

Chapter 10

Analysis of Mutations that Influence Pre-mRNA Splicing

Zhaiyi Zhang and Stefan Stamm

Abstract

A rapidly increasing number of human diseases are now recognized as being caused by the selection of wrong splice sites. In most cases, these changes in alternative splice site selection are due to single nucleotide exchanges in splicing regulatory elements. This chapter describes the use of bioinformatics tools to predict the influence of a mutation on alternative pre-mRNA splicing and the experimental testing of these predictions. The bioinformatic analysis determines the influence of a mutation on splicing enhancers and silencers, splice sites and RNA secondary structures. This approach generates hypotheses that are tested using splicing reporter constructs, which are then analyzed in transfection assays. We describe a recombination-based system that allows for the generation of splicing reporter constructs in the first week and their subsequent analysis in the second week.

Key words: Alternative splicing, mutation, splicing enhancer, splicing silencer, minigene analysis, single nucleotide polymorphisms.

1. Introduction

In the flow of genetic information, pre-messenger RNAs (pre-mRNAs) transcribed from DNA templates carry the information from genomic sequences to protein synthesis. The pre-mRNA is processed, and sequences known as *exons* are incorporated into the mRNA and exported into the cytosol. The remaining intervening sequences, known as *introns*, stay in the nucleus where they are eventually degraded. A pre-mRNA sequence can be recognized as an intron or an exon, depending on the cellular environment. This process is called alternative splicing. Almost all (95%) of human multi-exon genes undergo alternative splicing (1, 2). Unlike promoter activity that predominantly regulates the

abundance of transcripts, alternative splicing influences the structure of the mRNAs and their encoded proteins. As a result, it influences binding properties, intracellular localization, enzymatic activity, protein stability, and post-translational modification of numerous gene products (reviewed in (3)).

1.1. Patterns of Alternative Splicing

Alternative splicing events can be subdivided into five basic patterns, as shown in Fig. 10.1. Exons can be skipped or included, extended or shortened, or included in a mutually exclusive manner; introns can be either removed or retained. Cassette exons (or exon skipping) account for the majority of alternative splicing events conserved between human and mouse genomes. Less frequent are alternatively used 3' and 5' splice sites. Intron retention is the least frequently used pattern and is responsible for less than 3% of the alternative splice events conserved between human and mouse genomes. There are more complex events, such as mutually exclusive events, alternative transcription start sites, and multiple polyadenylation sites. In addition, these basic patterns can be combined resulting in highly complex transcripts (4). An estimated 75% of all alternative splicing patterns change the coding sequence (5), indicating that alternative splicing is a major mechanism for enhancing protein diversity (reviewed by (3)).

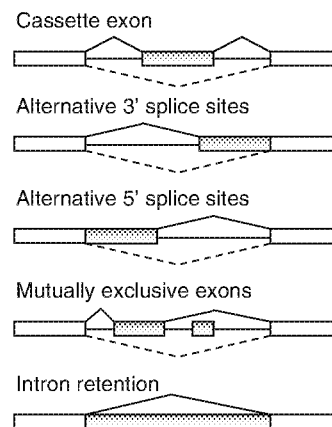


Fig. 10.1. Types of alternative splicing. Flanking constitutive exons are indicated as *white boxes*, and alternative spliced regions are shown in *grey gradient*; introns are shown by *solid lines*. The splicing patterns are indicated.

1.2. Mechanism of Splice Site Selection

The introns that are removed from the pre-mRNA are defined by the 5' and 3' splice sites located at their ends and by the branch point upstream of the 3' splice site (see Fig. 10.2). The molecular mechanism of the splicing reaction that connects these sites has been determined in great detail (6, 7). In contrast, it is not

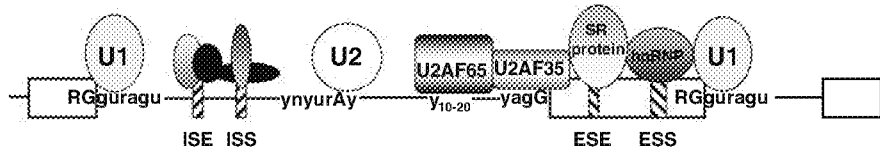


Fig. 10.2. Splicing elements and splicing factors. Exons are indicated as *boxes* and introns as *thick lines*. Splicing regulator elements (enhancers or silencers) are shown as boxes labeled as ESE/ESS in exons or as ISE/ISS in introns. The 5' splice site (RGguragu), 3' splice site (y)₁₀ncagG, and the branch point (ynyrAy) are indicated (r= a or g, y=c or u, n=a, t, c, or g). Two major groups of proteins, hnRNPs and SR or SR-related proteins bind to splicing regulator elements. ISE: intronic splicing enhancer, ISS: intronic splicing silencer, ESE: exonic splicing enhancer, ESS: exonic splicing silencer.

clear how splice sites are recognized in the large background of the pre-mRNA sequences. One major difficulty is the degeneracy of 5' and 3' splice sites and branch point. To precisely identify the relatively small exons and excise large introns, additional regulatory elements play an important role in splice site recognition. They are classified according to their location and their functional effect on splicing as exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), or intronic splicing silencer (ISS). These elements are again characterized by the loose consensus sequences (8). This sequence degeneracy prevents splicing regulatory elements from interfering with the coding capacity of the exons (9). As a result, the accurate splice site selection *in vivo* is achieved through a combinatorial regulatory mechanism by which the exonic or intronic auxiliary elements aid exon recognition by binding to regulatory proteins (10). Proteins binding to regulatory sequence elements can be classified into two groups: serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs). Generally, these proteins not only bind to RNA, but also to other regulatory proteins. The interaction between individual splicing factors and the regulatory sequence is weak, which allows easy dislodging of the proteins from processed RNA after the splicing reaction. As the protein:RNA interaction is weak, different SR and SR-like proteins can act through the same regulatory elements and influence the same splice sites. Higher specificity is achieved by protein-protein interactions that allow simultaneous binding of multiple proteins to RNA. Several SR and SR-like proteins bind to the catalytic component of the spliceosome, e.g., to the U1 and U2 snRNP. Therefore, the transient formation of these protein complexes facilitates splicing complex assembly (11). In addition, SR and SR-like proteins can bridge the introns by interacting with themselves and the classical spliceosomal components (9), (*see Fig. 10.2*). In summary, numerous factors binding to RNA elements influence splice site selection. This degeneracy makes it difficult to accurately predict splice site usage.

