric L; and, Yeo, Gene W (July 2008) Evolutionarily Conserved Intronic Splicing Regulatory Elements in the Human Genome. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0021005

functional nucleotides in the human genome. **See also**: Messenger RNA Splicing Signals

With the availability of multiple genomes, we can now search for functional regulatory elements with the assumption that these elements are more likely to be preserved through millions of years of evolution. In recent years, this approach has been combined with technologies that allow high-throughput identification of tissue-specific alternative splicing (AS) events to greatly speed up the search for splicing regulatory elements, enabling us to better understand the complex regulation of tissue-specific AS in humans. **See also**: Exonic Splicing Enhancers

Identification of Splicing Regulatory Elements

The degeneracy and length of these auxiliary sequence motifs (typically 5–10 bases for known splicing factors) makes them difficult to identify. **Figure 1** summarizes the major approaches that have proven successful in identification of intronic splicing regulatory elements. As depicted in **Figure 1a**, many of these sequences were previously found by first identifying exons that are alternatively spliced or misspliced, followed by cloning of the exon and flanking introns into a mini-gene construct. Mutations within sequences in the exon and surrounding intronic regions coupled with an assay for the amount of splice site usage can reveal the *cis*-regulatory elements controlling splicing of the exon. Alternatively, one can begin with a specific splicing factor, but with no knowledge of either the binding site or targets that are regulated by the factor (**Figure 1b**). In this case, SELEX (systematic evolution of ligands by exponential enrichment) can be performed to identify its strongest binding site. In this method, the splicing factor is first allowed to bind to a random library of short RNA oligonucleotides (approximately 20–30 nt). Immunoprecipitation with an antibody specific for the splicing factor will reveal RNA sequences bound by the factor. The eluted RNA can then be amplified (by reverse transcriptase-polymerase chain reaction (RT-PCR) and T7 transcription) and used as the starting material for a new round of SELEX. After multiple rounds, the RNA output will be highly enriched for sequences that are preferentially bound by the splicing factor. This method was applied to discover new classes of exonic splicing enhancers (Coulter *et al*., 1997; Liu *et al*., 1998).

Early computational efforts to identify *de novo* intronic *cis*-regulatory elements of tissue-specific AS achieved some success despite using small sets of sequences. For example, using a small set of human exons that had been previously described in the literature as specifically included in brain tissues, but skipped in other tissues, Brudno *et al*. were able to identify general sequence characteristics of flanking introns, including pyrimidine-rich sequences in the upstream introns that resembled the putative binding site for the *PTB* and *nPTB* splicing factors, and a lack of GGG motifs in the downstream intron. They additionally identified a UGCAUG hexamer as enriched in downstream introns, and further found an enrichment downstream of a second set of 12 exons specifically included in muscle tissue as well (Brudno *et al*., 2001).

However, it quickly became clear that such approaches were neither cost- nor time-effective for large-scale identification of new auxiliary elements. With the completion of the human genome sequence, new bioinformatics strategies to identify splicing sequence elements were introduced. As depicted in **Figure 1d**, genome-wide high-throughput identification of constitutive and alternative spliced exons was made possible by the large-scale sequencing of complementary deoxyribonucleic acids (cDNAs) (full-length mature mRNA transcripts) and ESTs (expressed sequence tags; short approximately 500–600 nt fragments of a mature mRNA transcript).

Early computational strategies focused on auxiliary elements located in exons, which could be identified by studying the relationship between certain six-nucleotide sequences and the 'strength' of the splice site, where this strength was measured as the match of the splice site to the canonical human splice sites (Fairbrother *et al*., 2002; Zhang and Chasin, 2004). Majewski and Ott used a variety of features to identify various sequence properties of internal exons and introns, including a rediscovery of the GGG triplet as playing a role in intronic splicing regulation (Majewski and Ott, 2002). This GGG triplet, as well as C-rich motifs, was also found to be significantly enriched in introns relative to exons, as well as in introns with weak (nonconsensus) 5′ or 3′ splice sites as compared to introns with strong (consensus) splice sites (Yeo *et al*., 2004). A similar analysis comparing exons with 'pseudo exons' (intronic regions flanked by canonical splice sites that are not recognized by the splicing machinery) identified C-rich and TG-rich motifs in upstream introns and C-rich and G-triplet motifs in downstream introns as significantly different between these two sets (Zhang *et al*., 2003).

