# Mechanisms of modulation of pre-mRNA 3D structural scaffolds for splicing substrate definition

Kaushik Saha* and Gourisankar Ghosh*

Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0375

*Correspondence: gghosh@ucsd.edu, kasaha@ucsd.edu

## ABSTRACT

Coordination of different serine-arginine-rich (SR) proteins – a class of critical splicing activators – facilitates recognition of the highly degenerate cognate splice signal sequences against the background sequences. Yet, the mechanistic details of their actions remain unclear. Here we show that cooperative binding of SR proteins to exonic and intronic motifs remodels the pre-mRNA 3D structural scaffold. The scaffold generated by pre-mRNA-specific combinations of different SR proteins in an appropriate stoichiometry is recognized by U1 snRNP. A large excess of U1 snRNP particles displaces the majority of the bound SR protein molecules from the remodeled pre-mRNA. A higher than optimal stoichiometry of SR proteins occludes the binding sites on the pre-mRNA, raising the U1 snRNP levels required for SR protein displacement and potentially impeding spliceosome assembly. This novel step is important for distinguishing the substrate and the non-substrate by U2AF65 – the primary 3′ splice site-recognizing factor. Overall, this work elucidates early regulatory steps of mammalian splicing substrate definition by SR proteins.

## INTRODUCTION

The early spliceosome defines the mammalian splicing substrate by recognizing four major splice signals namely the 5′ and 3′ splice sites (SS) located at the 5′- and 3′-ends of the intron as well as branch-point site (BS) and polypyrimidine tract (PPT) near the 3′SS. U1 small nuclear ribonucleoprotein (U1 snRNP), U2 auxiliary factor 35 (U2AF35), splicing factor 1 (SF1), and U2 auxiliary factor 65 (U2AF65), respectively, are reported to recognize them (Will and Luhrmann, 2011). U2AF65 is considered to be the primary factor that defines the 3′SS (Guth et al., 2001; Shao et al., 2014). These recognition events, however, are not governed just by the splice signal sequences but a combination of a variety of features embedded within the pre-mRNA strand, collectively known as the 'splicing code' (Baralle and Baralle, 2018). The sequence of the splice signals, the pre-mRNA secondary structure (Saha et al., 2020; Saldi et al., 2021), the splicing regulatory element (SREs) (Fu and Ares, 2014), and the pre-mRNA 3D structural scaffold (Saha et al., 2021) may contribute to the splicing code of a pre-mRNA. SREs recruit from a large repertoire of RNA binding proteins (RBP) for managing the recognition-potential of a splice signal (Fu and Ares, 2014). Serine-arginine-rich (SR) proteins are key players in regulation of early spliceosome assembly and deregulation of the process causes widespread splicing changes leading to a wide range of diseases (Das and Krainer, 2014; Karni et al., 2007; Kędzierska and Piekiełko-Witkowska, 2017; Kim et al., 2016; Luo et al., 2017; Sokół et al., 2018; Sun et al., 2010; Wu et al., 2010; Yang et al., 2018). Accurate correlation of their binding to the cognate SREs within the pre-mRNA with the splicing outcome is critical for elucidation of the mammalian 'splicing code (Zhou and Fu, 2013). Earlier we showed that strandedness of exonic segments immediately upstream of the 5′SS critically regulate binding of SR proteins for pre-mRNA structural remodeling (Saha et al., 2020). We observed that multiple copies of a specific SR protein could remodel the 3D structural scaffold of the pre-mRNA generating a recognizable substrate for the early spliceosomal components (Saha et al., 2021).

The success of this remodeling event by the specific SR protein also could depend on the initial 3D structure of the pre-mRNA. That is, just the presence of the binding motifs was not sufficient to exert the specific functions of the SR proteins. However, splicing regulation involves a complex interplay among multiple SR proteins bound at both exons and introns, which could be synergistic, antagonistic, and/or compensatory (Zhou and Fu, 2013). Currently, the mechanistic detail of this collaboration remains unclear.

