Nucleic Acids Research, 2024, **52**, 5959–5974
https://doi.org/10.1093/nar/gkae147
Advance access publication date: 1 March 2024
RNA and RNA-protein complexes

Splicing analysis of STAT3 tandem donor suggests non-canonical binding registers for U1 and U6 snRNAs

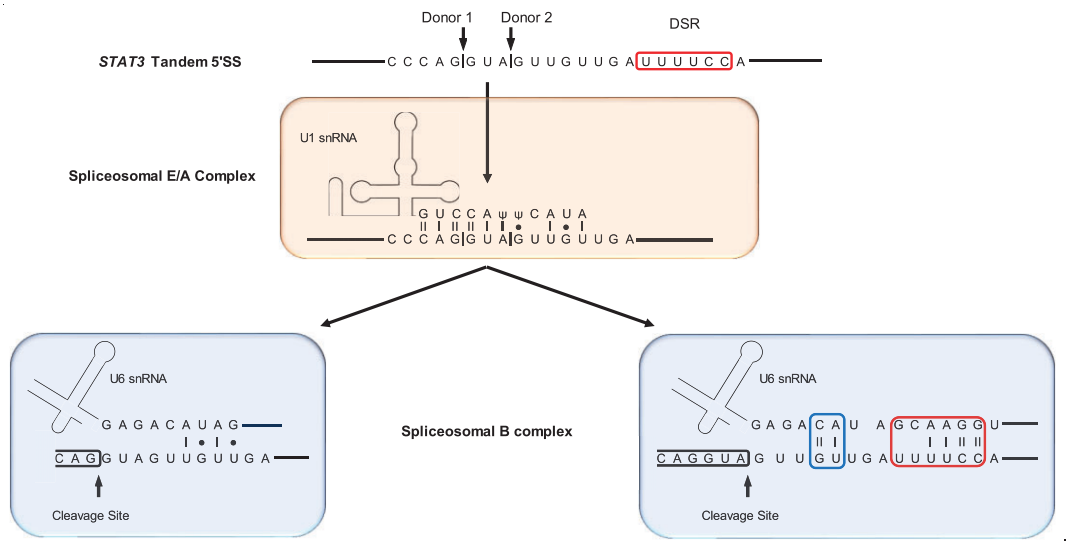
Michal Kramárek[1,2,3], Přemysl Souček[1,2,*], Kamila Réblová[4], Lucie Kajan Grodecká[1] and Tomáš Freiberger[1,2]

[1]Centre for Cardiovascular Surgery and Transplantation, 656 91 Brno, Czech Republic
[2]Faculty of Medicine, Masaryk University, 625 00 Brno, Czech Republic
[3]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 62500 Brno, Czech Republic
[4]Centre of Molecular Biology and Genetics, University Hospital and Masaryk University, Brno, Czech Republic
[*]To whom correspondence should be addressed. Tel: +420 549494560; Email: premysl.soucek@cktch.cz

Abstract

Tandem donor splice sites (5′ss) are unique regions with at least two GU dinucleotides serving as splicing cleavage sites. The Δ3 tandem 5′ss are a specific subclass of 5′ss separated by 3 nucleotides which can affect protein function by inserting/deleting a single amino acid. One 5′ss is typically preferred, yet factors governing particular 5′ss choice are not fully understood. A highly conserved exon 21 of the *STAT3* gene was chosen as a model to study Δ3 tandem 5′ss splicing mechanisms. Based on multiple lines of experimental evidence, endogenous U1 snRNA most likely binds only to the upstream 5′ss. However, the downstream 5′ss is used preferentially, and the splice site choice is not dependent on the exact U1 snRNA binding position. Downstream 5′ss usage was sensitive to exact nucleotide composition and dependent on the presence of downstream regulatory region. The downstream 5′ss usage could be best explained by two novel interactions with endogenous U6 snRNA. U6 snRNA enables the downstream 5′ss usage in *STAT3* exon 21 by two mechanisms: (i) binding in a novel non-canonical register and (ii) establishing extended Watson–Crick base pairing with the downstream regulatory region. This study suggests that U6:5′ss interaction is more flexible than previously thought.

Graphical abstract



Introduction

Pre-mRNA splicing is an essential RNA processing step that removes introns and joins exons to produce mature messenger RNA transcript (mRNA). An important hallmark of mammalian RNA processing is alternative splicing (AS). In humans, approximately 95% of human genes produce two or more splicing isoforms (1,2). Therefore, AS (along with alternative transcription initiation, alternative polyadenylation, and RNA editing) generates a considerable amount of variability in the human transcriptome and proteome despite the limited number of genes. There are three general ways in which AS can affect gene expression: (i) production of

Received: October 10, 2022. **Revised:** February 2, 2024. **Editorial Decision:** February 5, 2024. **Accepted:** February 16, 2024

protein or non-coding RNA isoforms that differ in structure or function, (ii) generation of mRNAs with differing 5′ or 3′ UTR that affects translation efficiency or mRNA stability and (iii) protein level regulation by introducing a premature stop codon (PTC) and subsequent mRNA degradation by nonsense-mediated mRNA decay (NMD).

Introns in humans are defined by four main splicing signals, 5′ splice site (5′ss or donors), branch point adenosine (BP), polypyrimidine tract (PPT), and 3′ splice site (3′ss or acceptors). While these sites are important for proper splicing, their sequence motifs are highly degenerate (3). Practically, the only conserved nucleotides in splice sites are GT and AG at the start and end of the intron and branch site adenosine near the 3′ss, respectively. 98.76% of all human introns are U2-type GT-AG introns with the next most represented group being GC-AG introns (0.86%) (3). The diversity of main splicing signals is made possible in part by the presence of *cis*-acting splicing regulatory elements (SRE), e.g. splicing enhancers or silencers, that affect splice site recognition (4,5). Splicing enhancer and silencer sequence motifs are even more degenerate since their function is mediated mainly by binding of two very diverse groups of RNA-binding proteins (RBPs), Serine/arginine-rich (SR) proteins (6) and heterogeneous nuclear ribonucleoproteins (hnRNPs) (7).

The most common AS type in humans is the skipping of one or more exons followed by the use of alternative acceptors and donors, intron retention, and inclusion of mutually exclusive exons (8). Interestingly, alternative donor pairs separated by 10 or fewer nucleotides (here referred to as tandem donor sites) constitute 23,7% of all alternative donor events (9). The Δ3n tandem donors are an important class of alternative donors separated by multiples of three nucleotides (Δ3n donors). While splicing of other tandem donors might result in a frameshift and completely change the downstream amino acid sequence, splicing of Δ3n donors results in adding or deleting a few amino acids and has the potential to slightly modulate the protein function.

A critical step in 5′ss recognition is the base pairing of U1 snRNA 5′ end with the exon/intron boundary of 5′ss (10,11). The U1 snRNA can establish up to 11 base pairs with 5′ss in the last 3 exonic and first 8 intronic nucleotides. While the last two intronic nucleotides of human 5′ss are not conserved, they can contribute to the recognition in some cases (12). However, there is considerable diversity among authentic human 5′ss sequences and only a small percentage of them fit the extended MAG/GURAGUAU consensus perfectly. On average, authentic human 5′ss contains 7 (out of 11) nucleotides that conform to the extended consensus sequence, and 25% of 5′ss match the consensus motif in 6 or fewer nucleotides. U1/5′ss duplex can also accommodate shifted, bulge, and asymmetric loop registers (13–15). Specifically, Roca and Krainer reported that in some human 5′ss, U1 snRNA binding was shifted by one nucleotide downstream from the canonical register (13). Additionally, various bulges in both strands of U1/5′ss duplex, and asymmetric loops were shown to exist by Tan and colleagues (15). These atypical 5′ splice sites looked like one nucleotide was inserted or deleted between positions +2 and +5 when compared to the consensus 5′ss motif. It is estimated that 1.5% of all human 5′ss use non-canonical registers. This means that some 5′ss that would be considered weak by traditional prediction algorithms such as MaxEntScan or H-bonds are in fact stronger when considering non-canonical U1 snRNA binding registers.

Signal transducer and activator of transcription 3 (STAT3) is a transcription factor that mediates cellular responses to interleukins and growth factors and regulates G1 to S phase progression. STAT3 also regulates embryogenesis, immunity and inflammation, haematopoiesis, cell migration and cell survival, and its persistent activation is implicated in multiple cancers (16,17). The *STAT3* gene produces four different isoforms (ΔSα, ΔSβ and Sα, Sβ) which result from alternative splicing of exon 21 (ΔS or S) and 23 (α or β). Isoforms α and β differ in the C-terminal transactivation domain and possess overlapping and distinct functions (16,18). Alternative splicing of the tandem 5′ss in exon 21 results in a ΔS or S isoform that lacks/contains serine in position 701. The splicing regulation and function of these two S variants is largely unknown, but the latest evidence suggests that both these isoforms are important for the proper STAT3 function. Zheng and colleagues (19) showed that the re-expression of isoforms in pairs (ΔSα + Sα or ΔSβ + Sβ) or all four splicing isoforms simultaneously led to better survival of the ABC subtype of diffuse large B-cell lymphoma cancer cell line (ABC DLBCL).

Factors that govern splice site choice within tandem donor splice sites are not well described. Our study aimed to identify and understand the role of these factors in tandem donor splice sites. We chose the tandem donor splice site in exon 21 of the human *STAT3* gene as an exemplary Δ3 donor type representative due to its high evolutionary conservation and well-described splicing pattern.

Materials and methods

Minigene constructs and characteristics

Minigene for *STAT3* gene splicing evaluation was created by inserting an amplified part of wild-type *STAT3* gene human genomic sequence into the pET01 vector (MoBiTec). The insert contained the whole *STAT3* exon 21 and 328 and 250 upstream and downstream flanking intronic nucleotides, respectively. The STAT3 insert was cloned into multiple cloning site in intron between two vector-specific exons giving rise to a three-exon minigene. The resulting minigene was named pET-STAT3. Expression vectors for human U1 snRNA (pGEM-U1) and U6 snRNA (pGEM-U6) were kindly provided by Emanuele Buratti from ICGEB Italy. Minigene variants were prepared by PCR-based site-directed mutagenesis using specific primers and PrimeSTAR® Max DNA Polymerase (Takara Bio). PCR primer sequences are provided in Supplementary Tables S1–S3. Minigenes were transfected into HeLa cells, whole RNA was isolated, and RT-PCR-amplified mRNA transcripts were analysed by capillary electrophoresis to properly distinguish and quantify all splicing isoforms. In wild-type pET-STAT3 (and mutants), we observed the use of three cryptic 3′ss which account for around 5–8% of all transcripts (data not shown). While these cryptic 3′ss have not been explicitly reported in the literature, they can be identified in the GTex and SRA datasets among the four most common mis-splicing events in the vicinity of STAT3 exon 21. These cryptic 3′ss are aggregated in the SpliceVault database at https://kidsneuro.shinyapps.io/splicevault/ (20). The results of the minigene analysis are presented without the transcripts that use these three cryptic 3′ss to make the graphs straightforward. These transcripts and their exact location within cDNA are depicted in Supplementary Figure S2B.

Transfections

HeLa cells (The European Collection of Authenticated Cell Cultures – ECACC) were maintained in RPMI 1640 medium (Sigma-Aldrich) and supplemented with 10% fetal bovine serum (Sigma-Aldrich). One day prior to transfection, 1.2×10^5 cells per well were seeded into a 12-well plate to achieve between 50% and 70% confluency. Depending on the nature of the experiment, pET-STAT3 alone or in combination with other plasmids (pGEM-U1 or pGEM-U6) were used for transfections. Single-construct transfections were carried out using 800 ng of pET-STAT3 minigene, while co-transfections were carried out using 200 ng of pET-STAT3 minigene and 900 ng of second plasmid (pGEM-U1 or pGEM-U6). Transfections for each unique plasmid combination were repeated three times.

