Activation of Cryptic 3′ Splice-Sites by SRSF2 Contributes to Cassette Exon Skipping

Heegyum Moon †, Ha Na Jang, Yongchao Liu, Namjeong Choi, Jagyeong Oh, Jiyeon Ha, Xuexiu Zheng * and Haihong Shen *

School of Life Sciences, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea
* Correspondence: xuexiuzheng@gist.ac.kr (X.Z.); haihongshen@gist.ac.kr (H.S.);
  Tel.: +82-62-715-2484 (X.Z. & H.S.)
† Current Affiliation: New Drug Development Center, Daegu-Gyeongbuk Medical Innovation Foundation, 88 Dongnae-ro, Dong-gu, Daegu 41061, Korea.

Abstract: Here we show that the serine/arginine rich splicing factor 2 (SRSF2) promotes cryptic 3′ splice-site (3′AG′) usage during cassette exon exclusion in survival of motor neuron (SMN2) minigenes. Deletion of the 3′AG′ (3′AG′1), its associated branch point (BP′) and polypyrimidine tract (PPT′) sequences directs SRSF2 to promote a second 3′AG′ (3′AG′2) with less conserved associated region for intron splicing. Furthermore, deletion of both 3′AG′1 and 3′AG′2 and their associated sequences triggered usage of a third 3′AG′3 that has very weak associated sequences. Interestingly, when intron splicing was directed to the 3′AG′ cryptic splice-sites, intron splicing from the canonical 3′AG splice-site was reduced along with a decrease in cassette exon inclusion. Moreover, multiple SRSF2 binding sites within the intron are responsible for 3′AG′ activation. We conclude that SRSF2 facilitates exon exclusion by activating a cryptic 3′AG′ and inhibiting downstream intron splicing.

Keywords: SRSF2; cryptic 3′ splice-site; exon exclusion; intron retention; pre-mRNA splicing

1. Introduction

Splicing occurs at the consensus sequences near the 5′ and 3′ ends of introns, known as 5′ and 3′ splice-sites (5′ss and 3′ss) by a large, dynamic RNA-protein complex called the spliceosome [1]. In the spliceosome, small nuclear ribonucleoproteins (snRNPs) and several other proteins are recruited to the pre-mRNA [2]. The assembly of spliceosome, U1 small nuclear ribonucleoprotein (snRNP) is recruited to the 5′ss, U2 snRNP is directed to branch point (BP) sequence of intron. In the recognition of 3′ss and its associated region, U2 auxiliary factor 65 (U2AF65) identifies polypyrimidine tract (PPT), U2AF35 recognizes AG, SF1 recognizes BP sequence, U2 snRNA is basepaired with BP sequence to stabilize the spliceosome [1,3–5]. The recruitment of U4/U5/U6 tri-snRNP into spliceosome leads to catalytic activation of the spliceosome. Mutations in 5′ss, 3′ss, BP sequence and PPT causes genetic diseases by alter splicing efficiency [6]. Cryptic 5′ss or 3′ss instead of canonical splice-sites are sometimes activated [7]. In humans, while 5′ss, 3′ss and BP have become more degenerate and less conserved compared to lower eukaryotes such as yeast, polypyrimidine tract are stronger [8,9]. Alternative splicing generates diversity of mRNA isoforms and protein variants by selecting different combinations of 5′ and 3′ splice-site pairs to mediate gene regulation. High-throughput and genome-wide technologies indicate that alternative splicing occurs within transcripts from ~95% human multi-exon genes [10]. 5′ss, 3′ss, BP and PPT are not sufficient to select correct splice-site to regulate alternative splicing. In addition to these core splicing signals, RNA motifs known as exonic/intronic splicing enhancers or silences are required for further regulating alternative splicing. Therefore, more abundant enhancer and inhibitor sequences are identified within exon and introns in human genomes. The splicing enhancers

or silencers are bound by trans-acting splicing factors such as serine/arginine-rich SR proteins [11]. SR proteins include N-terminal RNA recognition motif (RRM) that interacts with the pre-mRNA and C-terminal arginine/serine-rich (RS) protein interaction domain [12,13]. SRSF2 is a member of SR proteins, and first identified using a monoclonal antibody against purified spliceosome [14,15]. SRSF2 has been reported to interact with SSNG (S=C/G), G/A rich or purine-rich sequences from structure-based study, SELEX, CLIP-seq analysis [16,17]. In addition to the activator function in pre-mRNA splicing, SRSF2 was also reported to directly repress intron splicing [18]. In the alternative splicing, SRSF2 was demonstrated to either activate exon inclusion or exon skipping [19]. Furthermore, SRSF2 activates transcription elongation and couples transcription and splicing [20,21]. Spinal muscular atrophy (SMA) is a leading genetic cause of pediatric mortality, in which homozygous loss or mutation of survival motor neuron 1 (SMN1) gene occurred [22,23]. Humans also have a duplicate of SMN1 gene, SMN2, however, a mutation (C6T) in SMN2 causes predominant skipping of exon 7 to produce a truncated protein isoform that is degraded immediately [24]. An alternative cryptic exon in intron 6 of SMN1 and SMN2 was shown to be generated by exonization of an intronic Alu-like sequences of SMN1,2. [25,26]. HnRNP A1, SRSF1, SRSF2, SRSF10, TDP43, SRSF30c, PSF, U2AF65, hnRNP G and hnRNP M were shown to regulate exon 7 splicing in SMN2 [18,27–34]. SMN2 is present in all of the SMA patients but not able to compensate for the loss of SMN1 because of predominant exclusion of cassette exon 7 [27,28]. 29-*O*-(2-methoxyethyl) (MOE)-modified antisense oligonucleotides (ASOs) with phosphorotheioate backbone targeting splicing inhibitor was shown to rescues severe SMA mice [35].

