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)).

![Diagram of alternative splicing patterns](image)

**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
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 (hnRNP)s. 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 snRNPs. 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.
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.

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.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Effect</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADA</td>
<td>CTGTCACGCGC→CTGTCACACC</td>
<td>Exon 7 skipping</td>
<td>(81)</td>
</tr>
<tr>
<td>ATM</td>
<td>ATTCGATGG→ATTTGAGTG</td>
<td>Exon 9 skipping</td>
<td>(32)</td>
</tr>
<tr>
<td></td>
<td>ACTCAAACAT→ACTTAAACAT</td>
<td>Exon 9 skipping</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGGtaa→AGttaa</td>
<td>Intron 12 retention</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ttatatGT→ttatatGT</td>
<td>Exon 18 skipping</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AACGGTTTGA→AACAGTTTTA</td>
<td>Exon 26 skipping</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTCGAACA→CTCCAAACA</td>
<td>Exon 44 skipping</td>
<td></td>
</tr>
<tr>
<td>ATP7A</td>
<td>GATGGAAATC→GATC/AGAATC</td>
<td>Exon 4 skipping</td>
<td>(33)</td>
</tr>
<tr>
<td>BRCA1</td>
<td>ATCTTAAAGAG→ATCTTAAAGAG</td>
<td>Exon 18 skipping</td>
<td>(34)</td>
</tr>
<tr>
<td>FAH</td>
<td>AAGAAGAC→AAGAAGAGAC</td>
<td>Exon 8 skipping</td>
<td>(35, 36)</td>
</tr>
<tr>
<td></td>
<td>AAGCCAGAC→AAGCGGGAC</td>
<td>Exon 9 skipping</td>
<td></td>
</tr>
<tr>
<td>HEXB</td>
<td>GCGGCGGGGC→GCGCTGGGC</td>
<td>Exon 11 skipping</td>
<td>(37)</td>
</tr>
<tr>
<td>HPRT1</td>
<td>CATGGACTA→CATAGACTA</td>
<td>Exon 8 skipping</td>
<td>(38)</td>
</tr>
<tr>
<td></td>
<td>GAACGTCTT→GAACATCTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATTGTGGA→ATTATGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IKBKAP</td>
<td>graagtgc→graagegc</td>
<td>Exon 20 skipping</td>
<td>(39)</td>
</tr>
<tr>
<td>MAPT</td>
<td>ATTAATAAG→ATTAGAAGAG</td>
<td>Exon 10 inclusion</td>
<td>(40, 41)</td>
</tr>
<tr>
<td></td>
<td>GCCAGTgtg→GCCAGTgtg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GACATGCAG→GACATCGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATAATACAG→ATAACATCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLH1</td>
<td>GAGAAGAGAG→GAGTAGAGAG</td>
<td>Exon 12 skipping</td>
<td>(42)</td>
</tr>
<tr>
<td>NF1</td>
<td>CTTAAGAAC→CTTAAACAC</td>
<td>Exon 7 skipping</td>
<td>(43)</td>
</tr>
<tr>
<td>SMN2</td>
<td>GGTGTGAGAC→GGGTGTGAC</td>
<td>Exon 7 skipping</td>
<td>(44)</td>
</tr>
</tbody>
</table>

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 (25). 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 let 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(rB−, mB) supE44 ara14
Table 10.2
Overview of existing databases on alternative splicing events

