Review Article
Role of Pseudoexons and Pseudointrons in Human Cancer

Maurizio Romano,1 Emanuele Buratti,2 and Diana Baralle3,4

1 Department of Life Sciences, University of Trieste, Via A. Valerio 28, 34127 Trieste, Italy
2 International Centre for Genetic Engineering and Biotechnology (ICGEB), 34149 Trieste, Italy
3 Human Development and Health, University of Southampton, Duthie Building, Mailpoint 808, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK
4 Human Genetics Division, University of Southampton, Duthie Building, Mailpoint 808, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK

Correspondence should be addressed to Diana Baralle; d.baralle@soton.ac.uk

Received 17 June 2013; Accepted 9 August 2013

Copyright © 2013 Maurizio Romano et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In all eukaryotic organisms, pre-mRNA splicing and alternative splicing processes play an essential role in regulating the flow of information required to drive complex developmental and metabolic pathways. As a result, eukaryotic cells have developed a very efficient macromolecular machinery, called the spliceosome, to correctly recognize the pre-mRNA sequences that need to be inserted in a mature mRNA (exons) from those that should be removed (introns). In healthy individuals, alternative and constitutive splicing processes function with a high degree of precision and fidelity in order to ensure the correct working of this machinery. In recent years, however, medical research has shown that alterations at the splicing level play an increasingly important role in many human hereditary diseases, neurodegenerative processes, and especially in cancer origin and progression. In this minireview, we will focus on several genes whose association with cancer has been well established in previous studies, such as ATM, BRCA1/A2, and NF1. In particular, our objective will be to provide an overview of the known mechanisms underlying activation/repression of pseudoexons and pseudointrons; the possible utilization of these events as biomarkers of tumor staging/grading; and finally, the treatment options for reversing pathologic splicing events.

1. Introduction

Starting from the first description of alternative splicing and constitutive splicing processes in 1977 [1–3], the importance of this process that guarantees the correct flow of information from transcription to translation in eukaryotic cells has continued to grow exponentially.

In particular, one major branch of research in this area has the aim to investigate and characterize the cellular macromolecular machine (i.e., the spliceosome) that is physically responsible for the cutting and joining of intronexons by catalyzing two transesterification reactions [4, 5] and the mechanisms that ensure its fidelity [6]. As a result, research in spliceosome composition and functioning has been complemented by studies aiming to understand the sequences and molecules that determine under which conditions a particular exon or intron is selectively recognized and included in the mature transcript. Therefore, after the basic elements that define introns and exons and are composed by donor and acceptor splice sites plus the branch point sequence, there was the discovery of enhancer and silencer elements that can affect either positively or negatively the way these basic elements are recognized by the spliceosome [7–9]. Usually, enhancer and silencer elements are bound by members of the SR and hnRNP protein families, respectively.

The recruitment of these proteins to specific sites is crucial for their activity as depending on their position with respect to the basic elements their roles can be antagonistic or, in some case, agonistic [10, 11]. Global analyses of the ways these factors cooperate and influence each other in shaping the splicing process has only recently begun to shed light on how they promote or hinder exon recognition in a co-ordinated manner [12–14].
Rather unexpectedly, these studies have shown that SR proteins interact not only with alternatively spliced exons but also with constitutively spliced exons, and that their major role consists in recruiting the splicing machinery to splice sites [15, 16].

On the other hand, hnRNP proteins bind to nascent pre-mRNA and influence splicing decision through a complex and finely tuned network of protein/protein interactions. The mechanisms of hnRNPs actions on splicing is generally less defined than those of SR proteins [17].

In addition to these elements, we now also know that splicing choices are affected by a myriad of other factors, ranging from chromatin modifications [18, 19], transcriptional factors [20], RNA secondary structure [21], short noncoding RNAs [22, 23], and various cellular stresses [24].

In parallel to these mechanistic studies, another branch of splicing research has also addressed the functional importance of alternative splicing processes in biological pathways and in particular the way that alternative splicing isoforms of proteins can acquire different or even antagonistic biological properties [25, 26]. Because of this ability to expand the proteome of cells, alternative splicing has represented a very useful and powerful tool that allows cells to execute the various expression programs which underlie many fundamental needs of higher organisms: from general needs such as controlling normal development and tissue-specific expression of proteins, to highly specialized processes such as DNA damage response or microRNA biogenesis [27–33]. Moreover, the rearrangements that a pre-mRNA undergoes during the splicing process are also advantageous in terms of providing a longer half-life and better translational capacity, something that has recently begun to be exploited by the biotechnology industry [34].

Considering all these necessities and advantages, it is therefore not surprising that the number of genes that are subject to alternative splicing in eukaryotic organisms has been steadily growing. Indeed, it was recently estimated that more than 90% of mammalian protein coding genes can produce at least one splicing-derived isoform expressed at potentially significant biological levels [35].

The complexity of this system, however, also puts splicing under some functioning at risk of being impaired by the occurrence of single-point mutations in splicing regulatory elements, deletions/insertions, genomic rearrangements, and alterations at the splicing factor expression level [36]. Any of these alterations can result in a variety of aberrant splicing outcomes that usually include aberrant exon skipping, cryptic splice site selection, intron retention, and pseudexon activation [37, 38]. As expected, many of these changes can lead directly to the occurrence of disease in humans, and it has now been estimated that a sizable proportion of all gene mutations leading to disease can be directly connected with the presence of a splicing defect [7, 39].

