

Succession of splicing regulatory elements determines cryptic 5'ss functionality

Anna-Lena Brillen¹, Katrin Schöneweis², Lara Walotka¹, Linda Hartmann¹, Lisa Müller¹, Johannes Ptok¹, Wolfgang Kaisers³, Gereon Poschmann⁴, Kai Stühler^{4,5}, Emanuele Buratti⁶, Stephan Theiss^{7,*} and Heiner Schaal^{1,*}

¹Institute for Virology, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, ²Department of Infectious Diseases, Molecular Virology, University Hospital Heidelberg, 69120 Heidelberg, Germany, ³Department of Anesthesiology, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, ⁴Molecular Proteomics Laboratory, BMFZ, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, ⁵Institute for Molecular Medicine, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, ⁶International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, 34149 Trieste, Italy and ⁷Institute of Clinical Neuroscience and Medical Psychology, Heinrich-Heine-University Düsseldorf, 40225

Received July 26, 2016; Revised December 7, 2016; Editorial Decision December 18, 2016; Accepted December 19, 2016

ABSTRACT

A critical step in exon definition is the recognition of a proper splice donor (5'ss) by the 5' end of U1 snRNA. In the selection of appropriate 5'ss, cis-acting splicing regulatory elements (SREs) are indispensable. As a model for 5'ss recognition, we investigated cryptic 5'ss selection within the human fibrinogen B β -chain gene (FGB) exon 7, where we identified several exonic SREs that simultaneously acted on up- and downstream cryptic 5'ss. In the FGB exon 7 model system, 5'ss selection iteratively proceeded along an alternating sequence of U1 snRNA binding sites and interleaved SREs which in principle supported different 3' exon ends. Like in a relay race, SREs either suppressed a potential 5'ss and passed the splicing baton on or splicing actually occurred. From RNA-Seq data, we systematically selected 19 genes containing exons with silent U1 snRNA binding sites competing with nearby highly used 5'ss. Extensive SRE analysis by different algorithms found authentic 5'ss significantly more supported by SREs than silent U1 snRNA binding sites, indicating that our concept may permit generalization to a model for 5'ss selection and 3' exon end definition.

INTRODUCTION

Alternative 5' splice site selection is a highly regulated process involving degenerate sequence elements that are recog-

nized by a large intricate protein complex, the spliceosome, which is composed of five small nuclear ribonucleoprotein particles (snRNPs). Spliceosome assembly starts with the interaction of the U1 snRNP and the 5'ss at the exon–intron border. Since within vertebrates, relatively small exons are separated by much longer introns, splice site pairing is supposed to first occur across the exon through subsequent binding of the U2 snRNP to the branch point sequence of the upstream 3' splice site (3'ss or splice acceptor) (1,2). Both snRNPs interact with each other forming the 'exon-definition' complex (2), which is later converted into 'intron-definition' complexes (3,4), connecting U1 and U2 snRNPs across the intron and triggering the splicing reaction.

Splice donor choice, however, is not only directed by the spliceosome itself, recognizing the 11 nt long sequence of the 5'ss to form an RNA duplex with the 5' end of U1 snRNA (3,4), but also critically depends on RNA binding proteins that bind to splicing regulatory elements (SREs) in the vicinity of splice sites, like SR (serine-arginine-rich) or hnRNP (heterogeneous nuclear ribonucleoprotein) proteins (5). SR proteins are composed of one or two RNA binding domains (RRM) and an arginine-serine (RS)-rich domain that participate directly in the interaction with other proteins or with RNA itself. Both domains have been shown to be capable of participating in U1 snRNP recruitment to the 5'ss via the U1-specific protein U1-70K (6–9). Dependent on their binding position to the exon or intron, SR proteins can generally act in a position-dependent manner, either activating or silencing splice donor usage (10,11). Generally, SR proteins enhance 5'ss use, when they bind to the upstream exon, while they repress splicing from

*To whom correspondence should be addressed. Tel: +49 211 81 12393; Fax: +49 211 81 10856; Email: schaal@uni-duesseldorf.de
Correspondence may also be addressed to Stephan Theiss. Email: theiss@uni-duesseldorf.de

the downstream intron. During repression, proteins bound to inhibitory SREs interfere with further progression into late spliceosomal complexes and form so-called ‘dead-end’ complexes (10,12,13).

In human genetics, the computational identification of aberrant splice donor usage due to nucleotide exchanges is vitally needed for diagnostics, and evaluating a mutation’s biological relevance for clinical treatment of patients with hereditary disorders is indispensable (14,15).

By now, variations in *cis*- or *trans*-acting elements within protein coding genes have been associated with altering splicing patterns and thereby inducing genetic defects that cause human diseases (16). Estimates of the fraction of human inherited disease mutations that affect splicing range from 10% for mutations located directly within splice sites (17) and can even reach 22–25% if mutations within SREs were considered (18,19). Thus, roughly 1/3 of all nucleotide mutations leading to human disease result in exon skipping, use of cryptic splice sites or intron retention, leaving aside SREs that have not been discovered yet. However, since cryptic sites are splicing inactive as long as the authentic 5’ss is functional, it seems that splice site choice simply follows a ‘winner-takes-it-all’ rule. If the authentic 5’ss is weakened, however, it is generally unclear whether exon skipping or cryptic splicing occurs.

Although highly desirable, there is no single *in silico* tool available yet, providing reliable predictions of splice site usage. Algorithms like MaxEnt (20) and HBond (3,4) for 5’ss scoring, as well as e.g. Δ tESRseq (15) or HEXplorer-based (21) approaches calculating enhancing or silencing properties of regions in the vicinity of splice sites, greatly assist in this daunting task.

In this work we show that a tight cluster of alternating multiple SREs and U1 snRNA binding sites controls cryptic splice donor usage throughout the human fibrinogen B β -chain gene (FGB) exon 7. Based on HEXplorer profiles, we predicted several SREs that we confirmed by mutational analyses. Motifs identified in these *cis*-acting SREs exhibited some degeneracy with respect to the binding splicing regulatory proteins SRSF1 and Tra2 β , indicating a possible redundancy. Splicing regulatory proteins bound to these SREs acted in a strictly position-dependent manner, each functioning as a gateway that either terminated the exon or passed on an ‘exon end signal’ to the next U1 snRNA binding site.

MATERIALS AND METHODS

Single-intron splicing constructs

Constructs SV guanosine-adenosine-rich (GAR) SD4 Δ vpu env eGFP D36G, SV GAR[−] SD4 Δ vpu env eGFP D36G, SV GAR[−]ESE[−] SD4 Δ vpu env eGFP D36G are based on the HIV-1 glycoprotein/eGFP expression plasmid and have been described before (3,22). Inserting the neutral sequence CCAAACAA (23) was carried out by replacing GAR with a polymerase chain reaction (PCR) product obtained with primer pair #3378/#3379. All FGB exon 7-derived fragments were inserted into SV GAR SD4 Δ vpu env eGFP D36G, replacing the GAR element with DNA fragments obtained with primer

pairs #3168/#3169 (FGB7-A), #3166/#3167 (FGB7-B), #3170/#3171 (FGB7-C), #3172/#3173 (FGB7-D), #3326/#3327 (FGB7-E) and #3174/#3175 (FGB7-F), respectively. SV FGB7-D(8A) SD4 Δ vpu env eGFP D36G, SV FGB7-D(5C) SD4 Δ vpu env eGFP D36G and SV FGB7-D(5C+8A) SD4 Δ vpu env eGFP D36G were constructed by replacing GAR with PCR products resulting from primer pair #3479/#3480, #3481/#3482 and #3483/#3484, respectively.

Fibrinogen B β minigenes

The plasmids pT-B β -WT and pT-B β -IVS7+1G>T were previously described (24,25). pT-B β -IVS7+1G>A and pT-B β -IVS7+2T>A were cloned via mutagenesis PCR of pT-B β -WT using the primer pairs #5659/#5660, and #5661/#5660, respectively. pT-B β -IVS7+1G>T-mt-c1 was cloned via a mutagenesis PCR of pT-B β -IVS7+1G>T with primers #2619/#2622 and #2620/#2621; pT-B β -IVS7+1G>T-mt-c1/c2* #2619/#2647 and #2620/#2646; pT-B β -IVS7+1G>T-mt-c1/c2*/c3 with primers #2619/#2624 and #2620/#2623; pT-B β -WT-c1-15.8 was cloned via a mutagenesis PCR of pT-B β -WT using primers #2619/#2765 and #2620/#2764. pT-B β -WT-c1-18.8 was cloned via a mutagenesis PCR of pT-B β -WT-c1-15.8 using primers #2619/#2872 and #2620/#2871; pT-B β -WT-c1-20.8 using primers #2619/#2874 and #2620/#2873. pT-B β -WT-c3-15.8 was cloned via a mutagenesis PCR of pT-B β -WT with primers #2619/#2925 and #2620/#2924; pT-B β -WT-c3-18.8 with primers #2619/#2927 and #2620/#2926; pT-B β -WT-c3-20.8 with primers #2619/#2929 and #2620/#2928. HEXplorer-guided mutations of fragments B-D were inserted via mutagenesis PCR of pT-B β -WT or –IVS, respectively, with primers #5568/#2620 and #2619/#5569 (B), #5566/#2620 and #2619/#5567 (C), #3548/#2620 and #3549/#2619 (D), #5571/#2620 and #2619/#5569 (B/C), #5568/#2620 and #2619/#5569 (B/D), #5570/#2620 and #2619/#5567 (C/D), #5571/#2620 and #2619/#5569 (B/C/D). Exon 7 was replaced with only splicing neutral sequences (25) by using a customized synthetic gene from Invitrogen and inserted into pT-B β -IVS7+1G>T via EcoNI/Bpu10I. FGB7-derived fragments were inserted with PCR products resulting from primer pairs #4835/2620 (B), #5179/2620 (B MUT), #5581/#2620 (C), #5585/#2620 (C MUT), #4703/#2620 (D) and #4791/#2620 (D MUT). Fragments derived from the E1 α PDH gene were inserted with PCR products resulting from primer pairs #5497/#5498 (WT) and #5499/5500 (MUT) and fragments derived from the SNAPC4 gene (ENSG00000165684) with primer pairs #5498/#5490 (WT) and #5491/#5492 (MUT).

