Activation of the intronic cryptic 5′ splice site depends on its distance to the upstream cassette exon

Wei Liu a,1, Xia Li a,1, Shengjie Liao b,c, Kefeng Dou a,⁎, Yi Zhang b,c,⁎⁎

⁎⁎ Correspondence to: Y. Zhang, Center for Genome Analysis, ABLife Inc., Optics Valley International Biomedical Park, Building 9-4, East Lake High-Tech Development Zone, 388 Gaoxin 2nd Road, Wuhan, Hubei 430075, China
⁎ Corresponding author.

A R T I C L E  I N F O
Article history:
Received 11 January 2017
Received in revised form 13 March 2017
Accepted 17 March 2017
Available online 18 March 2017

Keywords:
Alternative splicing
Splice site selection
ESE
Positional effect
Cryptic splice site

A B S T R A C T
Splice site selection is a key step that determines the mRNA isoforms generated from a single transcript. The large diversity in splice site sequences emphasizes the plasticity of splice site recognition and selection. In this report, a cell-based reporter system using a SMN1/2 cassette exon was applied to study the roles governing the activation of a cryptic 5′SS from the intron 4 of the CT/GRP gene. We found that the cryptic site was activated when placed within 124 nt downstream the cassette exon, and the level of activation was negatively correlated with its distance from the exon. In addition, activation was not affected by PTB but was eliminated by an insertion extending the exon length. Activated cryptic 5′SSs in intron or exon could override the original alternative 5′SS, obeying the U1 base-pairing rule. These results suggest that the exon length itself could represent a factor in determining the splice site selection.

© 2017 Published by Elsevier B.V.

1. Introduction

Pre-mRNA alternative splicing is an essential post-transcriptional mechanism that is extensively used in higher eukaryotes to expand proteomic diversity. For each gene, altering the choice of splice site produces diverse mature mRNAs composed of different exons that encode different protein isoforms (Chen and Manley, 2009; Nilsen and Graveley, 2010). In contrast to yeast, metazoan pre-mRNAs are often extremely long and contain short exons and very long introns. The precise recognition of a splice site is a prerequisite for alternative splicing and the selection of short exons from long introns requires sophisticated mechanisms for splice site recognition (Reed, 1996, Wahl et al., 2013).

The consensus sequence elements of splice sites are extremely short and poorly conserved, and the mechanism by which the spliceosome distinguishes authentic splice sites from a huge number of cryptic splice sites remains a long-standing question (Black, 1995; Sun and Chasin, 2000). The exon definition model proposed more than two decades ago well explains the recognition of internal exons among long introns in vertebrates, and it invokes the concerted recognition of upstream 3′SSs and downstream 5′SSs across an exon (Robberson et al., 1990). The first step of pre-mRNA splicing is the formation of the E complex, in which U1 snRNA forms a short duplex with the 5′SS, with the U2AF65 protein binding to the polypyrimidine tract at the 3′SS (Reed, 1996, Wahl et al., 2009). The exon definition model relies on the crosstalk between 5′SS and 3′SS through protein-protein interaction, and the initial evidence was obtained from the binding of U1 snRNp at both 5′- and 3′SSs (Ruby and Abelson, 1988). Indeed, the binding of U1 snRNp to the 5′SS of an internal exon enhances upstream 3′SS recognition (Robberson et al., 1990), and a mutation in the 5′SS enhances or decreases U1 snRNA binding to a degree that is sufficient to cause the inclusion or exclusion of the entire exon (Kuo et al., 1991). U1 snRNp-promoted 3′SS recognition and usage is at least partially explained by its ability to target U2AF65 to the upstream 3′SS (Hoffman and Grabowski, 1992). However, the role of exon definition in the selection of authentic splice site is not yet well understood.