In this manuscript, we demonstrate that SRSF2 significantly stimulates intron splicing at a cryptic splice-site (3′AG′1) with the most conserved associated region located at upstream of exon 7 in a SMN2 minigene. 3′AG′1 activation reduced canonical 3′AG activation and also decreased cassette exon inclusion. Interestingly, deletion of 3′AG′1 and its associated sequences, the BP (hereafter called BP′) and PPT (hereafter called PPT′) induced intron splicing at a second cryptic 3′ splice-site (3′AG′2) with less conserved associated region. Furthermore, deletion of both 3′AG′1 and 3′AG′2 induced activation of the third 3′AG′ (3′AG′3) activation that has the least conserved associated sequence. Importantly, multiple SRSF2 binding sites within the intron is responsible for 3′AG′ activation. Our results reveal that 3′AG′ activation and inhibition of canonical 3′AG splicing contribute to regulation of cassette exon skipping.

2. Materials and Methods

2.1. Plasmids Construction

E6-7, E6-7 (20, 112, 226, 334, 441) and E6-C was generated using the E6-8 minigene of SMN2 gene [18] as a template. The Δ3′AG′1, Δ3′AG′1/2, Δ3′AG′1/2/3, ΔSRSF2, E7/8ex, E6-5′cons, E7-ss-mut and SRSF2-△1, 2, 3, 4 constructs were generated by site-directed mutagenesis using the E6-8 minigene as a template. The E6-A2 construct was generated using the 5′ half of E6-8 construct and the 3′ half of adenovirus major late (AdML) construct [18] as templates and inserted into NheI, EcoRI, and XhoI restriction sites of pcDNA3.1 vector. SRSF2 expression plasmid was constructed with coding region of SRSF2 sequence into pcDNA6/myc/His A vector using BamHI and XhoI restriction sites. All primer sequences used to construct the minigenes are listed in Table S1.

2.2. Cell Culture, Transfection and RT-PCR

We grew 293T cells in Dulbecco's Modified Eagle Medium (DMEM) media supplemented with antibiotics and 10% FBS at 37 °C and in 5% CO2. SMA patient cell line GM3813 was cultured in DMEM media with DMEM and 10% FBS by adding non-essential amino acids. We transfected all of the minigene plasmids into cells using polyethyleneimide (PEI) as previously described [18]. Total cellular RNA was extracted using the RiboEx reagent (GeneAll) and conventional RT-PCR was performed as described previously [18]. All primer sequences are listed in Table S1. Quantitation of RT-PCR result was performed by ImageJ software as described previously [29].

2.3. Knockdown of SRSF2 with Lentivirus-Mediated shRNA

The SRSF2 targeted shRNA lentivirus was generated by co-transfection of SRSF2-targeting shRNA plasmid (openbiosystems) and helper plasmids into 293T cells using polyethyleneimide (PEI) reagent. Lentivirus containing supernatant was harvested and added to the GM3813 cells with the supplement of 10 μg/mL polybrene.

3. Results

3.1. SRSF2 Promotes Cryptic 3′AG′ Activation

We have previously demonstrated that SRSF2 promotes a cryptic 3′ splice-site (3′AG′) that is located at 682 nt upstream of 3′AG at exon 7 while simultaneously suppressing splicing at the 3′AG splice-site in the SMN2 minigene [18]. In the minigene we studied, the cryptic exon that was previously described [25,26] was not included. To differentiate 3′AG′ from other putative cryptic splice-sites that are described in the other parts of the manuscript, we designated the 3′AG′ as 3′AG′1, a 22 nt polyuridine sequence that functions potentially as a polypyrimidine tract as PPT′, and a potential branch-point sequence upstream of the PPT as BP′ (summarized in Figure 1A).