Importantly, the introduction of novel technologies that allow fast profiling of the transcriptome seems promising for simplifying investigation of which genomic variability and plasticity events allow cancer cells to tailor specific functional units from the available exons of a gene. For example, recent RNA sequencing of the breast cancer transcriptome has revealed many new splicing alterations that were not previously described. In particular, for example, it was reported that 423 primary transcripts (derived from 377 genes) were differentially spliced in triple-negative breast cancer samples (generating 496 novel isoforms) [40]. By analyzing non-triple-negative breast cancer samples, the same study found that 270 and 460 primary transcripts (derived from 242 and 387 differentially spliced genes, resp.) generated 331 and 550 novel isoforms that were not present in normal breast tissue [40].

2. Aberrant Splicing Events in Cancer Genes

It has become clear that in several human pathologies, including cancer, the alternative splicing profile is aberrantly modified in a specific manner and in ways that can favor the growth and survival of cancer cells [41–44]. The generation of these aberrant splicing profiles can occur in many ways, such as through the re-expression of developmentally regulated isoforms that had previously been shut off following early developmental stages [45, 46], by affecting the splicing profiles of genes that are implicated in tumor progression [47, 48], or through other mechanisms that generally have anti-apoptotic/metastatic consequences and that can affect response to therapies [49, 50].

In addition, several studies have defined that cancer-associated splicing alterations arise not only from mutations in the cancer-related genes but derive also from variations in the expression and/or activity of splicing regulatory factors [51]. For example, it has been well established that the levels of SR- and hnRNP-proteins undergo changes associated with transformation and progression of cancers [44, 52]. Strikingly, it has been recently reported that some members of the SR protein family of splicing factors and other components of the spliceosomal cellular machinery can actually act as oncoproteins and play a direct role in promoting tumor origin and progression [53–55] and that external stimuli such as hypoxia can cause the aberrant redistribution of important splicing factors such as Tra2 and promote the expression of tumor-promoting splicing isoforms [56].

The case of Bcl-X (BCL2L1) gene, belonging to the BclII family and implicated in the control of mitochondrial breakdown during apoptosis, is emblematic of this concept. For example, two splicing Bcl-X isoforms can arise from the use of two alternative 5′ splice sites within exon 2 and lead to the synthesis of a short apoptosis-promoting protein (Bcl-XS) and to a long antiapoptotic form (Bcl-XL) [57]. Different splicing factors, including Sam68, hnRNPA1, SF2/ASF, hnRNP F/H, hnRNP K, SAP155, and SRp30c have been found to be involved in the selection of the two competing alternative 5′ splice sites that give rise to these two isoforms [58–61]. In addition, recent studies have shown that the elongation and splicing-related factor TCERG1 can bind to the Bcl-X pre-mRNA and promote the proapoptotic Bcl-XS 5′ splice site in a promoter-dependent manner [62].

Finally, the production of the proapoptotic Bcl-XS splice variant seems to be improved by the core (Y14 and eIF4A3) and auxiliary (RNPS1, Acinus, and SAP18) components of
the exon junction complex (EJC) [63], suggesting that EJC-associated components can regulate apoptosis at the alternative splicing level and represent a further level of vulnerability of cancer.

Therefore, one of the most interesting research areas in this field consists in the identification of cancer-specific splice variants or the aberrant expression of splicing-affecting proteins that could lead to their generation. Examples of both these events, in fact, have already been shown to occur in some individual types of cancer, such as breast and ovarian cancer [64, 65]. Another interesting research area in aberrant splicing events connected with tumors is the occurrence of particular types of splicing defects that involve the inclusion of “new” sequences (known as pseudoexons) in the mature mRNA of cancer-related genes. Rather more rarely, the opposite has also been shown to occur: the aberrant recognition of intronic sequences (pseudointrons) within normal exons. In this review, we have also decided to provide particular attention to these events as they are probably more common than previously considered and have not yet been the subject of particular attention.

One of the reasons why these two events are particularly interesting is that these types of defects are ideally suited for novel therapeutic effector molecules that are based in RNA biology. In the case of pseudoexons and pseudointrons, in fact, the major advantage of targeting this type of inclusion events is that the antisense oligonucleotides would be targeted against normal intronic sequences and thus would not remain bound to the mature mRNA (possibly to interfere with later stages of RNA processing such as export/translation).

3. Pseudoexon Activation in Cancer

In the pre-mRNA splicing field, the term “pseudoexon” has been introduced to describe exonic-like sequences that are present within intronic regions but are ignored by the spliceosomal machinery. A closer look at these sequences has often provided a reason for their inability to be recognized as normal exons: the presence of intrinsic defects in their apparently viable donor and acceptor sites [66] or of silencer elements [67–69] and the formation of inhibiting RNA secondary structures [70–72].