The recognition of the 5′SS in a pre-mRNA is a critical event in an alternative splicing decision and is initiated by the formation of a base-pairing interaction between the splice site sequence and particular sequences in the U1 snRNA of the spliceosome (Roca et al., 2013; Wahl et al., 2009). In general, greater sequence complementarity leads to U1 snRNp binding with a higher affinity and thus defines a stronger splice site (Roca et al., 2013). However, the compilation of 5′SS sequences in human has revealed more than 9000 sequence variants in the −3 to

http://dx.doi.org/10.1016/j.gene.2017.03.023
0378-1119/© 2017 Published by Elsevier B.V.
and antisense oligonucleotides were mixed for annealing:denaturation frameshift and consequent nonsense-mediated mRNA decay, some oligonucleotides containing full or truncated Saa1 and Sacl restriction sites (in pZWB8-SMN1 series) or XhoI and EcoRI restriction sites (in exon test reporter) were synthesized (Shanghai GeneChem Co., Ltd.) and diluted in linker buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 1 mM EDTA) to a final concentration of 100 nM. To avoid splicing-induced frameshift and consequent nonsense-mediated mRNA decay, some oligonucleotides contained truncated restriction sites. Each pair of sense and antisense oligonucleotides were mixed for annealing;denaturation at 95 °C for 2 min and annealing at 52 °C for 10 min and then chilled on ice. The product was diluted to a concentration of 10 nM and then ligated into the digested Original Reporter 1 or 2.

To construct the 1.10P (Del3) 3′ insertion and 1.10P (Del3) 5′ insertion plasmids, the inserted sequences were amplified from HeLa genome DNA using RNP-F/RNP-R and RPN-F/RPN-R primers (see Supplementary data). The 151-nt amplification originated from SMN1 intron 7, which is not included in the minigene sequence used. The fragments were then digested with SalI and SacI and inserted into the enzyme-treated SMN1 reporter.

2. Materials and methods

2.1. Plasmid constructs and mutagenesis

The minigenes SMN1 and SMN2 containing exon 7 (54 bp) and portions of the adjacent introns 6 and 7 from PCI-SMN1 and PCI-SMN2 (gifts from Dr. J. Zhou, Arizona State University) were cloned into the GPP-coding sequence of pZWB (a gift from Dr. C. B. Burge, Massachusetts Institute of Technology) to replace the original SIRT1 minigene and thus yield SMN1 and SMN2 reporters.

The two constructs were used as the backbone of the test system. To insert cis-elements at different positions in the intron, two restriction enzyme sites, namely, Sali and SacI, were cloned at different positions of the minigene introns (see Fig. 4A, lower) using the overlap PCR strategy (primer sequences are listed in Supplementary data S4). In total, 12 working constructs were obtained in both the SMN1 and SMN2 reporter backgrounds. Based on the position of the insertion and the backbone reporter, the resulting constructs are named 1.1-1.13 and 2.1-2.13, with the former series referring to the SMN1 backbone and the latter referring to the SMN2 backbone. Insertions at position 8 led to splice failure, likely due to an interruption of the original splice signal; thus, constructs 1.8 and 2.8 were not included in this study. To exclude the possibility that the insertion interrupted the original cis-elements or created a new enhancer motif, non-functional sequences (5′-GGCG) (Fairbrother et al., 2002) were inserted at each chosen position to construct control plasmids. All of the control constructs showed a splicing pattern that was similar to that of the parental construct (Supplementary Fig. S1).

To insert the test sequence into the exon sequence, two restriction enzyme sites, namely, XhoI and EcoRI, were cloned into exon 7 of original reporter 1 to replace the Tra2β-ESE sequence (5′-AAAGAAGGAA GCCUCUCCUUCCUTT; PTB1-HOMO-1618 siRNA, 5′-CCGUGUGUAA GGAUCAATT; and PTB1-HOMO-1279 siRNA, 5′-CCGUAAGAUCCUG UCAATT). Among these, the most efficient was PTB1-HOMO-1234, and this siRNA was thus selected for most of the experiments. siRNA transfection was performed following the standard protocol with Lipofectamine 2000, with 500 ng of plasmid per well. The cells were maintained for 24 h and then harvested for total RNA preparation or protein preparation. For PTB knockdown, several siRNAs were synthesized by Gene Pharma Co. (PTB1-HOMO-1234 siRNA, 5′-CCGUGUGUAA GGAUCAATT; and PTB1-HOMO-1279 siRNA, 5′-CCGUAAGAUCCUGU CAATT). Among these, the most efficient was PTB1-HOMO-1234, and this siRNA was thus selected for most of the experiments. siRNA transfection was performed following the standard protocol with Lipofectamine 2000 at a final concentration of 40 nM. The effect of PTB knockdown was assessed by western blotting and Q-PCR (Fig. 4).

