Cryptic U2-dependent Pre-mRNASplice Site Usage Induced by Splice Switching Antisense Oligonucleotides

Kristin Ham
Murdoch University
Niall Keegan
Murdoch University
Craig McIntosh
Murdoch University
May Aung-Htut
Murdoch University
Khine Zaw
Murdoch University
Kane Greer
Murdoch University
Sue Fletcher
Murdoch University
Stephen Wilton (s.wilton@murdoch.edu.au)
Centre for Molecular Medicine and Innovative Therapeutics, Health Futures Institute, Murdoch University, 8 Perth, Western Australia, 6150, Australia

Research Article

Keywords: Antisense oligomers (AOs), RNA splicing, cryptic splice site activation

Posted Date: January 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-144809/v1

Version of Record: A version of this preprint was published at Scientific Reports on July 23rd, 2021. See the published version at https://doi.org/10.1038/s41598-021-94639-x.

Cryptic U2-dependent pre-mRNA splice site usage induced by splice switching antisense oligonucleotides

Kristin A. Ham[1,2¶], Niall P. Keegan[1,2¶], Craig S. McIntosh[1,2], May T. Aung-Htut[1,2], Khine Zaw[1-3], Kane Greer[1,2], Sue Fletcher[1,2] and Steve D. Wilton[1,2*]

[1] Centre for Molecular Medicine and Innovative Therapeutics, Health Futures Institute, Murdoch University, Perth, Western Australia, 6150, Australia
[2] Perron Institute for Neurological and Translational Science, Centre for Neuromuscular and Neurological Disorders, The University of Western Australia, Perth, Western Australia, 6009, Australia
[3] Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand
* corresponding author s.wilton@murdoch.edu.au
¶ these authors contributed equally to this work

ABSTRACT

Antisense oligomers (AOs) are increasingly being used for modulating RNA splicing in live cells, both for research and for therapeutic purposes. While the most common intended effect of these AOs is to induce skipping of whole exons, rare examples are emerging of AOs that induce skipping of only part of an exon, through activation of an internal cryptic splice site. In this report, we examined seven such examples of AO-induced cryptic splice site activation – five new examples from our own experiments and three from reports published by others. We modelled the predicted effects that AO binding would have on the secondary structure of each of the RNA targets, and how these alterations would in turn affect the accessibility of the RNA to splice factors. We observed that a common predicted effect of AO binding was a disruption to the exon definition signal within the exon's excluded segment.

Introduction

The process of pre-mRNA splicing is a fundamental aspect of gene regulation and function in higher eukaryotes. Pre-mRNA consists of coding regions termed exons that are intersected by non-coding regions termed introns [1]. During maturation into mRNA, these non-coding regions are removed, and the exons ligated together to form a continuous message ready to be translated into a protein. Pre-mRNA splicing involves a multitude of splicing factors that interact with numerous splicing motifs on the transcript [2]. A large multi-protein complex called the spliceosome is responsible for the coordination of this complex set of transesterification reactions [3].

The major form of the spliceosome is composed of five small nuclear ribonucleoproteins (snRNPs; U1, U2, U5 and U4/U6), as well as numerous non-snRNP proteins [4,5]. The canonical 5′ splice site (5′ss) is defined by an AG|GURAGU sequence, while the 3′ splice site (3′ss) is denoted by a (Yn)-YAG| sequence (where; | = exon boundary; underlined sequence identifies invariant nucleotides; R = purine; Y = pyrimidine) [6]. The branchpoint sequence, typically located approximately 15 to 50 nucleotides (nt) upstream from the 3′ss, is required for U2 snRNA binding during spliceosome formation. This sequence is defined as YNC**U**R**A**Y (underlined sequence denotes branch formation region; bold nucleotides are highly conserved; N = any nucleotide) [6]. The major spliceosome (called spliceosome hereon), along with hundreds of associated splicing factors are responsible for over 95% of all splicing reactions including the phenomenon known as alternative splicing [7-10].

Alternative splicing is a process whereby multiple different transcript and protein isoforms can arise from a single protein-coding gene and is an essential element in spatial and temporal regulation of gene expression in higher eukaryotes [7]. In order to achieve alternative splicing, the spliceosome must recognize and select a splice site from a variety of alternative splice sites and branchpoints within the transcript. Typically, these splice sites are well defined and have evolutionarily conserved functions. However, sometimes sequences usually ignored by the spliceosome can become

activated as splice junctions. These are known as cryptic splice sites [11] and are most often activated by mutations or errors during transcription [12]. The most common causative mutations are those that abolish canonical splice sites, thus redirecting the spliceosome to either utilize a viable cryptic site nearby or exclude the exon completely from the mature mRNA [13]. Cryptic splice sites may be found within both exonic and intronic regions and typically include or exclude a proportion of the intron or exon [12]. Interestingly, recent data has shown that cryptic splice sites can also be activated by synthetic molecules such as antisense oligonucleotides.

Antisense oligonucleotides (AOs) are small, single-stranded RNA or DNA-like synthetic molecules used to modify gene expression. These AOs can be used to downregulate gene expression through RNA silencing, redirection of pre-mRNA splicing patterns, intron retention, inhibiting translation, or RNase H-induced degradation of the target gene transcript [14]. The sequence of maturing gene transcripts can also be altered by using AOs to induce removal or inclusion of an exon, as seen with current therapeutic strategies approved for the treatment of Duchenne muscular dystrophy and spinal muscular atrophy, respectively.

While most splice modulating AOs are designed with the intention to enhance exon selection or induce skipping of whole exons, the occasional activation of cryptic splice sites after *in vitro* AO treatment has also been observed. We have reported the activation of a cryptic donor splice site after treatment with an AO targeting *LMNA* pre-mRNA, promoting removal of 150 nt from the end of exon 11 [15]. Evers et al. [16] observed that an AO targeting exon 9 in *ATXN3* promoted a partial exon 9 skip, activating an alternative 5'ss. A partial exon 12 skip in the *HTT* transcript was also detected after treatment with an AO (World Patent WO2015053624A2); once again activating a cryptic donor splice site [17]. Lastly, we recently reported activation of two cryptic donor splice sites by AOs containing several locked nucleic acid residues, designed to enhance efficiency of exon skipping from the dystrophin transcript [18].

In addition to the established roles that splice site motifs and exon enhancer and silencer motifs play in directing RNA splicing, there is increasing evidence of a similar role for RNA secondary structure [19-22] and of its effect on splice factor binding [23,24]. While modelling the interactions of these phenomena presents a highly complex challenge, a reasonable starting point may be to assume that RNA secondary structure is generally antagonistic to splice factor binding within closed regions.

In our laboratory's quest to develop new therapeutics for debilitating genetic diseases, we have tested thousands of AOs targeted to numerous genes in a variety of cell types, but we have observed only a handful of AO-induced cryptic splicing events in the target transcripts in human cells, and a single example in mouse cells [25]. In this study, we investigated the possible mechanisms by which AOs may induce cryptic splicing. We analyzed 13 AOs targeting six different human gene transcripts and found that changes to the accessibility of enhancer and silencer motifs within the transcript secondary structure appeared to play a role in many cases. The diverse nature of these changes indicates that there may be multiple pathways to inducing cryptic splicing, sometimes within a single exon.

Results and Discussion

To explore the possible mechanisms behind cryptic splice site activation, we analyzed AO-induced cryptic splicing events in six different human transcripts: *COL7A1, SRSF2, ATXN3, USH2A, HTT,* and *LMNA*. Data for *HTT* and *LMNA* were obtained from the literature and analyzed together with those from the remaining transcripts.

Analysis of antisense oligonucleotide treatment

COL7A1 exon 15

Antisense oligonucleotides were transfected into healthy human fibroblasts as cationic lipoplexes at concentrations of 100 and 50 nM to induce skipping of exon 15 from the *COL7A1* pre-mRNA transcript, removing 144 nt from the full-length transcript (Fig. 1a). Subsequent RT-PCR analysis revealed both the full-length transcript and an unanticipated amplicon, smaller than full-length but larger than would be expected as a result of complete exon 15 removal. The unexplained amplicon was isolated and identified by Sanger sequencing to be missing the last 64 nucleotides from the 3' end of exon 15. Removing 64 nt from the *COL7A1* transcript would render the cryptically spliced product out-of-frame, and therefore produce a premature termination codon in exon 16. This discovery highlights the importance of investigating unexpecting splicing products after AO treatment. A new donor splice site was activated by treatment with an AO targeting *COL7A1* exon 15, H15A(+91+115), that resulted in cryptic splice site activation in 30% of the transcripts at both 100 nM and 50 nM. Treatment with this AO did not induce other aberrant splicing products. Transfection of cells with an AO covering the authentic donor site, H15D(+14-11), did not lead to cryptic donor site activation.

