Specificity and exon target space of splicing modifying compounds

Felina Lenkeit[1], Judith Knehr[1], Marc Altorfer[1], Andrea Byrnes[2], Wenjing Li[3], Jack Hsiao[3], Christian Kolter[1], Philip R. Skaanderup[1], Steve Mullin[2], Elizaveta Solovyeva[4], Andrew T. Krueger[2], Johannes Ottl[1], Caroline Gubser Keller[1], Ulrike Naumann[1], Philipp Ottis[1 # *], and Alejandro Reyes[1 # *]

[1]Discovery Sciences, Novartis Biomedical Research, Basel, Switzerland; [2]Discovery Sciences, Novartis Biomedical Research, Cambridge, MA, USA; [3]Disease Area Neuroscience, Novartis Biomedical Research, Cambridge, MA, USA; [4]Research Informatics, Novartis Biomedical Research, Basel, Switzerland; # Corresponding authors; [*] Contributed equally

Modulation of splicing has become an established therapeutic strategy, with proven clinical applications and continued potential to target specific exons to influence gene expression. Recent advancements led to the identification of small molecule splicing modifiers such as Risdiplam and Branaplam. These compounds induce the inclusion of exons that are typically skipped due to their weak 5' splice site. While Risdiplam has a preference to induce exons with a $\mathbf{N}_{-3}\mathbf{G}_{-2}\mathbf{A}_{-1}$ sequence at the 3' exon end, Branaplam has a proclivity towards introducing $\mathbf{A}_{-3}\mathbf{G}_{-2}\mathbf{A}_{-1}$ -ending exons. However, the variables that determine the selectivity and specificity of splicing modulators are still not completely understood, as evidenced by the hundreds of unaffected $\mathbf{N}_{-3}\mathbf{G}_{-2}\mathbf{A}_{-1}$ -ending exons present in the human genome. In this study, we delve into the molecular mechanisms governing the specificity of splicing-modifying compounds, focusing on their interactions with RNA structures at splice sites. Using biochemical assays, whole transcriptome analyses, and genetic perturbation approaches, our findings reveal contributions of primary sequence dependencies that help determine the required secondary structural conformation, thus governing responsiveness to splicing-modulator induction. Based on these learnings, we were able to reprogram the specificity of splicing modulators by genetic manipulation of the U1 snRNA component of the spliceosome. Our findings further the understanding of splicing modulators and might help to identify novel targets and to design new splicing-modifying compounds.

Introduction

Splicing is a fundamental biological process that generates mature mRNA by removing introns from pre-mRNA. Splicing is carried out by the spliceosome, a macromolecular complex which comprises various protein components and small nuclear RNAs (snRNAs), together termed small ribonucleoproteins (snRNPs). One of these snRNAs, as part of the U1 snRNP, guides the first step of the splicing reaction by base-pairing with the sequence spanning from the terminal three nucleotides at the 3' end of an exon (positions -3, -2, -1) to the eight nucleotides of the downstream intron (positions +1 to +8) (Wilkinson et al., 2020; Papasaikas and Valcarcel, 2016) (Figure 1A). This sequence stretch in the pre-mRNAs is known as 5' splice site (5'ss) and its consensus sequence is $5'$-$\mathrm{C}_{-3}\mathrm{A}_{-2}\mathrm{G}_{-1}\mathrm{G}_{1}\mathrm{U}_{2}\mathrm{A}_{3}\mathrm{A}_{4}\mathrm{G}_{5}\mathrm{U}_{6}\mathrm{A}_{7}\mathrm{U}_{8}$-$3'$. Similarly, the branchpoint sequence, the polypyrimidine tract and the 3' splice site are sequence elements that guide the splicing machinery to determine the 5' start of an exon. In most human genes, alternative selection of 5' and 3' splice sites enable the production of multiple isoforms from a single gene, contributing to protein diversity (Black, 2003; Lee and Rio, 2015).

Modification of splice site selection has been proven to be a promising approach in treating various diseases. This strategy addresses not only diseases caused by mutations that lead to aberrant splicing, but also diseases with other causes, where modulating the splicing process can influence the expression levels of particular isoforms (Tang et al., 2021). By directly targeting mRNAs, this approach enables the modulation of proteins, that may otherwise be challenging to drug or even considered undruggable. While antisense oligonucleotides (ASOs) to date constitute the majority of clinically validated splicing modifiers (Mercuri et al., 2018; Havens and Hastings, 2016; Mendell et al., 2013), RNA-targeting small molecules offer numerous advantages. Notably, these small molecules allow non-invasive oral delivery, ensure systemic distribution, and have the capacity to cross the blood–brain barrier, thus enhancing their potential as therapeutic agents (Warner et al., 2018; Singh et al., 2020).

To date, there are several small molecule splicing modulator compounds with available human clinical data, including Risdiplam, NVS-SM1 (Branaplam), PTC518, SKY-0515, RGT-61159 and REM-422. Risdiplam is being used for the treatment of spinal muscular atrophy (SMA) (Dhillon, 2020). SMA is caused by a mutation in the survival motor neuron 1 (*SMN1*) gene leading to a deficiency of functional SMN protein (Lefebvre et al., 1995). The gene *SMN2* is nearly identical to *SMN1*, however harbors a critical C to T substitution, which prevents efficient splicing of exon 7, featuring a non-canonical 5'ss

($5'$-$G_{-3}G_{-2}A_{-1}G_1U_2A_3A_4G_5U_6C_7U_8$-$3'$) (Figure 1B). As a result, most *SMN2* transcripts lack exon 7 and produce non-functional isoforms. By introducing exon 7 into transcripts of *SMN2*, Risdiplam restores the expression of functional SMN protein that compensates for the *SMN1* pathologic mutations (Ruggiu et al., 2012; Le et al., 2005; Ratni et al., 2018). Branaplam, here referred to as its development code NVS-SM1, was also found to promote exon 7 inclusion in *SMN2* transcripts. Additionally, this molecule demonstrated the capability to induce the inclusion of a frame-shifting pseudoexon between exon 49 and 50 in the *HTT* gene that introduces a pre-mature stop codon which triggers the mRNA to be degraded (Bhattacharyya et al., 2021; Keller et al., 2022; Chen et al., 2020). This *HTT* pseudoexon is not recognized by the spliceosome in the absence of NVS-SM1, as it contains a non-canonical 5'ss ($5'$-$A_{-3}G_{-2}A_{-1}G_1U_2A_3A_4G_5G_6G_7G_8$-$3'$).

Genome wide transcriptome analyses and mutational screening assays have shown that Risdiplam has a preference to induce exons with a $N_{-3}G_{-2}A_{-1}$ (N = any base) sequence at the 3' exon end, while NVS-SM1 has a proclivity towards introducing $A_{-3}G_{-2}A_{-1}$-ending exons (Krach et al., 2022; Ishigami et al., 2024) (Figure 1C). Without the presence of these compounds, these exons are typically being skipped due to their weak 5'ss, making recognition unlikely (Monani et al., 1999; Wong et al., 2018; Freund et al., 2005). A NMR structure of a chemical analog of Risdiplam, SMN-C5, showed that the compound binds the RNA helix formed by the 5'ss and the 5'end of the U1 snRNA, at the exon-intron junction in its major groove. Here, it stabilizes the bulged-out A_{-1} at the exon end, acting like a molecular glue that strengthens the interaction of both RNAs, compensating for the weak 5'ss to enable splicing initiation (Campagne et al., 2019; Ratni et al., 2016). Further studies using NMR spectroscopy showed that this bulge repair mechanism generalised to NVS-SM1 (Malard et al., 2024). Despite these mechanistic advances, the variables that determine the selectivity and specificity of splicing modulators are still not completely understood. For instance, a recent study predicted >58,000 potential exons with a 5'ss sequence of $5'$-$A_{-1}G_{-2}A_{-3}G_1U_2A_3A_4G_5$-$3'$ in human introns, however observed inclusion upon NVS-SM1 treatment for only 609 exons (Bhattacharyya et al., 2021). Similarly, transcriptome data have shown that $N_{-3}G_{-2}A_{-1}$ ending exons are induced at different levels upon treatment with NVS-SM1 or Risdiplam.

Here, we analyzed the contribution of sequence elements to the specificity of splicing modulators using a combination of biochemical assays, whole transcriptome analyses and genetic perturbation approaches. These findings revealed that the complementarity between the U1 snRNP and the target exons contributes quantitatively to the stability of the A_{-1}-bulge structure for compound-mediated splicing. Moreover, a crucial influence of the -4 position at the end of targeted exons in determining compound activity was observed. Finally, we quantitatively modulated compound specificity at $N_{-3}G_{-2}A_{-1}$ exons by genetically modifying the U1 snRNP in cells. Taken together, we refine parameters to explain the observed splicing modulator target space and to better understand the specificity of splicing modulators. Our learnings further the understanding of splicing modulator target space as well as help in rationally designing new splicing-modifying compounds.

Results

NVS-SM1 is known to preferentially induce the splicing of exons with 3' ends consisting of an $A_{-3}G_{-2}A_{-1}$ trinucleotide (Figure 1C), which we refer to as AGA exons for simplicity (Keller et al., 2022). First, we aimed to define the potential space of AGA target exons throughout the human genome. We used the large collection of human RNA-seq data from the Genotype-Tissue Expression Project available through the snaptron database (Consortium, 2013). We identified 8,718 potential 5'ss matching a $5'$-$A_{-3}G_{-2}A_{-1}G_1U_2$-$3'$ motif. Most of the corresponding exons were supported by unannotated exon-exon junctions and, when detected, were included at very low level, confirming that these non-canonical exons are usually not recognized by the splicing machinery under normal physiological conditions.

Variation of exon target space of NVS-SM1 across cell types

This large number of predicted target exons, however, is in strong contrast to the reported transcriptomic profiles in literature, where a few hundred of AGA exons are reported to be induced in response to splicing modifiers such as NVS-SM1 (Keller et al., 2022; Krach et al., 2022). Consistent with these reports, we detected the induction of 125 and 304 AGA-ending exons when we treated SH-SY5Y cells for 24 hours with 25nM and 100nM of NVS-SM1, respectively, in whole transcriptome profiles (10% false discovery rate [FDR]; Figure 1D/E). Despite focusing on non-canonical 5'ss, we detected 334 AGA exons with high levels of baseline inclusion in the absence of a compound (>0.6 percent spliced-in [PSI], Supplementary Figure S2A). Notably, the induction of these high baseline AGA exons was not further boosted upon compound treatment.