An important question was whether remote exons and introns are required for SRSF2-mediated cryptic 3′AG′ splice-site activation. To address the question, we constructed the E6-7 minigene (Figure 1B, upper panel) where intron 7 and exon 8 were deleted from the E6-8 minigene, which includes shorter length (1 kb) of intron 6 and full length of exon 8 [31]. As previously reported, E6-8 minigene produced the cryptic 3′AG′1 splice-site activated product (Figure 1B, lower panel, lane 2) [18]. Similarly, E6-7 minigene also produced 3′AG′1 activated product but to a much lesser extent (Figure 1B, lower panel, lane 4) compared to the E6-8 minigene (lane 2). Additionally, splicing at the canonical 3′AG splice-site was suppressed, but not as extensively as seen in the E6-8 minigene (Figure 1B, lower panel, lane 4), indicating that downstream introns and exons are necessary for SRSF2 activity. It was possible that the length of the downstream intron could also affect SRSF2 function. To address this possibility, we constructed a series of minigenes where 20 [E6-7 (20)], 112 [E6-7 (112)], 226 [E6-7 (226)], 334 [E6-7 (334)], and 441 [E6-7 (441)] nucleotides of intron 7 were retained, respectively (Figure 1B, upper panel). We found that SRSF2 stimulated splicing at 3′AG′1 in the E6-7 (20) minigene to similar levels as in the E6-7 minigene (Figure 1B, lane 6). These results reveal that the 20 nt intronic sequence is not important for intron splicing at the 3′AG′ cryptic splice-acceptor sites. Additionally, for SRSF2-expressing cells, more splicing at 3′AG′1 occurred in longer intronic sequences than do shorter ones (Figure 1B, lanes 8, 10, 12 and 14). Therefore, 3′AG′ activation occurs more readily with longer downstream introns in the presence of SRSF2 in SMN pre-mRNA. Finally, it was important to determine whether splicing occurs at cryptic splice-sites in a gene construct with only one exon and intron. Surprisingly, splicing at 3′AG′1 occurs in a minigene lacking exon 7 (E6-C) (Figure 1C, left panel) even if SRSF2 is absent (Figure 1C, right panel, lane 1). That cryptic splice-sites are more easily activated in minigenes lacking an exon was confirmed by constructing a chimeric minigene (E6-A2) composed of a part from exon 6 and upstream intron 6 sequences in the E6-7 minigene and the other part from exon 2 (A2) and upstream intron of the adenovirus major late (AdML) pre-mRNA derivative (Figure 1D, left panel). Splicing of cryptic exon was observed only when SRSF2 was expressed (Figure 1D, right panel, compare lanes 1 and 2). Therefore, we conclude that SRSF2 suppresses intron 6 splicing by activating cryptic 3′AG′. To consider the possibility that reduced SRSF2 expression regulates cryptic 3′AG′ activation, we performed shRNA-mediated reduction of SRSF2. As shown in Supplementary Figure S1, the cryptic 3′AG′ splice-site was not activated by SRSF2-targeting shRNA (Supplementary Figure S1).
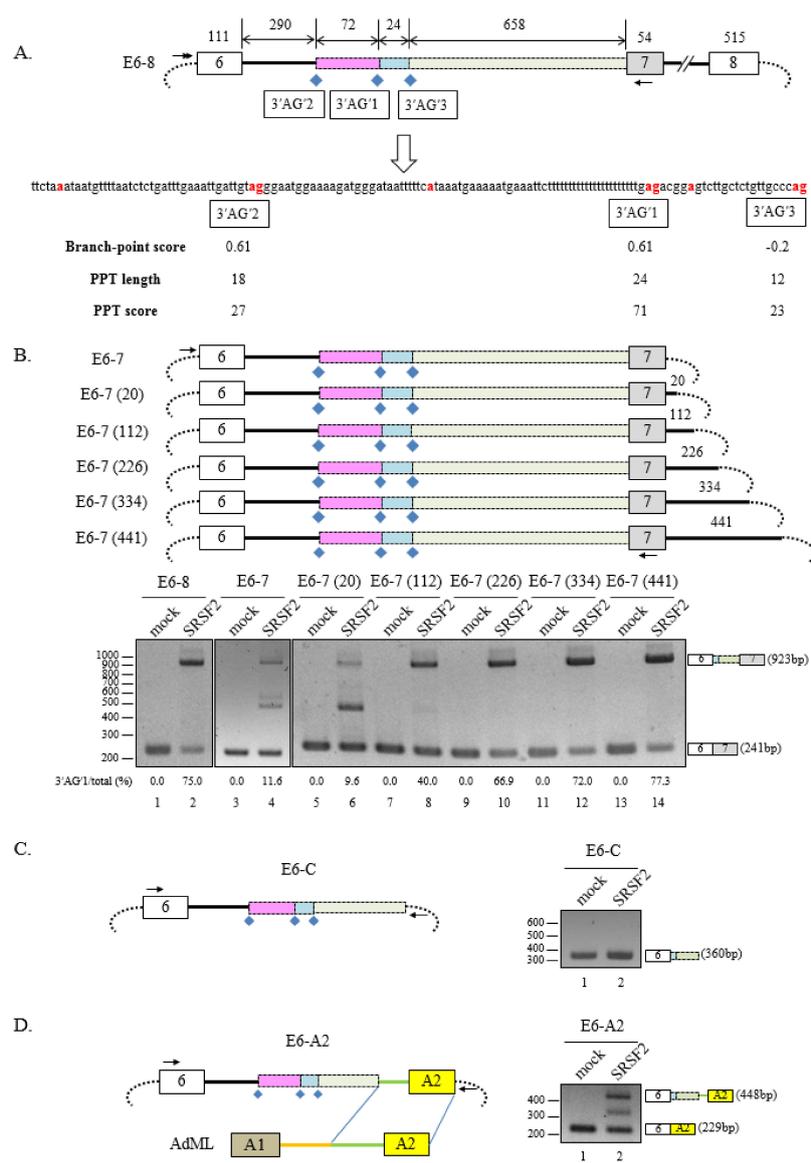
Figure 1. The serine/arginine rich splicing factor 2 (SRSF2) promotes cryptic 3′AG′ splice-site activation. (**A**) Schematic diagram of the E6-8 minigene. Exons are depicted as numbered gray and white boxes, introns as solid lines, vector sequence as dots and acryptic 3′AG′ sites as blue diamonds. The 3′AG′1, 3′AG′2 3′AG′3 and associated sequences are shown below the schematic diagram. The fragment from 3′AG′3 to the end of intron 6 is indicated in light green, the fragment from 3′AG′1 to 3′AG′3 is indicated in cyan, and the fragment between 3′AG′2 and 3′AG′1 is indicated in pink. Additionally, the 3′AG′ sequence and BP′ adenosine are highlighted in red, the primer binding sites used for RT-PCR reaction are indicated as arrows, and BP scores, PPT length, and PPT scores of each 3′AG′s are shown. (**B**) (Upper panel) Schematic diagram of the E6-7 minigene and various constructs. The primer bindings sites used for RT-PCR analysis are indicated as arrows. (Lower panel) RT-PCR analysis of intron 6 splicing within the E6-8, E6-7 minigene or various mutant constructs using RNAs extracted fromSRSF2-expressing or pcDNA transfected cells. The sizes of products are shown at the right of all figures here and below. (**C**) (Left panel) Schematic diagram of the E6-C minigene indicating primer binding sites. (Right panel) RT-PCR analysis of the E6-C minigene using RNAs extracted from SRSF2-expressing or pcDNA transfected cells. (**D**) (Left panel) Schematic diagram of the E6-A2 and AdML minigenes indicating primer binding sites. (Right panel) RT-PCR analysis of intron splicing within the E6-A2 minigene in pcDNA- or SRSF2-expressing cells.

