The evolving roles of alternative splicing

Liana F Lareau[1], Richard E Green[1], Rajiv S Bhatnagar[2,3] and Steven E Brenner[1,2]*

Alternative splicing is now commonly thought to affect more than half of all human genes. Recent studies have investigated not only the scope but also the biological impact of alternative splicing on a large scale, revealing that its role in generating proteome diversity may be augmented by a role in regulation. For instance, protein function can be regulated by the removal of interaction or localization domains by alternative splicing. Alternative splicing can also regulate gene expression by splicing transcripts into unproductive mRNAs targeted for degradation. To fully understand the scope of alternative splicing, we must also determine how many of the predicted splice variants represent functional forms. Comparisons of alternative splicing between human and mouse genes show that predominant splice variants are usually conserved, but rare variants are less commonly shared. Evolutionary conservation of splicing patterns suggests functional importance and provides insight into the evolutionary history of alternative splicing.

Addresses
Departments of [1]Molecular and Cell Biology, and [2]Plant and Microbial Biology, University of California, Berkeley, California 94720, USA
[3]Department of Dermatology, University of California, San Francisco, California 94143, USA
*e-mail: brenner@compbio.berkeley.edu

Current Opinion in Structural Biology 2004, **14**:273–282

This review comes from a themed issue on
Sequences and topology
Edited by Peer Bork and Christine A Orengo

Available online 19th May 2004

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2004.05.002

Abbreviations

Clk	CDC-like kinase
EST	expressed sequence tag
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
NMD	nonsense-mediated mRNA decay
PTB	polypyrimidine tract binding protein
PTC	premature termination codon
RUST	regulated unproductive splicing and translation
SR	serine/arginine-rich
TMV	tobacco mosaic virus
TNFR	tumor necrosis factor receptor
UTR	untranslated region

Introduction

It is now well established that many, if not most, human genes are alternatively spliced. Although estimates vary between studies, and increase as more data become available (reviewed in [1]), the prevalence of alternative splicing raises intriguing questions. How much of the observed alternative splicing is functional? What roles does alternative splicing play? How did alternative splicing evolve and what selective advantages does it confer?

The definition of functional splicing has evolved to match our increased understanding of gene expression. At the same time that we realized the human genome has far fewer genes than expected, we began to appreciate the full extent of alternative splicing. Many concluded that alternative splicing led to proteome expansion, bridging a perceived complexity gap. However, complexity is not determined simply by proteome size; it also encompasses interactions and regulation. Recently, we have begun to better understand alternative splicing as a regulatory process, contributing to biological complexity through its ability to control the expression of proteins. A recent paper provides a modern definition of functional splicing: ''An mRNA variant can be defined as being 'functional' if it is required during the life-cycle of the organism and activated in a regulated manner'' [2••]. In some cases, functional splice forms may not even be required in their own right, but their production is required to regulate active protein levels. Moreover, the meaning of 'required' can be generalized by defining functional splicing as that which conveys a selective advantage.

Well before alternative splicing was known to be widespread, studies showed that control of splicing could act as a general on/off switch to regulate gene expression [3]. Work on mRNA stability in *Caenorhabditis elegans* indicated that alternative splicing of serine/arginine-rich (SR) proteins, themselves involved in alternative splicing, could regulate their expression [4]. Recently, the combination of large-scale studies, enabled by the large data sets now available (Table 1), and studies of individual alternatively spliced genes has shed light on the regulatory roles and evolution of alternative splicing.

Large-scale studies of the regulatory impact of alternative splicing

The first large-scale studies of alternative splicing were inventory-style analyses, designed to determine the extent of alternative splicing. The results of those surveys, that alternative splicing is common in many eukaryotes, have inspired a spate of recent studies addressing the regulatory impact of alternative splicing.

Table 1

A selection of alternative splicing resources.

Database and URL	Description	References
NCBI Reference Sequences (RefSeq) Project http://www.ncbi.nih.gov/RefSeq/	Continuously updated database of confirmed and predicted genes, alternative isoforms and genome sequence from several organisms	[73]
Alternative Splicing Database (ASDB) http://hazelton.lbl.gov/~teplitski/alt/	Alignments of all annotated alternatively spliced DNA sequences from GenBank and protein sequences from Swiss-Prot	[74]
Alternative Splicing Annotation Project (ASAP) http://www.bioinformatics.ucla.edu/ASAP/	Human EST to genome mapping based on UniGene clusters with tissue specificity information	[75]
Alternative Splicing Database (ASD) Project http://www.ebi.ac.uk/asd/	Three separate human alternative splice data sources that differ by source of evidence. Data are organized by exon	[76]
Putative Alternative Splicing (PALS) Database http://binfo.ym.edu.tw/passdb/	Human EST to genome mapping based on UniGene clusters	[77]
Intron Sequence and Information (ISIS) Database http://isis.bit.uq.edu.au/front.html	Searchable database of intron sequence for several organisms	[78]
ProSplicer http://bioinfo.csie.ncu.edu.tw/ProSplicer/	Searchable database of UniGene EST clusters mapped to genome with supporting DNA and protein sequence to genome mappings	[79]
SpliceDB http://www.softberry.com/berry.phtml?topic=splicedb	Database and composition statistics for mammalian splice sites inferred from ESTs	[80]
spliceNest http://splicenest.molgen.mpg.de/	Searchable database of EST cluster to genome mappings from four organisms	[81]
Intronerator http://www.soe.ucsc.edu/~kent/intronerator/	Database and display tools for alternative splicing in *C. elegans*	[82]
Extended Alternatively Spliced EST Database (EASED) http://eased.bioinf.mdc-berlin.de/	EST-inferred alternative events in nine eukaryotes with coverage and tissue information	[83]
Plant Alternative Splicing Database (PASDB) http://pasdb.genomics.org.cn	Database of literature reporting alternative splicing events in *Viridiplantae*	[84]
Manually Annotated Alternatively Spliced Events (MAASE) http://splice.sdsc.edu	Searchable database of user-contributed alternative splicing events	[85]

Several studies ask whether alternative splicing occurs preferentially in certain classes of genes. An early analysis of 50 randomly chosen alternatively spliced human genes found that many were specific to the immune and nervous systems, and were commonly involved in signal transduction [5]. However, the FANTOM2 analysis of alternatively spliced mouse transcripts found that these functional categories were no more prevalent among alternatively spliced genes than other genes [6]. The FANTOM2 analysis did find that gene ontology (GO) terms [7] related to enzymes or enzymatic functions were underrepresented among alternatively spliced mouse genes, but an analysis of the alternative splicing described in reviewed RefSeq entries found that alternatively spliced genes function mostly as enzymes, signal transducers and receptors [8]. In addition to these apparently conflicting results, the usefulness of these studies is limited by the lack of a clear biological hypothesis. If alternative splicing is found to preferentially affect a certain functional class of gene, it could be explained in several different ways. For example, there may have been duplication of gene families whose ancestral members happen to have been alternatively spliced or a recent functional proliferation might have made use of alternative splicing for protein diversity.