To further refine the AGA exon target space, we sequenced the transcriptomes of SH-SY5Y, as well as iNGN2 cells, in a single experiment, following treatment with 25nM of NVS-SM1 for 24 hours. Compared to DMSO controls, we found 97 induced AGA exons in SH-SY5Y cells and 171 in iNGN2 cells, with 75 of these exons common to both (Figure 1F). When focusing on the list of genes containing AGA-ending exons induced only in SH-SY5Y cells, these genes, on average, demonstrated a 90-times higher expression in SH-SY5Y, compared to iNGN2 cells, and vice versa (Supplementary Figure S2B). These observations suggest that a large proportion of cell type specific differences in induced

AGA exons originate from differential baseline expression of genes containing potentially induceable AGA exons. Indeed, when we filtered the 8,718 potential AGA exons for those harbored by highly expressed genes in SH-SY5Y cells and, at the same time, demonstrating low baseline inclusion in the absence compound, the number of potentially inducible AGA exons in SY-SY5Y cells dropped to 628.

Although baseline exon inclusion and gene expression accounted for a large fraction of the uninduced AGA exon target space, still around half of the potentially inducible AGA exons were not induced upon NVS-SM1 treatment. We reasoned that if the induction of AGA exons introduced stop codons into transcripts, a rapid degradation of mRNA by nonsense mediated decay could prevent their detection by RNA-seq. However, only 8.5% of the genes carrying an uninduced AGA exon were downregulated (Figure 2A), indicating that this effect would explain only a minority of the unobserved AGA exons.

Impact of U1 snRNA complementarity on compound-mediated exon inclusion

Even when exons do check all the sequence-specificity, expression level, and baseline inclusion criteria, some still fail to show inclusion after compound treatment. One endogenous factor governing a 5' splice-site's efficiency is the overall complementarity of this splice-site to the U1 snRNA. High complementarity enhances the recognition of the splice site, whereas low complementarity favors exon skipping (Freund et al., 2005). We therefore rationalized that complementarity to U1 snRNA also plays a key role in compound-mediated splicing. Even if a modulator helps strengthen weak splice sites, a certain amount of baseline complementarity may still be required. To test this hypothesis, we calculated the maximum number of possible hydrogen bonds (H-bonds) formed between the 5'ss of AGA exons and the U1 snRNA. Accounting for the possibility of an A_{-1} bulge formation in the pre-mRNA, we included the -4 position of the exon in the score calculation (Figure 2B). Exemplary scores for some of the exons induced by NVS-SM1 can be found in Supplementary Table S1. This score, hereafter referred to as the complementarity score, was used as a proxy for the complementarity between the 5'ss and U1 snRNA. For simplicity, we assumed that each potential H-bond contributes equally to the stability of the U1–5'ss interaction, acknowledging that individual H-bonds can differ in strength. This approach provides a tractable estimate of sequence complementarity without accounting for such variations. Analyses highlighted a significant difference between the distributions of complementarity scores for AGA exons amenable to induction by NVS-SM1 compared to non-amenable AGA exons (p-value < 0.0001, Figure 2C). We observed the same difference in analogous analyses of induced and uninduced GA exons upon Risdiplam treatment (Supplementary Figure S3A). These findings suggested that the complementarity between the 5'ss of target exons and U1 snRNA contributes quantitatively to exon induction upon compound treatment and is a crucial determinant of exon responsiveness to the compounds.

To verify this, we designed an experiment to modify U1 snRNA – pre-mRNA complementarity while leaving the small-molecule binding site unchanged. We ectopically expressed a U1 snRNA, harboring 6A>G and 7U>C mutations to improve pairing with the intronic region of pre-mRNAs (Figure 3A). We refer to this mutant U1 snRNA as U1-G_6C_7, since the changes pair with positions +6 and +7 of the exon's U1 binding site. Whole transcriptome profiles were sequenced in biological triplicates and analyzed for exons induced by the combination of U1-G_6C_7 transfection and NVS-SM1 treatment at 100nM. Compared to the DMSO-treated cells transfected with U1-G_6C_7, 156 AGA exons were found significantly induced upon combining U1-G_6C_7 transfection and NVS-SM1 treatment (Supplementary Figure S1D). Of these, 29 were not detected as significantly induced upon the control combination of U1-wt transfection and NVS-SM1 treatment (FDR = 10%, dPSI >0.05). Analyses of the respective 5'ss in this set of induced exons highlighted a significant shift towards increased complementarity scores with the mutated U1-G_6C_7 compared to canonical U1 snRNA (Figure 3E).

One cryptic exon, located in the *TAMM41* gene and exclusively induced by the combination of U1-G_6C_7 expression and NVS-SM1 treatment, demonstrated particularly high inclusion rates, substantiated by clear exon-exon junction reads (dPSI=0.8, p-value < 0.001, Figure 3B). Another cryptic exon in gene *C2orf49* was also found significantly induced upon combination treatment (dPSI = 0.16, p-value < 0.001, Figure 3C). In *UBP1*, a cryptic exon demonstrated only a slight, non-significant increase with U1-wt plus NVS-SM1 (dPSI = 0.014), yet a much stronger induction with U1-G_6C_7 plus NVS-SM1 (dPSI = 0.52, p-value < 0.001, Figure 3D). Notably, all the 5'ss of these induced cryptic exons had a stronger complementarity with U1-G_6C_7 compared to canonical U1 snRNA (Figure 3E). Additional examples can be found in Supplementary Figure S4A-D.

To test whether this observation applies to other splicing modifiers as well, we repeated the experiment using 1µM of Risdiplam. We compared U1-G_6C_7–transfected cells treated with 1µM of Risdiplam to transfected cells treated with DMSO as a control and found 106 significantly induced GA exons (Supplementary Figure S1E). Of these, 23 exons were not significantly induced upon the combination of U1-wt transfection and 1µM Risdiplam. As previously observed for NVS-SM1, this set of exons demonstrated a shift to higher complementarity scores with U1-G_6C_7 compared to canonical U1 snRNA (Supplementary Figure S3B-F). Overall, these data show that baseline complementarity between U1 snRNA and the target exons contributes to compound-mediated exon induction.

Role of the -4 exon end position in compound-mediated splicing

In an attempt to unify the A_{-1} bulge formation model (Palacino et al., 2015; Campagne et al., 2019) and our observations on complementarity requirements, we tested how the -4 nucleotide at the exon's 3' end affects NVS-SM1 and Risdiplam activity. As illustrated in Figure 2B, a cytosine at position -4 could pair with a guanine in position 11 of the U1 snRNA, assuming an A-1 bulge forms in the pre-mRNA. Thus, restoring the canonical 5'ss initiation sequence CAG. To test this model in a cellular context, we analyzed the 125 AGA exons induced by NVS-SM1 in SH-SY5Y cells and grouped them according to their -4 position, resulting in 50 exons with a 5'-CAGA-3' ending and 75 exons with either UAGA, GAGA or AAGA ends. Strikingly, when focusing on the complementarity of the intronic base-pairing (positions +1 through +8) to a canonical U1 snRNA, we found that low-complementarity sites favor a cytosine at position -4 (p-value = 0.02, Figure 3F). This suggests that base-pairing at position -4 can help compensate for weak U1 binding in the intron.

To experimentally test the influence of the -4 position on the formation of an RNA duplex with a A_{-1} bulge, we employed a Fluorescence Polarization (FP) assay based on a TAMRA-labeled 16 nucleotide RNA representing the U1 snRNA sequence. With increasing concentrations of *HTT* pseudoexon 5'ss, we observed an increase in fluorescence polarization, indicating binding of the labelled U1 sequence (EC_{50} = 252nM, Figure 4A/B). Repeating the experiment with a target RNA harboring a C>G mutation at the -4 position (-4C>G, Figure 4B), we observed a significant decrease in FP signal compared to the wt RNA (EC_{50} = 627nM, Figure 4A). This finding corroborates the influence of the nucleotide at position -4 on the overall base pairing between U1 snRNA and 5'ss.

Primary sequence features of GA exons induced by NVS-SM1 and Risdiplam

Considering the influence of the -4 position on the pairing between U1 snRNA and non-canonical GA exons, we analyzed the primary sequence features of exons induced by NVS-SM1. In line with the established GA-specificities of these compounds, we observed an enrichment for GA exon ends (positions -2 and -1) in the 5'ss motifs of NVS-SM1 and Risdiplam responsive exons in SH-SY5Y cells. In NVS-SM1 treated cells, the $A_{-3}G_{-2}A_{-1}$ motif was enriched (Figure 1C). Analyzing the nucleotide distribution at position -2 of all induced A-ending exons, we found a near-exclusive occurrence of guanine (91%, i.e. 146 out of 160 upon NVS-SM1 treatment and 94%, i.e. 151 out of 161 upon Risdiplam treatment) (Figure 4C). Consistent with the observed motif, a robust enrichment for adenine in response to NVS-SM1 could be detected at the -3 position of induced GA-ending exons (86%, i.e. 125 out of 146). Interestingly, NVS-SM1 also induced 14 GGA and 7 UGA exons, but no CGA exons (Figure 4D). Similarly, Risdiplam disfavors CGA-ending exons; although four CGA-ending exons are induced in this experiment, they are clearly underrepresented. While AGA exons demonstrated an even distribution of nucleotides at position -4 (Figure 4E), GGA and UGA exons near exclusively featured an adenine in this position (Figure 4F). This pattern was also seen with Risdiplam-induced exons. The strong preference for inducing AGA-, AGGA- and AUGA-ending exons, suggests NVS-SM1 and Risdiplam to exhibit a preference for acting on exon ends with two adenines, the first one at position -1 and the second either at –3 or –4. As adenines in position -4 are unlikely to contribute to a stabilization of the U1-5'ss RNA duplex (Figure 4G), their presence may favor compound binding.

