SURVEY AND SUMMARY

Defective splicing, disease and therapy: searching for master checkpoints in exon definition

Emanuele Buratti, Marco Baralle and Francisco E. Baralle*

International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano 99, 34012 Trieste, Italy

ABSTRACT

The number of aberrant splicing processes causing human disease is growing exponentially and many recent studies have uncovered some aspects of the unexpectedly complex network of interactions involved in these dysfunctions. As a consequence, our knowledge of the various *cis*- and *trans*-acting factors playing a role on both normal and aberrant splicing pathways has been enhanced greatly. However, the resulting information explosion has also uncovered the fact that many splicing systems are not easy to model. In fact we are still unable, with certainty, to predict the outcome of a given genomic variation. Nonetheless, in the midst of all this complexity some hard won lessons have been learned and in this survey we will focus on the importance of the wide sequence context when trying to understand why apparently similar mutations can give rise to different effects. The examples discussed in this summary will highlight the fine 'balance of power' that is often present between all the various regulatory elements that define exon boundaries. In the final part, we shall then discuss possible therapeutic targets and strategies to rescue genetic defects of complex splicing systems.

INTRODUCTION

The study of the connections between defective splicing and occurrence of disease has become a central issue in the medical research field (1–7). This is a consequence of the growing list of pathological alterations that can be directly linked with aberrant splicing processes. In fact, for many genes, genomic variations which affect the splicing process may represent up to 50% of all mutations that lead to gene dysfunction (8,9). This awareness has also led to the critical reappraisal of many past observations, and it is now clear that substitutions which had for a long time been regarded as harmless synonymous changes in protein coding regions may have some very severe consequences on splicing processes, and thus on the appearance of disease (10).

The underlying biological reason for such an exuberant growth in the number of diseases derived from splicing derangements is due not only on the intrinsic properties of the exon itself (presence of regulatory elements within the exon, strength of the splice sites etc.) but also on the wider environment within which the exon is found, such as the presence of regulatory elements dispersed throughout the gene at much longer distances (11–14), the splicing events that have taken place before the exon of interest (15,16) or the type of promoter architecture and the consequent polymerase II processivity (17). This means that splicing spoilers can occur virtually everywhere within any given gene and many of the novel types of mutations identified have still poorly characterized functional mechanisms (18). Furthermore, alternative splicing processes (19) are responsible for increasing the functional diversity of the human genome (20–24). Therefore, it is very likely that variations anywhere in the genome may alter the delicate modulation of the splicing process and produce pathological alterations (3).

Numerous methodological developments have also aided researchers recently in the task of building connections between splicing and disease. For example, the refinement of minigene-based technologies for alternative splicing analysis initially described 20 years ago (25) has allowed a relatively fast approach to identify splicing spoilers and to study their underlying functional mechanisms (26,27). In addition, a great boost for the preliminary identification of general splicing regulatory sequences (SREs) has been the development of SELEX methodologies to characterize binding specificities of *trans*-acting factors involved in splicing regulation (28,29).

Last but not least, scientific interest in this subject has also been steadily growing as potential applications in the molecular therapy field have begun to be appreciated and have started to be tested, such as the recent advances in the use of bi-functional or inhibitory oligonucleotides in rescuing aberrant splicing events (3,30–35), in the engineering of specific compounds by chemical synthesis to target splicing factor protein families (36–38), or in the development of novel vectors based on U7snRNA to rescue protein functionality by exon skipping (39).

---

*To whom correspondence should be addressed. Tel: +39 040 3757337; Fax: +39 040 3757361; Email: baralle@icgeb.org

The fact that even single nucleotide changes in a pre-mRNA sequence away from the canonical AG/GU nucleotides at the splice junctions are ever more frequently been described as capable of affecting the splicing process (40) was rather unexpected. Indeed, computational analyses confirm that when most of the main determinants of exon definition (3'ss, 5' ss, branch site and pY tract) are close to the consensus minimal variations may be possible without impairing recognition by the spliceosome (41). This conclusion is also consistent with the rather loose sequence-specificities of many essential splicing factors. For example, *in vitro* studies have shown that a minimal amount of all the potential base pairings between the 5' splice site and U1snRNA is sufficient for its correct recognition by U1snRNP (42). Moreover, even in the absence of the 5' tail of U1snRNA, the U1snRNP particle is capable of selecting a 5'ss-like sequence probably through the U1-C subunit (43). Finally, the observation that over-expression of SR proteins could compensate for absence of U1snRNP *in vitro* (44,45) provides support to the view that the splicing process can be precise even in the absence of key participants.

Reality, however, has painted a much more complex picture, with the micro-environment having an important role in the ability of a splice site to tolerate a nucleotide transition (46). Furthermore, microarray technology and comparative expressed sequence tags analyses are now vividly depicting a world of widespread differences in alternative splicing levels, both within normal tissues or following pathological alterations (47–49). For this reason, particular effort has been devoted recently to characterize splicing abnormalities in cell transformation and cancer (50,51) as they may provide an additional therapeutic target for this kind of diseases.

The fact that splicing regulation appears to be a linked network (52–56) means that researchers and clinicians have to consider as a potential splicing mutation virtually every substitution that is encountered within a pre-mRNA sequence (18). The recent evidences that many splicing factors may also be actively involved in other important cellular processes, such as genome stability (57) or mRNA transport and translation (58–61), have just begun to unravel all the connections of the RNA processing system (62,63). In fact, rather than becoming simplified, splicing regulation in the new millenium could be best characterized as right on the edge of chaos. Figure 1 shows a simplified view of how the number of accepted players and functions have grown in recent years. It is for this reason that the connotation of 'difficult to characterize' is increasingly associated with many new splicing mutations.

The following observations exemplify the problems that the molecular machinery has to resolve to obtain proper splicing:

- Potential exon boundaries are particularly abundant in pre-mRNA molecules. For example, in the human *hprt* gene there are ∼10 times more consensus splice site than those effectively recognized by the splicing machinery (64),
- The splicing process occurs in a background of repression, as random human genomic sequences display a higher than average splicing inhibitory activity when tested in minigene systems (65–67),
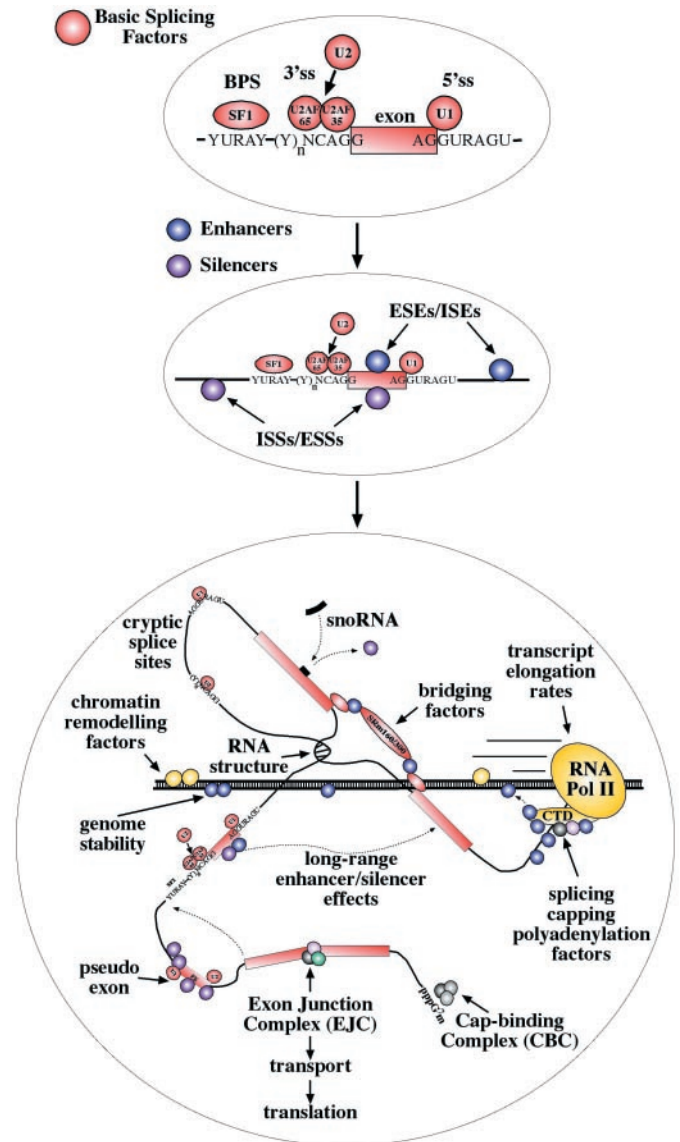

Figure 1. The three schematic representations show the increase during the last decade of the number of interactions reported to occur during RNA processing. The upper diagram depicts a classic pre-mRNA molecule in which the exon (closed box) is defined by a small number of relatively conserved sequences that bind a well-defined set of basic splicing factors. The complexity of the system begins to increase, as shown in the middle diagram, when the vast and still growing array of enhancer and silencer factors are to be added to the basic picture. These factors can bind particular enhancer and silencer sequences either within the exon itself (ESE and ESS) or in the nearby flanking regions (ISE and ISS) and help increase/decrease exon recognition. As shown in the bottom diagram, in these recent years the complexity of the regulatory elements that participate in splicing control has taken an enormous leap forward by the discovery of the potential connections between splicing, the specific structural features of pre-RNAs and many of the processes that participate in a their life-cycle such as genome stability, RNA processing speed, transport and translation.

- Positive or negative SREs are abundant in all pre-mRNAs, both within introns and exons regardless if they are constitutively or alternatively spliced (67–69).
- All the processes that deal with an RNA life cycle including transcription, splicing, transport, translation and degradation are no longer considered as separated from

each other but rather, in line with the literal translation of the Latin word '*complexus*', appear to be functionally entwined/twisted together (54,70,71). Indeed, just how close relationships within the spliceosome can grow is clearly highlighted by the recent demonstration of extensive molecular tethering between the pre-mRNA being processed and the splicing machinery, allowing processing to occur correctly even when the nucleotide chain of the pre-mRNA molecule is interrupted (72).

As a consequence, the sequence and genomic 'context' in which mutations are observed to occur are essential to explain the sometimes widely different outcomes of similar substitutions in different genes (18,55). Several excellent articles have reviewed the individual connections between aberrant splicing and disease in recent years (1–7). Our purpose will be that of using selected examples to highlight the importance of genomic and local context when investigating splicing outcomes. For example, the inactivation of donor site (5′ss) sequences produce a great variety of splicing outcomes that vividly depicts the mind boggling variety of exon/intron definition mechanisms.

WHAT MAY HAPPEN WHEN A DONOR SITE IS INACTIVATED?

The recognition of the donor site is one of the key events of the splicing process (73), and many of its peculiarities are still under investigation (42,43,74–76). Mutations that destroy natural donor sequences are usually observed to cause the skipping of their associated exon (77). However, in a considerable number of cases additional events can also take place, which include cryptic splice site activation and full intron inclusion. Now, whether any of these events takes place or not is often dependent on the individual context in which the substitution occurs. Sometimes evaluating this context may be fairly straightforward; however, in some cases it may also involve long-range genomic context interactions or be dependent on some kind of 'priming' effect to the splicing machinery by events happening upstream of the exon in question (see below).

Cryptic splice site activation

Activation of cryptic splice sites is probably the most prominent outcome (after exon skipping) that follows the inactivation of the original donor site. In fact, cryptic splice site activation in the β-globin gene was one of the first medically relevant splicing defects described (78,79). The usage of alternative splice sites following the inactivation of the proper site often leads to a frameshift in the protein coding sequence. It should be noted, however, that cryptic splice site sequences do not simply represent a harmful background to normal splicing pathways. Rather, it is now increasingly clear that they have played a very important role in the still very much debated evolutionary mechanisms of intron gain/loss (80) and recent work has suggested that in some cases they act as preferential sites for intron insertion (81,82).

Following the disruption of the normal donor site it would be convenient to predict beforehand which sites (if any) become activated by simply looking at all the nearby

sequences that match the consensus. In practice, however, there is not always a straight correlation between the chosen cryptic site and the calculated strength of the potential sites in the region (83). Contemporary studies on donor (83) and acceptor sites (84) have established that local context features may help the splicing machinery to actively discriminate against the usage of cryptic sites and in favor of natural 5′ and 3′ss. These local context features may play a role in cryptic splice site usage after disruption of the natural site.

For example, SR protein levels have been long known to affect relative cryptic splice site usage in the human β-globin gene both *in vitro* and *in vivo* (85,86). Now, two recent reports have demonstrated a critical role of SR-binding sequences on cryptic splice site selection: in the E1α Pyruvate Dehydrogenase pre-mRNA a single point mutation away from the natural ss strengthens a weak, pre-existing SC35-binding site and has the effect of shifting donor site usage to a previously silent cryptic site (87) (Figure 2A). This mutation has an important biological impact and is linked with mental retardation. In the second case, a strong SF2/ASF-binding region in *fibrinogen* exon 7 which is not normally needed for constitutive inclusion of this exon in the mature mRNA has the ability to determine an increase in the relative usage levels of a particular cryptic donor site when the natural donor site is inactivated by a IVS7 + 1G > U mutation (88). Interestingly, changing the local context by swapping the native SF2/ASF-binding sequence with a different SR-binding ESE sequence from the *fibronectin* gene (*fn*ESE) has the ability to change the relative cryptic site usage ratios with respect to the natural fibrinogen sequence (Figure 2B).

Finally, an additional general factor that has been described to affect cryptic splice site usage is represented by RNA secondary structure (12–14,89). For example, in the early region 3 of adenovirus type 2 a stem–loop structure allows the correct and exclusive usage of its D1 natural donor site rather than a cryptic donor site (Dcr1) located 74 nt away (90,91) (Figure 2C). Disruption of this structure results in a shift towards the Dcr1 site.