Large-scale experimental approaches did not lag behind, with researchers making use of splicing-sensitive constructs (in which either inclusion or exclusion of an exon could activate a fluorescent marker) to screen for sequences that, when inserted into these constructs, could cause a shift in exon usage (**Figure 1c**). One such effort identified 133 unique exon splicing silencer decamers, which were further clustered into 7 core motifs. These motifs were then further experimentally confirmed to confer silencing activity when inserted into exons (Wang *et al*., 2004). Computational analysis indicated a significant enrichment for these sequences in the hundreds of intronic sequences flanking exons, suggesting that the presence of these elements in introns may help to silence the usage of improper splice sites located in these introns.

Multiple Genomes Come in Handy

The availability of multiple vertebrate genomes has enabled the systematic search for human DNA and RNA

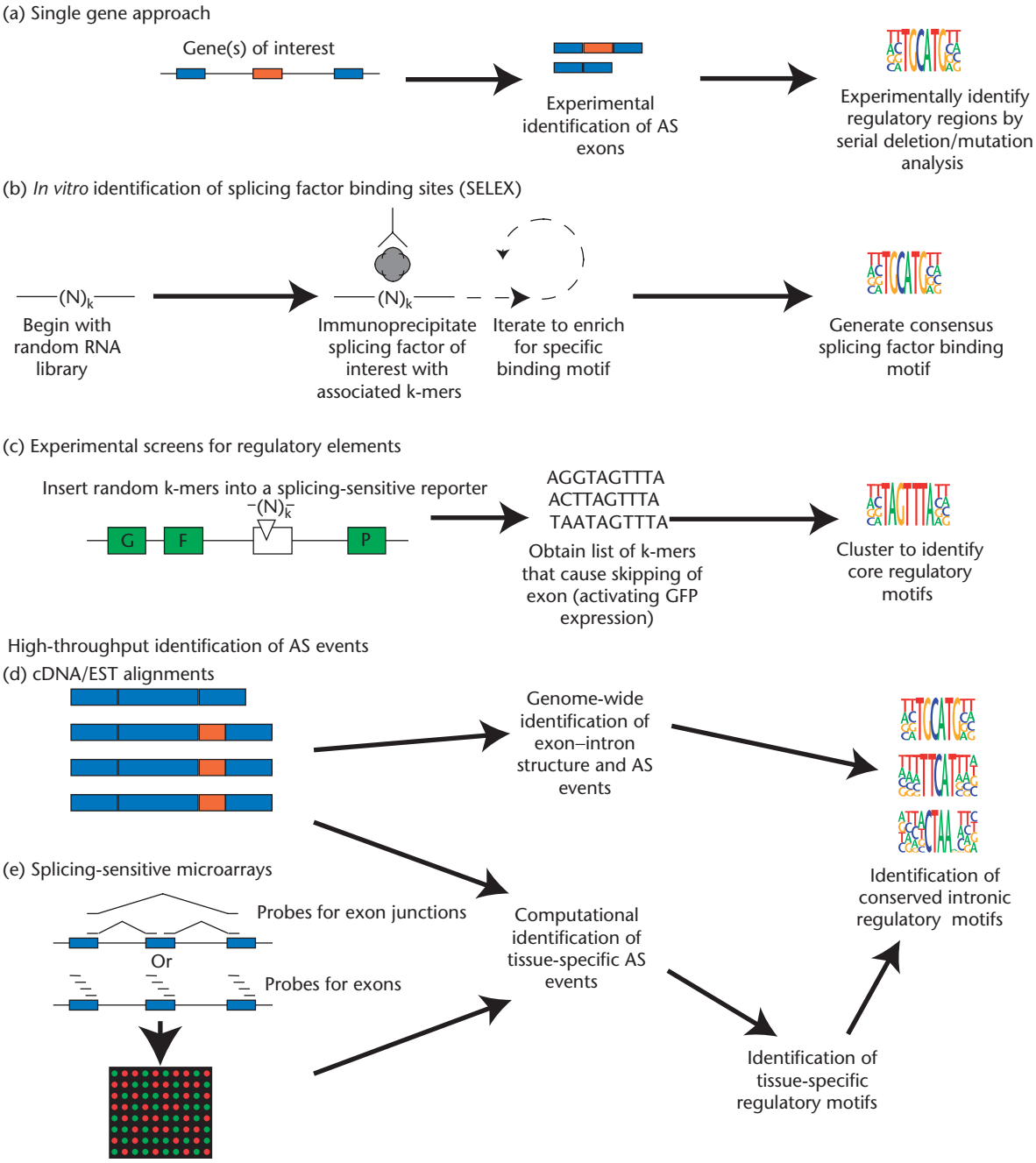

Figure 1 Various methods through which conserved (or nonconserved) intronic splicing regulatory elements can be identified. (a) Early work focused on identifying AS events for a specific gene of interest. Once an event was identified, the researchers could mutate or delete various intronic regions to identify specific regulatory regions that were critical for regulation of splicing of that specific exon. (b) Systematic evolution of ligands by exponential enrichment (SELEX) can be performed to identify the most preferred binding site of a splicing factor. A consensus binding site can be generated from the enriched RNAs after several rounds of SELEX. (c) Experimental screens for splicing regulatory elements can be conducted by inserting random k-mers into a splicing sensitive reporter, which enables the selection of k-mers that caused exon-skipping or inclusion. The list of k-mers can then be clustered to identify core regulatory elements. (d) The sequencing of mature mRNA transcripts isolated from various human tissues, either by full-length sequencing (cDNAs) or by sequencing expressed sequence tags (ESTs), alternatively spliced events can be identified in an unbiased genome-wide manner. These events could then either be directly analysed for enrichment of specific intronic motifs, or could be used to identify smaller subsets of apparent tissue-specific events that could then yield tissue-specific intronic elements. (e) Recent microarray technologies have allowed high-throughput identification of AS events either by probing for the presence of