Interactions of SR proteins within the assembled early spliceosome is well-characterized (Jobbins et al., 2018). There a single molecule of a specific type of SR protein bound at an exonic splicing enhancer (ESE) directly interacts with an early spliceosomal component bound at its cognate splice signal sequence. However, this does not highlight the collaborative nature of SR protein-mediated regulation of early recognition of the splice sites. Additionally, early splicing substrate definition by SR proteins may also occur without a detectable interaction between the SR protein and the early spliceosomal component and could primarily be mediated by SR protein-mediated pre-mRNA structural remodeling (Saha et al., 2021). Thus, mechanistic details of the collaborative actions of the SR proteins for the early recognition of a specific SS still remains unclear (Roca et al., 2013).

There are also additional aspects of SR protein actions that are mechanistically unclear. SR proteins may repress splicing by binding to intronic motifs (Zhou and Fu, 2013) through an as yet uncharacterized mechanism. Moreover, SR proteins may bind the pre-mRNAs in multiple copies prior to assembly of the early spliceosome (Jobbins et al., 2018; Saha et al., 2020) but only one or two are reported to remain bound within the assembled early spliceosome (Jobbins et al., 2018); their displacement mechanisms are not yet characterized.

In the current work, we characterized the regulatory collaboration of three SR proteins – SRSF1, SRSF2, and SRSF5 – for recruitment of U1 snRNP to full-length pre-mRNA substrates. We primarily used human *β-globin* IVS1 for dissecting these functions since each of these SR

proteins promote the splicing of this substrate (Screaton et al., 1995) and the technique for resolving the pre-mRNA complexes could clearly distinguish the SR protein-dependent and SR protein-independent U1 snRNP-*β-globin* complexes. Our data highlights the principles of regulatory collaboration among SR proteins, identifies a novel function of U1 snRNP critical for splicing substrate definition, and provides a greater understanding of the 3D structural scaffold and its modulation for splicing substrate definition.

## RESULTS

### SRSF1-RBD stabilizes the interactions between the global 3D structural scaffold of *β-globin* and U1 snRNP

First, we examined the specific features of *β-globin* that are recognized by U1 snRNP and whether SRSF1-RBD stabilizes this interaction. We titrated 10 pM *β-globin* in the presence as well as the absence of 60 nM SRSF1-RBD with U1 snRNP (Figure 1A). With increasing concentrations of U1 snRNP, the SRSF1-RBD-bound pre-mRNA complex was downshifted and eventually formed ~ 2 faster migrating, apparently stable, complexes. We confirmed the presence of SRSF1-RBD as well as U1 snRNP in the final WT 'ternary' complexes by carrying out antibody super-shift with anti-SRSF1 and anti-U1-C antibodies, respectively (compare lanes 5 & 9 and 18 & 19). Anti-U1C antibody also disintegrated a portion of the pre-mRNA complex likely because U1-C is important for pre-mRNA binding of U1 snRNP (Kondo et al., 2015). Surprisingly, *β-globin* Δ5′SS also formed very similar complexes with U1 snRNP in the presence as well as the absence of SRSF1-RBD (Figure 1A - compare lanes 4, 5 with 13, 14 and lanes 7, 8 with 16, 17, respectively). We hypothesized that the long 3′ exon of *β-globin* (206-nt in *β-globin*) might provide sufficient support to U1 snRNP to stabilize it on the pre-mRNA given that the global structural landscape of a pre-mRNA could be involved in U1 snRNP recruitment (Saha et al., 2021). Therefore, we generated a *β-globin* construct with truncated 3′ exon (*βg-ΔEx2*). EMSA indicates U1 snRNP does not exhibit distinct interaction with free *βg-ΔEx2*. In the presence of SRSF1-RBD, U1 snRNP formed distinct complexes with *βg-ΔEx2*; however, the complexes formed with *βg-ΔEx2* Δ5′SS was distinctively weaker than the former (Figure 1B – compare lanes 3 and 9). We also generated the hybridization-mutant (EH3+4) on *βg-ΔEx2* construct, in which the critical single-stranded segments immediately upstream of the 5′SS are hybridized; this mutant does not recruit SR proteins appropriately and is defective in splicing

(Saha et al., 2020). Distinct U1 snRNP-dependent complexes were not evident either in the presence or the absence of SRSF1-RBD (Figure 1B, lanes 15 and 17).

