



Universidad Autónoma  
de Madrid

**Biblos-e Archivo**  
Repositorio Institucional UAM

**Repositorio Institucional de la Universidad Autónoma de Madrid**

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:  
This is an **author produced version** of a paper published in:

Nucleic Acid Therapeutics (2022): 378-390

**DOI:** <http://doi.org/10.1089/nat.2021.0066>

**Copyright:** © 2022 Ainhoa Martínez-Pizarro et al. Published by Mary Ann Liebert, Inc.

El acceso a la versión del editor puede requerir la suscripción del recurso  
Access to the published version may require subscription

## ANTISENSE OLIGONUCLEOTIDE RESCUE OF DEEP-INTRONIC VARIANTS ACTIVATING PSEUDOEXONS IN THE *PTS* GENE

Ainhoa Martínez-Pizarro<sup>1</sup>, Fátima Leal<sup>1</sup>, Lise Lolle Holm<sup>2</sup>, Thomas K. Doktor<sup>2</sup>, Ulrika S. S. Petersen<sup>2</sup>, María Bueno<sup>3</sup>, Beat Thöny<sup>4</sup>, Belén Pérez<sup>1</sup>, Brage S Andresen<sup>2</sup>, Lourdes R Desviat<sup>1</sup>

<sup>1</sup>Centro de Biología Molecular Severo Ochoa UAM-CSIC, Centro de Diagnóstico de Enfermedades Moleculares (CEDEM), CIBERER, IdiPaz, Universidad Autónoma de Madrid, Madrid, Spain.

<sup>2</sup>Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

<sup>3</sup>Hospital Virgen del Rocio, Sevilla, Spain

<sup>4</sup>Division of Metabolism, University Children's Hospital Zürich, Zürich, Switzerland.

Correspondence should be addressed to:

Lourdes R Desviat

Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid

Nicolás Cabrera 1, 28049 Madrid, Spain

Phone: 34-911964566

Email: [lruiz@cbm.csic.es](mailto:lruiz@cbm.csic.es)

*Short title: Antisense therapy for PTS pseudoexons inclusions*

*Key words: pseudoexons, splicing, antisense oligonucleotides, tetrahydrobiopterin*

## Abstract

We report two new *PTS* splicing variants identified through genomic sequencing and transcript analysis in a patient with tetrahydrobiopterin deficiency, presenting with hyperphenylalaninemia and monoamine neurotransmitters deficiency. Variant c.243+3A>G causes exon 4 skipping. The deep-intronic c.164-672C>T variant creates a potential 5' splice site that leads to the inclusion of four overlapping pseudoexons, corresponding to exonizations of an antisense short interspersed nuclear element *AluSq* repeat sequence. Two of the identified pseudoexons have been reported previously, activated by different deep-intronic variants and were also detected at residual levels in control cells. Interestingly, the predominant pseudoexon is nearly identical to a disease causing activated pseudoexon in the *F8* gene, with the same 3' and 5' splice sites. Splice switching antisense oligonucleotides (SSOs) were designed to hybridize with splice sites and/or predicted binding sites for regulatory splice factors. Different SSOs corrected the aberrant pseudoexons inclusion, both in minigenes and in fibroblasts from patients carrying the new variant c.164-672C>T or the previously described c.164-716A>T. With SSO treatment PTPS protein was recovered, illustrating the therapeutic potential of the approach, for patients with different pseudoexon activating variants in the region. In addition, the natural presence of pseudoexons in the wild type context suggest the possibility of applying the antisense strategy in patients with hypomorphic *PTS* variants with the purpose of upregulating their expression to increase overall protein and activity.

## Introduction

The PTPS (6-pyruvoyltetrahydropterin synthase; EC 4.2.3.12) enzyme, encoded by the *PTS* gene (OMIM \*612719), catalyses the second step in tetrahydrobiopterin (BH<sub>4</sub>) synthesis from GTP in humans. BH<sub>4</sub> is the natural cofactor for phenylalanine, tyrosine and tryptophan hydroxylases, nitric oxide synthases, and glycerol ether monooxygenase [1]. PTPS deficient patients present with hyperphenylalaninemia and most of them present severe neurological manifestations including truncal hypotonia, bradykinesia, oculogyric crises, hyperthermia, seizures and impaired neurophysiological development. A subset of patients exhibits the less common atypical/peripheral form with minor or no changes in neurotransmitter levels. Treatment is based on oral BH<sub>4</sub> administration and supplementation with L-Dopa and 5-hydroxytryptophan, the precursors for catecholamines and serotonin, respectively, in order to normalize monoamine neurotransmitters in the central nervous system [2]. Despite treatment, many patients have complications with neurological and developmental problems [3].

In this work we describe a novel deep intronic variant affecting splicing in the *PTS* gene. It is well known that splice defects significantly contribute to human disease [4, 5]. Splice altering variants occur beyond disruption of the canonical splice sequences at exon-intron junctions (consensus 3' and 5' splice sites (ss), polypyrimidine tract and branch point), which are easily identified. Intronic and exonic splice silencers and enhancers (ISSs, ESSs, ISEs and ESEs, respectively), bound by regulatory splice factors, are less well conserved and sequence variants that disrupt them can be misclassified as silent or missense, unless transcript analysis is performed, which is not usual in a diagnostic setting [6, 7]. In addition, aberrant transcripts can result from deep intronic variants activating or creating potential splice sites, which are used in combination with other cryptic splice sites nearby resulting in the exonization of intronic regions known as pseudoexons [8, 9]. Pseudoexons share characteristics marking them as high-risk sites in the human genome, where SNVs may cause disease by splicing

activation [10, 11]. Again, these splice defects and the responsible variants are not detected in routine genetic analysis focusing in exome sequencing. Recent studies have revealed the potential of high-throughput transcriptome analysis for increasing the diagnostic rate in heterogeneous Mendelian disorders, allowing identification of splice-altering variants in both exonic and deep-intronic regions [12-16].

RNA targeted therapies such as splice switching antisense oligonucleotides (SSOs) provide an avenue towards therapy for specific splicing defects [17]. SSOs are designed to modulate splicing by sterically blocking splice sites or splicing regulatory sequences in the pre-mRNA. Blocking of consensus splice sites or splice enhancers will result in exon skipping, a strategy that has been used to bypass a mutated exon and/or to restore the open reading frame disrupted through genomic deletions, as is the case for Duchenne muscular dystrophy (DMD) [17]. On the other hand, blocking splice silencers may instead promote exon inclusion which is the therapeutic strategy used in spinal muscular atrophy (SMA) [18]. Therapeutic SSOs for DMD and SMA have been approved by the FDA/EMA, and several others are in clinical trials [19, 20].

SSOs have also been used to prevent aberrant pseudoexon inclusion in the mRNA. Many pseudoexons are transposable sequence elements derived from short or long interspersed nuclear elements (SINEs or LINES, respectively) [21] and their activation by point mutations has been described as pathogenic mechanism in many human disorders [8, 22-24]. In neuromuscular diseases, deep intronic variants have been identified in 77 different genes [14]. In inherited retinal diseases, therapeutic potential of SSOs have been shown for deep-intronic pseudoexon activating variants in the *CEP290*, *USH2A*, *OPA1* and *ABCA4* genes, with an ongoing phase 1/2 clinical trial using SSOs to treat *CEP290*-associated Leber congenital amaurosis (NCT03140969) [23]. Recently, an SSO (Milasen) targeting an activated pseudoexon identified in a single patient with Batten's disease was developed and approved by FDA for clinical testing

within one year, representing an extreme example of the use of SSOs as personalised medicines [25].

Previous studies by us and other groups reported the pathological activation by two different variants of overlapping pseudoexons in intron 2 of the *PTS* gene, which could be effectively reverted by use of a specific SSO [26]. We now report a novel deep intronic variant in the same region creating a 5'ss and activating the same pseudoexons plus two others that use the new splice site. The patient carries on the other allele an also novel variant affecting the 5'ss of exon 4. We have used transcript profiling in minigenes and in patients' fibroblasts to dissect the pathogenic mechanism for each variant and designed and tested a series of SSOs to block the insertion of the activated pseudoexons in the mature mRNA. Our results confirm SSO mediated pseudoexon skipping as a potential therapeutic approach for PTPS deficiency and point to the identification of high risk sites for the activation of *A/u*-derived pseudoexons.

## **Materials and methods**

**Patient' samples and genetic and biochemical analysis.** Patient 1 was referred to the laboratory for biochemical and genetic analysis after giving a positive result in the neonatal screening for hyperphenylalaninemia. Biochemical analysis (pterin levels in blood and urine, and monoamine neurotransmitter levels in CSF, measured by standard procedures [27]) indicated a clinical suspicion of PTPS deficiency. Genetic analysis was performed in DNA isolated from blood samples, by use of Massive Parallel Sequencing (MPS) using the Clinical-Exome Sequencing TruSight™ One Gene Panel (Illumina) and filtering for variants in all genes known to be responsible for hyperphenylalaninemia (*GCH1*, *PTS*, *PCBD1*, *QDPR*, *DNAJC12* and *PAH*). *PTS* transcript analysis was performed in RNA extracted from fibroblast samples. Potentially pathogenic variants were confirmed by Sanger sequencing. Patient 2 has been previously reported [26].

Ethical approval for the present study was granted by the Institutional Ethics Committee (Universidad Autónoma de Madrid).

**Fibroblasts and cell lines.** Patients' fibroblasts were referred to the lab for genetic analysis. Control fibroblasts CC2509 were obtained from Lonza. They were cultivated according to standard procedures and used with number of passes below 20. Briefly, cells were maintained in Minimum Essential Medium (MEM, Sigma Aldrich) supplemented with 1% (v/v) glutamine, 10% fetal bovine serum (FBS, Gibco) and 0.1% antibiotic mix (penicillin/streptomycin) under standard cell culture conditions (37°C, 95% relative humidity, 5% CO<sub>2</sub>). Cells were treated with puromycin (200 µg/ml) 5 hours before harvesting for transcript analysis, to inhibit nonsense mediated mRNA decay (NMD).

For minigene analysis, human hepatoma cells Hep3B and SH-SY5Y neuroblastoma cells were used. Cells were grown in MEM for Hep3B and Dulbecco's Modified Eagle Medium (DMEM, Gibco) for SH-SY5Y, both supplemented with 5% FBS, 1% glutamine and 0.1% antibiotic mix under standard cell culture conditions (37°C, 95% relative humidity, 5% CO<sub>2</sub>).

**Minigenes.** For *in vitro* evaluation of the splicing profile different minigene constructs were used. The previously described splicing reporter plasmid encompassing *PTS* exons 2 to 4 cloned in the pCR3.1 vector [26, 28] was used as template for deep intron 2 variants c.164-672C>T and c.164-716A>T. The pSPL3 vector (Life Technologies) was used in the analysis of the c.243+3A>G variant. Using this vector, we cloned *PTS* exon 4 and approximately 150 bp of the flanking intronic regions, amplified from genomic DNA using specific primers (Forward: 5'-GCTTCCATGCTGAGGTCAAT-3' and Reverse: 5'-ACTATTCCCCAACACCCACA-3'). Gene fragments were cloned into the pGEMT vector (Promega #A1360). The insert was excised with *EcoRI* and subsequently cloned into pSPL3. Variant minigenes containing the desired nucleotide changes were generated by site-directed mutagenesis with QuikChange Lightning Kit (Agilent Technologies, Santa

Clara, CA) using primers introducing the change and their reverse complement. Sanger sequencing confirmed the identity of the constructs.

**SSO design.** The sequence of the region encompassing the pseudoexons was analyzed and SSOs designed following published guidelines [29, 30]. Splice enhancer motifs were determined using ESE finder3.0. Pre-mRNA secondary structure was predicted using m-fold (<http://www.unafold.org/mfold/applications/rna-folding-form.php>) and RNAstructure software (<https://rna.urmc.rochester.edu/RNAstructureWeb/>) was used to predict secondary structure resulting from the interaction AON-target RNA. Specificity of the selected SSOs was verified using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). SSOs were 20- or 25-mers with a  $T_m$  above 60°C and a GC content between 50% and 60%. The exact sequences (Table S1) and their location are depicted in Figure 3A. A scrambled SSO was used as negative control. Given the abundance of *Alu* elements in the genome, we evaluated closely the predicted off-target binding sites, confirming they are far from adjacent exons. All oligonucleotides were synthesized as full 2'-OMe phosphorothioates by Sigma or Eurogentec.