specific exon junctions (top) that would indicate skipping of an internal exon, or by probing for the expression of each individual exon (bottom), where exon skipping will cause a single exon to have far lower expression than other exons in this gene. By repeating this analysis on multiple tissue samples tissue-specific AS events can thus be identified in an automated and high-throughput matter, and can then be analysed for intronic regulatory motifs in the same manner described in (d).

regulatory elements, based on the premise that blocks of nucleotide sequence in noncoding regions that have been conserved over millions of years are likely to be functional. Early on, most efforts focused on the conservation of general sequence features; in 2003, Sorek *et al*. used the newly released mouse genome to quantify the increased conservation of introns flanking AS events, where 77% of conserved AS exons were flanked by long conserved intronic sequences compared to only 17% of constitutive exons. In fact, these intronic regions were more conserved than many promoter regions, suggesting the presence of functional regulatory elements (Sorek and Ast, 2003).

On a small scale, various researchers have used conservation as a key feature in identifying regulatory elements for small sets of alternative exons. An early comparative genomics study identified intronic elements regulating the splicing of an alternative exon in the pre-mRNA for the splicing factor hnRNP (heterogeneous nuclear ribonuclearprotein) A1 by sequencing this region in mouse and aligning to the known human sequence. Further experiments identified a short conserved intronic element as critical for regulation of AS, and that this region was in fact bound and regulated by hnRNP A1 protein in an apparent autoregulatory loop (Chabot *et al*., 1997). Other groups made use of small sets of annotated AS events along with conservation to further confirm motifs that had been identified using human sequence alone (Miriami *et al*., 2003; Minovitsky *et al*., 2005). The role of the putative Fox-1/Fox-2 binding element UGCAUG that had previously been identified using only 25 brain-specific AS events (Brudno *et al*., 2001) was confirmed using a modified set of 27 exons that were again specifically included in brain and skipped in other tissues. In this new set, the authors confirmed that there indeed was a significant enrichment for conserved UGCAUG sites in the downstream introns of these exons, further confirming a critical role for Fox-1/Fox-2 regulation of splicing in the brain (Minovitsky *et al*., 2005).