To explore the lack of inducible CGA-ending exons, we transfected HeLa cells with a modified U1 snRNA harboring a -2U>C and a -1C>U mutation, thus stabilizing all GA ending exons (U1-$C_{-2}U_{-1}$, Figure 4G). We sequenced the transcriptomes of the transfected cells and recorded exons induced specifically by U1-$C_{-2}U_{-1}$ as compared to control cells transfected with the canonical U1 snRNA (Supplementary Figure S1F). Using this genetic stabilization of the U1 snRNP, we observed inclusion of 96 CGA exons (9% of all included GA-ending exons, Figure 4H). Although the frequency of inducible CGA exons upon U1-$C_{-2}U_{-1}$ transfection remains lower compared to other GA exons under these conditions, this experiment demonstrates that CGA-ending exons are, in principle, inducible. In contrast, when transfecting canonical U1 snRNA and treating cells with NVS-SM1 or Risdiplam, inclusion of CGA-ending exons is almost absent, with only one or two exons detected, respectively (less than 1% of induced GA-ending exons) (Figure 4H). These data demonstrate that CGA exons are generally amenable to splicing modulation by the stabilization of the U1 snRNP at the CGA 5'ss. This observation is in stark contrasts with the absence of induced CGA exons upon compound treatment (p-value < 0.001). One possible conclusion is that the sequence $C_{-3}G_{-2}A_{-1}$ prevents a bulge formation at A_{-1} due to an energetically favorable G-C pairing at position -3 of the 5'ss (Figure 4G). This, in turn, suggests the bulge stabilization at position -1 to not just be the result of compound binding, but actually a prerequisite for it.

NVS-SM1 interaction with A-1 bulge in RNA duplexes

To understand the dynamics of A_{-1} bulge structure stabilization, we designed a 13 nucleotide RNA probe containing the 5'ss sequence of the *HTT* pseudoexon (positions -4 to 9), in which we replaced the adenine at position -1 with the fluorescent analogue 2-aminopurine (2-AP). The fluorescence of 2-AP depends on its microenvironment and, while un-impacted by events such as hydrogen-bond formation, is being quenched by base-

stacking and clashes with neighboring nucleobases (Rachofsky et al., 2001). Therefore, 2-AP can act as a sensitive reporter to monitor changes to the secondary or tertiary structure of RNA, when incorporated into a nucleotide strand (Rau and Hall, 2015). Moreover, those characteristics can be applied to detect and monitor ligand-engagement to RNA (Bradrick and Marino, 2004; Chen et al., 2020). Thus, the 2-AP probe enabled monitoring both of A_{-1} bulge secondary structure formations.

We incubated the 2-AP probe with wild-type U1 snRNA or modified U1 snRNA oligonucleotides, respectively. The modified U1 oligo, harboring the mutations -3G>U, -2U>C and -1C>U, reverse complements the 2-AP probe and thus forms an RNA duplex without any mis-match induced secondary structures. Hence, we termed this modified U1 snRNA oligo "closed U1" (Figure 5A).

As expected, adding the closed U1 construct caused strong quenching of 2-AP fluorescence, indicating base stacking in the RNA duplex (Figure 5B). Conversely, when adding the canonical U1 RNA oligo, the 2-AP fluorescence was significantly higher compared to the closed construct, indicating that the 2-AP base was not stacked in the RNA duplex, thus confirming the presence of a bulge structure (Figure 5A). Next, we monitored the 2-AP fluorescence intensity (FI) changes in the presence of canonical U1 RNA oligo, with increasing concentrations of either NVS-SM1 or a structurally related, splicing-inactive compound (ctrl; Supplementary Figure S1G). While the inactive compound did not cause any fluorescence decrease, NVS-SM1 caused a concentration-dependent quenching of the 2-AP FI, confirming its interaction with the A_{-1} bulge in the RNA duplex (Figure 5C).

Primary and secondary RNA structures define the splicing modulator target exon space

To better understand how RNA structure and sequence affect compound binding, we adapted the 2-AP assay to test whether NVS-SM1 could bind the same secondary structures in the context of different primary sequences. Therefore, we modified the HTT 5'ss 2-AP probe with a -2G>C mutation to mimic an ACA-ending exon. To match this oligo probe, we designed a 12 base oligonucleotide with the U1 snRNA sequence harboring a -2C>G mutation, called "inverted U1" (Figure 5D). The combination of both the modified HTT 5'ss probe and the inverted U1 oligo allows for and favors the formation of a bulged-out 2-AP reporter nucleotide (Figure 5C/F). The baseline fluorescence intensity of the HTT-C_{-2} RNA complex was found similar to the FI observed for the HTT-wt complex (Figure 5F), suggesting a resembling formation of a bulged-out adenine reporter nucleotide at position -1 of the RNA duplex. Subjecting this reporter complex to increasing concentrations of NVS-SM1, we observed a dose-dependent quenching of the 2-AP probe, this time embedded into a different sequence context (Figure 5F). These data demonstrate that binding of NVS-SM1 to adenine bulge structures occurs independent of the primary RNA sequence. This observation, in line with our genetic perturbation experiments, highlights the effect of the highly conserved U1 snRNA sequence as a major determinant, governing the specificity of small molecule splicing modulators, such as NVS-SM1.

Conclusion

This study provides comprehensive insights into the mechanisms underlying the specificity of small molecule splicing modifiers, particularly focusing on NVS-SM1 (Branaplam) and its interaction with the splicing machinery. Although these compounds target specific exon-end sequences, not all matching exons are equally responsive. This suggests that additional determinants governing compound specificity are required to render an exon responsive to small molecule splicing modifiers. Through transcriptomic analysis of in-cell splice events induced by splicing modifier treatment, we elucidated several key factors governing the compound-mediated inclusion of exons. First, we highlighted the significance of gene expression profiles in identifying potential targets of splicing modifiers, emphasizing the importance of considering expression alongside primary sequence. Furthermore, our investigation into the complementarity between the 5'ss and the U1 snRNA revealed a crucial role that the snRNA plays in mediating compound-mediated exon inclusion. Exons with higher complementarity proved more likely to be targetable by splicing modifiers, whereas those with lower complementarity tend to be skipped, regardless of the intervention. Our experiments with mutated U1 snRNA showed that changing U1–5'ss pairing can reprogram compound specificity. This, in turn, highlights the strong dependence of the observed compound specificity on the highly conserved U1 snRNA sequence and explains why the actual specificity of these compounds tends to exceed theoretical expectations. To gain a deeper understanding of what governs the apparent primary sequence specificity of NVS-SM1 and Risdiplam, we took a closer look at potential bulge formation at the -1 position of the exon end. In line with previously reported biophysical data (Malard et al., 2024), our cellular and biochemical data corroborates the necessity of a bulge formation at the -1 position of the exon end. Normally, covariation helps identify structural specificity; however, in this case, covariation does not occur due to the highly conserved U1 snRNA sequence. So, we used mutated U1 snRNA in assays to test if we could shift compound sequence specificity. Notably, we were able to induce compound binding to previously a non-amenable CA-ending exon when the U1 snRNA was mutated to maintain a bulged-out adenine at the -1 position of the pre-mRNA 5'ss. This further emphasizes the role of the highly conserved U1 snRNA sequence in governing compound specificity via the determination of U1-5'ss RNA duplex secondary structure. Thus, our

findings lead us to propose that the AGA or ANGA sequence specificity arises from two key factors. Firstly, it is essential to maintain a specific secondary structure that allows for bulge formation, resulting in the requirement of a G at the -2 position. Secondly, both adenines appear to be essential. Since they do not contribute to the stability of a secondary structure, they are likely to form interactions with the compounds, thereby contributing to the observed primary sequence specificity. Overall, our findings shed light on the intricate interplay between primary sequence and secondary structure requirements in governing the overall observed specificity of small molecule splicing modifiers in cells. The specificity of a U1 snRNA can be influenced by the addition of a splicing-modulating compound. Genetic manipulation of the U1 snRNA, on the other hand, can change the specificity of a splicing modulating compound. These insights may aid future development of small molecules targeting RNA splicing as well as RNA secondary structures. Moreover, these findings will help to refine prediction algorithms, paving the way for identification of novel targets and rational design of new splicing-modifying compounds (Wilkinson et al., 2020).

Methods

Cell Culture: SH-SY5Y neuroblastoma cells were cultured using Advanced MEM media supplemented with 10% heat-inactivated fetal bovine serum (HI FBS) and 1X Glutamine (5 mL per 500 mL of media). Upon reaching confluency, the cells were sub-cultured and plated into 12-well plates at a density of 0.6 x 10^6 cells/mL, with 1 mL per well. On the second day post-plating, the cells were treated with compounds at different concentrations, DMSO treatment condition is used as control. Cells were harvested 24 hours post-treatment. For iNGN2 cells, day 3 iNGN2 neurons were thawed and plated onto six-well plates, pre-coated with poly-D-lysine (PDL) and a mixture of 1X Matrigel and laminin (1:300, Sigma-Aldrich, L2020-1MG). The coating procedure involved incubating the plates at 37℃ for at least 1 to 1.5 hours prior to cell plating. Neurons were maintained by providing half of the usual volume of culture medium every other day until day 10. On day 10, they were treated with compounds at different concentrations. DMSO treatment condition is used as control. Cells were harvested 24 hours post-treatment. HeLa 293 cells were maintained in DMEM with 10% fetal bovine serum (FBS) and 1% GlutaMax at 37 ℃ with 5% CO_2. They were plated on a 24-well plate (80.000 cells/well) and treated with potential compounds or DMSO as control the next day. Cells were harvested 24 hours post-treatment. **Expression of U1 snRNA:** For the ectopical expression of U1 snRNA, we used a plasmid either carrying the sequence of the canonical U1 snRNA or a mutated U1 snRNA version (e.g. U1-G_6C_7). HeLa cells were plated on a 24-well plate (80.000 cells/well) and transfected the following day with 0.3 µg DNA and 0.9 ml LipofectamineTM 2000 (Thermo Fisher Scientific), according to the supplier's protocol. After 24 h, medium was changed and potential compounds added or DMSO as a control. After 24 h incubation, RNA purification was performed using the *Quick*-RNA Miniprep Kit (Zymo Research). **RNA sample preparation and sequencing:** RNA samples were collected from SH-SY5Y and iNGN2 cells treated with DMSO or compounds using the RNeasy Plus 96 Kit (Qiagen) according to the manufacturer's instructions. RNA purification from HeLa cells purification was performed using the *Quick*-RNA Miniprep Kit (Zymo Research). RNA quality was assessed using the TapeStation 4200, with all samples exhibiting an RNA integrity number (RIN) greater than 9. Next-generation sequencing libraries were prepared with the TruSeq Stranded mRNA Sample Preparation Kit (Illumina) from 200 ng of input RNA, following the manufacturer's protocol. The libraries were then pooled and loaded onto a NovaSeq 6000 (Illumina) for paired-end 51 bp sequencing, yielding an average of 30-40 million reads per sample. **Detection of novel exons and compound-mediated splicing:** RNA-seq data were mapped to the hg38 human reference genome using STAR. To identify unannotated exons in the RNA-seq data, we used the split alignments reported by STAR. For a given RNA-seq sample, we mapped introns based on reads whose mapping to the reference genome was split between non-consecutive nucleotides (split reads). For two consecutive annotated exons, these reads would partially map to one exon and partially to the second exon, thus leaving an unmapped region in between that corresponds to the intronic coordinate. As the mapping of all intronic regions of a gene defines the coordinates of the non-terminal exons of a gene, these regions were considered potential novel exons. We considered introns for a given sample when they were supported by at least 2 reads and filtered for novel exons that were supported by at least one third of the samples in each experiment. These potential novel exons, if unannotated, were included to the reference transcript annotation for downstream analyses. For each exon in each sample, we counted the number of reads whose mapping supported the inclusion of the exon and the number of reads whose mapping supported the exclusion of the exon. We estimated percent spliced in values as described in (Schafer et al., 2015) and used the Bioconductor package DEXSeq to test for differences between conditions as described in (Reyes and Huber, 2018). **FP binding assay:** In the FP binding experiments, 20 nmol/L TAMRA-labeled U1 snRNA target oligonucleotides were incubated with increasing concentrations of probe oligonucleotides carrying the sequence of the U1 binding site of the pseudoexon in the *HTT* gene that is being included in response to NVS-SM1. 50 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 75 mM KCl and 0.05% Tween were used as buffer. The samples were thoroughly mixed, incubated for 5 min at 95