**Transfections and splicing analysis in minigenes.** For minigene assays, Hep3B or SH-SY5Y cells were seeded in six-well plates at a density of  $4 \times 10^5$  or  $8 \times 10^5$ , respectively, in 2 ml MEM or DMEM 5% FBS and grown overnight. Cells were transfected using Lipofectamine 2000 (Invitrogen) with 2  $\mu$ g per well of the corresponding plasmid minigene. For SSO treatment, cells were co-transfected with 2  $\mu$ g of wild type (wt) or mutant minigenes and 50 nM of each SSO. Cells were harvested by trypsinization 48 h after transfection. Total RNA was isolated using Trizol Reagent (ThermoFisher) and phenol-chloroform extraction. cDNA synthesis was performed using NZY First-Strand cDNA Synthesis Kit (NZYtech). Splicing analysis was carried out by PCR amplification with FastStart Taq DNA Polymerase (Roche) using specific primers SD6 (5'-TCTGAGTCACCTGGACAACC-3') and SA2 (5'-ATCTCAGTGGTATTTGTGAGC-3') for pSPL3 minigenes, and PL3 (5'-GGGAGACCCAAGCTGGCTA-3') and PL4b (5'-



AGTCGAGGCTGATCAGCGG-3') for the pCR3.1 minigene. The amplification products were analysed by 2% agarose gel electrophoresis. Pseudoexon bands were excised from the gel and purified using QIAquick Gel Extraction (Qiagen). The identity of each pseudoexon was confirmed by subcloning in pGEMT (Promega) and subsequent Sanger sequencing.

**Transfections and splicing analysis in fibroblasts.** For SSO treatment, fibroblasts were seeded in six-well plates at a density of  $3 \times 10^5$  in 2 ml MEM 10% FBS and reverse transfected at the same time with 50 nM SSOs using lipofectamine 2000. Cells were harvested by trypsinization after 48 h. RNA was isolated using Trizol Reagent (ThermoFisher) and phenol-chloroform extraction. RT-PCR analysis was performed using primers *PTS* cDNA F (5'-ATGAGCACGGAAGGTGGTG-3') and Exon 3-R (5'-CTCTCCATGTACTGTCACCACAA-3') to amplify a fragment from exon 1 to exon 3. The amplification products were analyzed by 2.5% agarose gel electrophoresis and their identity and quantification confirmed by deep sequencing.

**Western blot analysis.** Approximately  $4 \times 10^5$  fibroblasts were seeded per well in a 6-well plate with 50 nM of each SSO, which were reverse transfected using Lipofectamine 2000. Ninety-six hours after transfection cells were harvested by trypsinization and lysed by freeze-thawing in a lysis buffer (10 mM Tris-HCl pH 7.5, 10% glycerol, 150 mM NaCl and 0.1% triton X-100) with protease inhibitor (Roche). Protein concentration was determined using Bradford reagent (Biorad). 15 µg of protein extract from control sample and 75 µg from patients' samples were electrophoresed on 12% SDS-polyacrylamide gels, transferred to a nitrocellulose membrane in an iBlot transfer device (Invitrogen) and blocked in TBS-0.1%Tween supplemented with 5% non-fat dry milk. Membranes were incubated overnight at 4°C with the primary antibody (PTPS 1:500, Santa Cruz Biotechnologies, sc-514628) followed by a secondary HRP-conjugated anti-mouse IgG (1:2000, Cell Signalling, #7096). GAPDH was used as a loading control (using primary antibody at 1:5000, Abcam, ab8245). Proteins were detected using the Enhanced

Chemiluminescence reagent (ECL, GE Healthcare). Images were scanned with a densitometer GS-900 (Bio-Rad) and quantified with Image Lab 5.2 (Bio-Rad).

**Deep sequencing.** Amplicons were generated in a first PCR with fusion primers containing a specific sequence plus a common tag (**CS1** and CS2):

- Forward primer: 5'-**ACACTGACGACATGGTTCTACACTGTTTGGGAAATGCAACAA**-3'

- Reverse primer: 5'-

TACGGTAGCAGAGACTTGGTCTTCTCCATGTACTGTCACCACAA-3'

Libraries preparation and sequencing were carried out at Fundación Parque Científico de Madrid under protocols developed and optimized for NGS amplicon sequencing. Briefly: Amplicons were diluted 1/50 and 1 µl was used as input for a second PCR of 10 cycles performed with a High-Fidelity DNA Polymerase in the presence of primers (5'-AATGATACGGCGACCAACGAGATCT**ACACTGACGACATGGTTCTACA**-3' and 5'-CAAGCAGAAGACGGCATACGAGAT-[10 nucleotides barcode]-TACGGTAGCAGAGACTTGGTCT-3') of the Access Array Barcode Library for Illumina Sequencers (Fluidigm). The finally obtained amplicons were validated and quantified by Bioanalyzer and equimolecular pool was purified by AMPure XP beads (Beckman Coulter). This pool was titrated by quantitative PCR using the "Kapa-SYBR FAST qPCR kit forLightCycler480" and a reference standard for quantification. Amplicon pool was denatured prior to be seeded on a flow cell where clusters were formed and sequenced using a "MiSeq Reagent Kit v3", in a 2x300 pair-end sequencing run on a MiSeq sequencer.

**RNA pulldown.** Affinity purification of RNA-binding proteins was performed as previously described [31], using 3'-end biotinylated RNA oligonucleotides (LGC Biosearch Technologies, Risskov, Denmark): PTS1 (5'-CCAGGUUCGAGCGAUUCUCCAGCCU-3'), PTS1MUT1 (5'-

CCAGGUUCGUGCGAUUCUCCAGCCU-3'),	PTS1MUT2	(5'-
CCAGGUUCUAGCUAUUCUCCAGCCU-3'),	PTS1Rev	(5'-
AGGCUGGAGAAUCGCUCGAACCUGG-3'),	PTS5	(5'-
UUCUCCAGCCUCAGCCUCCCGAGUA-3'),	PTS5MUT	(5'-
UUCUCCAGCCUCAGCCUCUUGAGUA-3'),	PTS5Rev	(5'-
UACUCGGGAGGCUGAGGCUGGAGAA-3'),	PTS6	(5'-
UCCCGAGUAGCUGAGAUUACAGGUG-3'),	PTS6MUT	(5'-
UCUUGUGUAGCUUAUAUUUCUGGUG-3'),	and PTS6Rev	(5'-
CACCUGUAAUCUCAGCUACUCGGGA-3').	500 pmol of each biotinylated RNA	

oligonucleotide was immobilized in 25 µl Dynabeads M-280 Streptavidin magnetic beads (Invitrogen) and incubated with 293 cell nuclear extract (Protein One). Blank beads were added as a control for background binding. Proteins were eluted in 25 µl protein sample buffer and analyzed by western blotting with antibodies against SRSF1 (32-4500 from Invitrogen), hnRNP A1 (R9778 from Sigma Aldrich), SRSF2 (04-1550 from Millipore) and SR proteins (33-9300 from Invitrogen).

**Surface plasmon resonance imaging.** SPRi was carried out as previously described [32]. Briefly, the biotinylated RNA-oligonucleotides described above were immobilized onto a G-strep sensorchip (SSENS) for 20 min. The following recombinant proteins were injected for 8 minutes, followed by dissociation for 4 minutes; 25 – 800 nM SRSF1 (Genscript) and 6.25 – 200 nM hnRNPA1 (Abcam, ab123212). A continuous flow of SPR buffer (10 mM Tris-HCl, pH 7.9, 150 mM KCl, 3.4 mM EDTA, 0.005% tween-20). Nuclear extract was used as a control of the oligos binding efficiency. Binding was fitted to a 1:1 kinetics model with Scrubber2 (v. 2.1.1.0; Biologics inc.). For hnRNPA1 a biphasic 1:2 model was used in ClampXP (version 3.50; Biosensor Data Analysis).

**In silico analysis.** The effect of the variants on splice site strengths and the presence of putative splicing regulatory elements were predicted using Human Splice Finder (HSF)

(<http://www.umd.be/HSF/>)

[33],

MaxEntScan

([http://genes.mit.edu/burgelab/masent/Xmaxentscan\\_scoreseq.html](http://genes.mit.edu/burgelab/masent/Xmaxentscan_scoreseq.html)) [34], ESEFinder 3.0 software (<http://rulai.cshl.edu>)[35] and RBPmap (<http://rbpmap.technion.ac.il/>) [36].

The presence of repetitive sequences was analysed using RepeatMasker (<http://www.repeatmasker.org/>).

**Processing of CLIP datasets.** Raw SRSF1 eCLIP reads [37] were downloaded from the ENCODE website and trimmed twice using cutadapt v1.15 [38]. Raw PAR-CLIP reads were downloaded from the GEO repository using the accession number GSE71095 [39] and trimmed using cutadapt. Raw HITS-CLIP reads were downloaded from the GEO repository using the accession numbers GSE131745 [40] and trimmed using cutadapt. Subsequently, reads were mapped with STAR v2.7.8a [41] to the human genome using GENCODE v39 hg38 annotations allowing for up to 100 mapping locations. Run parameters and further details are described in Supplementary methods.

**Visualization of CLIP reads.** We used Gviz [42] to visualize the alignments along with GenomicFeatures [43] to parse the GENCODE GTF file with transcripts.

## Results

### Clinical, biochemical and genetic analysis

Patient 1 was detected in newborn screening due to high Phe levels (226  $\mu$ M) and later diagnosed as the severe form of PTPS deficiency based on urinary pterins (elevated neopterin, 22,9 mmol/mol creatinine and low biopterin, 0.3 mmol/mol creatinine) and monoamine neurotransmitter deficiency in CSF, corresponding to very low levels of homovanillic acid (HVA) and of 5-hydroxyindolacetic acid (5-HIAA), (140 nM and 75 nM, respectively; for reference values see [44]). Prolactin, which is physiologically inhibited by dopamine, showed very high values (655-1590  $\mu$ U/ml; reference values 127-637  $\mu$ U/ml). Treatment with sapropterin (BH<sub>4</sub>), 5-OH Trp and L-

Dopa/carbidopa, adjusted to body weight, was implemented. Currently, the patient is 6 years old and developing normally, with no overt clinical symptoms.

Genetic analysis by exome sequencing identified a single variant in the *PTPS* gene: c.243+3A>G, affecting the 5'ss of exon 4 (decrease in HSF splicing score from 97.71 to 81.74, and from 8.7 to 4.4 in MaxEnt score). PTPS deficiency is inherited in autosomal recessive fashion, so we sought to find the second pathogenic variant, which might reside within regions not covered by routine exome sequencing. We used Sanger sequencing to analyze a candidate region in intron 2 in which two different pathogenic variants were previously shown to activate a series of overlapping pseudoexons [26, 28]. This analysis revealed the presence of a novel variant c.164-672C>T that creates a potential 5'ss (HSF splicing score 75.85 and MaxEnt 5.8), downstream of the previously identified pseudoexons. RT-PCR analysis in patient's fibroblasts revealed several aberrant bands of higher and lower molecular weight compared to the correctly spliced band, which was detected in low amounts (Figure 1A). The lower sized bands were prominent in samples treated or not with puromycin, used as NMD inhibitor. After subcloning and sequencing we identified transcripts with exon 3 skipping (out of frame, p.(Val55Aspfs\*2), occurs naturally due to poor exon definition, see [45]), exon 4 skipping (in frame, p.(Ile63\_Glu81del)) and skipping of both exons 3 and 4 p.(Val55Glyfs\*8).

In puromycin treated samples faint high molecular weight bands were observed, which were identified by subcloning and sequencing analysis as insertion of several intronic sequences (pseudoexons) of 45, 70, 93 and 118 bp between exons 2 and 3. Pseudoexons of 45 and 70 bp (PE45 and PE70) have been previously reported to be activated by different variants in the region [26] and those of 93 and 118 bp (PE93 and PE118) correspond to the use of the newly created 5'ss by variant c.164-672C>T, in combination with upstream cryptic 3'ss (Figure 1B). All correspond to exonizations of an antisense *AluSq* sequence, as previously described [28]. The newly created 5'ss corresponds to *Alu* position 156, present in naturally exonized *Alus* [46, 47]. Interestingly,

by altering the conserved C at position 156 in the *AluY* consensus sequence the *PTS* c.164-672C>T variant alters this to a 5'ss, identical to that used in the disease associated *F8* pseudoexon [24, 48]. The *PTS* PE118 is nearly identical to the *F8* pseudoexon with the same 3'ss and 5'ss, consistent with the fact that this seems to be the most prominent *PTS* PE isoform. Insertion of *PTS* PE45 and PE93 result in extra amino acid insertions (p.(Val55\_Val56ins15) and p.(Val55\_Val56ins31), respectively), while insertion of PE70 and PE118 result in frameshift and introduction of premature stop codons (p.(Val55Serfs\*33) and p.(Val55Aspfs\*37), respectively).