Although computationally identifying novel regulatory sequences can yield high false positives, the use of conservation provides a way to filter away much of the nonfunctional sequence, vastly enriching for sequences that mediate some regulatory function. With this rationale in mind, searching for conserved regions within intronic regions proximal to exons should similarly increase the likelihood that these sequences are involved with splicing regulation. Yeo *et al*. used intronic conservation between human and mouse to aid in the identification of 5-nt sequence motifs that distinguished alternative exons from constitutive exons (Yeo *et al*., 2005). Using a training set of approximately 240 conserved AS events (exons which were annotated as being alternatively spliced in both human and mouse) and approximately 5 000 conserved constitutive exons, dozens of motifs were identified as significantly enriched or depleted in introns flanking exons in the conserved AS set. Although these motifs were not subjected to further analysis, it was noted that many appeared similar to previously characterized splicing regulatory elements. Also in 2005, Han *et al*. used alignments between the human and mouse genomes to associate the presence of exonic UAGG elements and GGGG elements in the 5′ splice site region with exon skipping, where 18.75% of exons containing both of these elements were annotated as skipped exons by EST and cDNA evidence, as opposed to only 4.6% of control exons. Interestingly, GGGG presence in bases 3–10 of the intron by itself was associated with exon skipping, with 7.8% of exons with this motif annotated as skipping events compared to 4.6% of those without a GGGG motif (Han *et al*., 2005).

Several comparative genomics approaches have also been applied to identify exonic regulatory elements that affect alternative and constitutive splicing in mammals (Goren *et al*., 2006), and intronic regulatory elements that are proximal to alternative exons in worms (Kabat *et al*., 2006). Leveraging the information contained in multiple genome alignments of four mammalian species (human, mouse, rat and dog), Yeo *et al*. used 200 nucleotides of sequence for all annotated internal exons to identify a set of 5- to 7-nucleotide intronic sequence motifs that were significantly more conserved than would be expected by chance in upstream and downstream flanking introns, respectively, suggesting that these motifs might be playing a functional regulatory role (Yeo *et al*., 2007a). After clustering highly similar sequences into a subset of core motifs, there were found to be 156 upstream and 158 downstream motifs, each of which was then further characterized by a battery of computational approaches. Many motifs were found to differ dramatically in their localization pattern within introns, with some only found near the splice sites while others were distributed more evenly across the entire flanking intron regions. In addition to a high degree of overlap with known intronic splicing regulatory motifs, these motifs were also validated in a splicing-sensitive construct in 293 T cells *in vivo*, where nearly 70% showed differential splice site usage. Further analysis indicated that genes with exons flanked by these motifs were commonly characterized by distinct patterns of tissue specificity, and often functioned in the same biological process (using Gene Ontology). Using a dataset of AS events from EST and cDNA transcripts, many of these intronic regulatory elements were also found to be enriched in introns flanking AS events, in agreement with their hypothesized role as regulators of AS. Many of these motifs were also enriched in intronic regions flanking tissue-specific AS events identified by splicing-sensitive microarrays, indicating the critical role that conserved intronic motifs may play in regulation of tissue-specific splicing. In a complementary study, Voelker and Berglund identified 4- to 7-nucleotide sequences that were significantly enriched for conservation in RefSeq-annotated internal introns (Voelker and Berglund, 2007). Using a graph clustering algorithm, they combined 7-mers with highly related sequences to identify 63 conserved motifs and 85 conserved motifs in the downstream and upstream flanking intron sets, respectively. The motifs identified by this approach were similarly enriched for previously identified intronic splicing regulators, and many of these conserved motifs were found to predominantly be

located in introns flanking either only AS events or only constitutive events.