3.2. Deletion of 3′AG′ Region Induces SRSF2-Dependent Usage of Alternate 3′AG′

Although there are several putative cryptic splice-sites within the SMN minigene (Figure 1A), it was unknown whether deletion of 3′AG′1 would prevent SRSF2 from influencing intron 6 splicing. To test this possibility, the Δ3′AG′1 minigene was generated with deletions of 3′AG′1, BP′1, and PPT′1 (Figure 2A). Cells expressing both SRSF2 and the Δ3′AG′1 minigene favored cassette exon exclusion (Figure 2A, lower panel, lane 2). Interestingly, another cryptic 3′ss (3′AG′2) located 72 nt upstream of 3′AG′1 was activated by SRSF2 (Figure 2A, lower panel, lane 4). The 3′AG′2 region includes a much weaker and shorter PPT compared with 3′AG′1 (Figure 1A) allowing for the better conserved 3′AG′1 to be activated in advance by SRSF2.

However, in the absence of a strong cryptic splice-site, alternative cryptic 3′ss splicing using a less conserved, weaker 3′AG′ might be possible when SRSF2 is present. These results raised the possibility that yet another cryptic splice-site, in the absence of both 3′AG′1 and 3′AG′2, could mediate alternative splicing of the SMN minigene. Therefore, two other minigenes with deletions of two or three cryptic splice-sites, Δ3′AG′1/2 and Δ3′AG′1/2/3, respectively, were generated to test for alternative splicing in SRSF2-expressing cells (Figure 2B,C, upper panel). As before, cassette exon exclusion was favored by both Δ3′AG′1/2 and Δ3′AG′1/2/3 minigenes (Figure 2B,C, lower panel, lane 2). However, when testing for splicing of intron 6, deletions of both 3′AG′1 and 3′AG′2 promoted alternative splicing from a third cryptic splice-site, 3′AG′3, only in cells overexpressing SRSF2 (Figure 2B, lower panel, lane 4). This was interesting because 3′AG′3 is located 24 nt downstream of 3′AG′1 and contains a much shorter PPT sequence and has less well-conserved BP sequences than 3′AG′2 (Figure 1A). Moreover, the absence of all three cryptic splice-sites (3′AG′1, 3′AG′2 and 3′AG′3) in the Δ3′AG′1/2/3 minigene resulted in only or mostly canonically spliced exons in the non-transfected or SRSF2-expressing cells, respectively (Figure 2C, lower panel, lane 4). The results imply that SRSF2 promotes usage of most conserved cryptic 3′AG′ based on branch-point score, PPT length, and PPT (SVM-BPfinder, http://regulatorygenomics.upf.edu/Software/SVM_BP/), summarized in Figure 2D.