### **Functional analysis in minigenes of the novel variants**

The individual effect on splicing of the novel variants was studied using minigenes transfected in Hep3B cells. Variant c.164-672C>T was analyzed in the previously described pCR3.1 minigene encompassing *PTS* exons 2 to 4 [26]. The mutant minigene produces, in addition to a normal sized, correctly spliced band, larger bands which after subcloning and sequencing were found to correspond to the insertion of PE70, PE93 and PE118, previously identified in patients' fibroblasts (Figure 2A). Thus, this variant is responsible for the activation of the intronic pseudoexons, although some residual normal splicing is produced. Variant c.243+3A>G was analyzed in a minigene with exon 4 and flanking intronic sequences cloned in the pSPL3 vector. As shown in Figure 2B the c.243+3A>G variant causes exon 4 skipping, consistent with the predicted decrease in splice score of the mutant 5'ss and with the results obtained in fibroblasts. A faint band corresponding to the correctly spliced transcript is barely detectable, indicating that the normally spliced band observed in patient's fibroblasts (Figure 1A) is probably mostly derived from the allele carrying the c.164-672C>T variant, which can be thus considered a "leaky" splicing variant.

The recognition of the pseudoexons by the splicing machinery may vary depending on the tissue analyzed, due to tissue specific differences in splice factors, as evidenced in several studies [12, 49]. Therefore, we subsequently analyzed the effect on

splicing of both variants c.243+3A>G and c.164-672C>T in SH-SY5Y neuroblastoma cells, given the relevant role of the *PTS* gene in the central nervous system. The results showed a similar splicing pattern for both variants (Figure 2A, B).

### **Pseudoexon exclusion with antisense oligonucleotides**

Pathogenic pseudoexon inclusion resulting from deep intronic variants has the potential to be reverted using antisense therapy with SSOs. We designed six SSOs targeting the intronic region in which the different pseudoexons are located (Figure 3A). A scrambled SSO was used as a non-target control to discriminate between specific and unspecific effects. In this study we used full 2-OMe PS oligonucleotides, with SSO 1 having the wild-type sequence of the morpholino oligonucleotide (PTS-AMO3) used in a previous study to target PE45 and PE70 activated by variant c.164-716A>T (designated as c.164-712A>T in the previous study) [26]. SSO1 and SSO4 cover the 5'ss used by both these pseudoexons. SSO2 and SSO6 were designed for targeting the new 5'ss created by the novel variant c.164-672C>T. SSO3 targets the TACAACCTC element previously suggested to be essential for PE45 inclusion [28]. SSO5 targets predicted potential binding sites for different SR splice factors. Minigenes carrying variant c.164-672C>T were co-transfected with SSOs at 50 nM and RT-PCR analysis showed that SSOs 1, 5 and 6 completely prevented pseudoexons inclusion (Figure 3B). For cells transfected with SSOs 2, 3 and 4 a faint band corresponding in size to PE70, which is also detectable in the wild-type untreated minigene, remained after treatment (Figure 3B).

The same SSOs were also used with the mutant minigene carrying the previously studied variant c.164-716A>T, that results in the inclusion of PE45 and PE70 [26]. Again, SSOs 1, 5 and 6 completely prevented pseudoexons inclusion and, in this case, SSO 2, 3 and 4 had no apparent effect (Figure 3C).

We next transfected patient 1 fibroblasts, treated with puromycin to inhibit NMD, with the six SSOs. For the sake of clarity, in the RT-PCR analysis we used a reverse

primer hybridizing to exon 3, to avoid detection of the exon skipping transcripts resulting from the allele with variant c.243+3A>G. The results were concordant with what we observe with minigenes, as SSO 1, 5 and 6 resulted in almost complete disappearance of the higher molecular weight bands, corresponding to pseudoexons inclusions, while with SSOs 2, 3 and 4 a band presumably corresponding to PE70 remained (Figure 4A).

The effect of the six SSOs was also analyzed in patient 2 fibroblasts, heterozygous for c.164-716A>T that causes insertion of PE45 and PE70 [26]. Again, SSOs 1, 5 and 6 prevented pseudoexons inclusions and SSOs 2, 3 and 4 showed no significant effect (Figure 4B). In both patient 1 and patient 2 fibroblasts, a scrambled SSO had no effect on the splicing pattern.

Examining the hit SSOs we observe that SSO1 covers a sequence shared by all pseudoexons; however, SSOs 5 and 6 target sequences present only in the longer PE93 and PE118. In order to precisely define and quantify the effect of each of the SSOs on the inclusion of the different pseudoexons we performed deep sequencing of the RT-PCR products from the transfected fibroblasts. The results are shown in Figure 4C, D. In patient 1 cells, we observe similar levels (estimated by the number of reads) for PE70, PE93 and PE118, plus a low amount of PE45 (Figure 4C). PE70 is predominant in patient 2 (Figure 4D) and is also detectable in low amounts in control fibroblasts. SSO1 and 6 almost completely prevent all pseudoexons inclusion in both patients' fibroblasts, with SSO5 exhibiting a slightly lower effect. SSOs 2, 3 and 4 prevalently hinder the inclusion of PE93 and PE118 and correspondingly have no effect in patient 2 cells (Figure 4B, D).

### **Analysis of splice factors binding to pseudoexon sequences**

SSO1 hybridizes to a region included in all pseudoexons, and its effect could be the result of blocking the binding of splice factors critical for their aberrant inclusion. The positive effect of SSOs 5 and 6 that lie downstream of the 5'ss (for PE45 and PE70) or



internal (PE93 and PE118) to the pseudoexons may also indicate the presence of splicing regulatory sites needed for pseudoexon recognition.

*In silico* analysis of the intronic sequences predict strong SRSF1 binding motifs in the region covered by hit SSOs 1, 5 and 6 (scores 4,3 and 4,6 according to ESEfinder) (Figure 3A). To further investigate which splice factors bind to sequences targeted by SSOs 1, 5 and 6, we performed RNA affinity pull downs on biotinylated RNA oligonucleotides, containing the wild-type target sequences or with mutations that abolish the SRSF1 motifs (Figure S1). After incubation in HeLa cell nuclear extract, SR and hnRNPA1 proteins bound to each oligonucleotide were analysed by SDS-PAGE and Western blotting. Using an anti-SR antibody there was no binding above background for SRSF2, SRSF3, SRSF4, SRSF5 or SRSF6 (data not shown). In addition, the pull downs with all RNA oligonucleotides showed little SRSF1 binding, although for PTS5 (antisense sequence to SSO5) we observe a stronger signal, which could be decreased by mutating the SRSF1 motif (Figure S1). SPRi analysis of PTS5 shows binding of SRSF1 above background. When the potential SRSF1 binding site in PTS5 is disrupted by point mutations this binding is disrupted (Figure S2). SPRi analysis of PTS1 (sequence targeted by SSO1) and PTS6 (sequence targeted by SSO6) confirm the very low binding of SRSF1 observed in the pull down experiments (Figure S2). Examination of the sequences employing recombinant hnRNPA1 in the SPRi analysis indicated binding for all RNA oligonucleotides, confirming that they are all functional (Figure S2).

To provide additional evidence of SRSF1 binding to the targeted sequences we examined publicly available SRSF1 CLIP datasets, including the K562 and HepG2 ENCODE eCLIP datasets [37], a PAR-CLIP dataset from HeLa cells [39] and a HITS-CLIP dataset from MDA-LM2 cells [40]. Normally, CLIP datasets are analyzed by considering only the uniquely aligned reads, but this overlooks real binding sites in duplicated sequences. Because the *PTS* PE is derived from the highly repeated *AluSq* sequence [28], reads would therefore normally be filtered away. We therefore aligned

CLIP reads allowing for up to 100 multiple mapping locations to allow for mapping of reads to duplicated regions, including *AluY* sequence elements. This resulted in reads aligning at or on top of the potential SRSF1 binding sites in *PTS* (Figure S3), providing further evidence that these binding sites may be functional.

### **PTPS protein recovery with SSO treatment**

SSOs 1, 5 and 6 corrected the missplicing defect in patients 1 and 2 fibroblasts carrying deep-intronic variants activating pseudoexons in intron 2. To illustrate the therapeutic potential of the approach, we analysed PTPS protein levels in cells treated with the different SSOs or a scrambled SSO. As using an NMD inhibitor may favour correctly spliced transcripts, interfering with the interpretation of the results, we analysed cells treated or not with puromycin. Both fibroblast cell lines are compound heterozygous, carrying missplicing variants in both alleles, although some residual normal splicing can be detected (Figure 1A). Western blot analysis revealed no detectable PTPS protein in untreated cells (Figure 5). This confirms that in fibroblasts the resulting misspliced transcripts do not result in stable protein and that the protein translated from the residual normally spliced transcripts escapes detection limits with the antibody used. After transfection with the hit SSOs PTPS protein is detected in both fibroblast cell lines, in cells treated or not with puromycin. Protein levels in patients' cells transfected with the different SSOs, estimated by densitometric analysis, ranged between 5 to 50%, relative to levels in control fibroblasts, depending on the SSO, treatment with or without puromycin and the cell line (Figure 5).

### **Discussion**

Deep intronic variants that alter splicing are missed with standard exon-based diagnostic sequencing approaches. However, evidence has accumulated that such variants are an important class of disease-causing variants mainly due to the activation

of pseudoexons [9, 16]. Thus, it has been proposed that RNA analysis should be incorporated into the diagnostic workflow for patients with specific phenotypes and with no or incomplete genetic diagnosis [13]. In the present study routine clinical exome sequencing identified only one pathogenic *PTS* variant in patient 1 with a biochemical suspicion of PTPS deficiency. We subsequently used a candidate region approach, sequencing a deep-intronic region where previous studies had identified variants activating pseudoexons, followed by *PTS* transcript analysis in patients' fibroblasts. Interestingly, the analysis identified a novel intronic variant, c.164-672C>T, adding to the previously reported variants c.163+1695\_1751del57 and c.164-716A>T in the same region [26]. The novel variant results in the exonization of previously described PE45 and PE70, plus two longer, overlapping PE93 and PE118. These two are defined by the newly created 5'ss and the 3'ss of the previously reported pseudoexons. Multiple pseudoexons inclusion due to a single intronic splice site coupled with different cryptic acceptor splice sites has already been described [50]. However, in this case, the novel variant also moderately increases the inclusion levels of upstream PE45 and PE70 compared to that observed in control fibroblasts, which could be due to the alleviation of some repressive mechanism, rather than to the occurrence of an activating event, as suggested in other cases [51]. In addition, some introns seem more prone to harbor disease-causing mutations than others, as observed in the *ABCA4* gene in which the clustering in certain introns of deep-intronic variants causing Stargardt disease has been reported [52]. In the present study, the pseudoexons correspond to transposable *Alu* elements, which may lead to the creation of new exons in the genome or to aberrant disease-causing transcripts due to single point variants [8, 21].

A very interesting aspect of our results is the fact that the new *PTS* c.164-672C>T variant creates a 5'ss identical to that used in the disease causing *Alu* derived *F8* pseudoexon that is activated by a deletion removing an inhibitory sequence in the 3'ss [24], and this 5'ss in *F8* is also activated when further strengthened by a point mutation

altering the +3G>A also leading to disease [24, 48]. Moreover, mutations in this sequence create the 5'ss most frequently used in naturally exonized *Alu*s [46, 47]. The *PTS* PE118, the most prominent pseudoexon activated by the c.164-672C>T variant, is nearly identical to the *F8* PE, with exactly the same 3'ss and 5'ss. This underscores that similar mutations in the potential 5'ss, as that here reported in *PTS* and previously in *F8*, in the numerous *AluY* sequence elements present in introns in other genes are likely to be a frequent cause of disease causing pseudoexon activation. Similarly, mutations in the same positions (205 and 206) in the *Alu* sequence as the 5' ss strengthened by the c.164-716A>T *PTS* variant have been reported to activate disease causing pseudoexons in *COL4A3* and *GUSB* genes [53, 54], further supporting that these positions in the *Alu* sequences are high risk sites for disease causing pseudoexon activation.

Generally, most of the reported variants that activate pseudoexons either create or strengthen a pre-existing suboptimal 5' or 3' ss [8, 48], although in other pseudoexons, different types of activating variants have been identified [10, 11]. In the *PTS* intron 2 region, one variant creates a novel 5'ss (c.164-672C>T), another strengthens a pre-existing one (c.164-716A>T) and another brings closer a branch point sequence to the 3'ss (c.163+1695\_1751del57) [26], revealing the spectrum of rare deleterious variants activating pseudoexons in this region.

Following the identification of the aberrant insertions, SSOs were designed with the aim of restoring normal splicing by forcing the skipping of the pathogenic pseudoexons. This strategy has been validated in minigene assays and in cultured patient-derived cells in a wide range of disorders, and has efficiently shown the rescue of these loss-of-function variants [8, 22, 30]. Our results showed that both in minigenes and in the endogenous gene context using patients' fibroblasts, SSOs 1 and 6 (and to a slightly lesser extent, SSO5) were most effective for all *PTS* intron 2 pseudoexons, providing support for the potential of using a single SSO to treat patients with different pathogenic variants in the region. The relative amounts of each pseudoexon in the final

transcript profile in patients' fibroblasts were measured using deep sequencing. This approach represents an efficient way to precisely quantify the efficacy of SSO-mediated splice correction for adjacent or overlapping pseudoexons.