An appropriate amount of each plasmid was added to the mix of 100 µl of RPMI medium and 2.4 µl of transfection reagent XtremeGene 9 (Roche) and the mixture was pipetted onto the cells after a 20-minute incubation. Total RNA was extracted 24 hours after transfection using RNeasy Plus Mini Kit (Qiagen) or Quick-RNA Miniprep Kit (Zymo Research). Isolated RNA was treated with a 1U of rDNase I from a DNA-free™ Kit (Thermo Fisher Scientific) and the enzyme was inactivated according to the manufacturer's protocol. The quality and quantity of RNA was checked by NanoDrop 2000c (Thermo Fisher Scientific).

Reverse transcription and capillary electrophoresis

Total RNA from transfected HeLa cells was reverse transcribed to cDNA by Transcriptor First Strand cDNA Synthesis Kit (Roche) or High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. To distinguish amplicons that differ in length by three nucleotides, capillary electrophoresis was used. To prevent overamplification in capillary electrophoresis, initial PCR for cycle optimization was performed by using Taq DNA polymerase (Thermo Fisher Scientific), forward primer pET_1A and reverse primer PS02. Subsequent PCR with optimized cycles using the same polymerase but with FAM-labelled forward primer pET_1FAM and reverse primer PS02 was carried out and 20 µl of this reaction was visualized on 1.5% agarose gel. For capillary electrophoresis, 1 µl of PCR product was mixed with 9.5 µl of HiDi formamide (Thermo Fisher Scientific) and 0.5 µl of GeneTrace 1000 ROX Size Standard (Carolina Biosystems). The mixture was then denatured for two minutes at 95°C and frozen at -80°C. The capillary electrophoresis was carried out on an ABI PRISM® 3500 Genetic Analyzer (Thermo Fisher Scientific) and analysed by GeneMapper 4.1 (Thermo Fisher Scientific) software. Alternatively, PCR products were sent to SEQme s.r.o and the received .fsa files were analysed by GeneMapper 4.1.

In silico analysis

The splice site strength was calculated using a Maximum Entropy Model implemented on the MaxEntScan website (http://hollywood.mit.edu/burgelab/maxent/). The propensity to form a secondary structure in RNA was evaluated by the UNAFold web server feature Quickfold with default settings using RNA 3.0 energy rules and only the lowest calculated change in Gibbs energy (ΔG) is reported here (21). The input sequence was 80 nucleotides long and was centred on the cleavage site of D2 5′ss. For the evaluation of enhancer/silencer activity, two scores based on hexamers were calculated. The total HZ_{EI} score was obtained from a Hexplorer website (https://rna.hhu.de/HEXplorer/) (22). ESRseq score was calculated based on scores obtained from the paper by Ke *et al.* (23). The input for both was the same as for UNAFold. Simple linear regression analysis, *P*-values and other statistics were calculated either by MS Excel, TIBCO Statistica (version 14) or NumPy package in Python. *P*-values were calculated by two sample unpaired t-tests.

Minigene construct for high-throughput splicing analysis

A STAT3 insert from pET-STAT3 was cloned into a vector described and previously used for high-throughput splicing analysis (24) and the new minigene was named pHT-STAT3 (Supplementary Figure S2). The pHT-STAT3 plasmid possesses several additional restriction sites compared to the pET-STAT3. The restriction sites for HindIII and BglII, located within exon 1, were used to introduce a barcode (BC1) that allowed us to match mRNA transcripts with mutations in the plasmid library. To match the BC1 with a mutation in the DNA mutagenesis library, restriction sites for PacI and AsiSI located at the ends of introns between exon 1 and exon 21 were used to shorten the intron for DNA sequencing. Due to read length limitations, *STAT3* exon 21 was shortened to 80 nucleotides by a deletion (e_del 3) spanning nucleotides from 8 through 140, numbered relative to the exon. This change resulted only in a minor increase in exon skipping (up to 8%) (Figure 5).

Constructing mutagenesis libraries

Before mutagenesis, nine unique barcodes were introduced into exon 2 of the pHT-STAT3 plasmid to improve the assignment of potential chimeric transcripts observed before (24). This was done by using primers pHT_BC2_mut_f and pHT_BC2_mut_r. None of these barcodes changed the pHT-STAT3 plasmid splicing. In the next step, two different mutagenesis libraries were constructed separately by PCR-based site-directed mutagenesis using PrimerSTAR® MAX DNA Polymerase (Takara Bio) and two degenerated mutagenesis primer pairs. Library 1 (L1) was generated by primers S3_L1mut_A and S3_L1mut_B which covered positions −7 to +8. Library 2 (L2) was generated by primers S3_L2mut_A and S3_L2mut_B which covered positions +9 to +19. Library L1 had a degeneration rate set as 88% of wild-type nucleotide and 12% of other at the N nucleotide. Library L2 had a degeneration rate set as 82% of wild-type nucleotide and 18% of other at the N nucleotide.

The entire volume of mutagenesis PCR for each library was divided into six tubes containing 50 µl of DH5α competent cells and then transformed by heat shock. The transformation reactions for each library were then pooled together into the Erlenmeyer bank with a total of 40 ml of LB medium with carbenicillin and incubated overnight. The next day, plasmid DNA was isolated and used as a backbone vector to introduce the BC1 barcode. To put the BC1 into the mutagenesis libraries, a portion of exon 1 from the wild-type pHT-STAT3 vector was amplified by PrimeSTAR® Max DNA Polymerase (Takara Bio) using pHT_BC1amp_f forward primer and a pHT_BC1mut_r reverse mutagenesis primer. This sequence was then cloned back into the mutagenesis library between HindIII and BglII restriction sites

using T4 DNA ligase (New England Biolabs). The resulting BC1 NNNNGNNNNGNNNN consisted of 12 degenerate N positions. Each mutagenesis library was barcoded separately, and the heat-shock transformation described above was repeated. Successful barcoding was checked by restriction enzymes NheI and XbaI and mutagenesis by Sanger sequencing. Where needed, the libraries were pooled together equimolarly and 400 ng was transfected into HeLa cells as described in the Transfections section.

Preparation of RNA libraries for next-generation sequencing

500 ng of total isolated RNA from HeLa cells was reverse transcribed to cDNA by Transcriptor First Strand cDNA Synthesis Kit (Roche) according to the manufacturer's protocol. To generate amplicons suitable for Next-generation sequencing, two PCR rounds were performed using Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs). The first PCR with PS03_ill_f.adapt and PS04_ill_r.adapt primers was used to specifically amplify the minigene mRNA transcripts and introduce Illumina adapters at both ends. The program for the first PCR was set as follows: 30 s at 98°C, 20× (10 s at 98°C, 20 s at 60°C, 20 s at 72°C), 30 s at 72°C and 20 s at 20°C. The PCR products were then cleaned up by AMPure XP beads (Beckman Coulter), diluted 100 times and used as a template for the second PCR. The second PCR utilized Illumina NEXTERA primers that annealed to Illumina adapters. The program for the second PCR was similar to the first, with 30 cycles. The PCR products were again cleaned up by AMPure XP beads, the library quality was checked by TapeStation (Agilent Technologies) and the concentration was determined by Qubit 2.0 (Thermofisher Scientific). All samples were then pooled equimolarly and run on NextSeq 500 (Illumina).

Preparation of DNA libraries for next-generation sequencing

The DNA from the three mutagenesis libraries was isolated by QIAprep Spin Miniprep Kit (Qiagen). For next-generation sequencing, the intron between exons 1 and 21 had to be shortened due to read length restriction. To this end, each sample was treated with PacI and AsiSI enzymes and then recircularized by T4 DNA ligase (New England Biolabs). Two rounds of PCR were performed using PrimeSTAR® Max DNA Polymerase (Takara Bio) to generate a suitable amplicon library. The first PCR used PS03_ill_f.adapt and STAT3_DNAseq_1R minigene-specific primers with Illumina adapters at 5′ ends. The program for both PCRs was set as follows: 30s at 98°C, 20× (10 s at 98°C, 20 s at 60°C, 15 s at 72°C), 30s at 72°C and 20 s at 20°C. The DNA was purified by AMPure XP beads (Beckman Coulter) and diluted 100 times. The second PCR employed Illumina NEXTERA primers that annealed to Illumina adapters.

Molecular dynamics (MD) simulations

The interaction between pre-mRNA and U1 snRNA was studied using MD simulations. RNA duplexes consisting of one strand from mRNA and a second one from U1snRNA were built by the NAB module of AMBER 16 (25). The duplexes were built as a right-handed A-RNA using a parmbsc0 OL3 force field (26). Mismatched base pairs were built in *cis*-Watson–Crick *cis*-WC/WC geometry as standard canonical pairs. Force field parameters for pseudouridine were taken from the modified residues library (27) and adjusted according to the force field used. Each RNA duplex was solvated by a periodic octahedral box of TIP3P water molecules extending 10 Å from the solute and neutralized with sodium counterions using AMBER's Xleap module. The equilibration and production setup used in this study corresponds to the standard protocol applied by the ABC consortium for large-scale MD nucleic acid studies (28). Equilibrated duplexes were simulated for 500 ns. The MD trajectories were processed using AMBER's Ptraj module and visualized using the VMD program (29). Stability of RNA duplex conformations was monitored using root-mean-square deviation (RMSD), which indicates the deviation in a superposition of MD structures and the initial A-RNA geometry.

Results

The downstream 5′ss is weakly defined

The tandem donor sequence in *STAT3* exon 21 (Figure 1A), CAG/GTA/GTTGTTGA, gives rise to two overlapping 5′ss motifs. Using a well-established prediction algorithm, MaxEntScan (30), we calculated the scores for the upstream 5′ss (D1) and downstream 5′ss (D2) as 6.3 and –17.47, respectively (Figure 2A). MaxEntScan (MES) score reflects, to a great extent, a given splice site's ability to be recognized by a spliceosome and it is generally assumed that a negative score is a sign of an extremely weak splice site. D1 and D2 also substantially differ in the potential number of base pairs (bp) they can create with U1 snRNA (10 versus 5 bp), which indicates that the D2 binds U1 snRNA much less effectively (Figure 2B). To support this, molecular dynamics was used to simulate the formation of various U1:5′ss RNA duplexes. The duplex formed by D1 and U1 was visually and structurally similar to the duplex formed by extended consensus 5′ss sequence and U1. In contrast, the duplex formed by D2 and U1 did not resemble a typical A-RNA structure nor the D1:U1 duplex and was the least stable structure among the three tested duplexes as illustrated by the highest root-mean-square deviation (RMSD) values (Figure 2C, D). Even though these in silico simulations do not account for U1-specific or other proteins that can stabilize various mismatches in the U1:5′ss duplex (31,32), they point out how distorted the duplex will become over time if not aided by proteins. Recognition of 5′ss by U1 snRNP in yeast is consistent with conformational proofreading model (32). If this also holds true for humans, the more distorted the U1:5′ss duplex becomes, the less likely it is to pass proofreading checkpoints and progress into catalytically active spliceosomal complexes. The huge distortion observed in the D2:U1 duplex suggests a low probability of functional association between endogenous U1 snRNA and D2. Combined, these *in silico* analyses suggest that the D2 donor is extremely weak and might not bind U1 snRNA effectively. This in turn implies that there is only one functional binding site (the D1 motif) for endogenous U1 snRNA in this tandem splice site.

Mutational analysis suggests only a single functional U1 snRNA binding site

To test whether there are one or multiple binding sites for endogenous U1 snRNA, we introduced single nucleotide substitutions into the pET-STAT3 minigene. All positions were numbered relative to the D2's GT dinucleotide, i.e.