Splicing-sensitive Microarrays Provide Context-specific Information

Identification of intronic conserved sequence elements that regulate AS in a tissue-specific manner has been greatly aided by the high-throughput identification of tissue-specific splicing events by splicing-sensitive microarrays (**Figure 1e**). There are currently two major microarray technologies that allow the identification of AS events, which detect events either directly (using probes that cross an exon–exon junction to directly assay for transcripts that skip a specific exon) (Johnson *et al*., 2003) or indirectly (by measuring the amount of every exon in the transcript individually, and then computationally identifying an exon with significantly different signals than the rest of the exons in that gene) (Clark *et al*., 2007). Once a set of tissue-specific AS events has been identified by this method, it is then possible to go back and identify putative regulators by searching for intronic motifs that are significantly enriched in introns flanking these events.

For example, Sugnet *et al*. performed microarray analysis in 22 adult mouse tissues to identify 171 brain-specific and 28 muscle-specific AS events (Sugnet *et al*., 2006). Using regions in introns flanking these alternative exons that were conserved in the human orthologous introns, they identified a significant enrichment for the Fox-1/Fox-2 binding motif GCAUG and the Nova-1/2 binding motif UCAUY, as well as a depletion of the Polypyrimidine tract binding proteins 1 and 2 (PTB1/PTB2) binding motif CUCUCU, in introns downstream of exons specifically included in brain. These datasets were also used to identify two novel conserved intronic motifs, including a branch point-like CUAAC motif that was significantly enriched downstream and depleted upstream of exons specifically included in muscle. A parallel study profiling 27 mouse cells and tissues by splicing-sensitive microarrays identified 110 nervous system-specific AS events, which were again characterized by high levels of conservation in the flanking introns (Fagnani *et al*., 2007). Motif analysis identified 26 motifs significantly enriched in these 110 exons and flanking introns, 17 of which were more frequently conserved in human sequence than expected by chance. These motifs included a number of C/U-rich sequences that are similar to the putative PTB1/PTB2 binding site, as well as a modest enrichment for Nova-1/2 and Fox-1/2 binding motifs. A number of novel motifs, including a CUAAUNC motif in the downstream intron region that resembles the consensus branch point sequence, were also identified by this *ab initio* search. Das *et al*. performed a similar analysis using 16 human tissues, which they used to identify 56 exons which were specifically included in muscle tissue (Das *et al*., 2007). Considering all possible hexamers, the authors identified 62 hexamers which were significantly correlated with the level of exon inclusion in muscle, 12 of which were also conserved in multiple species. Three known motif families were identified as enriched, including the Fox-1/2 and CUG-BP1 and ETR3 like factor (CELF) binding motifs in downstream introns and PTB-like binding sites in upstream introns, as well as a number of novel motifs. Many of these motifs were further shown to have specific positional biases within the introns, with most Fox-1/2 and CELF sites located within the first 100–200 bases of the downstream intron. Another recent study performed splicing-sensitive microarrays in human embryonic stem cells (hESCs) and neural progenitor cells, to characterize splicing regulation during neuronal differentiation (Yeo *et al*., 2007b). Using a novel outlier-based detection approach, the authors identified 1737 internal exons predicted to undergo AS in neuronal progenitors compared to hESCs. Performing an *ab initio* search for enriched pentamers that could potentially regulate splicing in these tissues identified 144 motifs enriched in the introns flanking exons included in neuronal progenitors and skipped in hESCs, and 52 motifs enriched in introns flanking exons skipped in neuronal progenitors but included in hESCs. These motifs included a number of sequences which resemble previously characterized splicing factor binding sites, including U-, GU- and CU-rich motifs upstream of exons included in neuronal progenitors and skipped in hESCs that resemble the putative binding sites for splicing factors hnRNP C, Embryonic Lethal, Abnormal Vision (ELAV)-type RNA binding protein 3 (ETR-3) and PTB1/PTB2, respectively. The putative Fox-1/2 GCAUG motif was also identified as enriched downstream of both sets of skipping events, suggesting that the Fox splicing factor family may play a general role in regulation of exon skipping in both of these cell-types.