The presence of SRSF1 motifs in the pseudoexon sequences indicated a possible mechanism of action of the hit SSOs, whereby they would block binding of this splice factor to the predicted sites, thus precluding pseudoexons inclusion, as previously described [55, 56]. Taken together, our results using RNA pulldowns and SPRi and analysis of CLIP data suggest that SSO5 could indeed function by blocking access to an SRSF1 ESE (CTCCCGAGTA) with two predicted overlapping binding sites (CTCCCGA - Score 4.56) and (CCGAGTA - Score 2.59) (Figure 3A). When these sites are abolished by point mutations, binding decreases (Figures S1 and S2). SSO6 covers most of the high score (4.56) SRSF1 motif, thus it is plausible that it will block SRSF1 binding and cause exon skipping, although as the entire motif is not present, experimentally we observe low SRSF1 binding to the target sequence (Figures S1 and S2).

SSO1 blocks the 5'ss of PE45 and PE70 which may contribute to the exclusion of these pseudoexons. In addition, it is known that secondary RNA structure can affect pre-mRNA splicing, regulating accessibility of cis-acting or trans-acting elements and the spatial distance in between [57]. In this case, SSOs binding may also influence local RNA structure favoring pseudoexon exclusion.

SSO6 targets a region that encompasses the mutated nucleotide at position c.164-672C>T and, therefore, contains a mismatch towards the sequence in patient 2. Contrary to what has been reported in some studies [52, 58], this single mismatch between the SSO and its target sequence apparently does not affect its splicing modulation capacity. Similarly, SSO1 with the wild-type sequence functions in patient 2 samples, despite the presence of a mismatch at the site of the c.164-716A>T variant. Of note, it can be envisaged that these same hit SSOs can be used for pseudoexon activating variants in similar *A/u*-derived activated pseudoexons, as the above mentioned

nearly identical in the *F8* gene. This implies for the first time, to our knowledge, that one single SSO could potentially be used for treating patients with different diseases.

Comparing the results obtained in a previous study [26], in which the PMO chemistry was used at concentrations 20-30  $\mu$ M, we observe similar splicing correction (pseudoexons exclusion) in patient 2 fibroblasts transfected with SSO1, same sequence as PTS-AMO3, but with 2-OMe-PS chemistry (see Figure 3 in [26], as compared to Figure 4, this work). Both chemistries have proven to be effective *in vitro*, although *in vivo*, the neutral charge of PMOs improves tolerability as it results in lower binding to plasma proteins. Currently, 3 SSOs with PMO chemistry have received FDA approval for the treatment of Duchenne muscular dystrophy [59].

Protein analysis confirmed the recovery of PTPS protein levels after transfection of patient-derived fibroblasts with hit SSOs 1, 5 and 6. Both patients are compound heterozygous, carrying the pseudoexon activating variant in one allele. Although PTPS enzymatic activity could not be determined in this study, the levels of protein achieved here have been previously shown to correlate with a therapeutically relevant increase in enzymatic activity [26].

It must be noted that for rare pathogenic variants, occurring only in one or two families, developing a personalised genetic therapy represents a challenge, although the proof of concept of the feasibility of such an approach for a single patient with a deep intronic variant was recently reported [25] and more n-of-1 RNA therapy trials are being developed in academic or pharmaceutical settings [60, 61]. Indeed, the recently founded n-Lorem Foundation aims to discover and develop antisense oligonucleotides for patients with unique mutations and to provide them for free as treatment for life [62]. Another important hurdle in preclinical development for SSOs targeting pseudoexons is the difficulty in generating adequate animal models, as intronic sequences are not conserved among species. However, with the advent of CRISPR/Cas gene editing technologies it is now feasible to easily generate humanized mouse models carrying the

desired intronic pseudoexon sequences, although reproducibility of the splicing defect may depend on the whole genomic context and thus should be first confirmed *in vitro* (e.g. using minigenes or a similarly edited mouse cell line).

PE70 that corresponds to the “natural” predicted *Alu* sequence [28] is detectable in low abundance in the wild type context, as evidenced in minigene assays (Figures 2 and 3) and in control fibroblasts (Figure 4). Residual pseudoexon inclusion leading to a non-functional transcript in control individuals has been detected for several genes involved in disease [63, 64]. It was initially considered an undesired unproductive splicing event. However, evidence has accumulated indicating it is actually a well-represented mechanism of gene expression regulation [65]. During specific dynamic cellular processes, genes may be subjected to alternative splicing, resulting in such non-functional transcripts rapidly regulating final mRNA and protein levels [66]. Recently, this knowledge was exploited applying antisense technology to prevent non-productive pseudoexon inclusion events, resulting in an increase in functional mRNA and protein levels in monogenic diseases characterized by haplo-insufficiency [67, 68]. The authors detected by bioinformatic analysis more than 1,000 non-productive splicing events in disease associated genes and using this approach, termed TANGO (targeted augmentation of nuclear gene output), confirmed an increase in productive mRNA and protein for different genes *in vitro* and *in vivo* [67]. The antisense approach targeting naturally occurring pseudoexon inclusion events could be applied to the *PTS* gene in the case of missense variants retaining residual activity. Over 100 variants have been reported in the *PTS* gene associated with PTPS deficiency, mostly missense variants [69], many of them affecting protein structure and stability [70-72]. It is also known that the milder, peripheral forms of PTPS deficiency are associated to partial activity [73]. Future experiments to investigate the feasibility of the TANGO approach can be considered using fibroblasts derived from patients with such genotypes.

The present work adds to the current knowledge of the relatively high prevalence of variant induced pseudoexon inclusion as a disease mechanism in inherited metabolic disorders [74]. These variants are promising candidates for personalized therapies using SSOs, which, along with other RNA-therapies, have registered a boost in the last few years, with more than a dozen drugs approved for clinical use and many other currently in different stages of development. It can be anticipated that this type of drugs may shortly be in use for one or very few patients (personalized medicine) or to treat common diseases, once delivery and toxicity issues are solved [19].

### **Acknowledgements**

The excellent technical assistance of E. Montalvo and M. Alvarez and of A. Rassi from the Division of Clinical Chemistry and Biochemistry at the University Children's Hospital Zürich, is gratefully acknowledged, as well as the expert advice of R. Ramos from Fundación Parque Científico de Madrid.

### **Conflict of interest**

The authors declare no conflict of interest.

### **Funding**

This work was funded by grant PID2019-105344RB-I00/AEI/10.13039/501100011033 from Spanish Ministry of Science and Innovation, COST (European Cooperation in Science and Technology) Action CA17103, Fundación Isabel Gemio, Fundación La Caixa (LCF/PR/PR16/11110018) and grants to BSA from the Danish Medical Research Council (FSS) (No. 9039-00281B) and the Novo Nordisk Foundation (NNF19OC0058588). Centro de Biología Molecular Severo Ochoa receives an institutional grant from Fundación Ramón Areces.



## REFERENCES

1. Werner, E.R. et al. (2011) Tetrahydrobiopterin: biochemistry and pathophysiology. *Biochem J* 438 (3), 397-414.
2. Blau, N. et al. (2001) Disorders of tetrahydrobiopterin and related biogenic amines. In *The Metabolic and Molecular Basis of Inherited Disease* (8th edn) (Scriver, C.R. et al. eds), pp. 1725-1776, McGraw-Hill.
3. Manzoni, F. et al. (2020) Retrospective analysis of 19 patients with 6-Pyruvoyl Tetrahydropterin Synthase Deficiency: Prolactin levels inversely correlate with growth. *Mol Genet Metab* 131 (4), 380-389.
4. Baralle, D. and Buratti, E. (2017) RNA splicing in human disease and in the clinic. *Clin Sci (Lond)* 131 (5), 355-368.
5. Scotti, M.M. and Swanson, M.S. (2016) RNA mis-splicing in disease. *Nat Rev Genet* 17 (1), 19-32.
6. Andresen, B.S. and Krainer, A.R. (2009) When the genetic code is not enough-how sequence variations can alter pre-mRNA splicing and cause (complex) disease. In *Genetics of Human Complex Diseases* (Almasy, L. and Al-Chalabi, A. eds), pp. 165-182, Cold Spring Harbor Laboratory Press.
7. Soemedi, R. et al. (2017) Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet* 49 (6), 848-855.
8. Dhir, A. and Buratti, E. (2010) Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *Febs J* 277 (4), 841-55.
9. Vaz-Drago, R. et al. (2017) Deep intronic mutations and human disease. *Hum Genet* 136 (9), 1093-1111.
10. Petersen, U.S.S. et al. (2022) Pseudoexon activation in disease by non-splice site deep intronic sequence variation - wild type pseudoexons constitute high-risk sites in the human genome. *Hum Mutat* 43, 103-127.

11. Keegan, N.P. et al. (2022) Analysis of pathogenic pseudoexons reveals novel mechanisms driving cryptic splicing. *Front Physiol* 12, 806946.
12. Cummings, B.B. et al. (2017) Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 9 (386).
13. Kremer, L.S. et al. (2017) Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* 8, 15824.
14. Gonorazky, H.D. et al. (2019) Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet* 104 (5), 1007.
15. Fresard, L. et al. (2019) Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med* 25 (6), 911-919.
16. Truty, R. et al. (2021) Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *Am J Hum Genet* 108 (4), 696-708.
17. Crooke, S.T. et al. (2018) RNA-Targeted Therapeutics. *Cell Metab* 27 (4), 714-739.
18. Aartsma-Rus, A. (2018) Genetic therapies for spinal muscular atrophy type 1. *Lancet Neurol* 17 (2), 111-112.
19. Hammond, S.M. et al. (2021) Delivery of oligonucleotide-based therapeutics: challenges and opportunities. *EMBO Mol Med* 13 (4), e13243.
20. Crooke, S.T. et al. (2021) Antisense Technology: A Review. *J Biol Chem*, 100416.
21. Vorechovsky, I. (2010) Transposable elements in disease-associated cryptic exons. *Hum Genet* 127 (2), 135-54.
22. Perez, B. et al. (2010) Present and future of antisense therapy for splicing modulation in inherited metabolic disease. *J Inherit Metab Dis* 33 (4), 397-403.
23. Sangermano, R. et al. (2019) Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* 21 (8), 1751-1760.
24. Jourdy, Y. et al. (2018) Recurrent F8 Intronic Deletion Found in Mild Hemophilia A Causes Alu Exonization. *Am J Hum Genet* 102 (2), 199-206.

25. Kim, J. et al. (2019) Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N Engl J Med* 381 (17), 1644-1652.
26. Brasil, S. et al. (2011) Pseudoexon exclusion by antisense therapy in 6-pyruvoyl-tetrahydropterin synthase deficiency. *Hum Mutat* 32 (9), 1019-27.
27. Bonafe, L. et al. (2001) Diagnosis of dopa-responsive dystonia and other tetrahydrobiopterin disorders by the study of biopterin metabolism in fibroblasts. *Clin Chem* 47 (3), 477-85.
28. Meili, D. et al. (2009) Disease-causing mutations improving the branch site and polypyrimidine tract: pseudoexon activation of LINE-2 and antisense Alu lacking the poly(T)-tail. *Hum Mutat* 30 (5), 823-31.
29. Aartsma-Rus, A. et al. (2009) Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms. *Mol Ther* 17 (3), 548-53.
30. Albert, S. et al. (2018) Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am J Hum Genet* 102 (4), 517-527.
31. Martinez-Pizarro, A. et al. (2018) Intronic PAH gene mutations cause a splicing defect by a novel mechanism involving U1snRNP binding downstream of the 5' splice site. *PLoS Genet* 14 (4), e1007360.
32. Holm, L.L. et al. (2022) Vulnerable exons, like ACADM exon 5, are highly dependent on maintaining a correct balance between splicing enhancers and silencers. *Hum Mutat* 43, 253-265.
33. Desmet, F.O. et al. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37 (9), e67.
34. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11 (2-3), 377-94.
35. Cartegni, L. et al. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31 (13), 3568-71.

36. Paz, I. et al. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* 42 (Web Server issue), W361-7.
37. Van Nostrand, E.L. et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13 (6), 508-14.
38. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17 (1), 10-12.
39. Xiao, W. et al. (2016) Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* 61 (4), 507-519.
40. Fish, L. et al. (2019) Nuclear TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay. *Mol Cell* 75 (5), 967-981 e9.
41. Dobin, A. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15-21.
42. Hahne, F. and Ivanek, R. (2016) Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol* 1418, 335-51.
43. Lawrence, M. et al. (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol* 9 (8), e1003118.
44. Blau, N., Spronsen FJ (2014) Disorders of phenylalanine and tetrahydrobiopterin metabolism. In *Physician's guide to the diagnosis, treatment and follow-up of inherited metabolic diseases* (Blau, N., Duran, M, Gibson MK, Dionisi-Vici C ed).
45. Liu, T.T. et al. (2001) Identification of three novel 6-pyruvoyl-tetrahydropterin synthase gene mutations (226C>T, IVS3+1G>A, 116-119delTGTT) in Chinese hyperphenylalaninemia caused by tetrahydrobiopterin synthesis deficiency. *Hum Mutat* 18 (1), 83.
46. Sela, N. et al. (2007) Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol* 8 (6), R127.