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Reference sequence</th>
<th>Species covered</th>
<th>Additional analysis tool</th>
<th>Integration with other resource</th>
<th>URL</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASAP</td>
<td>Database of human tissue-specific regulation of alternative splicing information through a genome-wide analysis of expressed sequence tags (ESTs)</td>
<td>ESTs</td>
<td>Human</td>
<td>Microarray analysis, tutorial download</td>
<td>Unigene, genbank database</td>
<td><a href="http://bioinfo.mbi.ucla.edu/ASAP/">http://bioinfo.mbi.ucla.edu/ASAP/</a></td>
<td>(45)</td>
</tr>
<tr>
<td>ASD (Alternative Splicing Database)</td>
<td>Database of manually annotated and computational generated data on alternative splicing events and the resultant isoform splice patterns of genes from model species. Database AltSplice and AEdb are included</td>
<td>Genomic sequence</td>
<td>Human and mouse</td>
<td>Intron analysis, scoring ATG context sequence, MZEF-SPC exon finder, regulatory sequences</td>
<td>Ensemble genome annotation project</td>
<td><a href="http://www.ebi.ac.uk/asd/">http://www.ebi.ac.uk/asd/</a></td>
<td>(46, 47)</td>
</tr>
<tr>
<td>ASDB (Alternative Splicing Database)</td>
<td>Database of alternatively spliced gene, their products and expression patterns</td>
<td>Genomic sequence</td>
<td>Human</td>
<td>Genbank database, Swiss Prot protein knowledge base</td>
<td></td>
<td><a href="http://hazelton.lbl.gov/~teplitski/alt/">http://hazelton.lbl.gov/~teplitski/alt/</a></td>
<td>(48, 49)</td>
</tr>
<tr>
<td>Database</td>
<td>Description</td>
<td>Reference sequence</td>
<td>Species covered</td>
<td>Additional analysis tool</td>
<td>Integration with other resource</td>
<td>URL</td>
<td>References</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>ECgene</td>
<td>Database that combined the genome-based EST clustering and transcript assembly procedures</td>
<td>Genomic sequence</td>
<td>Human</td>
<td>Gene structure and function analysis, tissue-specific transcripts expression level analysis</td>
<td>Genbank, HUGO gene</td>
<td>nomenclature, committee, Swiss-Prot</td>
<td><a href="http://genome.cwha.ac.kr/ECgene/">http://genome.cwha.ac.kr/ECgene/</a></td>
</tr>
</tbody>
</table>
Table 10.2
(continued)

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Sequence</th>
<th>Species Covered</th>
<th>Additional Analysis Tool</th>
<th>Integration with Other Resource</th>
<th>URL</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hollywood</td>
<td>Database built upon genomic annotation of splicing patterns of known genes derived from spliced alignments of cDNAs and ESTs, and provide various analysis tools</td>
<td>Genomic sequence</td>
<td>Human and mouse</td>
<td>Tool for searching alternative conserved exons (ACEs), splice site score, ESEs, and ESSs</td>
<td>Ensemble and gen bank database</td>
<td><a href="http://hollywood.mit.edu/Login.php">http://hollywood.mit.edu/Login.php</a></td>
<td>(58-60)</td>
</tr>
<tr>
<td>MAASE (the Manually annotated alternatively spliced events database)</td>
<td>Manually annotated alternatively spliced events database, designed for supporting splicing microarray applications</td>
<td>Genomic sequence</td>
<td>Human, mouse and Drosophila</td>
<td>Oligonucleotides design for microarray</td>
<td>Putative alternative splicing database, Swiss Prot protein knowledgebase</td>
<td><a href="http://maase.genomics.purdue.edu/">http://maase.genomics.purdue.edu/</a></td>
<td>(61)</td>
</tr>
<tr>
<td>PALS db (Putative Alternative Splicing Database)</td>
<td>A collection of all available putative alternative splicing and worm site finder</td>
<td>mRNA sequence</td>
<td>Human, mouse and worm</td>
<td>Unigene, HUGO gene nomenclature committee</td>
<td>Cancer Genome Anatomy Project (CGAP)</td>
<td><a href="http://ymc.tw/palsdb/">http://ymc.tw/palsdb/</a></td>
<td>(62)</td>
</tr>
<tr>
<td>Database</td>
<td>Description</td>
<td>Reference sequence</td>
<td>Species covered</td>
<td>Additional analysis tool</td>
<td>Integration with other resource</td>
<td>URL</td>
<td>References</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>ProSplicer</td>
<td>Database of putative alternate splicing information which are produced from variant proteins and expression patterns of genes</td>
<td>Genomic sequence</td>
<td>Human</td>
<td></td>
<td>Unigene, Ensemble and Swiss-Prot</td>
<td>[<a href="http://prosplicer">http://prosplicer</a>](<a href="http://prosplicer">http://prosplicer</a>. mbc.nctu.edu.tw/)</td>
<td>(63)</td>
</tr>
<tr>
<td>SpliceInfo</td>
<td>Database provides information on tissue-specific alternative splicing events</td>
<td>Genomic sequence</td>
<td>Human</td>
<td>Motif discovery tool for ESE, ESS and intronic splicing motifs</td>
<td>ProSplicer, ensemble, InterPro and gene ontology</td>
<td>[<a href="http://spliceinfo">http://spliceinfo</a>](<a href="http://spliceinfo">http://spliceinfo</a>. mbc.nctu.edu.tw/)</td>
<td>(63)</td>
</tr>
<tr>
<td>Database</td>
<td>Description</td>
<td>Reference sequence</td>
<td>Species covered</td>
<td>Additional analysis tool</td>
<td>Integration with other resource</td>
<td>URL</td>
<td>References</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>-----</td>
<td>------------</td>
</tr>
<tr>
<td>SpliceNest</td>
<td>A web-based graphical tool to visualize splicing, based on a mapping on the ESR consensus sequences from GeneNest database to the complete genome</td>
<td>Genomic sequence</td>
<td>Human, mouse, drosophila, and Arabidopsis</td>
<td>Unigene, and HUGO golden path assembly</td>
<td><a href="http://splicenest">http://splicenest</a>. (64) (65). molgen.mpg.de/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TassDB</td>
<td>Database stores information about alternative splicing events at GYNGYN donors and NAGNAG acceptors. TassDB allows searching genes containing tandem splice sites with specific features located in the UTR or in the CDS</td>
<td>Genomic sequence</td>
<td>Human, mouse, rat, dog, chicken, drosophila, and C. elegans</td>
<td></td>
<td><a href="http://helios">http://helios</a>. informatik.unifreiburg.de/TassDB/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 10.3