SRSF2 exon 2

Antisense oligonucleotides were transfected into healthy human fibroblasts as cationic lipoplexes at concentrations of 100, 50 and 25 nM to induce skipping of exon 2 from the *SRSF2* pre-mRNA transcript, removing 311 nt from the full-length transcript (Fig. 1b). Gel fractionation of the RT-PCR amplicons revealed several products confirmed by Sanger sequencing: full-length SRSF2-T204 (ENST00000452355.7); full-length SRSF2-T208 (ENST00000585202.5); and T208 missing 65 nt from the 3' end of exon 2. Multiple amplicons larger than 1000 nt were present, which correspond to the amplicon sizes of the transcripts SRSF2-T203 (ENST00000392485.2) and SRSF2-T202 (ENST00000359995.10) (Supplementary Fig. 1a). The splicing of T202 appears to be influenced by the AOs in the same manner (Supplementary Fig. 1a). However, we were unable to isolate and identify various amplicons to confirm this. The AOs did not appear to cause exon skipping or cryptic donor site activation within the T203 transcript, most likely due to the T203 isoform containing only two exons, making both "unskippable" [26].

Under normal conditions, *SRSF2* transcript isoforms T202 and T203 code for proteins while T208 and T204 undergo nonsense mediated decay (NMD). After AO treatment, the expression of the cryptically spliced T208 increased with a concomitant decrease in the full-length T202. The cryptic splicing of exon 2 removes the natural termination codon from T202, T204, and T208 and exposes a new in-frame termination codon in the following exon of each transcript (Supplementary Fig. 1b).

Mammalian NMD generally follows the '50 nucleotide rule', whereby termination codons more than 50 nt upstream of the final exon are determined premature and result in a reduction in mRNA abundance [27]. Cryptic splice site activation appears to stabilize T208 as a new termination codon is created within 50 nt of the penultimate 3' exon junction. Isoform T204 still appears to undergo NMD, as the new termination codon is exposed within the third exon of the five-exon isoform.

ATXN3 exon 9

Antisense oligonucleotides were transfected into healthy control human fibroblasts as cationic lipoplexes at concentrations of 400, 200, 100 and 50 nM to induce skipping of exon 9 from the *ATXN3* pre-mRNA, thereby removing 97 nt from the full-length transcript (Fig. 1c). Gel fractionation

of the RT-PCR amplicons revealed two full-length product bands representing the two transcripts in the untreated sample: a larger product (533 nt) containing 21 CAG (21Q) repeats, and a slightly smaller product containing eight CAG (8Q) repeats. Complete exon 9 skipping from the 8Q transcript was observed in healthy human fibroblasts treated with H9A(+38+62) at all concentrations tested. The same AO treatment also activated a cryptic donor site, resulting in removal of 55 nt from the 8Q transcript. Treatment with H9D(+20-05) resulted solely in partial exon 9 skipping from the 8Q transcript. All amplicons were isolated and identified by Sanger sequencing.

Complete and partial exon 9 skipping was observed only from the 8Q and not the 21Q transcript. Cryptic donor activation in the transcript with fewer CAG repeats dominates in some AO treatments but not others [28,29]. The CAG expansion occurs in the following exon 10 that is separated by a 10 kb intron from the AO target. Numerous studies assessing AO-mediated removal of exon 9 and/or exon 10 from the *ATXN3* transcript reported reduced exon skipping efficiencies the larger the expansion size. Although this phenomenon is directed more towards exon 10 removal, we speculate that the CAG repeat length may influence frequency of the cryptic splice site usage. The nature of the CAG repeat allows for numerous consecutive potential serine/arginine-rich splicing factor (SRSF) 2 (AGCAG) and SRSF5 (ACAGC) splice motifs. The fact that these positive exon selection sites are heavily repeated may influence exon 10, and potentially exon 9, selection and therefore susceptibility to AO-mediated exon skipping.

USH2A exon 13

Antisense oligonucleotides were transfected into a Huh7 cell line as cationic lipoplexes at concentrations of 200 and 50 nM to induce skipping of exon 13 from the *USH2A* pre-mRNA transcript (Fig. 1d). Subsequent RT-PCR analysis revealed multiple unanticipated amplicons larger than expected from the removal of exon 13 in its entirety. It was evident that multiple splicing events occurred: removal of the complete exon 13; activation of a cryptic donor; activation of a cryptic acceptor; or activation of both cryptic donor and acceptor sites within exon 13, after treatment with different AOs. Treatment with H13A(-05+20) and H13A(+70+94) resulted mainly in complete exon 13 exclusion, removing 642 nt from the full-length transcript, and the activation of a cryptic acceptor site, removing 527 nt from the full-length transcript. Treatment with H13A(+136+160) and H13A(+161+183) resulted in the activation of a cryptic donor site, both on its own (missing 513 nt from the 3' end of exon 13) and in conjunction with the cryptic acceptor site (missing 398 nt from the middle of exon 13), but did not remove the entire exon 13. We were unable to isolate and identify one of the amplicons by Sanger sequencing (labelled with an arrow in Fig. 1d). We speculate that this amplicon is a heteroduplex, which would explain why it could not be isolated.