Intron retention

An intron retention event is identified as the preservation of an entire intronic sequence within some forms of the final processed mRNA. Using this criteria, the frequency of normal intron retention events in the human genome has been estimated recently to be ~15% in a set of >21 000 annotated genes (92). Most of these events naturally occur in the untranslated regions of the pre-mRNA where the rate of retention is observed to be >10-fold greater with respect to protein coding sequence (92). The biological role of most of these events is currently unknown; however, retained introns within the coding region are significantly shorter than non-retained introns (92,93). Furthermore, their sequences are under selective pressure for coding potential (92), suggesting a potentially regulatory role in protein functions. In recent years, the biological significance of some physiological intron retention events has also been clearly established in several cases such as the generation of the P element and Msl2 transcripts in *Drosophila* (94,95) or in
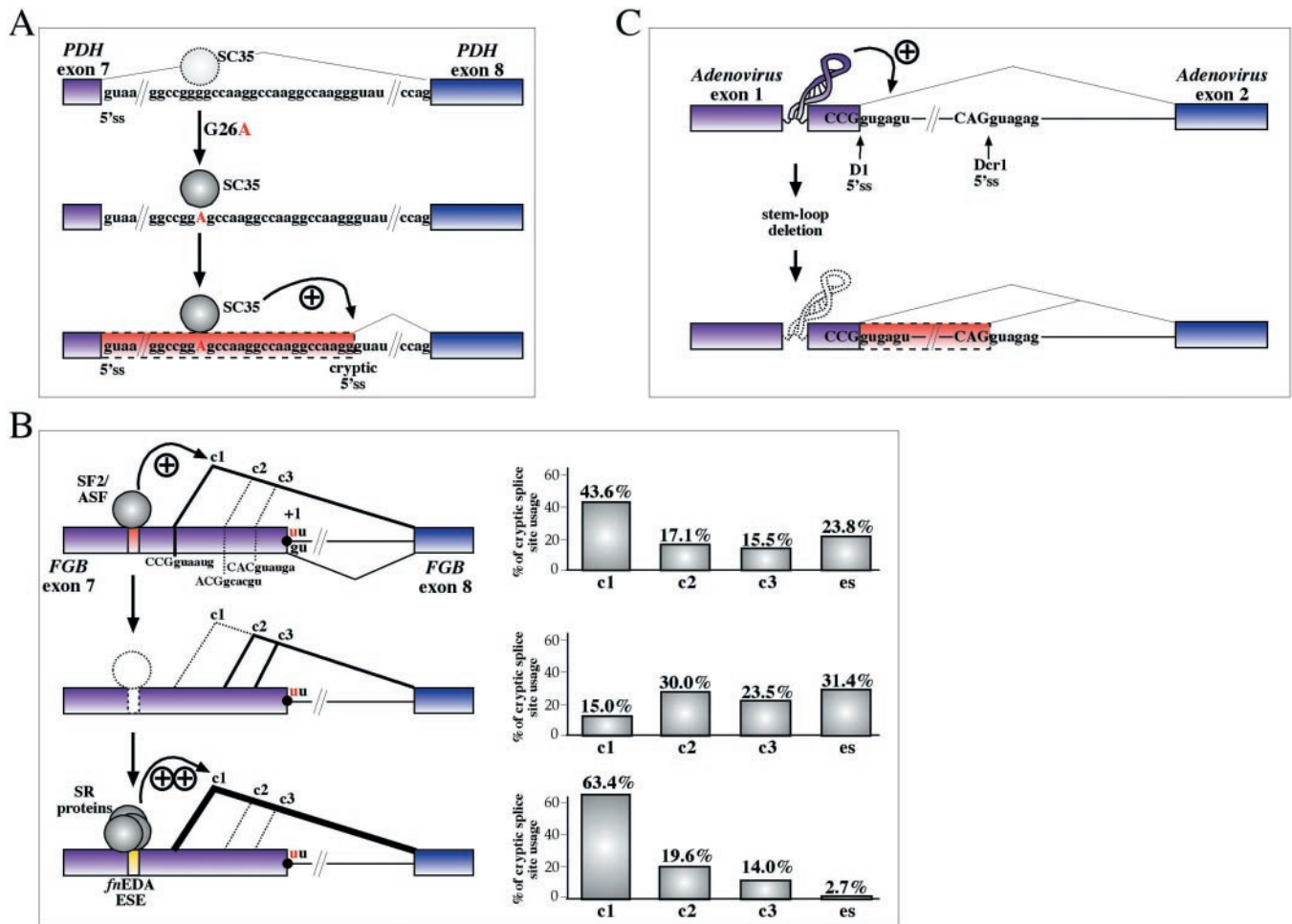
Figure 2. (**A**) Schematic diagram showing how local context arrangements can influence cryptic splice site activation. For example, in intron 7 of the *PDH* gene a G26A mutation greatly enhances a preexisting binding site for the SC35 protein (upper and middle panels). The increased binding of this protein causes the shift of donor site usage towards a cryptic donor site in the downstream sequence (lower panel). (**B**) Another example is represented by a strong SF2/ASF-binding sequence that is not normally needed by the splicing machinery to recognize *fibrinogen* exon 7 (upper panel). However, if the natural donor site is inactivated by a +1G > U substitution the relative usage of the three cryptic sites which become activated (c1, c2 and c3) and the levels of exon skipping (es) are profoundly influenced by its presence, especially c1 (middle panel). In this case, the importance of local context has also been demonstrated by replacing the natural SF2/ASF-binding sequence with a well-known ESE sequence from the *fibronectin* gene (*fn*ESE) and observing that the level of c1 usage are substantially increased with respect to the natural sequence (lower panel). (**C**) Finally, in the case of RNA secondary structure it has been shown that correct utilization of the natural donor site sequence (D1) can be achieved only in the presence of a highly conserved stem-loop structure within Adenovirus type 2 first exon (upper panel). Removal of this sequence has the consequence of activating usage of a cryptic donor site sequence (Dcr1) (lower panel).

the developmental regulation of the pro insulin messenger RNA in chicken embryos (96).

Most importantly, aberrant intron retention events have also been shown to be associated with several pathological splicing defects. For example, in the Ret tyrosine kinase gene (97) where it has been connected with the presence of pheocromocytomas, in the *HERG* gene where it has been identified as being responsible for the development of a long QT syndrome (LQT2 phenotype) in an individual which carried a +6T > C mutation in the donor site of intron 7 of this gene (98) and in a patient affected by Leigh syndrome where the retention of intron 3 in the *SURF1* gene follows a single-nucleotide mutation in its donor site (99). As similar mutations in other contexts cause exon skipping it is highly probable that local context features either in the retained intron itself or in the flanking exonic sequences determine this particular outcome. At the moment, very little

is known concerning this issue and it is to be hoped that future studies will target these particular issues.

An interesting phenomenon that also deals with intron retention/recognition has been observed by analyzing the last exon of the *TPO* gene. In 26% of the *TPO* transcripts a 116 n sequence located in the last exon is removed and hence treated as an 'alternative intron' (Figure 3A, upper panel) (15). Strikingly, this short intron behaves as such only if there has been an upstream splicing event. In fact, if the cDNA of the *TPO* gene is transfected in Hep3B cells removal of the 116 nt intron is completely abolished (Figure 3A, lower panel). The recognition of the small intron 5′ and 3′ss within *TPO*'s last exon seems to require spliceosome assembly on a previous intron, as the effect is obtained not only with the intact intron structure of the *TPO* gene but also by inserting any single intron in the cDNA sequence (Figure 3B). Although the mechanism responsible for these
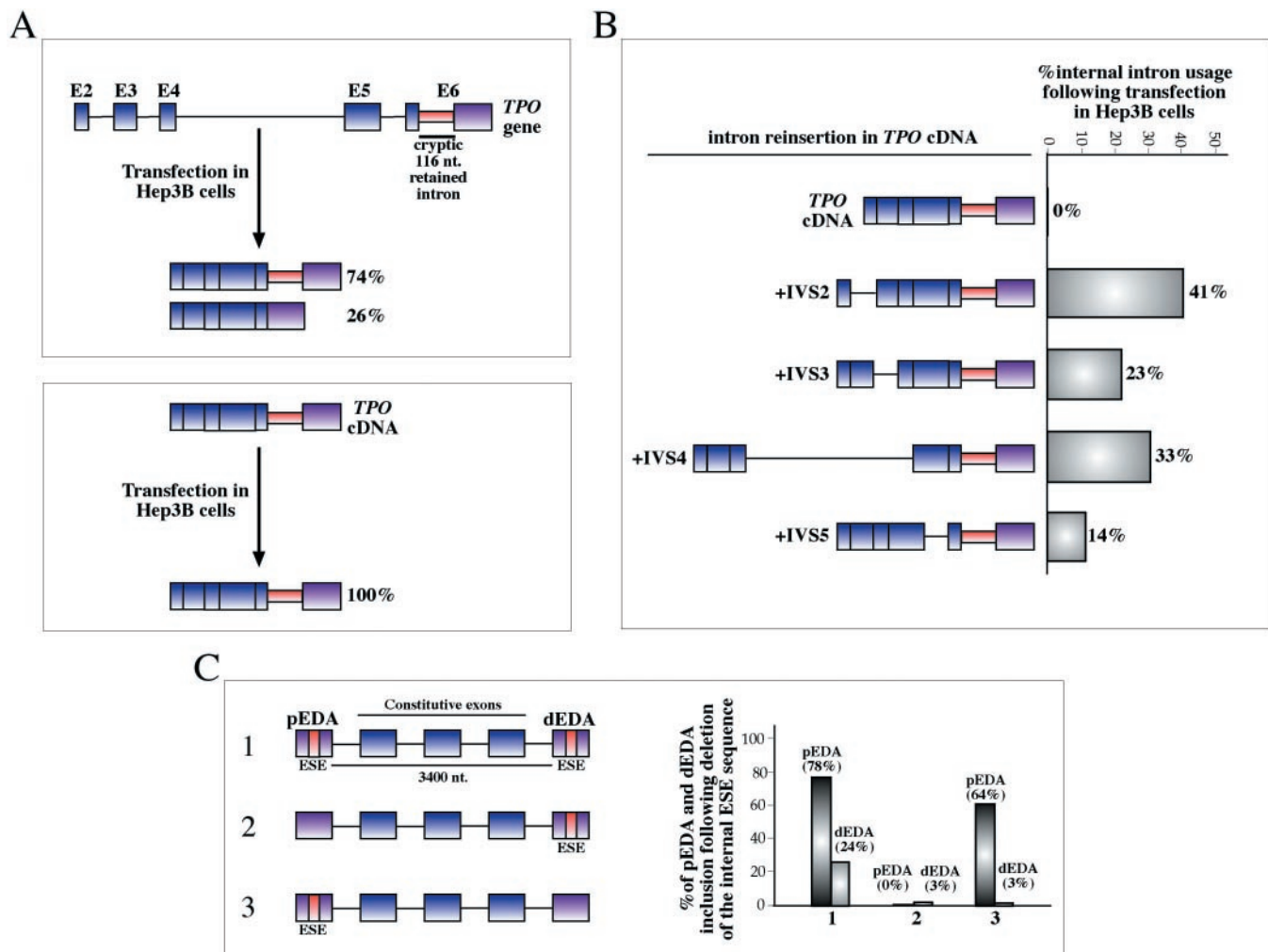
Figure 3. (**A**) This figure shows the effect on internal intron retention in exon 6 of the *TPO* gene. When the *TPO* gene is transfected in Hep3B cells only two processed mRNAs are produced, a major form lacking all the intervening introns (74%) and a minor form in which there is the removal of an additional 116 nt long sequence (26%) from the processed mRNA (upper panel, boxed in red). On the other hand, lower panel shows that when the *TPO* cDNA is transfected in Hep3B cells no processing of this internal intron can be observed. (**B**) The only way of recovering internal intron processing is that of reintroducing at least one of the normally processed introns of the *TPO* gene. An analogous situation in which the system has to be 'primed' before it can function efficiently is provided by the recent observation that coordinated splicing can occur in correspondence of distant alternatively spliced exons. (**C**) Shows such an experiment where two alternatively spliced EDA exons (pEDA and dEDA) separated by a 3400 nt long spacer containing three constitutively spliced exons are inserted in a minigene systems in either wild type or mutated form (lacking an important ESE sequence). The graph on the left shows that efficient inclusion of the proximal exon strictly correlates with inclusion efficiencies of the distal exon.

coordinated intron recognition events is currently unknown, it is clear that some kind of 'priming' event from the splicing machinery may be required for the processing of this intron. Intriguingly, a somewhat analogous mechanism, this time on exon recognition, has been proposed recently to exist in the coordination of splicing between two independent alternatively spliced regions within a fibronectin minigene (16) (Figure 3C). In this case, mutations that inhibit inclusion of the proximal exon (pEDA) correlate directly with inclusion levels of the distal exon (dEDA). Furthermore, existence of such a phenomenon has also been described *in vivo* in a mouse knock-in model (100) between two different alternatively spliced regions of the fibronectin gene (the EDA exon and the V region) (16).

These examples, together with the recent description of connections between chromatin factors (101), promoter usage and transcript elongation on the splicing process (17,102) clearly highlight how difficult it is to separate the functional effect of a given nucleotide variation from its wide genomic context, and further examples of connections between the upstream history of the transcription/splicing machinery and downstream exon definition will be discussed below.