Since the 3′ exon of *β-globin* stabilizes U1 snRNP on *β-globin*, we hypothesize that U1 snRNP recognizes a 3D structural scaffold of *β-globin*. To examine how *β-globin* structure regulates U1 snRNP binding even in the absence of 5′SS, we first examined the SHAPE reactivity of protein-free *β-globin* and its 5′SS-mutant and generated their SHAPE-derived secondary structure models (Supplementary Figure 1A, 1B). The largest structural differences were observed around the 5'SS between $60^{th}$ to $130^{th}$ positions (5′ exon ends at the $106^{th}$ position) (Figure 1C). However, strandedness of the nucleotides in the remainder of the pre-mRNA appeared somewhat similar. Moving averages (period = 6) of the SHAPE reactivity of the nucleotides between $140^{th}$ and $420^{th}$ positions of both RNAs showed a decent correlation (r=0.81, *p*= 9.77E-66). Moving averages of SHAPE reactivity of U1 snRNP-bound *β-globin* variants exhibited a slightly higher correlation even in a longer segment, between $120^{th}$ and $420^{th}$ position (r=0.83, *p*= 1.97E-78) (Figure 1D). This suggests that *β-globin* likely has a loosely organized 3D structure and is spatially brought together by the global engagement of U1 snRNP; hence, mutation in the authentic and cryptic 5′SS did not completely disrupt the 3D structural scaffold of *β-globin* and U1 snRNP recruitment to it. As expected, however, *β-globin* Δ5′SS is completely defective in splicing *in vivo* (Supplementary Figure 1C).

We then used two 14-mer short RNA oligonucleotides (*β-globin* 5'SS and *Ron* ESE, an SRSF1-binding ESE) to determine the stability of the *β-globin*:SRSF1-RBD:U1 snRNP ternary complex. Both the RNAs bind SRSF1-RBD with near equal efficiency due to the presence GGN and CN (N= any nucleotides) motifs (Cléry et al., 2021). We compared the stability of the ternary complexes with that of the *β-globin*:U1 snRNP binary complex by a challenge with the 14-nt long *β-globin* 5′SS RNA. The U1 snRNP-dependent complexes were stable in the presence of the short RNA (Figure 1E - compare lanes 3 and 6). However, the relative ratio of

the two shifted bands changed suggesting that these bands represent conformational variants that react to the 5'SS challenge differently. We observed similar phenomenon when the ternary complex was challenged with *Ron* ESE (Supplementary Figure 1D). As we noted before with short *AdML* 5′SS RNA (Saha et al., 2021), U1 snRNP did not bind strongly to short *β-globin* 5′SS RNA after displacing SRSF1-RBD (Supplementary Figure 1E). In contrast to the ternary complex, we found that the U1 snRNP:*β-globin* binary complexes were partially disrupted upon addition of excess 14-nt long *β-globin* 5′SS RNA (Figure 1F – compare lanes 1, 2, 3 with 4, 5, 6, respectively).

Overall, these results suggest that U1 snRNP recognizes a global 3D structural scaffold of *β-globin*, which is not severely impacted by the mutation in the authentic and cryptic 5′SS, and that SRSF1 enhances the global interaction of U1 snRNP to *β-globin* (Figure 1G).