Expression plasmids

pXGH5 (26) was cotransfected to monitor transfection efficiency.

Oligonucleotides

All oligonucleotides used were obtained from Metabion GmbH (Planegg, Germany) (see Supplementary Table S1).

Cell culture and RT-PCR analysis

HeLa cells were cultivated in Dulbecco's high-glucose modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf serum and 50 µg/ml penicillin and streptomycin each (Invitrogen). Transient-transfection experiments were performed with six-well plates at 2.5×10^5 cells per well by using TransIT[®]-LT1 transfection reagent (Mirus Bio LLC US) according to the manufacturer's instructions. Total RNA was isolated 24 h post-transfection by using acid guanidinium thiocyanate-phenol-chloroform as described previously (27). For (q)RT-PCR analyses, RNA was reversely transcribed by using Superscript III Reverse Transcriptase (Invitrogen) and Oligo(dT) primer (Invitrogen). For the analyses of the single-intron splicing constructs primer pair #3210/#3211 was used; for the analyses of the Fibrinogen B β minigenes, primer pair #2648/#2649 was used. Quantitative RT-PCR analysis was performed by using the qPCR MasterMix (PrimerDesign Ltd) and LightCycler 1.5 (Roche). For normalization, primers #1224/#1225 were used and the level of hGH present in each sample was monitored.

FACS analysis

Fluorescence-activated cell sorting (FACS) analysis for the measurement of quantitative eGFP expression was carried out using FACS Canto2 (BD Biosciences). First, cells were washed with PBS and incubated with trypsin for 5 min. After several washing steps (PBS + 3% FCS), samples were acquired to the cytometer. Next, data was edited using the FlowJo analysis software (Tree Star, Inc.).

Inhibition of translation by cycloheximide

In order to detect nonsense-mediated decay (NMD)-sensitive transcripts, cells were incubated with 50 µg/ml of the translational inhibitor cycloheximide (CHX) 6 h prior to harvesting. As control for CHX treatment, we amplified RNA encoding SRSF3 (SRP20) with specific primers binding within exon 1 and 5 (#4003/#4004). WT SRSF3-messages exclude the poison cassette exon 4, while transcripts including exon 4 contain a pre-mature stop codon and get degraded by NMD (28).

Protein isolation by RNA affinity chromatography

Three thousand picomoles of short RNA oligonucleotides for either wild-type (WT) or mutant version of FGB7-derived fragments B (#5170; #5173), C (#5572; #5573) and D (#5167; #5168) or the neutral sequence (#5169), respectively, were covalently coupled to adipic acid dihydrazide-agarose beads (Sigma). 60% of HeLa nuclear extract (Cilbiotech) was added to the immobilized RNAs. After stringent washing with buffer D containing different concentrations of KCl (20mM HEPES-KOH [pH 7.9], 5%[vol/vol] glycerol, 0.1-0.5 M KCl, 0.2 M ethylenediaminetetraacetic acid, 0.5 mM dithiothreitol, 0.4M MgCl₂), precipitated proteins were eluted in protein sample buffer. Samples were heated up to 95°C for 10 min and loaded onto sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for western blot analysis. Samples were transferred

to a nitrocellulose membrane probed with primary and secondary antibodies (SRSF1 (Invitrogen 32-4500), Tra2 β (abcam ab31353), MS2 (Tetracore TC-7004-002)) and developed with ECL chemiluminescence reagent (GE Healthcare).

HEXplorer score calculation

HEXplorer score profiles of pairs of WT and mutant sequences were calculated using the web resource https://www.hhu.de/rna/html/hexplorer_score.php (21).

Mass spectrometric analysis

Protein samples were shortly separated over about 4 mm running distance in a 4–12% polyacrylamide gel. After silver staining, protein containing bands were excised and prepared for mass spectrometric analysis as described (29). Briefly, samples were destained, reduced with dithiothreitol, alkylated with iodoacetamide and digested with trypsin. Resulting peptides were extracted from the gel piece and finally resuspended in 0.1% trifluoroacetic acid.

Initially, peptides were separated by liquid chromatography on an Ultimate 3000 Rapid Separation Liquid Chromatography system (RSLC, Thermo Scientific, Dreieich, Germany). A trap column (Acclaim PepMap100, 3 µm C18 particle size, 100 Å pore size, 75 µm inner diameter, 2 cm length, Thermo Scientific, Dreieich, Germany) was used for peptide pre-concentration at a flow rate of 6 µl/min for ten minutes using 0.1% trifluoroacetic acid as mobile phase. Subsequently, peptides were separated on a 25 cm length analytical column (Acclaim PepMapRSLC, 2 µm C18 particle size, 100 Å pore size, 75 µm inner diameter, Thermo Scientific, Dreieich, Germany) at a flow rate of 300 nl/min at 60°C using a 2 h gradient from 4 to 40% solvent B (0.1% (v/v) formic acid, 84% (v/v) acetonitrile in water) in solvent A (0.1% (v/v) formic acid in water). Peptides were injected into the mass spectrometer by distal coated Silica Tip emitters (New Objective, Woburn, MA, USA) via a nano electrospray ionization source using a spray voltage of 1.4 kV.

Tandem mass spectra were recorded in a data dependent setting with an Orbitrap Elite (Thermo Scientific, Dreieich, Germany) hybrid mass spectrometer in positive mode. Full scans (resolution 60 000) were recorded over a scan range of 350–1700 m/z with a maximal ion time of 200 ms and the target value for automatic gain control set to 1000 000 in profile mode in the orbitrap part of the instrument. Subsequently, up to twenty precursors at charge states two and three were isolated (isolation window 2 m/z), fragmented by collision induced dissociation and analyzed with a maximal ion time of 50 ms and the target value for automatic gain control set to 3000 (available mass range 50–2000 m/z, resolution 5400) in the linear ion trap part of the instrument. Already analyzed precursors were excluded from further isolation and fragmentation for 45 s.

Data analysis within the MaxQuant environment (version 1.5.5.1, Max Planck Institute of Biochemistry, Planegg, Germany) was performed independently for the two replicate sample batches with standard parameters if not otherwise stated. Spectra were searched against 70 615 entries from the UniProt KB homo sapiens proteome

UP000005640 (downloaded on 16 June 2016) with label-free quantification enabled as well as the 'match between runs' option. Tryptic cleavage specificity was chosen, as well as carbamidomethyl at cysteines as fixed and methionine oxidation and acetylation at protein n-termini as variable modifications. For precursor masses, the mass tolerances were set to 20 ppm (first search) and 4.5 ppm (second search after recalibration) and for fragment masses to 0.5 Da. Peptides and proteins were accepted at a false discovery rate of 1% and only proteins considered showing two or more identified different peptides.

RESULTS

Cryptic splice site activation is mediated by SREs acting in a strictly position-dependent manner

To exemplify the complexity of aberrant splicing and the difficulty of predicting the splicing outcome caused by human pathogenic mutations, we revisited cryptic splice site usage embedded in a splicing-regulatory network of the human FGB. Here, the FGB c.1244+1G>T (aka IVS7+1G>T) minigene analysis revealed that beside exon 7 skipping this mutation caused the activation of three cryptic splice donors localized in the upstream exon: two canonical at 106 nt (c1) and 24 nt (c3), and one non-canonical 40 nt (c2*, indicated by the asterisk) upstream of the physiological 5'ss, leading to a loss of functional fibrinogen (24,25). Additionally, we observed activation of the intron localized cryptic site p1 158 nt downstream of +1G>T.

Calculating the HBond scores (HBS) of all 11 nt long GT sequences within exon 7 and the downstream intron ((6); https://www.hhu.de/rna/html/hbond_score.php) confirmed that the physiological 5'ss had the highest HBS, indicative of the highest complementarity to U1 snRNA, followed by activated cryptic splice sites in the neighborhood of the physiological 5'ss (Figure 1).

To analyze whether activation of the three exonic cryptic splice sites caused by the mutant physiological 5'ss is solely mediated by the previously identified naturally silent SRSF1 (aka SF2/ASF) binding site (25), we mutated the exonic cryptic sites one after the other. The impact of these mutations on the splicing pattern was analyzed by RT-PCR following transient transfection assays of WT or +1G>T in a three exon minigene (25). As shown in Figure 2A, following +1G>T mutation c1, c2*, c3 and p1 were activated. Surprisingly, individual inactivation of c1, however, seemed to neither effectively shift the overall splice site use toward the other cryptic sites nor change the level of exon skipping (Figure 2A, cf. lanes 2 and 3). This obvious lack of competition between these cryptic splice sites suggested that the other cryptic sites might be regulated independently of c1. In line with this, inactivation of both c1 and c2* (Figure 2A, lane 4), or all three exonic cryptic splice sites (Figure 2A, lane 5) caused much less exon skipping as expected, strengthening our hypothesis that at least one additional SRE might be located downstream of c1.

Next, we investigated whether the canonical cryptic sites c1 or c3 could outcompete the physiological WT 5'ss, if they were modified to have a similar U1 snRNA complementarity as the physiological 5'ss, i.e. similar HBS (HBS 15.0). Adapting c1 from HBS 12.2 to 15.8 (Figure 2B) did

not change splice site usage (Figure 2C, cf. lanes 1 and 2), supporting our hypothesis that at least another SRE was localized within exon 7, between c1 and WT 5'ss. At the same time, such an SRE would repress c1 usage and enhance any downstream splice site (10). Thus, likewise adapting c3 from HBS 10.8 to 15.8 can be expected to switch splice site use from the physiological 5'ss to the 24 nt more proximal cryptic site c3. Indeed, increasing c3 from HBS 10.8 to 15.8 fully activated this cryptic splice site even in the presence of the physiological 5'ss (Figure 2C, cf. lanes 1 and 3), thereby shortening the exon by 24 nt. Interestingly, exclusive use of c1 was not observed even when it was increased to HBS 18.8 or even 20.8 (Figure 2C, lanes 4 and 6), speaking again for at least a second SRE downstream of c1 which would simultaneously repress the nearest upstream splice site and enhance the nearest downstream splice site, thereby extending the exon.