A related genre of large-scale analyses has also been performed: classification of the functional components of genes affected by alternative splicing. In many cases, understanding which components are affected (transmembrane regions, signal peptides, catalytic domains, etc.) and how they are affected (added, removed, altered) can provide important and easily interpretable biological conclusions. Taking this approach, Resch *et al.* [9] identified a set of functional domains that are removed more often than would be expected given the frequencies of each domain and of alternative splicing, including KRAB, ankyrin repeat and tubulin-binding domains. 56% of the members of this functionally diverse set fall into the higher level category of protein-interaction domains, yet this is no more than would be expected, given that this category is very common (64% of Swiss-Prot domains studied). They further showed, by way of several interesting examples, how this type of analysis leads to testable hypotheses regarding these isoforms' regulatory functions, as described below.

A final regulatory theme that has been explored in large-scale studies is localization, especially membrane association. Several genes have been shown to produce both

soluble and membrane-associated protein isoforms by selective incorporation of exons encoding transmembrane regions. In particular, several tumor necrosis factor receptor (TNFR) genes are alternatively spliced in a regulated manner to produce either soluble or membrane-associated signal-transducing forms [10–15]. Xing *et al*. [16] showed that alternative inclusion/exclusion of membrane-spanning regions is a common outcome of alternative splicing, a result that was corroborated in a mouse-specific study [17].

Detailed studies of alternative splicing reveal regulatory functions ranging from molecular interaction to cell apoptosis

Pioneering, detailed studies of specific alternatively spliced genes revealed that splicing could be used to regulate biological processes. Among the best-characterized classic examples are *Drosophila* Sex lethal, which binds its own pre-mRNA in female flies to repress inclusion of a male-specific exon that would introduce a premature stop codon (reviewed in [18]), and human Bcl-x, which is spliced to form either a short, pro-apoptotic protein or a long, anti-apoptotic protein [19]. These classic biochemical investigations have been followed by hundreds more that use a wide variety of methods to discover and detail the implications of alternative splicing for specific genes. As mentioned above, Resch *et al.* [9] demonstrate how appropriately designed, broad studies can lead to very specific and interpretable results. For example, they showed that, for the genes encoding the Kruppel family of transcription factors, alternative splicing preferentially removes the protein-interaction domain(s) and not the DNA-binding domains. Thus, the full-length isoforms are predicted to bind DNA and be responsive to interaction partners. Shorter isoforms that lack the protein-interaction domain(s) may bind DNA, but would be insensitive to interaction partners [9].

In many cases, alternative isoforms differ by small alterations of functional elements or domains, not by complete removal of functional elements [20]. In these cases, consideration of protein isoforms as simply the sum of their parts may not be sufficient to predict the biology of the resultant isoforms. For example, it was recently shown that Piccolo (Aczonin), a cytomatrix component of the presynaptic active zone, is alternatively spliced to remove a nine-residue region within its Ca^{2+}-binding C_2A domain [21•]. The shorter Piccolo isoform binds Ca^{2+} with higher affinity, but is incapable of the Ca^{2+}-dependent dimerization that occurs in the longer form. As revealed by NMR analysis, the nine extra residues of the long form unexpectedly displace a β strand present in the short form, providing a structural basis for the different biochemical properties of these isoforms.

Similar studies were performed to investigate the structural basis of the ligand-binding specificity of fibroblast growth factor receptor 2 (FGFR2) isoforms [22•,23]. Normal mammalian development and homeostasis depend on correct, specific interactions between the 23 known fibroblast growth factors (FGFs) and their receptors. In humans, several craniosynostosis syndromes, including Apert syndrome, are caused by mutations in FGFRs that abolish or alter ligand specificity [24]. For FGFR2, ligand specificity is achieved in part via tissue-specific alternative splicing of part of one of its three Ig-like domains, D3. FGFR2 co-crystal structures with FGF2 and FGF10 revealed that the alternatively spliced region of D3 lies in a binding cleft containing critical determinants of ligand specificity. These studies of Piccolo and FGFR2 underscore the importance of detailed structural analysis of the diversity generated by alternative splicing.

Over the past several years, scores of detailed studies have shown how alternative splicing is used in many organisms to regulate a great diversity of essential biological processes. For example, programmed cell death of lymphocytes must be tightly regulated for proper immune system function. In unactivated lymphocytes, which are insensitive to apoptotic stimulant, the alternatively spliced TNFR gene TNFRSF25 expresses at least 10 mRNA isoforms with truncated open reading frames (ORFs) [15]. After lymphocyte activation, a precise splicing shift results in production of a single, full-length isoform harboring the death domain, which renders the cells susceptible to apoptosis.

Another life-or-death regulatory role for alternative splicing is in host–pathogen interactions. In tobacco, regulated expression of *N* gene isoforms confers resistance to tobacco mosaic virus (TMV) infection [25]. In response to TMV infection, tobacco generates a stereotypical temporal profile of ratios of *N* alternative isoforms. This profile is necessary for effective TMV response. Expression of either isoform, or both, without the correct temporal regulation fails to confer proper TMV resistance. Conversely, in the final stages of the adenovirus replication cycle in human cells, the host splicing machinery is hijacked to produce the late category of adenovirus gene products via alternative splicing [26,27].

Finally, a virtually unexplored arena of alternative splicing regulation is deciphering roles for the roughly 20% of alternative isoforms that do not change protein coding potential. Many large-scale studies filter these in an early preprocessing step. Alternative mRNAs differing only in their 5′ untranslated region (UTR) are common and their expression may be regulated through alternative promoter usage [28•]. 3′ UTR elements may control mRNA subcellular localization, stability and translational efficiency [29,30•], although the signals responsible for this layer of regulation are incompletely understood.