Pseudoexon inclusion

The observation that consensus splice site sequences are also very abundant in intronic sequences indicates that introns have the potential ability to encode for 'false' exons (commonly referred to as pseudoexons) (64,66). Pseudoexons insertion in mature mRNA have been mainly reported in connection with human diseases and Table 1 shows a

Table 1. list of pseudoexon inclusion events connected with disease in humans

Gene	Pseudoexon length (bp)	Activating mutation	Reference
DMD	58	5′ss creation	182
DMD	67	5′ss creation	183
DMD	89	5′ss creation	183
DMD	90	5′ss creation	183
DMD	95	3′ss strengthening	182
DMD	149	3′ss creation	183
DMD	172/202	5′ss creation	184
CFTR	49	5′ss creation	185
CFTR	84	5′ss creation	186
CFTR	184	3′ss deletion	187
CFTR	214	5′ss creation	188
CYBB	56	5′ss strengthening	189
CYBB	94	3′ss creation	190
PHEX	50/100/170	5′ss creation	191
α-Gal A	57	SRE creation	107
ATM	65	SRE deletion	108
BRCA1	66	3′ss creation	192
GHER	69	5′ss creation	193
CHM	98	3′ss creation	194
GHR	102	5′ss creation	195
ATM	137	5′ss creation	196
OA1/GPR143	165	3′ss creation	197
β-globin	165	5′ss creation	78
NF-1	172	3′ss creation	198

comprehensive list of genes in which this kind of events has been described. Nonetheless, and possibly more important from a biological point of view, pseudoexon insertion may well represent a substrate for the evolution of new exons. Indeed, it has been proposed recently that 'exonization' of a class of particularly abundant intronic sequences (known as *Alu* repeats) may well represent a major and frequent source of 'new' exons and thus play a very important and useful role in evolution (103–105).

Despite the abundance in the genome of potential pseudoexons (intronic sequences between 50 and 200 nt in length with apparently viable 5′ss and 3′ss sites at either end) their inclusion does not seem to be a frequent event during normal pre-mRNA processing. Several different reasons have been put forward to account for this, e.g. it was originally noted that many of these 'non-spliced' pseudoexons contain multiple defects in splicing controlling regions despite their apparently good agreement with real exonic sequences (64). In keeping with this, three putative splicing silencer motifs have been found recently to be enriched in these non-spliced pseudoexons compared with their surrounding intronic regions (66). This observation is consistent with previous work which showed that human DNA is particularly rich in sequences which can inhibit splicing (65). Furthermore, an extended computational analysis of 8mer sequences that either promote or inhibit splicing in internal non-coding exons versus non-spliced pseudoexons has revealed that enhancer sequences are highly enriched in real exons whilst silencer sequences tend to cluster within pseudoexons (67). Finally, computational studies have suggested recently that pseudoexons flanking regions have a distinct tendency to form double-stranded structures that include the pseudo exon itself, providing an additional layer that differentiates

them from real exons, where flanking regions have the tendency to base pair between themselves (106).

Nonetheless, many intronic sequences seem poised on the brink of becoming exons because, as shown in Table 1, most of these disease-connected pseudoexon inclusion events originate from a single activating mutation. This mutation generally involves the *ex novo* creation of a strong donor or acceptor site, followed by the subsequent selection of weaker 'opportunistic' acceptor or donor site sequences. However, other mechanisms that cause the inclusion of pseudoexon sequences may also take place, and pseudoexon-activating mutations may also involve the creation/deletion of SREs. There are in fact at least two cases, in the α-*Gal A* (107) (Figure 4A) and the *ATM* gene (108) (Figure 4B), in which the creation/deletion of SREs has been observed to occur. In the *ATM* gene the inclusion of a 65 nt long pseudoexon is the consequence of a 4 nt deletion in *ATM* intron 20 that disrupts a high-affinity U1snRNP-binding site located within the pseudoexon sequence itself. This region was called ISPE (intron splicing processing element) and its disruption resulted in the activation of an immediately upstream cryptic acceptor site and of a downstream cryptic donor site (108). The comparison of this intronic sequence with its homologous region in the mouse has shown the importance of local context also in this case. In fact, an ISPE sequence is also present in the mouse and only a critically missing donor site downstream of this element is the reason why the mouse sequence cannot give rise to an analogous pseudoexon insertion event (109).

It is interesting to note that in recent years a lot of effort has gone into deciphering the code that underlies alternatively spliced exon recognition (29,53,69). However, the recent increase in pseudoexon and intron retention events reported in clinical investigations indicates that we should focus also on deciphering splicing codes within introns. In fact, investigations on pseudoexons may well acquire a wider importance in the future, as it has also been proposed recently that many of the these sequences which are thought to be ignored by the splicing machinery also participate in splicing regulation processes (110). The interesting notion is that the functional usage of many of these putative pseudoexon sequences may be often overlooked owing to the fact that their inclusion can lead to premature insertion of a termination codon in the mature mRNA, and its consequently rapid degradation by the nonsense-mediated decay (NMD) pathway (111). This intriguing phenomenon has been described to occur in the rat α-*tropomyosin* gene in which a pseudoexon sequence localized downstream of the mutually exclusive exons 2 (that is included only in smooth muscle tissue) and 3 (that is included in most cell types) can give rise to different splicing outcomes when it is joined to either exon (110) (Figure 4C). In particular, when joined to exon 2 the formation of an efficient ZLE (zero length exon) sequence allows to resplice exon 2 to the acceptor site of exon 4 (thus giving rise to a normally processed mRNA molecule). On the other hand, when the pseudoexon sequence is joined with exon 3 the resulting ZLE is not sufficient for the resplicing pathway and the pseudoexon 5′ss is used to join with the acceptor site of exon 4 (thus giving rise to a mRNA molecule which is rapidly degraded by NMD.
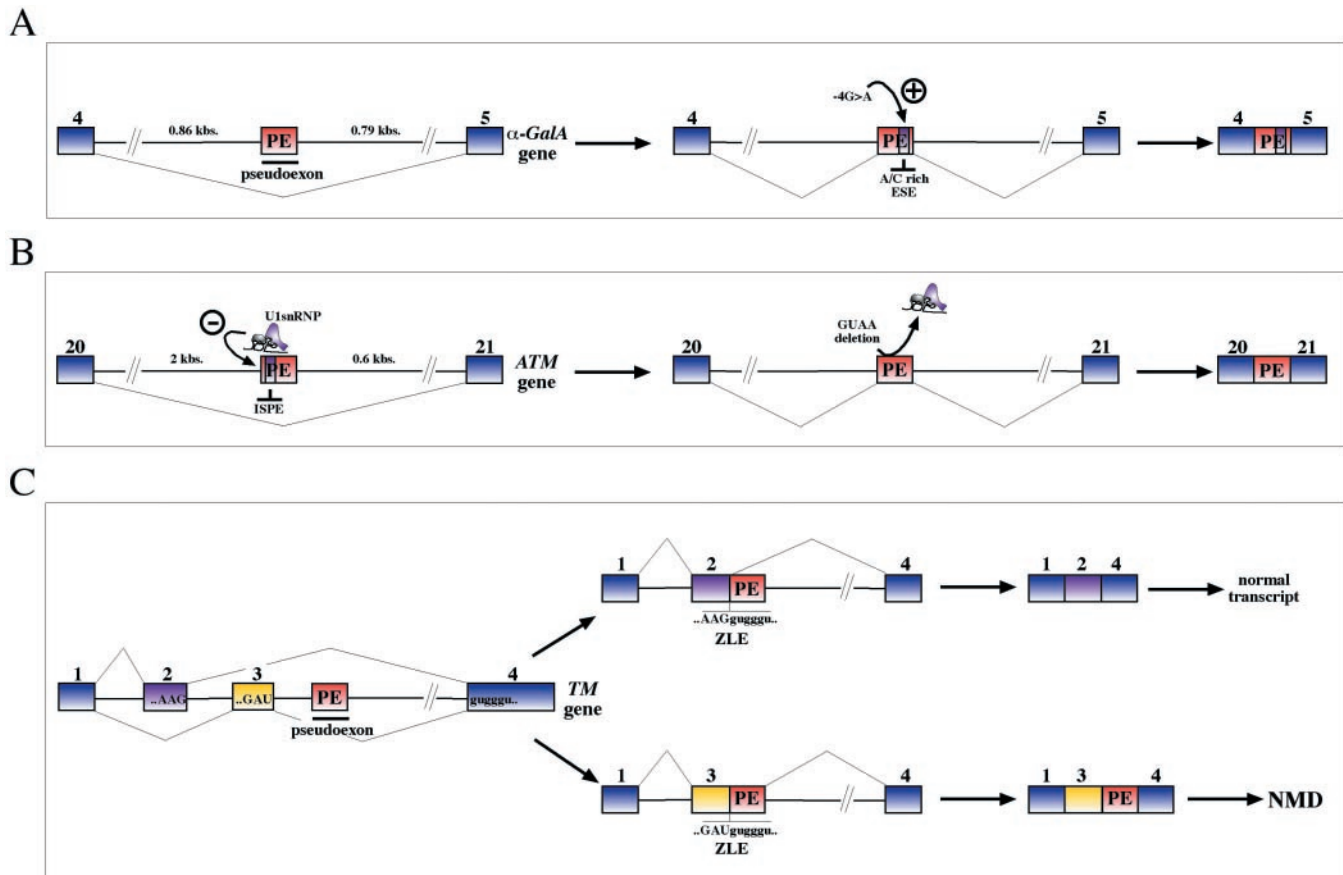
Figure 4. Shows several factors that determine pseudoexon inclusion aside the simple creation of *ex novo* functional splice sites. In the *a*-GalA gene (**A**), a single point mutation G > A observed to occur in position-4 with respect to a cryptic donor site has the effect of creating a novel A/C rich enhancer element that causes inclusion of a 57 bp. pseudoexon in the processed mRNA. In a second example, a 4 nt deletion (GUAA) in the intronic region of the *ATM* gene between exons 20 and 21 caused the insertion of a 65 nt long pseudoexon in the final mRNA (**B**). Functional analysis has demonstrated that this deletion abolished binding of an U1snRNP molecule in this position and activated a 3′ss lying 12 nt upstream of this element. Finally, (**C**) shows the effect of the inclusion of a pseudoexon sequence in the *a*-tropomyosin pre-mRNA with respect to two mutually exclusive exons 2 and 3. When the pseudoexon sequence (red box) is spliced to exon 2 a 5′ss ZLE sequence is regenerated, thus allowing splicing to occur again giving rise to a normal transcript bearing exon 2 joined to exon 4 (upper panel). On the other hand, when the pseudoexon is joined to exon 3 the ZLE sequence that is regenerated is insufficient to promote joining together of exon 3 to exon 4 and the resulting transcript is thus degraded by NMD (lower panel). This difference in resplicing activity has been proposed to be consistent with the match to consensus of the two ZLE elements (higher for exon 2 than exon 3) and for the higher exon spicing enhancer activity of exon 2.

GENOMIC CONTEXT IS IMPORTANT IN DEFINING EXONIC REGIONS

A practical example of the importance of genomic context on the splicing process is represented by the splicing regulation of the neurofibromatosis type-1 (*NF-1*) exon 37 (112). This exon is localized in an exon-rich genomic cluster (Figure 5A) and mutations within its sequence have been shown to be involved in the development of neurofibromatosis (113). It has been shown recently that this exon is badly recognized by the splicing machinery if it is placed with limited flanking intronic sequences in a minigene with heterologous genomic context. However, it becomes efficiently recognized when located within its natural milieu in the region spanning *NF-1* exons 34-38 (Figure 5B). Interestingly, mutations disrupting a SRE sequence within this exon (including some that have been described in neurofibromatosis patients such as C6792G) have genomic context-dependent effects (112). In fact, analysis of patient's lymphoblasts has revealed that this mutation causes both

the skipping of exons 36–37 together and the skipping of exon 37 alone. Also in this case, the use of a minigene system to mimic these effects has shown that, after taking in consideration the contribution of the unmutated allele in the patient's lymphoblasts, increasing the genomic context to comprise exons 31–38 allows to duplicate almost exactly the patient's *in vivo* splicing pattern (Figure 5C).