From a functional point of view, in most cases of pseudoexon insertion, the presence of an extraneous exon within the mature mRNA causes either the disruption of the translational reading frame or the insertion of novel amino acid sequences following translation. As a result, the normal biological properties of the resulting protein are very likely disrupted, and this can be associated with the development of disease.

Unfortunately, it is still quite hard to identify reliable pseudoexon insertion events in human genes implicated in cancers by just performing a general interrogation of databases (matching 22719 ENSEMBL protein coding genes versus 31057 entries of CanGEM Gene list) [73]. In fact, at present, it is only possible to retrieve strong candidates for alternative splicing events in protein coding genes (Table 1). As expected, a similar situation was seen when pseudo-genes were investigated (defined as genomic DNA sequences similar to normal genes but nonfunctional, although some can still be transcribed). In this case, inspection of 14775 pseudogenes returned a list of alternative splicing hits from which it is not easy to distinguish real events in expressed pseudogenes (Table 1).

More recently, some bioinformatic studies have tried to extrapolate the presence of pseudoexons in cancer tissues by developing new analysis methods of GeneChip gene expression array data [74, 75]. In this way, by comparing normal cerebellum and medulloblastoma tumors, it was possible to predict 811 significantly different expressed pseudoexons (derived from 577 genes). In addition, when nonmetastatic and metastatic medulloblastomas were compared, 13 pseudoexonic sequences were significantly expressed in a differential manner (derived from 8319 strong candidates). However, it should be noted that no experimental validation was carried out to support these predictions. Therefore, presently, manual annotations remain the most reliable system for identifying real pseudoexons.

Fortunately, the scientific community has recently identified a certain number of cases where pseudoexon inclusion has been validated in detail. For this reason, Table 2 reports all the cryptic exons described in the literature that are localized within genes whose expression was altered in cancer.

First of all, regarding the mechanisms underlying pseudoexon activation, it is interesting to note that these phenomena are caused by inherited mutations resulting in the insertion of intronic sequences in the mature mRNA [38]. In this respect, it is interesting to note that among the genes presenting pseudoexon “awakening” directly associated with cancer origin (Table 2), the creation of new splicing donor (12 events) and acceptor (7 events) sites represents the more frequent occurrence (48% and 28% of the listed events, resp.). Some of the most paradigmatic examples of cancer genes such as BRCA1, BRCA2, NF1, and ATM are included in these two sets.

On the other hand, among pseudoexons described in cancer-related genes (Table 3), the creation of 5’ splice sites and of 3’ splice sites comprise 55% and 15% of the listed records, respectively. Moreover, pseudoexon insertion can also be triggered by the deletion of nearby donor or acceptors splice sites (4 events in both lists), highlighting the importance of the genetic milieu in splicing decisions.

A deeper look into the mechanisms underlying pseudoexon activation has revealed the involvement of hnRNPs in regulation of these events. In particular, the partial inclusion of a pseudoexon of NF1 gene has been found to arise from a novel intronic mutation c.31–279A>G intron 30, creating a new acceptor splice site and activation of a cryptic 5’ splice site [76]. It has been found that both PTB and nPTB play an active role in the pseudoexon splicing by repressing its inclusion [76]. This finding is particularly interesting since it further supports the hypothesis that PTB/nPTB might have a general role in repressing weak exons and in particular most of pseudoexons [77].

Another example of mutation causing pseudoexon activation consists in the creation of novel branch site (Table 2, one event). This latter case is associated with a peculiar mutation (IVS5 ds +232 G–A) that leads to pseudoexon
Table 1: Alternative splicing events detected in protein coding genes and pseudogenes implicated in cancer.

<table>
<thead>
<tr>
<th>Alternative splicing event</th>
<th>Protein coding genes</th>
<th>CanGEM gene list (31057 entries)</th>
<th>% cancer protein coding genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cassette exon (CE)</td>
<td>13506</td>
<td>11771</td>
<td>87</td>
</tr>
<tr>
<td>Intron retention (IR)</td>
<td>8190</td>
<td>6775</td>
<td>83</td>
</tr>
<tr>
<td>Alternative 3's ss (A3SS)</td>
<td>6701</td>
<td>5685</td>
<td>85</td>
</tr>
<tr>
<td>Alternative 5's ss (A5SS)</td>
<td>6624</td>
<td>5640</td>
<td>85</td>
</tr>
<tr>
<td>Alternative first exons (AFE)</td>
<td>8050</td>
<td>7136</td>
<td>89</td>
</tr>
<tr>
<td>Alternative last exons (ALE)</td>
<td>3631</td>
<td>3233</td>
<td>89</td>
</tr>
<tr>
<td>Mutually exclusive exons (MXE)</td>
<td>3376</td>
<td>3006</td>
<td>89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alternative splicing event</th>
<th>Pseudogenes</th>
<th>CanGEM gene LIST (31057 entries)</th>
<th>% cancer pseudogenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cassette exon (CE)</td>
<td>434</td>
<td>139</td>
<td>32</td>
</tr>
<tr>
<td>Intron retention (IR)</td>
<td>465</td>
<td>187</td>
<td>40</td>
</tr>
<tr>
<td>Alternative 3's ss (A3SS)</td>
<td>241</td>
<td>91</td>
<td>38</td>
</tr>
<tr>
<td>Alternative 5's ss (A5SS)</td>
<td>218</td>
<td>82</td>
<td>38</td>
</tr>
<tr>
<td>Alternative first exons (AFE)</td>
<td>91</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>Alternative last exons (ALE)</td>
<td>104</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Mutually exclusive exons (MXE)</td>
<td>50</td>
<td>16</td>
<td>32</td>
</tr>
</tbody>
</table>

Homo sapiens genes (GRCh37.p10) dataset (62252 total genes) was used to filter alternative splicing events in protein coding genes (22719 entries) and pseudogenes (14775 entries) and verified their presence in the Cancer Genome Mine database (31057 entries—CanGEM, http://www.cangem.org/).