RUST: regulated unproductive splicing and translation

One common outcome of alternative splicing is down-regulation of the function of a gene by the production of non-functional isoforms of the active gene product. This can be accomplished by the alteration of domains necessary for catalysis, localization or association with other macromolecules, as discussed above. A distinct mechanism is the use of alternative splicing to induce nonsense-mediated mRNA decay (NMD) [4,31,32]. Gene expression can be regulated post-transcriptionally by the production of splice forms that will be degraded by NMD rather than translated into protein, a process we have termed regulated unproductive splicing and translation (RUST) [33••,34•].

NMD is an RNA surveillance function that recognizes mRNAs containing premature termination codons (PTC^+ mRNAs) and targets the transcripts for destruction rather than translation into protein (reviewed in [35]). Nonsense and frameshift mutations, errors in pre-mRNA processing and alternative splicing are among the many sources of PTC^+ mRNAs.

One example of the role of NMD in clearing the products of mis-splicing is observed in the splicing of FGFR2 (described above). Mutually exclusive inclusion of exons IIIb or IIIc, which encode critical determinants of ligand-binding specificity, is necessary for directional signal transduction. The cell creates correct mRNAs with a single alternative exon, but also produces mRNA species containing both or neither. These erroneous products are frameshifted and contain PTCs [36]; they appear to be degraded by NMD. In this case, NMD helps ensure the fidelity of directional signal transduction by preventing potentially harmful by-products of splicing from being translated.

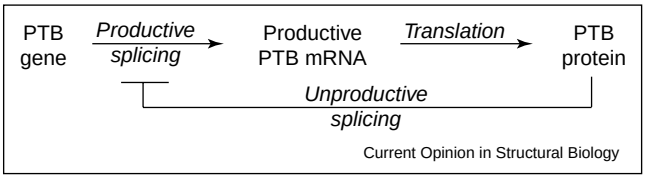
The process of NMD is conserved in all eukaryotes in which it has been examined (reviewed in [37]), but many details of PTC recognition and mRNA degradation remain to be elucidated. The mechanism of PTC recognition differs between mammals and *Saccharomyces cerevisiae*, and it is not as well established how PTCs are recognized in the transcripts of *Drosophila melanogaster* and *C. elegans* [38••]. In mammals, recognition of PTCs is intimately related to pre-mRNA splicing: a termination codon is recognized as premature if it is more than about 50 nucleotides upstream of the site of removal of an intron [35].

The elucidation of this simple rule determining whether a human mRNA is subject to NMD has allowed large-scale computational analyses of alternatively spliced transcripts. Studies of alternative splicing inferred from expressed sequence tags (ESTs) revealed that 45% of human alternatively spliced genes produce at least one mRNA isoform that contains a PTC and thus is a target of NMD [33••]. A similar computational screen of annotated alternatively spliced protein isoforms within Swiss-Prot revealed that 7.9% of 1463 human alternatively spliced genes amenable to analysis have at least one isoform that is predicted to undergo degradation by NMD [34•]. The surprisingly large proportion of alternative splicing that appears to destine mRNAs for degradation rather than translation led to the proposal that RUST is a widespread and generally applicable mechanism of post-transcriptional gene repression.

The kinetics and efficiency of NMD, and the amount and disposition of protein that may be translated before mRNA decay are incompletely understood and are active areas of research [35,39••]. The accuracy of computational predictions of the effect of NMD, generated from simple models of PTC recognition, is thus limited; Neu-Yilik *et al.* [40] review the limitations of this model in predicting NMD targets and functional consequences. Nonetheless, several recent studies have added to a growing list of experimentally characterized cases that validate the proposed broad role of RUST gene regulation. Polypyrimidine tract binding protein (PTB), a regulator of alternative splicing, was shown to autoregulate its protein level through RUST (Figure 1). When PTB protein levels increase, the protein alters splicing of its own pre-mRNA, favoring exclusion of exon 11 and thereby introducing a frameshift and thus a PTC. This PTC-containing mRNA is present at very low steady-state levels, which increase approximately tenfold when Upf1, an essential component of NMD, is knocked down by RNA interference (RNAi) [41••].

Similarly, the human alternative splicing regulatory SR-like protein TRA2-BETA autoregulates its level by inducing inclusion of a PTC-containing exon in its mRNA when the protein level is increased [42]. Another example is TIAR, a protein that affects regulation of both translation and alternative splicing events. TIAR is down-regulated by inclusion of a PTC-containing exon, forming a labile mRNA that is stabilized by treatment with the NMD inhibitor cycloheximide, suggesting degradation

Figure 1



Autoregulation of PTB expression. The PTB gene can be spliced productively to yield PTB protein. Rising levels of PTB protein favor unproductive splicing of PTB to generate an mRNA isoform that contains a PTC. This mRNA is then degraded by NMD. In this way, PTB autoregulates its own expression [41••].

by NMD. In this case, RUST is not autoregulatory. The PTC^+ splice form has not been demonstrated to be induced by TIAR protein itself, but by the related TIA-1 protein [43].

RUST may also occur in plants, although the process of NMD is not well characterized in the plant kingdom. The *Arabidopsis thaliana* AtGRP7 gene, which encodes an RNA-binding protein that can regulate alternative splicing, is under transcriptional regulation with circadian oscillation. Its expression is also post-transcriptionally autoregulated; AtGRP7 protein binds to its pre-mRNA (as well as the pre-mRNA of the related RNA-binding protein AtGRP8) and activates a cryptic splice site within the single intron (of each AtGRP7 and AtGRP8), resulting in a transcript with a PTC. The transcripts are rapidly degraded in a process that can be inhibited by cycloheximide [44].

Several well-characterized examples of RUST, including those described above, as well as autoregulation of the SR protein SC35 [45], involve genes that are explicitly involved in regulation of splicing. Whether this regulatory mechanism is used disproportionately by splicing factors or whether this observation reflects acquisition bias remains to be seen. Observations from our laboratory (D Soergel, SE Brenner, unpublished) suggest that RUST affects many protein groups other than splicing factors. Other examples, such as ribosomal proteins in *S. cerevisiae* [31] and *C. elegans* [46], and AUF1, a modulator of RNA stability [47], implicate RUST in regulation of proteins generally not involved in splicing. The ABCC4 gene, a multidrug resistance associated transporter, recently has been shown to undergo alternative splicing to produce isoforms containing one or two PTC-containing exons. These PTC^+ transcripts are degraded by NMD [48].