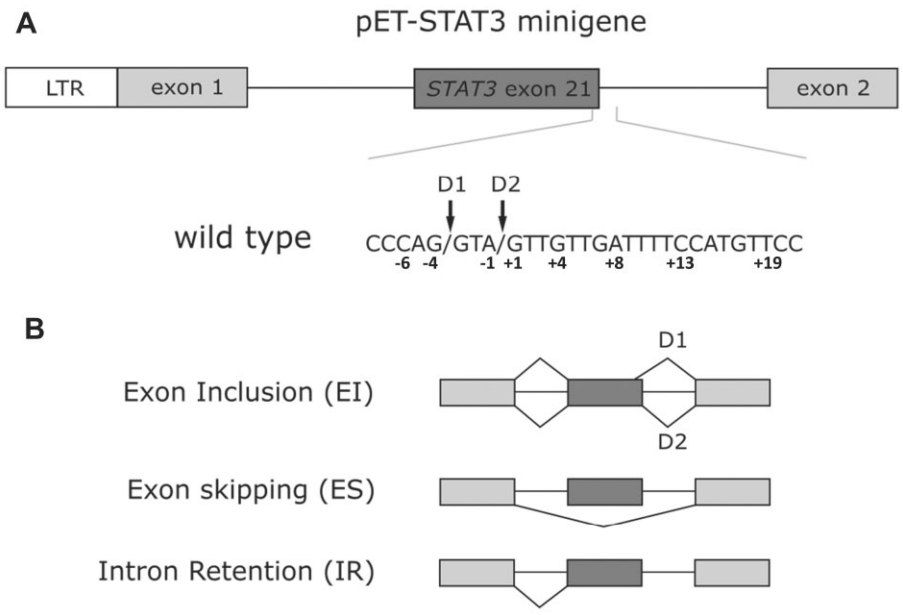
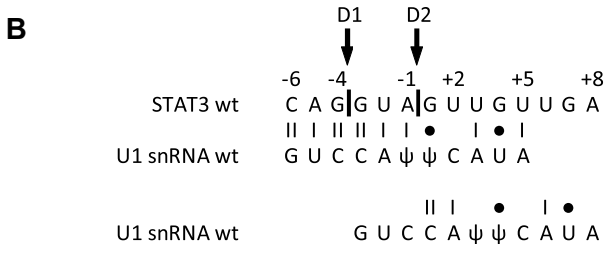
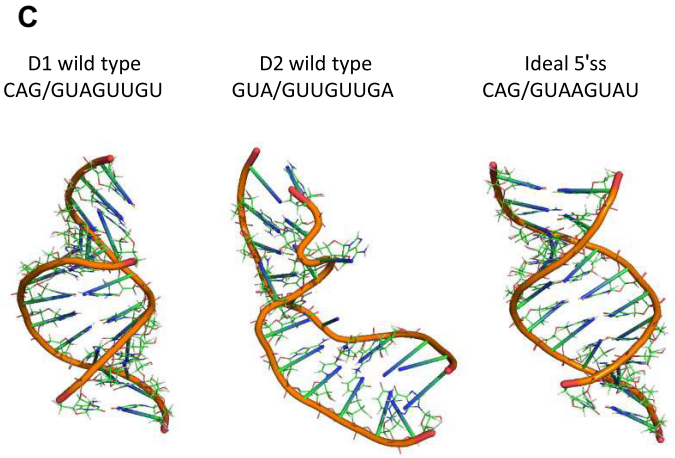


Figure 1. pET-STAT3 minigene used for splicing analysis. (**A**) Scheme of used minigene. (**B**) Transcripts produced from pET-STAT3 minigene. Transcripts including full-length exon 21 that use D1 or D2 donors are grouped under the 'Exon Inclusion' tag when appropriate, otherwise 'D1' and 'D2' tag are used.

A

	MaxEntScan score	Predicted U1 complementarity
D1 wild type	6.3	10 bp (8 WC + 2 wobble)
D2 wild type	-17.47	5 bp (3 WC + 2 wobble)





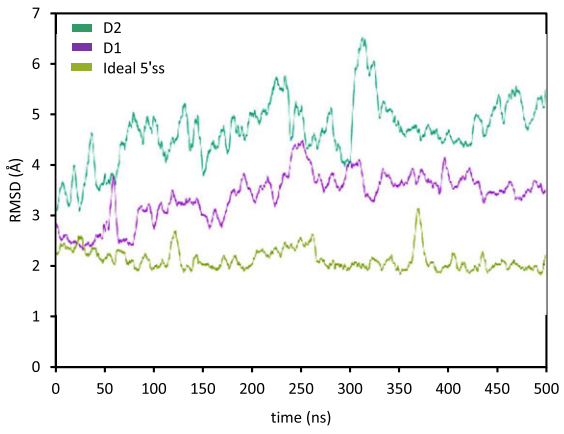


Figure 2. U1 snRNA pairing to the *STAT3* wild type tandem 5′ss in exon 21. (**A**) Predicting 5′ss strength with the use of MaxEntScan and complementarity between endogenous U1 snRNA and either of the two donors (D1 and D2). bp – base pair; WC – Watson–Crick base pair; wobble – wobble base pair (i.e. G:U or G: ψ). (**B**) Alignment and complementarity between endogenous U1 snRNA and both 5′ss (D1 and D2) in the *STAT3* tandem donor splice site. I – A:U pair; II – G:C pair; •, wobble base pair (G:U or G:ψ). (**C**) Molecular dynamics simulation of RNA duplex formation. The structures were obtained at the end of the experiment (500 ns). To generate the structures, two 11-nucleotide long oligomers were annealed, identical sequence to the 5′ end of endogenous U1 snRNA and one of the three different donor sequences – left to right: D1 wild type (CAG/guaguugu), D2 wild type (GUA/guuguuga) and human extended consensus donor motif (CAG/guaaguau). (**D**) Root-mean-square deviation (RMSD) value development during the molecular dynamics simulation.

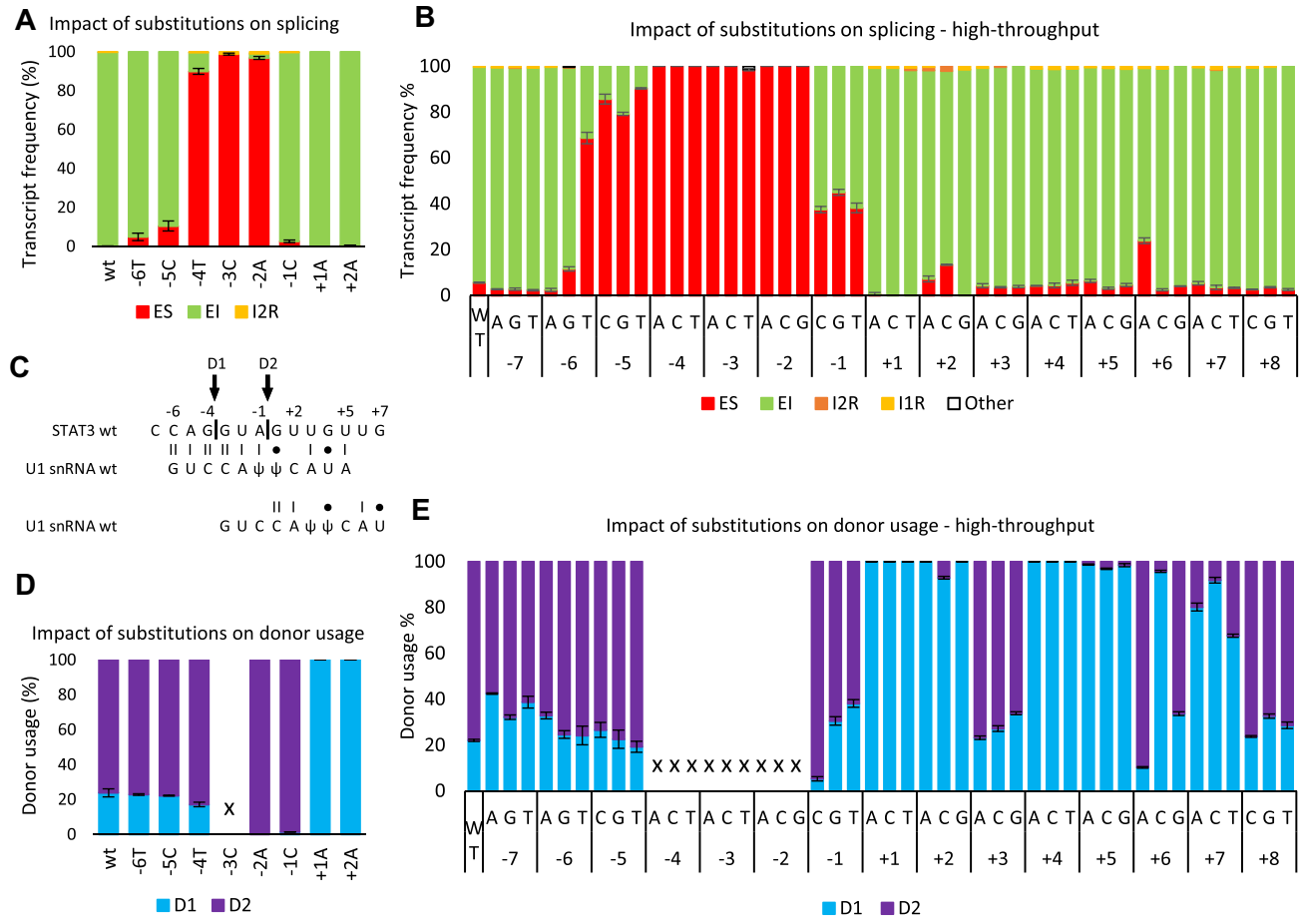


Figure 3. Tandem donor mutagenesis. Impact of single-nucleotide substitutions on *STAT3* splicing in regular pET-STAT3 minigene (A, D) and high-throughput pHT-STAT3 minigene (B, E). Transcripts produced from *STAT3* minigenes. ES, exon skipping; EI, exon inclusion; I1R, upstream intron retention; I2R, downstream intron retention. Impact of single nucleotide substitutions on all transcripts produced by regular (**A**) and high-throughput (**B**) *STAT3* minigenes. Impact of single nucleotide substitutions on donor usage in regular (**D**) and high-throughput (**E**) *STAT3* minigenes. (**C**) Scheme depicting base pairing between *STAT3* tandem 5′ss and endogenous U1 snRNA. (E) Donor usage in mutations in positions –4, –3 and –2 was not measurable (denoted by x) as exon inclusion was 0%. Standard deviation was calculated for exon skipping and D1 usage.

–1/+1 being the D2 exon/intron boundary. Nucleotides in positions important for potential U1 snRNA binding (–6 to +2) were mutated and characterized (Figure 3A, C). Four main spliced transcripts were detected: full-length *STAT3* exon 21 inclusion with the usage of one of the two donor splice sites (exon inclusion, EI), *STAT3* exon 21 skipping (ES), and downstream intron retention (IR) (Figure 1B). As expected, mutations in positions –3, –2 and +1, +2, which disrupted invariant GTs, resulted in non-usage of D1 or D2, respectively (Figure 3B). Mutations in positions –6 to –1 showed increased exon skipping, ranging from 2.6% to 99%. Mutations +1A, and +2A, that disrupted D2's GT dinucleotide, did not increase the exon 21 skipping despite resulting in the non-usage of the D2 (Figure 3). These changes in exon skipping agreed with the predicted base pairing changes between U1 and D1 only, and not with D2 (Supplementary Figure S1A). If there were two or more functional binding sites for U1 snRNA in this tandem 5′ss we would expect a partial redundancy in ensuring proper exon inclusion. However, mutations –3C and –2A that disrupted potential D1:U1 base pairing and improved D2:U1 base pairing resulted in basically 100% aberration. Additionally, mutations in +1A and +2A that disrupted potential D2:U1 base pairing did not affect exon inclusion substantially. This argues against any significant contribution of a potential U1 snRNA binding at the D2 to the exon inclusion. These results imply that the endogenous U1 snRNA either does not bind to the D2 or the U1 snRNA binding to the D2 motif is unimportant for exon 21 inclusion.