Table 2: Pseudoexon insertion events directly involved in cancer pathology.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Description</th>
<th>Entrez gene</th>
<th>Size (bp) pseudoexon</th>
<th>Activating mutation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>Adenomatosis polyposis coli</td>
<td>324</td>
<td>167</td>
<td>5'ss creation</td>
<td>[78]</td>
</tr>
<tr>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
<td>470</td>
<td>58</td>
<td>3'ss creation</td>
<td>[79]</td>
</tr>
<tr>
<td>ATM</td>
<td>65</td>
<td>SRE deletion</td>
<td>[80]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>137</td>
<td>5'ss creation</td>
<td>[81]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>212</td>
<td>5'ss creation</td>
<td>[82]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>25</td>
<td>Genomic rearrangement</td>
<td>[83]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCR-ABL</td>
<td>V-abl Abelson murine leukemia viral oncogene homolog 1 fusion</td>
<td>613</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td>Breast cancer 1, early onset</td>
<td>672</td>
<td>66</td>
<td>3'ss creation</td>
<td>[86]</td>
</tr>
<tr>
<td>BRCA2</td>
<td>Breast cancer 2, early onset</td>
<td>675</td>
<td>93</td>
<td>Downstream 3'ss deletion</td>
<td>[87]</td>
</tr>
<tr>
<td>ESRT1</td>
<td>Estrogen receptor 1</td>
<td>3467</td>
<td>69</td>
<td>5'ss creation</td>
<td>[89]</td>
</tr>
<tr>
<td>MSH2</td>
<td>MutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)</td>
<td>4436</td>
<td>75</td>
<td>5'ss creation</td>
<td>[90]</td>
</tr>
<tr>
<td>NFI</td>
<td>Neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)</td>
<td>4763</td>
<td>67/99</td>
<td>3'ss creation</td>
<td>[91]</td>
</tr>
<tr>
<td>NFI</td>
<td>70</td>
<td>5'ss creation</td>
<td>[92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>107</td>
<td>5'ss creation</td>
<td>[92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>172</td>
<td>3'ss creation</td>
<td>[93]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>58</td>
<td>3'ss creation</td>
<td>[94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>76</td>
<td>5'ss creation</td>
<td>[94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>54</td>
<td>5'ss creation</td>
<td>[95]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>177</td>
<td>5'ss creation</td>
<td>[92, 96, 97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>4771</td>
<td>106</td>
<td>BP creation</td>
<td>[98]</td>
<td></td>
</tr>
<tr>
<td>NFI</td>
<td>5925</td>
<td>103</td>
<td>3'ss creation</td>
<td>[99]</td>
<td></td>
</tr>
<tr>
<td>TSC2</td>
<td>Tuberous sclerosis 2</td>
<td>7249</td>
<td>89</td>
<td>5'ss creation</td>
<td>[100]</td>
</tr>
<tr>
<td>WRN</td>
<td>Werner syndrome</td>
<td>7486</td>
<td>106</td>
<td>5'ss creation</td>
<td>[101]</td>
</tr>
<tr>
<td>WRN</td>
<td>69</td>
<td>3'ss creation</td>
<td>[102]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Table 3: Pseudoexons described in cancer-related genes.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Description</th>
<th>Entrez gene</th>
<th>Size (bp)</th>
<th>Activating mutation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCC8</td>
<td>ATP-binding cassette, sub-family C (CFTR MRP), member 8</td>
<td>6833</td>
<td>76</td>
<td>5’s ss creation</td>
<td>[103]</td>
</tr>
<tr>
<td>ALDH7A1</td>
<td>Aldehyde dehydrogenase 7 family, member A1</td>
<td>501</td>
<td>36</td>
<td>5’s ss mutation</td>
<td>[104]</td>
</tr>
<tr>
<td>CD40LG</td>
<td>CD40 ligand (TNF superfamily, member 5, hyper-IgM syndrome)</td>
<td>959</td>
<td>59</td>
<td>5’s ss creation</td>
<td>[105]</td>
</tr>
<tr>
<td>CEP290</td>
<td>Centrosomal protein 290 kDa</td>
<td>80184</td>
<td>128</td>
<td>5’s ss creation</td>
<td>[106, 107]</td>
</tr>
<tr>
<td>CHM</td>
<td>Choroideremia (Rab escort protein 1)</td>
<td>1121</td>
<td>98</td>
<td>3’s ss creation</td>
<td>[108]</td>
</tr>
<tr>
<td>COLAA3</td>
<td>Collagen, type IV, alpha 3 (Goodpasture antigen)</td>
<td>1285</td>
<td>74</td>
<td>3’s ss creation</td>
<td>[109]</td>
</tr>
<tr>
<td>COL11A1</td>