2.2. Cell culture and plasmid/RNA transfection

HeLa (human cervical carcinoma) cells were maintained in Dulbecco’s modified Eagle medium supplemented with 10% foetal calf serum. The cells were plated in a 24-well plate at a concentration of 1 × 10⁵ cells per well and then transfected when they reached 60% confluence. Transfection was performed following the standard protocol with Lipofectamine 2000, with 500 ng of plasmid per well. The cells were maintained for 24 h and then harvested for total RNA preparation or protein preparation. For PTB knockdown, several siRNAs were synthesized by Gene Pharma Co. (PTB1-HOMO-1234 siRNA, 5′-CCGUGUGUAA GGAUCAATT; and PTB1-HOMO-1279 siRNA, 5′-CCGUAAGAUCCUGU CAATT). Among these, the most efficient was PTB1-HOMO-1234, and this siRNA was thus selected for most of the experiments. siRNA transfection was performed following the standard protocol with Lipofectamine 2000 at a final concentration of 40 nM. The effect of PTB knockdown was assessed by western blotting and Q-PCR (Fig. 4).

2.3. RNA preparation and RT-PCR analysis

The total RNA from the transfected cells was extracted using a one-step extraction method with TRIzol (Invitrogen). To assess the RNA quantity and quality, absorbance at 260 nm and 280 nm was measured using a microplate reader (Epoch, BioTek). For each sample, 500 ng of RNA was used as the template in a 10-μl cDNA transcription system. A standard reverse-transcription (RT) protocol supplied by Promega (utilizing M-MLV) was used with oligo (dt)18. This procedure was followed by PCR amplification and product separation on 2% agarose gels or Q-PCR amplification using an IQ5 (Bio-Rad). In all the PCR or Q-PCR systems, 1 μl of cDNA was used as the template. Gel images were captured using a GeneGenius Gel Imaging System (Syngene) and quantified with the associated GeneTools analysis software. All the results were confirmed at least three independent times.

2.4. Western analysis

All of the samples for RT-PCR were harvested at the same time. The samples were lysed using RIPA with a protease inhibitor cocktail at a concentration of 1:100 (Calbiochem, Cat. No. 539134). The samples were heated at 100 °C for 10 min and subjected to 12% SDS-PAGE. An anti-PTB antibody from ProteinTech Co., was used at 1:1000. All the results were confirmed at three independent times.

3. Results

3.1. An intronic cryptic 5′ SS was activated when it was cloned in close proximity to a competing 5′ SS of a cassette exon

To study the roles governing the selection of a cryptic splice site, we constructed a CMV promoter-driven plasmid containing an EGFP ORF...
that was interrupted by the SMN1/2 cassette exon, as previously reported (Liu et al., 2010). The splicing efficiency of the reporter gene transcripts in HeLa cells was assayed by RT-PCR 24 h after transfection. The cassette exon was included in the original SMN1 reporter, whereas it was excluded from the SMN2 reporter, as previously reported (Fig. 1B). A cryptic 5′SS plus the adjacent sequence of 42 nt located in intron 4 of the CT/CGRP gene was cloned at different positions of the intron downstream of the SMN1/2 cassette exon in our reporters (Lou et al., 1995). The sequence of the cryptic 5′SS forms canonical base pairs at the −3 to +5 positions with the +4 to +11 positions of U1 snRNA, indicating that the cryptic 5′SS represents a strong 5′SS. In contrast, the alternative 5′SS of the exon 7 of SMN1/2 was much weaker. Interestingly, the cryptic 5′SS is silent in human cells, and it has been identified as part of the intronic enhancer that increases the inclusion of upstream exon 4 (Lou et al., 1994, Lou et al., 1995). The enhancer sequence contains a strong polypyrimidine tract bound by PTB, and this binding is proposed to inhibit the usage of the cryptic splice site inside the enhancer sequence and to promote the usage of the upstream polyadenylation site (Lou et al., 1996).