**Cooperative multicopy binding of SRSF1-RBD to *β-globin* is important for stabilization of U1 snRNP**

To better understand how SRSF1-RBD stabilizes U1 snRNP, we tested concentration-dependent binding of SRSF1-RBD to 10 pM *β-globin* probe by EMSA. An apparent–binding saturation and an apparent super-saturation were observed at ~ 40 nM and ~ 70 nM protein concentrations, respectively (Figure 2A). In contrast to this high-affinity binding, $K_d$ of the SRSF1-RBD:ESE complex is around 0.5 μM (Cho et al., 2011a). Consistently, we also found that even at 240 nM, SRSF1-RBD could not saturate 14-nt long *Ron* ESE (Figure 2B), a strong binding site for SRSF1 (Cho et al., 2011a). This suggests a strong cooperation among SRSF1-RBD molecules for binding to *β-globin*. Under stoichiometric binding conditions, with 100 nM *β-globin* ( $>K_d$), the *β-globin* complex represented in the most compact band on the native gel in Supplementary Figure 2A have about 28 molecules of SRSF1-RBD bound to the RNA in consistence with a previous report (Saha et al., 2020). Similar result was also obtained with

*AdML* pre-mRNA (Supplementary Figure 2B), which binds about 10 molecules of SRSF1-RBD for an apparent saturation (Saha et al., 2020).

Then we examined if a lower level of SRSF1-RBD could form the stabilized ternary complex. Titration of 25 nM *β-globin* probe bound to 22X SRSF1-RBD with U1 snRNP indicated that about 60X U1 snRNP was required for generating the well-resolved ternary complexes (Supplementary Figure 2C). We showed later that excess U1 snRNP displaces majority of the SRSF1-RBD molecules from the pre-mRNA in a stoichiometry-dependent manner resolving the ternary complex on gel (Figure 3 and Supplementary Figure 3). Nonetheless, we next added 60X U1 snRNP to 25 nM *β-globin* probe bound to 2X, 6X, 10X, 14X, 18X, or 22X SRSF1-RBD molecules. At low SRSF1-RBD:*β-globin* ratio, smeary (i.e., kinetically unstable) complexes were formed, which migrated between the U1 snRNP-dependent ternary and the U1 snRNP-dependent binary complexes (Figure 2C, compares 7 and 13). It appears that roughly 10X SRSF1-RBD is required to form a stable ternary complex formation with U1 snRNP (Figure 2C, compare lanes 9 and 11). Thus, a certain SRSF1:*β-globin* ratio is required for SRSF1-RBD to stabilize U1 snRNP and 'lock' itself in a stable complex.

Taken together, these results suggest that multiple SRSF1 molecules could bind the pre-mRNAs in a cooperative mode and a minimum number of SRSF1 molecules are required for stabilization of U1 snRNP (Figure 2D).

**Multiple U1 snRNP particles displace majority of SRSF1-RBD molecules and specify the splicing substrate for U2AF65**

To further characterize the process of the ternary complex formation, we titrated 10 pM *β-globin* bound to about 25 molecules of SRSF1-RBD with U1 snRNP. Increasing concentrations of U1 snRNP gradually increased the mobility of the pre-mRNA complex, eventually forming the faster migrating complexes (marked with arrows in Figure 3A). Addition of anti-SRSF1 antibody to these complexes exhibited an exponential increase in the migration