℃ to unfold possible RNA structures and slowly cooled down to room temperature (RT). Splicing modifiers were added and incubated for another 1 h at RT. In white 384-well microplates with 30 µL per well using a PHERAstar, the parallel emission intensity (S) and perpendicular emission intensity (P) were measured. FP values were calculated using the following equation, as described previously (Zhang et al., 2006; Zhu et al., 2018): $FP(mP) = 1000 * (S - G * P)/(S + G * P)$ All samples were measured in triplicates. The FP values were plotted against the concentration of the probe RNA oligonucleotide in each experiment. Using GraphPad Prism 9.0, a sigmoidal dose-response curve was added to determine EC_{50} values. **2-AP assay:** Oligonucleotides carrying the sequence of the exon-binding part within the U1 snRNA and those with sequences corresponding to the U1 binding site that contain a 2-AP substitution at the -1 position, were mixed in a 1× Assay Buffer (0.185 M NaCl, 8 mM NaH_2PO_4, pH 6.3). In the absence or presence of potentially binding compounds, they were heated to 70 ℃ for 3 min and slowly cooled down to allow folding. In the absence of a compound DMSO is used as control. Fluorescence excitation and emission wavelengths (310nn and 390nm, respectively) were measured in a 384-well plate at RT with a Tecan plate reader (gain, 155; integration time, 20 µs). All assays were carried out in triplicates.

Data Availability

The RNA-seq data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1293306. The dataset includes raw FASTQ files and associated sample metadata. The data will be made publicly available upon publication.

Acknowledgment

We thank everyone involved in projects working with splicing modifiers. Moreover, we want to thank Gregory Marszalek for his support in the lab and Anne Granger for her support through the postdoctoral program at Novartis.

Bibliography

Bhattacharyya, A. et al. Small molecule splicing modifiers with systemic htt-lowering activity. *Nature Communications*, 12(1):7299, 2021. doi: 10.1038/s41467-021-27661-5.

Black, D. L. Mechanisms of alternative pre-messenger rna splicing. *Annual Review of Biochemistry*, 72:291–336, 2003. doi: 10.1146/annurev.biochem.72.121801.161720.

Bradrick, T. D. and Marino, J. P. Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to hiv-1 tar rna. *RNA*, 10(9):1459–1468, 2004. doi: 10.1261/rna.7650704.

Campagne, S. et al. Structural basis of a small molecule targeting rna for a specific splicing correction. *Nature Chemical Biology*, 15(12):1191–1198, 2019. doi: 10.1038/s41589-019-0377-2.

Chen, J. L. et al. Design, optimization, and study of small molecules that target tau pre-mrna and affect splicing. *Journal of the American Chemical Society*, 142(19):8706–8727, 2020. doi: 10.1021/jacs.0c02213.

Consortium, T. G. The genotype-tissue expression (gtex) project. *Nature Genetics*, 45(6): 580–585, 2013. doi: 10.1038/ng.2653.

Dhillon, S. Risdiplam: First approval. *Drugs*, 80(17):1853–1858, 2020. doi: 10.1007/s40265-020-01394-7.

Freund, M. et al. Extended base pair complementarity between u1 snrna and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Research*, 33(16):5112–5119, 2005. doi: 10.1093/nar/gki812.

Havens, M. A. and Hastings, M. L. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Research*, 44(14):6549–6563, 2016. doi: 10.1093/nar/gkw589.

Ishigami, Y. et al. Specificity, synergy, and mechanisms of splice-modifying drugs. *Nature Communications*, 15(1), 2024. doi: 10.1038/s41467-024-46090-5.

Keller, C. G. et al. An orally available, brain penetrant, small molecule lowers huntingtin levels by enhancing pseudoexon inclusion. *Nature Communications*, 13(1):1150, 2022. doi: 10.1038/s41467-022-28847-4.

Krach, F. et al. An alternative splicing modulator decreases mutant htt and improves the molecular fingerprint in huntington's disease patient neurons. *Nature Communications*, 13(1):6797, 2022. doi: 10.1038/s41467-022-34396-0.

Le, T. T. et al. Smndelta7, the major product of the centromeric survival motor neuron (smn2) gene, extends survival in mice with spinal muscular atrophy and associates with full-length smn. *Human Molecular Genetics*, 14(6):845–857, 2005. doi: 10.1093/hmg/ddi078.

Lee, Y. and Rio, D. C. Mechanisms and regulation of alternative pre-mrna splicing. *Annual Review of Biochemistry*, 84:291–323, 2015. doi: 10.1146/annurev-biochem-060614-034316.

Lefebvre, S. et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, 80(1):155–165, 1995. doi: 10.1016/0092-8674(95)90460-3.

Malard, F. et al. The diversity of splicing modifiers acting on a-1 bulged 5'-splice sites reveals rules for rational drug design. *Nucleic Acids Research*, 52(8):4124–4136, 2024. doi: 10.1093/nar/gkae200.

Mendell, J. R. et al. Eteplirsen for the treatment of duchenne muscular dystrophy. *Annals of Neurology*, 74(5):637–647, 2013. doi: 10.1002/ana.23982.

Mercuri, E. et al. Nusinersen versus sham control in later-onset spinal muscular atrophy. *New England Journal of Medicine*, 378(7):625–635, 2018. doi: 10.1056/NEJMoa1710504.

Monani, U. R. et al. A single nucleotide difference that alters splicing patterns distinguishes the sma gene smn1 from the copy gene smn2. *Human Molecular Genetics*, 8(7):1177–1183, 1999. doi: 10.1093/hmg/8.7.1177.

Palacino, J. et al. Smn2 splice modulators enhance u1-pre-mrna association and rescue sma mice. *Nature Chemical Biology*, 11(7):511–517, 2015. doi: 10.1038/nchembio.1834.

Papasaikas, P. and Valcarcel, J. The spliceosome: The ultimate rna chaperone and sculptor. *Trends in Biochemical Sciences*, 41(1):33–45, 2016. doi: 10.1016/j.tibs.2015.11.003.

Rachofsky, E. L., Osman, R., and Ross, J. B. Probing structure and dynamics of dna with 2-aminopurine: Effects of local environment on fluorescence. *Biochemistry*, 40(4):946–956, 2001. doi: 10.1021/bi001763s.

Ratni, H. et al. Specific correction of alternative survival motor neuron 2 splicing by small molecules: Discovery of a potential novel medicine to treat spinal muscular atrophy. *Journal of Medicinal Chemistry*, 59(13):6086–6100, 2016. doi: 10.1021/acs.jmedchem.6b00474.

Ratni, H. et al. Discovery of risdiplam, a selective survival of motor neuron-2 (smn2) gene splicing modifier for the treatment of spinal muscular atrophy (sma). *Journal of Medicinal Chemistry*, 61(15):6501–6517, 2018. doi: 10.1021/acs.jmedchem.8b00741.

Rau, M. J. and Hall, K. B. 2-aminopurine fluorescence as a probe of local rna structure and dynamics and global folding. *Methods in Enzymology*, 558:99–124, 2015. doi: 10.1016/bs.mie.2015.02.005.

Reyes, A. and Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Research*, 46(2):582–592, 2018. doi: 10.1093/nar/gkx1165.

Ruggiu, M. et al. A role for smn exon 7 splicing in the selective vulnerability of motor neurons in spinal muscular atrophy. *Molecular and Cellular Biology*, 32(1):126–138, 2012. doi: 10.1128/MCB.06077-11.

Schafer, S. et al. Alternative splicing signatures in rna-seq data: Percent spliced in (psi). *Current Protocols in Human Genetics*, 87:11.16.1–11.16.14, 2015. doi: 10.1002/0471142905.hg1116s87.

Singh, R. N., Seo, J., and Singh, N. N. Rna in spinal muscular atrophy: Therapeutic implications of targeting. *Expert Opinion on Therapeutic Targets*, 24(8):731–743, 2020. doi: 10.1080/14728222.2020.1799523.

Tang, Z. et al. Rna-targeting splicing modifiers: Drug development and screening assays. *Molecules*, 26(8), 2021. doi: 10.3390/molecules26082253.

Warner, K. D., Hajdin, C. E., and Weeks, K. M. Principles for targeting rna with drug-like small molecules. *Nature Reviews Drug Discovery*, 17(8):547–558, 2018. doi: 10.1038/nrd.2018.93.

Wilkinson, M. E., Charenton, C., and Nagai, K. Rna splicing by the spliceosome. *Annual Review of Biochemistry*, 89:359–388, 2020. doi: 10.1146/annurev-biochem-091719-064225.

Wong, M. S., Kinney, J. B., and Krainer, A. R. Quantitative activity profile and context dependence of all human 5' splice sites. *Molecular Cell*, 71(6):1012–1026.e3, 2018. doi: 10.1016/j.molcel.2018.08.015.

Zhang, T. T. et al. High-throughput fluorescence polarization method for identifying ligands of lox-1. *Acta Pharmacologica Sinica*, 27(4):447–452, 2006. doi: 10.1111/j.1745-7254.2006.00304.x.