**Computational tools to predict and analyze alternative splicing events**

<table>
<thead>
<tr>
<th>Tools</th>
<th>Description</th>
<th>URL</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alifold</td>
<td>Tool predicts consensus secondary structures for set of aligned RNA and DNA sequences</td>
<td><a href="http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi">http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi</a></td>
<td>(67)</td>
</tr>
<tr>
<td>ASD (Alternative Splicing Database)</td>
<td>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</td>
<td><a href="http://www.cbi.ac.uk/asd-srv/wb.cgi">http://www.cbi.ac.uk/asd-srv/wb.cgi</a></td>
<td>(46, 47)</td>
</tr>
<tr>
<td>AST (Analyzer Splice Tool)</td>
<td>Calculate splice site score, number of H-bond between U1 and the 5' splice site, and the ΔG of U1/5' splice site pairing</td>
<td><a href="http://ast.bioinfo.tau.ac.il/SpliceSiteFrame.htm">http://ast.bioinfo.tau.ac.il/SpliceSiteFrame.htm</a></td>
<td>(68, 69)</td>
</tr>
<tr>
<td>BDGP (Berkeley Drosophila Genome Project)</td>
<td>Splice site prediction using neural network recognizers, provide single nucleotide polymorphism (SNP) maps</td>
<td><a href="http://www.fruitfly.org/seq_tools/splice.html">http://www.fruitfly.org/seq_tools/splice.html</a></td>
<td>(70)</td>
</tr>
<tr>
<td>ESEfinder</td>
<td>Analysis of sequence to find ESE motifs</td>
<td><a href="http://rulai.cs.tamu.edu/cgi-bin/tools/ESE3/eSEfinder.cgi">http://rulai.cs.tamu.edu/cgi-bin/tools/ESE3/eSEfinder.cgi</a></td>
<td>(71–76)</td>
</tr>
<tr>
<td>GENE BEE</td>
<td>RNA secondary structure prediction using energy model</td>
<td><a href="http://www.genebee.msu.su/services/rna2_reduced.html">http://www.genebee.msu.su/services/rna2_reduced.html</a></td>
<td>(77)</td>
</tr>
<tr>
<td>Tools</td>
<td>Description</td>
<td>URL</td>
<td>References</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Hollywood</td>
<td>5' and 3' splice site score calculation, ESE, ESS motif search in both human and mouse</td>
<td><a href="http://hollywood.mit.edu/Dx">http://hollywood.mit.edu/Dx</a> exon.php</td>
<td>(58, 78, 79)</td>
</tr>
<tr>
<td>HSF (Human Splicing Finder)</td>
<td>Analyses of mutation, branch point sequence, splice site, and multiple transcripts</td>
<td><a href="http://www.umd.edu/HSF/">http://www.umd.edu/HSF/</a></td>
<td>(72, 80-82)</td>
</tr>
<tr>
<td>ImageJ</td>
<td>Image analysis tool</td>
<td><a href="http://rsb.info.nih.gov/ij/">http://rsb.info.nih.gov/ij/</a></td>
<td>N/A</td>
</tr>
<tr>
<td>NIPU web server</td>
<td>Analyses of splicing regulatory motifs and single-stranded regions</td>
<td><a href="http://biwww2.informatik.uni-freiburg.de/Software/NIPU/">http://biwww2.informatik.uni-freiburg.de/Software/NIPU/</a></td>
<td>(83)</td>
</tr>
<tr>
<td>Primer 3</td>
<td>Optimization of PCR primers</td>
<td><a href="http://trodo.wi.mit.edu/">http://trodo.wi.mit.edu/</a></td>
<td>N/A</td>
</tr>
<tr>
<td>RegRNA</td>
<td>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</td>
<td><a href="http://regRNA.mbc.nctu.edu.tw/php/browse.php">http://regRNA.mbc.nctu.edu.tw/php/browse.php</a></td>
<td>(84)</td>
</tr>
<tr>
<td>RNA Bioinformatics</td>
<td>Collection of minigenes</td>
<td><a href="http://www.stammslab.net/minigenes.htm">http://www.stammslab.net/minigenes.htm</a></td>
<td>N/A</td>
</tr>
<tr>
<td>Splice site score calculation</td>
<td>Calculation of 5' splice site and 3' splice site score. The statistical data were calculated using the sequence compilation for GENIE program</td>
<td><a href="http://rulai.cshl.edu/new_alt_exon_db2/HTML/score.html">http://rulai.cshl.edu/new_alt_exon_db2/HTML/score.html</a></td>
<td>(86)</td>
</tr>
<tr>
<td>SplicePredictor</td>
<td>Identify potential splice in plant pre-mRNA using Bayesian statistical models</td>
<td><a href="http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi">http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi</a></td>
<td>(87)</td>
</tr>
<tr>
<td>StrataSplice</td>
<td>A human splice site predictor software that combines local GC content with a first-order dependence weight array model to predict splice sites</td>
<td><a href="http://www.sanger.ac.uk/Software/analysis/strataslice/">http://www.sanger.ac.uk/Software/analysis/strataslice/</a></td>
<td>(87)</td>
</tr>
</tbody>
</table>
galK2 lacY1 proA2 rpsL20(Sm<sup>R</sup>) xyl5 Δleu mtl1) to clone inserts that contain the bacterial ccdB marker.