To obtain proper sites for analysing the activation of the cryptic splice site, a non-functional sequence (10 nt) was inserted into 8 different positions in the upstream intron and 5 positions in the downstream intron of the SMN1/2 cassette exon. We found that the insertion at position 8 completely abolished SMN1/2 inclusion (data not shown), likely due to its interruption of the branch point function. All of the other 12 insertion sites were suitable for intronic insertion (Supplementary Fig. S1). A series of splicing enhancer sequences for SR proteins Tra2β, SRSF5, SRSF2, SRSF7, and SRSF6 (Goren et al., 2006) were cloned and inserted at these positions within the SMN2 splicing context, and most showed no obvious enhancing effect (Supplementary Fig. S2), with a few exceptions. In fact, Tra2β and SRP55 demonstrated an enhancing function at the −84-nt intronic position upstream the SMN2 exon; whereas the SRP40 enhancer sequence at the +24-nt intronic position downstream the SMN2 exon resulted in near-complete exon inclusion. These results indicate that the enhancer sequence in intronic regions close to an alternative exon could be recognized and bound by SR proteins to promote exon inclusion.

We then cloned the 42-nt insert containing the cryptic 5′SS at the 12 suitable intronic positions (Fig. 1A). The cryptic sites containing insertions at positions 9, 10 and 11 of the SMN1 splicing reporter, which correspond to intronic insertion locations of 24 nt, 64 nt and 124 nt, resulted in a spliced product that was larger than that obtained with the inclusion of SMN1, and the spliced product varied in lengths corresponding to activation of the cryptic 5′SS (Fig. 1B, upper panel). The RT-PCR products were sequenced, and the results showed that the new spliced products originated from the selection of the cryptic 5′SS. The sequencing result of the spliced product at position 10 is shown in Supplementary Fig. S3. The splicing pattern of the reporter pre-mRNA

![Fig. 1](image_url)

**Fig. 1.** Activation of an intronic cryptic 5′ splice site. (A) Diagram of the 42-nt insert sequence components and its location in the CT/CGRP gene (upper); diagram of the splice site sequence in the reporter plasmid containing the SMN1/2 cassette exon (middle); diagram of the insertion positions of the 42-nt insert (Lower). (B) RT-PCR analysis of the spliced products from the constructs containing the PTB binding site and cryptic splice site shown as in diagram A. The spliced products with different exon compositions are diagrammed on the right. ‘In’ indicates the product containing the SMN1/2 cassette exon. ‘Ex’ indicates the skipping of this cassette exon. ‘*’ indicates the extended cassette exon resulting from the selection of the cryptic splice site.
resulted from the 42-nt insertion at all of the other intronic sites was similar to that of the background SMN1 reporter (Fig. 1B, upper panel). Interestingly, the activation of the cryptic splice site was also evident for the insertion at locations of 24 nt, 64 nt and 124 nt within the SMN2 splicing context (Fig. 1B, lower panel).

In contrast to its activation when being placed at the close proximity of the alternative 5′SS, the cryptic 5′SS close to the upstream constitutive 5′SS was silent (Fig. 1, positions 1–3). This result could be explained either by the constitutive 5′SS possessing a highly conserved splicing signal sequence of CAGGUA or a slightly stronger affinity for U1 snRNA than the cryptic 5′SS does. Alternatively, the splice site selection in long terminal exons could be different from those in short internal exons.

The activation of the cryptic 5′SS as an intron-proximal splice site has demonstrated some interesting features. First, the extended inclusion of SMN1 resulted from the use of cryptic 5′SS was the exclusive spliced product when the cryptic 5′SS was located at positions 24 nt and 64 nt (Fig. 1B, upper panel), consistent with its much higher strength in base pairing the U1 snRNA than the authentic 5′SS of exon 7 in SMN. As the SMN exon is 54 nt in length, the choice of the inserted cryptic 5′SS from intronic positions 24 nt and 64 nt resulted in an extended exon of 100 nt and 140 nt (taking the 17-nt sequence from the cryptic 5′SS) does. Alternately, the splice site selection in long terminal exons could be different from those in short internal exons.