1.3. Single Nucleotide Changes Can Influence Splicing Pattern

Single nucleotide changes in the pre-mRNA can disturb the fragile balance of multiple weak interactions governing exon recognition and alter pre-mRNA splicing pattern. Mutations of *cis*-elements can be classified into four categories according to their location and effect. Type I and type II mutations occur in the splice sites. They either destroy known splice sites or create novel ones, which leads to exon skipping or inclusion. Type III and type IV mutations take place in exons or introns, respectively, and alter exon usage. From all the mutations that are annotated in the human genome, about 10% of the 80,000 reported affect canonical splice site sequences (12). This number is most likely an underestimation since most mutations that affect the intronic and exonic splicing elements are not included in the statistic. Exonic nucleotide exchanges that influence splice site selection are often synonymous (13) and polymorphisms located in introns can influence splice site selection (14). Because these single nucleotide polymorphisms (SNPs) do not change the predicted reading frame they were considered ‘noise’ without functional effects. However, detailed analyses of synonymous exonic and intronic mutations revealed a strong effect on splicing that leads to frameshifts, subsequent loss of protein function and human disease (15). Recent array analysis suggests that a large number of SNPs associate with changes in splicing (16). An increased awareness concerning the role of alternative splicing in the etiology of human diseases has led to a strong increase in the number of diseases reported to be associated with changes in alternative splicing (reviewed in (17–21)). It is estimated that up to 50% of the mutations that cause human disease alter the efficiency and pattern of splicing (16). Analysis of splicing mutations causing cystic fibrosis revealed that the splicing pattern caused by a SNP can differ between individuals. This most likely reflects that splice site selection is regulated by multiple factors that work in combination and that a mutated allele responds only to a certain combination. These findings suggest that alternative splicing is a genetic modifier (18, 22) and imply that a large fraction of mutations affect exon usage. A list of single nucleotide mutations in splicing regulatory elements is given in **Table 10.1**. The widespread influence of SNPs on alternative splicing is the reason for their further bioinformatics and experimental analysis.

1.4. Bioinformatics Resources for Alternative Splicing

The completion of numerous genomic sequencing projects has provided a wealth of information revealing the abundant usage of alternative splicing in metazoan organisms. A number of groups have created resources by collecting splice variants and alternative transcript structures. These resources can be classified into two categories: computer and manually generated databases on alternative splicing events and computational tools to decipher the

Table 10.1

Examples of single nucleotide mutations change splicing pattern. The table lists examples of mutations in the intronic or exonic sequences that cause aberrant splicing

Gene	Mutation	Effect	References
ADA	CTGTCCACGEC→CTGTCCACACC	Exon 7 skipping	(31)
ATM	ATTCGAGTG→ATTTGAGTG ACTCAACAT→ACTTAACAT AGgtaa→Agtraa tttagGT→tttaaGT AAGGTTTTA→AAGATTTTA CTCGAAACA→CTCAAAACA	Exon 9 skipping Exon 9 skipping Intron 12 retention Exon 18 skipping Exon 26 skipping Exon 44 skipping	(32)
ATP7A	GATGGAAATC→GATC/AGAATC	Exon 4 skipping	(33)
BRCA1	ATCTTAGAG→ATCGTAGAG	Exon 18 skipping	(34)
FAH	ATGAACGAC→ATCAATGAC AAGCAGGAC→AAGCGGGAC	Exon 8 skipping Exon 9 skipping	(35, 36)
HEXB	GCGCCGGGC→GCGCTGGGC	Exon 11 skipping	(37)
HPRT1	CATGGACTA→CATAGACTA GAACGTCCT→GAACATCCT ATTGTGGAA→ATTATGGAA	Exon 8 skipping	(38)
IKBKAP	gtaagtgc→gtaagegc	Exon 20 skipping	(39)
MAPT (tau)	ATTAATAAG→ATTAAGAAG GGCAGTgtg→GGCAATgtg GATCTTAGC→GATCTCAGC ATAATATCA→ATAACATCA	Exon 10 inclusion	(40, 41)
MLH1	GAGAAGAGA→GAGTAGAGA	Exon 12 skipping	(42)
NF1	CTTAAGAAC→CTTAAAAAC	Exon 7 skipping	(43)
SMN2	GGTTTTAGAC→GGTTTCAGAC	Exon 7 skipping	(44)

Large letters indicate exonic sequences; small letters indicate intronic sequences. The sequence on the left side of the arrow is wild type, and the right side is mutant. The bold letter indicates the single nucleotide mutation.

splicing signals. Most of the databases are based on the continuously increasing amount of expressed sequence tag (EST) data. These EST data provide the major information source for computational detection of alternative splicing patterns.

In order to predict alternative splicing events with accuracy, several algorithms were specially devised. The first computational approach is based on EST and mRNA comparison (23). However, EST-mRNA comparison has limited power because the intronic information is not included (24). To address the problems caused by EST-mRNA comparison, algorithms based on genome-EST pair-wise alignment are used in computational prediction programs. Several popular programs, such as BLAT, are designed using alignment of cDNA and genomic segments. These programs perform both genomic mapping and alignment (25). However, genome-EST comparison algorithms do not provide direct

exon–intron gene structures and may not be reliable. Hence, an algorithm based on EST-genome multiple alignment comparison has been devised in order to overcome this limitation. The algorithm minimizes the false splice site prediction due to incorrect EST-gene alignment that is not supported by the majority of EST data (26). **Table 10.2** lists database resources for alternative splicing.

Splicing regulatory RNA sequences and their *trans*-acting factors have been individually studied in great detail, both by experimental and bioinformatics approaches. The result of these studies led to the development of programs that predict splice sites, splicing regulatory sequences, their binding partners, as well as possible RNA secondary structure. **Table 10.3** lists these computational tools. It should be emphasized that due to the complexity of splice site selection the programs are currently fairly inaccurate. However, the usage of several programs allows the generation of hypotheses that can be experimentally tested.

1.5. Reporter Gene Analysis of Splicing Events

The most common technique to analyze exon usage and especially the effect of a mutation on splice site selection is reporter minigene analysis. In this method, an exon of interest, as well as its flanking exons are cloned into an expression vector and analyzed after transfecting this construct into eukaryotic cells (reviewed in (27, 28)). Currently about 200 such constructs have been reported in the literature. The comparison of two minigenes that were mutated to reflect naturally occurring alleles allows one to determine how a SNP influences alternative splicing.

The system can be expanded for the analysis of *trans*-acting factors by cotransfecting an increasing amount of factor-expressing cDNA constructs with the reporter minigene. To facilitate cloning of splicing reporter constructs, a recombination-based system that allows rapid generation of minigenes from PCR products containing the alternative exon has been developed (29). The system allows generating reporter minigenes within 1 week (*see Fig. 10.3*). Reporter minigenes then permit experimental tests concerning the influence of a mutation on splicing.

2. Materials

The computational analysis requires only internet access, as all programs are freely available.