To complicate matters further, it has been observed that whether an exon is recognized as such and to what an extent is also decided by the overall architecture of the pre-mRNA. Two models exist in mRNA processing, the exon definition model and the intron definition model where, as their names suggest, in the former the exon is defined by the acceptor and donor sites that flank it whilst in the latter it is the acceptor and donor sites on the same intron which define its recognition (114,115). Recent experimental evidence has allowed to determine that intron recognition ceases to occur when the intron size increases beyond the 200–250 nt range (116). As splice site recognition across introns of similar
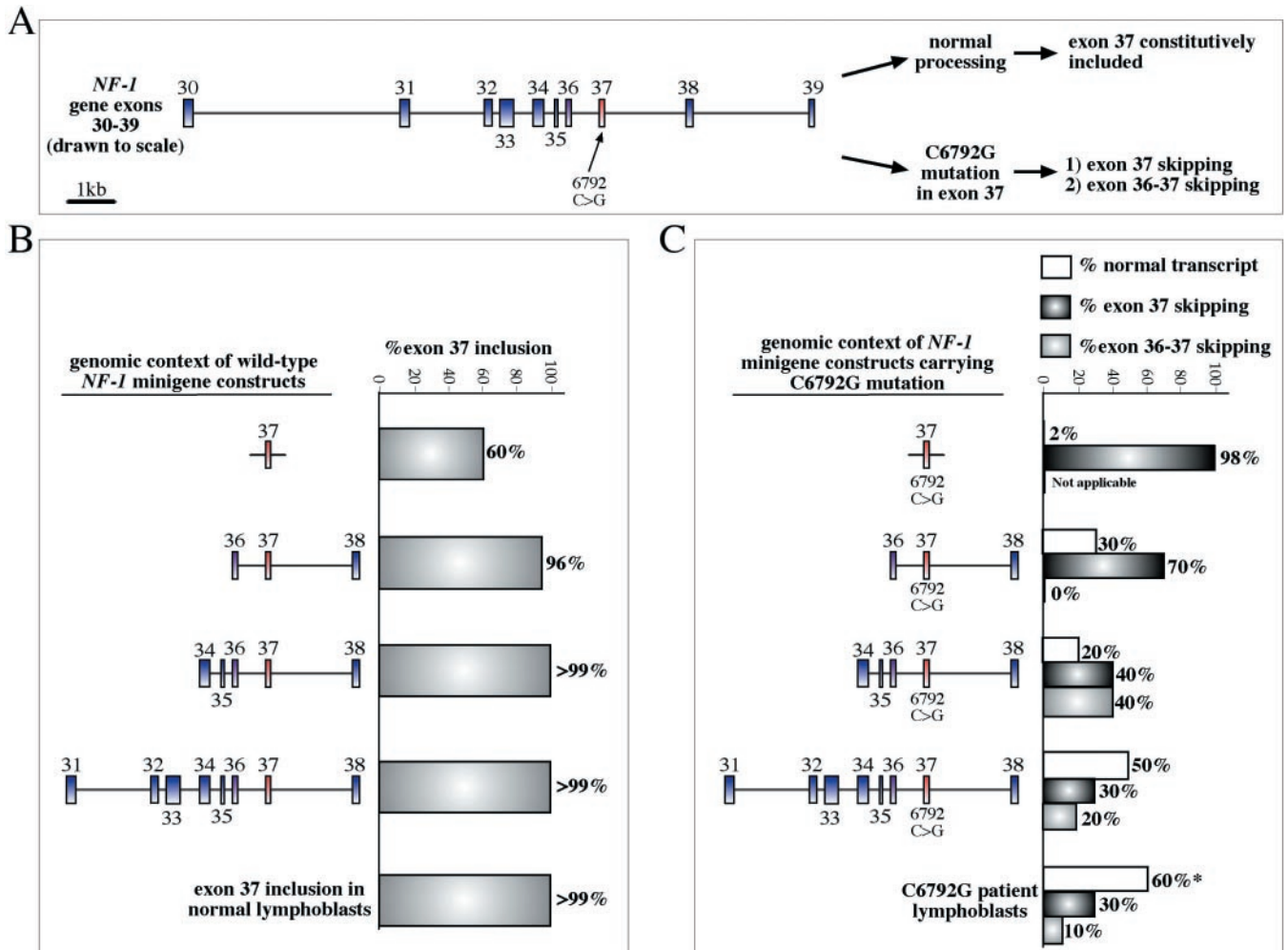
Figure 5. (**A**) Shows a schematic diagram (drawn to scale) of the exon-rich genomic region surrounding exon 37 of the neurofibromatosis (*NF-1*) gene. This exon is normally included constitutively in the mature mRNA but in the presence of a pathological substitution in position 6792 (a C is replaced by a G) the analysis from patient lymphoblasts reveals both the double skipping of exons 36 and 37 and also skipping of only exon 37. The influence of the genomic context has been tested by engineering a series of minigene constructs that gradually included larger portions of this genomic region, and their functionality compared with the *in vivo* pattern. (**B**) Wild-type exon 37 alone with a small portion of flanking intronic regions is not sufficient to yield the observed >99% inclusion in normal lymphoblasts. However, this result can be readily achieved when larger portions of this genomic region are included in the genomic construct. In addition to the normal situation, an increased genomic context also provides a closer correspondence with regards to the exon skipping effects of the C6792G substitution (**C**). In fact, only the construct carrying most of the *NF-1* exons 31 to 38 together with a mutated exon 37 yields a percentage of normal transcript, exon 37 skipping and exon 36–37 skipping that reflects the observed percentage in a patient's lymphoblasts that carry the C6792G substitution (the asterisk indicates the levels of normal mRNA from the C6792G-carrying allele after deducting the estimated contribution of the un mutated allele).

size to exons tends to be significantly more efficient than across the exons (116), it follows that exonic sequences with similar exon/intron junction sequences can be recognized with widely different efficiencies depending on the local pre-mRNA architecture where they happen to be present.

SREs ADD A LAYER OF REDUNDANCY FOR THE DEFINITION OF WEAK SPLICE SITES

The first observation that exonic sequences were modulating alternative splicing in a cellular gene was done in 1987 (117). At that time this observation appeared to most researchers as an exceptional case, difficult to interpret. However, in the last 10 years it has been shown that far from being

exceptional the presence of SREs in exonic sequences is the rule and their importance in the modulation of pre mRNA processing is now widely recognized.

These SREs reinforce the limited information in the splice sites and, depending on their position of and their effect on splicing, have been generally classified as exonic splicing enhancer and silencer (ESE/ESS) or intronic splicing enhancer and silencer (ISE/ISS) sequences (Figure 1). Although the number of reported proteins recognizing specific enhancer and silencer elements is steadily growing (118), most enhancer sequences have been found to bind members of the SR protein family (119–125) whilst many silencing elements have been found to bind members of the hnRNP protein family (55,126–128). There are, however, several difficulties when trying to evaluate the action and the presence of SREs in many splicing events.

Firstly, it should be noted that there are many exceptions to the rough division stated above. For example, classical SR proteins have also been described to be involved in splicing repression in a few number of cases (129–131) and some members of this protein family (SRp38) act as splicing repressors upon dephosphorylation (132,133). Analogously, some well-represented hnRNP proteins (hnRNP L) may also act as splicing enhancers (134) whilst others, such as hnRNP H, may be involved either in splicing repression or enhancement depending on context (46,135,136). Hence, the simple belonging to a particular family of splicing factors is not a guarantee of function conservation and experimental studies are often required to clarify the role played by even well known splicing factor found in new systems under study.

Secondly, the RNA-binding sequences of many of these splicing factors can be highly degenerate, making it difficult to spot them beforehand on the basis of the primary sequence alone. For example, SELEX analysis of SR protein-binding motifs showed that the major family members recognized fairly degenerate consensus sequences varying from 5 to 7 nt (137) and, although this finding is entirely consistent with the great diversity of ESE sequences mapped within exons (119,138,139), it obviously provides an additional hurdle when trying to identify likely binding sites. To these difficulties it has to be added that much of the same problems apply when investigating common splicing inhibitors, such as hnRNP A1 (140). Moreover, it has to be noted that local context factors, such as RNA secondary structure, may profoundly influence the way that optimal binding motifs are recognized by their specific binding factors (141). Therefore, the simple presence of a high-score-binding motif may not necessarily translate in a strong binding affinity.

In order to obviate to these difficulties several computational approaches have been proposed based on functional SELEX methodologies to identify potential SR protein-binding sites (142), on statistical analysis of genomic exonic sequences to identify splicing-enhancing sequences (67, 68,143,144), or on the information content of the splice site sequences (145). The reason why so much effort is devoted to the development of faster and more efficient ways of predicting SRE presence/action is justified by the clear indication that disruption/creation of these elements has become a very prominent cause of splicing related alterations in disease (4). Indeed, it is probable that many mutations which were previously thought to be harmless polymorphisms may in reality be very powerful modifiers of normal splicing processes (10,146–148).

At present, the accuracy of these methods to successfully predict changes in binding of splicing factors has yielded in some cases encouraging results (143,149,150). However, in several genes such as *dystrophin* (151), *CFTR* (147,148) and *ATR* (152) the predictive power of these *in silico* approaches has failed to establish a clear connection between splicing and abolishment (or creation) of SREs. Thus, experimental validation of predictions is still a mandatory step in most splicing studies. Nonetheless, because of the realization that accurate *in silico* approaches will be very useful in the diagnostic and therapeutic field (34,35), studies of this kind are destined to increase. Indeed, statistical approaches have also been aimed recently at characterizing inhibitory sequences (66,69,153) or other kinds of repeated motifs

such as UAGG or G-runs which are often found near or within alternatively spliced exons and which may contribute heavily to inclusion/skipping decisions by the splicing machinery (153,154).

Finally, the mechanisms through which splicing factors bind SRE sequences has been the subject of many studies, especially regarding the interplay between classic antagonistic factors such as ASF/SF2 (a SR protein family member that promotes inclusion) and hnRNP A1 (a classic hnRNP protein that has been historically associated with exon skipping) (155,156). In many instances, however, the number of 'duelling partners' increases considerably and a complex number of SRE sequences and binding factors is involved in regulating tissue- or developmentally specific alternatively spliced exons. A schematic diagram of four well-studied regulatory mechanisms is shown in Figure 6. The example given are the *c-src* exon N1 (157,158) (Figure 6A), the *CD44* exon v5 (159–161) (Figure 6B), chicken *cTNT* exon 5 (162) (Figure 6C) and *HipK3* 'T' exon (163) (Figure 6D). In all four cases it should be noted that in order to achieve tissue- and/or temporal-precise accuracy a complex arrangement of *cis*-acting elements (with either positive or negative influence on exon recognition) is complemented by a correspondingly high number of splicing factors.

SPLICE SITE STRENGTHS CAN EITHER MASK OR UNCOVER ADDITIONAL SPLICING REGULATORY REGIONS PRESENT IN EXONS

A well-studied working example in which the interplay between splice sites and regulatory sequences has been analyzed in depth is represented by exon 7 in the *SMN1/SMN2* genes. This exon plays a key role in a common autosomal recessive disorder, spinal muscular atrophy. In humans, there are two copies of the *SMN* gene, named *SMN1* and *SMN2*, and inactivation of the *SMN1* gene leads to disease because the *SMN2* gene cannot compensate for its absence. The reason for this is a critical C to U substitution in exon 7 of the *SMN2* gene that does not change the codon but either disrupts an ASF/SF2-binding site (149,164), creates an hnRNP A1-binding site (165), or both (although recent evidence supports the ASF/SF2 model whilst confirming an indirect role for hnRNP A1 in splicing repression). As a result of this substitution, the *SMN2* gene produces a transcript where exon 7 is predominantly skipped, leading to the production of a truncated protein and inability to compensate for *SMN1* inactivation. In addition to this region, several other SRE sequences both within the exon and in the upstream intron have been identified. These elements include a Tra2-binding site in the region adjacent to the SF2/ASF interaction (166) and a highly structured ISE region in the IVS7 intron a short distance away from the donor site (167) (Figure 7A). Recently, Singh *et al.* (168) have used a functional selection technique with the aim of identifying additional controlling elements that regulate the splicing of an entire exon. As shown in Figure 7B, they have achieved this purpose by engineering a pre-mRNA template which contained the full sequence of *SMN* exon 6 and 8 with some intronic sequences and a partially randomized (30%) exon 7 sequences in order to substantially preserve the
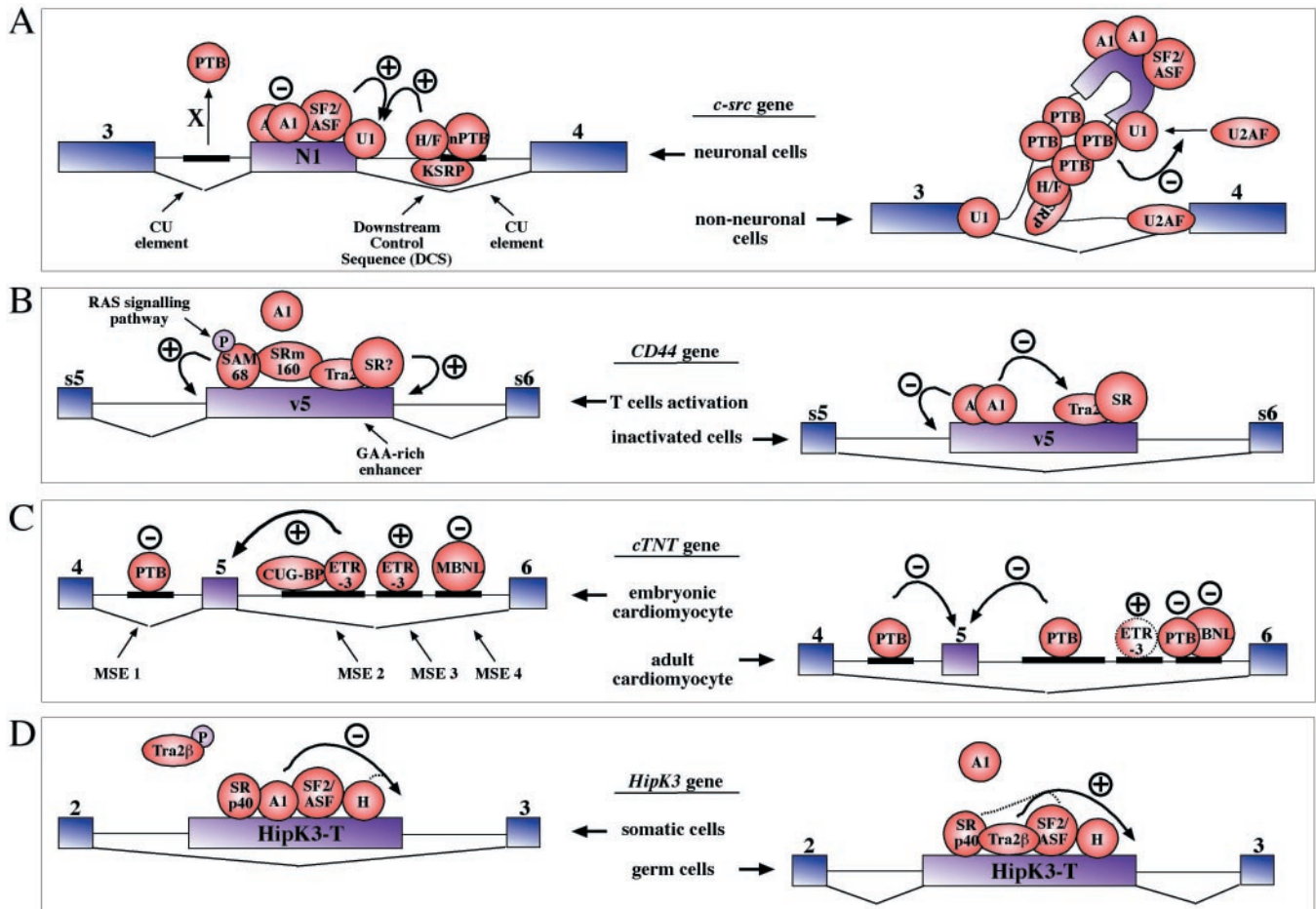
Figure 6. Modulation of tissue- and developmental-specific expression of four different exons. (**A**) Shows that in non-neuronal cells PTB binds to two CU elements which are located both upstream and downstream *c-src* exon N1 and form a silencing complex that interferes with the interaction between U1snRNP bound at the donor site of this exon and U2AF. This complex does not interfere with binding of U2AF with the U1snRNP present at the donor site of exon 3 and this allows splicing of exon 3 to exon 4, thereby excluding exon N1 from the mature mRNA. On the other hand, in neuronal cells PTB is removed from the CU elements by an unknown factor 'X' in the presence of ATP and exon inclusion is promoted by the SF2/ASF protein binding to this exon (which is also bound by its antagonistic molecule, hnRNP A1) and by the DCS complex (which contains hnRNP H/F (H/F), KSRP, plus a neuronal form of PTB known as nPTB or WERI PTB). (**B**) Shows the Ras signaling pathway-dependent regulation of splicing of the variable v5 exon in the *CD44* gene that occurs during T cell activation. In normal situations the enhancer complex in correspondence to a GAA-rich motif near the donor site of this exon is repressed by a strong hnRNP A1-binding sequence. During T cell activation, however, the repression can be overcome by either inhibition or displacement of A1 by phosphorylation of Sam 68 (Src-associated in mitosis) and involvement of powerful splicing coactivators such as SRm160. (**C**) Represents a model of *cTNT* exon 5 alternative splicing in the chicken developing heart. In the embryonic cardiomyocite CUG-BP and ETR-3 are expressed at high levels in the nucleus and they bind to the regulatory regions downstream of exon 5 (MSE2 to MSE4) to promote its inclusion. On the other hand, in the adult cardiomyocite CUG-BP and ETR-3 expression is downregulated and skipping is favored by PTB and MBNL proteins that are now free to bind on either side of the exon. Finally, (**D**) contains a model of *HipK3* 'T' (testis-specific) differential exon inclusion in both somatic cells and in germ lines. In somatic cells, hnRNP A1 (A1) is principally responsible for the inhibition of the T exon donor site, probably with some help by hnRNP H (H). This is favored by the fact that in these cells the Tra2β protein tends to be in low amounts and is mostly phosphorylated. On the other hand, in germ cells Tra2β expression is upregulated and tends to be in its hypo-phosphorylated form whilst hnRNP A1 is downregulated. This arrangement allows the Tra2β protein to bind in place of A1 and to promote T exon inclusion with the assistance of SF2/ASF and SRp40 which also bind within this exon.