Multiple exonic splicing enhancers are located within FGB exon 7

To examine the validity of this concept, we analyzed splice site recognition in more detail and searched for additional functional SREs. First, we experimentally determined the impact of the putative exonic splicing enhancers on splice site recognition. For this, we used our well-characterized enhancer-dependent single intron eGFP splicing reporter (3,30), permitting to measure eGFP fluorescence intensity proportional to U1 snRNP binding to the 5'ss (21). In this way, splice site recognition can not only be measured via (q)RT-PCR, but also be quantified in an independent experimental setup by flow cytometry. Furthermore, the leader sequence of this enhancer reporter can be substituted with any putative SRE. As reference for grading downstream enhancer impact we used the HIV-1 GAR splicing enhancer, containing two SRSF1- and one SRSF5-binding sites and its mutations GAR⁻ and GAR⁻ESE⁻ (3,22,30) (Supplementary Figure S1A). To confirm the GAR inactivating mutations, we additionally substituted the inactive GAR⁻ESE⁻ with a CCAAACAA repeat previously shown to be splicing neutral (23) and determined their impact on 5'ss recognition. In fact, using both qRT-PCR and flow cytometry, we could measure an up to 230-fold increase in splice site recognition mediated by the GAR element, an up to 7-fold increase mediated by the GAR⁻, but none for GAR⁻ESE⁻ (0.5-fold activation) (Supplementary Figure S1B). In summary, the enhancer reporter allows to functionally rank strong, intermediate and not enhancing properties of SREs in comparison to the splicing neutral reference sequence.

Next, we examined six FGB exon 7-derived fragments (named FGB7-A to FGB7-F), where FGB7-B corresponded to the naturally silent SRSF1 binding site (25), and inserted each fragment into this enhancer reporter. Here, B, C and D showed an increase in splice site recognition of more than 100 to even 1000 times compared to GAR⁻ESE⁻ (Figure 3A). Interestingly, fragment D, but not B showed the highest splicing enhancing activity and was further subjected to mutational analyses to identify the splicing regulatory protein binding to it. We used the HEXplorer algorithm (21) to predict the most promising

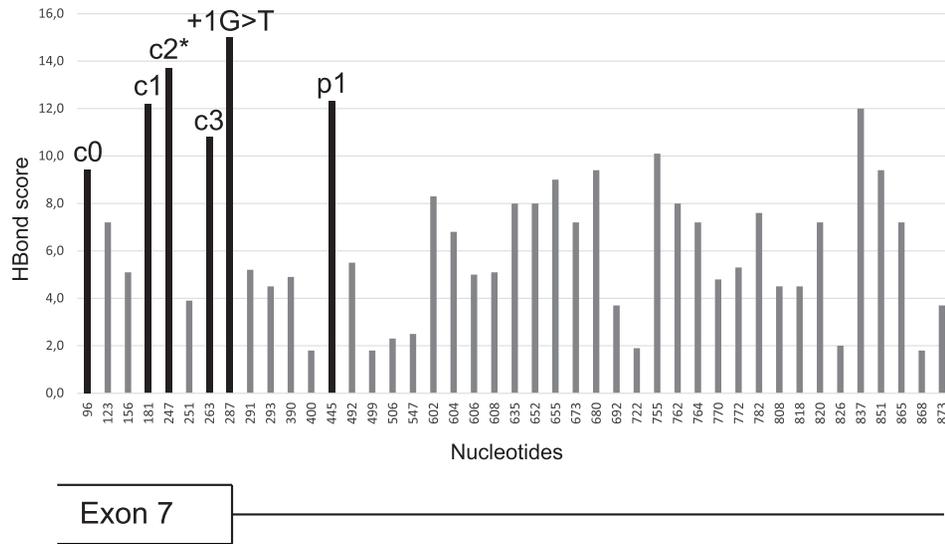


Figure 1. HBond score (HBS) profile of all 11-nt long GT sequences within the human fibrinogen B β -chain gene exon 7 and its downstream intron. Additionally, the GC-splice site c2* (with a substituted GT for HBS calculation) is considered and indicated by an asterisk. Numbers on the x-axis describe positions starting from the beginning of exon 7.

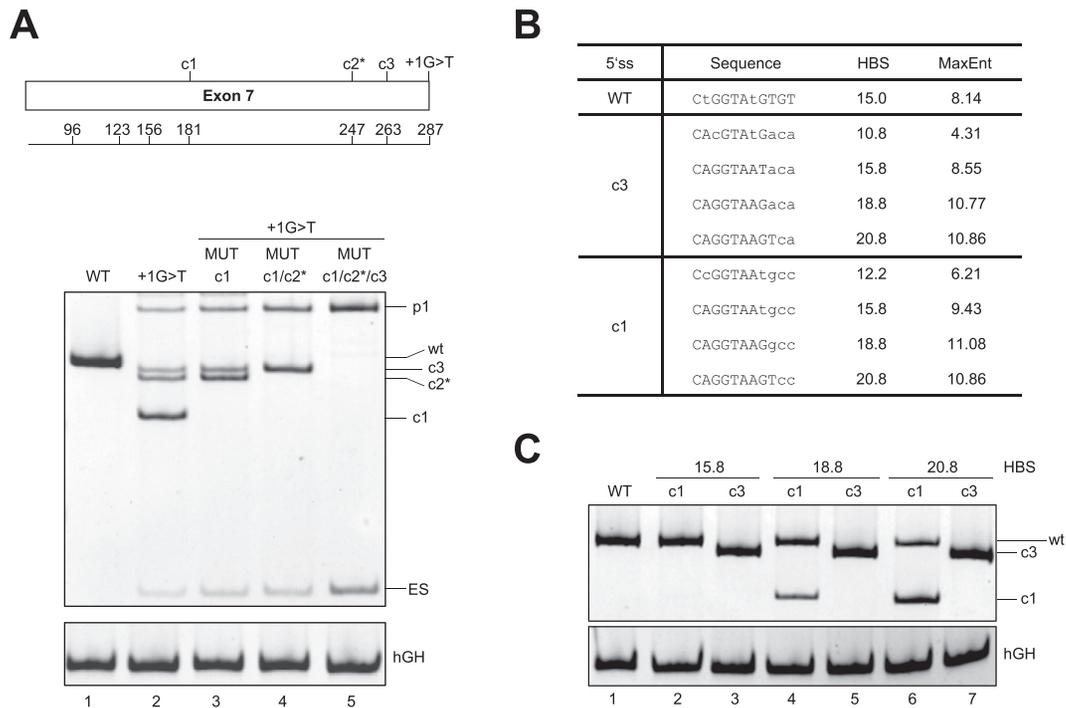


Figure 2. Cryptic splice donor activation. (A) Schematic drawing of the 287-nt long human fibrinogen B β -chain gene exon 7 and its cryptic splice sites (top). 2.5×10^5 HeLa cells were transiently transfected with 1 μ g of each construct together with 1 μ g of pXGH5 (hGH) to monitor transfection efficiency. Twenty-four hours after transfection, RNA was isolated and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR products were separated by 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide (bottom). Exonic (c1–c3) and intronic (p1) cryptic splice donor sites as well as the skipped exon (ES) are depicted on the right hand side. (B) Sequences of the wild-type 5'ss (WT), c1 and c3 variants including their different HBond/MaxEnt scores. (C) RT-PCR analysis of splicing patterns of cryptic splice sites mutated to have higher HBS according to (B).

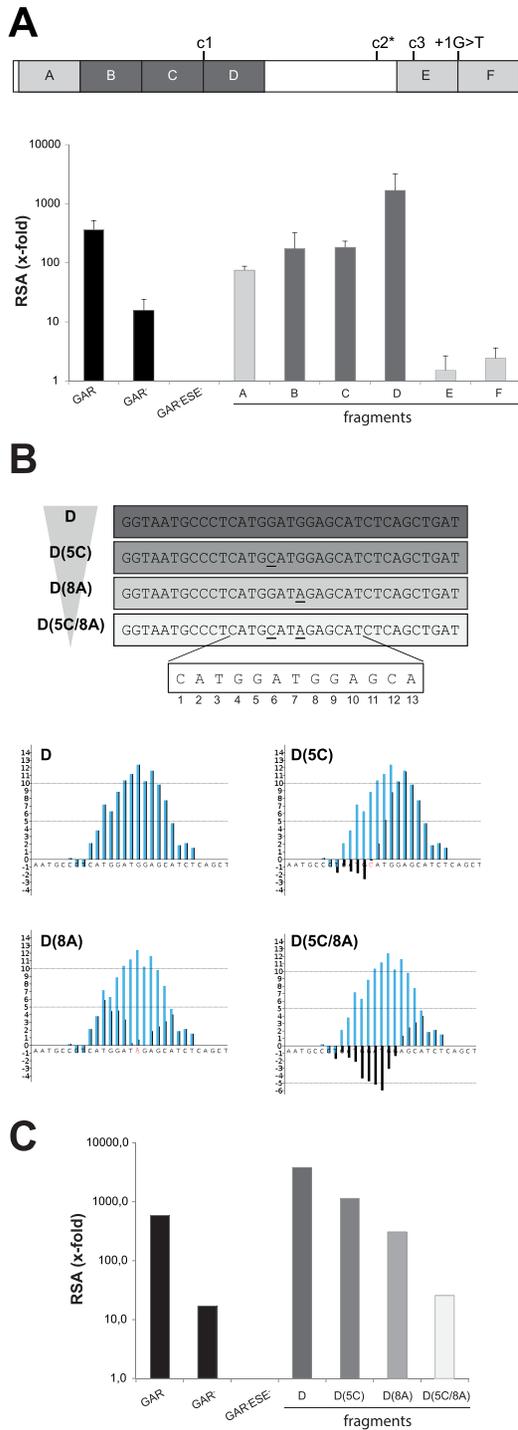


Figure 3. SREs within FGB exon 7. (A) Schematic of the localization of FGB exon 7 fragments used in SRE analysis. 2.5×10^5 HeLa cells were transiently transfected with 1 μ g of each construct and 1 μ g of pXGH5. At 24 h after transfection, total-RNA samples were collected and used for qRT-PCR with primer pair #3210/#3211 and normalized to hGH (#1224/#1225). Relative splicing activity (RSA). (B) Sequences (top) and HEXplorer profiles (bottom) of fragment D and its mutations. The WT profile is shown in blue and mutant profiles in black. (C) Real-time PCR of transcripts expressed from the enhancer reporter. cDNA samples were prepared as described for panel A and used in real-time PCR assays to specifically quantitate the relative abundances of spliced mRNA. Relative splicing activity (RSA).

inactivating mutations for fragment D. The HEXplorer is based on a RESCUE-type approach (31), calculating the different distributions of hexamer frequencies within introns versus exons. The profiles of genomic regions depict exonic enhancing and silencing properties, while HEXplorer score (HZ_{EI}) differences can assess mutational effects within SREs. Here, the sequence CATGGATGGAGCA was shown to have the longest contiguous HZ_{EI} -positive stretch, reflecting splicing enhancing properties (Figure 3B). In both the proximal and the distal parts of fragment D, we selected point mutations (5G>C, 8G>A) strongly decreasing the HZ_{EI} -positive area. The double mutation 5G>C/8G>A was predicted to maximally neutralize the enhancing properties of FGB7-D (Figure 3B). To examine this prediction, the mutations were tested within the eGFP enhancer reporter and inserted into FGB7-D upstream of the reporter 5'ss, and HeLa cells were transfected to monitor splice site activity using semi-quantitative RT-PCR. Indeed, the predicted mutations turned out to drastically impair the enhancing functionality of this fragment confirming its activity as another SRE within FGB exon 7 (Figure 3C).