3.3. Multiple SRSF2-Binding Sequences and Clusters Are Necessary for 3′AG′ Activation by SRSF2

Because SRSF2 is an RNA-binding protein with degenerate sequence-specificity (unlike other SR proteins), the pre-mRNA usually includes many potential SRSF2 binding sequences [19]. We observed that 12 potential SRSF2 binding sequences or clusters are located downstream of 3′AG′1 (Figure 3A, middle panel). In contrast, only one potential SRSF2 binding site is located around the canonical 3′AG. Therefore, it was possible that the abundance of potential SRSF2-binding sequences downstream of 3′AG′1 were responsible for activation of cryptic splicing. To address this possibility, we deleted 412 nucleotides from the E6-8 minigene to remove potential SRSF2 binding sites (ΔSRSF2, Figure 3A, upper panel). Although SRSF2 mediated exon skipping as before (Figure 3A, lower panel, lane 2), we also observed mostly the unspliced isoform (873bp) but no cryptically spliced isoforms of the ΔSRSF2 minigene (Figure 3A, lower panel, lane 4). These data provide strong evidence that the region of the E6-8 minigene encompassing the 412 nucleotides deletion is important for SRSF2 function in cryptic 3′AG′ activation. To further narrow the binding target sequences of SRSF2, we generated several deletion mutants of the ΔSRSF2 minigene with 103, 32, 26 or 40 nucleotide deletions (SRSF2-Δ1, SRSF2-Δ2, SRSF2-Δ3 and SRSF2-Δ4, respectively, Supplementary Figure S2A). With these smaller deletions, SRSF2 was still able to activate cryptic splicing at 3′AG′1 in all the mutant constructs (Supplementary Figure S2B). Therefore, it appears likely that several of these SRSF2-binding sequences together are responsible for cryptic splice-site activation by SRSF2. Our next task was to determine whether 3′AG′ could still be activated by SRSF2 if intron 6 is located downstream of the cassette exon (exon 7). Therefore, we swapped intron 6/exon 7 with intron 7/exon 8 (E7/8ex, Figure 3B, upper panel) and found that while intron 7 splicing was inhibited (Figure 3B, lower panel, lane 2), cleavage at 3′AG′1 was still possible (Figure 3B, lower panel, lane 4). Therefore, we concluded that 3′AG′ activation is not due to the location of the intron containing the cryptic splice-site.
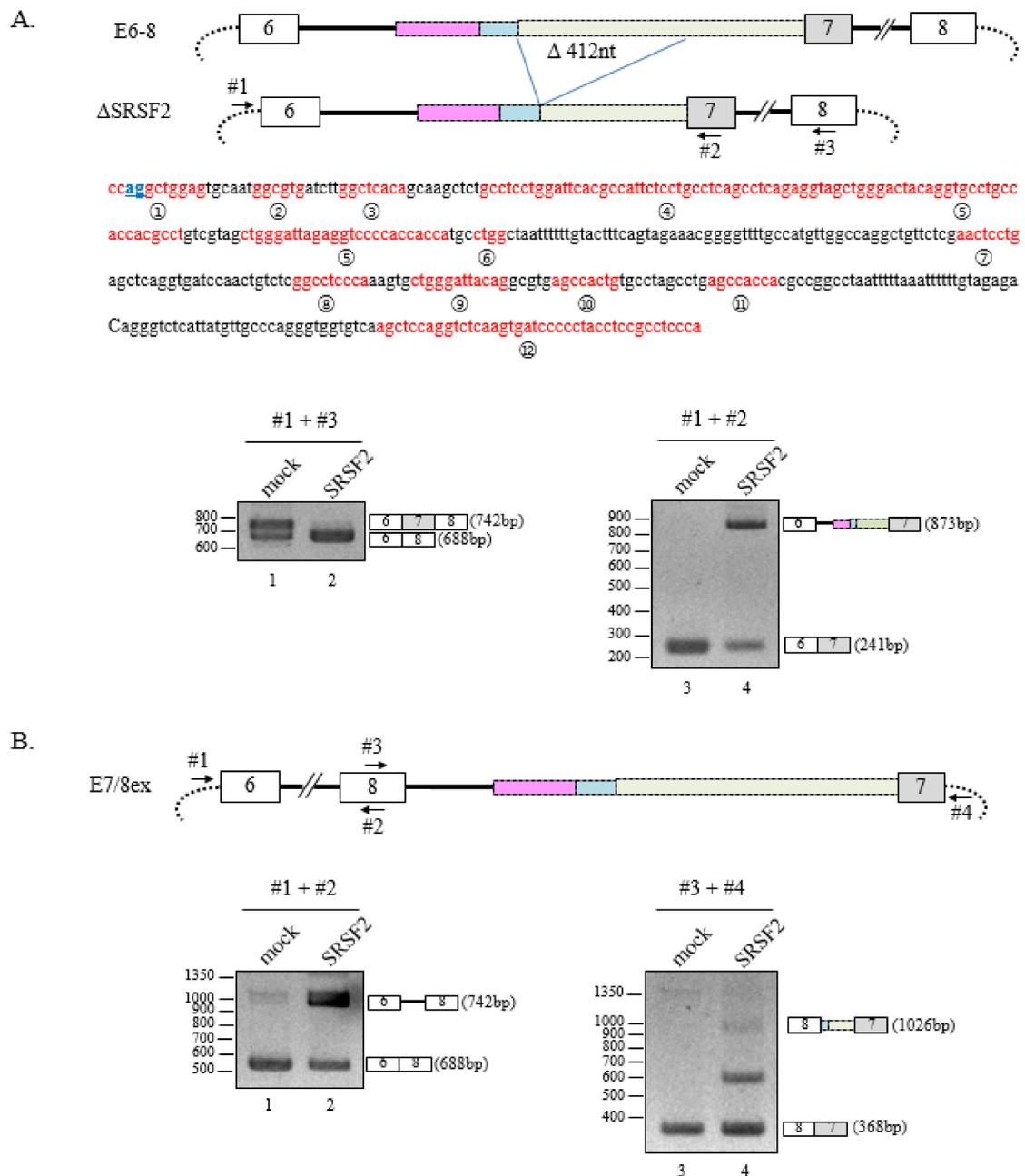
Figure 2. Deletion of 3′AG′1 region induces SRSF2-dependent usage of alternate 3′AG′. (**A**) (Upper panel) Schematic diagram of the Δ3′AG′1 minigene. The nucleotide deletion is indicated at top but is otherwise the same as Figure 1A. (Lower panel) RT-PCR analysis of the Δ3′AG′1 minigene using RNA extracted from pcDNA- or SRSF2-expressing cells with primer pairs #1 and #3 (left) or #1 and #2 (right). (**B**) (Upper panel) Schematic diagram of the Δ3′AG′1/2 minigene. Both nucleotide deletions are indicated at top but is otherwise the same as Figure 1A. (Lower panel) RT-PCR analysis of the Δ3′AG′1/2 minigene using RNA extracted from pcDNA- or SRSF2-expressing cells with primer pairs #1 and #3 (left) or #1 and #2 (right). (**C**) (Upper panel) Schematic diagram of the Δ3′AG′1/2/3 minigene. Nucleotide deletions encompassing all three cryptic splice-sites are indicated at top but is otherwise the same as in Figure 1A. (Lower panel) RT-PCR analysis of the Δ3′AG′1/2/3 minigene using primers #1 and #3 (left) or #1 and #2 (right). (**D**) Summary of the order of activation (splice-site usage) of the various cryptic splice-sites based on BP score and PPT strength.

Figure 3. Multi SRSF2-binding sequences and clusters are responsible for 3′AG′ activation by SRSF2. (**A**) (Upper panel) Schematic diagram of the ΔSRSF2 minigene. The nucleotide deletion is indicated using blue lines. The primer binding sites used in lower panels are indicated with arrows. (Middle panel) The sequence of deleted 412 nt is shown. Potential SRSF2 binding sites and clusters are shown in red and numbered. (Lower panel) RT-PCR analysis of the ΔSRSF2 minigene in SRSF2-expressing cells using primers #1 and #3 (left) or #1 and #2 (right). (**B**) (Upper panel) Schematic diagram of the E7/8ex minigene where a portion including intron 6 though exon 7 was swapped with a region containing intron 7 through exon 8. The primer binding sites used in lower panels are indicated with arrows. (Lower panel) RT-PCR analysis of the E7/8ex minigene in SRSF2-expressing cells using primers #1 and #2 (left) or #3 and #4 (right).