wild-type characteristics of this exon but at the same time allow mutability of each nucleotide. This pool of pre-mRNA molecules was then transfected in a human cervical carinoma cell line and the spliced products were purified to select the exon 7 sequences that displayed correct recognition by the splicing machinery (Figure 7B). After successive rounds of *in vivo* selection the different exon 7 sequences were subsequently analyzed to evaluate the range and type of nucleotide substitutions that had occurred with respect to the initial wild-type sequence. Beside confirming previous

results with regards to the SF2/ASF and Tra2-binding regions the results of this analysis allowed the identification of two novel regulatory regions, one near the 3' end of the exon (3'-Cluster) and a conserved CT tract in the middle. Most importantly, however, the weak 5'ss consensus sequence was also found to represent a critical factor in regulating *SMN* exon 7 recognition, as 82% of the spliced exon 7 clones identified following the functional SELEX contained a A54G substitution (Figure 7B). It is interesting to note that this substitution alone was quite capable of overcoming the negative
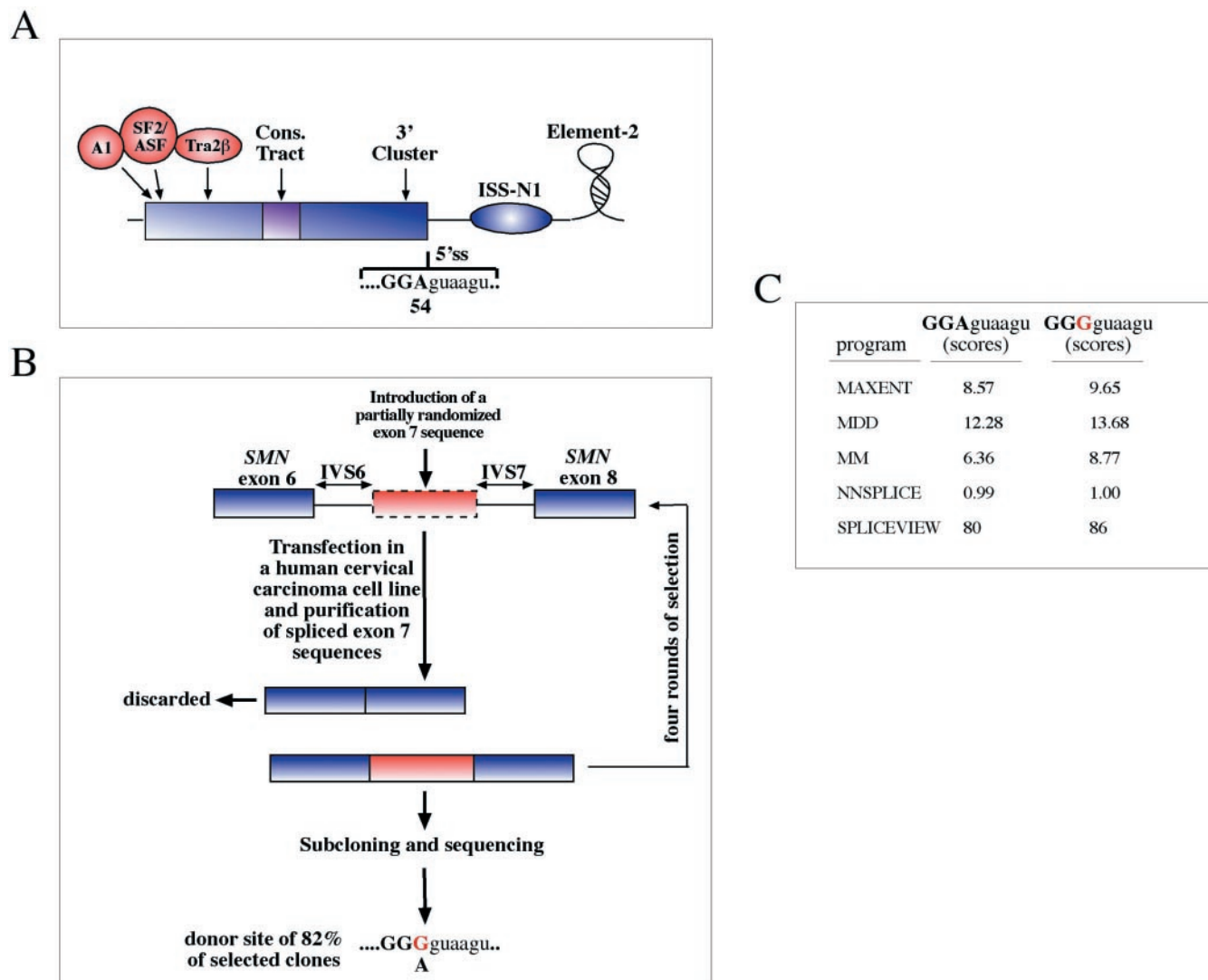
Figure 7. (**A**) Shows a schematic diagram of the splicing regulatory regions that have been described to affect recognition of *SMN1/SMN*2 exon 7. Recently, a functional SELEX analysis has been applied to this whole exonic sequence. (**B**) Shows the minigene system based on the human *SMN* gene used for this assay in which the entire exon 7 is partially randomized. These sequences can then be selected for inclusion in the final spliced products by subjecting them to rounds of selection *in vivo*. The borders of this randomized exon are dotted in (B) to indicate the fact that these sequences will not always be used as an exon in this technique. After each round of selection the spliced exon 7 sequences are recovered by RT–PCR and reinserted in the minigene splicing cassette. This selection process has allowed the precise mapping, at the single nucleotide level, of all the major determinants involved in the control of exon 7 splicing. Interestingly, 82% of all the final clones contained the same substitution in correspondence to the last position of this exon, and this change was quite capable of overcoming the effects of other negative regulatory mutations. (**C**) Shows a comparison of the two donor site strengths (the wild type and the one carrying the A54G substitution) using a panel of splice site prediction programs.

effect of several mutations within other conserved elements such as the first position of the exon or single mutations within the Tra2-binding site (168).

It is also interesting to note that several splice site prediction programs such as MaxEnt, MDD and MM (169), NNSplice (170) and Spliceview (171) fail to detect a major improvement in donor site score following the A54G substitution (Figure 7C). As these programs are normally very accurate in detecting splice site strength variations the most likely conclusion is that only the slightly weakening of exon 7 splice site can 'highlight' several sequences that would normally not be able to influence its splicing process. This is consistent with the observation that alternatively spliced exons possess splice site sequences which are on average weaker than splice sites belonging to constitutively recognized exons and stresses the importance of the context (172). In addition, it has also to be noted that splice site prediction programs may not pick up additional protein–RNA interactions at the 5′ splice site that may substantially contribute in determining its relative strength. For example, in the *NF-1* exon 3 donor site a G-four stretch in correspondence to the donor site has been identified as a binding site of hnRNP H that severely constrained the 5′ss sequences that could vary without disrupting splicing (46).

Nonetheless, the result of SELEX analysis suggests that many skipping events originating from various mutations within the exon may be rescued by even small improvements in splice site quality of the affected exon(s). However,

changing the donor site in exons of patients affected by mutations within SRE elements would probably involve some too costly and uncertain investments in gene therapy schemes and, for this reason, alternative targets should be preferred.

INACTIVATION OF STRONG NEGATIVE ELEMENTS CAN RESCUE SPLICING DEFECTS ORIGINATING FROM OTHER NEARBY REGULATORY ELEMENTS

Therefore, in order to optimize therapeutic strategies an important question that needs to be addressed is the existence of hierarchical levels between these splicing regulatory elements, as this may decide which region should be preferentially targeted by the effector molecule of choice. Recently, the presence of a human-specific ISS near to the 5′ss of

SMN1/SMN2 exon 7 has been described to have analogous compensatory effects as the A54G substitution (173). Indeed, Figure 8A shows that the effects of mutating three well-characterized SRE elements in *SMN1* exon 7 (the first nucleotide, the Tra2 binding sequence, and the conserved tract) can all be overcome by deleting the ISS-N1 sequence (Figure 8B). In keeping with this result, much the same effect can be obtained by the deletion of the ISS-N1 region in the *SMN2* sequence that carries the C6U substitution that abrogates SF2/ASF binding.