Cryptic splice donor selection is highly dependent on each single SRE

On the basis of the above findings, we extended our analyses to determine whether the newly identified SRE was essential for splice site selection in the physiological exonic context. First, to test if the c.1244+1G>T mutation not only disrupted U1 snRNP binding but by itself might have created an SRE, we analyzed two additional splicing inactivating mutations within the three-exon minigene (c.1244+1G>A, c.1244+2T>A). However, only marginal differences in the splicing pattern could be observed, so that creation of a new SRE can be ruled out as the main cause for the observed splicing pattern (Supplementary Figure S2A). The slight increase in p1 usage for c.1244+1G>A and c.1244+2T>A was compatible with a formation of a moderate putative SRE located directly upstream of the exon/intron boundary (Supplementary Figure S2B). We therefore used the WT as well as the pathogenic FGB c.1244+1G>T three-exon minigenes for further analyses. To complete the picture, we performed HEXplorer-based mutational analyses for fragment C, but also B in order to compare HEXplorer-based inactivation to deletion of the naturally silent SRSF1 binding site (25) (Figure 4A). In agreement with Spena *et al.* (24), we did not observe any effect on cryptic 5'ss activation for an individual mutation as long as the physiological 5'ss was present (Figure 4B, lanes 1–4). Combining, however, either mutations within B and C (Figure 4B, lane 5) or all three parts at the same time (Figure 4B, lane 8), but not B and D (Figure 4B, lane 6) or C and D (Figure 4B, lane 7) resulted in activation of a cryptic 3'ss (Figure 4B, lane 5 and 8; Figure 4C, a2 (**)). This, however, could simply be explained by the accidental upregulation of this cryptic 3'ss (MaxEnt score from -6.23 to 2.39) located within C, and therefore also be present in the combined fragments B and C (Figure 4D). Aside from this, this cryptic 3'ss usage might also be supported by the changed sequence profile after HEXplorer-guided mutagenesis (Figure 4E). Indeed, the sequence environment preceding the AG is composed of a

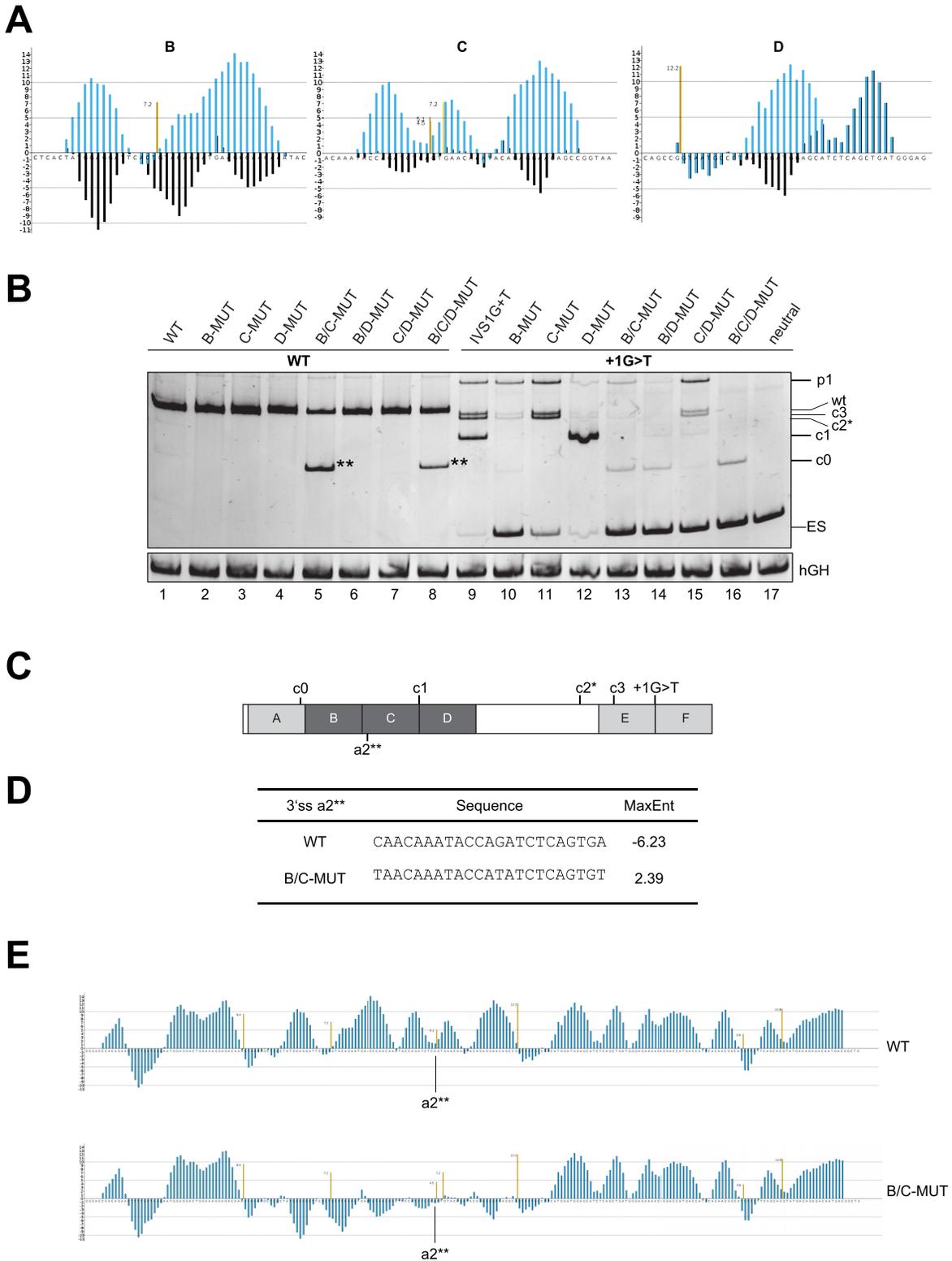


Figure 4. Splicing pattern of the FGB minigenes. **(A)** HEXplorer profiles of WT fragments B, C and D (blue) and mutant profiles (black). **(B)** RT-PCR analysis of splicing patterns of WT and c.1244+1G>T minigenes. Neutral sequence is CCAAACAA-repeat. 2.5×10^5 HeLa cells were transiently transfected with 1 μ g of each construct and 1 μ g of pXGH5. Twenty-four hours after transfection RNA was isolated and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR products were separated by 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide. **(C)** Positions of newly identified cryptic splice donor c0 and acceptor site a2** within FGB exon 7. **(D)** Sequences of the cryptic WT 3'ss a2** and the cryptic 3'ss generated upon mutation B/C-MUT, together with their MaxEnt scores. **(E)** HEXplorer profiles of FGB exon 7 of WT and B/C-MUT.

HZ_{EI}-negative stretch of hexamers reflecting intronic rather than exonic sequences (21).

As seen before, as soon as the physiological canonical 5' ss was rendered non-canonical (c.1244+1G>T), all cryptic splice sites c1, c2*, c3 and p1 were activated but still almost no exon skipping could be observed (Figure 4B, lane 9).

As expected, fragments B and C seemed to activate their proximal downstream splice donor c1. Strikingly, even mutating only one of these fragments completely abolished c1 donor usage and concomitantly enhanced exon skipping (Figure 4B, lanes 10 and 11), demonstrating that both fragments had to act in concert to activate c1. However, they did not differentially affect activation of c2* and c3, indicating that these two sites are independently regulated by another SRE upstream of both c2* and c3.

In agreement with the individual fragments' splicing regulatory activity (Figure 3A), changing the enhancing properties of D had the strongest effect on splice site selection, leading to an almost exclusive c1 donor usage and very little exon skipping, thereby shortening the exon (Figure 4B, lane 12). Further mutation of any combination of fragments drastically reduced exon 7 recognition (Figure 4B, lanes 13–16), and also activated the fourth exonic cryptic 5' ss c0 with an HBS of 9.4 (Figure 4B, lanes 13–16; Figure 4C). Since fragment A increased splice donor recognition 75-fold within the enhancer reporter (Figure 3A), it is likely that c0 was activated when there was no concurrent position-dependent inhibition by B or C.

Eventually, we inserted HEXplorer-guided point mutations into B instead of deleting B (25) to maintain constant exon length. Inactivating B by point mutations resulted in complete loss of c1 usage and an increase in exon skipping, whereas deleting fragment B only moderately impacted the splicing pattern (Supplementary Figure S3). This apparent discrepancy might be explained by the circumstances that the deletion brings fragments A and C in juxtaposition with each other, increasing the overall enhancing properties of this area.

We also treated WT and c.1244+1G>T mutant minigenes with the protein synthesis inhibitor CHX to examine if the observed mutation-induced splicing pattern also depended on NMD. However, as no difference in the splicing patterns could be observed, we exclude NMD as being responsible for the pattern of mutation-induced transcript isoforms (Supplementary Figure S4).