3.4. 5′ Splice-Site Mutations do not Affect SRSF2-Mediated 3′AG′ Activation

It remained unknown whether 5′ splice-site mutations affect cryptic 3′ splice-site activation by SRSF2. First, we tested the E7-5′cons minigene where the 5′ss of the cassette exon (exon 7) was mutated

to a more well-conserved sequence (Figure 4A, left panel). Although cassette exon exclusion was not observed, splicing at 3′AG′1 was observed when SRSF2 was overexpressed (Figure 4A, right panel). Therefore, a better conserved 5′ss did not affect SRSF2-mediated cryptic splicing. Next, we examined the E6-5′cons minigene where the 5′ss of exon 6 was mutated to a conserved sequence (Figure 4B, left panel). As before, alternative splicing of the minigene at 3′AG′ was observed but both cells, overexpressing SRSF2 or not, were now able to activate 3′AG′1 (Figure 4B, right panel). Finally, we wondered whether disruption of 5′ splice-donor and 3′ splice-acceptor site in the cassette exon would affect SRSF2-mediated 3′AG′ activation. We tested another mutant minigene, E7-ss-mut, where the splice-site sequences of exon 7 were abolished, effectively converting E6-8 to a two exon minigene (Figure 4C, left panel). Here, alternative splicing at 3′AG′ was still possible in the absence of splicing signals but the majority of transcripts had excluded exon 7 (Figure 4C, right panel, lane 2). Together, these results provide strong evidence that 3′AG′ can always be activated by SRSF2 regardless of 5′ss strength.
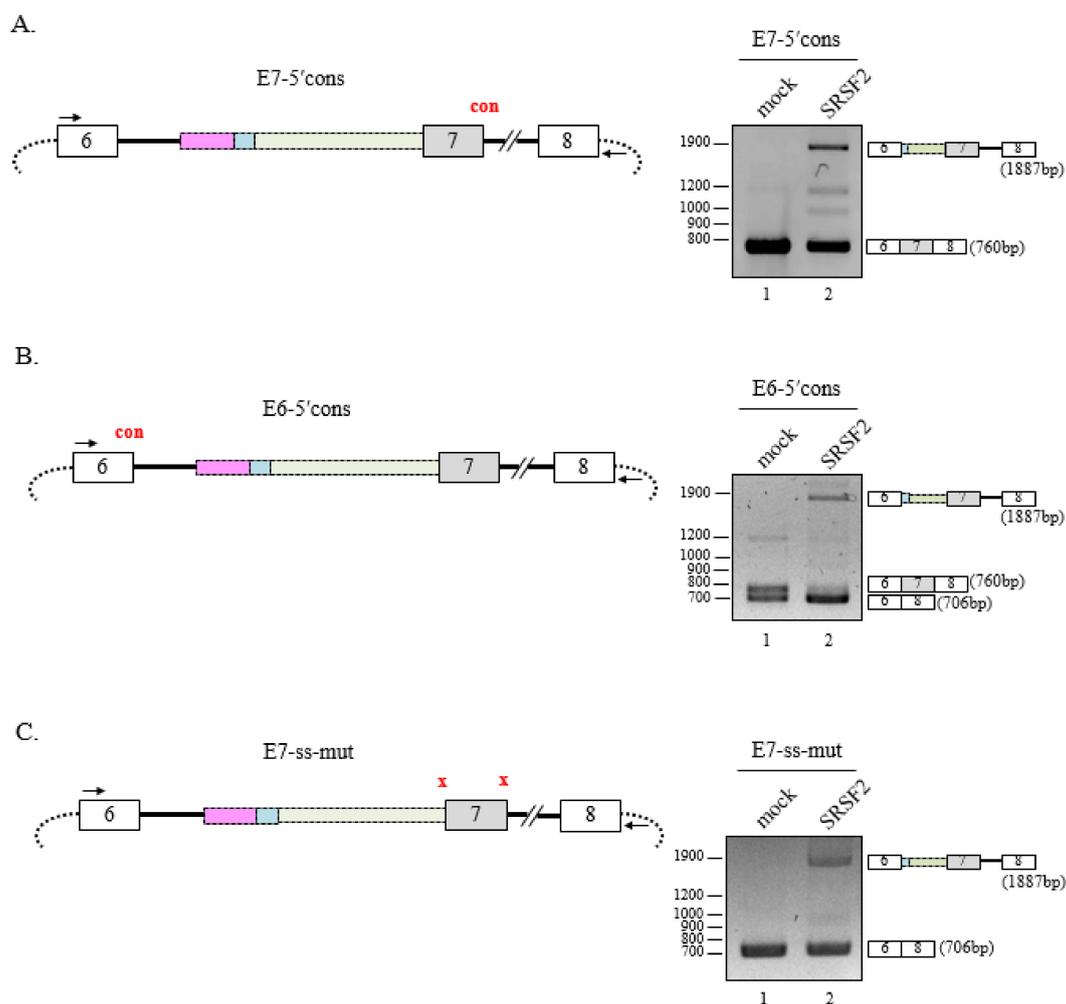

Figure 4. 5′ splice-site mutations do not affect SRSF2-mediated 3′AG′ activation. (**A**) (Left panel) Schematic diagram of the E7-5′cons minigene where the 5′ss has been changed to a conserved sequence. (Right panel) RT-PCR analysis of the E7-5′cons minigene in SRSF2-expressing cells. (**B**) (Left panel) Schematic diagram of the E6-5′cons minigene including the conserved 5′ss sequence of exon 6. (Right panel) RT-PCR analysis of the E6-5′cons minigene in SRSF2-expressing cells. (**C**) (Left panel) Schematic diagram of the E7-ss-mut minigene where the splice signals of exon 7 have been abolished. (Right panel) RT-PCR analysis of the E7-ss-mut minigene in SRSF2-expressing cells.

4. Discussion

In this study, our data showed that among three cryptic 3′AGs, the 3′AG′1 with most conserved associated sequence of intron 6 is first activated by SRSF2 under conditions favoring exon skipping, and 3′AG′ activation reduced splicing from the canonical 3′AG to inhibit cassette exon splicing. Interestingly, when we deleted 3′AG′1 and its upstream PPT′ and BP′, we found that the second 3′AG′2 having less conserved associated region among two remaining 3′AG's was activated. Furthermore, deletion of both 3′AG′ (3′AG′1, 3′AG′2) induced the activation of the third remaining 3′AG′3 by SRSF2. Importantly, 3′AG′ activation requires multiple SRSF2 binding sites within the intron. It was previously reported that a cryptic exon in intron 6 is included in a mRNA to produce another isoform of SMN. Splicing of the cryptic exon occurred by an exonization event of Alu element [25,26]. Here we identified novel three cryptic 3′AG′ in intron 6 of SMN2 gene, that was activated by SRSF2 in a SMN2 minigene.