rate of the super-shifted complexes with increasing concentrations of U1 snRNP up to 80 nM (lanes 3, 5, and 7 of Figure 3B). This suggests that SRSF1-RBD molecules are depleted from *β-globin* pre-mRNA with increasing concentrations of U1 snRNP. We also observed that U1 snRNP displaces SRSF1-RBD from short SRSF1-binding RNA such as 14-nt long *β-globin* 5′SS RNA (Supplementary Figure 1D) and 14-nt long *Ron* ESE (Supplementary Figure 3A). To more clearly demonstrate whether U1 snRNP displaces SRSF1-RBD from the full-length pre-mRNA, we wanted to identify a U1 snRNP variant that can displace SRSF1-RBD from the pre-mRNA without itself getting recruited, which would release the free pre-mRNA that can be easily detected by EMSA. Since different protein components of U1 snRNP play important roles in U1 snRNP binding to the RNA (Kondo et al., 2015), we assembled U1 snRNP with varied components to identify a variant that would not bind the pre-mRNA. We found that assembling U1 snRNP with Sm $B_{1\text{-}174}$ and without U1-A (U1 snRNP $B_{174}$ ΔA) served this purpose. We then titrated SRSF1-RBD-saturated *β-globin* with U1 snRNP $B_{174}$ and U1 snRNP $B_{174}$ ΔA (Figure 3C). While U1 snRNP $B_{174}$ formed complexes of similar migration pattern as observed with U1 snRNP (lanes 5, 6, 7), U1 snRNP $B_{174}$ ΔA at high concentrations displaced all SRSF1-RBD molecules from *β-globin* releasing free probe (lane 12).

Finally, we examined how titration with U1 snRNP impacts specific binding of U2AF65 that recognizes the PPT and dictates recognition of the 3′SS. Therefore, we added U2AF65 to the ternary complex of *β-globin* formed with 160 nM U1 snRNP (Figure 3D - lanes 4-6). Addition of U2AF65 upshifted all U1 snRNP-dependent bands to a single band (lane 4). Titration of the quaternary complex with increasing concentrations of U1 snRNP slightly downshifted the complex likely due to displacement of additional SRSF1-RBD molecules. Interestingly, with *β-globin* ΔPPT, we observed that with increasing concentrations of U1 snRNP, the quaternary complex becomes weaker likely at the expense of large aggregate-like dead-end/non-productive complexes that did not enter the gel (Hoskins et al., 2011; Jobbins et al., 2018) (Figure 3D -

lanes 10-12). We did not observe any detectable binding of 25 nM U2AF65 to protein-free *β-globin* pre-mRNA (Supplementary Figure 3B)

Overall, these results suggest that although a large number of SRSF1-RBD molecules could bind each *β-globin* molecule, only a sub-set of these molecules are essential for U1 snRNP recruitment and even a fewer are 'locked' in the early spliceosomal complexes (Figure 3E). We further found that additional copies of U1 snRNP could release the loosely bound SRSF1-RBD molecules and enhances specificity for U2AF65.

**Different SR proteins regulate the efficiency of U1 snRNP recruitment to *β-globin* through collaboration**

We wished to understand if SRSF1, SRSF2, and SRSF5, all of which promote the splicing *β-globin* (Screaton et al., 1995), could cooperate with each other to stably recruit U1 snRNP. For the assembly experiments, we used the RNA binding domains (RBD) of SRSF1 and SRSF5, which are known to be their functional variants (Cho et al., 2011b; Saha et al., 2021), and SRSF2-RBD in chimera with the fully phosphomimetic RS domain of SRSF1 (SRSF2-RE) (Chandler et al., 1997) (Supplementary Figure 4A). SRSF2-RE cooperatively bound *β-globin* at <3 nM concentration (Supplementary Figure 4B, lanes 1-9) while its affinity for its cognate ESE is weaker ($K_d$ = ~ 0.3 µM) (Zhang et al., 2015). The complex forming the most compact band on the native gel has about 10 molecules of SRSF2-RE bound to it (Saha et al., 2020). Similarly, at least a few copies of SRSF5-RBD bound to 10 pM *β-globin* below 6 nM concentration (Supplementary Figure 4B, lane 11). Binding of SRSF5-RBD to 25 nM unlabeled *β-globin* traced with radiolabeled *β-globin* indicates that the most compact band is formed with about 15X SRSF5-RBD (Supplementary Figure 4B, lane 23). For all three SR proteins, at high SR protein to pre-mRNA probe ratios, a portion of the probe was caught in the well likely due to formation larger complexes that could not enter the gel.