<td>Collagen, type XI, alpha 1</td>
<td>1301</td>
<td>50</td>
<td>5’s ss creation</td>
<td>[110]</td>
</tr>
<tr>
<td>CTDPI</td>
<td>CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) phosphatase, subunit 1</td>
<td>9150</td>
<td>95</td>
<td>5’s ss creation</td>
<td>[111]</td>
</tr>
<tr>
<td>CYBB</td>
<td>Cytochrome b-245, beta polypeptide (chronic granulomatous disease)</td>
<td>1536</td>
<td>56</td>
<td>5’s ss creation</td>
<td>[112]</td>
</tr>
<tr>
<td>CYBB</td>
<td>Cytochrome b-245, beta polypeptide</td>
<td>1586</td>
<td>94</td>
<td>Upstream 5’s ss mutation</td>
<td>[113]</td>
</tr>
<tr>
<td>CYPI7A1</td>
<td>Cytochrome P450, family 17, subfamily A, polypeptide 1</td>
<td>5860</td>
<td>152</td>
<td>5’s ss creation</td>
<td>[114]</td>
</tr>
<tr>
<td>QDPR</td>
<td>Quinoid dihydropteridine reductase</td>
<td>1806</td>
<td>44</td>
<td>5’s ss creation</td>
<td>[115]</td>
</tr>
<tr>
<td>DPYD</td>
<td>Dihydropyrimidine dehydrogenase</td>
<td>2200</td>
<td>73</td>
<td>5’s ss creation</td>
<td>[116]</td>
</tr>
<tr>
<td>FBN1</td>
<td>Fibrillin</td>
<td>2690</td>
<td>56</td>
<td>5’s ss creation</td>
<td>[117]</td>
</tr>
<tr>
<td>GHR</td>
<td>Growth hormone receptor</td>
<td>2990</td>
<td>68</td>
<td>5’s ss creation</td>
<td>[118, 119]</td>
</tr>
<tr>
<td>GUSB</td>
<td>Glucuronidase, beta</td>
<td>3033</td>
<td>141</td>
<td>5’s ss creation</td>
<td>[120]</td>
</tr>
<tr>
<td>HADH</td>
<td>Hydroxyacyl-coenzyme A dehydrogenase</td>
<td>3032</td>
<td>141</td>
<td>5’s ss creation</td>
<td>[121]</td>
</tr>
<tr>
<td>HSPG2</td>
<td>Heparan sulfate proteoglycan 2</td>
<td>3339</td>
<td>141</td>
<td>5’s ss creation</td>
<td>[122]</td>
</tr>
<tr>
<td>SMARCB1</td>
<td>SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1</td>
<td>6598</td>
<td>72</td>
<td>5’s ss creation</td>
<td>[123]</td>
</tr>
<tr>
<td>ISCU</td>
<td>Iron-sulfur cluster scaffold homolog (E. coli)</td>
<td>23479</td>
<td>86</td>
<td>3’s ss creation</td>
<td>[124–125]</td>
</tr>
<tr>
<td>SLCC1A1</td>
<td>Solute carrier family 14 (urea transporter), member 1 (Kidd blood group)</td>
<td>64087</td>
<td>136</td>
<td>Internal 7 kb deletion</td>
<td>[126]</td>
</tr>
<tr>
<td>MCCC2</td>
<td>Methylcrotonyl-coenzyme A carboxylase 2 (beta)</td>
<td>4240</td>
<td>102</td>
<td>SRE deletion</td>
<td>[127]</td>
</tr>
<tr>
<td>MLCI</td>
<td>Megalencephalic leukoencephalopathy with subcortical cysts 1</td>
<td>23209</td>
<td>246</td>
<td>5’s ss creation</td>
<td>[128]</td>
</tr>
<tr>
<td>MTRR</td>
<td>5-methyltetrahydrofolate-homocysteine methyltransferase reductase</td>
<td>4552</td>
<td>140</td>
<td>SRE creation</td>
<td>[129]</td>
</tr>
<tr>
<td>GPR143</td>
<td>G protein-coupled receptor 143</td>
<td>5095</td>
<td>84</td>
<td>SRE creation</td>
<td>[130]</td>
</tr>
<tr>
<td>OTC</td>
<td>Ornithine carbamoyltransferase</td>
<td>5009</td>
<td>135</td>
<td>3’s ss creation</td>
<td>[131]</td>
</tr>
<tr>
<td>OFD1</td>
<td>Ornithine aminotransferase (gyrate atrophy)</td>
<td>4942</td>
<td>142</td>
<td>5’s ss creation</td>
<td>[132]</td>
</tr>
<tr>
<td>OTC</td>
<td>Ornithine carbamoyltransferase</td>
<td>5095</td>
<td>135</td>
<td>5’s ss creation</td>
<td>[133]</td>
</tr>
<tr>
<td>PCCA</td>
<td>Propionyl-coenzyme A carboxylase, alpha polypeptide</td>
<td>5096</td>
<td>72</td>
<td>SRE creation</td>
<td>[134]</td>
</tr>
<tr>
<td>PCCB</td>
<td>Propionyl-coenzyme A carboxylase, beta polypeptide</td>
<td>5251</td>
<td>100</td>
<td>5’s ss creation</td>
<td>[135]</td>
</tr>
<tr>
<td>PHEX</td>
<td>Phosphate regulating endopeptidase homolog, X-linked (hypophosphatemia, vitamin D resistant rickets)</td>
<td>5251</td>
<td>170</td>
<td>5’s ss creation</td>
<td>[136]</td>
</tr>
</tbody>
</table>
inclusion in NF2 gene by creating a consensus branch point. In particular, it has been found that the G>A transition, occurring at position −18 from the acceptor site of NF2 IVS5, works in combination with existing consensus splice acceptor and donor sites and causes the formation of an extra exon 5a in the NF2 gene that introduces a premature stop codon [98].