Intronic Regulatory Elements Aid in *De novo* Alternative Splicing Prediction

In 2004–2005, two groups demonstrated that conserved sequence elements within introns were a key feature in the prediction of novel AS events. Dror *et al*. showed that using the frequency of certain triplet sequences in upstream and downstream introns as a feature greatly increased their ability to correctly predict whether an exon would be skipped (Dror *et al*., 2005). In parallel, Yeo *et al*. utilized dozens of 5-nt sequence elements to perform *de novo* prediction of exon skipping events, many of which were similar to known splicing factor binding motifs (Yeo *et al*., 2005). We expect similar progress to be made in predicting context-specific AS in the near future.

References

Brudno M, Gelfand MS, Spengler S, Zorn M and Dubchak I (2001) Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Research* **29**(11): 2338–2348.

Chabot B, Blanchette M, Lapierre I and La Branche H (1997) An intron element modulating 5′ splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Molecular and Cell Biology* **17**(4): 1776–1786.

Clark TA, Schweitzer AC, Chen TX *et al*. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biology* **8**(4): R64.

Coulter LR, Landree MA and Cooper TA (1997) Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Molecular and Cell Biology* **17**(4): 2143–2150.

Das D, Clark TA, Schweitzer A *et al*. (2007) A correlation with exon expression approach to identify *cis*-regulatory elements for tissue-specific alternative splicing. *Nucleic Acids Research* **35**(14): 4845–4857.

Dror G, Sorek R and Shamir R (2005) Accurate identification of alternatively spliced exons using support vector machine. *Bioinformatics* **21**(7): 897–901.

Fagnani M, Barash Y, Ip JY *et al*. (2007) Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biology* **8**(6): R108.

Fairbrother WG, Yeh RF, Sharp PA and Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* **297**(5583): 1007–1013.

Goren A, Ram O, Amit M *et al*. (2006) Comparative analysis identifies exonic splicing regulatory sequences: the complex definition of enhancers and silencers. *Molecular Cell* **22**(6): 769–781.

Han K, Yeo G, An P, Burge CB and Grabowski PJ (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biology* **3**(5): e158.

Johnson JM, Castle J, Garrett-Engele P *et al*. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**(5653): 2141–2144.

Kabat JL, Barberan-Soler S, McKenna P *et al*. (2006) Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLoS Computational Biology* **2**(7): e86.

Liu HX, Zhang M and Krainer AR (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & Development* **12**(13): 1998–2012.

Majewski J and Ott J (2002) Distribution and characterization of regulatory elements in the human genome. *Genome Research* **12**(12): 1827–1836.

Minovitsky S, Gee SL, Schokrpur S, Dubchak I and Conboy JG (2005) The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Research* **33**(2): 714–724.

Miriami E, Margalit H and Sperling R (2003) Conserved sequence elements associated with exon skipping. *Nucleic Acids Research* **31**(7): 1974–1983.

Sorek R and Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Research* **13**(7): 1631–1637.

Sugnet CW, Srinivasan K, Clark TA *et al*. (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Computational Biology* **2**(1): e4.

Voelker RB and Berglund JA (2007) A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Research* **17**(7): 1023–1033.

Wang Z, Rolish ME, Yeo G *et al*. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell* **119**(6): 831–845.

Yeo G, Hoon S, Venkatesh B and Burge CB (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proceedings of the National Academy of Sciences of the USA* **101**(44): 15700–15705.

Yeo GW, Nostrand EL and Liang TY (2007a) Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genetics* **3**(5): e85.

Yeo GW, Van Nostrand E, Holste D, Poggio T and Burge CB (2005) Identification and analysis of alternative splicing events conserved in human and mouse. *Proceedings of the National Academy of Sciences of the USA* **102**(8): 2850–2855.

Yeo GW, Xu X, Liang TY *et al*. (2007b) Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLoS Computational Biology* **3**(10): e196.

Zhang XH and Chasin LA (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes & Development* **18**(11): 1241–1250.

Zhang XH, Heller KA, Hefter I, Leslie CS and Chasin LA (2003) Sequence information for the splicing of human pre-mRNA identified by support vector machine classification. *Genome Research* **13**(12): 2637–2650.

Further Reading

Ladd AN and Cooper TA (2002) Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biology* **3**(11): reviews0008.1–reviews0008.16.