Then we tested if a suboptimal SRSF1-RBD:pre-mRNA ratio, which fails to stably recruit U1 snRNP, could be compensated for by other SR proteins. For this, we compared the U1 snRNP-dependent *β-globin* complexes formed with different stoichiometry of SRSF1-RBD and SRSF2-RE. It was evident that even 2-3X molar excess of SRSF1-RBD could stably recruit U1 snRNP if at least 7X SRSF2-RE was present (Figure 4A, lanes 8-13). To further test if the presence of a third SR protein helps stabilize the ternary complex at a different ratio, we added SRSF5-RBD to the binding reaction mixture. In the presence of SRSF5-RBD, SRSF2-RE could be lowered to obtain similar stable ternary complex (Supplementary Figure 4C). We then examined if similar phenomenon could be recapitulated with low concentrations of components in accord with the *in vivo* milieu. SRSF2-RE and SRSF5-RBD compensated for low concentrations of SRSF1-RBD for formation of the stable U1 snRNP-dependent complexes (Figure 4B). Finally, we observed that the *β-globin* complexes formed with U1 snRNP and different SR proteins were not disintegrated by 80,000X molar excess of free *β-globin* 5'SS RNA suggesting its stability (Supplementary Figure 4D).

We also tested if SRSF2-RE or SRSF5-RBD alone could stabilize U1 snRNP on *β-globin*. Titration of *β-globin* saturated with either SRSF2-RE or SRSF5-RBD with U1 snRNP gradually downshifted the *β-globin* complexes forming smears (Supplementary Figure 4E). A smear indicates kinetically unstable complexes unlike as observed with SRSF1-RBD. Thus, it appears that U1 snRNP displaces both SRSF2-RE and SRSF5-RBD, but they alone cannot stabilize U1 snRNP. We also repeated the same experiment with U2AF65 as in Figure 3D but instead of SRSF1-RBD, we used SRSF2-RE (Supplementary Figure 4F). The complexes formed with the *β-globin* variants did not enter the gel efficiently. Given the pore size of a 4% gel is >100 nm (Holmes and Stellwagen, 1991), we conclude that SRSF2-RE promoted the formation of predominantly the aggregate-like complexes even with WT *β-globin*. Then we examined the U2AF65-dependent complex formation with a combination of 10 nM SRSF2-RE and 20 nM

SRSF1-RBD (Figure 4C). The non-aggregated complex formed most with WT substrate (lane 2) and to a lesser extent with the ΔBS and Δ3′SS mutant (lanes 6 and 10). This complex was more weakened with ΔPPT and Δ5′SS. This observation correlates well with *in vivo* splicing efficiency of the *β-globin* mutants (Figure 4D). The Δ5′SS and ΔPPT mutants were completely defective in splicing while ΔBS spliced weakly into WT mRNA likely using a cryptic BS (Pineda and Bradley, 2018) while Δ3′SS mutant spliced weakly using a cryptic 3′SS 26-nt downstream of the authentic 3′SS (determined by Sanger sequencing).

We then examined if the used SR proteins could form pre-mRNA-independent complexes with U1 snRNP to examine if, upon displacement, they could sequester U1 snRNP into an SR:U1 complex. We previously observed detectable interactions of U1 snRNP with SRSF1-RBD but not SRSF5-RBD in a pre-mRNA-independent manner (Saha et al., 2021). Further GST pull-down assay indicates that U1 snRNP also forms pre-mRNA-independent complexes with SRSF2-RBD (Figure 4E).

Overall, these results suggest that the stoichiometry of individual SR proteins dictates U1 snRNP stabilization efficiency and U1 snRNP level required for SR protein displacement (Figure 4F).