To gain a more detailed understanding of how changes in the tandem donor sequence motif affect splicing and GT choice, a specialized pHT-STAT3 minigene suited for high-throughput (HT) splicing analysis was created (Supplementary Figure S2). The 213 nt *STAT3* exon 21 was shortened to 80 nt (due to read length limitations) with minimal influence on wild-type splicing (refer to mutation e_del 3 in Figure 5). We obtained more than 300 different mutations (up to 6n substitutions) in the positions –7 to +8 and termed this library L1. Comparing identical mutations –6T, –5C and –1C in both experiments reveals that the pHT-STAT3 minigene is more susceptible to exon skipping than the pET-STAT3 minigene (Figure 3A, B). Nevertheless, based on simple linear regression analysis, changes in exon inclusion seen in the single nucleotide substitutions were highly correlated with the predicted MES score changes in the D1 ($R^2 = 0.838$, $P << 0.001$) but not in the D2 ($R^2 = 0.072$, $P = 0.168$) (Supplementary Figure S3C). Thus, the MES score of D1 can

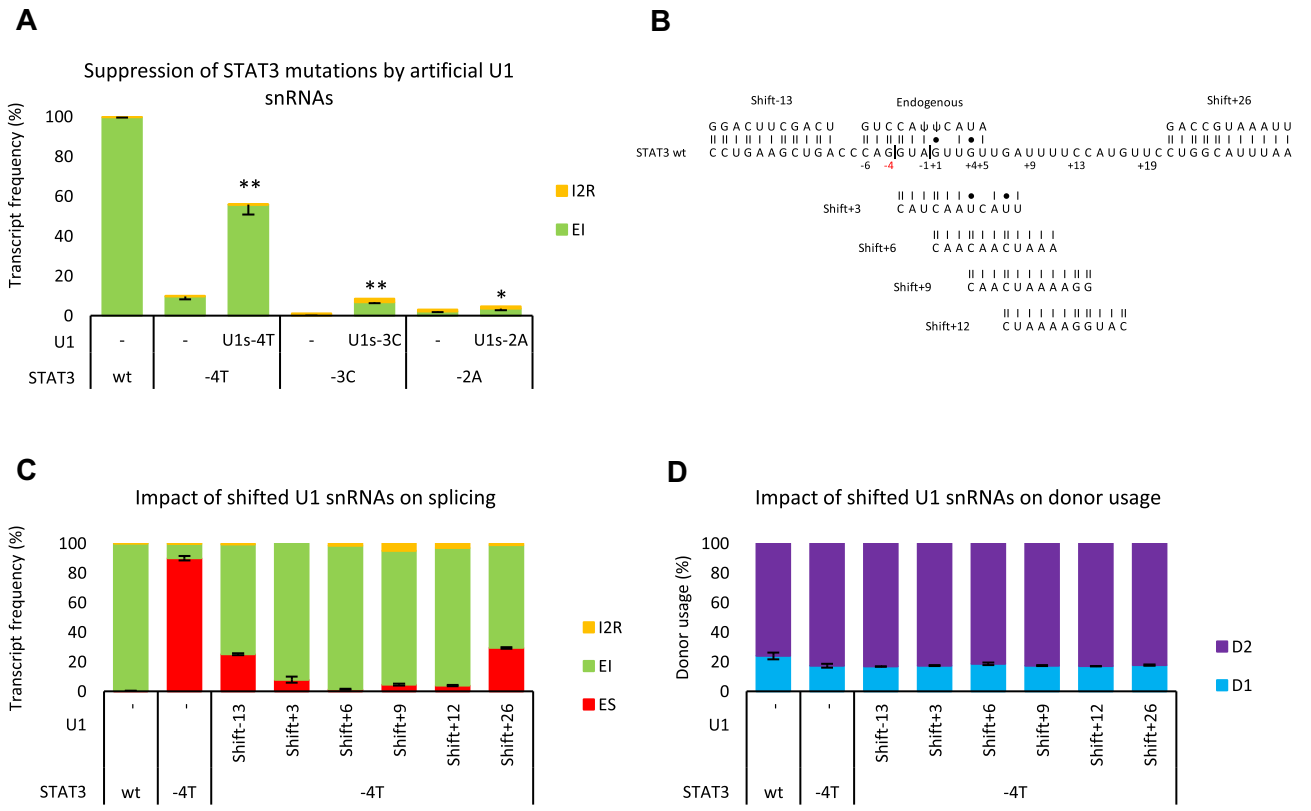


Figure 4. pET-STAT3 mutation rescue by suppressor or shifted U1 snRNAs. (**A**) three *STAT3* mutations rescued by suppressor U1 snRNA that re-established D1 base pairing. **P*-value < 0.05; ***P*-value < 0.01. (**B**) Alignment and base pairing of endogenous and shifted U1 snRNAs with the sequence of wild-type *STAT3* pre-mRNA. II, C:G base pair; I, A:U base pair; •, wobble base pair (i.e. G:U or G:ψ). (**C**) Splicing rescue of –4T mutation by shifted U1 snRNAs. (**D**) Impact of shifted U1 snRNAs on donor use in –4T mutation. Standard deviations are depicted for exon inclusion and D1 usage frequency.

explain 83.8% of the exon inclusion variance. As U1 snRNA binding is a crucial contributor to this score, this again supports the idea that only the D1 and not the D2 binds endogenous U1 snRNA. We also tried to explain the observed exon inclusion and 5′ss usage by changes in potential RNA secondary structure formed at the tandem 5′ss (ΔG, mfold) (21) or in silencer/enhancer activity at the tandem 5′ss using two scoring systems, total HZ_{EI} score (22) and ESRseq score (23). The total HZ_{EI} and ESRseq scores represent the potential of a given sequence to bind splicing enhancer or silencer proteins. Ultimately, none of these three methods were able to explain the observed variance and therefore probably have minimal or no impact on the splicing of *STAT3* tandem 5′ss (Supplementary Figures S16 and S17).

Exon inclusion is rescued by re-establishment of U1 snRNA base pairing to D1 motif

The splicing of human 5′ss is not always dependent on the presence of U1 snRNP and in some cases, the depletion of U1 snRNP can be overcome by the addition of high concentrations of SR proteins (33–37). Interestingly, *Cyanidioschyzon merolae* might even lack the U1 snRNP and associated proteins completely and the spliceosome assembly might begin with the recognition of 5′ss by tri-snRNP (38). Therefore, it is theoretically possible that splicing of human *STAT3* tandem 5′ss is U1-independent. To exclude this possibility and confirm that the endogenous U1 snRNA plays a role in the splicing of *STAT3* tandem 5′ss, three suppressor U1 snRNAs (U1s) were designed to suppress –4T, –3C and –2A *STAT3* mutations in the pET-STAT3 minigene. Each U1s harboured a single nucleotide change in its 5′ end but otherwise was identical to the endogenous wild-type U1 snRNA. These suppressor U1 snRNAs re-established base pairing for the respective *STAT3* mutant only if bound across the D1 (Supplementary Figure S1B). All three U1 snRNAs rescued exon inclusion, although, U1s-3C and U1s-2A did so only slightly but still significantly (Figure 4A). Molecular dynamics simulations supported the minigene results and showed that suppression with U1s substantially decreased average D1:U1 duplex RMSD only in the -4T mutation (Supplementary Figure S4). Additionally, the RNA duplexes formed by –3C (D1) and –2A (D1) *STAT3* mutants with their respective U1s showed similarly distorted structures to duplexes formed with endogenous U1 snRNA and thus might explain the low observed rescue of –3C and –2A mutants. The –3C and –2A mutations also disrupted D1 dinucleotide which might further prevent progression to the later spliceosomal complexes resulting in spliceosome disassembly and high exon skipping even in the presence of suppressor U1 snRNAs. Since all suppressor U1 snRNAs differed from endogenous U1 snRNA only in a single nucleotide at the 5′ end, it is highly probable that it is the endogenous wild-type U1 snRNA that binds to the *STAT3* tandem 5′ss. Taken together, the information and results presented so far—*in silico* predictions, molecular dynamics simulations, single nucleotide mutations and their rescue by U1s, and high-throughput experiments—can be best explained by the presence of only one functional binding site for endogenous U1

snRNA which is located directly on top of the D1 in the *STAT3* tandem donor 5′ss.

The binding position of U1 snRNA does not regulate 5′ss choice within the tandem donor

As there is only one functional binding site for endogenous U1 snRNA exactly at the D1, preferential D1 usage over D2 would be expected due to D2 being shifted three nucleotides downstream. In reality, the D1 usage in the wild type STAT3 minigenes was only 22–24% (Figure 3D, E), which suggested that the U1 snRNA binding site was not the main determinant of 5′ss choice. To test how the exact U1 snRNA binding position affects the splicing in this tandem donor, we designed shifted U1 snRNAs that bound to the region around the tandem donor. All but one (Shift +3) of the shifted U1 snRNAs were fully complementary (11 base pairs) to the pET-STAT3 sequence and the distances varied from –13 nucleotides upstream to +26 nucleotides downstream from the D1 (Figure 4B). The shifted U1 snRNAs were co-transfected with the –4T mutant, instead of wild type STAT3 minigene. This experimental setup enabled us to detect an increase in exon inclusion upon successful shifted U1 snRNA binding. All shifted U1 snRNAs rescued exon inclusion proportionally to its distance from the D1 up to 97%. None of the applied shifted U1 snRNAs changed the D1 usage significantly, and in fact its usage stayed consistently around 18% in these experiments, virtually the same as the -4T mutant (Figure 4D). Importantly, the shift +3 U1 snRNA that imitated functional binding of U1 snRNA directly on top of D2 (base pairing in D2:shift +3 U1 was equal to the D1:U1 wt) restored exon inclusion to near wild type level but did not affect splice site choice. This demonstrates that even if there was a functional binding of U1 snRNA at the D2 it would not affect the D2 usage.

To test if D1 usage is not generally dependent on the U1 binding site, the L1 library was co-transfected separately with three different artificial shifted U1 snRNAs and control U1 snRNA (identical to endogenous U1 snRNA). These experiments recapitulated the splicing pattern seen in the pET-STAT3 minigene (Figure 4, Supplementary Figures S5, S6). Specifically, for a given mutation, the D1 usage remained virtually identical regardless of U1 snRNA binding position (Supplementary Figure S6). Additionally, the changes in exon inclusion seen in the individual HT experiments could not explain the changes in D1 usage (Supplementary Figure S7). This strongly suggests that the functional U1 snRNA binding position does not regulate the 5′ss choice in the *STAT3* tandem donor site. Overall, all our experiments argue that the *STAT3* tandem 5′ss is dependent on U1 snRNA that binds exclusively to the D1 5′ss and that there is another mechanism that regulates choice between D1 and D2.

The downstream regulatory region enables D2 usage

We hypothesized that splicing regulatory elements might be involved in the 5′ss choice. To screen for potential splicing elements, multiple large and small deletions were introduced into the exon 21 and the downstream intron (Figure 5A). Only deletions targeting the immediate downstream region, named downstream regulatory region (DSR), affected splicing. These deletions (denoted DSR_del) did not affect exon inclusion but greatly diminished the D2 usage (down to 0%) suggesting this region is essential for the use of D2 but not for exon inclusion (Figure 5B, C). The inclusion of the exon 21 therefore does not seem to require additional splicing elements. To better understand the importance of this region, a new L2 library for HT analysis was generated by mutating positions from +9 to +19 in the pHT-STAT3 minigene (Supplementary Figure S8A). The L2 library contained around 500 mutations, mostly double and triple substitutions, and displayed highly variable D2 usage ranging from 0 to 92% (Supplementary Figure S8B). For further analysis, we only considered mutations with up to triple change. Exon inclusion ranged only from 79.7% to 100% and did not correlate with D2 usage (Supplementary Figure S8C). While only 10% of mutations resulted in D2 usage higher than in the wild type, 51% of mutations showed a higher exon inclusion rate than the wild-type minigene (Supplementary Figure S8B).

Single and double mutations in nucleotides +11 to +14 caused the biggest decrease in D2 usage, while most mutations in positions +15 and further caused only little change in D2 usage (Figure 5D, E and Supplementary Figure S9B, C). Comparing the nucleotide frequency of two mutation subsets, the top 10 and bottom 10 percentile based on D2 usage revealed a strong preference for cytidine in the positions +13 and +14 and a preference for thymidine in positions +10 and +11 (Figure 5E, F and Supplementary Figure S10). Nucleotide over/under-representation in a particular position thus reflected a positive/negative preference for certain nucleotides (Figure 5F). Most mutations in the positions +15 to +19 modulated the D2 usage only slightly (Supplementary Figure S9B, C). Overall, these experiments demonstrate that, at least in the presence of the wild-type D1, the DSR is essential for the D2 usage and that most introduced mutations decreased the D2 usage.