2.1. Cloning of the Minigenes

1. Bacterial strain DB3.1 (Invitrogen, *E. coli* F⁻ gyrA462 endA1 (sr1-recA) mcrB mrr hsdS20(r_B⁻, m_B) supE44 ara14

Table 10.2
Overview of existing databases on alternative splicing events

Database	Description	Reference sequence	Species covered	Additional analysis tool	Integration with other resource	URL	References
ASAP	Database of human tissue-specific regulation of alternative splicing information through a genome-wide analysis of expressed sequence tags (ESTs)	ESTs	Human	Microarray analysis tutorial download	UniGene, genbank database	http://bioinfo.mbi.ucla.edu/ASSAP/	(45)
ASD (Alternative Splicing Database)	Database of manually annotated and computational generated data on alternative splicing events and the resultant isoform splice patterns of genes from model species. Database Alt-Splice and AEdb are included	Genomic sequence	Human and mouse	Intron analysis, scoring ATG-context sequence, MZEF-SPC exon finder, regulatory sequences	Ensemble genome annotation project	http://www.cbi.ac.uk/asd/	(46, 47)
ASDB (Alternative Splicing Database)	Database of alternatively spliced gene, their products and expression patterns	Genomic sequence	Human		Genbank database, Swiss-Prot protein knowledge base	http://hazelton.fbi.gov/~teplitski/alt/	(48, 49)

Table 10.2
(continued)

Database	Description	Reference sequence	Species covered	Additional analysis tool	Integration with other resource	URL	References
ASTRA (Alternative Splicing and Transcription Archives)	Database of elementary patterns of alternative splicing and transcriptional initiation	Genomic sequence	Human, mouse, <i>Drosophila</i> , <i>C. elegans</i> , arabidopsis, and Japanese rice		Genbank database	http://alterna.ebrc.jp/index.php?sp=mm#search	(50, 51)
ECgene	Database that combined the genome-based EST clustering and transcript assembly procedures	Genomic sequence	Human	Gene structure and function analysis, tissue-specific transcripts expression level analysis	Genbank, HUGO gene nomenclature committee, Swiss-Prot	http://genome.ewha.ac.kr/ECgene/	(52, 53)
H-DBAS (Human-transcriptome Database for Alternative Splicing)	Database of genome-wide representative alternative splicing variants generated from H-Inv full-length cDNAs and all transcripts datasets	Genomic sequence	Human	Motif sequence search, ESE prediction, protein subcellular localization and transmembrane domain prediction	HUGO gene nomenclature committee, ensemble genome annotation project, Genbank database	http://www.invitational.jp/h-dbas/	(54-57)

Table 10.2
(continued)

Database	Description	Reference sequence	Species covered	Additional analysis tool	Integration with other resource	URL	References
Hollywood	Database built upon genomic annotation of splicing patterns of known genes derived from spliced alignment of cDNAs and ESTs, and provide various analysis tools	Genomic sequence	Human and mouse	Tool for searching alternative conserved exons (ACEs), splice site score, ESEs and ESSs	Ensemble and gen bank database	http://hollywood.mit.edu/Login.php	(58-60)
MAASE (the Manually Annotated Alternatively Spliced Events Database)	Manually annotated alternatively spliced events database, designed for supporting splicing microarray applications	Genomic sequence	Human, mouse and Drosophila	Oligonucleotides design for microarray	Putative alternative splicing database, Swiss-Prot protein knowledgebase	http://maase.genomics.purdue.edu/	(61)
PALS db (Putative Alternative Splicing Database)	A collection of all available putative alternative splicing information hidden in biological sequence databases	mRNA	Human, mouse and worm	Putative alternative splice site finder	UniGene, HUGO gene nomenclature committee and Cancer Genome Anatomy project (CGAP)	http://ymbc.ym.edu.tw/palsdb/	(62)

Table 10.2
(continued)

Database	Description	Reference sequence	Species covered	Additional analysis tool	Integration with other resource	URL	References
ProSplicer	Database of putative alternative splicing information which are produced from variant proteins and expression patterns of genes	Genomic sequence	Human		UniGene, Ensembl and SwissProt	http://prosplicer.mbc.nctu.edu.tw/	(63)
SpliceInfo	Database provides information on tissue-specific alternative splicing events	Genomic sequence	Human	Motif discover tool for ESE, ESS and intronic splicing motifs	ProSplicer, Ensembl, InterPro and gene ontology	http://spliceinfo.mbc.nctu.edu.tw/	(63)

Table 10.2
(continued)

Database	Description	Reference sequence	Species covered	Additional analysis tool	Integration with other resource	URL	References
SpliceNest	A web based graphical tool to visualize splicing based on a mapping on the ESR consensus sequences from Genes database to the complete genome	Genomic sequence	Human, mouse, drosophila and arabidopsis		Umgene, and HUGO golden path assembly	http://splice.nest.molgen.mpg.de/	(64) (65)
TassDB	Database stores information about alternative splicing events at GYNGYN donors and NAGNAG acceptors. TassDB allows searching genes containing tandem splice sites with specific features location in the UTR or in the CDS	Genomic sequence	Human, mouse, rat, dog, chicken, drosophila, and <i>C. elegans</i>			http://helios.informatik.uni-freiburg.de/TassDB/	(66)

Table 10.3
Computational tools to predict and analyze alternative splicing events

Tools	Description	URL	References
Alifold	Tool predicts consensus secondary structures for set of aligned RNA and DNA sequences	http://ma.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi	(67)
ASD (Alternative Splicing Database)	Provide various tools for intron analysis, donor/acceptor score analysis, Polypyrimidine Tract (PPT) position analysis, branch point position and score analysis. ASD also can be used to identify potential exons in human genes, to identify and score alternative Open Reading Frames (ORF), to identify binding sites for splicing factor in RNA sequences and to check the known regulatory motifs in the sequences	http://www.ebi.ac.uk/asd-srv/wb.cgi	(46, 47)
ASp (Analyzer Splice Tool)	Calculate splice site score, number of H-bond between U1 and the 5' splice site, and the ΔG of U1/5' splice site pairing	http://ast.bioinfo.tau.ac.il/SpliceSiteFrame.htm	(68, 69)
BDGP (Berkeley Drosophila Genome Project)	Splice site prediction using neural network recognizers Provide single nucleotide polymorphism (SNP) maps	http://www.fruitfly.org/seq_tools/splice.html	(70)
CMfinder	RNA motif prediction tool. Perform well on unaligned sequences with long extraneous flanking regions	http://wingless.cs.washington.edu/hitbin-post/unrestricted/CMfinderWeb/CMfinderInput.pl	N/A
ESEfinder	Analysis of sequence to find ESE motifs	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi	(71-76)
GENE-BEE	RNA secondary structure prediction using energy model	http://www.genetec.msu.su/services/rna2_reduced.html	(77)
H-DBAS (Human-transcriptome Database for Alternative Splicing)	Coding-sequence (CDS) prediction, and ESE search	http://www.h-invitational.jp/h-dbas/adv_search.jsp	(55, 56)

Table 10.3
(continued)