**SR proteins collectively remodel the 3D structural scaffold of the pre-mRNA**

To examine if the differential impacts of varying stoichiometries of SR proteins are mediated by pre-mRNA structural remodeling, we carried out SHAPE analyses of *β-globin* bound to 15X SRSF1-RBD (15S1), 4X SRSF1-RBD (4S1), 4X SRSF1-RBD + 12X SRSF2-RE (S1S2), 4X SRSF1-RBD + 8X SRSF5-RBD (S1S5), 4X SRSF1-RBD + 6X SRSF2-RE + 4X SRSF5-RBD (S1S2S5), 15X SRSF2-RE (15S2), and 15X SRSF5-RBD (15S5). The 4S1, 15S2, and 15S5 provide suboptimal conditions for U1 snRNP stabilization while 15S1, S1S2, S1S5, and S1S2S5 are optimal for the same (Figure 4, Supplementary Figure 4).

We calculated fold-changes in the SHAPE reactivity of *β-globin* under all experimental conditions compared to *β-globin*+SR protein buffer. Figure 5A indicates the fold changes in SHAPE reactivity in the optimal samples (15S1, S1S2, S1S5, and S1S2S5). We then identified 5-nt long segments across the pre-mRNA within which at least one nucleotide exhibits >1000-fold enhancement of SHAPE reactivity and the central nucleotides of these segments were given a score of 1 (Supplementary Figure 5A). Where at least 1000-fold enhancement of SHAPE reactivity was not observed upon SR protein binding within two-nucleotides flanking the central nucleotide, it was given a score of zero. The four optimal samples exhibited a score of 1 at the same 105 sites (identified with blue circles in the plot in Figure 5B). We then compared the fold-changes in SHAPE reactivity in the sub-optimal samples and any of these 105 sites that did not exhibit the same property as the optimal samples were given a score of negative 1 and any site that did exhibit a similar property as the optimal samples were given a score of positive 1. All other sites were given a score of zero in the sub-optimal samples. 33 sites in Sample 4S1 (Figure 5B), 20 in 15S5 (Figure 5C), and 11 in 15S2 (Figure 5D) scored negative 1.

Since at least a small level of SRSF1-RBD was required for a stable binding of U1 snRNP to *β-globin* under all tested SR protein combinations, we asked if SRSF1-RBD exhibits any common difference from both SRSF2-RE and SRSF5-RBD. We found that binding of multicopy SRSF1-RBD to *β-globin* intron is more cooperative and efficient than that of SRSF2-RE as well as SRSF5-RBD (Supplementary Figure 5B, C). This suggests that SRSF1-RBD remodels the 3D structural scaffold of *β-globin* differently from the two other SR proteins.

In all, these results suggest that different combinations of SR proteins could collectively remodel the pre-mRNA to generate a pre-mRNA 3D scaffold with common structural features.

**RNP reactivity verifies SR protein-mediated pre-mRNA structural remodeling for the stable recruitment of U1 snRNP**

We examined the secondary structural features of *β-globin* pre-mRNA *in vivo* prior to binding of U1 snRNP. To accomplish this, DMS reactivity (DMS modifies single-stranded A and C residues not bound to proteins) was measured after blocking U1 snRNP binding using an antisense morpholino oligonucleotide complementary to the 5′ end of U1 snRNA (Saha et al., 2021) to observe the effect of binding of splicing regulatory RNA binding proteins. The analysis of the DMS-MaP-seq data exhibited a high signal-to-noise ratio with zero reactivity for all nucleotides in the untreated sample. Upon DMS-treatment, out of 179 A and C residues, only five nucleotides had a zero reactivity and only 35 were modified in less than 2% of the RNA molecules; the remaining nucleotides were more heavily modified (Supplementary Figure 6A). Given that an open base has a probability of 2-10% of being modified by DMS (Tomezsko et al., 2020), this reactivity profile of *β-globin* suggests that majority of the nucleotides are open *in vivo* prior to U1 snRNP recruitment.