Zhu, M. R. et al. Development of a high-throughput fluorescence polarization assay for the discovery of ezh2-eed interaction inhibitors. *Acta Pharmacologica Sinica*, 39(2):302–310, 2018. doi: 10.1038/aps.2017.117.

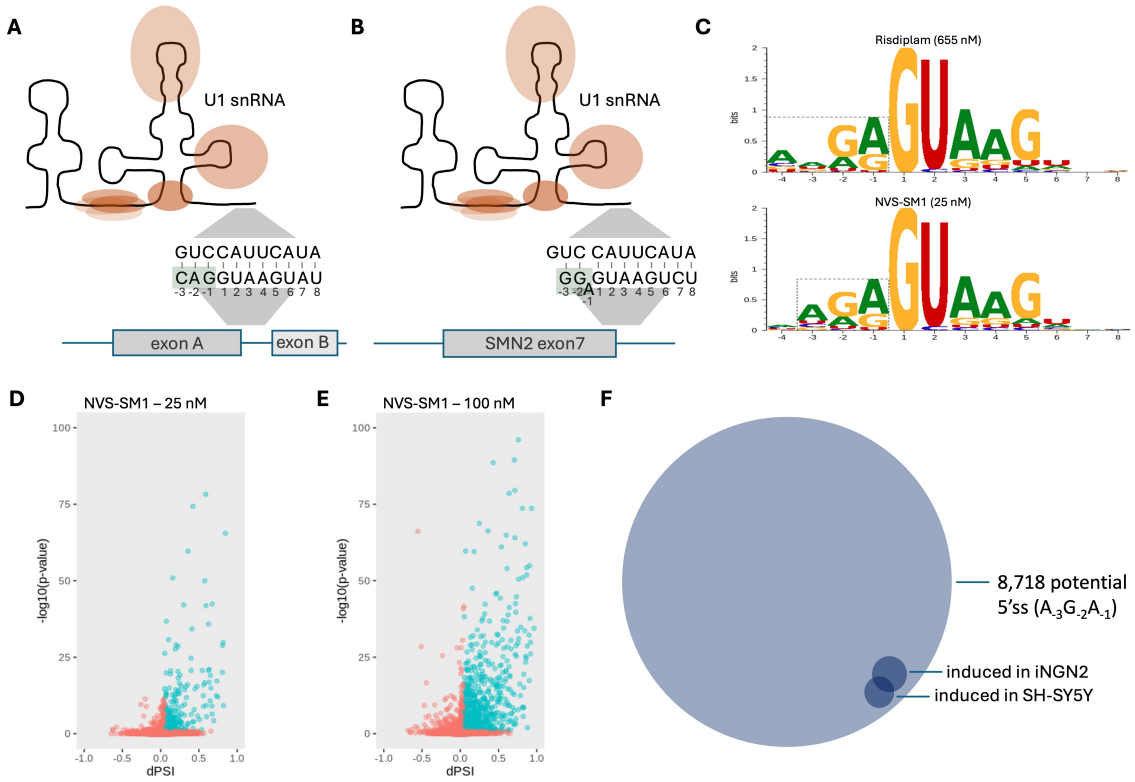

Figure 1. Sequence specificity of splicing compounds. (A) Splice site recognition by base-pairing between the U1 snRNA within the spliceosome and an exon carrying a consensus U1 binding site (here called exon A). The U1 binding site spanes from the three nucleotides at the 3' end of exon (positions -3, -2, -1) to the eight nucleotides of the downstream intron (positions +1 to +8). (B) Splice site recognition of exon 7 in the *SMN2* gene in the presence of a splicing modifier like Risdiplam. (C) RNA sequence logos to represent sequence conservation at the U1 binding site of all exons included in response to either 655nM Risdiplam, or 25nM NVS-SM1. The last 4 nucleotides of the exon (position -4 to -1) and the first 8 nucleotides of the following intron (position +1 to +8) are included in the logo. (D) Differential splicing analysis between whole transcriptome profiles of SH-SY5Y cells treated with 25 nM NVS-SM1 and SH-SY5Y cells treated with DMSO. Exons highlighted in turquoise are considered as included in response to compound treatment (dPSI > 0.05, adjusted p-value < 0.1) (E) Differential splicing analysis between whole transcriptome profiles of SH-SY5Y cells treated with 100 nM NVS-SM1 and SH-SY5Y cells treated with DMSO. (F) Venn diagram to highlight gap between 8,718 potential 5'ss matching primary sequence preference of NVS-SM1 and the actual number of detected induced AGA-ending exons in response to NVS-SM1 treatment in SH-SY5Y and iNGN2 cells.

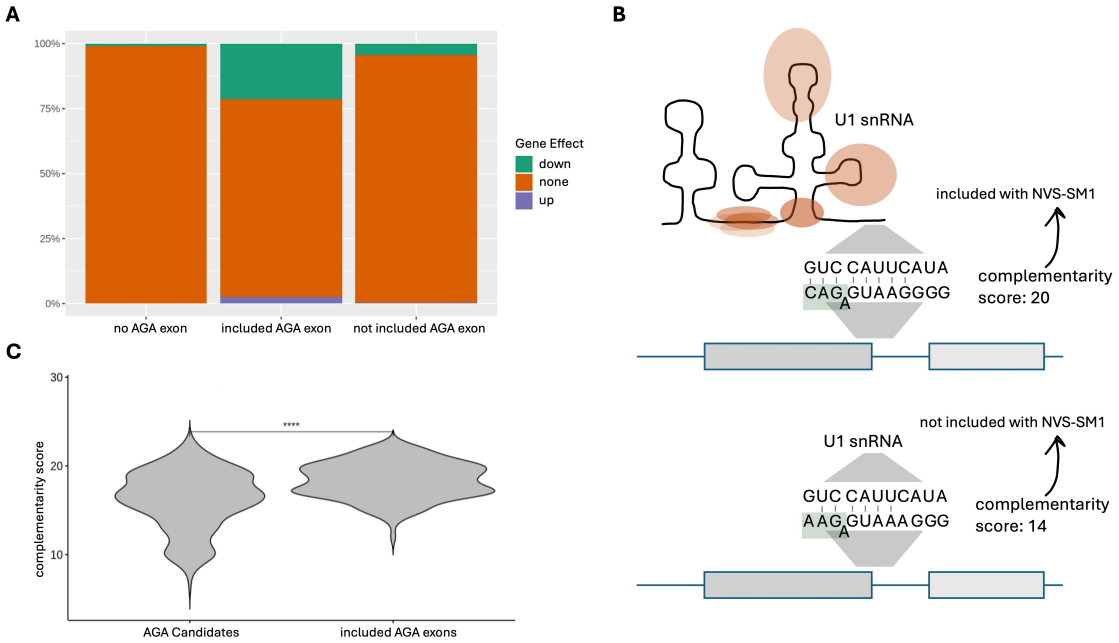

Figure 2. Compound effects on gene expression and U1 snRNA complementarity in exon inclusion. (A) Effect on gene expression (upregulation, downregulation or no effect) of NVS-SM1 treatment of SH-SY5Y cells. 8.5% of the genes carrying an AGA exon, not inducible upon NVS-SM1, are downregulated. (B) Violin plot showing the complementarity between the U1 snRNA and U1 binding sites of AGA-ending exons included in response to NVS-SM1 (included AGA exons) and those 5'ss matching a 5'-$A_{-3}G_{-2}A_{-1}G_1U_2$-3' motif, thus being potential candidates for NVS-SM1 mediated inclusion (AGA Candidates). The complementarity score was calculated by summing up all H-bonds possibly formed between the two RNAs. (C) Example of two AGA exons that seem to be good candidates for NVS-SM1 mediated inclusion due to their primary sequence at the exon end. Here shown as complementarity score is the number of maximal H-bonds between the U1 binding site and the U1 snRNA.