In several of the examples discussed above, analysis of orthologous and paralogous sequences suggests that splicing to generate PTC^+ alternative isoforms, and thus RUST regulation, is conserved across species and across protein families. For PTB, the sequence and upstream regulatory elements of alternatively included PTC-containing exon 11 are very similar in the *Fugu rubripes* ortholog of the analyzed human gene, as well as in a human neuronal-specific paralog [41••]. Mouse and monkey orthologs of human ABCC4 share highly conserved PTC-containing exons that are orthologous to the alternatively included exons of ABCC4 [48]. The chicken ortholog of the human TIAR gene likewise has an orthologous PTC^+ exon [43].

Strong evidence of conservation of RUST is found in the CDC-like kinases (Clks), which phosphorylate SR proteins to regulate splicing [49]. Alternative splicing to exclude exon 4, introducing a frameshift and PTC, is conserved among three human paralogs (Clk1–3), the three orthologs of these genes in mouse and even the sole copy of the gene in the sea squirt *Ciona intestinalis* [34•]. For Clk1, RUST appears to regulate the protein through an indirect feedback mechanism. Clk1 has been shown to indirectly modify splicing of its own transcript, most likely through phosphorylation of SR proteins [50]. Thus, as a variation of the autoregulatory circuits described above, increased Clk1 activity results in changes in the activity of one or more SR proteins, resulting in altered splicing of Clk1 pre-mRNA to favor a PTC^+ transcript that is predicted to undergo NMD [34•].

Autoregulation of Clk1 through regulation of SR proteins probably affects the splicing of numerous other SR protein substrate pre-mRNAs. Conversely, Clk1 splicing may be regulated by conditions such as the phosphorylation state of other splicing factors. As noted above, not all RUST involving splicing regulators is autoregulatory; TIAR/TIA-1 and AtGRP7/AtGRP8 illustrate examples of heterologous RUST regulation. Splicing factors that have been demonstrated to be autoregulated by RUST may well be subject to RUST triggered by heterologous factors as well. Indeed, the splice regulator proteins raver1 and CELF4 have been demonstrated to affect the splicing decision for PTB exon 11 [41••], and SRp46 can induce production of the PTC^+ splice variant of SC35 mRNA [45]. The ability to use alternative splicing to regulate the factors that control alternative splicing of other gene products suggests the possibility of regulatory cascades of alternative splicing. An upstream splicing (or phosphorylation) event might lead to changes in the splicing of a few splicing regulatory factors that can affect a broad range of substrates. Thus, RUST may represent both a mode of regulation *by* alternative splicing and a mechanism of regulation *of* alternative splicing. Such a cascade of regulated splicing events would enable limited upstream input to effect a large change in the proteome.

Function versus noise

How much of the alternative splicing predicted by ESTs is functional, rather than noise? Some noise is merely data contamination, but noise can also be of biological origin. Rare transcripts representing splicing errors could arise from mutations disrupting splice signals. Mutations in intronic regions can have significant effects on splicing, yet are invisible after intron removal. Errors could also come from malfunction of the splicing machinery itself. Though the spliceosome distinguishes true splice sites from nearby cryptic sites with impressive fidelity, splice site recognition is a complex process and errors must occur at some low rate. EST libraries contain millions of transcript sequences, so even extremely rare events may be represented. Further noise could come from neutral splicing decisions, with no significant effect on the cellular function [51].

To avoid noise from data contamination, the reliability of the data set can be increased by filtering ESTs with poor alignment to the genome sequence, checking for good donor and acceptor splice sites, and ignoring ESTs likely to represent incompletely processed transcripts, as summarized by Modrek and Lee [52]. Some studies also filter splicing events that are seen only in a few ESTs. All filtering can exclude legitimate splice forms, including rare forms or forms produced by events that are indistinguishable from incomplete processing, such as intron retention [53]. An extension of this concept uses a statistical model to see whether a splice form is likely to occur above some frequency cutoff, based on its occurrence in the observed data [54]. Further support for function comes from tissue or temporal specificity of an isoform [55]. Specific appearance of an isoform is often taken as evidence that its production is regulated and this regulation could suggest function. However, differential expression need not be functional. The cell might have a background level of producing non-functional mRNAs; different conditions in various tissues could lead to differences in this background splicing.

Conservation and evolution of alternative splicing

With genome sequence and EST libraries now available for multiple organisms, many recent studies focus on cross-species conservation of alternative splicing, particularly between human and mouse. Evolutionary conservation over long time periods suggests positive selection and implies function; thus, it can be used both to study the evolution of alternative splicing and to filter noise. The flurry of reports in the past year seem at first glance to introduce more uncertainty, as estimates of the conservation of alternative splicing between human and mouse range from 8% [54] to 98% [56••]. To untangle the situation, it is important to consider the data sets, methods and definitions used by each group (Table 2).

The most straightforward approach to studying conservation of alternative splicing uses EST information from two species to identify equivalent alternative splice events in orthologous genes [57]. EST libraries are incomplete, so this method underestimates the extent of conservation. Two groups have increased the estimate of conserved alternative splicing by using statistics to extrapolate from genes with high EST coverage [54,58]. Missing EST data can also be recovered by examining the underlying genome sequence. If ESTs show inclusion of a cassette exon in human, but not mouse, examination of the genomic sequence might show that the exon sequence is indeed conserved in the correct location, with acceptable splice sites [56••,59,60]. This does not confirm that the exon is ever included; it could represent a recent loss of splicing. Some groups choose to rely solely on experimentally characterized splicing events, not ESTs, leading to small, high-confidence data sets that can be studied in detail but may not be completely representative [61].

Many of the analyses distinguish between the major or predominant splice form, and the minor or rare splice form of a gene. These distinctions are based on the number of ESTs found for each splice form; although precise definitions vary, 'major' is often defined as the splice form found in at least half the ESTs for the gene

Table 2

Studies of alternative splicing conservation between human and mouse.