STAT3 tandem 5′ss is evolutionarily conserved

While little is known about the importance of this alternative splicing event in exon 21 in humans and other mammals, and the role of the two protein isoforms, we have some clues to think that it is essential for the proper function of STAT3 protein. The genomic alignment of 24 primate species and 16 non-primate mammalian species revealed high evolutionary conservation in the last exonic nucleotides and downstream intronic region (Supplementary Figure S19A, B). There are occasional variations in positions -5, +3, +9 and +14 and substantial variation could be found in the further downstream region starting at positions +15 and further (Supplementary Figure S19A, B). Interestingly, all but one variation (–5G) in the region –6 to +19, that we could find in our high-throughput data, showed no substantial change in exon inclusion or D2 usage (±6%) (Supplementary Figure S19C, D). However, since the pHT-STAT3 minigene used for high-throughput analysis was more susceptible to exon skipping, we expect that the –5G should behave similarly to the –5C which displayed wild-type-like exon inclusion (90%) in regular pET-STAT3 minigene. It is therefore highly possible that the *STAT3* tandem 5′ss is spliced similarly in all these mammals and the expression of both transcript variants is somehow important for general gene function.

The importance of distance between D1, D2 and downstream regulatory region

We hypothesized that the distance between D1, D2 and the DSR might affect the splicing of the tandem donor site. To

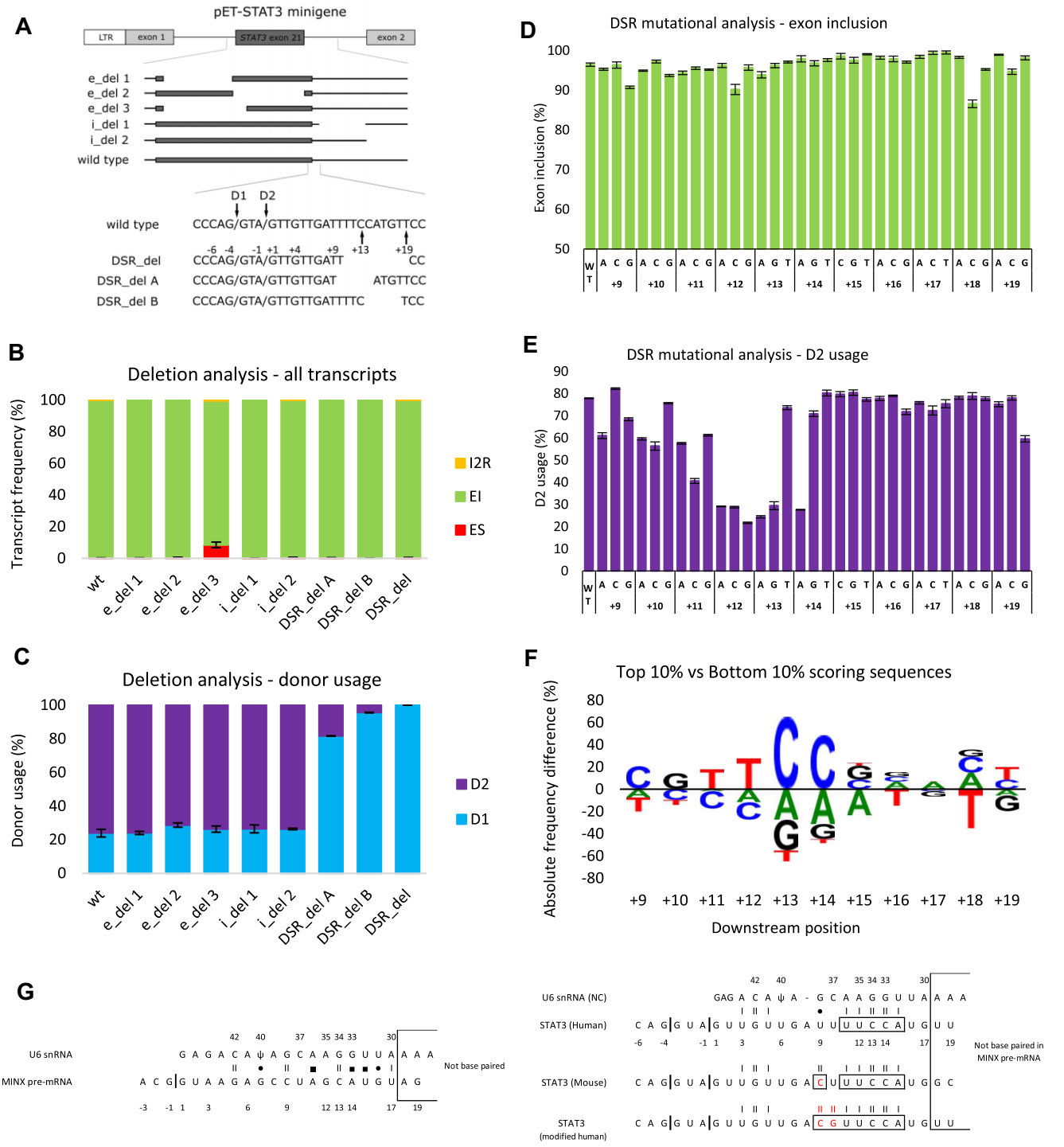


Figure 5. Identifying D2 usage drivers by mutational analysis. (**A**) Deletions introduced into pET-STAT3 minigene. **(B, C)** Impact of deletions on splicing isoforms and donor usage in pET-STAT3 minigene. ES, exon skipping; EI, exon inclusion; I2R, downstream intron retention. Standard deviation calculated for exon skipping and D1 usage. (D–F) Mutational analysis of DSR in high-throughput minigene. Analysis of single nucleotide substitutions **(D, E)** and up to triple substitutions (F). (**F**) Nucleotide frequency differences between two mutated DSR sequence subgroups sorted by D2 usage: top 10% versus bottom 10% scoring sequences. Generally, the letters above the y-axis (0) represent nucleotides supporting D2 usage, while the letters below the y-axis represent decreasing D2 usage. (**G**) Extended pairing between human U6 snRNA and *MINX* pre-mRNA based on the Cryo-EM structure 6AHD. Proposed extended base pairing between U6 snRNA and *STAT3* 5′ss in mouse or human. NC, non-canonical register; II, G:C base pair; I, A:U base pair; •, wobble base pair (i.e. G:U or G:ψ); ■, non-Watson–Crick base pair.

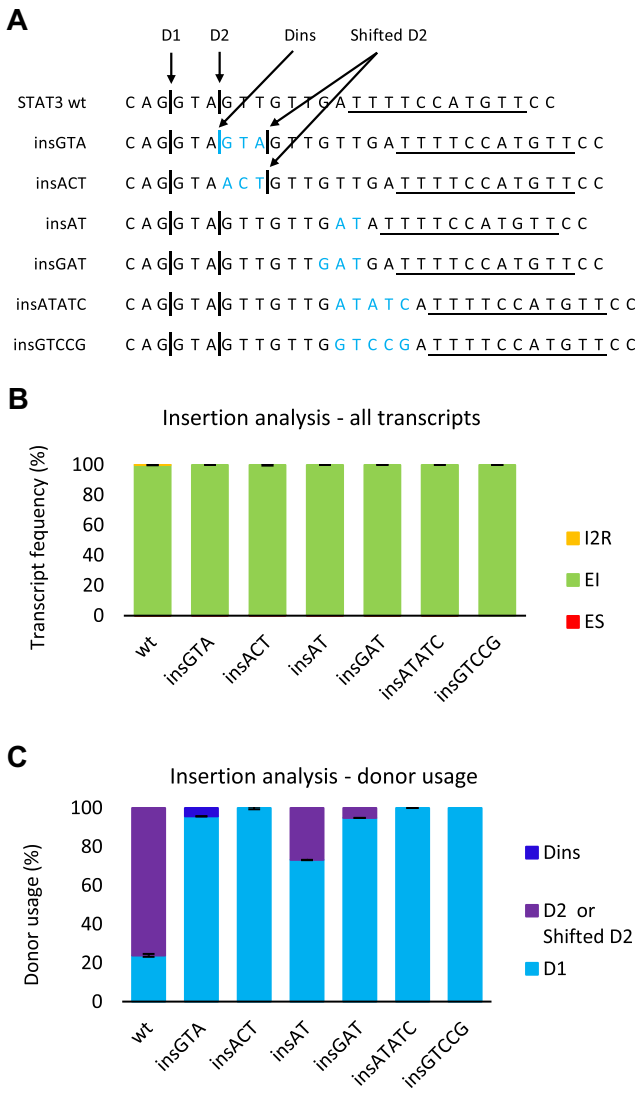


Figure 6. Impact on D2 usage by shifting the DSR. (**A**) Insertions introduced into regular pET-STAT3 minigene. Underscored nucleotides indicate DSR studied in previous high-throughput experiments. (**B**) Impact of insertions on splicing and (**C**) D2 usage. ES, exon skipping; EI, exon inclusion; I2R, downstream intron retention. Error bars represent standard deviations; Dins, Donor inserted in insGTA mutant; D2 or shifted D2, the original or shifted D2 donor.

test this hypothesis, up to five nucleotides were inserted between these elements in the pET-STAT3 minigene so as not to disturb the U1 snRNA and U6 snRNA base pairing (Figure 6A, Supplementary Figure S11). In the insGTA and insACT mutants, three nucleotides were inserted between positions –1 and +1. This change increased the D1–D2 and D1–DSR distance by three nucleotides. No change in exon inclusion or intron retention was observed, but both mutants showed complete non-usage of the shifted D2 (Figure 6B and C). Next, up to five nucleotides were inserted between intronic positions +7 and +8, which left the D1-D2 distance unchanged but altered the D1–DSR and D2–DSR distance (Figure 6A, Supplementary Figure S11). Again, no significant change in exon inclusion or intron retention was observed, and all mutations decreased the D2 usage substantially (Figure 6B, C). The D2 usage gradually decreased down to 0% with an increasing insertion length. Overall, these experiments showed that the distance between all three elements (D1–D2, D1–DSR and D2–DSR) significantly affected the D2 usage.

Evidence for a novel non-canonical U6 snRNA binding register that drives D2 usage

After the initial recognition of 5′ss by U1 snRNP, the U1 snRNA interaction is replaced by U6 snRNA which ultimately leads to spliceosome activation and definition of the 5′ss cleavage site (39). While multiple alternative pairing registers of U1 snRNA with 5′ss have been reported (13–15), no alternative pairing registers of U6 snRNA with 5′ss have yet been described. Human 5′ss consensus motif has another GT in positions +5 and +6 that base pairs with U6 snRNA positions 42 and 41, which is important for splice site choice (40,41). Interestingly, the D2 in the *STAT3* tandem donor has a GT shifted one nucleotide upstream to the positions +4 and +5 (Figure 7A). All mutations in positions +4 and +5 significantly decreased the D2 usage to nearly 0% (Figures 3D, 7C, E). This was surprising, since even mutations +4A and +5G that introduced consensus nucleotides into the D2 almost eliminated the D2 usage. These results would make sense if U6 snRNA nucleotides 42C and 41A (that usually base pair with donor positions +5G and +6U) base paired with D2 positions +4G and +5U in a non-canonical (NC) register (–1 nucleotide upstream) (Figure 7A). For this non-canonical pairing, the +4 and +5 positions already match the consensus 5′ss nucleotides and any mutation in these positions would disrupt the U6 snRNA base pairing with the D2, thus decreasing its use.