The activation of the cryptic 5′SS as an intron-proximal splice site has demonstrated some interesting features. First, the extended inclusion of SMN1 resulted from the use of cryptic 5′SS was the exclusive spliced product when the cryptic 5′SS was located at positions 24 nt and 64 nt (Fig. 1B, upper panel), consistent with its much higher strength in base pairing the U1 snRNA than the authentic 5′SS of exon 7 in SMN. As the SMN exon is 54 nt in length, the choice of the inserted cryptic 5′SS from intronic positions 24 nt and 64 nt resulted in an extended exon of 100 nt and 140 nt (taking the 17-nt sequence from the cryptic 5′SS) does. Alternately, the splice site selection in long terminal exons could be different from those in short internal exons.

The above-described results are well consistent with the exon length-dependent model in which the distance of an intronic cryptic 5′SS splice site to the 3′SS across an exon plays a role in determining its selection efficiency. To further validate this model, we inserted a 151-nt sequence that originated from intron 7 of the SMN1 gene either upstream or downstream of the activated cryptic 5′SS in the 1.10-Δ3′ construct to generate two reporter plasmids denoted 1.10-Δ3′-5′ ins and 1.10-Δ3′-3′ ins (Fig. 3A). In the 1.10-Δ3′-5′ ins plasmid, the distances from the cryptic 5′SS to the upstream exon and to the across-exon 3′SS were increased to 215 nt and 293 nt, respectively, whereas the distance to the downstream exon was not changed. In contrast, in the 3′ insertion plasmid, the distance from the cryptic 5′SS to the upstream exon was not changed, whereas the distance to the downstream exon was increased by 151 nt. Fig. 3B demonstrates that cryptic 5′SS activation was abolished when the internal exon length was increased to

3.2. Activation depends on the splice site strength rather than the adjacent sequence context

To study whether the cryptic 5′SS activation depends on the sequence context of the 42-nt insert, we constructed a set of reporter plasmids with inserted deletions and/or mutations at position +64 nt (the 10th insertion site) of the SMN1 cassette exon (Fig. 2A). Fig. 2B shows that the deletion of the 5′ PTB-binding site (1.10-Δ5′), the 3′ part of the enhancer sequence (1.10-Δ3′), or both (1.10-Δ5′Δ3′) did not inhibit the activation of the cryptic 5′SS, suggesting that activation does not depend on the enhancer sequence context. Consistently, mutation of the PTB-binding sites did not alter the activation activity (1.10-m5′ and 1.10-m3′ Δ3′).

Consistent with the model that the activation of cryptic 5′SSs depends on the strength of the splice site itself, mutation of the critical CAG 5′SS consensus to CUG within the context of the full-length 42-nt insertion (1.10-mSS) resulted in a substantial decrease in the usage of the cryptic 5′SS and an increase in the usage of the upstream competing 5′SS (Fig. 2B). A similar shift in 5′SS usage was also evident when a CAG-to-CUG mutation was created in the 1.10-del3′ background (1.10-Δ3′ mSS). Furthermore, the CAG-to-CUG mutation in the 1.10-Δ5′Δ3′ background 1.10-Δ5′Δ3′mSS decreased the usage of the cryptic 5′SS but did not result in evident usage of the upstream competing 5′SS. Instead, we observed an increase in the usage of the constitutive 5′SS, which resulted in a skipped exon.

3.3. Activation depends on the proximity of the splice site to the upstream alternative exon

The above-described results are well consistent with the exon length-dependent model in which the distance of an intronic cryptic 5′SS splice site to the 3′SS across an exon plays a role in determining its selection efficiency. To further validate this model, we inserted a 151-nt sequence that originated from intron 7 of the SMN1 gene either upstream or downstream of the activated cryptic 5′SS in the 1.10-Δ3′ construct to generate two reporter plasmids denoted 1.10-Δ3′-5′ ins and 1.10-Δ3′-3′ ins (Fig. 3A). In the 1.10-Δ3′-5′ ins plasmid, the distances from the cryptic 5′SS to the upstream exon and to the across-exon 3′SS were increased to 215 nt and 293 nt, respectively, whereas the distance to the downstream exon was not changed. In contrast, in the 3′ insertion plasmid, the distance from the cryptic 5′SS to the upstream exon was not changed, whereas the distance to the downstream exon was increased by 151 nt. Fig. 3B demonstrates that cryptic 5′SS activation was abolished when the internal exon length was increased to
293 nt, whereas a change in the length of the downstream intron sequence had no effect on the activation of the cryptic 5′ SS. These results further support a negative correlation between the length of the spliced internal exon and the selection of cryptic 5′ SSs.