Tools	Description	URL	References
Hollywood	5' and 3' splice site score calculation, ESE, ESS motif search in both human and mouse	http://hollywood.mit.edu/Dexon.php	(58, 78, 79), (72, 80-82)
HSF (Human Splicing Finder) <i>Image</i>	Analyses of mutation, branch point sequence, splice site, and multiple transcripts	http://www.umd.be/HSF/	N/A
NIPU web server	Image analysis tool	http://rsb.info.nih.gov/ij/	(83)
Primer 3	Analyses of splicing regulatory motifs and single-stranded regions	http://biwww2.informatik.uni-freiburg.de/Software/NIPU/	N/A
RegRNA	Optimization of PCR primers	http://frodo.wi.mit.edu/	(84)
RNA-Bioinformatics	A regulatory RNA motifs and elements finder, include motifs in 5' and 3' UTR, motifs involved in mRNA splicing, motifs involved in transcriptional regulation, riboswitches, splice site prediction, RNA structural features, and miRNA target sites	http://regna.mbc.nctu.edu.tw/php/browse.php	N/A
RNA-Motif Search Tool)	Collection of miRNAs	http://www.stamms-lab.net/miRNAs.htm	(85)
Splice site score calculation	An efficient and flexible RNA motif search tool for RNA structural homologs	http://bioinfo.csie.ncu.edu.tw/~rnamst/search.php	(86)
SplicePredictor	Calculation of 5' splice site and 3' splice site score. The statistical data were calculated using the sequence compilation for GENIE program	http://rulai.cshl.edu/new_alt_exon_db2/HTML/score.html	(87)
StrataSplice	Identify potential splice in plant pre-mRNA using Bayesian statistical models	http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi	(87)
	A human splice site predictor software that combines local GC content with a first-order dependence weight array model to predict splice sites	http://www.sanger.ac.uk/Software/analysis/stratasplice/	(87)

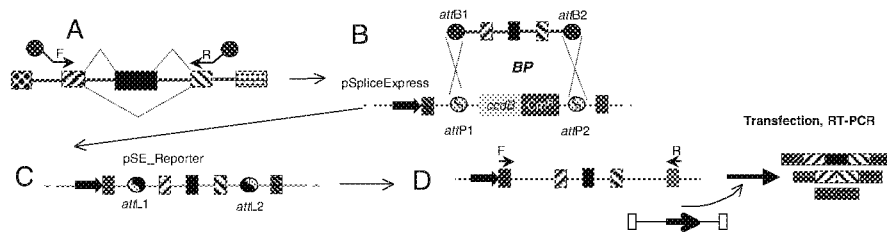


Fig. 10.3. Construction and analysis of splicing reporter genes using pSpliceExpress. A PCR product encompassing the gene region of interest is directly converted in to a splicing reporter gene by cloning it into pSpliceExpress. **a** Amplification of the region of interest. Two primers F and R are used to amplify a part of the genomic DNA that harbors the alternative exon (*black, splicing patterns* are indicated). The primers have recombination sites that are indicated by *circles*. **b** Construction of the splicing reporter using pSpliceExpress. The PCR fragment is recombined *in vitro* with pSpliceExpress vector. The vector contains Cm and ccdB selection markers that are used to isolate recombined clones. **c** Structure of the final construct using pSpliceExpress. The inserted DNA is flanked by two constitutive rat insulin exons, indicated by a *dotted pattern*. The transcript is driven by a CMV promoter (*Arrow*) and the subcloned genomic fragment is flanked by attL sites, generated by the recombination of attB and attP sites. **d** The analysis of the reporter occurs in cotransfection assays using expression constructs for splicing factors, siRNAs and other regulatory factors. The analysis is done by RT-PCR using primers in the constitutive insulin exons that are indicated by *small pointed arrows*.

galK2 lacY1 proA2 rpsL20(Sm^R) xyl5 Δ leu mtI1) to clone inserts that contain the bacterial ccdB marker.

2. Bacterial strain TOP10: (Invitrogen, *E. coli* F⁻ mcrA Δ (mrr-hsdRMS-mcrBC) Δ 80lacZ Δ M15 Δ lacX74 recA1 araD139 Δ (ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG) for all other constructs.

3. Primers

attB1F	5'-GGGG-ACAAGTTTCTACAAAAAAGCAGGCT - (template specific sequence)-3'
attB2R	5'-GGGG-ACCACTTTGTACAAGAAAGCTGGGT - (template specific sequence)-3'
attB1nestedF	5'-AAAAAGCAGGCT-template-specific sequences-3'
attB2nestedR	5'-AGAAAGCTGGGT-template-specific sequences-3'
attB1adapterF	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT -3'
attB2adapterR	5'-GGGGACCACTTTGTACAAGAAAGCTGGGT -3'

The primers are used as 10 pmol/ μ L working solutions in dH₂O. Store at -20°C.

4. *Pfx* DNA polymerase (5 U/ μ L) and buffer (Invitrogen). Store at -20°C.
5. BP clonase enzyme mix (Invitrogen). Store at -20°C.
6. *DpnI* restriction enzyme (20 U/ μ L) and buffer (New England Biolabs). Store at -20°C.
7. Proteinase K (Invitrogen, 20 mg/mL). Store at -20°C.

2.2. Transfection

1. Dulbecco's Modified Eagle's Medium (DMEM) (Gibco) supplemented with 10% fetal calf serum (Gibco).
2. 1 M CaCl₂.
3. 2× HEPES-buffered saline (2× HBS): 50 mM HEPES, 280 mM NaCl, 1.5 mM Na₂HPO₄, pH 6.95. Store at 4°C.

2.3. In Vivo Splicing Assay

1. RNaseasy kit (Qiagen).
2. Superscript III reverse transcriptase (200 U/μL) and buffer (Invitrogen). Store at -20°C.
3. *DpnI* restriction enzyme (20 U/μL) and buffer (New England Biolabs). Store at -20°C.
4. *Taq* DNA polymerase (5 U/μL) and buffer (New England Biolabs). Store at -20°C.
5. dNTP mix (Invitrogen, 100 mM)

3. Methods

As an example, we show the analysis of a mutation in an exon. In the first step, the effect of the mutation on alternative splicing is investigated bioinformatically. In the second step of analysis, these predictions are validated experimentally using transfection assays. The analysis strategy is shown in **Fig. 10.4**.

3.1. Bioinformatics Analysis

1. The sequence of interest is entered into several prediction programs. We routinely employ the “splicing rainbow” that predicts binding to regulatory factors and the “ESResearch” tool that predicts exonic regulatory elements using algorithms developed by three different laboratories.

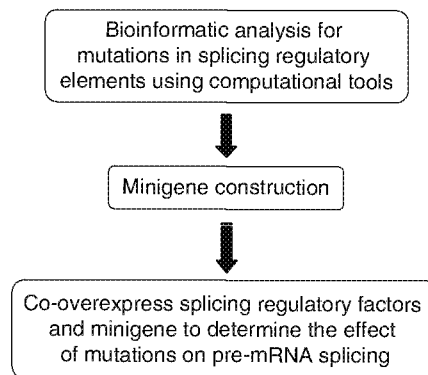


Fig. 10.4. Flowchart of the analysis.