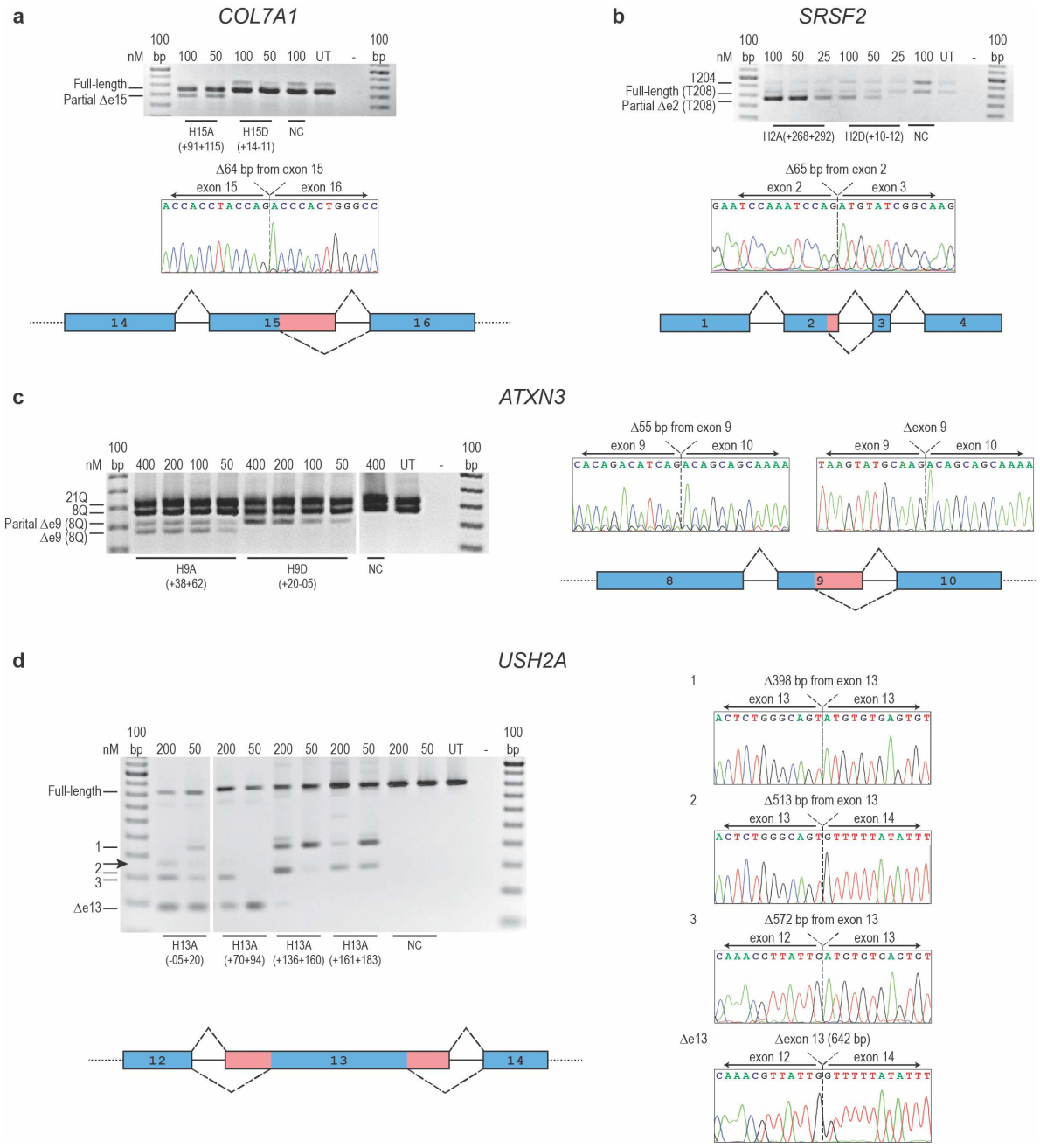


Figure 1. Activation of cryptic splice sites by AO-mediated splice switching in four different gene transcript targets. (**a**) *COL7A1* exon 15. (**b**) *SRSF2* exon 2. (**c**) *ATXN3* exon 9. (**d**) *USH2A* exon 13**.** Reverse transcription-PCR analysis after transfection with antisense oligonucleotides (AOs), at various nM concentrations indicated above the gel image. Sanger sequencing data identifies the smaller amplicon(s) resulting from AO treatment. Blue boxes represent exons, lines between the boxes represent introns, dashed lines above and below represent various splicing events, pink boxes represent the portion of exon removed after the activation of a new cryptic splice site. Arrow indicates an amplicon that could not be succesfully isolated and sequenced. NC, negative control sequence synthesized as 2'-OMe PS; UT, untreated; 100 bp, 100 base pair DNA ladder; nM, nanomolar.

Analysis of splice site scores and exonic splicing enhancer motifs masked by the examined antisense oligonucleotides

Two models were employed to calculate the scores of both the canonical and cryptic splice sites activated after AO treatment: a weight matrix model, Human Splice Finder 3.1 [30], and a maximum entropy model, MaxEntScan [31]. No discernable pattern became evident using either model (Table 1), indicating splice site scores are not the only factor influencing splice site usage.

Table 1. Comparing canonical and cryptic splice site scores using two different modeling approaches.

Gene (exon)	Splice site	HSF canonical splice site score	HSF cryptic splice site score	MaxEnt canonical splice site score	MaxEnt cryptic splice site score	Position relative to beginning of exon
USH2A (13)	Acceptor	88.04	80.44	8.95	-1.01	+527
	Donor	97.66	82.16	10.77	4.88	+129
COL7A1 (15)	Donor	88.19	78.49	4.01	2.97	+80
ATXN3 (9)	Donor	74.37	76.7	1.6	7.09	+42
SRSF2 (2)	Donor	73.19	72.69	-0.64	5.46	+246
*HTT** (12)	Donor	83.43	92.8	7.16	8.54	+206
*LMNA** (11)	Donor	98.84	88.33	8.07	2.93	+120

Exonic splicing enhancer (ESE) motifs masked by AO binding sites were counted using ESEFinder3.0 [32]; (Table 2). Motifs were included when one or more motif nucleotides were masked by the targeting AO. The examined AOs were found to consistently mask SRSF1 motifs, with exception of the AO H2D(+10-12) targeting the *SRSF2* donor site.

Table 2. Exonic splicing enhancer motifs masked by the antisense oligonucleotides examined in this study.

Gene	AO nomenclature	**SRSF1** (SF2)	**SRSF2** (SC35)	**SRSF5** (SRp40)	**SRSF6** (SRp55)
USH2A	H13A(-05+20)	1	1	1	0
	H13A(+70+94)	1	2	0	0
	H13A(+136+160)	2	0	1	0
	H13A(+161+183)	4	1	1	1
COL7A1	H15A(+91+115)	5	1	1	0
	H15D(+14-11)	2	2	0	1
ATXN3	H9A(+38+62)	1	0	1	1
	H9A(+65+85)*	3	0	1	0
	H9D(+20-05)	1	0	2	0
SRSF2	H2A(+268+292)	4	1	0	0
	H2D(+10-12)	0	1	1	0
HTT	H12A(+269+297)*	3	4	3	0
LMNA	H11A(+221+245)*	3	3	1	1

*Not tested in this study; published results

The splicing factor SRSF1 is necessary for several splicing processes, including lariat formation and 5'ss cleavage [33]. In addition, SRSF1 assists in modulating 5'ss selection [33]. The addition of purified SRSF1 to cultured cells favored 5'ss located more proximally to the 3'ss while lower levels of SRSF1 favored 5'ss located distal to the 3'ss [34]. In our study, AOs can mask the availability of ESE motif binding sites, therefore reducing the amount of SRSF1 that can bind to the pre-mRNA. Fewer SRSF1 binding sites may drive the 5'ss preference away from the canonical splice site towards a more distal cryptic splice site.

Analysis of AO-induced changes to exonic splicing enhancer/silencer access within cryptically spliced exons

It is notable that in all seven of the above examples, the cryptic splice sites observed fell within the exon, between the canonical splice sites, rather than downstream or upstream. We suggest that this is a logical consequence of the 'exon definition' paradigm under which the human spliceosome is thought to operate, whereby transcript sequence between the first and last exons is processed as intron unless specifically defined as being part of an internal exon [35]. Because 'intron' is the default sequence identity under this paradigm, AO binding is therefore much more likely to diminish an existing exon signal than it is to spontaneously extend it.

Because four of the seven cryptic splice sites had MaxEnt scores lower than their canonical counterparts, it was clear that our analysis would need to encompass other variables in order to explain the activation of these sites – specifically, those variables that could plausibly be altered by AO binding. We therefore attempted to model the effect that AO binding would have on both the local secondary structure of the transcript and the subsequent change in accessibility to ESE and exon splicing silencer (ESS) motifs.

The ESE and ESS motifs for each cryptically spliced exon were overlaid to generate enhancer and silencer scores at each nucleotide position. These values were then "masked" by the predicted secondary structure for the exons, effectively resetting the ESE and ESS scores to zero for all nucleotides predicted to bind other nucleotides. This masking was repeated with the altered structures predicted for on-target AO binding, and the two plots were vertically aligned to allow comparison between them (Fig. 2a-e). Because the size of *USH2A* exon 13 (642 nt) made it impractical to visually compare changes in its ESE and ESS access in the same manner as for the other exons, we elected to present only the net changes in ESE and ESS access as a result of AO binding (Fig. 2f-g).

In *COL7A1* exon 15 (Fig. 2a), AO binding was predicted to increase ESE access in the retained 5′ segment, as well as directly competing with ESEs in the excised 3′ segment. The net effect was a much stronger exon signal from the 5′ segment that improved the profile of the cryptic donor site. This example demonstrates that blocking an authentic donor site does not automatically activate a cryptic donor site; additional elements, including secondary structure and exon and intron definition motifs, are necessary to define the exon boundary.