Finally, in spite of being poorly represented, gross genomic rearrangements can bring together or expose splice site sequences that would normally be very distant from each other or absent from the original sequence (one event).

Pseudoe exon activation can also be caused by mutations that cause the creation/deletion of splicing regulatory elements (SREs, 6 events). Among these events, one of the most representative examples of the complex regulatory networks that can underlie pseudoe xon insertion is represented by the identification of an unusual splicing defect directly related with neoplasia (Figure 1). In normal conditions, the repression of the 3' ss in the wild-type pseudoe xon sequence (ATMWT) is shown to be a direct consequence of U1snRNP binding in correspondence to the ISPE sequence. The importance of this U1snRNP binding was twofold: first of all, to sterically hinder 3' ss recognition by U2snRNP and secondly, to also inhibit binding of the SRSF1 splice factor to the stem-loop sequence formed by this pseudoe xon (Figure 1(b)) [71, 72]. As a result, this internal U1snRNP binding event to the ISPE sequence causes an unproductive U2snRNP association with the 3' ss that results in the complete inhibition of pseudoe xon inclusion in normal conditions (Figure 1(b)). Following the deletion of the GUAA motif observed in the patient, U1snRNP binding to the ISPE is relieved, and SRSF1 is also free to bind to the internal enhancer site that becomes available (Figure 1(c)). Taken together, these two events result in efficient recruitment of U2snRNP to the branch site and a better recruitment of U1snRNP to the rather inefficient 5' gc splice site that lead to efficient inclusion of the pseudoe xon in the mature ATM mRNA [72].

This example clearly shows the complexity of pseudoe xon activation events that relies on a variety of factors well beyond the simple initial mutation (i.e., 5' ss or 3' ss creation), and that can include the eventual presence of enhancer and silencer elements within the cryptic exon or intro, RNA secondary structures, and probably several other mechanisms that have not yet been described in detail.

4. Pseudoe xtron Activation in Cancer

Pseudoe xtrons (PSIs) represent an intriguing set of intron-like sequences localized within exons that can undergo alternative splicing and that therefore can be included in or excluded from the ORF within the final mRNA species. Although the examples of pseudoe xtrons reported in the review are not associated with activating mutations, we have included them because of the peculiarity of events insofar that they do not represent simple intron retention events, but rather alternatively spliced intraxonic sequences whose behaviour can resemble that of introns under particular circumstances (e.g., occurrence of previous splicing events in the processed transcript). It is this regulated use of the intraxonic splice sites that fits well with the idea that these intraxonic splice site represent “false” introns, even in the absence of genetic alterations.

In general, however, pseudoe xtrons are less well described compared with pseudoe xons probably because of the difficulty in detecting them using available technology (a situation that may well soon change for the better because of the introduction of more sensitive and comprehensive technical approaches such as RNA sequencing). At the moment, only three examples of PSI have been characterized, and for at least two of these pseudoe xtrons there are data supporting their relevance in the pathogenesis of cancer.

4.1. Fibronectin: IIICS-C5. Fibronectin (FN) is an extracellular matrix protein whose functions range from cell adhesion and migration to wound healing and oncogenic transformation [145]. The functional complexity of FN is
related to the structural diversity arising from cell type-specific alternative splicing [146]. The fibronectin pre-mRNA undergoes alternative splicing primarily at three sites: two extra domain exons encoding extra structural repeats and a region of nonhomologous sequence called the type-III connecting segment (IIICS).

The IIICS domain is divided into three subdomains of 75 bp, 192 bp, and 93 bp, respectively, that can be alternatively spliced, thus generating up to five variant mRNAs in humans [147, 148]. The subdomains of 75 bp and 93 bp encode for two cell-specific binding sites, CS1 (residues 1–25 of the IIICS), and CS5 (residues 90–109 of the IIICS), respectively, that interact with the integrin α4β1 with different affinity [149, 150]. Indeed, the CS1 site has approximately 20-fold higher affinity for integrins than the CS5 site [151].

Because of its intraexonic localization and splicing behavior, the 93 bp segment can be considered a pseudointron (Figure 2(a)). Different tissue-specific and disease-associated changes in the variants of the IIICS region have been reported [152–154], and the CS5 pseudointron is upregulated in foetal tissue and in adult liver [155].