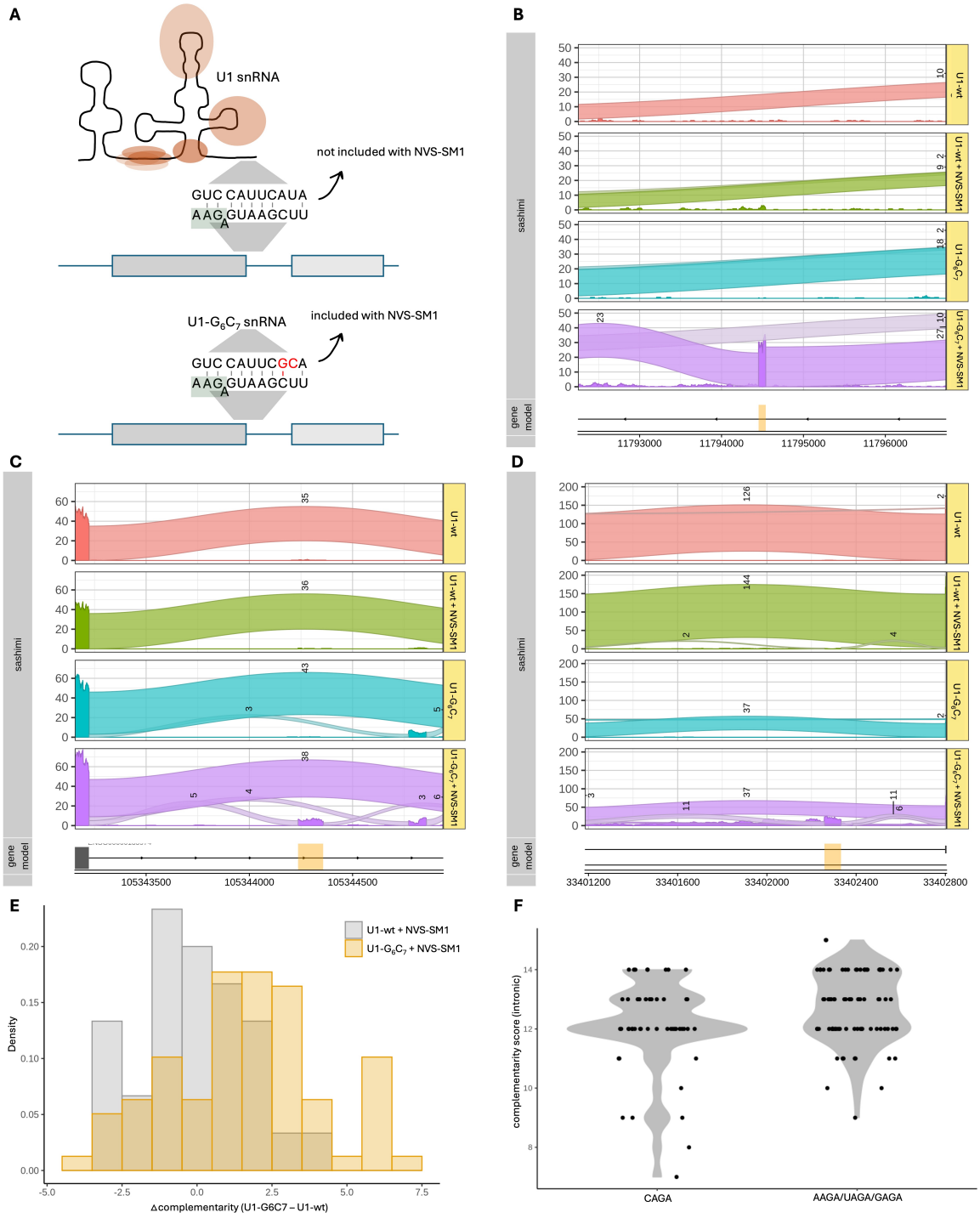

Figure 3. Complementarity between U1 snRNA and the U1 binding site is a critical factor for exon responsiveness. (A) Cartoon showing the complementarity of the exon highlighted in (B) with the canonical and the mutated U1 snRNA. Complementarity score with U1-wt was 19 and the score calculated for U1-G_6C_7 was 22. (B) Sashimi plots of RNA-seq results deriving from HeLa cells either transfected with canonical U1 snRNA (U1-wt) or with mutated U1 snRNA (U1-G_6C_7), grown in the absence or presence of NVS-SM1, each performed in triplicates. Plot highlights a cryptic exon on chromosome 3 within the gene *TAMM41*. Significant inclusion only detected in cells transfected with U1-G_6C_7 and treated with 100nM NVS-SM1 (dPSI = 0.8, p-value < 0.001). (C) Sashimi plots of RNA-seq results deriving from the same experiment as explained in A. Plot highlights an exon on chromosome 2 within the gene *C2orf49*. Significant inclusion only detected in cells transfected with U1-G_6C_7 and treated with 100nM NVS-SM1 dPSI = 0.16, p-value < 0.001). (D) Sashimi plots of RNA-seq results deriving from the same experiment as explained in A. Plot highlights an exon on chromosome 3 within the gene *UBP1*. Significant inclusion only detected in cells transfected with U1-G_6C_7 and treated with 100nM NVS-SM1 (dPSI = 0.52, p-value < 0.001). (E) Δcomplementarity between canonical and mutated U1 snRNA of U1 binding sites from exons included in response to 100nM NVS-SM1 and canonical U1 snRNA (U1-wt + NVS-SM1) and exons only included when transfected with mutated U1 snRNA and NVS-SM1 (U1-G_6C_7 + NVS-SM1). (Δcomplementarity was calculated by subtraction of the complementarity score calculated using the canonical U1 snRNA sequence from the score calculated using the mutated U1 snRNA. (F) Complementarity of intronic base-pairing (positions +1 through +8) of exons included upon NVS-SM1 treatment, grouped them according to their -4 position. Right group shows all exons with a $C_{-4}A_{-3}G_{-2}A_{-1}$ sequence while the left group shows all other AGA exons ($A_{-4}A_{-3}G_{-2}A_{-1}$, $U_{-4}A_{-3}G_{-2}A_{-1}$ or $G_{-4}A_{-3}G_{-2}A_{-1}$). Significant co-dependence of low-complementarity sites with the position -4 detected (p-value = 0.02).

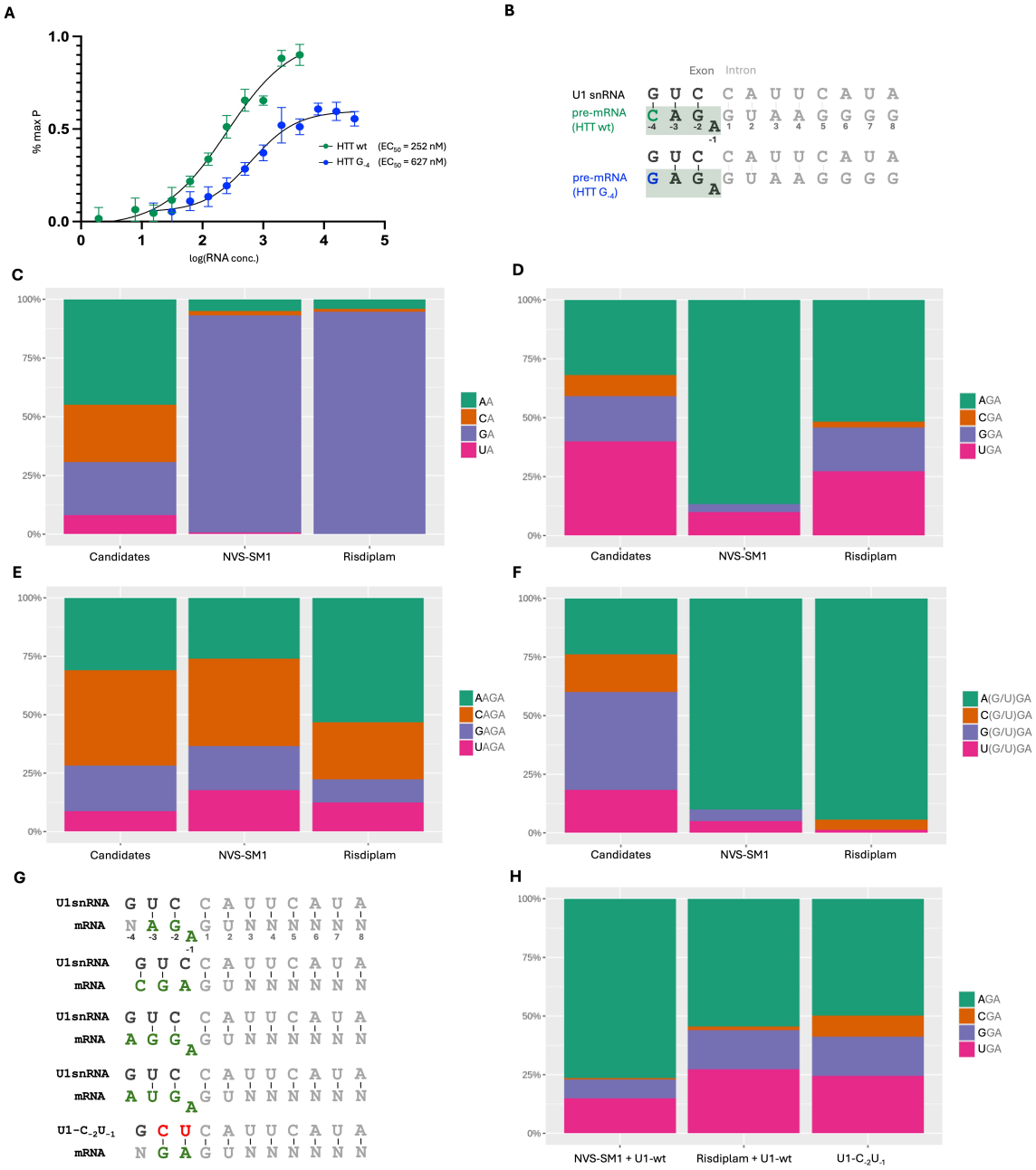

Figure 4. Bulge formation and sequence features of exons induced by NVS-SM1 and Risdiplam. (A) Fluorescence polarization (FP) binding curves of TAMRA-labeled U1 snRNA incubated with serial dilutions of the U1 binding site of the *HTT* pseudoexon, either the wild-type sequence (wt), or a mutated version (HTT G_{-4}). Samples were measured in triplicates, and the data analyzed in GraphPad Prism 9.0. A trendline was generated using a sigmoidal dose-response curve fit to determine an EC_{50} value. (B) Constructs used for the FP-assay. (C) Frequency of each nucleotide at position -2 within all A-ending exons (Candidates) or all induced A-ending exons upon NVS-SM1 or Risdiplam treatment. (D) Frequency of each nucleotide at position -3 within all GA-ending exons (Candidates) or all induced GA-ending exons upon NVS-SM1 or Risdiplam treatment in SH-SY5Y cells. (E) Frequency of each nucleotide at position -4 within all AGA exons (Candidates) or all induced AGA exons upon NVS-SM1 or Risdiplam treatment. (F) Frequency of each nucleotide at position -4 of all UGA or GGA exons (Candidates) or all induced UGA or GGA exons upon NVS-SM1 or Risdiplam treatment. (G) Potential bulge formation in RNA duplexes formed between U1 snRNA and the exon-intron junction of GA-ending exons. U1-$C_{-2}U_{-1}$ shows sequence of the mutated U1 snRNA, modified to directly induce GA-ending exons in the absence of a compound. (H) Frequency of each nucleotide at position -3 of GA-ending exons induced in response to NVS-SM1 or Risdiplam treatment in HeLa cells transfected with wt U1 snRNA or in cells transfected with U1-$C_{-2}U_{-1}$ and treated with DMSO.