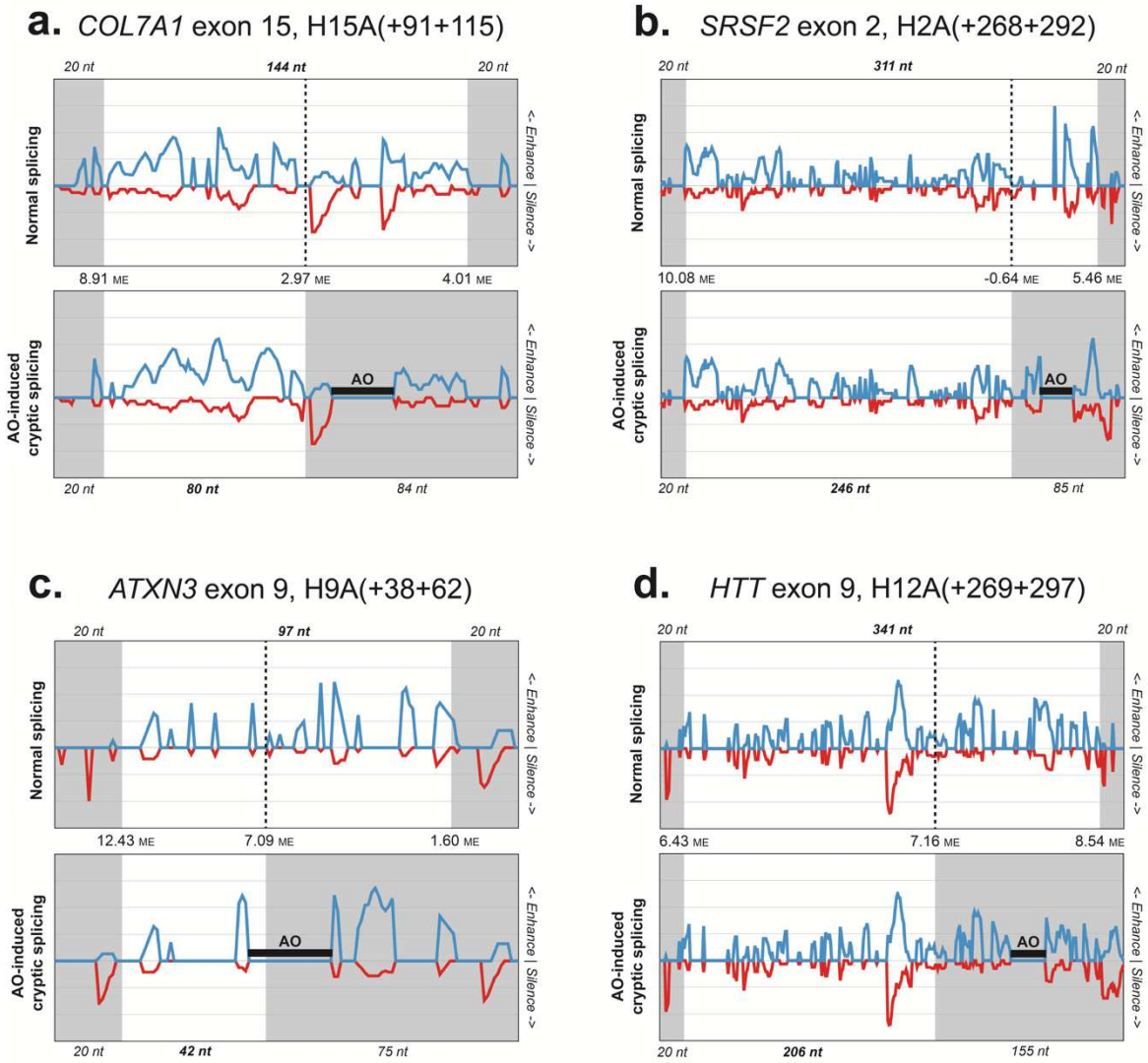
For *SRSF2* exon 2 (Fig. 2b), the AO directly obscured the strongest enhancer peak in the excised 3' segment and induced a moderate increase in ESE access within the retained 5' segment. We also observed that, in the absence of AO binding, the enhancer signal in the excised 3' segment of the exon was substantially stronger than in the rest of the exon. This may be a positively selected feature to ensure inclusion of this segment and avoidance of the cryptic splice site, though it is not clear why the very poor MaxEnt score of the cryptic donor is not a sufficient deterrent alone.

In *ATXN3* exon 9 (Fig. 2c), the AO binding site overlapped the cryptic donor site and caused loss of ESE access 3' of cryptic donor and a slight increase of ESE access immediately 5′ of the cryptic donor. This, combined with the much stronger MaxEnt score of the cryptic site, may have been enough to shift exon definition to the 5' region of the exon.

In *HTT* exon 12 (Fig. 2d), the changes in secondary structure did not clearly favor either enhancement or silencing of the excised segment. However, ESS access was increased both 5' and 3' of the canonical donor site, and this appears to have been sufficient to tip the balance towards the comparably strong cryptic donor splice site. A similar change appears to have occurred in *LMNA* exon 11 (Fig. 2e), with the exception that the cryptic donor site in this exon was much stronger than its canonical neighbor.

For *USH2A* exon 13, there was almost no change to predicted secondary structure induced by H13A(+70+94), apart from that at the AO binding site (Fig. 2f). It therefore appears that steric blocking alone is the reason for cryptic splice site inducement in this case, though it is notable that this relatively minor change to the exon's secondary and tertiary structure is sufficient to redirect the spliceosome to an alternative 3'ss, 490 nt downstream. This suggests that coordination between the canonical acceptor and donor sites is essential for sustaining the exon definition signal across this exon's entire 642 nt span. However, we cannot explain why H13A(+70+94) drove use of the cryptic splice site removing 527 nt from exon 13 specifically, especially when compared to the case of the

H13A(+136+160) AO that induced use of a much stronger cryptic acceptor site only 37 nt downstream (Fig. 2g). We suggest that there may be some aspects of exon definition that are unique to internal exons as large as *USH2A* exon 13 and that these can only be properly understood by studying splicing in similarly sized exons from other genes. Exons longer than 500 nt, such as *USH2A* exon 13, typically rely on intron definition rather than exon definition in order to achieve correct splicing, but this intron-defined splicing can become inefficient when the intron size exceeds 500 nt [36,37]. It is possible that sporadic splice site activation in this larger exon is partly due to the inability of the spliceosome to utilize intron definition, and thus inefficiently creates exon isoforms of less than 500 nt by activating various internal splice sites regardless of their strength.



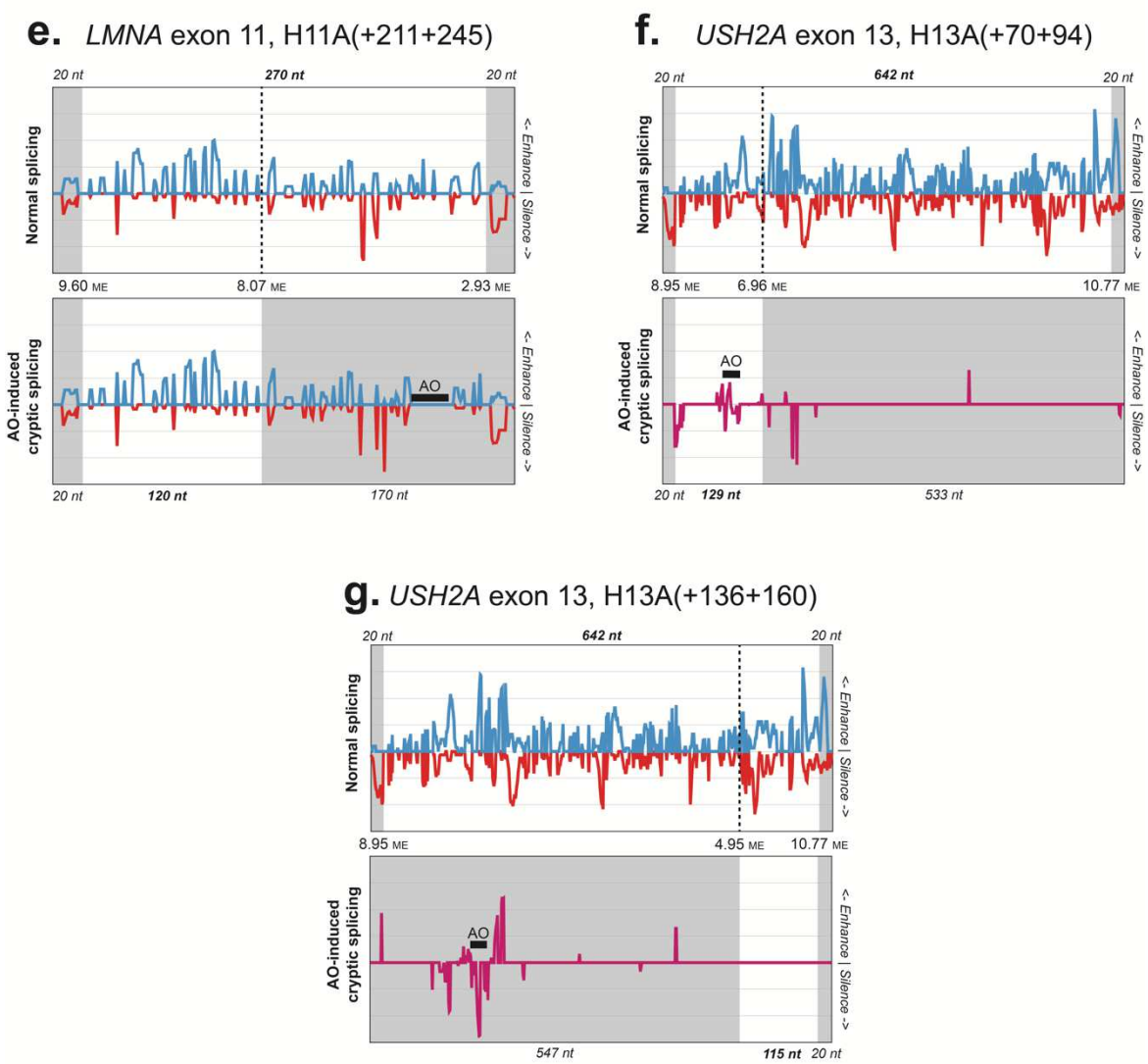


Figure 2. Changes to predicted exon splicing enhancer/silencer (ESE/ESS) access in seven examples of antisense oligonucleotide (AO)-induced cryptic splicing of canonical exons. Blue lines indicate ESE access and red lines indicate ESS access (**a**-**e**), while for the 642 nt *USH2A* exon 13, purple indicates the net change in ESE and ESS access as a result of AO binding (**f**-**g**). Grey shading indicates pre-mRNA sequence excluded from the mature transcript. Region sizes and Maximum Entropy scores for cryptic and canonical splice sites are also shown.