For the purpose of this review, it is important to note that alternative splicing of the IIICS-CS5 subdomain can influence the onset and progression of cancers by affecting
Figure 2: Pseudointrons whose alternative splicing can be altered in tumors. Schemes of pseudointron alternative splicing events shown to occur within the Fibronectin (IIICS, (a)), Acetylcholinesterase (AchE, (b)), and Thrombopoietin (THPO, (c)) genes. Pseudointrons are shown as white boxes in IIICS (CS5-93 nt), AchE (4'R), and THPO (116 nt). Solid lines indicate introns and dotted lines distinguish the different splicing events. The C-terminal sequence of each splicing variant for each gene is also shown. For IIICS and THPO, additional exonic elements undergoing alternative splicing are shown as gray boxes. In the AchE gene, gray triangles indicate stop codon within different ORFs. See text for definition of each splice variant.

Fibronectin activities related to tissue organization and cell-ECM interactions. In keeping with this hypothesis, in vitro studies have shown that CS5 can modulate the spreading of melanoma cells [151].

4.2. Acetylcholinesterase: AChE-R. Acetylcholinesterase is a serine specific hydrolase that hydrolyzes acetylcholine, and its main function is related to the clearance of the neurotransmitter from the synaptic cleft [156]. Nonetheless, the observation that the expression of AChE is not limited to cholinergic tissues has suggested the presence of additional functions [157].

From the splicing process point of view, mammal AChE has been shown to undergo alternative splicing resulting in expression of 3 isoforms, called R (“readthrough”), H (“hydrophobic”), and T/S (“tailed” or “synaptic”), that differ in their subcellular localization, tissue distribution, and developmental pattern of expression [158]. Whereas, AChE-H or AChE-T is the most abundant and common isoforms, the AChE-R expression is low and restricted in specific cell lines, cell differentiation stage, or muscle development [159, 160].

Interestingly, the AChE-R mRNA arises by retention of pseudointron 4, that leads to generation of the E1-E2-E3-E4-I4-E5 mRNA transcript (Figure 2(b)). Since pseudointron 4 includes a stop codon when it is included in mature RNA, the resulting protein will be prematurely terminated by a C-terminus translated from the 5’ region of I4. This variant is expected to remain monomeric [161].

In general, the AChE-R transcript can be coexpressed with the AChE-T or AChE-H transcripts, and variations in their relative levels seem to be important for modulation of AChE functions, as well as for the pathogenesis of neurological and autoimmune disorders [161, 162]. Intriguingly, recent studies have found that AChE-R mRNA accumulates in primary human astrocytomas, and that its presence is correlated with their grade of aggressiveness [163]. In human U87MG glioblastoma cells, it was found that AChE-R protein variant can support proliferation of glioblastoma tumors by forming a complex with the scaffold protein RACK1 and protein kinase Ce [164].

In addition, increased levels of an N-terminally extended N-AChE-R isoform were detected in human testicular tumors indicating that the generation of the AChE-R variant can be associated with the utilization of an alternative promoter in the transformed cells.

Finally, related functional studies have suggested that both AChE-R variants (AChE-R and N-AChE-R) might be crucial for increasing the cellular ATP levels and might support selective metabolic advantages as well as genotoxic resistance by altering p73 gene expression [165]. Although it is not yet clear whether the AChE-R (as well as the other AChE splicing variants) can be related directly to tumorigenesis, the observed downregulation of all these splicing isoforms...
in colorectal carcinoma suggests a crucial role for tumor development [166].

4.3. Thrombopoietin: THPO-3, THPO-5, and THPO-6. A third example of pseudointron possibly implicated in cancer pathogenesis is found within the Thrombopoietin (THPO) gene. THPO is the most important cytokine for regulation of platelet production, and its expression has also been reported in skeletal muscle, ovary, testis, and central nervous system (CNS) [167–169].

The THPO gene shows a complex pattern of alternative splicing [170]; among the six splice variants that can be generated, three of them (THPO-3, THPO-5, and THPO-6) show the peculiar alternative splicing of 116 nucleotides within exon 6 (that can therefore be classified as a pseudointron) (Figure 2(c)) [171].

Contrary to the Fibronectin IIICS-C5 and AchE-R pseudointrons, the involvement in cancerogenesis of the variants deriving from the alternative splicing of the 116 nt pseudointron is less characterized. Nonetheless, functional studies have demonstrated that the protein translated from mouse THPO-3 is not secreted efficiently [172–174]. These results have suggested that THPO-3 might be implicated in the fine regulation of THPO full-length levels and in the modulation of several biological functions beyond thrombopoiesis. Interestingly, parallel studies have also shown an altered expression pattern of the THPO-3, THPO-5, and THPO-6 splicing variants in human carcinomas [175]. As a result, these observations have led researchers to hypothesize that alterations in alternative splicing of the 116 nt-THPO pseudointron might be employed as biomarker of tumor formation or progression [170].