47. Sorek, R. et al. (2004) Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. *Mol Cell* 14 (2), 221-31.
48. Dericquebourg, A. et al. (2020) Identification of new F8 deep intronic variations in patients with haemophilia A. *Haemophilia* 26 (5), 847-854.
49. Palhais, B. et al. (2016) The prevalent deep intronic c. 639+919 G>A GLA mutation causes pseudoexon activation and Fabry disease by abolishing the binding of hnRNPA1 and hnRNP A2/B1 to a splicing silencer. *Mol Genet Metab* 119 (3), 258-269.
50. Ikezawa, M. et al. (1999) Newly recognized exons induced by a splicing abnormality from an intronic mutation of the dystrophin gene resulting in Duchenne muscular dystrophy. *Mutations in brief* no. 213. Online. *Hum Mutat* 13 (2), 170.
51. Grodecka, L. et al. (2017) Detailed molecular characterization of a novel IDS exonic mutation associated with multiple pseudoexon activation. *J Mol Med (Berl)* 95 (3), 299-309.
52. Khan, M. et al. (2020) Detailed Phenotyping and Therapeutic Strategies for Intronic ABCA4 Variants in Stargardt Disease. *Mol Ther Nucleic Acids* 21, 412-427.
53. Knebelmann, B. et al. (1995) Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum Mol Genet* 4 (4), 675-9.
54. Vervoort, R. et al. (1998) A mutation (IVS8+0.6kbdeITC) creating a new donor splice site activates a cryptic exon in an Alu-element in intron 8 of the human beta-glucuronidase gene. *Hum Genet* 103 (6), 686-93.
55. Palhais, B. et al. (2015) Splice-shifting oligonucleotide (SSO) mediated blocking of an exonic splicing enhancer (ESE) created by the prevalent c.903+469T>C MTRR mutation corrects splicing and restores enzyme activity in patient cells. *Nucleic Acids Res* 43 (9), 4627-39.
56. Homolova, K. et al. (2010) The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to

- pseudoexon activation and causes the cbIE type of homocystinuria. *Hum Mutat* 31 (4), 437-44.
57. Buratti, E. and Baralle, F.E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 24 (24), 10505-14.
58. Aartsma-Rus, A. et al. (2004) Comparative analysis of antisense oligonucleotide analogs for targeted DMD exon 46 skipping in muscle cells. *Gene Ther* 11 (18), 1391-8.
59. Gupta, A. et al. (2021) Nucleic acid delivery for therapeutic applications. *Adv Drug Deliv Rev* 178, 113834.
60. Mullard, A. (2020) N-of-1 drugs push biopharma frontiers. *Nat Rev Drug Discov* 19 (3), 151-153.
61. Kaiser, J. (2021) RNA drugs custommade for rare diseases face wider test. *Science* 374 (6568), 672.
62. Crooke, S.T. (2021) Addressing the Needs of Patients with Ultra-Rare Mutations One Patient at a Time: The n-Lorem Approach. *Nucleic Acid Ther.*
63. Broeders, M. et al. (2020) A Generic Assay to Detect Aberrant ARSB Splicing and mRNA Degradation for the Molecular Diagnosis of MPS VI. *Mol Ther Methods Clin Dev* 19, 174-185.
64. Sakaguchi, N. and Suyama, M. (2021) In silico identification of pseudo-exon activation events in personal genome and transcriptome data. *RNA Biol* 18 (3), 382-390.
65. McGlincy, N.J. and Smith, C.W. (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem Sci* 33 (8), 385-93.
66. Mendell, J.T. et al. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 36 (10), 1073-8.

67. Lim, K.H. et al. (2020) Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. *Nat Commun* 11 (1), 3501.
68. Han, Z. et al. (2020) Antisense oligonucleotides increase Scn1a expression and reduce seizures and SUDEP incidence in a mouse model of Dravet syndrome. *Sci Transl Med* 12 (558).
69. Himmelreich, N. et al. (2021) Molecular and metabolic bases of tetrahydrobiopterin (BH4) deficiencies. *Mol Genet Metab* 133 (2), 123-136.
70. Oppliger, T. et al. (1995) Structural and functional consequences of mutations in 6-pyruvoyltetrahydropterin synthase causing hyperphenylalaninemia in humans. Phosphorylation is a requirement for in vivo activity. *J Biol Chem* 270 (49), 29498-506.
71. Oppliger, T. et al. (1997) Identification of mutations causing 6-pyruvoyl-tetrahydropterin synthase deficiency in four Italian families. *Hum Mutat* 10 (1), 25-35.
72. Muniz, J.R.C. et al. (2019) Role of protein structure in variant annotation: structural insight of mutations causing 6-pyruvoyl-tetrahydropterin synthase deficiency. *Pathology* 51 (3), 274-280.
73. Niederwieser, A. et al. (1987) "Peripheral" tetrahydrobiopterin deficiency with hyperphenylalaninaemia due to incomplete 6-pyruvoyl tetrahydropterin synthase deficiency or heterozygosity. *Eur J Pediatr* 146 (3), 228-32.
74. Perez, B. et al. (2014) Antisense mediated splicing modulation for inherited metabolic diseases: challenges for delivery. *Nucleic Acid Ther* 24 (1), 48-56.

## LEGENDS TO FIGURES

**Figure 1. *PTS* gene splicing profile in patient 1 fibroblasts.** A) RT-PCR analysis in control (C) and patient 1 fibroblasts treated with (+) or without (-) puromycin. Shown on the right are the schematics of the identified bands confirmed by sequencing. B) Sequence of the intron 2 region where the pseudoexons are located, showing the conserved GT and AG splice sites (in bold) used by the pseudoexons and the location of the novel variant. Shown below is the schematics of the identified pseudoexons along with the score of the splice sites according to Human Splice Finder (HSF).

**Figure 2. Minigene analysis of the novel splicing variants identified in patient 1.** A) scheme of the pCR3.1 minigene used and RT-PCR analysis of wild-type (WT) and mutant (c.164-672C>T) minigenes in Hep3B and SH-SY5Y cells. B) Scheme of the pSPL3 minigene used and RT-PCR analysis of wild-type (WT) and mutant (c.243+3A>G) minigenes in Hep3B and SH-SY5Y cells. The star indicates the location of the variants.

**Figure 3. SSO treatment in minigenes.** A) Sequence of the intron 2 region showing the identified pseudoexons, the conserved GT and AG splice sites used, the location of the c.163-716A>T and c.164-672C>T variants and the region targeted by the designed SSOs (black boxes). Binding sites for SRSF1 predicted by ESE finder are shown in light blue boxes along with their scores. B) and C) RT-PCR analysis in wild-type (WT) and mutant minigenes untransfected (-) or transfected with 50 nM of each SSO. SCR: scrambled SSO.

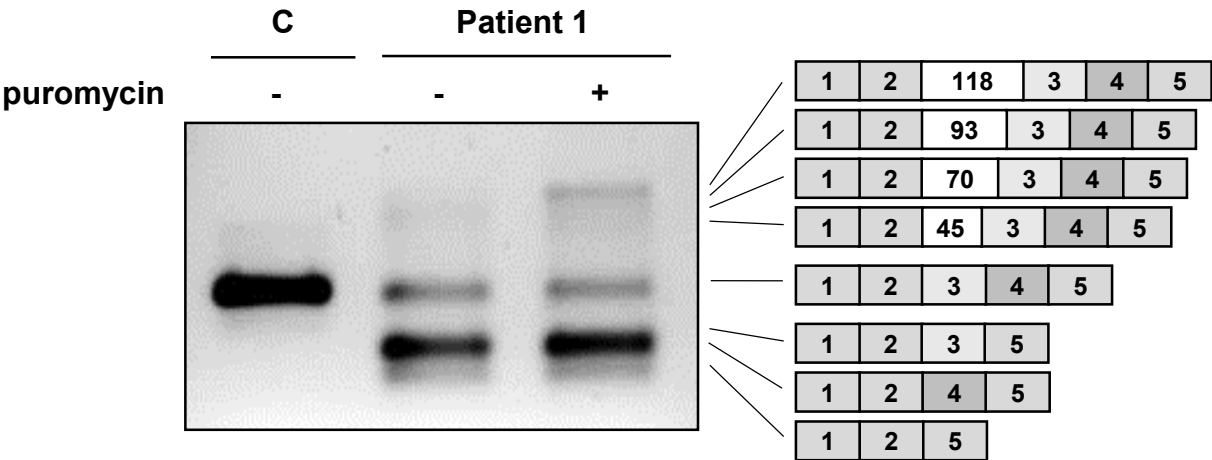
**Figure 4. SSO treatment in patients' fibroblasts.** RT-PCR analysis, using primers to amplify a fragment from exon 1 to exon 3, in control and patient 1 (A) or patient 2 (B) fibroblasts, transfected with 50 nM of each SSO. Puromycin was used as NMD inhibitor. SCR: scrambled SSO. C) and D) show the graphical representation of the quantification of each band corresponding to the transcripts without pseudoexons (non-PE), and to the



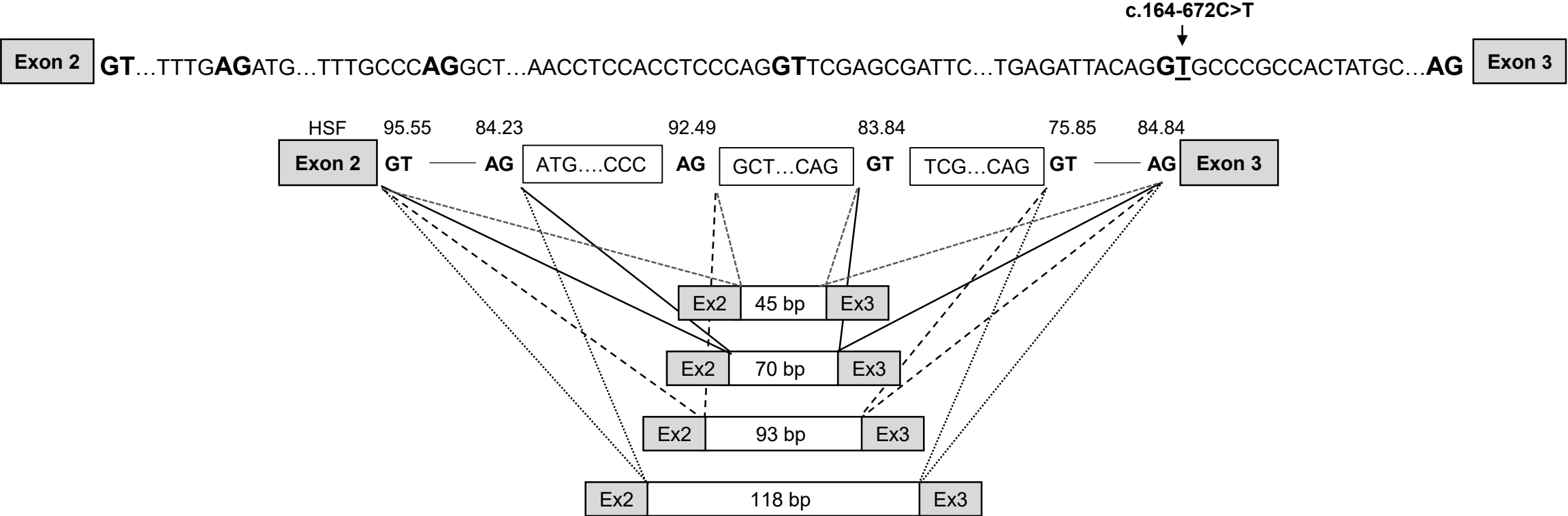
insertion of each pseudoexon, expressed as percentage of reads obtained after deep sequencing the region in puromycin treated samples.

**Figure 5. PTPS protein analysis in patients' fibroblasts.** Western blot analysis of the PTPS protein in untransfected (-) or transfected fibroblasts from patients 1 and 2, untreated (A) or treated (B) with puromycin. GAPDH was used as loading control. C, control fibroblasts. 15 µg of protein extract from control sample and 75 µg of patients' samples were loaded, to facilitate densitometric analysis.