Conclusions

Despite the small number of examples of AO-induced cryptic splicing, we observed considerable diversity in the etiology of this phenomenon. However, a common feature appears to be disruption of the exon definition signal.

It is clear that canonical exon definition is achieved not by any single motif, but by the cumulative signal of multiple enhancers binding with regularity and consistency along the entire exon span. Furthermore, continuity of this enhancing signal appears to be just as important, if not more important, than its overall strength. This continuity is especially crucial when the exon contains a cryptic splice site, as this is often the only metric by which the spliceosome can distinguish the cryptic site from its canonical neighbor.

Methods

Antisense oligonucleotides (AOs)

Antisense oligonucleotides (AOs) comprising of 2′-*O*-methyl modified bases on a phosphorothioate backbone (2'-OMe PS) were synthesized by TriLink BioTechnologies (San Diego, CA) or synthesized in-house on an Expedite 8909 Nucleic Acid synthesizer (Applied Biosystems, Melbourne, Australia) using the 1 µmol thioate synthesis protocol, as described previously [38]. After synthesis, the oligonucleotides were cleaved from the support following incubation in ammonium hydroxide for a minimum of 24 h at room temperature. The 2′-OMe PS AOs were subsequently desalted under sterile conditions on NAP-10 columns (GE Healthcare, Sydney, Australia) according to manufacturer's instructions. The 2'-OMe PS AOs used in this study are listed in Table 3. Oligonucleotide nomenclature is based on that described by Aung-Htut et al. (2019) and Mann et al. (2002), indicating the intron:exon, exon or exon:intron annealing coordinates in the target gene pre-mRNA [39,40].

Table 3. Information for AOs

Gene	AO nomenclature	Sequence (5' to 3')
USH2A	H13A(-05+20)	GCAAUGAUCACACCUAAGCCCUAAA
	H13A(+70+94)	GAGCCAUGGAGGUUACACUGGCAGG
	H13A(+136+160)	UGAAGUCCUUUGGCUUCUUUUUUGC
	H13A(+161+183)	AGUUUUCUCUGCAGGUGUCACAC
COL7A1	H15A(+91+115)	CCCUCCUCUCUGCCUCGCAGUACCG
	H15D(+14-11)	CAGGGCCUGACCCGUUCGAGCCACG
ATXN3	H9A(+38+62)	UUCUGAAGUAAGAUUUGUACCUGAU
	H9A(+65+85)*[16]	GCUUCUCGUCUCUUCCGAAGC
	H9D(+20-05)	UUUACUUUUCAAAGUAGGCUUCUCG
SRSF2	H2A(+268+292)	UCCUUCCUCUUCAGGAGACUUG
	H2D(+10-12)	CCCAGACAUUACUUAAGAGGAC
HTT	H12A(+269+297)*[17]	CGGUGGUGGUCUGGGAGCUGUCGCUGAUG
LMNA	H11A(+221+245)*[15]	AGGAGGUAGGAGCGGGUGACCAGAU

* Not tested in this study; published results

Cell culture and transfection

All cell culture reagents were purchased from Gibco, (ThermoFisher Scientific, Scoresby, Australia), unless otherwise stated. Primary dermal fibroblasts derived from a healthy volunteer after informed consent (The University of Western Australia Human Research Ethics Committee approval RA/4/1/2295; Murdoch University Human Research Ethics Committee approval 2013/156) were propagated in Dulbecco's modified Eagle's medium (DMEM) supplemented with 1% GlutaMax™-I and 10% foetal bovine serum (FBS) (Scientifix, Cheltenham, Australia) at 37°C in a 5% CO_2 atmosphere. Cells were seeded in 24-well plates (1.8 x 10^4 cells/well) in DMEM supplemented with 10% FBS for 24 hours before transfection. The human hepatocarcinoma cell line, Huh7, was supplied by the JCRB Cell Bank (Osaka, Japan) and purchased from CellBank Australia (Westmead, NSW, Australia). These cells were maintained in DMEM supplemented with 10% FBS at 37°C in a 5% CO_2 atmosphere. Huh7 cells were seeded in 24-well plates (5 x 10^4 cells/well) in propagation media for 24 hours before transfection.

Fibroblast and Huh7 cells were transfected with 2′-OMe PS AO-Lipofectamine 3000 (Thermo Fisher Scientific) lipoplexes in Opti-MEM (Gibco) according to the manufacturer's instructions, at various concentrations in duplicate wells, and the cells were then incubated at 37°C for 24 hours before RNA extraction. The negative control oligo (sequence from Gene Tools, LLC synthesized as

a 2´-OMe PS AO), which targets a human beta-globin intron mutation, was used as a negative control.

Molecular analysis

After harvesting the cells, total RNA was extracted using MagMax™ nucleic acid isolation kit (AM1830; Thermo Fisher Scientific) according to manufacturer's instructions and included the DNase treatment step. Molecular analyses were accomplished using three different systems optimized for different gene targets. SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* DNA Polymerase (Thermo Fisher Scientific) was used to synthesize and amplify cDNA from 50 ng of total RNA in a single step. Nested PCR was necessary to amplify the *USH2A* transcripts. Briefly, after 20 cycles of amplification, one µl aliquot was removed and subjected to nested PCR for 25 cycles using AmpliTaq Gold (Thermo Fisher Scientific) and an inner primer set. For regions with a high GC-content that are more difficult to amplify, SuperScript™ IV First-Strand Synthesis System and random hexamers (Thermo Fisher Scientific) were used to synthesize cDNA from harvested total RNA, and approximately 50 ng of cDNA was used as a template for PCR amplification using the TaKaRa LA Taq® DNA Polymerase with GC Buffer II system (Takara Bio USA, Inc., Clayton, Australia). PCR systems, conditions and primers used to assess splice modulation across the different gene transcripts are summarized in Table 4.

Amplified RT-PCR products were resolved on 2% agarose gels by electrophoresis in Tris-acetate ethylenediaminetetraacetic acid buffer, compared to a 100 bp DNA size standard (Geneworks, Adelaide, Australia). Relative transcript abundance was estimated by densitometry on images captured by the Fusion FX system (Vilber Lourmat, Marne-la-Vallée, France) using Fusion-Capt software and ImageJ (version 1.8.0_112) software for densitometry analysis. To identify RT-PCR products, the amplicons were first isolated by bandstab [41], followed by template preparation using Diffinity RapidTip for PCR Purification (Diffinity Genomics, Inc., West Henrietta, NY) and DNA sequencing, performed by the Australian Genome Research Facility Ltd. (Nedlands, Australia).

Table 4. List of primers, PCR system and conditions used in this study.