Interestingly, an analogous situation has been described for exon 9 of the *CFTR* gene. Skipping of this exon is involved in the occurrence of several forms of cystic fibrosis depending on the presence of particular polymorphisms in a (UG)m(U)n repeated motif near to its 3′ splice site (174). Recently, several studies have confirmed that a nuclear
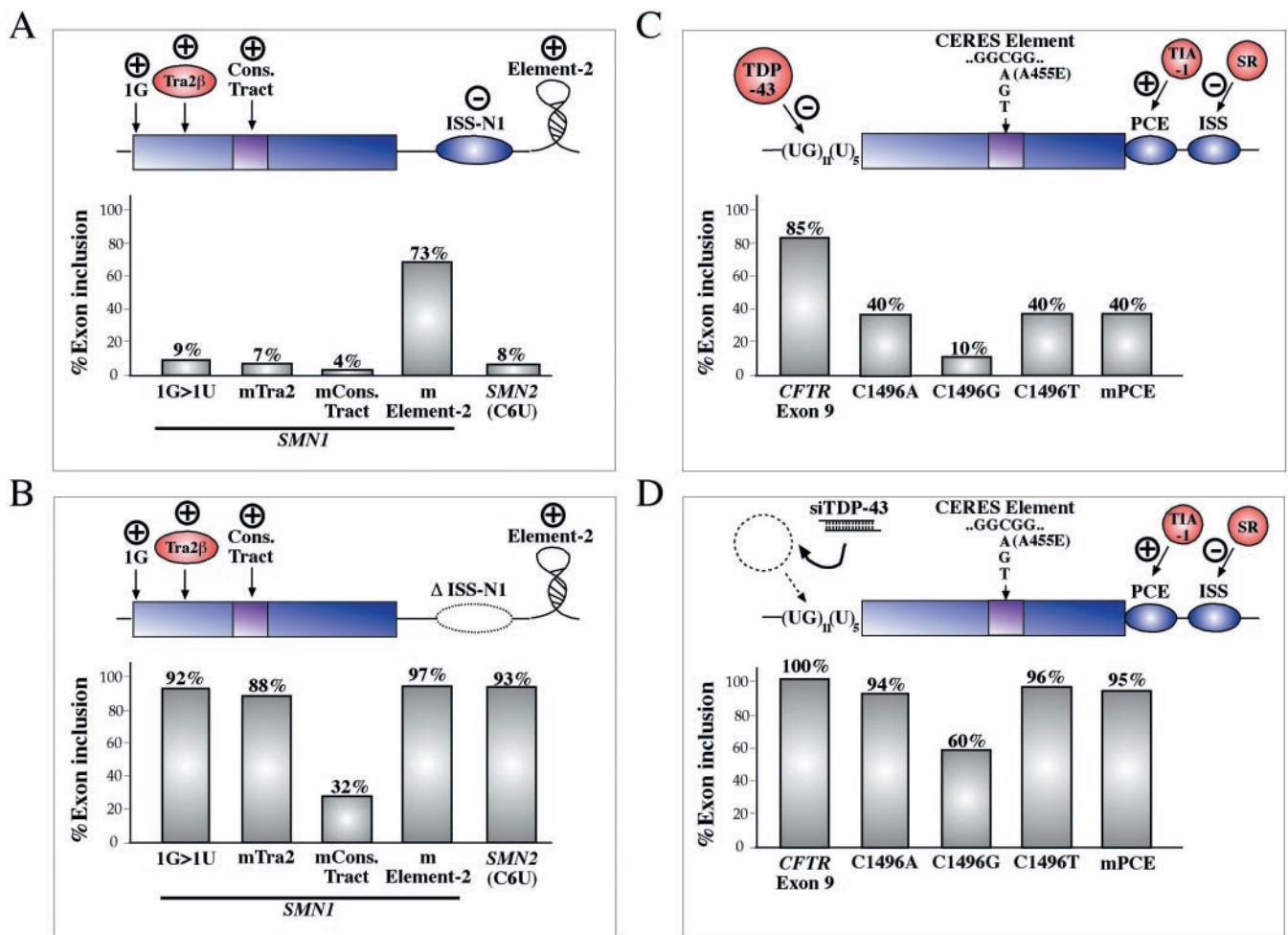

Figure 8. (**A**) Shows a diagrammatic representation of several positive elements that regulate *SMN1/SMN*2 exon 7 inclusion. The elements analyzed in this experiment are the first nucleotide of exon 7 (1G), the Tra2β-binding site (Tra2), the conserved tract in the middle of the exon (ConsTract) and the intronic splicing enhancer (element-2) in IVS7. Also shown is the dominant ISS element (ISS-N1). The graphical representation underneath shows the levels of exon 7 inclusion following mutation of all of these elements in a suitable minigene construct. Also included are the basal exon 7 inclusion levels from *SMN2* exon 7 that carries a natural C6U substitution. (**B**) Shows the effect of the same mutations/substitutions when the ISS-N1 sequence has been deleted from the minigene constructs. An analogous situation which regards the *CFTR* exon 9 sequence is schematically represented in (**C**) in which the (UG)m(U)n repeated motif near the 3′ss of this exon has been shown to bind a powerful negative regulator of splicing, TDP-43. Several other *cis*-acting elements are also shown in this diagram: a CERES element within the exon, a positive polypyrimidine sequence near the 5′ (PCE) that binds TIA-1, and a human-specific ISS in IVS9 that binds members of the SR protein family. The graph below shows that mutations within these elements are strongly inhibitory with respect to the basal inclusion level on a TG11T5 background (lane *CFTR* exon 9). Nonetheless, also in this case the removal of TDP-43 using siRNA technology rescues the effect of all these negative mutations (**D**).

protein, TDP-43, binds to the (UG)m sequence and can inhibit its inclusion in the final mRNA (175–177). However, in addition to this regulatory sequence several other exonic and intronic SREs have been described to take part in splicing regulation of exon 9: a composite exonic regulatory element of splicing (CERES) (147), a positive polypyrimidine rich element near to the 5′ss that binds the TIA-1 splicing factor (PCE) (178), and an ISS further down in IVS9 that binds several members of the SR protein family (129) (Figure 8C). Mutations in all these sequences have modulating effects in *CFTR* exon inclusion in the mRNA. In particular, the CERES element identified by Pagani *et al.* (129) lies in correspondence to a previously described C1496A missense mutation (A455E) which should be therefore also considered a splicing mutation. The recent use of siRNA technology to knockdown TDP-43 expression (179) has resulted in a generalized improvement in the inclusion of exon 9 even in the presence of many of these mutations (Figure 8D), demonstrating that removal of this key inhibitory factor may be a useful way of repairing negative mutations also in other controlling elements. In this respect, it is important to note that depletion of this factor has also been described to override the need for several ESEs and ISEs of *Apo AII* exon 3 (180). Obviously, knowledge of other critical functions of this or any other factor is essential before a general inhibitory approach is used in therapy.

CONCLUDING REMARKS

Taken together, recent results clearly highlight that any pre-mRNA sequence is embedded with regulatory elements, some of which may only be uncovered by disrupting mutations or by other facets of mRNA synthesis (such as transcript elongation rates or secondary structure). The practical consequences of such an arrangement is that many splicing events are the result of a fine 'balance of power' between a myriad of controlling factors which can vary from tissue to tissue, during development or in the presence of external stimuli. Moreover, the recent discovery that snoRNAs can also participate actively in the regulation of splicing (181) emphasizes that we are still ignorant of entire classes of splicing regulators. This complexity makes the job of predicting the splicing outcomes of given genomic variations exceedingly difficult for the researcher. Nonetheless, not all these factors possess the same weight when determining splicing outcomes, and the *SMN1/SMN2* exon 7 and *CFTR* exon 9 examples demonstrate that identifying master checkpoints may, if acted upon, compensate for negative effects in several other regulatory elements. This conclusion provides a considerable encouragement towards the search for novel therapeutic strategies aimed at covering as great an amount as possible of splicing mutations. Paradoxically, though, it is very probable that effective cures for many splicing mutations will be developed long before the functional mechanisms that underlie them may be satisfactorily explained.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Nissim-Rafinia,M. and Kerem,B. (2005) The splicing machinery is a genetic modifier of disease severity. *Trends Genet.*, **21**, 480–483.
2. Caceres,J.F. and Kornblihtt,A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.*, **18**, 186–193.
3. Garcia-Blanco,M.A., Baraniak,A.P. and Lasda,E.L. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.*, **22**, 535–546.
4. Cartegni,L., Chew,S.L. and Krainer,A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet.*, **3**, 285–298.
5. Faustino,N.A. and Cooper,T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.*, **17**, 419–437.
6. Tazi,J., Durand,S. and Jeanteur,P. (2005) The spliceosome: a novel multi-faceted target for therapy. *Trends Biochem. Sci.*, **30**, 469–478.
7. Stoilov,P., Meshorer,E., Gencheva,M., Glick,D., Soreq,H. and Stamm,S. (2002) Defects in pre-mRNA processing as causes of and predisposition to diseases. *DNA Cell. Biol.*, **21**, 803–818.
8. Ars,E., Serra,E., Garcia,J., Kruyer,H., Gaona,A., Lazaro,C. and Estivill,X. (2000) Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet.*, **9**, 237–247.
9. 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.
10. Chamary,J.V., Parmley,J.L. and Hurst,L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Rev. Genet.*, **7**, 98–108.
11. Lenasi,T., Peterlin,B.M. and Dovc,P. (2006) Distal regulation of alternative splicing by splicing enhancer in equine beta-casein intron 1. *RNA*, **12**, 498–507.
12. Graveley,B.R. (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell*, **123**, 65–73.
13. Kreahling,J.M. and Graveley,B.R. (2005) The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam pre-mRNA. *Mol. Cell. Biol.*, **25**, 10251–10260.
14. Buratti,E. and Baralle,F.E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.*, **24**, 10505–10514.
15. Romano,M., Marcucci,R. and Baralle,F.E. (2001) Splicing of constitutive upstream introns is essential for the recognition of intra-exonic suboptimal splice sites in the thrombopoietin gene. *Nucleic Acids Res.*, **29**, 886–894.
16. Fededa,J.P., Petrillo,E., Gelfand,M.S., Neverov,A.D., Kadener,S., Nogues,G., Pelisch,F., Baralle,F.E., Muro,A.F. and Kornblihtt,A.R. (2005) A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol. Cell*, **19**, 393–404.
17. Kornblihtt,A.R. (2005) Promoter usage and alternative splicing. *Curr. Opin. Cell. Biol.*, **17**, 262–268.
18. Pagani,F. and Baralle,F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Rev. Genet.*, **5**, 389–396.
19. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
20. Modrek,B. and Lee,C. (2002) A genomic view of alternative splicing. *Nature Genet.*, **30**, 13–19.
21. Maniatis,T. and Tasic,B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, **418**, 236–243.
22. Stetefeld,J. and Ruegg,M.A. (2005) Structural and functional diversity generated by alternative mRNA splicing. *Trends Biochem. Sci.*, **30**, 515–521.

23. Graveley,B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.*, **17**, 100–107.

24. Stamm,S., Ben-Ari,S., Rafalska,I., Tang,Y., Zhang,Z., Toiber,D., Thanaraj,T.A. and Soreq,H. (2005) Function of alternative splicing. *Gene*, **344**, 1–20.

25. Vibe-Pedersen,K., Kornblihtt,A.R. and Baralle,F.E. (1984) Expression of a human alpha-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing. *EMBO J.*, **3**, 2511–2516.

26. Baralle,D. and Baralle,M. (2005) Splicing in action: assessing disease causing sequence changes. *J. Med. Genet.*, **42**, 737–748.

27. Cooper,T.A. (2005) Use of minigene systems to dissect alternative splicing elements. *Methods*, **37**, 331–340.

28. Buratti,E. and Baralle,F.E. (2005) Another step forward for SELEXive splicing. *Trends Mol. Med.*, **11**, 5–9.

29. Fu,X.D. (2004) Towards a splicing code. *Cell*, **119**, 736–738.

30. Buratti,E., Baralle,F.E. and Pagani,F. (2003) Can a 'patch' in a skipped exon make the pre-mRNA splicing machine run better? *Trends Mol. Med.*, **9**, 229–232.

31. Garcia-Blanco,M.A. (2005) Making antisense of splicing. *Curr. Opin. Mol. Ther.*, **7**, 476–482.

32. Wilton,S.D. and Fletcher,S. (2005) RNA splicing manipulation: strategies to modify gene expression for a variety of therapeutic outcomes. *Curr. Gene Ther.*, **5**, 467–483.

33. Sazani,P. and Kole,R. (2003) Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J. Clin. Invest.*, **112**, 481–486.

34. Cartegni,L. and Krainer,A.R. (2003) Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nature Struct. Biol.*, **10**, 120–125.

35. Skordis,L.A., Dunckley,M.G., Yue,B., Eperon,I.C. and Muntoni,F. (2003) Bifunctional antisense oligonucleotides provide a *trans*-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc. Natl Acad. Sci. USA*, **100**, 4114–4119.

36. Soret,J., Bakkour,N., Maire,S., Durand,S., Zekri,L., Gabut,M., Fic,W., Divita,G., Rivalle,C., Dauzonne,D. *et al.* (2005) Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors. *Proc. Natl Acad. Sci. USA*, **102**, 8764–8769.

37. Hagiwara,M. (2005) Alternative splicing: a new drug target of the post-genome era. *Biochim. Biophys. Acta*, **1754**, 324–331.

38. Graveley,B.R. (2005) Small molecule control of pre-mRNA splicing. *RNA*, **11**, 355–358.

39. Goyenvalle,A., Vulin,A., Fougerousse,F., Leturcq,F., Kaplan,J.C., Garcia,L. and Danos,O. (2004) Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science*, **306**, 1796–1799.

40. Maquat,L.E. (2001) The power of point mutations. *Nature Genet.*, **27**, 5–6.

41. Zhang,M.Q. (1998) Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.*, **7**, 919–932.

42. Lund,M. and Kjems,J. (2002) Defining a 5′ splice site by functional selection in the presence and absence of U1 snRNA 5′ end. *RNA*, **8**, 166–179.

43. Du,H. and Rosbash,M. (2002) The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. *Nature*, **419**, 86–90.

44. Tarn,W.Y. and Steitz,J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes. Dev.*, **8**, 2704–2717.

45. Crispino,J.D., Blencowe,B.J. and Sharp,P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**, 1866–1869.

46. Buratti,E., Baralle,M., De Conti,L., Baralle,D., Romano,M., Ayala,Y.M. and Baralle,F.E. (2004) hnRNP H binding at the 5′ splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res.*, **32**, 4224–4236.

47. Relogio,A., Ben-Dov,C., Baum,M., Ruggiu,M., Gemund,C., Benes,V., Darnell,R.B. and Valcarcel,J. (2005) Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J. Biol. Chem.*, **280**, 4779–4784.

48. Pan,Q., Shai,O., Misquitta,C., Zhang,W., Saltzman,A.L., Mohammad,N., Babak,T., Siu,H., Hughes,T.R., Morris,Q.D. *et al.* (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell*, **16**, 929–941.

49. Yeo,G., Holste,D., Kreiman,G. and Burge,C.B. (2004) Variation in alternative splicing across human tissues. *Genome Biol.*, **5**, R74.

50. Kalnina,Z., Zayakin,P., Silina,K. and Line,A. (2005) Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes Cancer.*, **42**, 342–357.