2. Bacterial strain TOP10: (Invitrogen, E. coli F<sup>−</sup> mcrA Δ(mrr-hsdRMS-mcrBC) Δ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str<sup>R</sup>) endA1 nupG) for all other constructs.

3. Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>artB1F</td>
<td>5'-GGGG-ACAAGTTTCTACAAAAAAGCAGGCT-3'</td>
</tr>
<tr>
<td>artB2R</td>
<td>5'-GGGG-ACCACCTTGTACAAGAAAGCTGGGT-3'</td>
</tr>
<tr>
<td>artB1nestedF</td>
<td>5'-AAAAAGCAGGCT-template specific sequences-3'</td>
</tr>
<tr>
<td>artB2nestedR</td>
<td>5'-AGAAGGCTGGGT-template specific sequences-3'</td>
</tr>
<tr>
<td>artB1adapterF</td>
<td>5'-GGGACAAGTTTGTACAAGAAAGCAGGCT-3'</td>
</tr>
<tr>
<td>artB2adapterR</td>
<td>5'-GGGACCACTTTGTACAAGAAAGCAGGCT-3'</td>
</tr>
</tbody>
</table>

The primers are used as 10 pmol/μL working solutions in dH<sub>2</sub>O. Store at −20°C.

4. Pf<sub>K</sub> DNA polymerase (5 U/μL) and buffer (Invitrogen). Store at −20°C.

5. BP clonase enzyme mix (Invitrogen). Store at −20°C.

6. DpnI restriction enzyme (20 U/μL) 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× Heps-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. RNAeasy 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.
2. 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).

3. 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.

4. 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.

1. Set up a standard PCR reaction using a proofreading DNA polymerase such as Pfu 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 fmol of the attB containing PCR product
   b. 25 fmol of pSpliceExpress vector
   c. 1 μL of 5-fold BP clonase reaction buffer mixture
   d. TE buffer, pH 8–5 μL
   The 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 minipreparation. The recombination site is flanked by KpnI sites. Digesting the minipreparation 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.

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
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 μg. To avoid “squelching” 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× 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