In summary, all four fragments (A–D) regulated both exon recognition and splice site selection by inhibiting upstream splice donor usage and simultaneously stimulating downstream splice donor usage. They were required to repress weak 5' ss along the way to the physiological 3' exon end.

Variation of 5' ss complementarity systematically controls FGB exon 7 inclusion in the presence of various SREs

To examine the impact on exon recognition and splice site selection of a single SRE and the 5' ss it supports, we investigated splice site activation of fragments B, C and D individually. To this end, the FGB-7 sequence was fully substituted with neutral sequences maintaining only c1, c3

and c.1244+1G>T. Each fragment was then individually replaced back into this simplified splicing neutral exon at its physiological position either upstream of c1 or c3 (Figure 5A). Additionally, the HBS of c1 or c3 were stepwise increased to examine the interaction between the splice site proper and surrounding SREs.

Gradually increasing the HBS either for c1 from 12.2 up to 20.8 or for c3 from 10.8 up to 20.8 led to a strong increase in splice site recognition in an otherwise fully splicing-neutral environment (Figure 5B–D, cf. lanes 1–4). In this particular neutral context, an HBS threshold of 18.8 was required for splice site recognition (Figure 5B–D, lane 3). Exon skipping, however, could not be totally eliminated even by increasing the HBS up to 20.8. As expected, inserting either B, C or D into this neutral exon substantially increased exon recognition and combined with an increase in complementarity of the supported splice site fully restored exon recognition (Figure 5B–D, cf. lanes 5–8). In the same way, the mutant versions of all individual fragments supported exon recognition much less, confirming the splicing enhancing activity of fragments B, C, D (Figure 5B–D, cf. lanes 5–8 with lanes 9–12). Furthermore, this experimental setting also allowed to estimate that e.g. fragment C contributed equally to exon recognition as an increase in splice site complementarity from HBS 15.8 to 20.8 (Figure 5C, cf. lanes 4 and 6).

Multiple SR proteins bind to FGB exon 7

To identify splicing regulatory proteins binding to RNA fragments B, C or D we performed RNA affinity purification assays, extending the work of Spina *et al.* on fragment B binding SRSF1 (25). We therefore incubated short WT or mutant sequence RNA oligonucleotides with HeLa nuclear extract (32). After several washing steps, the remaining specifically bound proteins were eluted, separated by SDS-PAGE and analyzed via mass spectrometry analysis. For B and C, 13 out of 14 SR protein abundance ratios 'mutant/WT' were below 1, indicating that a significant number of SR proteins showed higher affinities for WT sequences. For fragment B, 5 out of 6 SR protein intensities were lower in the mutant sequence, covering a range of 0.36–0.82 ($p = 0.04$, Fisher's exact test). For fragment C, all 8 SR protein intensities were lower in the mutant sequence, covering a range of 0.11–0.93 ($p = 0.0022$, Fisher's exact test) (Supplementary Table S2). For fragment D, a diverse picture emerged: 4 out of 8 SR protein intensities—including SRSF1—were lower in the mutant sequences (range 0.54–0.94, n.s.), although specifically Tra2 β was found at slightly elevated levels of 1.14. In particular, SRSF1 showed the highest counts of unique peptides of all SR proteins and was decreased in all three mutant fragments.

For SRSF1 and Tra2 β , we additionally carried out western blot analyses. As expected, not only control RNA B, but also both RNA oligo C and D were bound by SRSF1, and the respective mutations clearly impaired binding, confirming the mass spectrometry results. Tra2 β was also bound to all individual WT oligos B–D, and the reduction in binding to mutant sequences was even more pronounced than for SRSF1 (Figure 6A). Analyzing the sequence composition of the individual fragments revealed

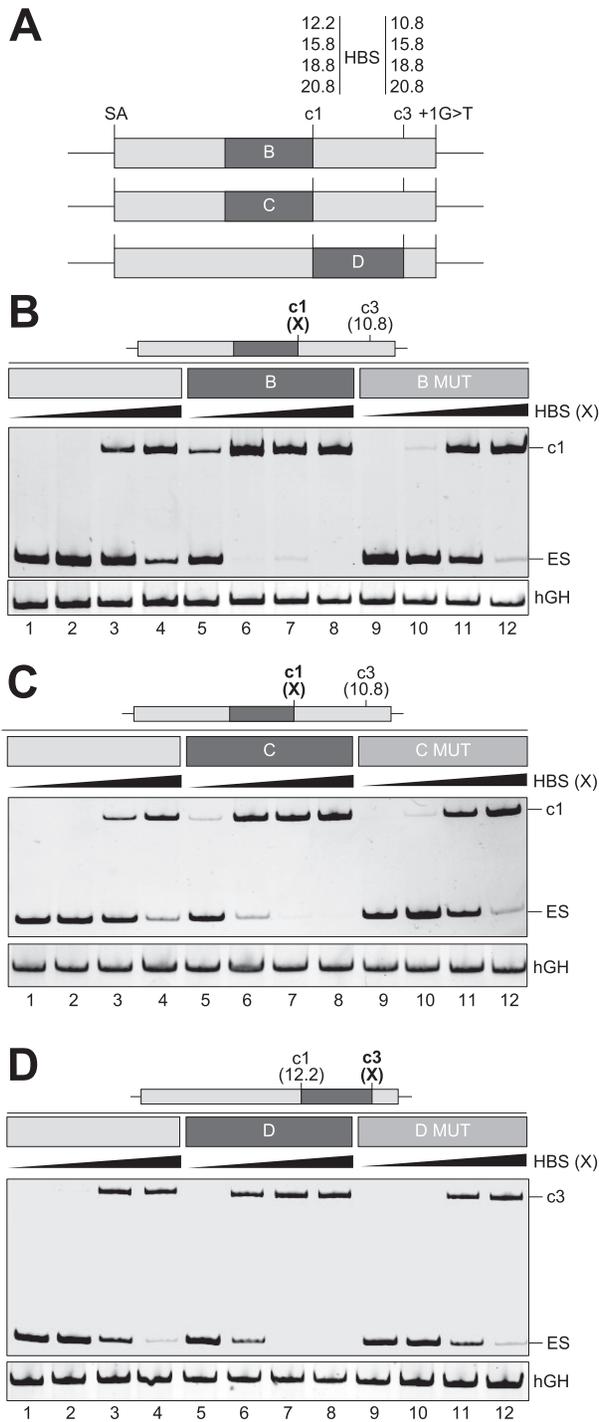


Figure 5. Impact of SREs on exon 7 recognition. (A) Schematic overview of the modified c.1244+1G>T minigene containing only neutral sequences (CCAAACA-repeats, light gray boxes), c1, c3, c.1244+1G>T and either fragment B, C or D, respectively. Additionally, the HBS of c1 and c3 were stepwise increased to the values depicted above. (B–D) RT-PCR analyses of the splicing pattern of the minigenes as shown in (A). HBS (X) at the right hand side of the black wedges above lanes 1–4, 5–8, 9–12 indicates for either cryptic site c1 or c3 (marked as bold and with X) increasing HBS values given in (A). 2.5×10^5 HeLa cells were transiently transfected with 1 μ g of each construct and 1 μ g of pXGH5. Cells were subjected to RT-PCRs using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR amplicons were separated on a non-denaturing 10% polyacrylamide gel and stained with ethidium bromide. Exon skipping (ES).

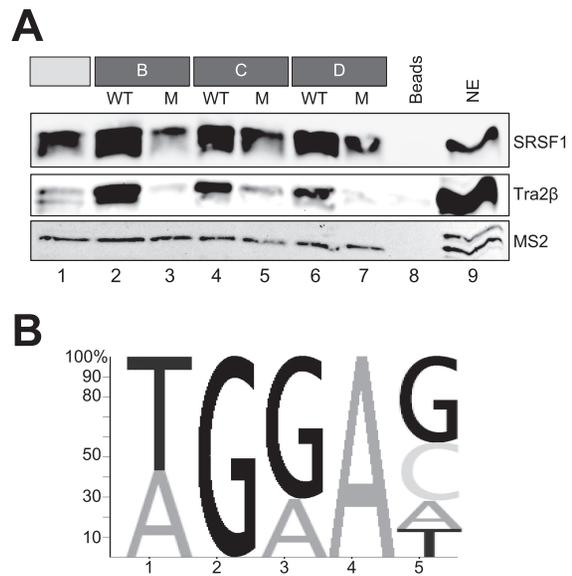


Figure 6. Western blot of SRSF1 and Tra2β binding to each fragment but not to the mutant variants. (A) RNAs including MS2 loops were immobilized using agarose beads, and analyzed for proteins binding by western blot. After the precipitated proteins have been resolved by SDS-PAGE (12%), specific antibodies directed against SRSF1 and Tra2β were used. MS2 coat protein added to the nuclear extract served as a loading control. (B) Sequence logo generated from a sequence motif generated by manual alignment of fragments A, B, C and D. The size of the letters reflects the relative frequency of the nucleotides at the position in the alignment.

that all were enriched in common purine-rich sequence motifs ((A/T)GGA; TGAA) (Figure 6B), previously shown to be bound by SRSF1 (33–35) or Tra2β (36–39).

Based on these results, FGB exon 7 seems to be regulated by multiple SR protein binding sites, for the most part by SRSF1, which regulates cryptic splice donor usage.

HEXplorer-guided mutations beyond FGB exon 7 induce cryptic splice site activation

In order to examine whether the splice site selection concept depending on both 5'ss complementarity and position dependent activity of up- and downstream SREs can be extended beyond FGB exon 7, we tested two examples outside the FGB gene. In particular, in both selected examples the competing U1 snRNA binding site and the WT splice site have similar U1 snRNA complementarity.

First, we revisited the well-documented pathogenic intronic SRE mutation (G to A substitution at +26, termed '759+26G>A') downstream of E1α PDH exon 7, which was found in a patient suffering from encephalopathy and lactic acidosis. This mutation has been shown to create a *de novo* SRSF2 binding site leading to activation of a cryptic splice site located within E1α PDH intron 7 (40,41). This cryptic splice site has even higher U1 snRNA complementarity (HBS 13.7) than the weak physiological 5'ss (HBS 12.2).