- In addition, we use the “NIPU” server that predicts whether a nucleotide is in an enhancer or silencer by neighborhood interference (NI) and whether this nucleotide is in a single stranded or double stranded region (PU: probability unpaired).
- Finally, we determine the splice site strength of the exon hosting the mutation using “splice site score calculation.” The internet links to the programs are listed in **Table 10.3** and screenshots of the programs are shown in **Fig. 10.5**. Depending on the outcome of these predictions, we use additional programs, such as “ESE finder,” which determines binding to a subset of splicing regulatory proteins. Since different algorithms frequently give conflicting results, we combine the output of numerous programs.
- If several programs indicate that a nucleotide exchange influences a splicing regulatory sequence, we test these predictions experimentally using reporter gene analysis.

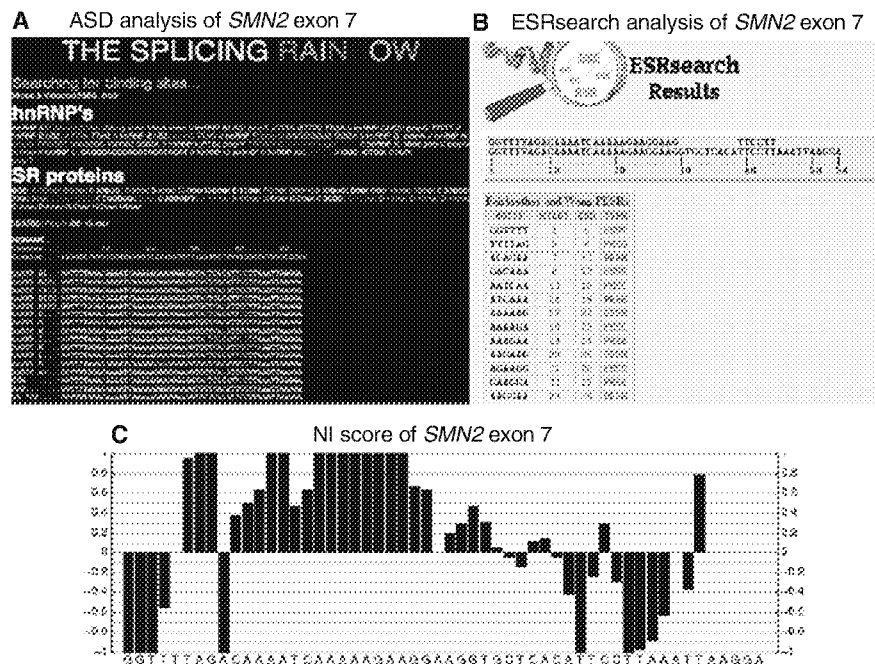


Fig. 10.5. Example of exonic elements prediction using computational tools. The sequence of *SMN2* exon 7 is used as a model sequence. The internet addresses of the programs are listed in **Table 10.3**. Analyses were done with (a) “splicing rainbow,” (b) “ESRsearch,” (c) NIPU.

3.2. Experimental Testing of the Bioinformatics Prediction

Most splicing reporter genes (minigenes) are constructed by cloning the alternative exon flanked by its constitutive exons into a eukaryotic expression vector. The resulting construct is transfected into cells and the splicing products are analyzed by

RT-PCR; frequently, an exon-trap vector is used. These vectors already contain two constitutive exons that flank a multiple cloning site. An exon of interest is inserted into this site and the construct is analyzed via transfection assays. Exon-trap vectors can be used when the exon of interest is flanked by large introns. The construction and analysis of minigenes has been previously reviewed (27, 28, 30). A list of currently employed minigenes is annotated on the web (*see* **Table 10.3** for address).

The cloning of reporter constructs is time-consuming and a major impediment of the technique. We therefore developed a cloning system that relies on site-specific recombination and allows generation of reporter minigenes within 1 week (29). The system is based on pSpliceExpress, a vector that contains two strong, constitutively used insulin exons. The insulin exons ensure that pre-mRNA splicing occurs in these constructs. The system is fast, allowing to generate reporter minigenes within 1 week. Numerous comparisons between conventional cloned minigenes and reporter genes with pSpliceExpress have shown that both systems behave similar (29). An overview of the technique is shown in **Fig. 10.3**.

3.2.1. Generation of Vectors with pSpliceExpress

1. Set up a standard PCR reaction using a proofreading DNA polymerase such as Pfx DNA polymerase and genomic DNA or a cloned piece of genomic DNA as template. For amplification primers, AttB1F and AttB2R are used (*see* **Section 2.1** for a list of primers) (*see* **Note 1**).
2. Add 5–10 units of *DpnI* to the PCR reaction and incubate at 37°C for 2 h to remove contaminating DNA originating from the genomic clone (*see* **Note 2**).
3. Set up a reaction to clone the PCR fragment into pSpliceExpress by mixing:
 - a. 20–30 fmoles of the attB containing PCR product
 - b. 25 fmoles of pSpliceExpress vector
 - c. 1 µL of 5-fold BP clonase reaction buffer mixture
 - d. TE buffer, pH 8–5 µLThe reaction is incubated at 25°C for 1 h (preferably overnight for fragments larger than 3 kb).
4. Add 0.5 µL of Proteinase K (2 mg/mL) solution to the reaction in order to inactivate the enzyme. Incubate at 37°C for 10 min.
5. Use the recombination mixture to transform Top10 bacteria. Any *recA*, *endA* *E. coli* strain including OmniMAX™ 2-T1R, TOP10, DH5α™, DH10B™ or equivalent can be used for transformation; however, no strains with the F episome should be used.

6. Isolate colonies, inoculated in LB amp-medium, and extract DNA by standard miniprep. The recombination site is flanked by *KpnI* sites. Digesting the miniprep DNA with *KpnI* or its isoschizomer Asp718I is used to identify clones with inserts. All constructs subject to further analysis should be verified by sequencing.

3.2.2. *Transfection and Analysis of Minigenes*
(see Note 3)

1. Use 1–2 µg of the minigene plasmid to transfect eukaryotic cells (see Note 4). Cells are seeded in 6-well plates and transfection is performed 24 h after plating (see Notes 5 and 6).
2. After incubation for 14–17 h at 3% CO₂, isolate total RNA from the cells using RNA columns (RNAeasy kit).
3. Set up a reverse transcription reaction for RT-PCR using 400 ng of RNA. The reverse primer used for RT is specific for the vector in which the minigene was cloned. This prevents amplification of endogenous RNA.
4. Add 5–10 units of *Dpn I* to the PCR reaction and incubate at 37°C for 2 h to avoid the problem of the amplification of minigene DNA (see Note 2). A control reaction with H₂O instead of RNA served as a contamination control.
5. 1/8 of the reverse transcription reactions is used for PCR with minigene-specific primers. The primers are selected to amplify alternatively spliced minigene products. A control reaction with no template (RNA instead of cDNA) is included in the PCR. The PCR programs should be optimized for each minigene in trial experiments. We alter the annealing temperature, elongation time, and cycle number (see Notes 7 and 8).
6. PCR reactions are resolved on a 0.3–0.4 cm thick 1–2% agarose TBE gel and the image are analyzed using “ImageJ” analysis software (see Table 10.3 for internet address). A typical analysis is shown in Fig. 10.6.