51. Venables,J.P. (2006) Unbalanced alternative splicing and its significance in cancer. *Bioessays*, **28**, 378–386.

52. Smith,C.W. and Valcarcel,J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.*, **25**, 381–388.

53. Matlin,A.J., Clark,F. and Smith,C.W. (2005) Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell. Biol.*, **6**, 386–398.

54. Maniatis,T. and Reed,R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499–506.

55. Singh,R. and Valcarcel,J. (2005) Building specificity with nonspecific RNA-binding proteins. *Nature Struct. Mol. Biol.*, **12**, 645–653.

56. Shin,C. and Manley,J.L. (2004) Cell signalling and the control of pre-mRNA splicing. *Nature Rev. Mol. Cell. Biol.*, **5**, 727–738.

57. Li,X. and Manley,J.L. (2005) New talents for an old acquaintance: the SR protein splicing factor ASF/SF2 functions in the maintenance of genome stability. *Cell Cycle*, **4**, 1706–1708.

58. Nott,A., Le Hir,H. and Moore,M.J. (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.*, **18**, 210–222.

59. Sanford,J.R., Gray,N.K., Beckmann,K. and Caceres,J.F. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.*, **18**, 755–768.

60. Huang,Y., Yario,T.A. and Steitz,J.A. (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proc. Natl Acad. Sci. USA*, **101**, 9666–9670.

61. Huang,Y., Gattoni,R., Stevenin,J. and Steitz,J.A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell*, **11**, 837–843.

62. Graveley,B.R. (2005) Coordinated control of splicing and translation. *Nature Struct. Mol. Biol.*, **12**, 1022–1023.

63. Blaustein,M., Pelisch,F., Tanos,T., Munoz,M.J., Wengier,D., Quadrana,L., Sanford,J.R., Muschietti,J.P., Kornblihtt,A.R., Caceres,J.F. *et al.* (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nature Struct. Mol. Biol.*, **12**, 1037–1044.

64. Sun,H. and Chasin,L.A. (2000) Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.*, **20**, 6414–6425.

65. Fairbrother,W.G. and Chasin,L.A. (2000) Human genomic sequences that inhibit splicing. *Mol. Cell. Biol.*, **20**, 6816–6825.

66. Sironi,M., Menozzi,G., Riva,L., Cagliani,R., Comi,G.P., Bresolin,N., Giorda,R. and Pozzoli,U. (2004) Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res.*, **32**, 1783–1791.

67. Zhang,X.H. and Chasin,L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.*, **18**, 1241–1250.

68. Zhang,X.H., Kangsamaksin,T., Chao,M.S., Banerjee,J.K. and Chasin,L.A. (2005) Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol. Cell. Biol.*, **25**, 7323–7332.

69. Wang,Z., Rolish,M.E., Yeo,G., Tung,V., Mawson,M. and Burge,C.B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell*, **119**, 831–845.

70. Mendes Soares,L.M. and Valcarcel,J. (2006) The expanding transcriptome: the genome as the 'Book of Sand'. *EMBO J.*, **25**, 923–931.

71. Proudfoot,N.J., Furger,A. and Dye,M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.

72. Dye,M.J., Gromak,N. and Proudfoot,N.J. (2006) Exon tethering in transcription by RNA polymerase II. *Mol. Cell*, **21**, 849–859.

73. Horowitz,D.S. and Krainer,A.R. (1994) Mechanisms for selecting 5′ splice sites in mammalian pre-mRNA splicing. *Trends Genet.*, **10**, 100–106.

74. Freund,M., Hicks,M.J., Konermann,C., Otte,M., Hertel,K.J. and Schaal,H. (2005) Extended base pair complementarity between U1 snRNA and the 5′ splice site does not inhibit splicing in higher eukaryotes, but rather increases 5′ splice site recognition. *Nucleic Acids Res.*, **33**, 5112–5119.

75. Roca,X., Sachidanandam,R. and Krainer,A.R. (2005) Determinants of the inherent strength of human 5′ splice sites. *RNA*, **11**, 683–698.

76. Carmel,I., Tal,S., Vig,I. and Ast,G. (2004) Comparative analysis detects dependencies among the 5′ splice-site positions. *RNA*, **10**, 828–840.

77. Krawczak,M., Reiss,J. and Cooper,D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.*, **90**, 41–54.

78. Treisman,R., Orkin,S.H. and Maniatis,T. (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature*, **302**, 591–596.

79. Wieringa,B., Meyer,F., Reiser,J. and Weissmann,C. (1983) Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit beta-globin gene. *Nature*, **301**, 38–43.

80. Roy,S.W. and Gilbert,W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Rev. Genet.*, **7**, 211–221.

81. Sadusky,T., Newman,A.J. and Dibb,N.J. (2004) Exon junction sequences as cryptic splice sites: implications for intron origin. *Curr. Biol.*, **14**, 505–509.

82. Stoltzfus,A. (2004) Molecular evolution: introns fall into place. *Curr. Biol.*, **14**, R351–352.

83. Roca,X., Sachidanandam,R. and Krainer,A.R. (2003) Intrinsic differences between authentic and cryptic 5′ splice sites. *Nucleic Acids Res.*, **31**, 6321–6333.

84. Kralovicova,J., Christensen,M.B. and Vorechovsky,I. (2005) Biased exon/intron distribution of cryptic and *de novo* 3′ splice sites. *Nucleic Acids Res.*, **33**, 4882–4898.

85. Krainer,A.R., Conway,G.C. and Kozak,D. (1990) The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Cell*, **62**, 35–42.

86. Caceres,J.F., Stamm,S., Helfman,D.M. and Krainer,A.R. (1994) Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science*, **265**, 1706–1709.

87. Gabut,M., Mine,M., Marsac,C., Brivet,M., Tazi,J. and Soret,J. (2005) The SR protein SC35 is responsible for aberrant splicing of the E1alpha pyruvate dehydrogenase mRNA in a case of mental retardation with lactic acidosis. *Mol. Cell. Biol.*, **25**, 3286–3294.

88. Spena,S., Tenchini,M.L. and Buratti,E. (2006) Cryptic splice site usage in exon 7 of the human fibrinogen Bβ-chain gene is regulated by a naturally silent SF2/ASF binding site within this exon. *RNA*, **12**, 948–958.

89. Smith,C.W. (2005) Alternative splicing—when two's a crowd. *Cell*, **123**, 1–3.

90. Domenjoud,L., Gallinaro,H., Kister,L., Meyer,S. and Jacob,M. (1991) Identification of a specific exon sequence that is a major determinant in the selection between a natural and a cryptic 5′ splice site. *Mol. Cell. Biol.*, **11**, 4581–4590.

91. Domenjoud,L., Kister,L., Gallinaro,H. and Jacob,M. (1993) Selection between a natural and a cryptic 5′ splice site: a kinetic study of the effect of upstream exon sequences. *Gene Expr.*, **3**, 83–94.

92. Galante,P.A., Sakabe,N.J., Kirschbaum-Slager,N. and de Souza,S.J. (2004) Detection and evaluation of intron retention events in the human transcriptome. *RNA*, **10**, 757–765.

93. Stamm,S., Zhu,J., Nakai,K., Stoilov,P., Stoss,O. and Zhang,M.Q. (2000) An alternative-exon database and its statistical analysis. *DNA Cell Biol.*, **19**, 739–756.

94. Laski,F.A. and Rubin,G.M. (1989) Analysis of the cis-acting requirements for germ-line-specific splicing of the P-element ORF2-ORF3 intron. *Genes Dev.*, **3**, 720–728.

95. Gebauer,F., Merendino,L., Hentze,M.W. and Valcarcel,J. (1998) The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *RNA*, **4**, 142–150.

96. Mansilla,A., Lopez-Sanchez,C., de la Rosa,E.J., Garcia-Martinez,V., Martinez-Salas,E., de Pablo,F. and Hernandez-Sanchez,C. (2005) Developmental regulation of a proinsulin messenger RNA generated by intron retention. *EMBO Rep.*, **6**, 1182–1187.

97. Le Hir,H., Charlet-Berguerand,N., de Franciscis,V. and Thermes,C. (2002) 5′-End RET splicing: absence of variants in normal tissues and intron retention in pheochromocytomas. *Oncology*, **63**, 84–91.

98. Zhang,L., Vincent,G.M., Baralle,M., Baralle,F.E., Anson,B.D., Benson,D.W., Whiting,B., Timothy,K.W., Carlquist,J., January,C.T. *et al.* (2004) An intronic mutation causes long QT syndrome. *J. Am. Coll. Cardiol.*, **44**, 1283–1291.

99. Pequignot,M.O., Dey,R., Zeviani,M., Tiranti,V., Godinot,C., Poyau,A., Sue,C., Di Mauro,S., Abitbol,M. and Marsac,C. (2001) Mutations in the SURF1 gene associated with Leigh syndrome and cytochrome C oxidase deficiency. *Hum. Mutat.*, **17**, 374–381.

100. Muro,A.F., Chauhan,A.K., Gajovic,S., Iaconcig,A., Porro,F., Stanta,G. and Baralle,F.E. (2003) Regulated splicing of the fibronectin EDA exon is essential for proper skin wound healing and normal lifespan. *J. Cell. Biol.*, **162**, 149–160.

101. Batsche,E., Yaniv,M. and Muchardt,C. (2006) The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nature Struct. Mol. Biol.*, **13**, 22–29.

102. Kornblihtt,A.R. (2006) Chromatin, transcript elongation and alternative splicing. *Nature Struct. Mol. Biol.*, **13**, 5–7.

103. Kreahling,J. and Graveley,B.R. (2004) The origins and implications of Aluternative splicing. *Trends Genet.*, **20**, 1–4.

104. Sorek,R., Lev-Maor,G., Reznik,M., Dagan,T., Belinky,F., Graur,D. and Ast,G. (2004) Minimal conditions for exonization of intronic sequences: 5′ splice site formation in alu exons. *Mol. Cell*, **14**, 221–231.

105. Lei,H., Day,I.N. and Vorechovsky,I. (2005) Exonization of AluYa5 in the human ACE gene requires mutations in both 3′ and 5′ splice sites and is facilitated by a conserved splicing enhancer. *Nucleic Acids Res.*, **33**, 3897–3906.

106. Zhang,X.H., Leslie,C.S. and Chasin,L.A. (2005) Dichotomous splicing signals in exon flanks. *Genome. Res.*, **15**, 768–779.

107. Ishii,S., Nakao,S., Minamikawa-Tachino,R., Desnick,R.J. and Fan,J.Q. (2002) Alternative splicing in the alpha-galactosidase A gene: increased exon inclusion results in the Fabry cardiac phenotype. *Am. J. Hum. Genet.*, **70**, 994–1002.

108. Pagani,F., Buratti,E., Stuani,C., Bendix,R., Dork,T. and Baralle,F.E. (2002) A new type of mutation causes a splicing defect in ATM. *Nature Genet.*, **30**, 426–429.

109. Lewandowska,M.A., Stuani,C., Parvizpur,A., Baralle,F.E. and Pagani,F. (2005) Functional studies on the ATM intronic splicing processing element. *Nucleic Acids Res.*, **33**, 4007–4015.

110. Grellscheid,S.N. and Smith,C.W. (2006) An apparent pseudo-exon acts both as an alternative exon that leads to nonsense-mediated decay and as a zero-length exon. *Mol. Cell. Biol.*, **26**, 2237–2246.

111. Maquat,L.E. (2005) Nonsense-mediated mRNA decay in mammals. *J. Cell. Sci.*, **118**, 1773–1776.

112. Baralle,M., Skoko,N., Knezevich,A., De Conti,L., Motti,D., Baralle,D., Buratti,E. and Baralle,F.E. (2006) NF1 mRNA biogenesis: effects of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS Lett.*, doi:10.1016/j.febslet.2006.07.018.

113. Vandenbroucke,I., Callens,T., De Paepe,A. and Messiaen,L. (2002) Complex splicing pattern generates great diversity in human NF1 transcripts. *BMC Genomics*, **3**, 13.

114. Berget,S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.

115. Ast,G. (2004) How did alternative splicing evolve? *Nature Rev. Genet.*, **5**, 773–782.

116. Fox-Walsh,K.L., Dou,Y., Lam,B.J., Hung,S.P., Baldi,P.F. and Hertel,K.J. (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl Acad. Sci. USA*, **102**, 16176–16181.

117. Mardon,H.J., Sebastio,G. and Baralle,F.E. (1987) A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res.*, **15**, 7725–7733.

118. Jurica,M.S. and Moore,M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.

119. Bourgeois,C.F., Lejeune,F. and Stevenin,J. (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **78**, 37–88.

120. Graveley,B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197–1211.

121. Blencowe,B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.*, **25**, 106–110.

122. Fu,X.D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.

123. Sanford,J.R., Ellis,J. and Caceres,J.F. (2005) Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem. Soc. Trans.*, **33**, 443–446.

124. Zhang,Z. and Krainer,A.R. (2004) Involvement of SR proteins in mRNA surveillance. *Mol. Cell*, **16**, 597–607.

125. Huang,Y. and Steitz,J.A. (2005) SRprises along a messenger's journey. *Mol. Cell*, **17**, 613–615.

126. Krecic,A.M. and Swanson,M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell. Biol.*, **11**, 363–371.

127. Dreyfuss,G., Kim,V.N. and Kataoka,N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell. Biol.*, **3**, 195–205.

128. Spellman,R. and Smith,C.W. (2006) Novel modes of splicing repression by PTB. *Trends. Biochem. Sci.*, **31**, 73–76.

129. Pagani,F., Buratti,E., Stuani,C., Romano,M., Zuccato,E., Niksic,M., Giglio,L., Faraguna,D. and Baralle,F.E. (2000) Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a non-evolutionary conserved intronic element. *J. Biol. Chem.*, **275**, 21041–21047.