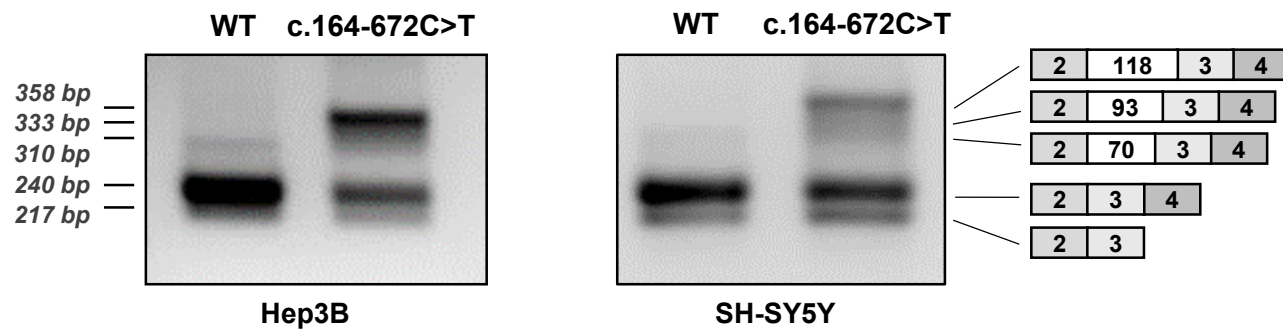
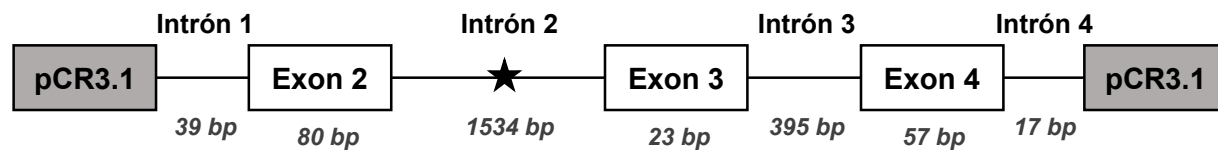
A)



B)



**A)**



**B)**

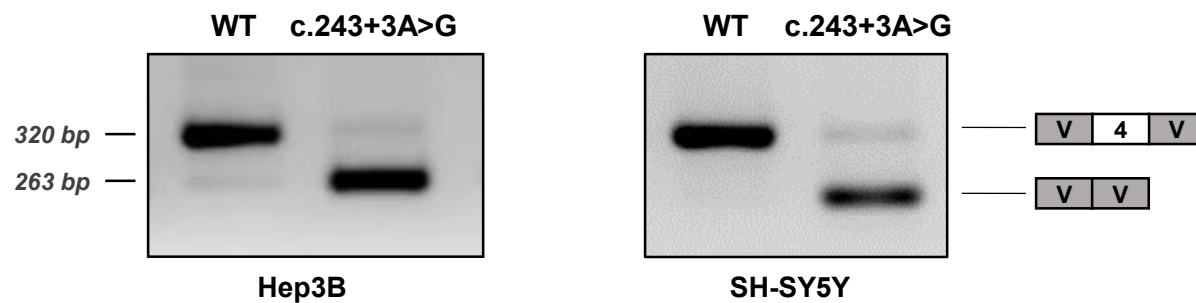
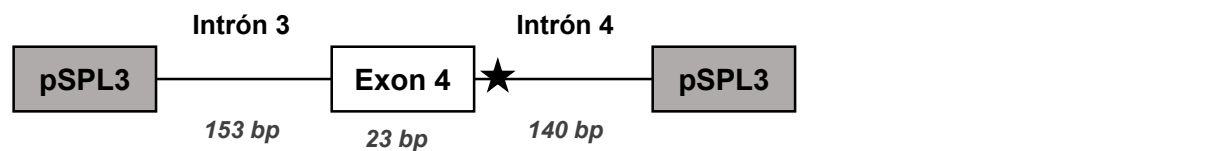
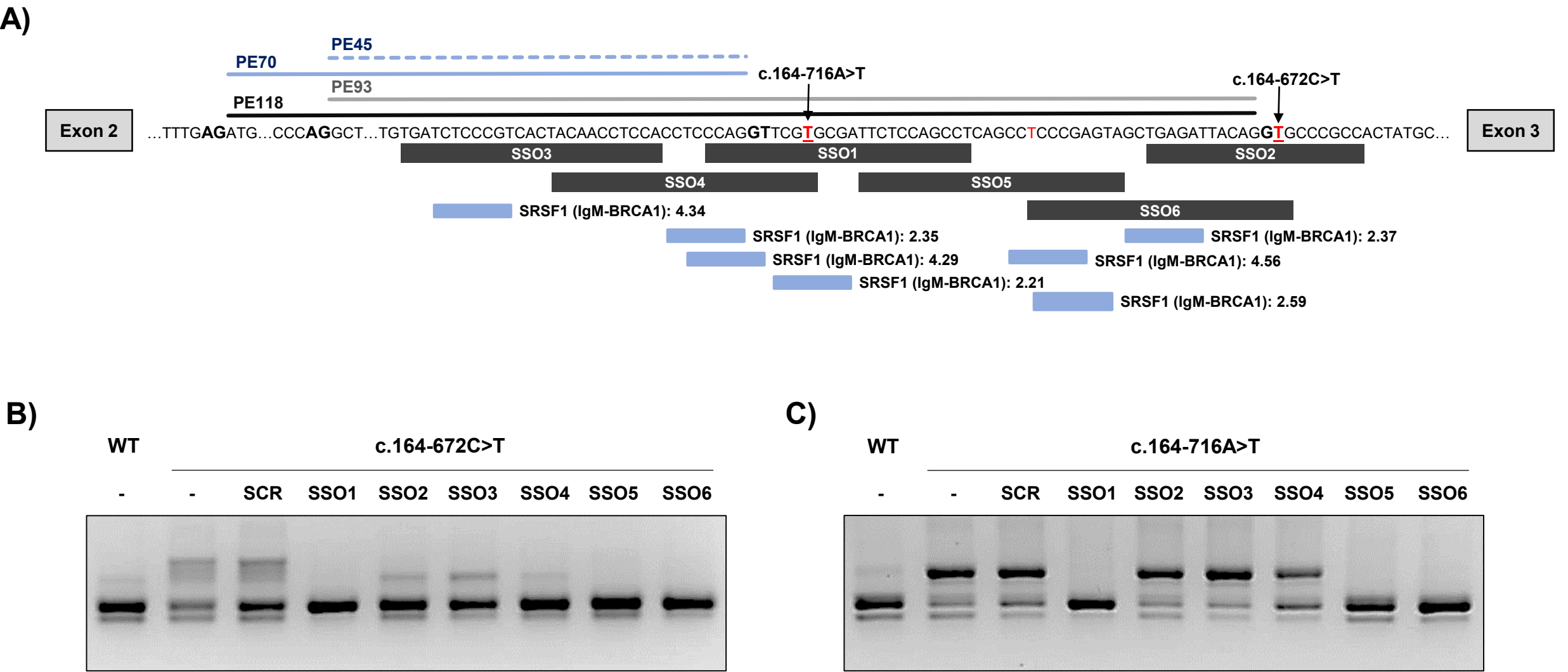
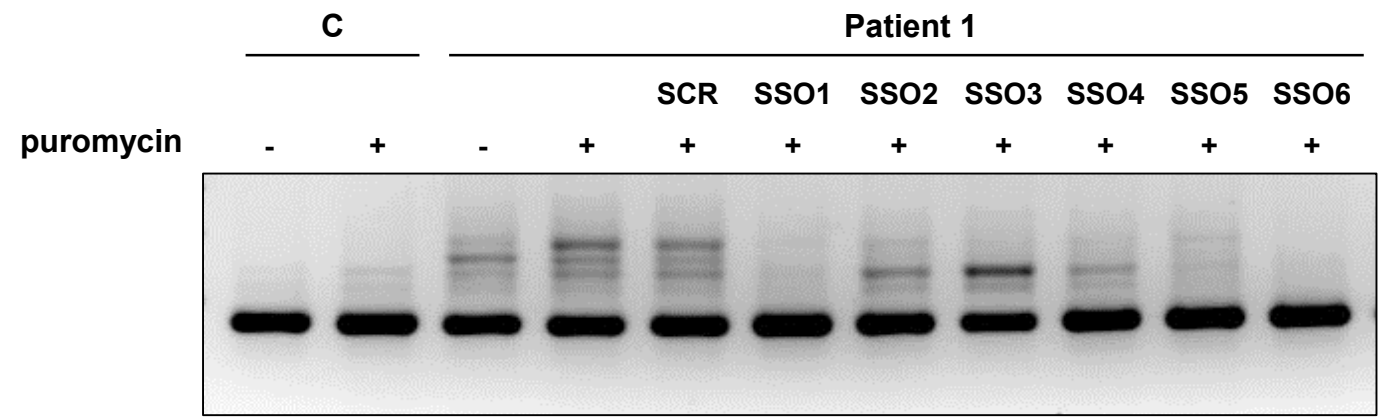


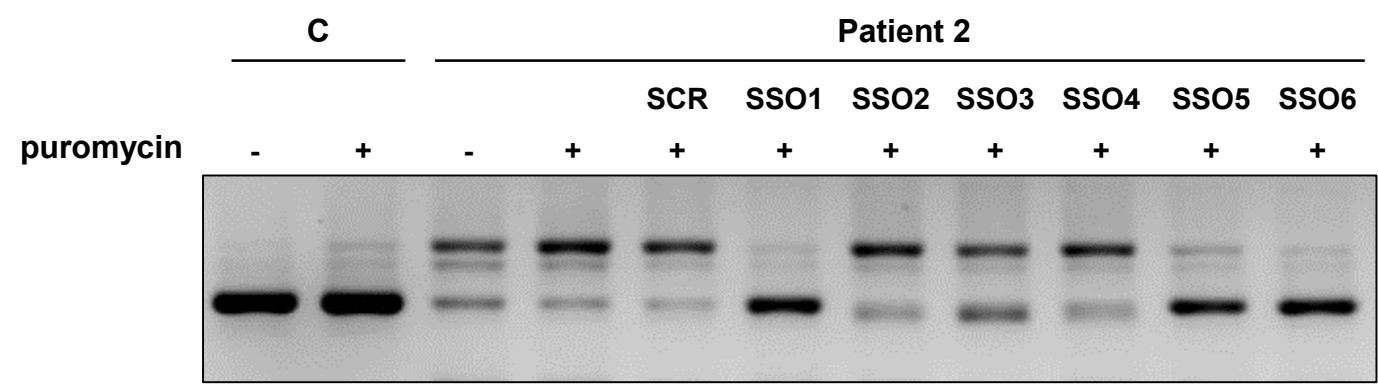
Figure 3



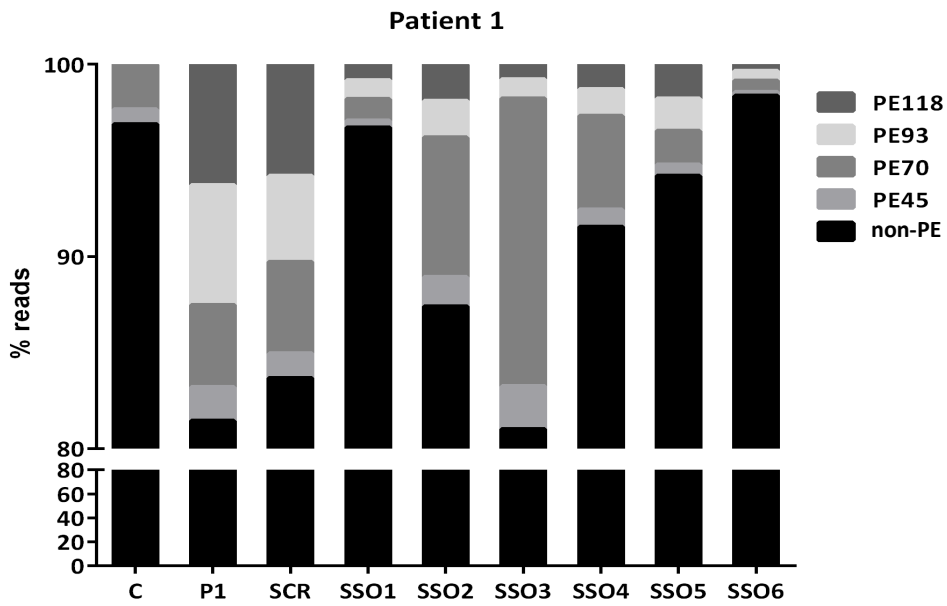
A)