Study	Evidence	Scope	Observed conservation	Notes
Sugnet *et al*. [57]	EST	All events	~27% of human genes with alternative splicing have a conserved alternative form in mouse	Overall, 10% of human genes have conserved alternative splicing
Kan *et al*. [54]	EST	All events	61% of major form events 8% of minor form events	Statistical extrapolation increases to 99% major, 42% minor forms conserved
Thanaraj *et al*. [58]	EST	All events	67% of major form events 15% of minor form events	Statistical extrapolation increases to 74% major, 61% minor forms conserved
Modrek and Lee [56••]	EST, genome sequence	Cassette exons	98% of major form events 25% of minor form events	Major form exons are conserved at the same level as constitutive exons
Nurtdinov *et al*. [61]	Annotated splice events, genome sequence	All events	69–83% of events	Half of alternatively spliced genes have at least one non-conserved isoform
Resch *et al*. [60]	EST, genome sequence	All events	12% of cassette exons (EST evidence only)	Conserved alternative splicing is likely to preserve frame, especially minor forms
Sorek *et al*. [2••]	EST, genome sequence	Cassette exons	25% of events	Conserved alternative exons are much less likely to have a frameshift or stop

and 'minor' encompasses all others. Though the coverage in EST libraries may not show the true expression level of a splice form, results from multiple independent studies show that there are meaningful differences between these classes of alternative splice forms.

Despite differences in the actual frequencies observed by different groups, there is overwhelming evidence that minor forms are less conserved than major forms. Conservation of major forms can be as high as 98%, equivalent to the conservation level of constitutive exons, in the case of cassette exons [56••] and overall seems to be at least two-thirds [54,58]. Minor forms are only conserved a fraction, perhaps one-quarter, of the time [54,56••,58]. Although this may indicate that some minor forms do indeed represent noise, minor forms are generally supported by multiple ESTs and thus may represent legitimate species-specific splicing. On the other hand, statistical extrapolation shows that conservation of minor forms may be as high as 60% [58] and further EST evidence may confirm that a substantial number of minor forms are conserved. More evidence that the distinction between major and minor forms is biologically meaningful comes from the observation that the major form in human is usually the major form in mouse as well. Similarly, minor forms in one species are generally also minor forms in the other species [56••].

The results also show that alternative splicing conserved between species differs from non-conserved alternative splicing in significant ways. Conserved alternative splicing events are much more likely than random to preserve frame, especially in minor forms, whereas constitutive exons and non-conserved alternative splicing are no more likely to preserve frame than expected randomly [60]. Similar results show that conserved cassette exons are less likely to have stop codons or introduce frameshifts than non-conserved cassette exons. Interestingly, the same result held when looking at non-conserved cassette exons with at least five ESTs [2••].

Also of interest is the ancestral state of alternatively spliced genes, which may not be equivalent to the distinction between major and minor forms in all cases. By aligning alternatively spliced human genes with their prokaryotic and yeast orthologs, Kondrashov and Koonin [62] showed that, in 48 of 73 cases, the longer form was ancestral; this is most often associated with exon skipping in the derived form. Although it may be difficult for derived forms to gain novel coding sequence as exon insertions, some of the insertions may serve to introduce PTCs for regulation by RUST. Cases of ancestral short forms show more evidence of *de novo* emergence than exon duplication [62], although exon duplication does play a significant role in gene evolution [63]. In humans, new exon sequence can also arise from *Alu* repeats inserted in introns. The *Alu* sequence includes motifs that can be recognized as splice sites, causing the repeat to act as an alternative exon [64]. Furthermore, single mutations in these sequences can result in constitutive inclusion [65•]. Although these mutations may be detrimental, they provide a source of potentially advantageous novel forms.

New experimental methods quantify splicing on a large scale

Recent experimental innovations can help confirm that an isoform actually exists and is produced in a regulated way. A microarray using exon junction sequences was used to examine 10 000 human genes in 52 tissues, to determine tissue specificity of alternative splicing events [66•]. Because of practical constraints, only known splice junctions are included on the array. All-exon probe arrays allow limited discovery of new forms. For instance, if an exon is skipped in some conditions, it will be less abundant than constitutive exons of the same gene. Exon probe arrays are still unable to show specific patterns of exon use, such as mutually exclusive exons. Microarrays using a combination of exon and junction sequences have been used to determine the relative abundance of known isoforms [67]. The actual isoforms produced by the *Drosophila* DSCAM gene, a popular example of extensive combinatorial alternative splicing, were assessed using an exon probe microarray and single-cell RT-PCR [68]. Arrays can show the frequencies of different exons or splice junctions, but may not indicate which combinations occur in individual transcripts. Single-molecule methods such as polymerase colonies (polonies) detect each individual mRNA and can show exon use correlations, rare isoforms and novel splice forms [69•]. Unlike EST abundance, which gives a very crude approximation of actual transcript levels, single-molecule experiments can pinpoint expression levels. A transcript produced only once in experiments from hundreds of cells could be identified as a rare mis-splicing event. Other single-molecule techniques, such as 'BEAMing', in which DNA molecules are amplified and attached to magnetic beads for sorting [70], could perhaps be applied to alternative splicing with similar results.

Conclusions

The prevalence of alternative splicing shows that it may play important roles in many biological processes, but many questions must be answered before its true scope and significance are known. Recent work has advanced our understanding of alternative splicing; at the same time, it has raised interesting and important questions.

By affecting regions involved in interaction or localization, alternative splicing can generate isoforms that differ in subtle or dramatic ways. Many of these isoforms have been shown to play key regulatory roles in a diverse array of biological processes. Functional coupling of alternative splicing and NMD, in which transcripts are spliced into forms that are targets for NMD, is another mode of regulation. Examples of this potentially widespread

mechanism of gene regulation can be found in fungi, animals and perhaps plants, indicating either that the process was found in their common ancestor or that it has evolved independently multiple times. Indeed, in *S. cerevisiae*, as many characterized examples of alternative splicing are used for RUST (via regulated intron retention) as are potentially used for proteome diversity [71]. Although some fraction of the predicted human NMD targets probably represent splicing errors, and others may escape NMD and play other roles in the cell [40], the wide conservation of RUST indicates that it may be one of the major roles of alternative splicing.

Although EST data show that more than half of all human genes are alternatively spliced, the fraction that can be considered functional is unknown. Some of the transcripts almost certainly represent biological noise. The best evidence of function comes from individual experiments showing the beneficial effect of regulation, but this approach is impractical or even impossible on a large scale [51]. Even expression data from microarray experiments only provide second-hand evidence of regulation of a given form. In the absence of a direct test of selective advantage, the nearest facsimile is to learn from the experiment of evolution. The studies discussed above show that approximately a quarter of minor alternative splice forms are conserved between human and mouse. What should be concluded about the non-conserved forms? Some may represent species-specific functional splicing; some may be noise. The transcripts representing aberrant splicing may not be functional, but the existence of such noise might be evidence of flexibility of the splicing mechanism that could enable the evolution of new functional forms.