Gene target (accession numbers)	Primer orientation	Sequence (5' – 3')	Length (nt)	PCR system	Cycling conditions
ATXN3 (NM_004993.6)	Exon 7F	GTCCAACAGATGCATCGACCAA	522 (21Q) 483 (8Q)	SSIII One-Step	55˚C (30 min) and 94˚C (2 min); 28 cycles of 94˚C (30s), 55˚C (30s) and 68˚C (1.5 min)
	Exon 11R	AGCTGCCTGAAGCATGTCTTCTT			
COL7A1 (NM_000094.4)	Exon 13F	CTTAGCTACACTGTGCGGGT	765	SSIII One-Step	55˚C (30 min) and 94˚C (2 min); 30 cycles of 94˚C (30s), 60˚C (30s) and 68˚C (1.5 min)
	Exon 19R	TGGGAGTATCTGGTGCCTCA			
SRSF2 (XR_429913.4)	Exon 1F	CCCAGAGCTGAGGAAGCC	850	SSIV TaKaRa GC I	94˚C (1 min); 32 cycles of 94˚C (30s), 62˚C (30s) and 72˚C (4 min)
	Exon 4R	CTCAACTGCTACACAACTGC			
USH2A (NM_206933.4)	Exon 12F	AAGAGTTGGATCCTGATGGCTGC	993	SSIII One-Step	55˚C (30 min) and 94˚C (2 min); 20 cycles of 94˚C (15s), 60˚C (30s) and 68˚C (1 min)
	Exon 15R	GACAGGTTTCATTCAAGGCTCC			
	Exon 12F	CTGTAACTGCAATACCTCTGG	837	AmpliTaq Gold	94˚C (5 min); 25 cycles of 94˚C (30s),
	Exon 14R	CAAACACACTGACCAGTCAGG			

					60˚C (30s) and 72˚C (1 min); 72˚C (5 min)

In silico analysis

Basic Local Alignment Search Tool (BLAST) [42] was used to compare amplicon sequences to the reference mRNA sequences (accession numbers: Table 4). Sequences for each cryptically spliced exon and +/-20 nt of flanking intron were input to Human Splice Finder [30] which generated a JSON file with the locations of every detected ESE and ESS motif, as well as predicted acceptor and donor splice sites. Raw text from this JSON file was then imported into a custom-made spreadsheet (see Supplementary Material) that used this data to assign an ESE and an ESS score to each nucleotide of the sequence, under the following rationale:

ESE score: +1/n for each overlapping ESE motif, where n = ESE motif length;

ESS score: -1/n for each overlapping ESS motif, where n = ESS motif length.

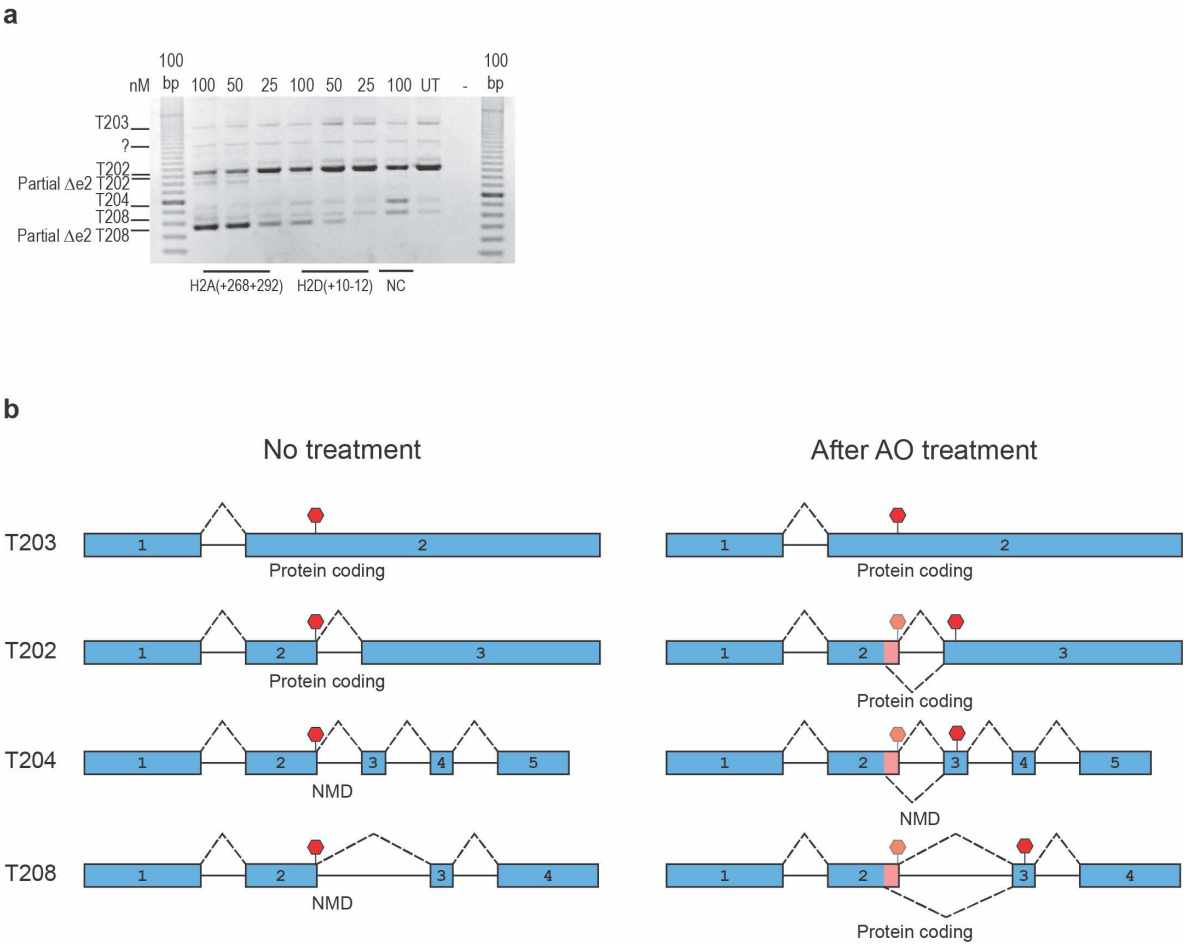
For example, a nucleotide that fell within two six nt ESE motifs and one eight nt ESS motif would be assigned an ESE score of 0.333 (2 x 1/6) and an ESS score of -0.125 (1 x -1/8).

Predicted centroid normal RNA folding was calculated for the sequence of each cryptically spliced exon with +/-70 nt flanking intron, using RNAfold [43] with the “avoid isolated base pairs” option. Predicted centroid AO-induced folding was calculated for each exon using the same sequence and settings as for normal folding, but with an additional constraint mask that prohibited binding within the AO target sites.

Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information file).

Supplementary materials



Supplementary Figure 1. **Analysis of antisense oligonucleotide-mediated splice switching in *SRSF2* gene transcripts.** (**a**) Full gel image of reverse transcription-PCR analysis after transfection with antisense oligonucleotides (AOs), at various nM concentrations indicated above the gel image. (**b**) Alternative transcript exon composition before and after AO treatment. Blue boxes represent exons, lines between the boxes represent introns, dashed lines above and below represent various splicing events, red polygons represent termination codons, pink boxes represent the exon portion removed after the activation of a cryptic splice site, pink polygons represent termination codons removed after cryptic splice site activation. Multiple transcript isoforms noted as T### according to Ensembl. Question mark (?) indicates an amplicon that could not be succesfully isolated and sequenced. NC, negative control sequence synthesized as 2'-OMe PS; UT, untreated; 100 bp, 100 base pair DNA ladder; nM, nanomolar; NMD, nonsense mediated decay.

References

1 Ward, A. J. & Cooper, T. A. The pathobiology of splicing. *J. Pathol.* **220**, 152-163, doi:10.1002/path.2649 (2010).

2 Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349**, 1191-1198, doi:10.1126/science.aac8159 (2015).

3 Sperling, R. The nuts and bolts of the endogenous spliceosome. *WIRES RNA* **8**, e1377, doi:10.1002/wrna.1377 (2017).

4 Papasaikas, P. & Valcarcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem. Sci.* **41**, 33-45, doi:10.1016/j.tibs.2015.11.003 (2016).

5 Turunen, J. J., Niemela, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. *WIRES RNA* **4**, 61-76, doi:10.1002/wrna.1141 (2013).

6 Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108-121, doi:10.1038/nrm3742 (2014).

7 Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437-451, doi:10.1038/nrm.2017.27 (2017).

8 Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514**, 1-30, doi:10.1016/j.gene.2012.07.083 (2013).

9 Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **84**, 291-323, doi:10.1146/annurev-biochem-060614-034316 (2015).

10 Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am. J. Hum. Genet.* **102**, 11-26, doi:10.1016/j.ajhg.2017.11.002 (2018).

11 Nelson, K. K. & Green, M. R. Mechanism for cryptic splice site activation during pre-mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6253-6257, doi:10.1073/pnas.87.16.6253 (1990).