To test the non-canonical register hypothesis, we designed multiple artificial U6 snRNAs (artU6s) and applied them individually to each +4 and +5 mutant. A group of artU6s specific for each mutant contained: at least one suppressor U6 snRNA that re-established base pairing with the D2 in the NC register (Supplementary Figure S12), one U6 snRNA that improved base pairing with the D2 in canonical register and two non-suppressor U6 sRNAs that should not rescue the D2 usage. Compared to the endogenous U6 snRNA, all these artificial U6 snRNAs harboured only a single nucleotide substitution (Supplementary Figure S12). A particular artU6 can be considered a suppressor or non-suppressor depending on the +4 or +5 mutant it is applied to. All U6 snRNAs that have been designed to suppress a specific mutation in the shifted register successfully increased the D2 usage ($P < 0.01$) (Figure 7C, E). Conversely, artU6s that improved the canonical base pairing with the D2 in +4 and +5 mutants failed to improve the D2 usage. When non-suppressor U6 snRNAs, which should not impact D2 usage, were applied to appropriate +4 or +5 mutants, D2 usage remained unchanged except when U6 41C was paired with a +5C mutation. In this case, the D2 change was statistically significant ($P < 0.01$) but the magnitude was much lower (+ 3.8% versus + 32% difference) than the suppressor U6 41G. The low rescue of D2 usage by U6 42A in the +4T mutant (Figure 7C) may be explained by a simultaneous increase in base pairing between the U6 42A and D1 in the canonical register (Supplementary Figure S12B).

As failure to rescue the D2 usage can also result from non-functional artU6 (e.g. insufficient incorporation into spliceosomes or defective interaction with other spliceosomal components), the proper functionality of suppressor U6 snRNA variants (U6 41 and U6 42) was tested. To this end, we modified the D2 intronic sequence (mD2) and forced canonical

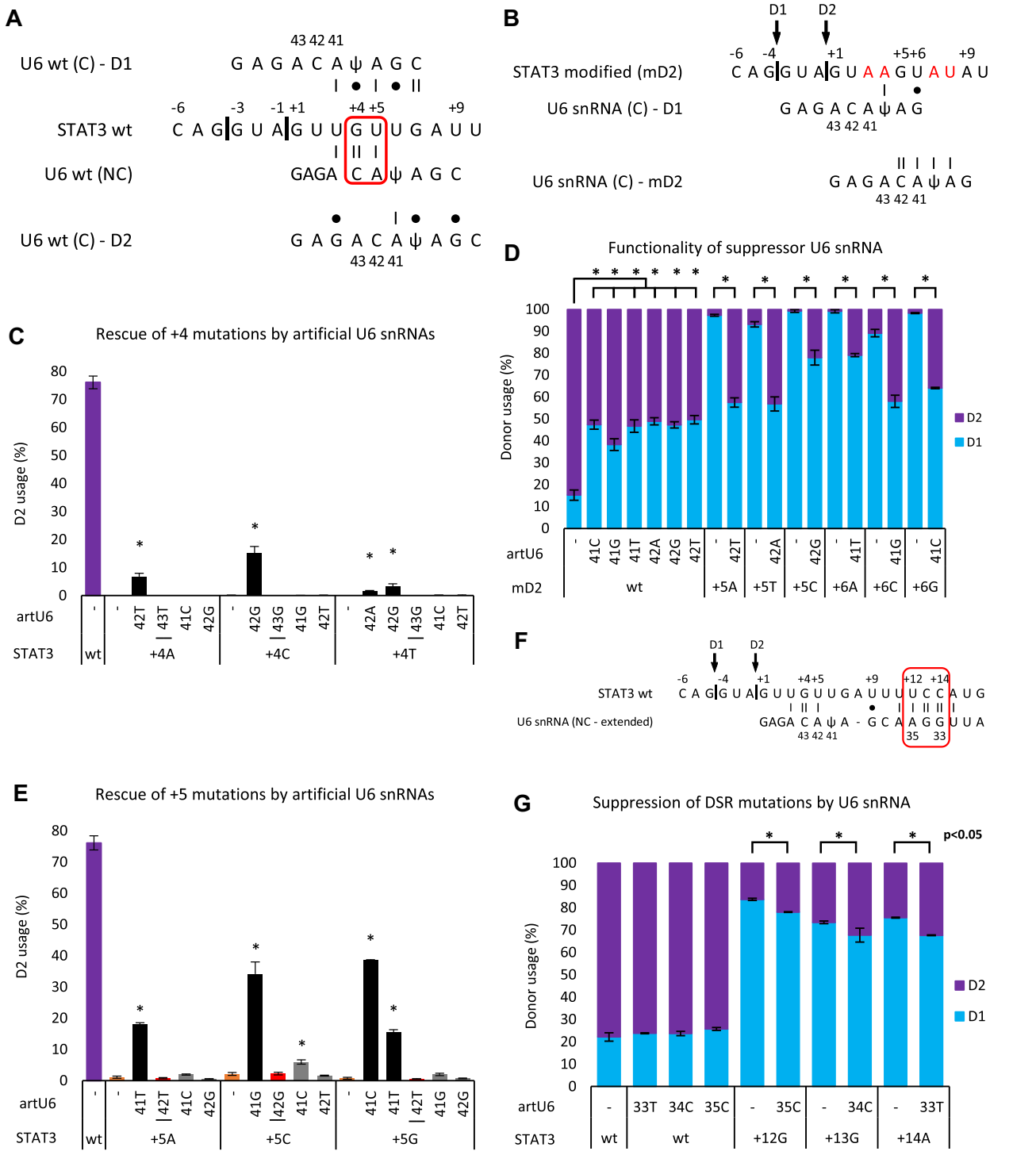


Figure 7. Rescue of +4 and +5 mutations in pET-STAT3 minigene by artificial U6 snRNAs. (**A**) Alignment and base pairing between endogenous U6 snRNA and the wild type *STAT3*. (**B**) Depiction of modified *STAT3* tandem 5′ss and base pairing between endogenous wild type U6 snRNA and both 5′ss. Red letters denote introduced mutations from wild-type *STAT3*. **(C, E)** D2 usage rescue in +4 and +5 mutants. +4 or +5 mutant with no overexpressed artU6 (orange bars), suppressor U6 snRNAs (black bars), artU6 that improve canonical base pairing in D2 (red bars), non-suppressor U6 snRNAs (grey bars). **P* was generally < 0.01, in (G) *P* < 0.05. U6-41C was the only non-suppressor U6 snRNA that produced a statistically significant (*P* < 0.01) change in D2 usage when applied to the +5C mutant. The magnitude of change was nevertheless much smaller than that produced by suppressor U6-41G. (**D**) Testing the functionality of suppressor U6 snRNAs (mutants in positions 41 and 42) in modified mD2 STAT3 minigene. (**F**) Depiction of potential extended base pairing between U6 snRNA and *STAT3* tandem 5′ss. Rectangles denote positions that were mutated in suppressor experiments. C, canonical register; NC, non-canonical register; II, G:C base pair; I, A:U base pair; •, wobble base pair (i.e. G:U or G:ψ). (**G**) Suppression of mutations in DSR by artificial U6 snRNA in the context of wild-type STAT3. * *P*-value < 0.05.

alignment with endogenous U6 snRNA (Figure 7B). This modification distinguished D1 and D2 sequences, preserved similar D2 usage and allowed us to attribute the changes in mD2 usage to changes in U6:mD2 base pairing. Figure 7D shows that the application of all U6 41 and 42 mutants to this modified minigene decreased mD2 usage, which was consistent with the disruption of canonical alignment in mD2. Next, we introduced the equivalent of +4 and +5 mutations into the mD2. The mutations were now in positions +5 and +6 and disrupted canonical U6 alignment. Importantly, all these suppressor U6 snRNAs robustly rescued the mD2 usage for the mutation they were designed to suppress. This result was again consistent with the improvement in mD2:U6 base pairing and confirmed that these artificial U6 snRNAs were functional.

Similarly, none of the artU6s used in these experiments improved the D2 usage when co-transfected with the unmodified pET-STAT3 wild-type minigene (Supplementary Figures S13, S14). Additionally, suppressor U6 snRNAs that improved the D2 donor usage in their respective mutants failed to significantly improve the D2 usage in other mutants they were not designed to suppress (Figure 7C, E). This indicates that the rescue of D2 usage seen in +4 and +5 mutants after applying suppressor U6 snRNAs is specific for the mutation and not due to a non-specific or global suppressor effect. These experiments with artificial U6 snRNAs suggest that the D2 in the *STAT3* exon 21 employs a novel non-canonical interaction with endogenous U6 snRNA to ensure its efficient usage.

Evidence for a functional extended base pairing between U6 snRNA and D2

Recently, an extended base pairing between U6 snRNA and 5′ss has been demonstrated for MINX pre-mRNA using Cryo-EM (42,43) (Figure 5G). In the structure of the human spliceosomal pre-catalytic B complex, the human U6 snRNA aligned and base paired with the MINX 5′ss up to the downstream intronic position +17 (6AHD). This structure has also been reported in the C* complex (44,45). However, the function and importance of this extended pairing remain to be determined. Based on the mutational analysis of DSR in *STAT3*, we identified a possible extended base pairing between U6 snRNA and D2 (Figure 5E, F). The proposed base pairing involves pre-mRNA positions from +11 to +15 and U6 snRNA positions from 36 to 32 (Figure 5G). Moreover, +9C and +10G mutations also supported D2 usage (Figure 5F), which is consistent with the creation of +9C:38G and +10G:37C base pairs. We further observed that nucleotides supporting D2 usage in pre-mRNA positions +9 to +14 could establish Watson–Crick (WC) base pairing with U6 snRNA (Figure 5F, G). In double substitutions, +9C and +10G mutants displayed the least amount of detrimental effect on D2 usage compared to other mutations in the same positions, which is consistent with new C:G/G:C base pairing creation (Supplementary Figure S9). Similarly, +13T and +14T mutations that replaced a C:G with a wobble U:G base pair showed the least detrimental effect on D2 usage (either separately or in combination with other mutations) compared to other mutations in the same positions. To test the extended base pairing hypothesis, three different mutations (+12G, +13G and +14A) that disrupted WC base pairing in U6:D2 and decreased D2 usage were separately introduced into the pET-STAT3 minigene (Figure 7F, G). These mutations could indeed be rescued by suppressor U6 snRNAs (U6 35C, 34C and 33T) by restoring the disrupted base pairing. In all three cases, the D2 usage was improved in a statistically significant way. Finally, we tested whether improvement in non-canonical U6:D2 interaction could rescue insertion or deletion mutations in the DSR. Indeed, the D2 usage in compound mutants +5C insGAT and +5G insGAT, which combined shifting of DSR (Figure 6B) and non-canonical U6 snRNA interaction disruption (Figure 7E), could partially be rescued by applying suppressor U6 snRNA (non-canonical register) (Supplementary Figure S15). However, the suppressor capacity of artU6 was lost following DSR deletion (e.g. +5C DSR_del) (Supplementary Figure S15). These data showed that improvement in base pairing in the non-canonical U6:D2 register can partially rescue mutations in the DSR unless this region is completely disrupted. Based on this data, we propose that the downstream regulatory region stabilizes the U6:D2 non-canonical binding register by establishing extended Watson–Crick base pairing with U6 snRNA in the pre-mRNA positions +11 to +14.

Discussion

Initially, splicing of tandem 5′ss and 3′ss used to be considered a simple noise tolerated by cells (46,47). However, functional diversification in proteins encoded by both splicing isoforms (48) and high evolutionary conservation in flanking intronic regions (49) point to the importance of these alternative splicing events. Recently, our group analysed the impact of exact nucleotide composition on the splicing of Δ3 tandem 3′ss (NAGNAG) (50) and demonstrated that the NAG choice is also affected by the upstream 5′ss sequence. In the current study, we analysed the impact of nucleotide composition on 5′ss choice in the Δ3 tandem 5′ss. We presented an in-depth splicing analysis of Δ3 tandem 5′ss in exon 21 of *STAT3* and provided evidence for a novel regulatory mechanism of 5′ss choice.

Splicing of *STAT3* tandem 5′ss might be evolutionarily conserved

The sequence of *STAT3* tandem 5′ss is evolutionarily conserved in mammals and our high-throughput data suggests that the splicing pattern might also be preserved (Supplementary Figure S19). Specifically, all variations and partially matching substitutions in the region -6 to +19 showed wild-type-like exon inclusion and D2 usage (Supplementary Figure S19). Therefore, the mechanism of *STAT3* tandem 5′ss recognition and regulation might be conserved in these mammalian species as well. For example, mice and humans differ in three positions (+9C, +18G and +19C) and have a very similar ratio of the Δ3 variants (51). Interestingly, the ratio of STAT3 ΔS/S variants is remarkably stable across 16 human tissues and is more constant than the STAT3 α/β ratio which was reported to impact tumorigenesis (52). Furthermore, re-expression of ΔS and S isoforms in pairs (ΔSα + Sα or ΔSβ + Sβ) or all four splicing isoforms simultaneously led to better survival of the ABC subtype of human diffuse large B-cell lymphoma cancer cell line (ABC DLBCL) (19). This analysis suggests that the exact ratio of ΔS/S is important for the basic function of all cells.