4. Notes

1. Since the amplification primers contain significant amounts of non-target sequences, in some genes we encountered undesired PCR products. This problem was especially apparent when we used genomic DNA and can often be avoided by using BAC clones. If the problem persists, we perform a two-step PCR procedure. First, the reaction is performed with a primer that is template specific and contains a part of the *attB* sequence at the 5' end. The first PCR is then used as a template for the second PCR with

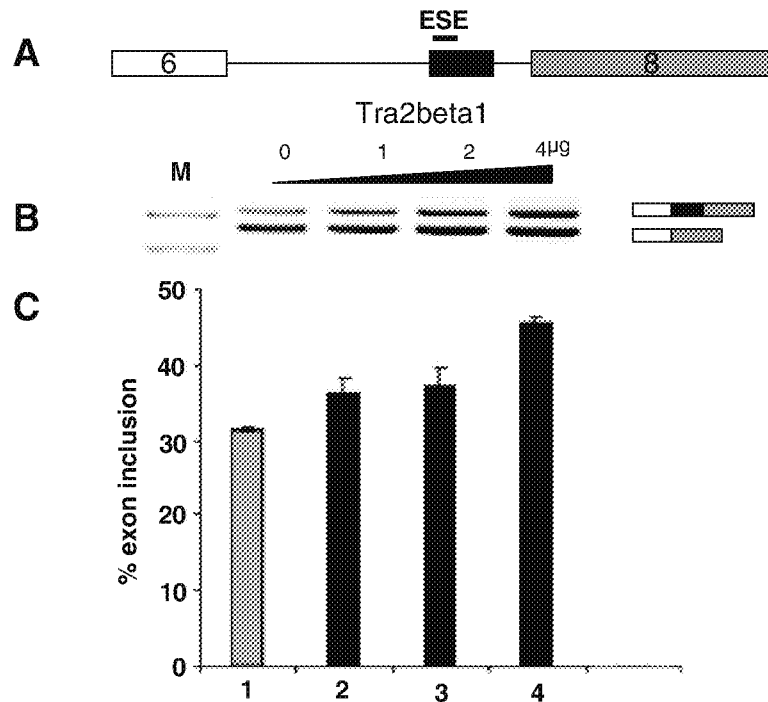


Fig. 10.6. Example for a minigene analysis. (a) Structure of the SMN2 minigene. (b) Cotransfection analysis of the SMN2 minigene with an increasing amount of the splicing factor tra2-beta1. (c) Quantification of the results.

adapter primers having a complete *attB* sequence. Template-specific primers for the first PCR reaction are designed with twelve bases of the attB1 or attB2 site on the 5' end of each primer (attB1nestedF and attB1nestedR, see primers). For the second PCR reaction, adapter primers are designed to generate the complete *attB* sequences (attB1adapterF and attB2adapter). The identity between adapter primers and template-specific primers has been underlined in **Table 10.1**. This alternative method allows smaller primers to be synthesized. Only the first set of primers (template-specific primers) is specific for a new minigene. The second set of primers (adapter primers) is used repeatedly for different minigene cloning projects.

2. The *DpnI* treatment degrades the contaminating plasmid DNA as *DpnI* recognizes methylated GATC sites. The treatment reduces background in the subsequent BP recombination reaction associated with template contamination. Purification of the PCR-amplified DNA is not required if a strong single band is obtained. In those cases where there is a high background, PCR purification of the products is performed by agarose gel electrophoresis followed by crystal violet staining and gel isolation of the relevant PCR product.

3. More detailed experimental details have been published (30).
4. In order to determine the effect of a single mutation, two versions of the minigene are generated and compared in transfection assays. Often, splicing patterns are cell-type dependent and we therefore test variant minigenes in different cell lines. The influence of predicted *trans*-acting factors can be assessed by cotransfecting an expression construct together with the reporter minigene. The expression construct can either encode a splicing factor or an shRNA targeted against a splicing factor. Usually, a concentration-dependent effect is analyzed. The expression construct is transfected in increasing amounts, in the range of 0–3 μ g. To avoid “squenching” effects, the “empty” parental expression plasmid containing the same promoter is added in decreasing amounts to ensure a constant amount of transfected DNA.
5. To obtain the best result, cells should be in optimal physiological conditions. HEK293 cells should be 60–80% confluent at the day of transfection.
6. The pH of transfection reagent 2 \times HBS is crucial. It should be 6.95 and tested with a pH meter. After filtering the transfection reagents under sterile conditions, these reagents should be tested by transfecting empty EGFP vectors into HEK293 cells. 24 h later, the transfection rate will be determined by observing the green cells ratio under fluorescent microscope.
7. In order to prevent amplification of endogenous genes, a vector-specific primer should be applied in RT reaction, and a gene-specific primer and a vector-specific primer should be used in PCR reaction.
8. We adjust the annealing temperature, which can be calculated using free online “Primer 3 program” (*see Table 10.3*), and the elongation time is if there are any difficulties of PCR amplification.

Acknowledgment

This work was supported by the EURASNET (European Alternative Splicing Network of Excellence), NIH (National Institutes of Health; P20 RR020171 from the National Center for Research Resources), BMBF (Federal Ministry of Education and Research, Germany), and the DFG (Deutsche Forschungsgemeinschaft; SFB 473).