Alternative splicing has almost certainly increased its scope over the course of evolution; its functional importance may have evolved to fill new roles as well. As Jacob noted, nature is a tinkerer [72]. In alternative splicing, which almost certainly evolved by tinkering with the splicing machinery and its modulators, nature has found an efficient tool for tinkering both with proteins and with regulation of gene expression.

Acknowledgements

We would like to thank Emma Hill, Marcin Joachimiak and Nicholas Ingolia for helpful comments. This work was supported by a Searle Scholarship (1-L-110), and National Institutes of Health grants K22 HG00056, T32 GM07127 and T32 HG00047.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Boue S, Letunic I, Bork P: **Alternative splicing and evolution**. *Bioessays* 2003, **25**:1031-1034.

2. •• Sorek R, Shamir R, Ast G: **How prevalent is functional alternative splicing in the human genome?** *Trends Genet* 2004, **20**:68-71.
The authors demonstrate that conserved cassette exons have features distinct from those of non-conserved cassette exons. In particular, conserved cassette exons are much less likely to introduce stop codons or frameshifts. They conclude from this difference that some non-conserved splicing events may represent aberrant splicing.

3. Bingham PM, Chou TB, Mims I, Zachar Z: **On/off regulation of gene expression at the level of splicing**. *Trends Genet* 1988, **4**:134-138.

4. Morrison M, Harris KS, Roth MB: **smg mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans***. *Proc Natl Acad Sci USA* 1997, **94**:9782-9785.

5. Modrek B, Resch A, Grasso C, Lee C: **Genome-wide detection of alternative splicing in expressed sequences of human genes**. *Nucleic Acids Res* 2001, **29**:2850-2859.

6. Zavolan M, Kondo S, Schonbach C, Adachi J, Hume DA, Hayashizaki Y, Gaasterland T: **Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome**. *Genome Res* 2003, **13**:1290-1300.

7. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C *et al.*: **The Gene Ontology (GO) database and informatics resource**. *Nucleic Acids Res* 2004, **32**:D258-D261.

8. Liu S, Altman RB: **Large scale study of protein domain distribution in the context of alternative splicing**. *Nucleic Acids Res* 2003, **31**:4828-4835.

9. Resch A, Xing Y, Modrek B, Gorlick M, Riley R, Lee C: **Assessing the impact of alternative splicing on domain interactions in the human proteome**. *J Proteome Res* 2004, **3**:76-83.

10. Cheng J, Zhou T, Liu C, Shapiro JP, Brauer MJ, Kiefer MC, Barr PJ, Mountz JD: **Protection from Fas-mediated apoptosis by a soluble form of the Fas molecule**. *Science* 1994, **263**:1759-1762.

11. Lainez B, Fernandez-Real JM, Romero X, Esplugues E, Canete JD, Ricart W, Engel P: **Identification and characterization of a novel spliced variant that encodes human soluble tumor necrosis factor receptor 2**. *Int Immunol* 2004, **16**:169-177.

12. Cascino I, Fiucci G, Papoff G, Ruberti G: **Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing**. *J Immunol* 1995, **154**:2706-2713.

13. Michel J, Langstein J, Hofstadter F, Schwarz H: **A soluble form of CD137 (ILA/4-1BB), a member of the TNF receptor family, is released by activated lymphocytes and is detectable in sera of patients with rheumatoid arthritis**. *Eur J Immunol* 1998, **28**:290-295.

14. Tone M, Tone Y, Fairchild PJ, Wykes M, Waldmann H: **Regulation of CD40 function by its isoforms generated through alternative splicing**. *Proc Natl Acad Sci USA* 2001, **98**:1751-1756.

15. Screaton GR, Xu XN, Olsen AL, Cowper AE, Tan R, McMichael AJ, Bell JI: **LARD: a new lymphoid-specific death domain containing receptor regulated by alternative pre-mRNA splicing**. *Proc Natl Acad Sci USA* 1997, **94**:4615-4619.

16. Xing Y, Xu Q, Lee C: **Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains**. *FEBS Lett* 2003, **555**:572-578.

17. Cline MS, Shigeta R, Wheeler RL, Siani-Rose MA, Kulp D, Loraine AE: **The effects of alternative splicing on transmembrane proteins in the mouse genome**. *Pac Symp Biocomput* 2004:17-28.

18. Graveley BR: **Sex, AGility, and the regulation of alternative splicing**. *Cell* 2002, **109**:409-412.

19. Boise LH, Gonzalez-Garcia M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nunez G, Thompson CB: **bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death**. *Cell* 1993, **74**:597-608.

20. Kriventseva EV, Koch I, Apweiler R, Vingron M, Bork P, Gelfand MS, Sunyaev S: **Increase of functional diversity by alternative splicing**. *Trends Genet* 2003, **19**:124-128.

21. Garcia J, Gerber SH, Sugita S, Sudhof TC, Rizo J:
• **A conformational switch in the Piccolo C2A domain regulated by alternative splicing**. *Nat Struct Mol Biol* 2004, **11**:45-53.
This structural study of two alternatively spliced versions of Piccolo showed that the longer form, which differs by the insertion of nine amino acids, has markedly different Ca^{2+}-binding properties. Surprisingly, the extra residues of the long form induce a displacement of a secondary structural element, providing the basis for the biochemical difference between these isoforms.

22. Yeh BK, Igarashi M, Eliseenkova AV, Plotnikov AN, Sher I, Ron D,
• Aaronson SA, Mohammadi M: **Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors**. *Proc Natl Acad Sci USA* 2003, **100**:2266-2271.
A crystal structure of the FGFR2b isoform bound to FGF10 reveals the importance of two b-specific loops in conferring ligand specificity.

23. Plotnikov AN, Hubbard SR, Schlessinger J, Mohammadi M: **Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity**. *Cell* 2000, **101**:413-424.

24. Yu K, Herr AB, Waksman G, Ornitz DM: **Loss of fibroblast growth factor receptor 2 ligand-binding specificity in Apert syndrome**. *Proc Natl Acad Sci USA* 2000, **97**:14536-14541.

25. Dinesh-Kumar SP, Baker BJ: **Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance**. *Proc Natl Acad Sci USA* 2000, **97**:1908-1913.