3.4. PTB is not required for the cryptic 5′ SS activation

The PTB-binding site in the 42-nt insert from intron 4 of the CT/CGRP gene aids in the binding of U1 snRNP to the cryptic 5′ SS, which is essential for the activity of this intronic enhancer (Lou et al., 1999). However, PTB is generally known as a splicing repressor, the binding of which correlates with the negative regulation of the recruitment of spliceosomal components to their nearby splice sites (Wagner and Garcia-Blanco, 2001). Our recent recognition of the enhancer function of PTB is attributed to its binding to the polypyrimidine tracts near the constitutive splice site, which allows PTB to modulate the competitiveness of alternative and constitutive splice sites (Xue et al., 2009). A recent study showed that PTB binds to a pyrimidine-rich sequence between a pair of alternative 5′ SSs promotes usage of the upstream 5′ SS (Hamid and Makheyev, 2014). These previous studies prompted us to further analyse the role of PTB in the activation of cryptic 5′ SSs in this study.

The results presented in Fig. 2 show that the PTB-binding site is not required for the activation of cryptic 5′ SSs. To address whether PTB proteins have any function in splicing activation by binding to some other intronic elements in the constructs used, HeLa cells were pre-treated with siRNA for 24 h to downregulate the expression level of the PTB protein, and we then transfected different reporter plasmids to assay the choice of 5′ SS (Fig. 4A). Western blot and RT-PCR analyses showed that PTB expression was successfully downregulated (Fig. 4A). The splicing of two endogenous genes, namely, EI4G2 and RBM27, which has been proven to be regulated by PTB (Xue et al., 2009), was analysed and shown to respond to PTB knockdown (Fig. 4B). Nevertheless, the splicing patterns of all of the tested reporter mRNAs with or without the cryptic 5′ SS did not respond to PTB knockdown (Fig. 4A and C). Taken together, the results obtained in this study suggest that PTB does not play a role in regulating the alternative splicing of the two competitive 5′ SSs investigated, as will be discussed below.

3.5. The intronic cryptic 5′ SS is activated in the cassette exon

The results obtained in this study generally obey the previously reported proximity rule, whereby the intron-proximal 5′ SS is preferentially selected (Hicks et al., 2010, Reed and Maniatis, 1986). Nevertheless, there are substantial differences. According to our results, the distance between the two competing 5′ SSs could not reach a distance of 200–300 nt, as previously reported. Indeed, the closer the two splice sites, the stronger the selection of the intron-proximal splice site (Fig. 1).

To further investigate the intron-proximal rule, we inserted cryptic 5′ SS-containing sequences into the reporter SMN1 exon to replace the original Tra2β/SE, allowing the construction of new plasmids, with one harbouring the 3′-deleted cryptic 5′ SS (Ex-Δ3′) and the other harbouring a splice site truncation (Ex-Δ3′ΔSS) (Fig. 5A). Strikingly, the cryptic 5′ SS was exclusively activated during splicing after transfection of the plasmid containing an intact splice site into HeLa cells, whereas transfection of the plasmid containing the splice site truncation abolished this activation (Fig. 5B). The spliced products were confirmed by sequencing, and mutation of the CAG consensus of the cryptic splice site to CUG or CAC completely abolished the activation of the cryptic 5′ SS (Fig. 5C). These results argue against the intron-proximal rule, indicating that the application of this rule could be restricted. For example, the rule may be more applicable to the splicing of terminal exons (Hicks et al., 2010, Yu et al., 2008), but not the internal cassette exon used in this study.