We previously reported that SRSF2 directly inhibits intron 7 splicing and BP surrounding 10 nt is essential for the SRSF2-mediated intron repression, and the intron 7 inhibition contribute directly to cassette exon exclusion [18]. In this manuscript, we present that SRSF2 suppresses another flanking intron 7 splicing by promoting a cryptic 3′AG′ splice-site. The previous and current results demonstrate that SRSF2 simultaneously inhibits flanking 5′ss and 3′ss splicing by either directly inhibiting downstream intron splicing or activating cryptic 3′AG′. These observations provide an insight for the regulation of SR proteins.

Although 3′AG′ activation by other SR proteins, such as SRSF1 and SRSF5, was described previously [36], 3′AG′ activation by SRSF2 has never been reported. An important observation is that SRSF2 could activate the 3′AG′3 with much less conserved associated BP′ and PPT′. Among three potential 3′AG's, the 3′AG′ with most strong associated sequences, including conserved BP′ and abundance of pyrimidine nucleotides in PPT′, are always activated. 3′AG′ activation did not occur in a location-dependent manner. Notably, deletion of all 3′AG's did not totally abrogate SRSF2 function as exon skipping was still possible. It is possible that the presence of many SRSF2-binding sequences could still bring the splicing machinery around to the 3′AG's thus deletion of 3′AG's reduced canonical 3′AG splicing. Others have suggested that SR proteins interact with RNA in one location to activate splicing. However, it was also reported that SR proteins coordinate with each other to activate splicing [37,38]. In agreement with these reports, we demonstrated that multiple SRSF2-binding sites are required for 3′AG′ activation.

Compared with other SR proteins, SRSF2 has more diverse binding consensus [16]. SSNG (S = C/G), GA-rich or purine-rich sequences were identified as the consensus sequences in different approaches [16,17]. Although we performed shRNA treatment for SRSF2, we could not observe endogenous 3′AG′ activation by reduced SRSF2 expression. The difference of the results is also caused by experimental differences: minigenes or endogenous pre-mRNA, overexpression or shRNA treatment. Another possibility is that the minigene we tested include only ~1 kb sequence, but the other ~4.8 kb RNA has important roles in the endogenous splicing. Furthermore, additional regulatory mechanisms play roles in the endogenous 3′AG′. In addition, we performed all experiments with SRSF2-overexpressing cells, and the shRNA effect is not always opposite to that of overexpression. The degenerate consensus of SRSF2 binding targets probably make endogenous SRSF2 forms a multi-molecular complex with different roles. SRSF2 mutations occur frequently in patients with myelodysplastic syndromes (MDS) [39,40]. How these mutations in SRSF2 regulate 3′AG′ in MDS still needs to be determined.

Author Contributions: H.S. designed the concept of the present study and supervised it. H.M., H.N.J., Y.L., N.C., J.O., J.H. acquired, analyzed and interpreted the data. H.M., X.Z. and H.S. wrote the manuscript. All authors read and approved the manuscript.

References

1. Nelson, K.K.; Green, M.R. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev.* **1989**, *3*, 1562–1571. [CrossRef] [PubMed]

2. Matera, A.G.; Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 108–121. [CrossRef] [PubMed]

3. Berglund, J.A.; Chua, K.; Abovich, N.; Reed, R.; Rosbash, M. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **1997**, *89*, 781–787. [CrossRef]

4. Wu, S.; Romfo, C.M.; Nilsen, T.W.; Green, M.R. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **1999**, *402*, 832–835. [CrossRef] [PubMed]

5. Zamore, P.D.; Patton, J.G.; Green, M.R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **1992**, *355*, 609–614. [CrossRef] [PubMed]

6. Faustino, N.A.; Cooper, T.A. Pre-mRNA splicing and human disease. *Genes Dev.* **2003**, *17*, 419–437. [CrossRef]

7. Buratti, E.; Chivers, M.; Hwang, G.; Vorechovsky, I. DBASS3 and DBASS5: Databases of aberrant 3'- and 5'-splice sites. *Nucleic Acids Res.* **2011**, *39*, D86–D91. [CrossRef]

8. Keren, H.; Lev-Maor, G.; Ast, G. Alternative splicing and evolution: Diversification, exon definition and function. *Nat. Rev. Genet.* **2010**, *11*, 345–355. [CrossRef]

9. Schwartz, S.H.; Silva, J.; Burstein, D.; Pupko, T.; Eyras, E.; Ast, G. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* **2008**, *18*, 88–103. [CrossRef]

10. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415. [CrossRef]

11. Graveley, B.R. Sorting out the complexity of SR protein functions. *RNA* **2000**, *6*, 1197–1211. [CrossRef] [PubMed]

12. Moon, H.; Cho, S.; Loh, T.J.; Oh, H.K.; Jang, H.N.; Zhou, J.; Kwon, Y.S.; Liao, D.J.; Jun, Y.; Eom, S.; et al. SRSF2 promotes splicing and transcription of exon 11 included isoform in Ron proto-oncogene. *Biochim. Biophys. Acta* **2014**, *1839*, 1132–1140. [CrossRef] [PubMed]

13. Fu, X.D. Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* **1993**, *365*, 82–85. [CrossRef] [PubMed]

14. Fu, X.D.; Maniatis, T. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **1990**, *343*, 437–441. [CrossRef] [PubMed]

15. Fu, X.D.; Maniatis, T. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1725–1729. [CrossRef] [PubMed]

16. Daubner, G.M.; Clery, A.; Jayne, S.; Stevenin, J.; Allain, F.H. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.* **2012**, *31*, 162–174. [CrossRef]