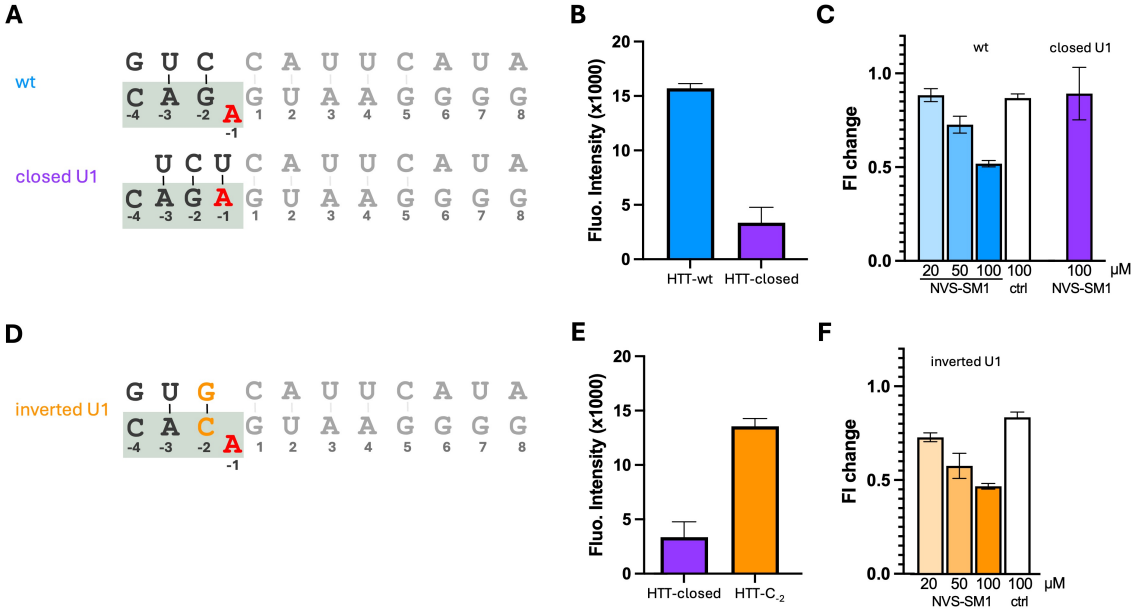

Figure 5. Biochemical Assays to investigate bulge formation at U1 binding site of the *HTT* pseudoexon. (A) RNA constructs used for the 2-AP assay shown in (B) and (C). RNA constructs carrying either wild-type sequence of the RNA complex formed when *HTT* pseudoexon is recognized (wt, top) or a mutated version, forming a closed duplex (bottom, closed U1). Position of the 2-AP mutation highlighted in red. (B) Fluorescence Intensity (FI) measured in the 2-Aminopurine (2-AP) assay with wt or closed U1 RNA construct (sequences shown in (A)). (C) Changes in FI upon titration of NVS-SM1 or inactive compound (ctrl, shown as neg.ctrl in Supplementary Figure S1G) to RNA duplexes, relative to the DMSO condition. (D) RNA constructs used for the 2-AP assay shown in (E) and (F). RNA construct carries a mutated version of the RNA complex formed when *HTT* pseudoexon is recognized, changing primary sequence while maintaining possibility of bulge formation. 2-AP mutation highlighted in red. (E) Fluorescence Intensity (FI) measured in the 2-Aminopurine (2-AP) assay with the closed U1 or inverted U1 RNA construct (sequences shown in (A) and (D)). (F) Changes in FI upon titration of NVS-SM1 or inactive compound (ctrl, shown as neg.ctrl in Supplementary Figure S1G) to RNA duplexes, relative to the DMSO condition.

Supplementary Information

Table S1. List of some NVS-SM1-induced exons with corresponding complementarity scores. Scores were calculated by summing the maximum number of possible H-bonds formed between the 5'ss of AGA exons and U1 snRNA. We also considered the possibility of an A_{-1} bulge formation, thus including the -4 position of the exon in the score calculation.

Gene name	chr.	start	end	U1 binding site	Complementarity score
HTT	4	3213622	3213736	CAGAGTAAGGGG	20
KCNT2	1	196317148	196317173	CAGAGTAAGAAA	21
LRRC42	1	53947862	53947981	TAGAGTAAGTTA	19
MLLT10	10	21728632	21728675	GAGAGTAAGAGA	18
ARHGAP1	11	46697021	46697069	CAGAGTAAGAGT	20
WNK1	12	895173	895196	GAGAGTAGGTGA	19

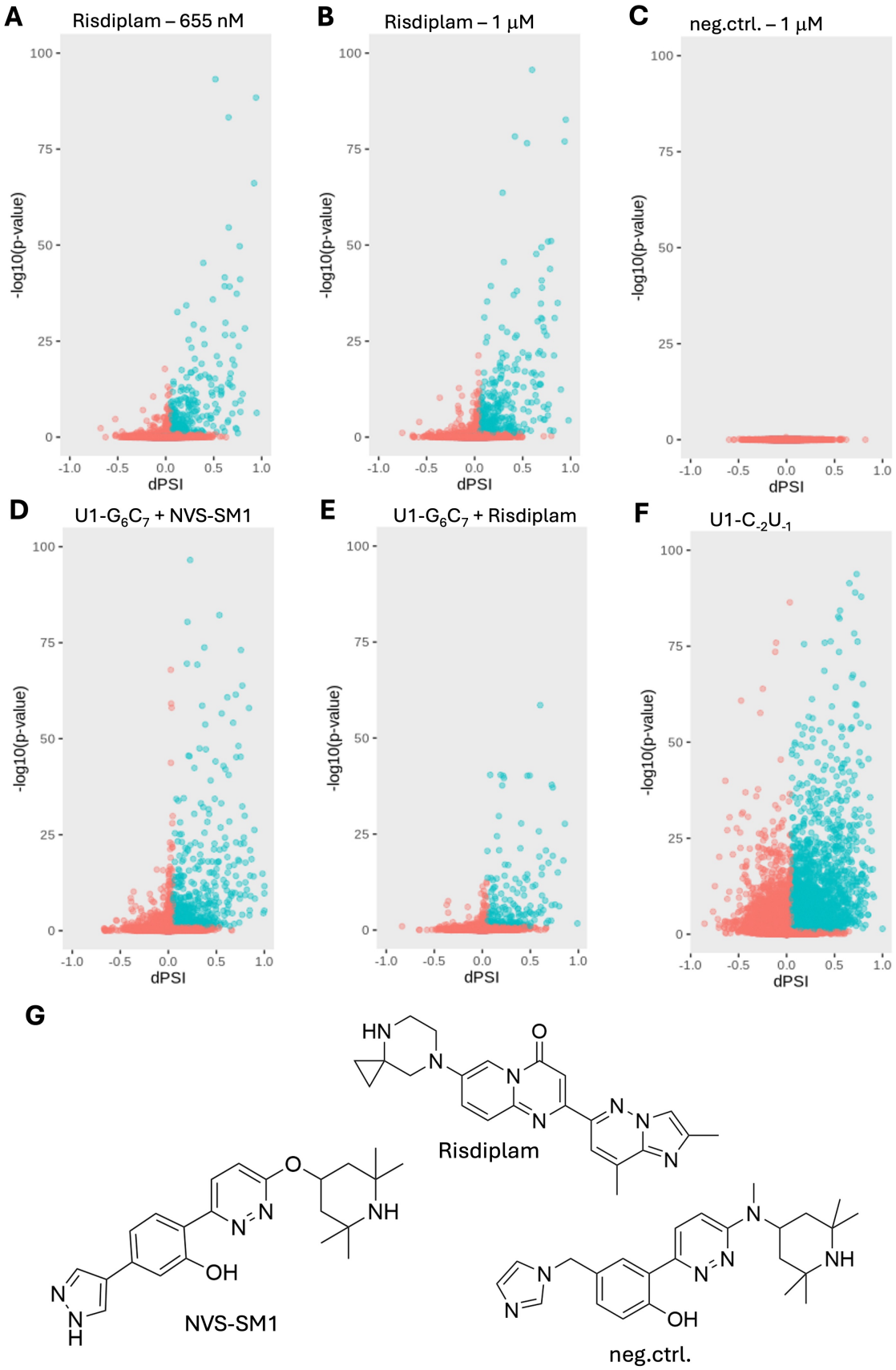

Figure S1. Differential splicing analysis between whole transcriptome profiles of SH-SY5Y cells treated with (A) 655 nM Risdiplam, (B) 1 µM Risdiplam or (C) 1 µM of inactive compound, and SH-SY5Y cells treated with DMSO. Differential splicing analysis between whole transcriptome profiles of HeLa cells transfected with U1-G_6C_7 and treated with (D) 100 nM NVS-SM1 or (E) 1 µM Risdiplam, and transfected HeLa cells treated with DMSO. (F) Differential splicing analysis between whole transcriptome profiles of HeLa cells transfected with U1-$C_{-2}U_{-1}$ and treated with DMSO and HeLa cells transfected with canonical U1. Exons highlighted in turquoise are considered as included in response to compound treatment (dPSI >0.05, adjusted p-value <0.1). (G) Chemical structures of the reference compounds used in this study: NVS-SM1, Risdiplam, and an inactive compound used as control (neg.ctrl.).

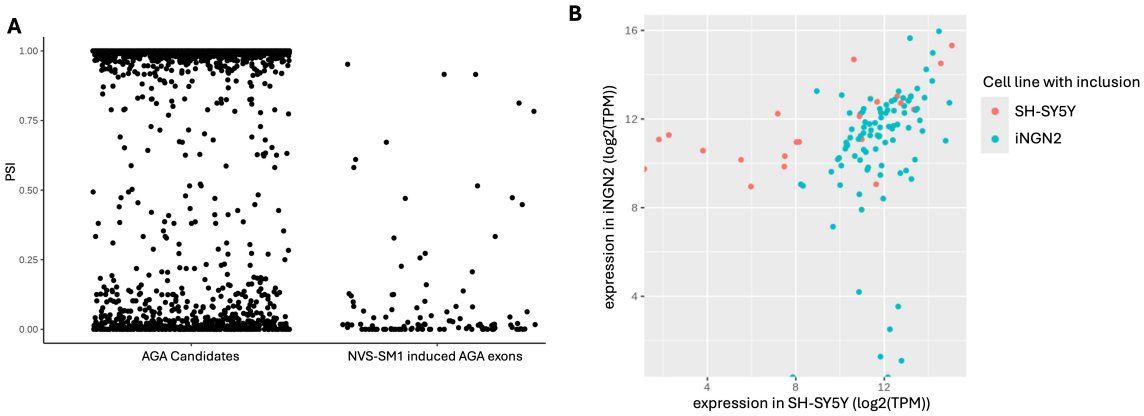

Figure S2. (A) Detected baseline inclusion in SH-SY5Y cells in the absence of a compound. Left group (AGA Candidates) represents all 5'ss matching an $5'$-A$_{-3}$G$_{-2}$A$_{-1}$G$_1$U$_2$-$3'$ motif, identified as potential targets for NVS-SM1 in the human RNA-seq data collection from the Genotype-Tissue Expression Project. Right group (NVS-SM1 induced AGA exons) contains all AGA exons detected to be significantly induced in respond 25nM NVS-SM1 treatment. (B) Expression levels (log2(TPM)) of genes containing AGA-ending exons significantly induced only in SH-SY5Y cells or only in iNGN2 cells, respectivly. Expression in iNGN2 cells shown on y-axis and expression in SH-SY5Y cells shown on x-axis.