From E1α PDH, we derived the physiological 5'ss, the cryptic 5'ss and the intronic region in between containing the SRSF2 binding site, and inserted these into our FGB splicing neutral three-exon minigene. We furthermore inserted two copies of SRSF7 binding sites upstream of the

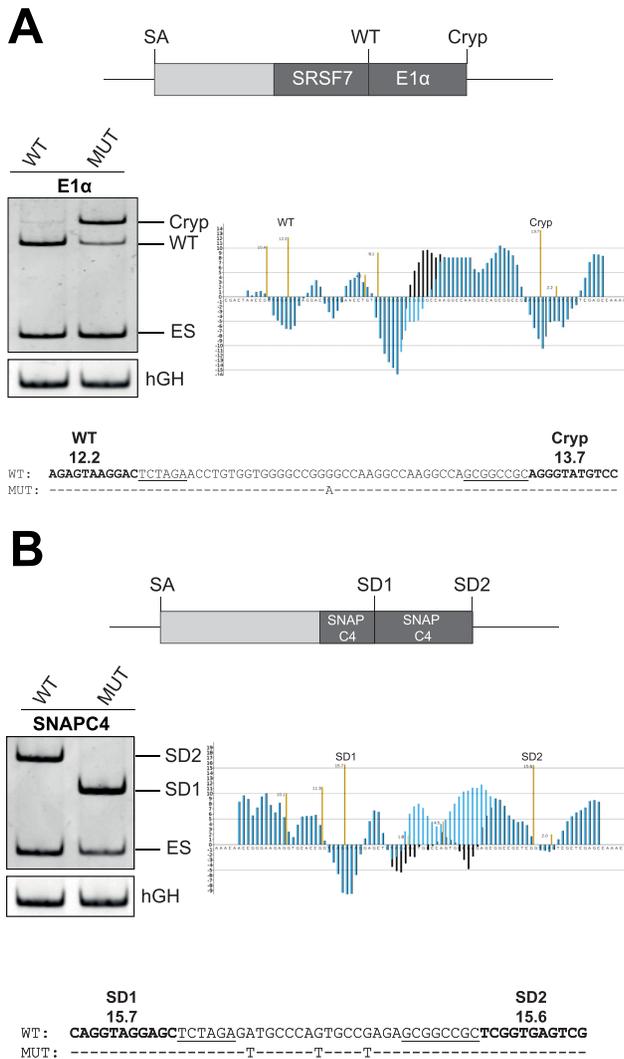


Figure 7. Extension of splice site selection concept from FGB exon 7 three-exon minigene to two other genes. (A) The middle exon (top) contains a fragment derived from the E1 α PDH gene including the two corresponding splice donor sites (WT, Cryp) at their authentic positions, and else neutral sequences (CCAAACAA-repeats, light gray boxes) and two SRSF7 binding sites (61). HEXplorer profile of mutant sequence (black) shows stronger splice enhancing properties than WT (blue). 2.5×10^5 HeLa cells were transiently transfected with 1 μ g of the construct and 1 μ g of pXGH5. At 24 h after transfection, total-RNA samples were collected and used for RT-PCR with primer pair #2648/#2649 and normalized to hGH (#1224/#1225). (B) Same minigene containing two fragments derived from the SNAPC4 gene including the two corresponding splice donor sites at their authentic positions in between restriction sites (top). HEXplorer profiles of mutant sequence (black) shows weaker splice enhancing property than WT (blue). RT-PCR of transcripts amplified from the enhancer reporter (left hand side). Samples were prepared as described for panel A.

physiological 5'ss, since it has been shown that recognition of this rather weak E1 α PDH splice donor is dependent on a strong SRE (41) (Figure 7A, top). In agreement with previous results, the 759+26G>A mutation led to a switch in 5'ss usage (wt to crypt, Figure 7A, left). Furthermore, HEXplorer analysis of the sequence between wt and crypt confirmed this observed phenotype: the pathogenic

759+26G>A mutation positively shifted the HEXplorer profile ($\Delta HZ_{EI} = 72$) indicative of increased downstream SRE activity (Figure 7A, right).

Second, we randomly selected an exon with an unused upstream U1 snRNA binding site of comparable complementarity (HBS 15.7) as the physiological 5'ss (HBS 15.6) from our fibroblast transcriptome dataset (see below): exon 12 in the SNAPC4 transcript (ENST00000298532.2). We inserted the following four segments into our FGB splicing neutral three-exon minigene: (i) upstream of the unused U1 snRNA binding site, to account for possible natural SRE context, (ii) exonic U1 snRNA binding site, (iii) physiological 5'ss, (iv) region between these sites (Figure 7B, top).

Following insertion of these SNAPC4 segments into the splicing reporter, we transfected HeLa cells and analyzed the splicing pattern confirming exclusive usage of the physiological 5'ss (wt; Figure 7B, left). HEXplorer-guided mutagenesis decreased the enhancing properties in the region between the U1 snRNA binding site and the 5'ss ($\Delta HZ_{EI} = -162$; Figure 7B, right), and splicing completely switched from the physiological 5'ss to the further upstream located U1 snRNA binding site (crypt; Figure 7B, left). Again, splice site usage seemed to be regulated by promoting downstream splice donor usage and simultaneously repressing upstream splice donor usage.

Taken together, we have confirmed our SRE dependent splice site selection concept in two examples beyond FGB exon 7: each had a pair of physiological 5'ss and U1 snRNA binding site with similar complementarity—one exonic and one intronic.

Can SREs explain 5'ss selection between GT sites of similar U1 snRNA complementarity?

We independently tested our 5'ss selection concept on individual pairs of a 5'ss and a nearby rarely used exonic U1 snRNA binding site with even higher complementarity, systematically selected from a dataset of 54 human RNA-Seq samples. These samples were derived from short term cultivated *in vivo* aged human dermal fibroblasts, collected from 30 healthy control subjects. Alignment with STAR (Ensembl 82) identified 2,050,307 multiply covered exon-exon junctions (E-MTAB-4652; Kaisers *et al.* PLoS One, in revision). From these, we selected exons with highly used canonical 5'ss (>10 000 exon junction reads). In this subset, we additionally selected only exons containing a U1 snRNA binding site within 35 nucleotides upstream that (i) had high complementarity (HBS > 14), (ii) had higher complementarity than the authentic 5'ss, but (iii) was silent (# reads < 1.3% of authentic 5'ss; median # reads 3). To allow for a putative SRE hexamer between the GT sites but not overlapping with either one, we furthermore required at least 17 (8+6+3) nucleotides in between. Application of these strict selection criteria left only 19 such exons from 19 different genes.

In each of these 19 hits, we scanned three separate regions of equal size for SREs: (i) upstream the silent GT-site, (ii) between silent GT-site and real 5'ss, (iii) downstream the authentic 5'ss. Regions (i) and (ii) each included the full 11 nucleotides of the silent GT-site and 5'ss. The size of these re-

gions was individually determined by the distance between the pair of silent GT-site and 5'ss.

We assessed these three regions for SREs using web resources for the common algorithms ESEfinder 3.0, RESCUE-ESE, FAS-ESS-hex3, PESX, ESRsearch and HEXplorer (31,42–47), applying respective default settings for SRE detection.

In order to find a semi-quantitative measure for the 'overall enhancer effect' in a given upstream exonic neighborhood of a GT-site, we first assigned a weight of +1 or –1 for each enhancer or silencer motif in this region predicted by any of the algorithms. To account for the direction dependence of enhancer action, we performed the same calculation for a mirror region of equal size downstream the GT-site and defined the *splice site enhancer weight* as the difference between sum of upstream weights and sum of downstream weights (Figure 8A). In this way, we treated exonic splicing enhancer (ESEs) occurring downstream of a GT-site as exonic splicing silencers with the same weight. This splice site enhancer weight was designed to capture both enhancing and silencing properties of equally sized regions up- and downstream of any GT-site, and its construction is analogous to the 'exonic splicing motif difference' defined by Ke *et al.* (48). In the same way (21), we calculated splice site weights using the ESRseq and HEXplorer algorithms as the average ESRseq and HEXplorer difference between the up- and downstream regions of any GT-site.

In the 19 genes containing exons with highly used 5'ss and nearby silent GT-sites with higher complementarity, all three splice site enhancer weights were significantly higher for authentic 5'ss than for silent GT-sites (enhancer/silencer weights $p = 0.009$, ESRseq $p = 0.006$, HEXplorer $p = 0.0002$; Figure 8B–D). Indeed, 17 out of 19 genes contained predicted SREs expected to repress U1 snRNA binding sites with even higher complementarity in favor of the authentic 5'ss, which is consistent with our 5'ss selection concept derived from the FGB exon 7 model system. These results suggest that this concept may not be limited to FGB exon 7.

DISCUSSION

In this study, we provide evidence for multiple SREs within the human FGB exon 7. Predictions obtained by both HEXplorer (21) and HBS (3,4), combined with the position-dependence of SREs (10), allow us a glimpse at understanding specific splicing outcomes of human pathogenic 5'ss mutations, based on the model exon FGB7. Here, we propose that, starting from the 3'ss, 5' splice site selection iteratively proceeds along an alternating sequence of U1 snRNA binding sites and interleaved SREs which can in principle support different 3' exon ends. Like in a relay race, SREs can either suppress a potential 5'ss and pass the splicing baton on or splicing actually occurs. This picture may permit generalization to a model for 5'ss selection and 3' exon end definition.

Binding of the 5' end of U1 snRNA to 5'ss initiates spliceosome formation. A higher base pair complementarity to U1 snRNA thereby supports splice site recognition to a higher degree (4,49). In many cases it has been shown that cryptic splice sites are significantly weaker than their

authentic 5'ss counterparts (50). In line with this, assessment of the hydrogen bonding patterns of potential U1 snRNA binding sites (http://www.uni-duesseldorf.de/rna/html/hbond_score.php) within FGB exon 7 revealed that the authentic 5'ss had a higher HBS than the cryptic sites and that in turn the cryptic sites were stronger than the remaining potential U1 snRNA binding sites. Incidentally, we found the splice site p1 within the downstream intron, which was used in the presence of the c.1244+1G>T mutation and which seemed to be below detection limit in previous experiments but had an HBS comparable to c1 (25). However, even slightly different expression levels of splicing regulatory proteins under these different experimental conditions might have caused this difference.