References

1. Pan, Q., Shai, O., Lee, L. J., Frey, B. J., Blencowe, B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413–1415.
2. Wang, E. T., Sandberg, R., Luo, S., Khreb-tukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., Burge, C. B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
3. Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A., Soreq, H. (2005) Function of alternative splicing. *Gene* 344C, 1–20.
4. Ast, G. (2004) How did alternative splicing evolve? *Nat Rev Genet* 5, 773–782.
5. Zavolan, M., Kondo, S., Schonbach, C., Adachi, J., Hume, D. A., Hayashizaki, Y., Gaasterland, T. (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res* 13, 1290–1300.
6. Jurica, M. S., Moore, M. J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12, 5–14.
7. Nilsen, T. W. (2003) The spliceosome: the most complex macromolecular machine in the cell?. *Bioessays* 25, 1147–1149.
8. Hertel, K. J. (2008) Combinatorial control of exon recognition. *J Biol Chem* 283, 1211–1215.
9. Black, D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72, 291–336.
10. Smith, C. W., Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* 25, 381–388.
11. Maniatis, T., Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature* 416, 499–506.
12. Cooper, D. N., Stenson, P. D., Chuzhanova, N. A. (2006) The Human Gene Mutation Database (HGMD) and its exploitation in the study of mutational mechanisms. *Curr Protoc Bioinformatics* Chapter 1, Unit 1.13, 1.13.1 to 1.13.20.
13. Cooper, T. A., Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet* 61, 259–266.
14. Hull, J., Campino, S., Rowlands, K., Chan, M. S., Copley, R. R., Taylor, M. S., Rockett, K., Elvidge, G., Keating, B., Knight, J., Kwiatkowski, D. (2007) Identification of common genetic variation that modulates alternative splicing. *PLoS Genet* 3, e99.
15. Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., Baralle, F. E. (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 12, 1111–1120.
16. Graveley, B. R. (2008) The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population. *Trends Genet* 24, 5–7.
17. Orengo, J. P., Cooper, T. A. (2007) Alternative splicing in disease. *Adv Exp Med Biol* 623, 212–223.
18. Nissim-Rafinia, M., Kerem, B. (2002) Splicing regulation as a potential genetic modifier. *Trends Genet* 18, 123–127.
19. Buratti, E., Baralle, M., Baralle, F. E. (2006) Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res* 34, 3494–3510.
20. Faustino, N. A., Cooper, T. A. (2003) Pre-mRNA splicing and human disease. *Genes Dev* 17, 419–437.
21. Tazi, J., Bakkour, N., Stamm, S. (2009) Alternative splicing and disease. *Biochim Biophys Acta* 1792, 14–26.
22. Rave-Harel, N., Kerem, E., Nissim-Rafinia, M., Madjar, I., Goshen, R., Augarten, A., Rahat, A., Hurwitz, A., Darvasi, A., Kerem, B. (1997) The molecular basis of partial penetrance of splicing mutations in cystic fibrosis. *Am J Hum Genet* 60, 87–94.
23. Brett, D., Hanke, J., Lehmann, G., Haase, S., Delbrück, S., Krueger, S., Reich, J., Bork, P. (2000) EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett* 474, 83–86.
24. Heber, S., Alekseyev, M., Sze, S. H., Tang, H., Pevzner, P. A. (2002) Splicing graphs and EST assembly problem. *Bioinformatics* 18 Suppl 1, S181–S188.
25. Kent, W. J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12, 656–664.
26. Grasso, C., Modrek, B., Xing, Y., Lee, C. (2004) Genome-wide detection of alternative splicing in expressed sequences using partial order multiple sequence alignment graphs. *Pac Symp Biocomput* 29–41.
27. Tang, Y., Novoyatleva, T., Benderska, N., Kishore, S., Thanaraj, T. A., Stamm, S. (2005) in *Analysis of Alternative Splicing in Vivo using Minigenes*. Chapter 46 (Westhof, E., Bindereif, A., Schön, A., Hartmann, K. eds.), *Handbook of RNA Biochemistry*. Wiley-VCH, Weinheim, pp. 755–782.

28. Cooper, T. A. (2005) Use of minigene systems to dissect alternative splicing elements. *Methods* 37, 331–340.
29. Kishore, S., Khanna, A., Stamm, S. (2008) Rapid generation of splicing reporters with pSpliceExpress. *Gene* 427, 104–110.
30. Stoss, O., Stoilov, P., Hartmann, A. M., Nayler, O., Stamm, S. (1999) The in vivo minigene approach to analyze tissue-specific splicing. *Brain Res Protoc* 4, 383–394.
31. Ozsahin, H., Arredondo-Vega, F. X., Santisteban, I., Fuhrer, H., Tuchscheid, P., Jochum, W., Aguzzi, A., Lederman, H. M., Fleischman, A., Winkelstein, J. A., Seger, R. A., Herschfield, M. S. (1997) Adenosine deaminase deficiency in adults. *Blood* 89, 2849–2855.
32. Teraoka, S. N., Telatar, M., Becker-Catania, S., Liang, T., Onengut, S., Tolun, A., Chessa, L., Sanal, O., Bernatowska, E., Gatti, R. A., et al. (1999) Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am J Hum Genet* 64, 1617–1631.
33. Das, S., Levinson, B., Whitney, S., Vulpe, C., Packman, S., Gitschier, J. (1994) Diverse mutations in patients with Menkes disease often lead to exon skipping. *Am J Hum Genet* 55, 883–889.
34. Yang, Y., Swaminathan, S., Martin, B. K., Sharan, S. K. (2003) Aberrant splicing induced by missense mutations in BRCA1: clues from a humanized mouse model. *Hum Mol Genet* 12, 2121–2131.
35. Ploos van Amstel J. K., Bergman A. J., van Beurden E. A., Roijers J. F., Peelen T., van den Berg I. E., Poll-The B. T., Kvittingen E. A., Berger R. (1996) Hereditary tyrosinemia type I: novel missense, nonsense and splice consensus mutations in the human fumarylacetoacetate hydrolase gene; variability of the genotype-phenotype relationship. *Hum Genet* 97, 51–59.
36. Dreumont, N., Poudrier, J. A., Bergeron, A., Levy, H. L., Baklouti, F., Tanguay, R. M. (2001) A missense mutation (Q279R) in the fumarylacetoacetate hydrolase gene, responsible for hereditary tyrosinemia, acts as a splicing mutation. *BMC Genet* 2, 9.
37. Wakamatsu, N., Kobayashi, H., Miyatake, T., Tsuji, S. (1992) A novel exon mutation in the human beta-hexosaminidase beta subunit gene affects 3' splice site selection. *J Biol Chem* 267, 2406–2413.
38. Torres, R. J., Mateos, F. A., Molano, J., Gathoff, B. S., O'Neill, J. P., Gundel, R. M., Trombley, L., Puig, J. G. (2000) Molecular basis of hypoxanthine-guanine phosphoribosyltransferase deficiency in thirteen Spanish families. *Hum Mutat* 15, 383.
39. Carmel, I., Tal, S., Vig, I., Ast, G. (2004) Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* 10, 828–840.
40. Lee, V. M., Goedert, M., Trojanowski, J. Q. (2001) Neurodegenerative tauopathies. *Annu Rev Neurosci* 24, 1121–1159.
41. Wang, J., Gao, Q.-S., Wang, Y., Lafyatis, R., Stamm, S., Andreadis, A. (2004) Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. *J Neurochem* 88, 1078–1090.
42. Stella, A., Wagner, A., Shito, K., Lipkin, S. M., Watson, P., Guanti, G., Lynch, H. T., Fodde, R., Liu, B. (2001) A nonsense mutation in MLH1 causes exon skipping in three unrelated HNPCC families. *Cancer Res* 61, 7020–7024.
43. Fahsold, R., Hoffmeyer, S., Mischung, C., Gille, C., Ehlers, C., Kucukcylan, N., Abdel-Nour, M., Gewies, A., Peters, H., Kaufmann, D., et al. (2000) Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am J Hum Genet* 66, 790–818.
44. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., et al. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80, 155–165.
45. Kim, N., Alekseyenko, A. V., Roy, M., Lee, C. (2007) The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Res* 35, D93–D98.
46. Stamm, S., Riethoven, J. J., Le Texier, V., Gopalakrishnan, C., Kumanduri, V., Tang, Y., Barbosa-Morais, N. L., Thanaraj, T. A. (2006) ASD: a bioinformatics resource on alternative splicing. *Nucleic Acids Res* 34, D46–D55.
47. Thanaraj, T. A., Stamm, S., Clark, F., Riethoven, J. J., Le Texier, V., Muilu, J. (2004) ASD: the Alternative Splicing Database. *Nucleic Acids Res* 32, Database issue, D64–D69.
48. Dralyuk, I., Brudno, M., Gelfand, M. S., Zorn, M., Dubchak, I. (2000) ASDB: database of alternatively spliced genes. *Nucleic Acids Res* 28, 296–307.
49. Gelfand, M. S., Dubchak, I., Dralyuk, I., Zorn, M. (1999) ASDB: database of