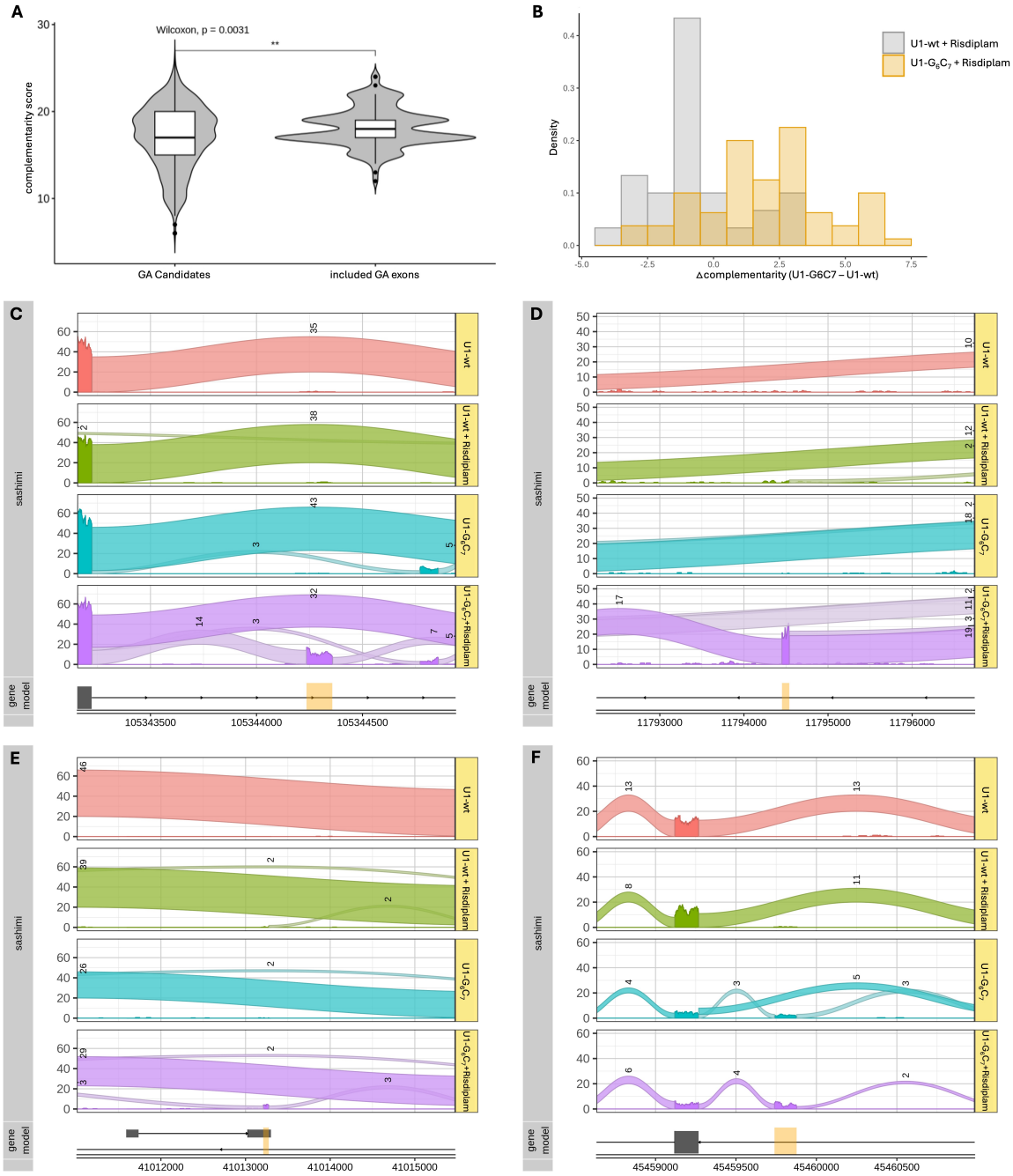

Figure S3. (A) Violin plot showing the complementarity between the U1 snRNA and U1 binding sites of GA-ending exons included in response to Risdiplam (included GA exons) and those GA exons that were potential targets for Risdiplam (GA Candidates) (p-value < 0.0001). The complementarity score was calculated by summing up all H-bonds possibly formed between the two RNAs. (B) Δcomplementarity between canonical and mutated U1 snRNA. Grey shows Δcomplementarity exons included in response to 1µM Risdiplam and transfection with canonical U1 snRNA (U1-wt+Risdiplam). Exons only included when transfected with mutated U1 snRNA and Risdiplam (U1-G_6C_7+Risdiplam) are shown in yellow. (C) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene ERGIC1. Significant inclusion only detected in cells transfected with U1-G_6C_7 and treated with 1µM Risdiplam NVS-SM1 dPSI = 0.36, p-value < 0.001). (D) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene *TAMM41*. Significant inclusion only detected in cells transfected with U1-G_6C_7 and treated with 1µM Risdiplam NVS-SM1 dPSI = 0.73, p-value < 0.001). (E) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene *INO80*. Low, but not significant inclusion detected in cells transfected with canonical U1 snRNA and 1µM Risdiplam. Significant inclusion only detected in cells transfected with U1-G_6C_7 and Risdiplam treatment (dPSI = 0.17, p-value < 0.001). (F) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene MARCHF8. Significant inclusion detected in cells transfected with U1-G_6C_7 (dPSI = 0.79, p-value < 0.001) and boosted inclusion in cells transfected with U1-G_6C_7 and additional Risdiplam treatment (dPSI = 1.0, p-value < 0.001).

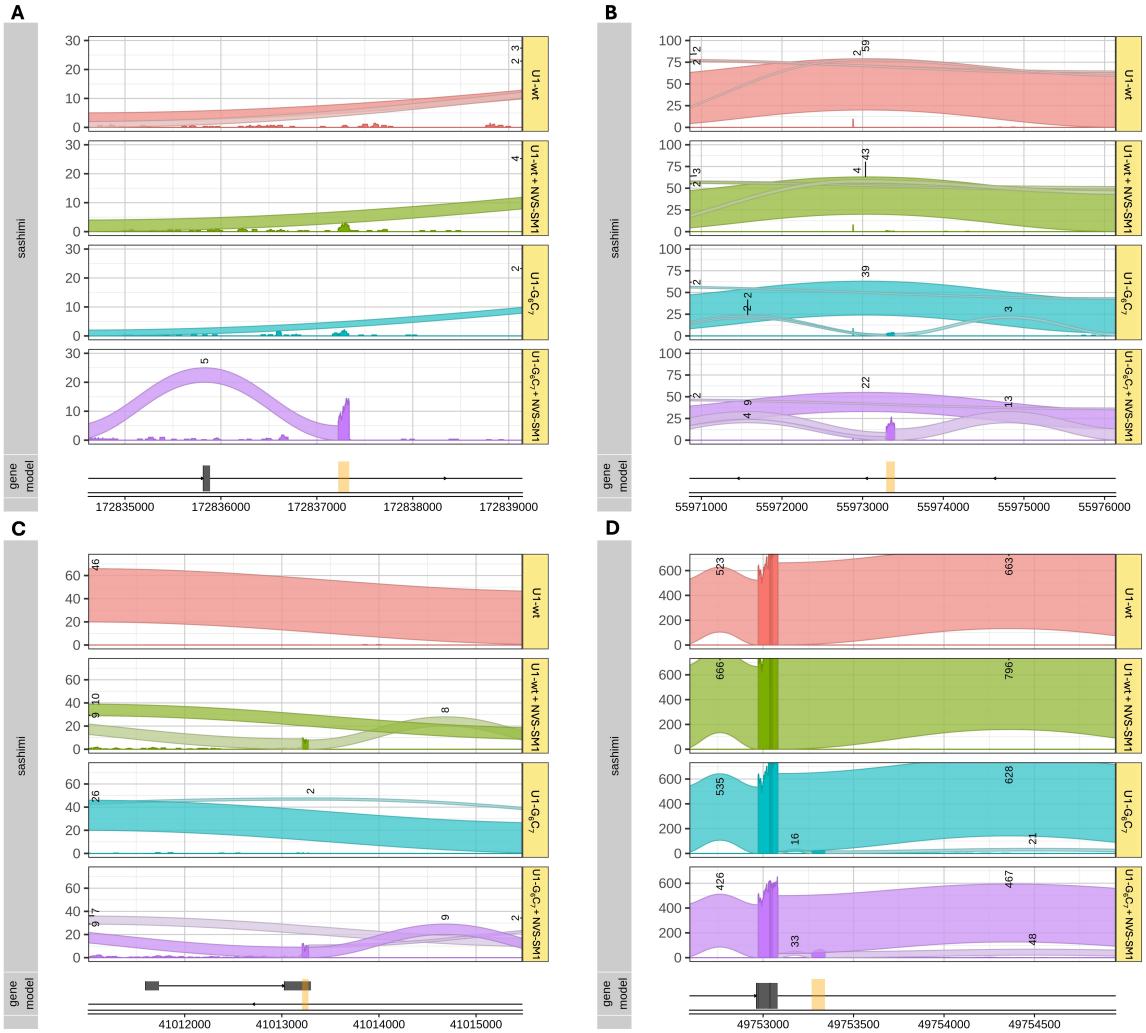

Figure S4. (A) Sashimi plots of RNA-seq results deriving from the same experiment as in Figure 3A. Plot highlights an exon within the gene *ERGIC1*. Significant inclusion only detected in cells transfected with U1-G_6C_7 and treated with 100nM NVS-SM1 dPSI = 0.13, p-value < 0.001). (B) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene *IL6ST*. Significant inclusion detected in cells transfected with U1-G_6C_7 (dPSI = 0.1, p-value < 0.001) and boosted inclusion in cells additionally treated with 100nM NVS-SM1 (dPSI = 0.43, p-value < 0.001). (C) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene *INO80*. Significant inclusion detected in cells transfected with canonical U1 snRNA (U1-wt) and NVS-SM1 treatment (dPSI = 0.6, p-value < 0.001) and boosted inclusion in cells transfected with U1-G_6C_7 and NVS-SM1 treatment (dPSI = 0.7, p-value < 0.001). (D) Sashimi plots of RNA-seq results deriving from the same experiment as explained in Figure 3A. Plot highlights an exon within the gene *TMBIM6*. Significant inclusion detected in cells transfected with U1-G_6C_7 (dPSI = 0.06, p-value < 0.001) and boosted inclusion in cells additionally treated with 100nM NVS-SM1 (dPSI = 0.12, p-value < 0.001).