Generally, mutations targeting an SRE may not only activate cryptic splice sites but can also lead to complete loss of exon recognition (17). We showed that either inserting SREs or strengthening the cryptic splice sites c1 and c3 within our simplified splicing neutral exon led to a gradual increase in exon recognition, which depended both on cryptic 5'ss HBS and support by splicing regulatory proteins.

By using an enhancer reporter, we identified multiple elements (A–D) within exon 7, which are each able to promote downstream splice donor usage. This is not surprising, since at least 3/4 of all nucleotides within a normal exon have been shown to be involved in splicing regulation (51). Furthermore, it could be shown that multiple enhancer elements increase the overall rate of splicing (10,52) which was attributed to the fact that more SREs improve the chance of an enhancer element promoting U1 snRNP binding to a 5'ss.

An important part in selecting a 5'ss in the presence of multiple simultaneously acting SREs is played by the strict position-dependency of splicing regulatory proteins. Plenty of work has shown that the same splicing regulatory proteins can activate splice donor usage from upstream positions as well as inhibit from downstream positions (18,53–55). Minigene analyses could show that the position-dependency seems to be a common mechanism of several SR and hnRNP proteins (10,11). This is in line with our findings showing that mutating fragment B and C upstream of c1 led to an impaired c1 donor usage, whereas mutating D upstream of c3 reduced c3 usage. However, mutating e.g. fragment D at the same time led to an upregulation of the upstream located splice donor site c1.

SR proteins are composed of one or two RRMs and one RS domain that participate directly in the interaction with other proteins or with the RNA itself. Until now, there is controversial data about the exact mechanism by which SR proteins promote splice site recognition. It has been shown that SRSF1 targets the U1-specific protein U1-70K to facilitate recruitment of the spliceosome to a splice donor site via RS-RS domain interactions (6,7). However, it also has been shown that the RS domain is not responsible for the interaction with U1 snRNP but rather the RRM (8,9). Notwithstanding, for the SR-related protein Tra2 it has been shown that splice site repression and activation occur via different effector domains (56). Therefore, it is tempting to speculate that SRSF1 acts in the same fashion; whether it is the RRM or the RS domain that is involved in activation or repression needs to be subject of closer examination. Fur-

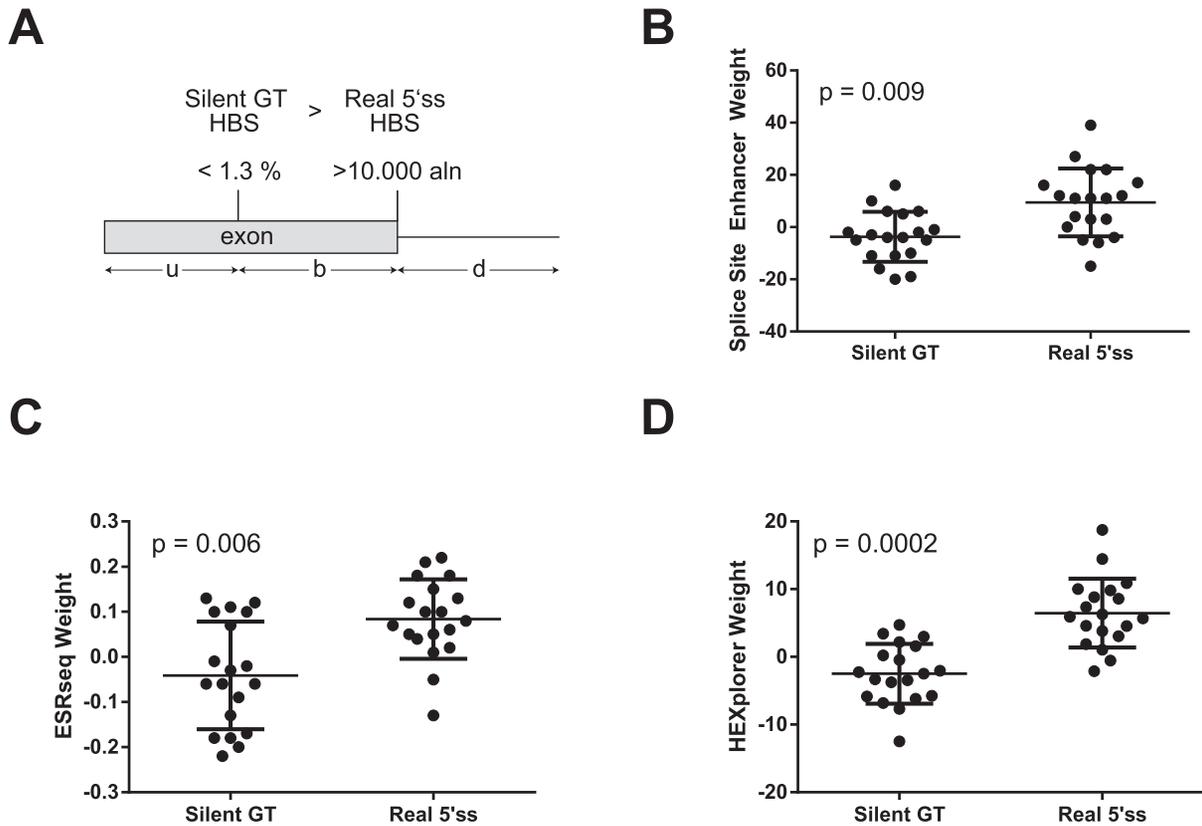


Figure 8. SREs support highly used 5'ss more than nearby silent GT-sites with similar or higher U1 snRNA complementarity. We screened 19 exons with pairs of a real 5'ss (>10 000 exon junction reads, *aln*) and a nearby silent exonic U1 snRNA binding site (HBS > 14, > 5'ss, <1.3% 5'ss *aln*) for splicing regulatory elements. (A) For each of these 19 exons, we scanned three separate regions of equal size: upstream the silent GT-site (u), between silent GT-site and real 5'ss (b), downstream the real 5'ss (d). (B) *Splice site enhancer weights*, (C) ESRseq weights and (D) HEXplorer weights, calculated as differences between sum of upstream weights and sum of downstream weights (A; 'u-b', 'b-d') were significantly higher for real 5'ss than for silent GT-sites (Wilcoxon signed rank test).

thermore, both functions might act simultaneously to inhibit the use of upstream cryptic splice sites. This perfectly matches with our results revealing that each SRE activated downstream donor usage but at the same time inhibited upstream located 5'ss.

Furthermore, it was assumed that SREs can be recognized by more than one SR protein (34) and that a purine-rich enhancer element can enhance splicing if bound by a protein complex (57). This is in accordance with our mass spectrometric analysis of proteins bound to SREs where we found a couple of SR proteins showing a selective higher abundance in wt sequence based affinity purifications. Pandit *et al.* (34) suggested that different kinds of SR proteins exist: some SR proteins like SRSF1 bind rather loosely to exonic positions, while others bind to more distinct binding motifs (58). It was additionally shown that SR proteins can directly interact with each other, like Tra2 and SRSF1 (59) and that Tra2 recruits other splicing factors to ESE sequences (60). From our data we cannot decide whether SR proteins bind directly to SREs or indirectly in complex with other SR proteins. Therefore, we can extend our hypothesis that the simultaneous inhibition or promotion of splice donor usage by SREs is facilitated by a set of splicing regu-

latory proteins which only together ensure specific binding as it has been shown for FUS and hnRNP H (11).

Aberrant splicing is one major cause of human genetic disease, and does not only involve mutations within splice sites but also within SREs, which makes computational mutation assessment within SREs highly important. All investigated elements within FGB exon 7 identified using the HEXplorer algorithm (21) could be experimentally validated as enhancer elements. Only recently, the HEXplorer has been proven a powerful tool to predict not only mutational effects within SREs but also their severity (15). Therefore, we propose that for a 'functional 5'ss usage prediction' both the splice site complementarity to U1 snRNA and in addition the sequence environment must be considered (21). The data presented in this work suggest a model in which SREs surrounding cryptic splice sites act in a strictly position-dependent manner, possibly supporting or antagonizing each other. This effect, however, is invisible in the presence of the strong authentic splice site (Figure 9, WT). As soon as the authentic splice site is disrupted, the SRE effects within FGB exon 7 become visible, leading to exon shortening. Even though c1 is supported by the SREs in B and C, fragment D lowers the enhancing potential for this cryptic splice site, letting the SREs in B and C appear

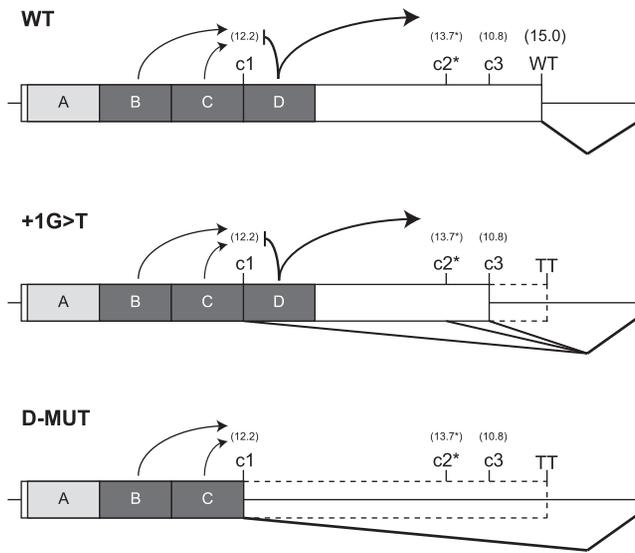


Figure 9. Model for FGB exon 7 recognition. Authentic splice donor usage is facilitated by the WT 5'ss with an HBS of 15.0 exceeding all cryptic splice sites and by the bi-directional properties of SREs that antagonize each other (B, C versus D) (WT). Inactivation of the WT 5'ss (GT>TT) leads to the usage of cryptic splice sites within exon 7, thereby shortening the exon (c.1244+1G>T) to either c1, c2* or c3. Mutation of fragment D shortens the exon exclusively to c1, as it is the only remaining SRE-supported splice site which is no longer repressed by the downstream localized SRE (D-MUT).