26. Gama-Carvalho M, Condado I, Carmo-Fonseca M: **Regulation of adenovirus alternative RNA splicing correlates with a reorganization of splicing factors in the nucleus**. *Exp Cell Res* 2003, **289**:77-85.

27. Estmer Nilsson C, Petersen-Mahrt S, Durot C, Shtrichman R, Krainer AR, Kleinberger T, Akusjarvi G: **The adenovirus E4-ORF4 splicing enhancer protein interacts with a subset of phosphorylated SR proteins**. *EMBO J* 2001, **20**:864-871.

28. Zhang T, Haws P, Wu Q: **Multiple variable first exons:**
• **a mechanism for cell- and tissue-specific gene regulation**. *Genome Res* 2004, **14**:79-89.
The authors found that several thousand human genes have variable first exons. The alternative versions of these genes may be under independent transcriptional control.

29. Bratu DP, Cha BJ, Mhlanga MM, Kramer FR, Tyagi S: **Visualizing the distribution and transport of mRNAs in living cells**. *Proc Natl Acad Sci USA* 2003, **100**:13308-13313.

30. Wickens M, Bernstein DS, Kimble J, Parker R: **A PUF family**
• **portrait: 3′UTR regulation as a way of life**. *Trends Genet* 2002, **18**:150-157.
The PUF family of RNA-binding proteins, whose primordial function was probably to sustain mitotic proliferation of stem cells, has evolved to regulate several steps of gene expression.

31. Vilardell J, Chartrand P, Singer RH, Warner JR: **The odyssey of a regulated transcript**. *RNA* 2000, **6**:1773-1780.

32. Hilleren P, Parker R: **Mechanisms of mRNA surveillance in eukaryotes**. *Annu Rev Genet* 1999, **33**:229-260.

33. Lewis BP, Green RE, Brenner SE: **Evidence for the widespread**
•• **coupling of alternative splicing and nonsense-mediated mRNA decay in humans**. *Proc Natl Acad Sci USA* 2003, **100**:189-192.
This large-scale analysis used carefully filtered EST data to reliably infer alternative splicing and provided the first evidence of the widespread production of PTC^{+} alternative isoforms.

34. Hillman RT, Green RE, Brenner SE: **An unappreciated role for**
• **RNA surveillance**. *Genome Biol* 2004, **5**:R8.
This study found >100 human protein isoforms in Swiss-Prot that are likely targets for NMD based on the 50 nucleotide rule. Conservation of the NMD-inducing alternative splicing pattern of Clks in mice and sea squirt suggests that this is a regulated event.

35. Maquat LE: **Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics**. *Nat Rev Mol Cell Biol* 2004, **5**:89-99.

36. Jones RB, Wang F, Luo Y, Yu C, Jin C, Suzuki T, Kan M, McKeehan WL: **The nonsense-mediated decay pathway and mutually exclusive expression of alternatively spliced FGFR2IIIb and -IIIc mRNAs**. *J Biol Chem* 2001, **276**:4158-4167.

37. Frischmeyer PA, Dietz HC: **Nonsense-mediated mRNA decay in health and disease**. *Hum Mol Genet* 1999, **8**:1893-1900.

38. Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E:
•• **Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways**. *EMBO J* 2003, **22**:3960-3970.
This study shows that the system for PTC recognition in flies has evolved differently from that in mammals, in that it is not splicing dependent. Splicing and exon junction complex components are dispensable for NMD in flies.

39. Cao D, Parker R: **Computational modeling and experimental**
•• **analysis of nonsense-mediated decay in yeast**. *Cell* 2003, **113**:533-545.
The authors developed a computational model that parameterizes the key steps in mRNA biogenesis and decay. They used this tool to generate and test predictions about PTC recognition and the polar NMD effect. They found strong support for a model of NMD that includes efficient PTC recognition followed by a position-dependent rate of decapping.

40. Neu-Yilik G, Gehring NH, Hentze MW, Kulozik AE: **Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife**. *Genome Biol* 2004, **5**:218.

41. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA,
•• Smith CW: **Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay**. *Mol Cell* 2004, **13**:91-100.
This carefully conducted study most convincingly demonstrates the combined action of alternative splicing and NMD in autoregulation.

42. Stoilov P, Daoud R, Nayler O, Stamm S: **Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA**. *Hum Mol Genet* 2004, **13**:509-524.

43. Le Guiner C, Gesnel MC, Breathnach R: **TIA-1 or TIAR is required for DT40 cell viability**. *J Biol Chem* 2003, **278**:10465-10476.

44. Staiger D, Zecca L, Wieczorek Kirk DA, Apel K, Eckstein L: **The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA**. *Plant J* 2003, **33**:361-371.

45. Sureau A, Gattoni R, Dooghe Y, Stevenin J, Soret J: **SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs**. *EMBO J* 2001, **20**:1785-1796.

46. Mitrovich QM, Anderson P: **Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans***. *Genes Dev* 2000, **14**:2173-2184.

47. Wilson GM, Sun Y, Sellers J, Lu H, Penkar N, Dillard G, Brewer G: **Regulation of AUF1 expression via conserved alternatively spliced elements in the 3′ untranslated region**. *Mol Cell Biol* 1999, **19**:4056-4064.

48. Lamba JK, Adachi M, Sun D, Tammur J, Schuetz EG, Allikmets R, Schuetz JD: **Nonsense mediated decay downregulates conserved alternatively spliced ABCC4 transcripts bearing nonsense codons**. *Hum Mol Genet* 2003, **12**:99-109.

49. Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI: **The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution**. *EMBO J* 1996, **15**:265-275.

50. Duncan PI, Stojdl DF, Marius RM, Bell JC: ***In vivo* regulation of alternative pre-mRNA splicing by the Clk1 protein kinase**. *Mol Cell Biol* 1997, **17**:5996-6001.

51. Graveley BR: **Alternative splicing: increasing diversity in the proteomic world**. *Trends Genet* 2001, **17**:100-107.

52. Modrek B, Lee C: **A genomic view of alternative splicing**. *Nat Genet* 2002, **30**:13-19.

53. Sakabe NJ, de Souza JE, Galante PA, de Oliveira PS, Passetti F, Brentani H, Osorio EC, Zaiats AC, Leerkes MR,

Kitajima JP *et al.*: **ORESTES are enriched in rare exon usage variants affecting the encoded proteins**. *C R Biol* 2003, **326**:979-985.