17. Tacke, R.; Manley, J.L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **1995**, *14*, 3540–3551. [CrossRef]

18. Moon, H.; Cho, S.; Loh, T.J.; Jang, H.N.; Liu, Y.; Choi, N.; Oh, J.; Ha, J.; Zhou, J.; Cho, S.; et al. SRSF2 directly inhibits intron splicing to suppresses cassette exon inclusion. *BMB Rep.* **2017**, *50*, 423–428. [CrossRef]

19. Pandit, S.; Zhou, Y.; Shiue, L.; Coutinho-Mansfield, G.; Li, H.; Qiu, J.; Huang, J.; Yeo, G.W.; Ares, M., Jr.; Fu, X.D. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell* **2013**, *50*, 223–235. [CrossRef]

20. Lin, S.; Coutinho-Mansfield, G.; Wang, D.; Pandit, S.; Fu, X.D. The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* **2008**, *15*, 819–826. [CrossRef]

21. Das, R.; Yu, J.; Zhang, Z.; Gygi, M.P.; Krainer, A.R.; Gygi, S.P.; Reed, R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **2007**, *26*, 867–881. [CrossRef] [PubMed]

22. Crawford, T.O.; Pardo, C.A. The neurobiology of childhood spinal muscular atrophy. *Neurobiol. Dis.* **1996**, *3*, 97–110. [CrossRef] [PubMed]

23. Lefebvre, S.; Burglen, L.; Reboullet, S.; Clermont, O.; Burlet, P.; Viollet, L.; Benichou, B.; Cruaud, C.; Millasseau, P.; Zeviani, M.; et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **1995**, *80*, 155–165. [CrossRef]

24. Lorson, C.L.; Hahnen, E.; Androphy, E.J.; Wirth, B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6307–6311. [CrossRef] [PubMed]

25. Yoshimoto, S.; Harahap, N.I.; Hamamura, Y.; Ar Rochmah, M.; Shima, A.; Morisada, N.; Shinohara, M.; Saito, T.; Saito, K.; Lai, P.S.; et al. Alternative splicing of a cryptic exon embedded in intron 6 of SMN1 and SMN2. *Hum. Genome Var.* **2016**, *3*, 16040. [CrossRef] [PubMed]

26. Seo, J.; Singh, N.N.; Ottesen, E.W.; Lee, B.M.; Singh, R.N. A novel human-specific splice isoform alters the critical C-terminus of Survival Motor Neuron protein. *Sci. Rep.* **2016**, *6*, 30778. [CrossRef] [PubMed]

27. Kashima, T.; Manley, J.L. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* **2003**, *34*, 460–463. [CrossRef] [PubMed]

28. Cartegni, L.; Krainer, A.R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* **2002**, *30*, 377–384. [CrossRef]

29. Cho, S.; Moon, H.; Loh, T.J.; Jang, H.N.; Liu, Y.; Zhou, J.; Ohn, T.; Zheng, X.; Shen, H. Splicing inhibition of U2AF65 leads to alternative exon skipping. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9926–9931. [CrossRef]

30. Bose, J.K.; Wang, I.F.; Hung, L.; Tarn, W.Y.; Shen, C.K. TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *J. Biol. Chem.* **2008**, *283*, 28852–28859. [CrossRef]

31. Cho, S.; Moon, H.; Loh, T.J.; Oh, H.K.; Cho, S.; Choy, H.E.; Song, W.K.; Chun, J.S.; Zheng, X.; Shen, H. hnRNP M facilitates exon 7 inclusion of SMN2 pre-mRNA in spinal muscular atrophy by targeting an enhancer on exon 7. *Biochim. Biophys. Acta* **2014**, *1839*, 306–315. [CrossRef] [PubMed]

32. Cho, S.; Moon, H.; Loh, T.J.; Oh, H.K.; Williams, D.R.; Liao, D.J.; Zhou, J.; Green, M.R.; Zheng, X.; Shen, H. PSF contacts exon 7 of SMN2 pre-mRNA to promote exon 7 inclusion. *Biochim. Biophys. Acta* **2014**, *1839*, 517–525. [CrossRef]

33. Hofmann, Y.; Wirth, B. hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum. Mol. Genet.* **2002**, *11*, 2037–2049. [CrossRef] [PubMed]

34. Young, P.J.; DiDonato, C.J.; Hu, D.; Kothary, R.; Androphy, E.J.; Lorson, C.L. SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. *Hum. Mol. Genet.* **2002**, *11*, 577–587. [CrossRef] [PubMed]

35. Hua, Y.; Sahashi, K.; Rigo, F.; Hung, G.; Horev, G.; Bennett, C.F.; Krainer, A.R. Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **2011**, *478*, 123–126. [CrossRef] [PubMed]

36. Buratti, E.; Stuani, C.; De Prato, G.; Baralle, F.E. SR protein-mediated inhibition of CFTR exon 9 inclusion: Molecular characterization of the intronic splicing silencer. *Nucleic Acids Res.* **2007**, *35*, 4359–4368. [CrossRef] [PubMed]

37. Crawford, J.B.; Patton, J.G. Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol. Cell. Biol.* **2006**, *26*, 8791–8802. [CrossRef] [PubMed]

38. Lynch, K.W.; Maniatis, T. Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev.* **1996**, *10*, 2089–2101. [CrossRef] [PubMed]

39. Zhang, J.; Lieu, Y.K.; Ali, A.M.; Penson, A.; Reggio, K.S.; Rabadan, R.; Raza, A.; Mukherjee, S.; Manley, J.L. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4726–E4734. [CrossRef] [PubMed]

40. Kim, E.; Ilagan, J.O.; Liang, Y.; Daubner, G.M.; Lee, S.C.; Ramakrishnan, A.; Li, Y.; Chung, Y.R.; Micol, J.B.; Murphy, M.E.; et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* **2015**, *27*, 617–630. [CrossRef]