Role of endogenous U1 snRNAs in *STAT3* tandem 5′ss

The binding of U1 snRNA to the 5′ss motif is crucial for the use of a given 5′ss as illustrated by the consensus motif MAG/GTRAGT in humans. This should also be true for the use of tandem 5′ss. Generally, in tandem donors, the 5′ss with the best complementarity to the U1 snRNA is preferred (data pulled from the database of tandem splice sites, TassDB2) (53). Interestingly, the situation in human *STAT3* tandem 5′ss is reversed and the D2 is used preferentially despite it being weakly defined, as only the D1 has substantial complementarity to U1 snRNA.

There are 16 variant U1 snRNAs (vU1) in humans (54,55). Although expressed at very low levels in differentiated cells (∼0.02% of wild-type U1 snRNA), they are upregulated in human embryonic stem cells (hESC) and were implicated in the maintenance of pluripotency and pathology of neuromuscular disease (56). Six of these vU1s have unique variations in the 5′ ends that base pair with 5′ss and four of those potentially improve base pairing with *STAT3* D2 5′ss (Supplementary Figure S18A and B). However, at best only one additional base pair is created in D2:vU1 compared to D2:U1 (wt) (Supplementary Figure S18C and D). The base pairing in D2:vU1 is always predicted to be worse than in D1:U1 (wt) duplex. Additionally, the situation for vU1s is very similar to the wild-type U1 snRNA because the base pairing in D1:vU1 duplex is always predicted to be much better than in D2:vU1. Importantly, co-transfection of shift +3 U1 snRNA, which imitated functional binding of highly complementary vU1 (equal to D1:U1 wt) directly on top of D2 did not affect D2 usage at all (Figure 4). Therefore, variant U1 snRNAs are not expected to play a role in splicing of *STAT3* tandem 5′ss.

We ultimately provided evidence that endogenous wild-type U1 snRNA most likely bound only to the D1 and was responsible for recognizing the tandem 5′ss as one unit. This was mainly argued by suppressor U1 snRNAs that rescued exon inclusion by re-establishing base pairing with the D1 motif (Figure 4A). In silico predictions, molecular dynamics simulations and linear regression analysis of high-throughput mutagenesis were also consistent with only a single functional bindings site for endogenous U1 snRNA. Specifically, the majority of exon inclusion changes could be explained by changes in D1 and not D2 5′ss strength (Supplementary Figure S3C and D) as indicated by the MES score. Additionally, predicted RNA secondary structure or changes in enhancer/silencer properties at the tandem 5′ss sequence had very low explanatory power in general as well (Supplementary Figure S16 and S17). Further experiments with shifted U1 snRNAs confirmed that the exact U1 snRNA binding position did not affect the splice site choice in *STAT3*. Our findings are consistent with the well-established notion that U1 snRNA only roughly defines the 5′ss region and the cleavage site is then defined by tri-snRNP components U5 and U6 snRNA (57–59). The flexibility in the U1 snRNA binding to the 5′ss has previously been proposed and demonstrated (11,13–15). Similar to our work, shifted U1 snRNA binding was used to rescue the splicing of several genes (60).

Novel non-canonical binding register between U6 snRNA and D2

While U1 snRNA binding to 5′ss can accommodate non-canonical binding registers (13–15) the base pairing of U6 snRNA and 5′ss was not thought to be very flexible as proper alignment is crucial for splice site choice (40,57). Additionally, intronic 5′ss nucleotides +4U, +5G and +6U were confirmed to form base pairs with U6 snRNA positions 49A, 48C and 47A in yeast (corresponding to 43A, 42C and 41A in human U6 snRNA) (41,61). The 5′ss cleavage site always occurred 5 nucleotides upstream from the +5G:48C (yeast) and +5G:42C (human) 5′ss:U6 interaction. Recent Cryo-EM structures of human spliceosome also reported canonical alignment for the pre-mRNAs studied (39,42–45,62). However, these studies used 5′ss that could not even theoretically establish a non-canonical binding register with U6 snRNA. Here we provide the first evidence that the cleavage site can also occur four nucleotides upstream from the proposed 5′ss:U6 42C interaction.

Even though we did not directly assay the binding of U6 snRNA, our experiments with suppressor U6 snRNAs provide compelling evidence that the D2 employed a new non-canonical binding register with U6 snRNA to ensure its proper use. Firstly, mutations in the STAT3 D2 intronic positions +4 and +5 but not in the position +6 eliminated the D2 usage (Figure 3E). Secondly, all artificial U6 snRNAs designed to rescue the disrupted base pairing in the non-canonical register improved D2 usage. Thirdly, these U6 snRNAs decreased (or did not alter in one case) the D2 usage when applied to wild-type *STAT3* minigene, and to the modified D2, (Figure 7D and Supplementary Figure S14) demonstrating the suppression was specific for mutations with disrupted D2:U6 interaction. Lastly, all but one artificial U6 snRNA designed not to improve the non-canonical pairing register indeed failed to improve the D2 usage.

The overall shape (and not just the base pairing) of 5′ss:U6 might be especially important near the catalytic centre of the spliceosome due to multiple pre-mRNA, snRNAs and protein interactions guiding the spliceosomal transitions and transesterification reactions (63,64)). A particular D2:U6 duplex conformation might therefore be required as the non-canonical pairing register in *STAT3* makes it more difficult for the pre-mRNA to reach the catalytic centre. Specifically, if we considered the U6 snRNA structure unchanged, *STAT3* D2 would need one more nucleotide to physically reach the catalytic centre as the D2:U6 interaction is essentially shifted by one nucleotide.

Extended base pairing between U6 snRNA and downstream regulatory region

Recent Cryo-EM structure of human spliceosomal pre-catalytic B complex assembled on MINX pre-mRNA has shown an extended base pairing between U6 snRNA and 5′ss (42,43). This extended structure has also been reported in the C* complex (44,45). The structure showed that the human U6 snRNA aligned with MINX 5′ss up to the +17 downstream intronic position (6AHD). Interestingly, not all positions were base paired or formed WC base pairs and the U6 snRNA position C37 bulged out from the RNA–RNA duplex (Figure 5G). While the function and importance of the extended pairing remain to be determined, there is evidence that the increased U6 snRNA and 5′ss complementarity in the positions +8 and +9 does improve 5′ss use (57). As the dataset of all human 5′ss shows no significant conservation past the intronic position +6, we expect that the alignment and exact

shape of the extended U6:5′ss duplex might change based on the sequence identity of the 5′ss.

There are multiple lines of evidence showing that the DSR in the *STAT3* tandem 5′ss establishes extended base pairing with U6 snRNA. The DSR is located in the same intronic region as the extended base pairing with U6 snRNA observed in the Cryo-EM structure (42). Furthermore, mutational analysis of the DSR revealed a preference for CGTTCC in the positions +9 to +14 (Figure 5F). This motif nicely corresponded to a possible extended WC base pairing with U6 snRNA in positions 33 to 38 (Figure 5G). This was further supported by double substitutions in positions +9, +10, +13 and +14, where mutations that introduced a new G:C or converted G:C to G:U pair displayed the least amount of detrimental effect on D2 usage compared to other mutations in the same positions (either separately or in combination with other mutations) (Supplementary Figure S9). Overall, our data support the presence of functional extended WC base pairing between *STAT3* D2 and U6 snRNA. We propose that this interaction in the region from +11 to +14 further improves the stability of the novel non-canonical D2:U6 interaction.

Splice site choice and competition between D1 and D2

Both D1 and D2 lack G in their relative +5 intronic position which predisposes them to be weakly recognized by U6 snRNA in a canonical register. Generally, this would result in the inefficient conversion of pre-catalytic spliceosomal complexes into catalytic ones as recognition of the intronic 5′ss part by U6 snRNA is thought to trigger spliceosome activation (39). In D2, this is possibly offset by establishing non-canonical U6:5′ss interaction and the presence of downstream regulatory region. As a result, the D2 is the dominant 5′ss in the *STAT3* wild type. However, if either the non-canonical binding register or DSR is disrupted, D1 outcompetes the D2 completely. Recently, Artemyeva-Isman and Porter (59) provided compelling evidence that the U5 and U6 snRNA base pairing with human 5′ss cooperate to ensure proper splice site recognition. The D1 can potentially establish more consecutive base pairs with U6 snRNA than the D2 (5 versus 3) (Figure 7A). Additionally, the D1 possibly establishes more WC base pairs with U5 snRNA than the D2 (2 versus 0). Therefore, possibly better U5 and U6 snRNA interaction with the D1 might explain the competitive advantage of D1 over D2 in the absence of non-canonical U6 pairing and/or DSR.

Finally, our data suggest a mechanism of splice site choice in the *STAT3* as well. Exon inclusion in single nucleotide substitutions in the D1 dinucleotide could not be rescued by shifted U1 snRNAs as effectively as in substitutions in D2 dinucleotide or other positions (Supplementary Figure S5). As suppression with one particular shifted U1 snRNA should ideally result in similar rescue across various mutations, it follows that D1 dinucleotide disruption further prevents the use of both splice sites by inducing exon skipping. The importance of D1 for D2 usage was also supported by insGTA mutation (Figure 6). In the insGTA mutation, the D2 was not utilized even though the D2 motif was essentially unchanged (only moved by three nucleotides downstream). This again showed that some sort of interaction between D1 and D2 is required for proper D2 use. Based on this, we propose that, in the wild type *STAT3*, the tri-snRNP interacts initially with the D1 motif which then enables the use of the D2. Interestingly, one of the earliest steps where the initial selection of 5′ss in *STAT3* could happen is during the pre-B to B spliceosomal complex transition. During this phase, the 5′ss is transferred from U1 snRNA to the mobile ACAGAGA loop of U6 snRNA and other nearby spliceosomal components such as U5 snRNA loop 1 and specific residues in hPrp8 (39,42). The 5′ss that is closest to these recognition elements, in this case the D1, might initially be preferentially selected. However, our experiments with shifted U1 snRNAs argue against simple proximity-based selection of 5′ss (Figure 4B, D). For example, shift U1 snRNAs (+3, +6, +9 and +12) that changed the proximity of tandem 5′ss to the U6 snRNA did not affect the 5′ss selection or exon inclusion. We suggest that the initial and ultimate 5′ss selection is based on the affinity of either 5′ss to the tri-snRNP and is modulated by distance between these two 5′ss. We suggest that tri-snRNP is initially bound to the D1 due to the stronger D1:U5 snRNA/snRNP interaction and the D1 to D2 shift happens due to the combination of novel non-canonical U6 snRNA interaction and the DSR. Additionally, other proteins, such as hDim1, FBP21, Prp38 and Snu23 that contact the newly formed U6:5′ss duplex in the B spliceosomal complex might also contribute to the splice site choice, possibly by regulation of other spliceosomal proteins such as hBrr2 helicase (42,65).

Our work provides evidence that the 5′ss:U6 interaction is much more flexible than previously thought and can potentially accommodate a non-canonical binding register if specific conditions are met. We hypothesize that U5 and U6 snRNA explore multiple alignments and binding registers with pre-mRNA, pick the 5′ss that is most suitable for catalysis and neglect the other.

Data availability

Raw RNAseq data are available at the National Center for Biotechnology (NCBI) and accessible by BioProject ID PRJNA942059. Further data are available at: https://zenodo.org/doi/10.5281/zenodo.10654359.

Supplementary Data

Supplementary Data are available at NAR Online.