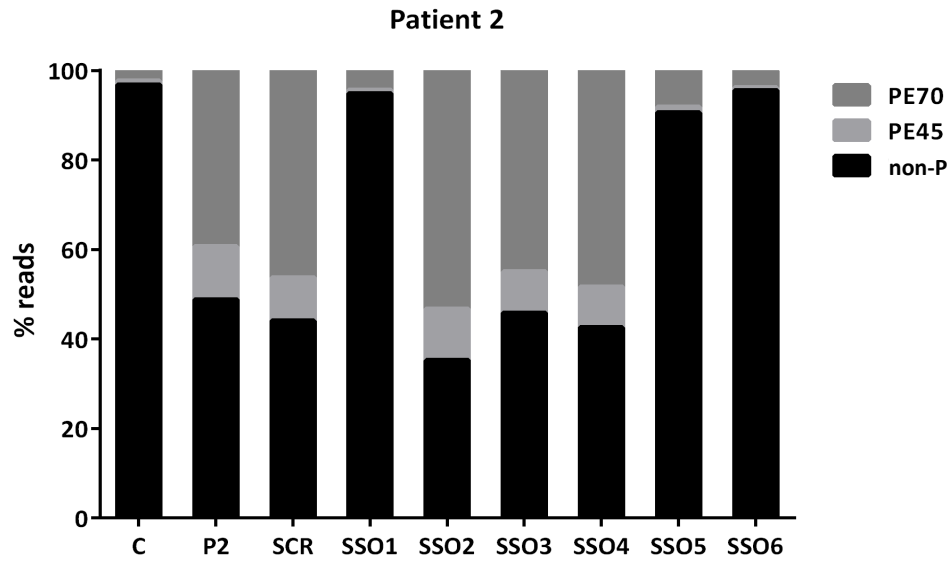
B)



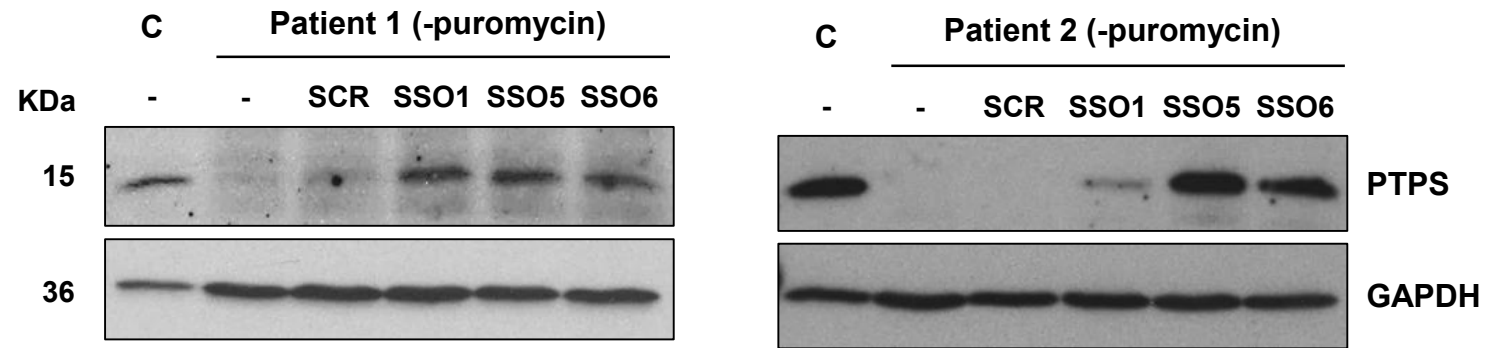
C)



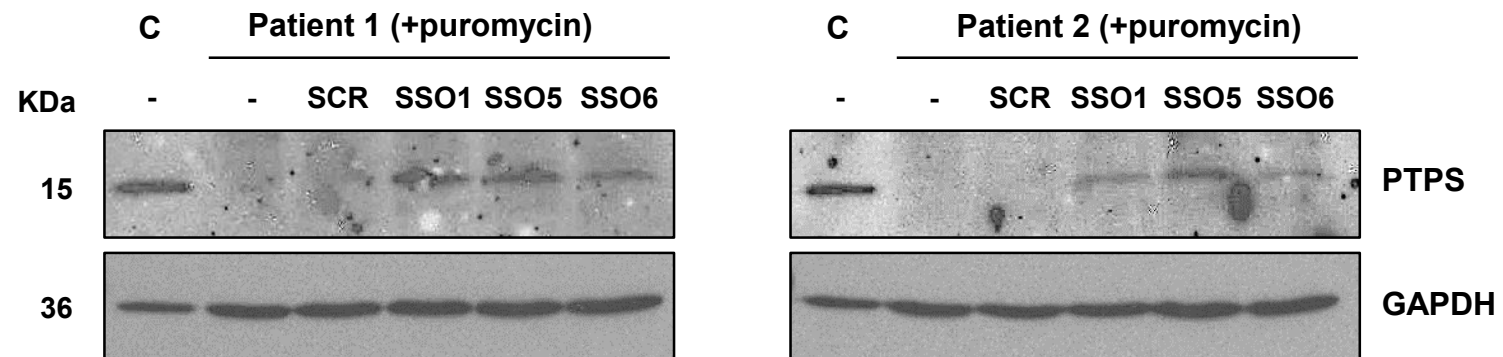
D)

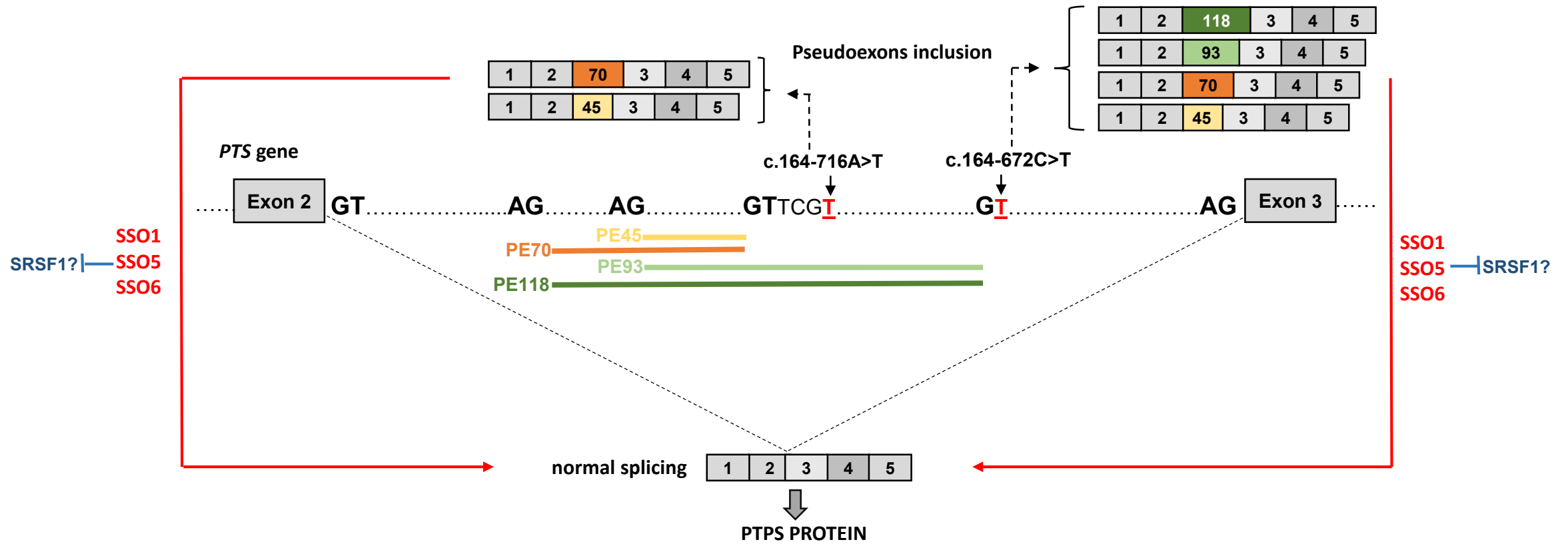


**A)**



**B)**





## SUPPLEMENTARY METHODS

**Processing of eCLIP datasets.** Raw SRSF1 eCLIP reads [1] were downloaded from the ENCODE website and trimmed twice using cutadapt v1.15 [2]. Details of the run parameters are described in Supplementary methods. for the first round of trimming:

```
cutadapt -f fastq --match-read-wildcards --times 1 -e 0.1 -O 1 --quality-cutoff 6 -m 18 -a
NNNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -g
CTTCCGATCTACAAGTT -g CTTCCGATCTTGGTCCT -A AACTTGTAGATCGGA -A
AGGACCAAGATCGGA -A ACTTGTAGATCGGAA -A GGACCAAGATCGGAA -A
CTTGTAGATCGGAAG -A GACCAAGATCGGAAG -A TTGTAGATCGGAAGA -A
ACCAAGATCGGAAGA -A TGTAGATCGGAAGAG -A CCAAGATCGGAAGAG -A
GTAGATCGGAAGAGC -A CAAGATCGGAAGAGC -A TAGATCGGAAGAGCG -A
AAGATCGGAAGAGCG -A AGATCGGAAGAGCGT -A GATCGGAAGAGCGTC -A
ATCGGAAGAGCGTCG -A TCGGAAGAGCGTCGT -A CGGAAGAGCGTCGTG -A
GGAAGAGCGTCGTGT -o reads.R1.trimmed.fastq.gz -p reads.R2.trimmed.fastq.gz
reads.R1.fastq.gz reads.R2.fastq.gz
```

The second round of trimming was run with the following parameters:

```
cutadapt -f fastq --match-read-wildcards --times 1 -e 0.1 -O 5 --quality-cutoff 6 -m 18 -A
AACTTGTAGATCGGA -A AGGACCAAGATCGGA -A ACTTGTAGATCGGAA -A
GGACCAAGATCGGAA -A CTTGTAGATCGGAAG -A GACCAAGATCGGAAG -A
TTGTAGATCGGAAGA -A ACCAAGATCGGAAGA -A TGTAGATCGGAAGAG -A
CCAAGATCGGAAGAG -A GTAGATCGGAAGAGC -A CAAGATCGGAAGAGC -A
TAGATCGGAAGAGCG -A AAGATCGGAAGAGCG -A AGATCGGAAGAGCGT -A
GATCGGAAGAGCGTC -A ATCGGAAGAGCGTCG -A TCGGAAGAGCGTCGT -A
```



```
CGGAAGAGCGTCGTG -A GGAAGAGCGTCGTG -o reads.R1.trimmed2.fastq.gz -p  
reads.R2.trimmed2.fastq.gz reads.R1.trimmed.fastq.gz reads.R2.trimmed.fastq.gz
```

For each read file, we then prepared the barcode for PCR-duplication detection with UMI tools by appending it to the read ID:

```
awk -v l=10 'BEGIN{OFS=FS=" "} substr($1, 1, 1) == "@" {print "@" substr($1, (l+3), 500)  
"_" substr($1, 2, l) " " $2 }; substr($1, 1, 1) != "@" {print}; ' <(zcat  
reads.R1.trimmed2.fastq.gz) | gzip > reads.R1.trimmed2.bc.fastq.gz
```

Subsequently, reads were mapped with STAR v2.7.8a [3] to the human genome using GENCODE v39 hg38 annotations allowing for up to 100 mapping locations with the following run parameters:

```
star --outSAMtype BAM SortedByCoordinate --runThreadN 40 --genomeDir star_index -  
-readFilesIn reads.R1.trimmed2.bc.fastq.gz reads.R2.trimmed2.bc.fastq.gz --  
readFilesCommand zcat --outFilterType BySJout --outFilterMultimapNmax 100 --  
alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --  
outFilterMismatchNoverLmax 0.04 --scoreDelOpen -1 --alignIntronMin 20 --  
alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignEndsType EndToEnd --  
outMultimapperOrder Random --winAnchorMultimapNmax 100
```

Using UMI-tools [4], PCR duplicates were removed with the following command:

```
umi_tools dedup -l Aligned.sortedByCoord.out.bam --paired -S  
Aligned.sortedByCoord.rm-pcr-dup.bam
```

Subsequently, only read 2 was retained, to mark sites of protein binding, using samtools [5]:

```
samtools view -hb -f 130 Aligned.sortedByCoord.rm-pcr-dup.bam -o  
Aligned.sortedByCoord.rm-pcr-dup.R2.bam
```

Finally, we merged the two replicates for each cell type using samtools merge.

**Processing of PAR-CLIP dataset:** Raw PAR-CLIP reads were downloaded from the GEO repository using the accession number GSE71095 [6] and trimmed using cutadapt with the following run parameters:

```
cutadapt -f fastq -u 5 -q 15 -m 20 -a TGGAATTCTCGGGTGCCAAGG -o  
reads.trimmed.fastq.gz reads.fastq.gz
```

We then mapped the reads to the human genome using GENCODE v39 hg38 annotations allowing for up to 100 mapping locations with the following run parameters:

```
star --outSAMtype BAM SortedByCoordinate --runThreadN 40 --genomeDir star_index -  
-readFilesIn reads.trimmed.fastq.gz --readFilesCommand zcat --outFilterType BySJout  
--outFilterMultimapNmax 100 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --  
outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --scoreDelOpen -1 --  
alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --  
alignEndsType Local --outMultimapperOrder Random --winAnchorMultimapNmax 100
```

Subsequently, we removed PCR duplicates using samtools rmdup.