In conclusion, although currently there are few examples of pseudointron removal, at least three of these events have been associated with different levels of cancerogenesis, ranging from transformation to cancer progression. Therefore, it is advisable that future high throughput screening analysis should be able to better associate the alternative splicing process of these pseudointrons with tumoral events. In particular, with regard to their use as possible biomarkers of transformation or of cancer staging/grading.

5. Therapeutic Outlooks

Regarding therapies, it is likely that all this information on the importance of alternative splicing for tumor development and progression may soon become extremely important for the development of novel therapeutic effectors in the fight against cancer [176, 177]. Indeed, one of the reasons why these defects are particularly interesting is the opportunity they provide for RNA therapies and the chance to actually make the jump from the bench to the bedside.

Currently, the field of RNA therapy is rapidly growing especially in the inhibition of undesirable splice site choices by the use of antisense oligos [178–180]. This technique, pioneered by Dominski and Kole [181], is based on the use of suitably modified antisense oligonucleotides that target specific sequences within introns and exons and block the recognition by their cellular binding factors. Therefore, they are ideally suited to block unwanted splice sites that may become activated following their creation or activation. It is interesting that the role of using antisense strategies to regulate exon inclusion also occurs in physiological conditions. An example of this is snoRNA HBII-52 in the regulation of exon Vb inclusion in the serotonin receptor 2C [182].

To this moment, a few pilot studies aimed at inhibiting the inclusion of a pseudoxeon in patient cell lines have already been reported. In particular, antisense oligonucleotides have been used to target newly created 5’ or 3’ splice sites deep within intronic regions of the NFI gene to restore the normal splicing profile [91, 92]. Similar approaches have also been successfully used to target deep intronic mutations causing pseudoxeon activation within BRCA2 intron 12 (c.6937+594T>G) [88], ATM intron 11 (c.1236–405C>T) [82], and ATM IVS19 (c.2639–384A>G) [79].

Recent advances in antisense-mediated exon skipping for DMD suggest that not only the splice sites but also exonic splicing enhancer sites or branch points might represent potential targets for inducing pseudoxeon skipping [183, 184], and that the action of ASOs might be potentiated by coadministration of drugs, as shown for Duchenne muscular dystrophy [185].

An intriguing variation on the theme of the antisense oligonucleotides, successfully tested for inhibiting the 5’ splice site of Bcl-XL, might consist in the use of tailed oligonucleotides containing a portion annealing to sequences immediately upstream of the target donor splice site joined to a nonhybridizing 5’ tail that includes binding sites for the hnRNP A1/A2 proteins [186].

In spite of being prominent, antisense technologies are not the only possibility available for targeting these aberrant splicing events. An alternative to antisense usage is the development of siRNA molecules capable of targeting the newly created system. In this manner, only the mRNAs containing the pseudoxeon might be selectively degraded leaving any residual normal mRNA to be properly processed [187]. The disadvantage of simply degrading the pseudoxeon-bearing transcript (rather than acting on the splicing event itself) is that everything depends on the level of pseudoxeon inclusion (less than 100%) and on the status of the other allele (i.e., whether it is normally expressed or not).

Finally, small molecule compounds might be also considered as possible therapeutic options in order to prevent the recognition of pseudoxeons or modulate the recognition of pseudointrons. For example, it has been shown that sodium butyrate, an histone deacetylase inhibitor known to upregulate the expression of Htra2-beta1 and SC35, promotes skipping of the pseudoxeon activated by the CFTFR 3849+10kbC>T mutation and can restore functional CFTR channels [188]. In addition, the splicing of different NFI skipped exons as a result of mutations in cis-acting sequences has been restored with administration of the small molecule kinetin [189].

More recently, a growing body of research indicates that TALE nucleases (TALENs) have been used with great success in a number of organisms to generate site-specific DNA variations [190]. As a result, this approach is another good
candidate for an innovative therapeutic anticancer strategy for correction of deep intronic mutations that create novel splice sites.

Of course, crossing the difficult gap between the bench to the bedside will not be immediate even in the presence of highly efficient therapeutic molecules, whether antisense oligos, siRNAs, or others. Indeed, before any of these can be used in humans to treat cancer, several additional factors need to be considered. First of all, there is the development of an efficient and safe carrier system to deliver these compounds into the human body. Secondly, these systems will need to be optimized in order to achieve the best balance between successful delivery, intrinsic toxicity (if any), and avoidance of undesired immune responses (in the case of antisense oligos). Finally, even when all these achievements are met, there will still be the need to optimize recurrent-administration protocols to determine the uptake levels, clearance, and accumulation in various tissues (this is an often overlooked, since none of these methods will cause permanent correction of mRNA splicing defects).

Nonetheless, the first results of this exciting news strategy are available, and there is the distinct possibility that RNA-based treatment of cancer may soon enter the application stage.

Acknowledgments

This work was supported by the University of Trieste-Finanziamento per Ricercatori di Ateneo to Maurizio Romano and the Higher Education Founding Council for England (HEFCE) and Cancer research UK (CRUK) to Diana Baralle.

References


[63] L. Michelle, A. Cloutier, J. Toutant et al., “Proteins associated with the exon junction complex also control the alternative...


E. Ars, E. Serra, J. Garcia et al., “Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1,” Human Molecular Genetics, vol. 9, no. 2, pp. 257–247, 2000.