12 Haj Khelil, A., Deguillien, M., Moriniere, M., Ben Chibani, J. & Baklouti, F. Cryptic splicing sites are differentially utilized in vivo. *FEBS J.* **275**, 1150-1162, doi:10.1111/j.1742-4658.2008.06276.x (2008).

13 Krawczak, M. *et al.* Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* **28**, 150-158, doi:10.1002/humu.20400 (2007).

14 Aartsma-Rus, A. *et al.* Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms. *Mol. Ther.* **17**, 548-553, doi:10.1038/mt.2008.205 (2009).

15 Luo, Y.-B. *et al.* Antisense Oligonucleotide Induction of Progerin in Human Myogenic Cells. *PLoS ONE* **9**, e98306, doi:10.1371/journal.pone.0098306 (2014).

16 Evers, M. M. *et al.* Ataxin-3 protein modification as a treatment strategy for spinocerebellar ataxia type 3: removal of the CAG containing exon. *Neurobiol. Dis.* **58**, 49-56, doi:10.1016/j.nbd.2013.04.019 (2013).

17 van Roon-Mom, W. M., Evers, M. M., Pepers, B. A., Aartsma-Rus, A. & Van Ommen, G. J. Antisense oligonucleotide directed removal of proteolytic cleavage sites, the hchwa-d mutation, and trinucleotide repeat expansions. *WIPO* (2014).

18 Zaw, K. *et al.* Consequences of Making the Inactive Active Through Changes in Antisense Oligonucleotide Chemistries. *Front. Genet.* **10**, 1249, doi:10.3389/fgene.2019.01249 (2019).
19 Jin, Y., Yang, Y. & Zhang, P. New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *RNA Biol.* **8**, 450-457, doi:10.4161/rna.8.3.15388 (2011).
20 Shilo, A., Tosto, F. A., Rausch, J. W., Le Grice, S. F. J. & Misteli, T. Interplay of primary sequence, position and secondary RNA structure determines alternative splicing of LMNA in a pre-mature aging syndrome. *Nucleic Acids Res.* **47**, 5922-5935, doi:10.1093/nar/gkz259 (2019).
21 Soemedi, R. *et al.* The effects of structure on pre-mRNA processing and stability. *Methods* **125**, 36-44, doi:10.1016/j.ymeth.2017.06.001 (2017).
22 Zhang, J., Kuo, C. C. & Chen, L. GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC Genomics* **12**, 90, doi:10.1186/1471-2164-12-90 (2011).
23 Hiller, M., Zhang, Z., Backofen, R. & Stamm, S. Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* **3**, e204, doi:10.1371/journal.pgen.0030204 (2007).
24 Saha, K. *et al.* Structural disruption of exonic stem-loops immediately upstream of the intron regulates mammalian splicing. *Nucleic Acids Res.* **48**, 6294-6309, doi:10.1093/nar/gkaa358 (2020).
25 Mitrpant, C. *et al.* Rational Design of Antisense Oligomers to Induce Dystrophin Exon Skipping. *Mol. Ther.* **17**, 1418-1426, doi:10.1038/mt.2009.49 (2009).
26 Lee, Y. *et al.* Variants Affecting Exon Skipping Contribute to Complex Traits. *PLoS Genet.* **8**, e1002998, doi:10.1371/journal.pgen.1002998 (2012).
27 Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5**, R8, doi:10.1186/gb-2004-5-2-r8 (2004).
28 McIntosh, C. S., Aung-Htut, M. T., Fletcher, S. & Wilton, S. D. Removal of the Polyglutamine Repeat of Ataxin-3 by Redirecting pre-mRNA Processing. *Int. J. Mol. Sci.* **20**, doi:10.3390/ijms20215434 (2019).
29 Toonen, L. J. A., Schmidt, I., Luijsterburg, M. S., Van Attikum, H. & Van Roon-Mom, W. M. C. Antisense oligonucleotide-mediated exon skipping as a strategy to reduce proteolytic cleavage of ataxin-3. *Sci. Rep.* **6**, 35200, doi:10.1038/srep35200 (2016).
30 Desmet, F. O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67, doi:10.1093/nar/gkp215 (2009).
31 Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377-394, doi:10.1089/1066527041410418 (2004).
32 Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31**, 3568-3571 (2003).
33 Zuo, P. & Manley, J. L. Functional domains of the human splicing factor ASF/SF2. *EMBO J.* **12**, 4727-4737 (1993).
34 Krainer, A. R., Conway, G. C. & Kozak, D. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62**, 35-42, doi:10.1016/0092-8674(90)90237-9 (1990).

35 De Conti, L., Baralle, M. & Buratti, E. Exon and intron definition in pre-mRNA splicing. *WIRES RNA* **4**, 49-60, doi:10.1002/wrna.1140 (2013).
36 Fox-Walsh, K. L. *et al.* The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *PNAS* **102**, 16176-16181, doi:10.1073/pnas.0508489102 (2005).
37 Sterner, D. A., Carlo, T. & Berget, S. M. Architectural limits on split genes. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15081-15085, doi:10.1073/pnas.93.26.15081 (1996).
38 Adams, A. M. *et al.* Antisense oligonucleotide induced exon skipping and the dystrophin gene transcript: cocktails and chemistries. *BMC Mol. Biol.* **8**, 57, doi:10.1186/1471-2199-8-57 (2007).
39 Aung-Htut, M. *et al.* Systematic Approach to Developing Splice Modulating Antisense Oligonucleotides. *Int. J. of Mol. Sci.* **20**, 5030, doi:10.3390/ijms20205030 (2019).
40 Mann, C. J., Honeyman, K., McClorey, G., Fletcher, S. & Wilton, S. D. Improved antisense oligonucleotide induced exon skipping in the mdx mouse model of muscular dystrophy. *J. Gene Med.* **4**, 644-654, doi:10.1002/jgm.295 [doi] (2002).
41 Wilton, S. D., Lim, L., Dye, D. & Laing, N. Bandstab: A PCR-Based Alternative to Cloning PCR Products. *BioTechniques* **22**, 642-645, doi:10.2144/97224bm14 (1997).
42 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Med. Biochem.* **215**, 403-410, doi:10.1016/s0022-2836(05)80360-2 (1990).
43 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26, doi:10.1186/1748-7188-6-26 (2011).

Author contributions

Conceptualization, K.A.H., N.P.K., S.D.W.; methodology, K.A.H., N.P.K., C.S.M., K.Z., K.G., S.D.W.; formal analysis, K.A.H., N.P.K., S.D.W.; investigation, K.A.H., N.P.K., C.S.M., K.Z., K.G.; writing—original draft preparation, K.A.H., N.P.K., C.S.M.; writing—review and editing, K.A.H., N.P.K., C.S.M., M.T.A-H., K.Z., K.G., S.F., S.D.W.; supervision, S.D.W., S.F., M.T.A-H.; resources, S.D.W., S.F.; funding acquisition, S.D.W., S.F. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Health and Medical Research Council [grant number 1144791].

Conflicts of interest

S.D.W. is a consultant to Sarepta Therapeutics; S.D.W. and S.F. are named inventors on patents licensed through the University of Western Australia to Sarepta Therapeutics and as such are entitled to milestone and royalty payments; K.A.H., C.S.M., M.T.A-H., K.G. receive salary support from Sarepta Therapeutics. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. N.P.K and K.Z declare no competing interests.