Acknowledgements

We acknowledge the CF Genomics supported by the NCLG research infrastructure (LM2018132 funded by MEYS CR) for their support with obtaining scientific data presented in this paper. Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic

Funding

Centre for Cardiovascular Surgery and Transplantation [201903]; Ministry of Education, Youth and Sports [MUNI/A/1244/2021, LM2018132, 90254]; Ministry of Health Czech Republic [FNBr65269705]. Funding for open access charge: Faculty of Medicine, Masaryk University, Brno, Czech Republic.

Conflict of interest statement

None declared.

References

1. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
2. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
3. Sheth,N., Roca,X., Hastings,M.L., Roeder,T., Krainer,A.R. and Sachidanandam,R. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.*, **34**, 3955–3967.
4. Ule,J. and Blencowe,B.J. (2019) Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell*, **76**, 329–345.
5. Lee,Y. and Rio,D.C. (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.*, **84**, 291–323.
6. Zhou,Z. and Fu,X.D. (2013) Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, **122**, 191–207.
7. Geuens,T., Bouhy,D. and Timmerman,V. (2016) The hnRNP family: insights into their role in health and disease. *Hum. Genet.*, **135**, 851–867.
8. Sammeth,M., Foissac,S. and Guigó,R. (2008) A general definition and nomenclature for alternative splicing events. *PLoS Comput. Biol.*, **4**, e1000147.
9. Chern,T.M., Van Nimwegen,E., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y. and Zavolan,M. (2006) A simple physical model predicts small exon length variations. *PLoS Genet.*, **2**, e45.
10. Séraphin,B., Kretzner,L. and Rosbash,M. (1988) A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5′ cleavage site. *EMBO J.*, **7**, 2533–2538.
11. Siliciano,P.G. and Guthrie,C. (1988) 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev.*, **2**, 1258–1267.
12. 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.
13. Roca,X. and Krainer,A.R. (2009) Recognition of atypical 5′ splice sites by shifted base-pairing to U1 snRNA. *Nat. Struct. Mol. Biol.*, **16**, 176–182.
14. Roca,X., Akerman,M., Gaus,H., Berdeja,A., Bennett,C.F. and Krainer,A.R. (2012) Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.*, **26**, 1098–1109.
15. Tan,J., Ho,J.X.J., Zhong,Z., Luo,S., Chen,G. and Roca,X. (2016) Noncanonical registers and base pairs in human 5′ splice-site selection. *Nucleic Acids Res.*, **44**, 3908–3921.
16. Dewilde,S., Vercelli,A., Chiarle,R. and Poli,V. (2008) Of alphas and betas: distinct and overlapping functions of STAT3 isoforms. *Front. Biosci.*, **13**, 6501–6514.
17. Hu,X., li,J., Fu,M., Zhao,X. and Wang,W. (2021) The JAK/STAT signaling pathway: from bench to clinic. *Signal Transduct. Target Ther.*, **6**, 402.
18. Aigner,P., Just,V. and Stoiber,D. (2019) STAT3 isoforms: alternative fates in cancer? *Cytokine*, **118**, 27–34.
19. Zheng,M., Turton,K.B., Zhu,F., Li,Y., Grindle,K.M., Annis,D.S., Lu,L., Drennan,A.C., Tweardy,D.J., Bharadwaj,U., *et al.* (2016) A mix of s and δs variants of stat3 enable survival of activated b-cell-like diffuse large b-cell lymphoma cells in culture. *Oncogenesis*, **5**, e184.
20. Dawes,R., Bournazos,A.M., Bryen,S.J., Bommireddipalli,S., Marchant,R.G., Joshi,H. and Cooper,S.T. (2023) SpliceVault predicts the precise nature of variant-associated mis-splicing. *Nat. Genet.*, **55**, 324–332.
21. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
22. 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.
23. 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**, 13060–13074.
24. Souček,P., Réblová,K., Kramárek,M., Radová,L., Grymová,T., Hujová,P., Kováčová,T., Lexa,M., Grodecká,L. and Freiberger,T. (2019) High-throughput analysis revealed mutations' diverging effects on SMN1 exon 7 splicing. *RNA Biol*, **16**, 13064–13076.
25. Case,D.A.D.A., Betz,R.M.R., Cerutti,D.D.S., T.E.,C.I., Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W., Homeyer,N., *et al.* (2016) In: *Amber 2016*. University of California, San Francisco.
26. Zgarbová,M., Otyepka,M., Šponer,J., Mládek,A., Banáš,P., Cheatham,T.E. and Jurečka,P. (2011) Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
27. Aduri,R., Psciuk,B.T., Saro,P., Taniga,H., Schlegel,H.B. and SantaLucia,J. (2007) AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theory Comput.*, **3**, 1464–1475.
28. Lavery,R., Zakrzewska,K., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T., Dixit,S., Jayaram,B., Lankas,F., Laughton,C., *et al.* (2009) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.
29. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
30. 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.
31. Kondo,Y., Oubridge,C., van Roon,A.M.M. and Nagai,K. (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife*, **4**, e04986.
32. Hansen,S.R., White,D.S., Scalf,M., Corrêa,I.R., Smith,L.M. and Hoskins,A.A. (2022) Multi-step recognition of potential 5′ splice sites by the *Saccharomyces cerevisiae* U1 snRNP. *eLife*, **11**, e70534.
33. Crispino,J.D., Blencowe,B.J. and Sharp,P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**, 1866–1869.
34. Tarn,W.Y. and Steitz,J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev.*, **8**, 2704–2717.
35. Crispino,J.D. and Sharp,P.A. (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev.*, **9**, 2314–2323.
36. Crispino,J.D., Mermoud,J.E., Lamond,A.I. and Sharp,P.A. (1996) Cis-acting elements distinct from the 5′ splice site promote U1-independent pre-mRNA splicing. *RNA*, **2**, 664–673.
37. Fukumura,K., Taniguchi,I., Sakamoto,H., Ohno,M. and Inoue,K. (2009) U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res.*, **37**, 1907–1914.
38. Black,C.S., Whelan,T.A., Garside,E.L., Macmillan,A.M., Fast,N.M. and Rader,S.D. (2023) Spliceosome assembly and regulation: insights from analysis of highly reduced spliceosomes. *RNA*, **29**, 531–550.

39. Charenton,C., Wilkinson,M.E. and Nagai,K. (2019) Mechanism of 5′ splice site transfer for human spliceosome activation. *Science*, **364**, 362–367.
40. Lesser,C.F. and Guthrie,C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*, **262**, 1982–1988.
41. Kandels-Lewis,S. and Séraphin,B. (1993) Role of U6 snRNA in 5′ splice site selection. *Science*, **262**, 2035–2039.
42. Zhan,X., Yan,C., Zhang,X., Lei,J. and Shi,Y. (2018) Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res.*, **28**, 1129–1140.
43. Bertram,K., Agafonov,D.E., Dybkov,O., Haselbach,D., Leelaram,M.N., Will,C.L., Urlaub,H., Kastner,B., Lührmann,R. and Stark,H. (2017) Cryo-EM structure of a pre-catalytic Human spliceosome primed for activation. *Cell*, **170**, 701–713.
44. Bertram,K., Agafonov,D.E., Liu,W.T., Dybkov,O., Will,C.L., Hartmuth,K., Urlaub,H., Kastner,B., Stark,H. and Lührmann,R. (2017) Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, **542**, 318–323.
45. Zhang,X., Yan,C., Hang,J., Finci,L.I., Lei,J. and Shi,Y. (2017) An atomic structure of the Human spliceosome. *Cell*, **169**, 918–929.
46. Hiller,M., Huse,K., Szafranski,K., Jahn,N., Hampe,J., Schreiber,S., Backofen,R. and Platzer,M. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat. Genet.*, **36**, 1255–1257.
47. Hiller,M., Szafranski,K., Backofen,R. and Platzer,M. (2006) Alternative splicing at NAGNAG acceptors: simply noise or noise and more? [1]. *PLoS Genet.*, **2**, e207.
48. Zavolan,M., Kondo,S., Schönbach,C., Adachi,J., Hume,D.A., Arakawa,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Gaasterland,T. (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res.*, **13**, 1290–1300.
49. Hiller,M. and Platzer,M. (2008) Widespread and subtle: alternative splicing at short-distance tandem sites. *Trends Genet.*, **24**, 246–255.
50. Hujová,P., Souček,P., Radová,L., Kramárek,M., Kováčová,T. and Freiberger,T. (2021) Nucleotides in both donor and acceptor splice sites are responsible for choice in NAGNAG tandem splice sites. *Cell. Mol. Life Sci.*, **78**, 6979–6993.
51. Hiller,M., Huse,K., Szafranski,K., Rosenstiel,P., Schreiber,S., Backofen,R. and Platzer,M. (2006) Phylogenetically widespread alternative splicing at unusual GYNGYN donors. *Genome Biol.*, **7**, R65.
52. Turton,K.B., Annis,D.S., Rui,L., Esnault,S. and Mosher,D.F. (2015) Ratios of four STAT3 splice variants in human eosinophils and diffuse large B cell lymphoma cells. *PLoS One*, **10**, e0127243.
53. Sinha,R., Lenser,T., Jahn,N., Gausmann,U., Friedel,S., Szafranski,K., Huse,K., Rosenstiel,P., Hampe,J., Schuster,S., *et al.* (2010) TassDB2 - A comprehensive database of subtle alternative splicing events. *BMC Bioinf.*, **11**, 216.
54. O'Reilly,D., Dienstbier,M., Cowley,S.A., Vazquez,P., Drozdz,M., Taylor,S., James,W.S. and Murphy,S. (2013) Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.*, **23**, 281–291.
55. Mabin,J.W., Lewis,P.W., Brow,D.A. and Dvinge,H. (2021) Human spliceosomal snRNA sequence variants generate variant spliceosomes. *RNA*, **27**, 1186–1203.
56. Vazquez-Arango,P., Vowles,J., Browne,C., Hartfield,E., Fernandes,H.J.R., Mandefro,B., Sareen,D., James,W., Wade-Martins,R., Cowley,S.A., *et al.* (2016) Variant U1 snRNAs are implicated in human pluripotent stem cell maintenance and neuromuscular disease. *Nucleic Acids Res.*, **44**, 10960–10973.
57. Hwang,D.Y. and Cohen,J.B. (1996) U1 snRNA promotes the selection of nearby 5′ splice sites by U6 snRNA in mammalian cells. *Genes Dev.*, **10**, 338–350.
58. Brackenridge,S., Wilkie,A.O.M. and Screaton,G.R. (2003) Efficient use of a 'dead-end' GA 5′ splice site in the human fibroblast growth factor receptor genes. *EMBO J.*, **22**, 1620–1631.
59. Artemyeva-Isman,O.V. and Porter,A.C.G. (2021) U5 snRNA interactions with exons ensure splicing precision. *Front. Genet.*, **12**, 676971
60. Alanis,E.F., Pinotti,M., Mas,A.D., Balestra,D., Cavallari,N., Rogalska,M.E., Bernardi,F. and Pagani,F. (2012) An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.*, **21**, 2389–2398.
61. Sawa,H. and Abelson,J. (1992) Evidence for a base-pairing interaction between U6 small nuclear RNA and the 5′ splice site during the splicing reaction in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 11269–11273.
62. Zhang,X., Yan,C., Zhan,X., Li,L., Lei,J. and Shi,Y. (2018) Structure of the human activated spliceosome in three conformational states. *Cell Res.*, **28**, 307–322.
63. Wilkinson,M.E., Charenton,C. and Nagai,K. (2020) RNA splicing by the spliceosome. *Annu. Rev. Biochem.*, **89**, 359–388.
64. Wan,R., Bai,R., Zhan,X. and Shi,Y. (2020) How is precursor messenger RNA spliced by the spliceosome? *Annu. Rev. Biochem.*, **89**, 333–358.
65. Henning,L.M., Santos,K.F., Sticht,J., Jehle,S., Lee,C.T., Wittwer,M., Urlaub,H., Stelzl,U., Wahl,M.C. and Freund,C. (2017) A new role for FBP21 as regulator of Brr2 helicase activity. *Nucleic Acids Res.*, **45**, 7922–7937.

Received: October 10, 2022. **Revised:** February 2, 2024. **Editorial Decision:** February 5, 2024. **Accepted:** February 16, 2024