![Fig. 4. Effect of PTB expression on the activation of cryptic splice sites. (A) Western (left lower) and Q-PCR (right) results obtained after PTB knockdown. RT-PCR results of the spliced product of the 1.10-Δ3′ construct without and with PTB knockdown (left upper). (B) Response of PTB-sensitive genes EI4G2 and RBM27 to PTB knockdown. (C) RT-PCR results showing the selection of the cryptic splice site in different constructs with or without PTB knockdown.](image)

4. Discussion

In mammals, the precise selection of a splice site is a challenging procedure under highly organized regulation. In fact, the activation of an inappropriate splice site can trigger severe disease (Cooper et al., 2009, Muntoni and Wood, 2011, Roca et al., 2008). A large number of cryptic splice sites with no sequence deficiency are repressed, but the underlying mechanisms still remain obscure. It has been recently reported that various cis-acting splicing regulatory elements and noncanonical uridine-pseudouridine interaction in the 5′ SS/U1 helix can contribute to the recognition of some 5′ splice sites (Brillen et al., 2016, Kralovicova et al., 2011, Tan et al., 2016). Consistently, the recent study on structure also demonstrates additional contacts for the U1 recognition of 5′ splice sites in addition to the known base pairing (Kondo et al., 2015).

We have shown that a naturally unused cryptic 5′ SS can be activated when it is placed in an intron close to the upstream cassette exon. The
cryptic 5'SS used in this study is a well-characterized enhancer sequence involved in PTB binding and the activation of the upstream 5'SS and polyadenylation site (Lou et al., 1996, Lou et al., 1999, Lou et al., 1998, Lou et al., 1995). Interestingly, activation of the cryptic 5'SS strongly depends on the consensus sequence of the splice site but not on the adjacent sequences that bind PTB or other factors (Fig. 2).

In addition, knockdown of the PTB level does not affect the activation of the cryptic 5'SS (Fig. 4). We showed that shortening the exon sequence from 54 nt to 24 nt and increasing the exon length from 54 nt to 118 nt using a pure intronic sequence have no effect on the exclusive activation of a cryptic 5'SS in reporter plasmids when they are expressed in HeLa cells. However, an increase in the distance beyond 200 nt (by different sequences) completely abolishes its activation (Figs. 1 and 3). These results suggest that, in addition to the well-recognized roles of the exon sequence in splice site selection, exon length itself might function as a determinant as well.

Recent studies have linked nucleosome structure and modifications to alternative splicing control in metazoans (Keren et al., 2010, Schwartz et al., 2009). In human, nucleosome occupancy at different regions of a gene are correlated with the level of inclusion of these regions in mature mRNAs: constitutive exons have the highest occupancy, followed by alternative exons that are both frequently and rarely used; introns have the lowest nucleosome occupancy (Schwartz et al., 2009, Tilgner et al., 2009). The nucleosome could act as a fluctuating barrier that pauses RNAPII, a process termed “speed bump”, which may create a time window for splice site recognition and selection (Hodges et al., 2009, Keren et al., 2010). Exonic nucleosomes have a specific set of histone modifications that lead to interactions with the splicing machinery and enable more efficient recognition of the exon. In both Caenorhabditis elegans and humans, exons are preferentially marked by histone 3 lysine 36 trimethylation (H3K36me3) (Kolasinska-Zwierz et al., 2009, Schwartz et al., 2009, Tilgner et al., 2009). This H3K36me3 modification could be bound by a multifunctional protein that also binds splicing factors to impact splice site recognition and alternative splicing (Luco et al., 2010).

The link between nucleosome structure and alternative splicing could represent mechanisms for the exon definition model. The requirement of the distance of a strong cryptic 5'SS to its upstream 3'SS to be within a nucleosome shown in this study is consistent with both the exon definition model and the nucleosome impact of pre-mRNA splicing in metazoan. Nevertheless, we cannot rule out the possibility that the distance-related regulation of the cryptic 5'SS selection is caused by the presence of unknown splicing regulatory sequences or structures residing in the adjacent sequence.

Funding
This work was supported by a grant from the China 973 Program (2015CB554100) awarded to K.D. and grants from the China 973 Program (2005CB724604) and ABL2015-AS001 awarded to Y.Z.

Conflicts of interest
The authors have no conflicts of interest to declare.

Acknowledgements
We thank Dr. Chris Burge (MIT) for sending us the PCI-SMN1 plasmid as a gift; Dr. Tao Sun (Wuhan University) for construction of the PCI-SMN1 backbone plasmid; and Dr. Qian Li (Wuhan University) for assisting with the transfection experiments.

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2017.03.023.

References
Hoffman, R.E., Grabowski, P.J., 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3’ splice site by a network of interactions spanning the exon. Genes Dev. 6, 2554–2568.