**Processing of HITS-CLIP dataset:** Raw HITS-CLIP reads were downloaded from the GEO repository using the accession numbers GSE131745 [7] and trimmed using cutadapt with the following run parameters:

```
cutadapt -f fastq -u 5 -q 15 -m 20 -a GTGTCAGTCACTTCCAGCGG -o  
reads.trimmed.fastq.gz reads.fastq.gz
```

We then mapped the reads to the human genome using GENCODE v39 hg38 annotations allowing for up to 100 mapping locations with the following run parameters:

```
star --outSAMtype BAM SortedByCoordinate --runThreadN 40 --genomeDir star_index -  
-readFilesIn reads.trimmed.fastq.gz --readFilesCommand zcat --outFilterType BySJout  
--outFilterMultimapNmax 100 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --  
outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --scoreDelOpen -1 --
```

```
alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --  
alignEndsType Local --outMultimapperOrder Random --winAnchorMultimapNmax 100
```

Subsequently, we removed PCR duplicates using samtools rmdup and merged the two replicate experiments using samtools merge.

## References

1. Van Nostrand, E.L. et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13 (6), 508-14.
2. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17 (1), 10-12.
3. Dobin, A. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15-21.
4. Smith, T. et al. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 27 (3), 491-499.
5. Li, H. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16), 2078-9.
6. Xiao, W. et al. (2016) Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* 61 (4), 507-519.
7. Fish, L. et al. (2019) Nuclear TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay. *Mol Cell* 75 (5), 967-981 e9.

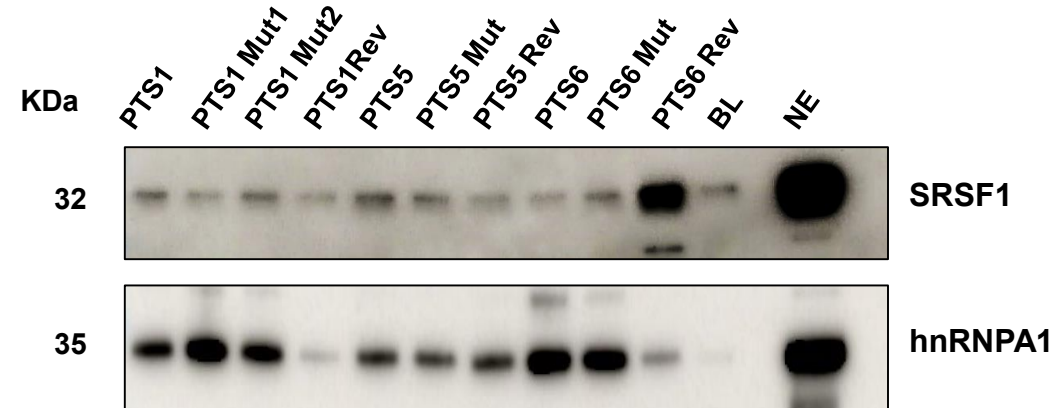
**Table S1.** SSO length and sequence.

<b>SSOs</b>	<b>Length (bp)</b>	<b>Sequence 5'-3'</b>
SSO Scrambled	20	CTCAATATGCTACTGCCATG
SSO1	25	AGGCTGGAGAATCGCTCGAACCTGG
SSO2	20	GTGGCGGGCACCTGTAATCT
SSO3	25	TGGAGGTTGTAGTGACGGGAGATCA
SSO4	25	TCGAACCTGGGAGGTGGAGGTTGTA
SSO5	25	TACTCGGGAGGCTGAGGCTGGAGAA
SSO6	25	CACCTGTAATCTCAGCTACTCGGGA

**A)**

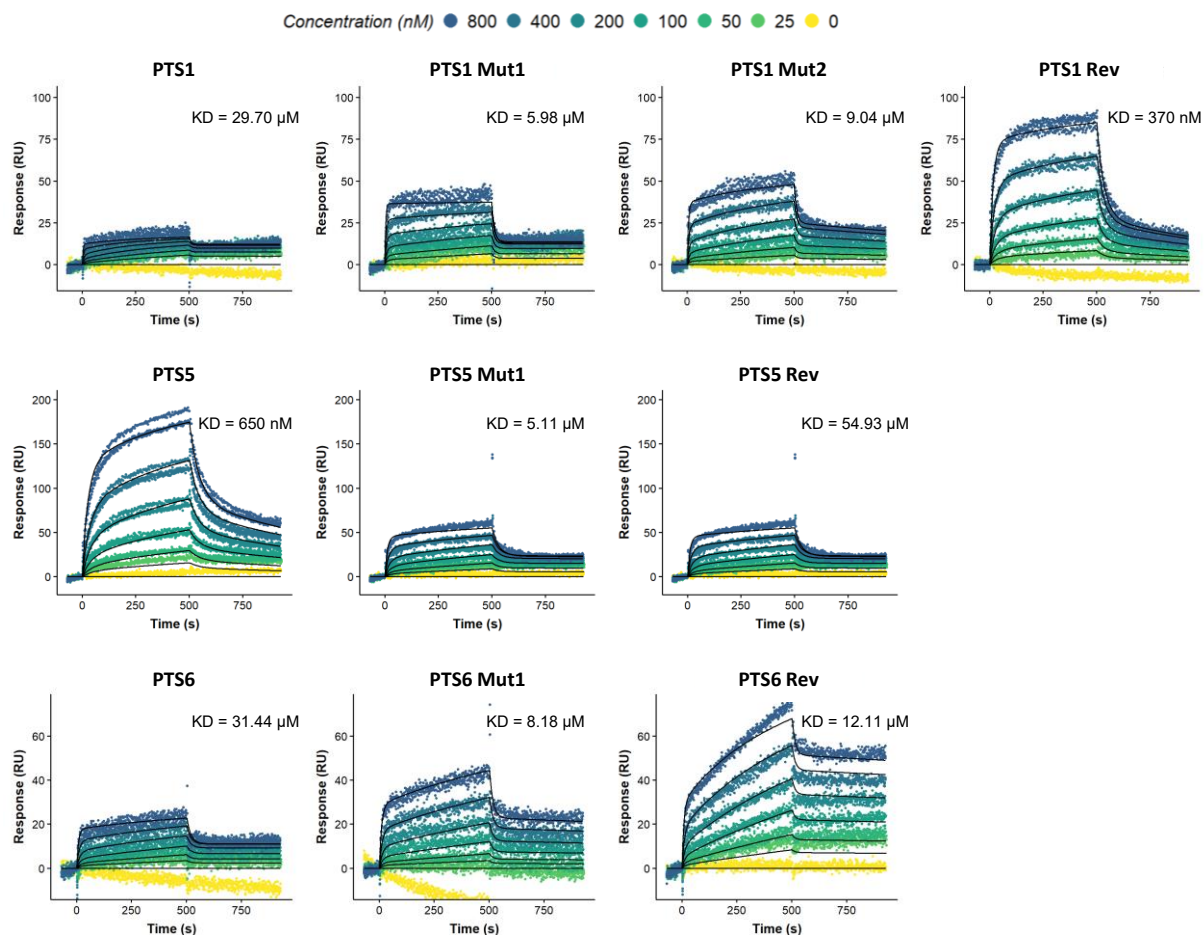
PTS1	CCAGGUUCGAGCGAUUCUCCAGCCU-biotin
PTS1 Mut1	CCAGGUUCG <u>U</u> GCGAUUCUCCAGCCU-biotin
PTS1 Mut2	CCAGGUUC <u>U</u> AGCUAUUCUCCAGCCU-biotin
PTS1 Rev	AGGCUGGAGAAUCGCUCGAACCUGG-biotin
PTS5	UUCUCCAGCCUCAGCCUCCCGAGUA-biotin
PTS5 Mut	UUCUCCAGCCUCAGCCUC <u>UU</u> GAGUA-biotin
PTS5 Rev	UACUCGGGAGGCUGAGGCUGGAGAA-biotin
PTS6	UCCCGAGUAGCUGAGAUUACAGGUG-biotin
PTS6 Mut	UCUUGUGUAGCU <u>UAUAUUUC</u> UGGUG-biotin
PTS6 Rev	CACCUGUAAUCUCAGCUACUCGGGA-biotin

**B)**

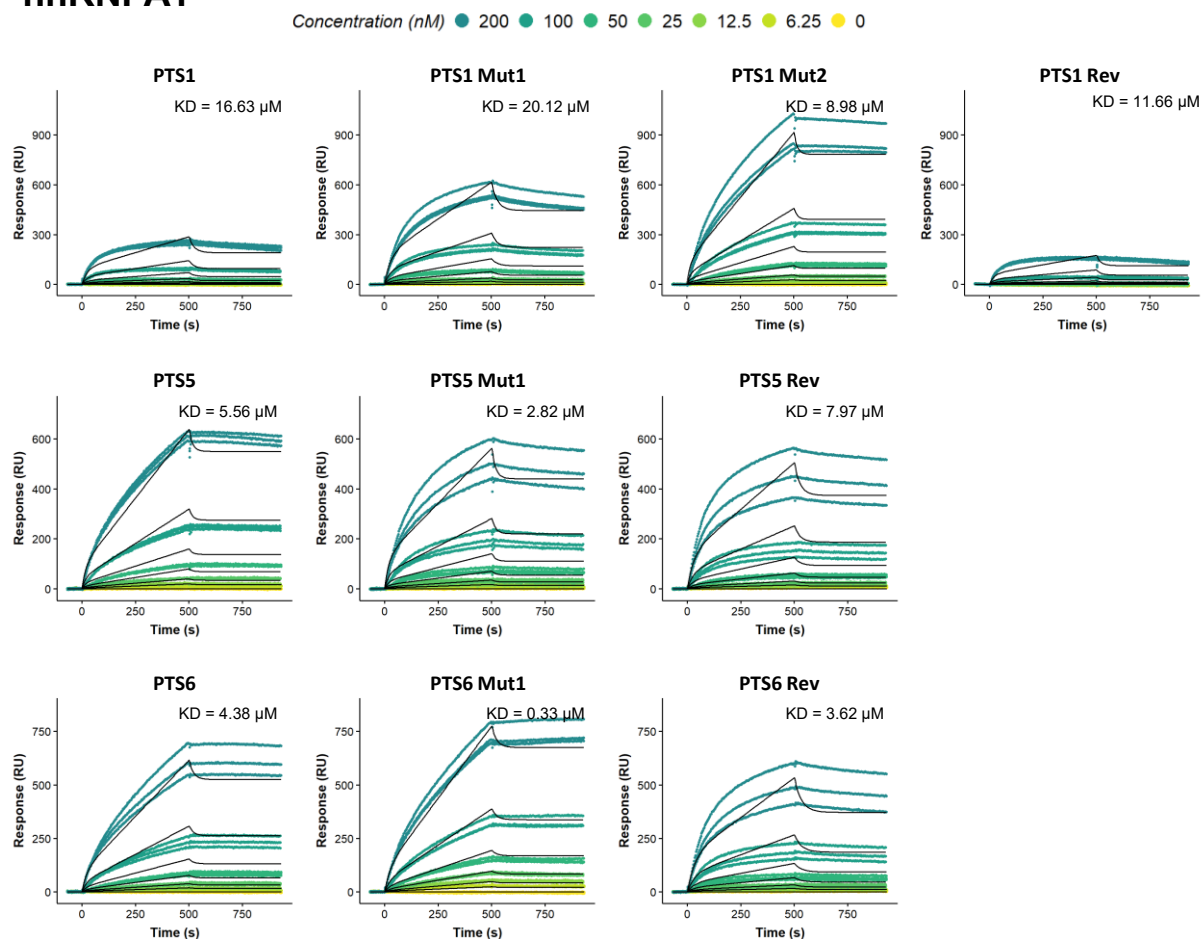


**Figure S1.** RNA affinity purification. A) Sequences of biotin-conjugated RNA oligonucleotides covering the three SSO1, 5 and 6 binding sites of *PTS* intron 2. Mutations disrupting the SRSF1 binding motifs are shown in red. B) Western blot of SRSF1 and hnRNPA1 proteins purified by RNA-affinity chromatography. Results are representative for two experiments. BL and NE indicate control lanes without RNA oligonucleotides or with nuclear extract alone, respectively.

A) SRSF1



B) hnRNPA1



**Figure S2.** Surface plasmon resonance imaging (SPRi) of SRSF1 and hnRNP A1 recombinant proteins. SRSF1 and hnRNP A1 were injected in increasing 2-fold concentrations from 25 to 800 nM or 6.25 to 200 nM, respectively. Measurements were fitted to a bimodal 1:1. For hnRNPA1 a biphasic 1:2 model was used in ClampXP.