54. Kan Z, States D, Gish W: **Selecting for functional alternative splices in ESTs**. *Genome Res* 2002, **12**:1837-1845.

55. Xu Q, Modrek B, Lee C: **Genome-wide detection of tissue-specific alternative splicing in the human transcriptome**. *Nucleic Acids Res* 2002, **30**:3754-3766.

56. •• Modrek B, Lee CJ: **Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss**. *Nat Genet* 2003, **34**:177-180.

The authors used EST evidence to identify reliable alternative splice forms in human, mouse and rat, and then studied the conservation of alternative splicing events in orthologous genes. They show that cassette exons included only in minor splice forms in one species are unlikely to be found in the other species.

57. Sugnet CW, Kent WJ, Ares M Jr, Haussler D: **Transcriptome and genome conservation of alternative splicing events in humans and mice**. *Pac Symp Biocomput* 2004:66-77.

58. Thanaraj TA, Clark F, Muilu J: **Conservation of human alternative splice events in mouse**. *Nucleic Acids Res* 2003, **31**:2544-2552.

59. Sorek R, Ast G: **Intronic sequences flanking alternatively spliced exons are conserved between human and mouse**. *Genome Res* 2003, **13**:1631-1637.

60. Resch A, Xing Y, Alekseyenko A, Modrek B, Lee C: **Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation**. *Nucleic Acids Res* 2004, **32**:1261-1269.

61. Nurtdinov RN, Artamonova II, Mironov AA, Gelfand MS: **Low conservation of alternative splicing patterns in the human and mouse genomes**. *Hum Mol Genet* 2003, **12**:1313-1320.

62. Kondrashov FA, Koonin EV: **Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences**. *Trends Genet* 2003, **19**:115-119.

63. Letunic I, Copley RR, Bork P: **Common exon duplication in animals and its role in alternative splicing**. *Hum Mol Genet* 2002, **11**:1561-1567.

64. Sorek R, Ast G, Graur D: **Alu-containing exons are alternatively spliced**. *Genome Res* 2002, **12**:1060-1067.

65. • Lev-Maor G, Sorek R, Shomron N, Ast G: **The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons**. *Science* 2003, **300**:1288-1291.

The authors previously showed that 5% of human alternatively spliced exons are derived from mobile *Alu* elements. In this study, they analyze the sequences within *Alu* elements that are used as splice sites. They show how minor changes in sequence can cause a shift from alternative to constitutive splicing of *Alu*-derived exons.

66. • Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD: **Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays**. *Science* 2003, **302**:2141-2144.

The authors describe the first large-scale use of microarrays to monitor alternative splicing, using an array incorporating known splice junctions from 10 000 genes. They tested 52 conditions to observe tissue-specific alternative splicing, inferring new forms from the hybridization of transcripts to junctions in some tissues but not others.

67. Wang H, Hubbell E, Hu JS, Mei G, Cline M, Lu G, Clark T, Siani-Rose MA, Ares M, Kulp DC *et al.*: **Gene structure-based splice variant deconvolution using a microarray platform**. *Bioinformatics* 2003, **19(suppl 1)**:i315-i322.

68. Neves G, Zucker J, Daly M, Chess A: **Stochastic yet biased expression of multiple Dscam splice variants by individual cells**. *Nat Genet* 2004, **36**:240-246.

69. • Zhu J, Shendure J, Mitra RD, Church GM: **Single molecule profiling of alternative pre-mRNA splicing**. *Science* 2003, **301**:836-838.

The authors use polymerase colonies (polonies) to amplify and detect individual mRNA molecules. They show how this method can distinguish different combinations of exons in different transcripts and can be used to quantify the levels of inclusion of different exons.

70. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B: **Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations**. *Proc Natl Acad Sci USA* 2003, **100**:8817-8822.

71. Davis CA, Grate L, Spingola M, Ares M Jr: **Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast**. *Nucleic Acids Res* 2000, **28**:1700-1706.

72. Jacob F: **Evolution and tinkering**. *Science* 1977, **196**:1161-1166.

73. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence project: update and current status**. *Nucleic Acids Res* 2003, **31**:34-37.

74. Dralyuk I, Brudno M, Gelfand MS, Zorn M, Dubchak I: **ASDB: database of alternatively spliced genes**. *Nucleic Acids Res* 2000, **28**:296-297.

75. Lee C, Atanelov L, Modrek B, Xing Y: **ASAP: the Alternative Splicing Annotation Project**. *Nucleic Acids Res* 2003, **31**:101-105.

76. Clark F, Thanaraj TA: **Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human**. *Hum Mol Genet* 2002, **11**:451-464.

77. Huang YH, Chen YT, Lai JJ, Yang ST, Yang UC: **PALS db: Putative Alternative Splicing database**. *Nucleic Acids Res* 2002, **30**:186-190.

78. Croft L, Schandorff S, Clark F, Burrage K, Arctander P, Mattick JS: **ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome**. *Nat Genet* 2000, **24**:340-341.

79. Huang HD, Horng JT, Lee CC, Liu BJ: **ProSplicer: a database of putative alternative splicing information derived from protein, mRNA and expressed sequence tag sequence data**. *Genome Biol* 2003, **4**:R29.

80. Burset M, Seledtsov IA, Solovyev VV: **Analysis of canonical and non-canonical splice sites in mammalian genomes**. *Nucleic Acids Res* 2000, **28**:4364-4375.

81. Krause A, Haas SA, Coward E, Vingron M: **SYSTERS, GeneNest, SpliceNest: exploring sequence space from genome to protein**. *Nucleic Acids Res* 2002, **30**:299-300.

82. Kent WJ, Zahler AM: **The intronerator: exploring introns and alternative splicing in *Caenorhabditis elegans***. *Nucleic Acids Res* 2000, **28**:91-93.

83. Pospisil H, Herrmann A, Bortfeldt RH, Reich JG: **EASED: Extended Alternatively Spliced EST Database**. *Nucleic Acids Res* 2004, **32**:D70-D74.

84. Zhou Y, Zhou C, Ye L, Dong J, Xu H, Cai L, Zhang L, Wei L: **Database and analyses of known alternatively spliced genes in plants**. *Genomics* 2003, **82**:584-595.

85. Zheng CL, Nair TM, Gribskov M, Kwon YS, Li HR, Fu XD: **A database designed to computationally aid an experimental approach to alternative splicing**. *Pac Symp Biocomput* 2004:78-88.