Abbreviations

AO -	Antisense oligonucleotide
SnRNP -	Small nuclear ribonucleoproteins
5'ss -	5' splice site
3'ss -	3' splice site
Nt -	Nucleotide
ESE -	Exonic splicing enhancer
ESS -	Exonic splicing silencer
SRSF -	Serine/arginine-rich splicing factor
2'-OMe PS -	2´-*O*-methyl modified bases on a phosphorothioate backbone
DMEM -	Dulbecco's modified Eagle's medium
FBS -	Foetal bovine serum

520 NC - Negative control
521 NMD - Nonsense mediated decay

Figures

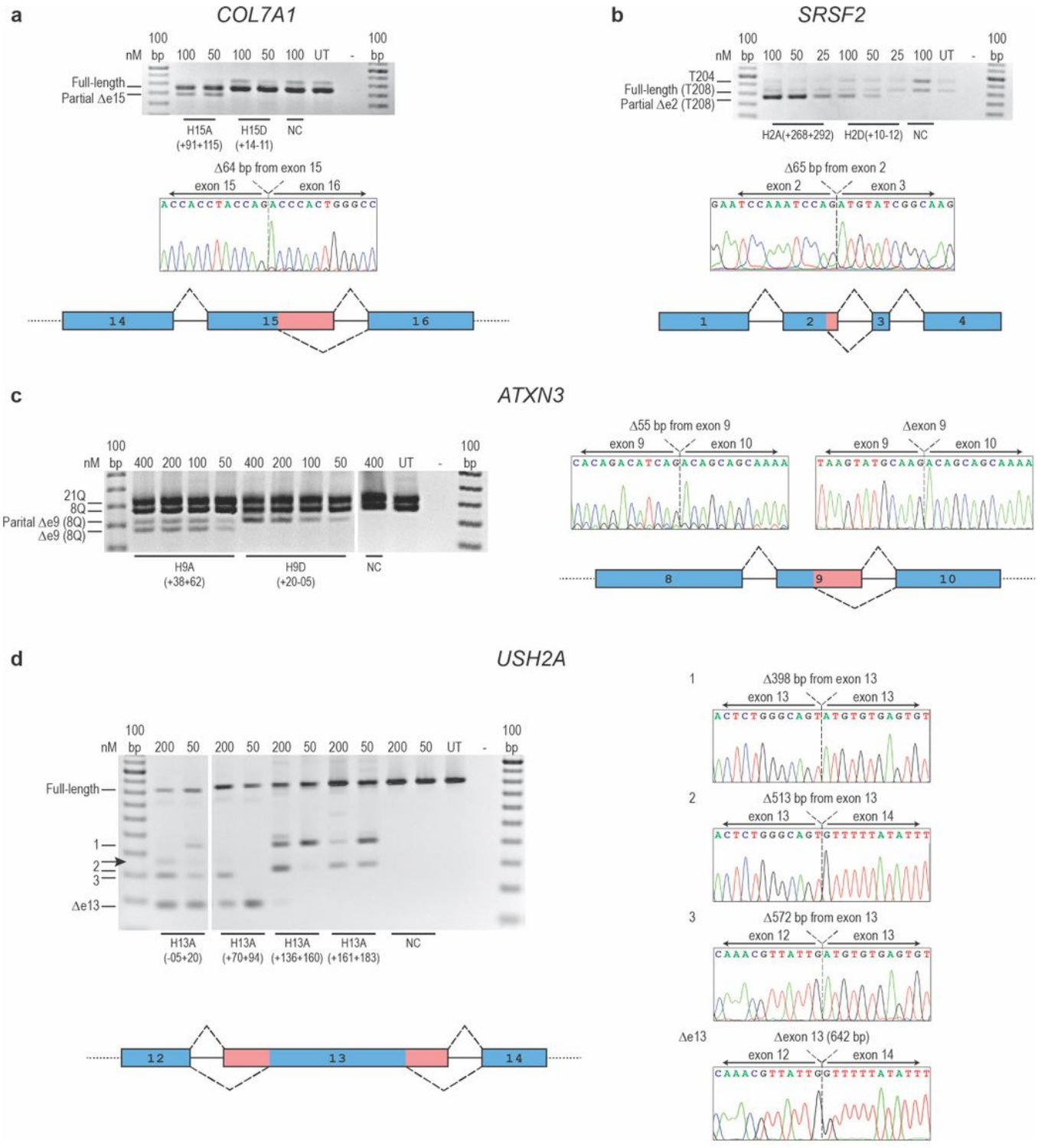


Figure 1

Activation of cryptic splice sites by AO-mediated splice switching in four different gene transcript targets. (a) COL7A1 exon 15. (b) SRSF2 exon 2. (c) ATXN3 exon 9. (d) USH2A exon 13. Reverse transcription-PCR analysis after transfection with antisense oligonucleotides (AOs), at various nM concentrations indicated

above the gel image. Sanger sequencing data identifies the smaller amplicon(s) resulting from AO treatment. Blue boxes represent exons, lines between the boxes represent introns, dashed lines above and below represent various splicing events, pink boxes represent the portion of exon removed after the activation of a new cryptic splice site. Arrow indicates an amplicon that could not be succesfully isolated and sequenced. NC, negative control sequence synthesized as 2'-OMe PS; UT, untreated; 100 bp, 100 base pair DNA ladder; nM, nanomolar.

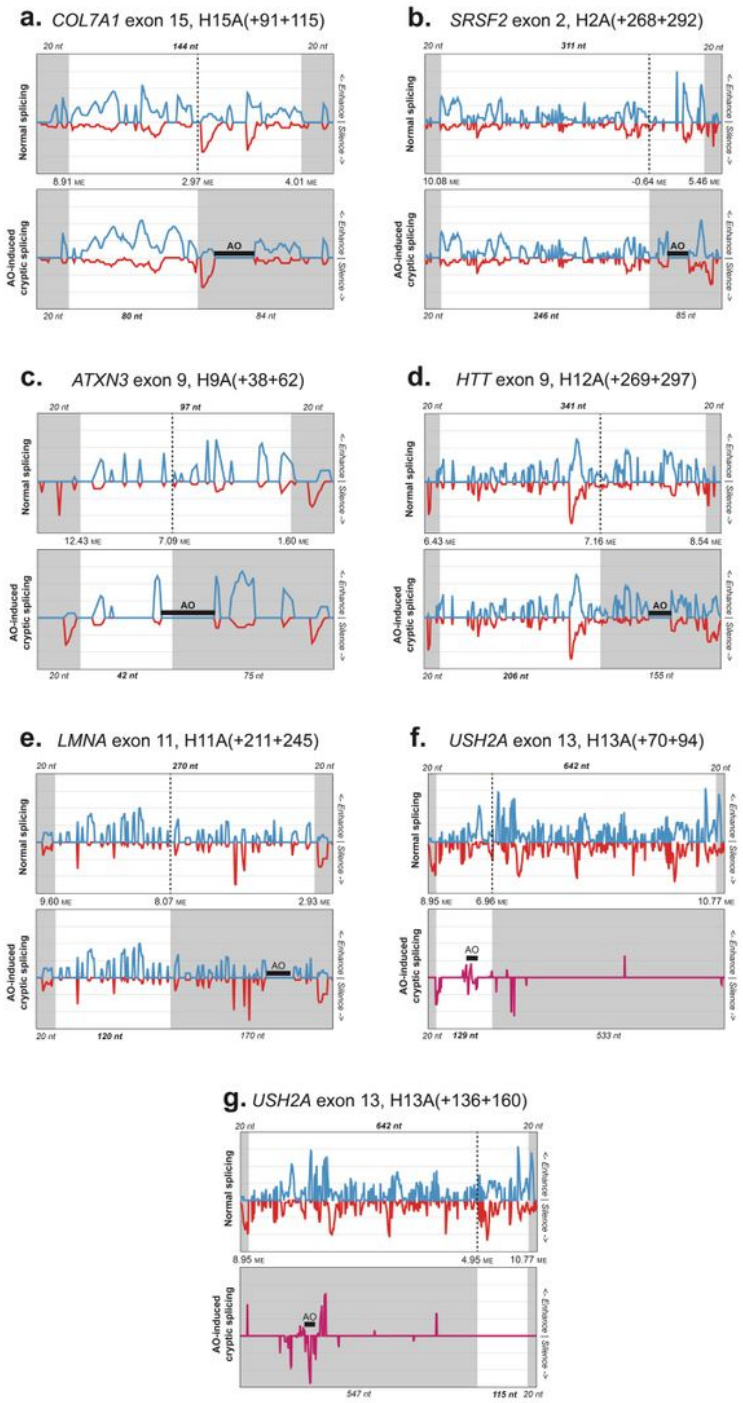


Figure 2

Changes to predicted exon splicing enhancer/silencer (ESE/ESS) access in seven examples of antisense oligonucleotide (AO)-induced cryptic splicing of canonical exons. Blue lines indicate ESE access and red lines indicate ESS access (a-e), while for the 642 nt USH2A exon 13, purple indicates the net change in ESE and ESS access as a result of AO binding (f-g). Grey shading indicates pre-mRNA sequence excluded from the mature transcript. Region sizes and Maximum Entropy scores for cryptic and canonical splice sites are also shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 1SupplementaryInformation.pdf
- 2SupplementaryDataAnalysis.xlsx
- 3SupplementaryRNASecondaryStructures.docx
- FigS1.jpg