130. Wang,Y., Wang,J., Gao,L., Lafyatis,R., Stamm,S. and Andreadis,A. (2005) Tau exons 2 and 10, which are misregulated in neurodegenerative diseases, are partly regulated by silencers which bind a SRp30c.SRp55 complex that either recruits or antagonizes htra2beta1. *J. Biol. Chem.*, **280**, 14230–14239.

131. Kanopka,A., Muhlemann,O. and Akusjarvi,G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre- mRNA. *Nature*, **381**, 535–538.

132. Nilsen,T.W. (2004) Too hot to splice. *Nature Struct. Mol. Biol.*, **11**, 208–209.

133. Shin,C., Feng,Y. and Manley,J.L. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature*, **427**, 553–558.

134. Hui,J., Stangl,K., Lane,W.S. and Bindereif,A. (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nature Struct. Biol.*, **10**, 33–37.

135. Martinez-Contreras,R., Fisette,J.F., Nasim,F.U., Madden,R., Cordeau,M. and Chabot,B. (2006) Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.*, **4**, e21.

136. Chen,C.D., Kobayashi,R. and Helfman,D.M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.*, **13**, 593–606.

137. Liu,H.X., Zhang,M. and Krainer,A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.*, **12**, 1998–2012.

138. Ladd,A.N. and Cooper,T.A. (2002) Finding signals that regulate alternative splicing in the post-genomic era. *Genome. Biol.*, **3**, reviews 0008.

139. Zheng,Z.M. (2004) Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J. Biomed. Sci.*, **11**, 278–294.

140. Abdul-Manan,N. and Williams,K.R. (1996) hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base- specific binding. *Nucleic Acids Res.*, **24**, 4063–4070.

141. Buratti,E., Muro,A.F., Giombi,M., Gherbassi,D., Iaconcig,A. and Baralle,F.E. (2004) RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol.*, **24**, 1387–1400.

142. Cartegni,L., Wang,J., Zhu,Z., Zhang,M.Q. and Krainer,A.R. (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.*, **31**, 3568–3571.

143. Fairbrother,W.G., Yeh,R.F., Sharp,P.A. and Burge,C.B. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science*, **297**, 1007–1013.

144. Fairbrother,W.G., Yeo,G.W., Yeh,R., Goldstein,P., Mawson,M., Sharp,P.A. and Burge,C.B. (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res.*, **32**, W187–W190.

145. Nalla,V.K. and Rogan,P.K. (2005) Automated splicing mutation analysis by information theory. *Hum. Mutat.*, **25**, 334–342.

146. Pagani,F., Raponi,M. and Baralle,F.E. (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl Acad. Sci. USA*, **102**, 6368–6372.

147. Pagani,F., Buratti,E., Stuani,C. and Baralle,F.E. (2003) Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J. Biol. Chem.*, **278**, 26580–26588.

148. Pagani,F., Stuani,C., Tzetis,M., Kanavakis,E., Efthymiadou,A., Doudounakis,S., Casals,T. and 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.

149. Cartegni,L. and Krainer,A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature Genet.*, **30**, 377–384.

150. Liu,H.X., Cartegni,L., Zhang,M.Q. and Krainer,A.R. (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nature Genet.*, **27**, 55–58.

151. Sironi,M., Cagliani,R., Pozzoli,U., Bardoni,A., Comi,G.P., Giorda,R. and Bresolin,N. (2002) The dystrophin gene is alternatively spliced throughout its coding sequence. *FEBS Lett.*, **517**, 163–166.

152. O'Driscoll,M., Ruiz-Perez,V.L., Woods,C.G., Jeggo,P.A. and Goodship,J.A. (2003) A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nature Genet.*, **33**, 497–501.

153. Han,K., Yeo,G., An,P., Burge,C.B. and Grabowski,P.J. (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.*, **3**, e158.

154. Kralovicova,J. and Vorechovsky,I. (2006) Position-dependent repression and promotion of DQB1 intron 3 splicing by GGGG motifs. *J. Immunol.*, **176**, 2381–2388.

155. Zhu,J., Mayeda,A. and Krainer,A.R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell*, **8**, 1351–1361.

156. Eperon,I.C., Makarova,O.V., Mayeda,A., Munroe,S.H., Caceres,J.F., Hayward,D.G. and Krainer,A.R. (2000) Selection of alternative 5′ splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.*, **20**, 8303–8318.

157. Sharma,S., Falick,A.M. and Black,D.L. (2005) Polypyrimidine tract binding protein blocks the 5′ splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell*, **19**, 485–496.

158. Chou,M.Y., Underwood,J.G., Nikolic,J., Luu,M.H. and Black,D.L. (2000) Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol. Cell*, **5**, 949–957.

159. Matter,N., Herrlich,P. and Konig,H. (2002) Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*, **420**, 691–695.

160. Matter,N., Marx,M., Weg-Remers,S., Ponta,H., Herrlich,P. and Konig,H. (2000) Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. *J. Biol. Chem.*, **275**, 35353–35360.

161. Cheng,C. and Sharp,P.A. (2006) Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. *Mol. Cell. Biol.*, **26**, 362–370.

162. Ladd,A.N., Stenberg,M.G., Swanson,M.S. and Cooper,T.A. (2005) Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev. Dyn.*, **233**, 783–793.

163. Venables,J.P., Bourgeois,C.F., Dalgliesh,C., Kister,L., Stevenin,J. and Elliott,D.J. (2005) Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. *Hum. Mol. Genet.*, **14**, 2289–2303.

164. Cartegni,L., Hastings,M.L., Calarco,J.A., de Stanchina,E. and Krainer,A.R. (2006) Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.*, **78**, 63–77.

165. Kashima,T. and Manley,J.L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nature Genet.*, **34**, 460–463.

166. Hofmann,Y., Lorson,C.L., Stamm,S., Androphy,E.J. and Wirth,B. (2000) Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proc. Natl Acad. Sci. USA*, **97**, 9618–9623.

167. Miyaso,H., Okumura,M., Kondo,S., Higashide,S., Miyajima,H. and Imaizumi,K. (2003) An intronic splicing enhancer element in survival motor neuron (SMN) pre-mRNA. *J. Biol. Chem.*, **278**, 15825–15831.
168. Singh,N.N., Androphy,E.J. and Singh,R.N. (2004) *In vivo* selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes. *RNA*, **10**, 1291–1305.
169. Yeo,G. and Burge,C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
170. Reese,M.G., Eeckman,F.H., Kulp,D. and Haussler,D. (1997) Improved splice site detection in Genie. *J. Comput. Biol.*, **4**, 311–323.
171. Rogozin,I.B. and Milanesi,L. (1997) Analysis of donor splice sites in different eukaryotic organisms. *J. Mol. Evol.*, **45**, 50–59.
172. Zheng,C.L., Fu,X.D. and Gribskov,M. (2005) Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA*, **11**, 1777–1787.
173. Singh,N.K., Singh,N.N., Androphy,E.J. and Singh,R.N. (2006) Splicing of a critical exon of human survival motor neuron is regulated by a unique silencer element located in the last intron. *Mol. Cell. Biol.*, **26**, 1333–1346.
174. Groman,J.D., Hefferon,T.W., Casals,T., Bassas,L., Estivill,X., Des Georges,M., Guittard,C., Koudova,M., Fallin,M.D., Nemeth,K. *et al.* (2004) Variation in a repeat sequence determines whether a common variant of the cystic fibrosis transmembrane conductance regulator gene is pathogenic or benign. *Am. J. Hum. Genet.*, **74**, 176–179.
175. Buratti,E., Dork,T., Zuccato,E., Pagani,F., Romano,M. and Baralle,F.E. (2001) Nuclear factor TDP-43 and SR proteins promote *in vitro* and *in vivo* CFTR exon 9 skipping. *EMBO J.*, **20**, 1774–1784.
176. Buratti,E., Brindisi,A., Giombi,M., Tisminetzky,S., Ayala,Y.M. and Baralle,F.E. (2005) TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J. Biol. Chem.*, **280**, 37572–37584.
177. Wang,H.Y., Wang,I.F., Bose,J. and Shen,C.K. (2004) Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics*, **83**, 130–139.
178. Zuccato,E., Buratti,E., Stuani,C., Baralle,F.E. and Pagani,F. (2004) An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. *J. Biol. Chem.*, **279**, 16980–16988.
179. Ayala,Y.M., Pagani,F. and Baralle,F.E. (2006) TDP43 depletion rescues aberrant CFTR exon 9 skipping. *FEBS Lett.*, **580**, 1339–1344.
180. Mercado,P.A., Ayala,Y.M., Romano,M., Buratti,E. and Baralle,F.E. (2005) Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Res.*, **33**, 6000–6010.
181. Kishore,S. and Stamm,S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.
182. Tuffery-Giraud,S., Saquet,C., Chambert,S. and Claustres,M. (2003) Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. *Hum. Mutat.*, **21**, 608–614.
183. Beroud,C., Carrie,A., Beldjord,C., Deburgrave,N., Llense,S., Carelle,N., Peccate,C., Cuisset,J.M., Pandit,F., Carre-Pigeon,F. *et al.* (2004) Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene. *Neuromuscul. Disord.*, **14**, 10–18.
184. Ikezawa,M., Nishino,I., Goto,Y., Miike,T. and Nonaka,I. (1999) Newly recognized exons induced by a splicing abnormality from an intronic mutation of the dystrophin gene resulting in Duchenne muscular dystrophy. Mutations. in. brief. no. 213. Online. *Hum. Mutat.*, **13**, 170.
185. Chillon,M., Dork,T., Casals,T., Gimenez,J., Fonknechten,N., Will,K., Ramos,D., Nunes,V. and Estivill,X. (1995) A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA→G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. *Am. J. Hum. Genet.*, **56**, 623–629.
186. Highsmith,W.E., Burch,L.H., Zhou,Z., Olsen,J.C., Boat,T.E., Spock,A., Gorvoy,J.D., Quittell,L., Friedman,K.J., Silverman,L.M. *et al.* (1994) A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. *N. Engl. J. Med.*, **331**, 974–980.
187. Will,K., Dork,T., Stuhrmann,M., Meitinger,T., Bertele-Harms,R., Tummler,B. and Schmidtke,J. (1994) A novel exon in the cystic fibrosis transmembrane conductance regulator gene activated by the nonsense mutation E92X in airway epithelial cells of patients with cystic fibrosis. *J. Clin. Invest.*, **93**, 1852–1859.
188. Monnier,N., Gout,J.P., Pin,I., Gauthier,G. and Lunardi,J. (2001) A novel 3600+11.5 kb C>G homozygous splicing mutation in a black African, consanguineous CF family. *J. Med. Genet.*, **38**, E4.
189. Rump,A., Rosen-Wolff,A., Gahr,M., Seidenberg,J., Roos,C., Walter,L., Gunther,V. and Roesler,J. (2006) A splice-supporting intronic mutation in the last bp position of a cryptic exon within intron 6 of the CYBB gene induces its incorporation into the mRNA causing chronic granulomatous disease (CGD). *Gene.*, **371**, 174–181.
190. Noack,D., Heyworth,P.G., Newburger,P.E. and Cross,A.R. (2001) An unusual intronic mutation in the CYBB gene giving rise to chronic granulomatous disease. *Biochim. Biophys. Acta*, **1537**, 125–131.
191. Christie,P.T., Harding,B., Nesbit,M.A., Whyte,M.P. and Thakker,R.V. (2001) X-linked hypophosphatemia attributable to pseudoexons of the PHEX gene. *J. Clin. Endocrinol. Metab.*, **86**, 3840–3844.
192. Balz,V., Prisack,H.B., Bier,H. and Bojar,H. (2002) Analysis of BRCA1, TP53, and TSG101 germline mutations in German breast and/or ovarian cancer families. *Cancer. Genet. Cytogenet.*, **138**, 120–127.
193. Wang,M., Dotzlaw,H., Fuqua,S.A. and Murphy,L.C. (1997) A point mutation in the human estrogen receptor gene is associated with the expression of an abnormal estrogen receptor mRNA containing a 69 novel nucleotide insertion. *Breast Cancer Res. Treat.*, **44**, 145–151.
194. van den Hurk,J.A., van de Pol,D.J., Wissinger,B., van Driel,M.A., Hoefsloot,L.H., de Wijs,I.J., van den Born,L.I., Heckenlively,J.R., Brunner,H.G., Zrenner,E. *et al.* (2003) Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. *Hum. Genet.*, **113**, 268–275.
195. Metherell,L.A., Akker,S.A., Munroe,P.B., Rose,S.J., Caulfield,M., Savage,M.O., Chew,S.L. and Clark,A.J. (2001) Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. *Am. J. Hum. Genet.*, **69**, 641–646.
196. McConville,C.M., Stankovic,T., Byrd,P.J., McGuire,G.M., Yao,Q.Y., Lennox,G.G. and Taylor,M.R. (1996) Mutations associated with variant phenotypes in ataxia-telangiectasia. *Am. J. Hum. Genet.*, **59**, 320–330.
197. Vetrini,F., Tammaro,R., Bondanza,S., Surace,E.M., Auricchio,A., De Luca,M., Ballabio,A. and Marigo,V. (2006) Aberrant splicing in the ocular albinism type 1 gene (OA1/GPR143) is corrected in vitro by morpholino antisense oligonucleotides. *Hum. Mutat.*, **27**, 420–426.
198. Raponi,M., Upadhyaya,M. and Baralle,D. (2006) Functional splicing assay shows a pathogenic intronic mutation in neurofibromatosis type 11 (NF1) due to intronic sequence exonization. *Hum. Mutat.*, **27**, 294–295.