- alternatively spliced genes. *Nucleic Acids Res* 27, 301–312.
50. Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., Gotoh, O. (2005) Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes. *Gene* 364, 53–62.
 51. Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., Gotoh, O. (2006) Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns. *Bioinformatics* 22, 1211–1216.
 52. Kim, N., Shin, S., Lee, S. (2005) ECgene: genome-based EST clustering and gene modeling for alternative splicing. *Genome Res* 15, 566–576.
 53. Kim, P., Kim, N., Lee, Y., Kim, B., Shin, Y., Lee, S. (2005) ECgene: genome annotation for alternative splicing. *Nucleic Acids Res* 33, D75–D79.
 54. Takeda, J., Suzuki, Y., Nakao, M., Barrero, R. A., Koyanagi, K. O., Jin, L., Motono, C., Hata, H., Isogai, T., Nagai, K., et al. (2006) Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res* 34, 3917–3928.
 55. Takeda, J., Suzuki, Y., Nakao, M., Kuroda, T., Sugano, S., Gojobori, T., Imanishi, T. (2007) H-DBAS: alternative splicing database of completely sequenced and manually annotated full-length cDNAs based on H-Invitational. *Nucleic Acids Res* 35, D104–D109.
 56. Takeda, J., Suzuki, Y., Sakate, R., Saro, Y., Seki, M., Irie, T., Takeuchi, N., Ueda, T., Nakao, M., Sugano, S., et al. (2008) Low conservation and species-specific evolution of alternative splicing in humans and mice: comparative genomics analysis using well-annotated full-length cDNAs. *Nucleic Acids Res* 36, 6386–6395.
 57. Imanishi, T., Itoh, T., Suzuki, Y., O'Donovan, C., Fukuchi, S., Koyanagi, K. O., Barrero, R. A., Tamura, T., Yamaguchi-Kabata, Y., Tanino, M. (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol* 2, e162.
 58. Fairbrother, W. G., Yeh, R. F., Sharp, P. A., Burge, C. B. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007–1013.
 59. Fairbrother, W. G., Yeo, G. W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P. A., Burge, C. B. (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res* 32, W187–W190.
 60. Fairbrother, W. G., Holste, D., Burge, C. B., Sharp, P. A. (2004) Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol* 2, E268.
 61. Zheng, C. L., Nair, T. M., Gribskov, M., Kwon, Y. S., Li, H. R., Fu, X. D. (2004) A database designed to computationally aid an experimental approach to alternative splicing. *Pac Symp Biocomput* 78–88.
 62. Huang, Y. H., Chen, Y. T., Lai, J. J., Yang, S. T., Yang, U. C. (2002) PALS db: putative alternative splicing database. *Nucleic Acids Res* 30, 186–190.
 63. Huang, H. D., Horng, J. T., Lin, F. M., Chang, Y. C., Huang, C. C. (2005) Splice-Info: an information repository for mRNA alternative splicing in human genome. *Nucleic Acids Res* 33, D80–D85.
 64. Gupta, S., Zink, D., Korn, B., Vingron, M., Haas, S. A. (2004) Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing. *BMC Genomics* 5, 72.
 65. Gupta, S., Vingron, M., Haas, S. A. (2005) T-STAG: resource and web-interface for tissue-specific transcripts and genes. *Nucleic Acids Res* 33, W654–W658.
 66. Hiller, M., Nikolajewa, S., Huse, K., Szafranski, K., Rosenstiel, P., Schuster, S., Backofen, R., Platzer, M. (2007) TassDB: a database of alternative tandem splice sites. *Nucleic Acids Res* 35, D188–D192.
 67. Hofacker, I. L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31, 3429–3431.
 68. Norton, P. A. (1994) Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res* 22, 3854–3860.
 69. Coolidge, C. J., Seely, R. J., Patton, J. G. (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* 25, 888–896.
 70. Reese, M. G., Eeckman, F. H., Kulp, D., Haussler, D. (1997) Improved splice site detection in Genie. *J Comput Biol* 4, 311–323.
 71. Blencowe, B. J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25, 106–110.
 72. Cartegni, L., Chew, S. L., Krainer, A. R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3, 285–298.
 73. Caceres, J. F., Kornblihtt, A. R. (2002) Alternative splicing: multiple control mechanisms

- and involvement in human disease. *Trends Genet* 18, 186–193.
74. Hastings, M. L., Krainer, A. R. (2001) Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* 13, 302–309.
 75. Graveley, B. R. (2000) Sorting out the complexity of SR protein function. *RNA* 6, 1197–1211.
 76. Graveley, B. R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 17, 99–108.
 77. Brodsky, L. I., Drachev, A. L., Leontovich, A. M., Feranchuk, S. I. (1993) A novel method of multiple alignment of biopolymer sequences. *Biosystems* 30, 65–79.
 78. Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., Burge, C. B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell* 119, 831–845.
 79. Lim, S. R., Hertel, K. J. (2001) Modulation of SMN pre-mRNA splicing by inhibition of alternative 3' splice site pairing. *J Biol Chem* 276, 2.
 80. Garcia-Blanco, M. A., Baraniak, A. P., Lasda, E. L. (2004) Alternative splicing in disease and therapy. *Nat Biotechnol* 22, 535–546.
 81. Wang, G. S., Cooper, T. A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8, 749–761.
 82. Wang, Z., Burge, C. B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14, 802–813.
 83. Stadler, M. B., Shomron, N., Yeo, G. W., Schneider, A., Xiao, X., Burge, C. B. (2006) Inference of splicing regulatory activities by sequence neighborhood analysis. *PLoS Genet* 2, e191.
 84. Huang, H. Y., Chien, C. H., Jen, K. H., Huang, H. D. (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* 34, W429–W434.
 85. Chang, T. H., Huang, H. D., Chuang, T. N., Shien, D. M., Horng, J. T. (2006) RNAMST: efficient and flexible approach for identifying RNA structural homologs. *Nucleic Acids Res* 34, W423–W428.
 86. Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., Zhang, M. Q. (2000) An alternative-exon database and its statistical analysis. *DNA Cell Biol* 19, 739–756.
 87. Brendel, V., Xing, L., Zhu, W. (2004) Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. *Bioinformatics* 20, 1157–1169.