'naturally silent' and leading to activation of c1, c2* and c3 in comparable amounts (Figure 9, c.1244+1G>T). This is supported by the fact that mutating fragment D leads to a drastic increase in c1 and loss of c2* and c3 usage, further shortening the exon (Figure 9, D-MUT).

Moreover, we could substantiate our findings of SRE-dependent splice site selection by evaluating two further examples beyond FGB exon 7 in the same reporter system. Within both, E1 α and SNAPC4, competing U1 snRNA binding sites exist that are regulated through position-dependent SREs. Mutational analyses could switch splicing in either direction, matching the calculated HEXplorer scores by interrupting or creating a sequence with exonic enhancer properties, respectively (Figure 7). Extending these examples of competing U1 snRNA binding sites, we specifically selected exons from 19 genes with highly used 5'ss downstream of nearby silent GT-sites with similar complementarity (average HBS difference 1.2) from our fibroblast RNA-Seq transcriptome dataset. This choice of exons permitted focusing on predicted SRE effects on 5'ss selection rather than 5'ss strength. Simultaneously taking potential SREs in regions both up- and downstream of a GT-site into account, as suggested by our extensive FGB exon 7 analyses, we found authentic, highly used 5'ss significantly more supported by SREs than silent GT-sites. This result was consistently found in analyses using eight different SRE identifying algorithms, indicating that the proposed concept may be more generally valid beyond our FGB exon7 model system.

This concept of potentially iterated position-dependent SRE action may highlight the important role of SREs not

only in alternative splice site selection, but also as key for constitutive splicing in exons containing internal U1 snRNA binding sites that must be ignored to obtain an appropriate exon end.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Imke Meyer and Björn Wefers for excellent technical assistance and Philipp Peter for implementing the HEXplorer algorithm on our RNA website.

FUNDING

Deutsche Forschungsgemeinschaft (DFG) [SCHA 909/4-1]; Jürgen Manchot Stiftung (to A.L.B., L.W., L.H., H.S.); Stiftung für AIDS-Forschung, Düsseldorf (to H.S.). Funding for open access charge: Heinrich-Heine University. *Conflict of interest statement.* None declared.

REFERENCES

- Will, C.L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, **3**, a003707.
- Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H. and Luhrmann, R. (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like pre-catalytic splicing complexes. *Mol. Cell*, **38**, 223–235.
- Kammler, S., Leurs, C., Freund, M., Krummheuer, J., Seidel, K., Tange, T.O., Lund, M.K., Kjems, J., Scheid, A. and Schaal, H. (2001) The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *RNA*, **7**, 421–434.
- Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J., Hipp, M., Meyer, I., Gierling, W., Theiss, S., Preuss, T. et al. (2003) A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.*, **31**, 6963–6975.
- Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
- Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.
- Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
- Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R. and Ghosh, G. (2011) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 8233–8238.
- Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schoneweis, K., Hertel, K.J. and Schaal, H. (2013) Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA*, **19**, 96–102.
- Reber, S., Stettler, J., Filosa, G., Colombo, M., Jutzi, D., Lenzken, S.C., Schweingruber, C., Bruggmann, R., Bachi, A., Barabino, S.M. et al. (2016) Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. *EMBO J.*, **35**, 1504–1521.
- Domsic, J.K., Wang, Y., Mayeda, A., Krainer, A.R. and Stoltzfus, C.M. (2003) Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. *Mol. Cell. Biol.*, **23**, 8762–8772.

13. Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C. and Black, D.L. (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat. Struct. Mol. Biol.*, **15**, 183–191.
14. Hartmann, L., Theiss, S., Niederacher, D. and Schaal, H. (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front. Biosci.*, **13**, 3252–3272.
15. Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frebourg, T., Tosi, M. and Martins, A. (2016) Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet.*, **12**, e1005756.
16. Daguene, E., Dujardin, G. and Valcarcel, J. (2015) The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep.*, **16**, 1640–1655.
17. Krawczak, M., Thomas, N.S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J. and Cooper, D.N. (2007) Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.*, **28**, 150–158.
18. Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J. and Fairbrother, W.G. (2011) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11093–11098.
19. Sterne-Weiler, T., Howard, J., Mort, M., Cooper, D.N. and Sanford, J.R. (2011) Loss of exon identity is a common mechanism of human inherited disease. *Genome Res.*, **21**, 1563–1571.
20. 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.
21. Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J.O. and Schaal, H. (2014) Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. *Nucleic Acids Res.*, **42**, 10681–10697.
22. Asang, C., Hauber, I. and Schaal, H. (2008) Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic Acids Res.*, **36**, 1450–1463.
23. Zhang, X.H., Arias, M.A., Ke, S. and Chasin, L.A. (2009) Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA*, **15**, 367–376.
24. Spena, S., Duga, S., Asselta, R., Malcovati, M., Peyvandi, F. and Tenchini, M.L. (2002) Congenital afibrinogenemia: first identification of splicing mutations in the fibrinogen Bbeta-chain gene causing activation of cryptic splice sites. *Blood*, **100**, 4478–4484.
25. Spena, S., Tenchini, M.L. and Buratti, E. (2006) Cryptic splice site usage in exon 7 of the human fibrinogen B beta-chain gene is regulated by a naturally silent SF2/ASF binding site within this exon. *RNA*, **12**, 948–958.
26. Selden, R.F., Howie, K.B., Rowe, M.E., Goodman, H.M. and Moore, D.D. (1986) Human growth hormone as a reporter gene in regulation studies employing transient gene expression. *Mol. Cell. Biol.*, **6**, 3173–3179.
27. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
28. Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. and Brenner, S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.
29. Poschmann, G., Seyfarth, K., Besong Agbo, D., Klafki, H.W., Rozman, J., Wurst, W., Wiltfang, J., Meyer, H.E., Klingenspor, M. and Stuhler, K. (2014) High-fat diet induced isoform changes of the Parkinson's disease protein DJ-1. *J. Proteome Res.*, **13**, 2339–2351.
30. Caputi, M., Freund, M., Kammler, S., Asang, C. and Schaal, H. (2004) A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J. Virol.*, **78**, 6517–6526.
31. 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.
32. Erkelenz, S., Hillebrand, F., Widera, M., Theiss, S., Fayyaz, A., Degrandi, D., Pfeffer, K. and Schaal, H. (2015) Balanced splicing at the Tat-specific HIV-1 3' splice site is critical for HIV-1 replication. *Retrovirology*, **12**, 29.
33. Anczukow, O., Akerman, M., Clery, A., Wu, J., Shen, C., Shirole, N.H., Raimer, A., Sun, S., Jensen, M.A., Hua, Y. et al. (2015) SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell*, **60**, 105–117.
34. Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M. Jr and Fu, X.D. (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell*, **50**, 223–235.
35. Ray, D., Kazan, H., Chan, E.T., Pena Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q. and Hughes, T.R. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.*, **27**, 667–670.
36. Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139–148.
37. Grellscheid, S., Dalglish, C., Storbeck, M., Best, A., Liu, Y.L., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K. et al. (2011) Identification of evolutionarily conserved exons as regulated targets for the splicing activator Tra2 beta in development. *PLoS Genet.*, **7**, e1002390.
38. Tsuda, K., Someya, T., Kuwasako, K., Takahashi, M., He, F., Unzai, S., Inoue, M., Harada, T., Watanabe, S., Terada, T. et al. (2011) Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. *Nucleic Acids Res.*, **39**, 1538–1553.
39. Erkelenz, S., Poschmann, G., Theiss, S., Stefanski, A., Hillebrand, F., Otte, M., Stuhler, K. and Schaal, H. (2013) Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of vpr mRNA. *J. Virol.*, **87**, 2721–2734.
40. Mine, M., Brivet, M., Touati, G., Grabowski, P., Abitbol, M. and Marsac, C. (2003) Splicing error in E1alpha pyruvate dehydrogenase mRNA caused by novel intronic mutation responsible for lactic acidosis and mental retardation. *J. Biol. Chem.*, **278**, 11768–11772.
41. 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.
42. 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.
43. Zhang, X.H. and Chasin, L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.*, **18**, 1241–1250.
44. Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q. and Krainer, A.R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.*, **15**, 2490–2508.
45. 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.
46. Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T. and Ast, G. (2006) Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. *Mol. Cell*, **22**, 769–781.
47. Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J. and Chasin, L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.*, **21**, 1360–1374.
48. Ke, S., Zhang, X.H. and Chasin, L.A. (2008) Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res.*, **18**, 533–543.
49. 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.
50. Roca, X., Sachidanandam, R. and Krainer, A.R. (2003) Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res.*, **31**, 6321–6333.
51. Chasin, L.A. (2007) Searching for splicing motifs. *Adv. Exp. Med. Biol.*, **623**, 85–106.
52. Hertel, K.J. and Maniatis, T. (1998) The function of multisite splicing enhancers. *Mol. Cell*, **1**, 449–455.
53. Cereda, M., Pozzoli, U., Rot, G., Juvan, P., Schweitzer, A., Clark, T. and Ule, J. (2014) RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol.*, **15**, R20.

54. Ule,J., Stefani,G., Mele,A., Ruggiu,M., Wang,X., Taneri,B., Gaasterland,T., Blencowe,B.J. and Darnell,R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.
55. Llorian,M., Schwartz,S., Clark,T.A., Hollander,D., Tan,L.Y., Spellman,R., Gordon,A., Schweitzer,A.C., de la Grange,P., Ast,G. *et al.* (2010) Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.*, **17**, 1114–1123.
56. Shen,M. and Mattox,W. (2012) Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. *Nucleic Acids Res.*, **40**, 428–437.
57. Yeakley,J.M., Morfin,J.P., Rosenfeld,M.G. and Fu,X.D. (1996) A complex of nuclear proteins mediates SR protein binding to a purine-rich splicing enhancer. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 7582–7587.
58. Anko,M.L., Muller-McNicoll,M., Brandl,H., Curk,T., Gorup,C., Henry,I., Ule,J. and Neugebauer,K.M. (2012) The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.*, **13**, R17.
59. Amrein,H., Hedley,M.L. and Maniatis,T. (1994) The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell*, **76**, 735–746.
60. Lynch,K.W. and Maniatis,T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes Dev.*, **10**, 2089–2101.
61. Cavaloc,Y., Bourgeois,C.F., Kister,L. and Stevenin,J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5**, 468–483.