Then we wanted to investigate the sites of stoichiometric binding of specific SR proteins and its effects on pre-mRNA structure and U1 snRNP recruitment *in vitro*. Purification of the pre-mRNA:SR:U1 snRNP complexes was required for its SHAPE and RNP-MaP analysis (see Methods), which could not be done efficiently for *β-globin*. Therefore, we tested the mechanisms with *AdML* by SHAPE and RNP-MaP, whose complexes could be efficiently purified as described before (Saha et al., 2021). U1 snRNP binding to *AdML* apparently saturated with SRSF1-RBD also involves displacement of SRSF1-RBD as was evident from an EMSA result we showed earlier (Saha et al., 2021). We compared fold changes in SHAPE reactivity of *AdML* upon binding of 5X SRSF1-RE (SHAPE_S1, the optimal SRSF1 stoichiometry (Saha et al., 2021)), 2X SRSF1-RE (SHAPE_S1_low – suboptimal SRSF1 stoichiometry), U1 snRNP (SHAPE_U1), and 5X SRSF1-RE + U1 snRNP (SHAPE_S1+U1). Fold-changes in SHAPE reactivity in SHAPE_S1 and SHAPE_S1_low compared to protein-free *AdML* were different (Figure 6A). SHAPE_S1 exhibits more efficient enhancement of nucleotide flexibility across the

pre-mRNA than SHAPE_S1_low, particularly immediately upstream of the 5′SS (marked with a rectangle above the plot). Nucleotide flexibility of this segment was lowered in SHAPE_S1+U1 (Figure 6B). This suggests that to promote the likely contact between U1 snRNP and this segment, SRSF1-mediated remodeling enhances flexibility of the segment immediately upstream of the 5′SS, a segment whose single-strandedness is important for recruitment of SR proteins (Saha et al., 2020) and U1 snRNP [(Saha et al., 2021), this work]. Finally, we compared the fold changes in SHAPE reactivity in SHAPE_U1 and SHAPE_S1+U1 (Figure 6C). SHAPE_S1+U1 exhibited a much higher level of flexibility-gain across the pre-mRNA compared to SHAPE_U1. These current observations is consistent with our previous report which conform to the fact that splicing competent pre-mRNAs are highly flexible *in vivo* (Saha et al., 2020) (Supplementary Figure 6A) correlating the SR protein-mediated pre-mRNA structural remodeling observed *in vitro* with the *in vivo* structural state of the pre-mRNA.

We then examined the protein contacts with the RNA by RNP-MaP-seq, which identifies the nucleotides within ~ 12 Å of the central carbon atom of an interacting lysine residue, under three different binding conditions: 5XSRSF1 alone (RNP_S1), U1 snRNP alone (RNP_U1), and 5X SRSF1 and U1 snRNP (RNP_S1+U1) (Figure 6D). RNP_S1 exhibits lysine contacts at 44 sites, RNP_S1+U1 at 40 sites, and RNP_U1 at 23 sites. While RNP_U1 exhibits RNP reactivity across the exon 1 and intron, the reactivity is extended into the entire length of the pre-mRNA including exon 2 in RNP_S1 and RNP_S1+U1. Although in RNP_S1+U1, it is not possible to discern if the cross-linking to a nucleotide is mediated by SRSF1 or a protein subunit of U1 snRNP, nucleotides at positions 38, 55, 85, 94, 106, 108, 122, 133, 149, 159, 179, 180, 181, and 189 exhibit RNP-reactivity only in RNP_S1+U1. The unique contacts in RNP_S1+U1 could be mediated by U1 snRNP components in the presence of SRSF1. This, in turn, suggests that SRSF1-indcued structural remodeling enhances U1 snRNP binding. Additionally, RNP_S1 exhibited RNP reactivity at 43 positions, 24 of which did not show a reactivity in RNP_ S1+U1,

which could be due to displacement of SRSF1 from the pre-mRNA upon addition of U1 snRNP. Moreover, 17 sites (25, 31, 41, 43, 60, 61, 62, 83, 87, 124, 148, 151, 170, 199, 202, 205, 216) exhibit RNP reactivity in both RNP_S1 and RNP_S1+U1 but the reactivity is higher in RNP_S1+U1 suggesting a higher stability of the latter. Similarly, 9 sites (40, 41, 54, 59, 62, 93, 97, 124, 139) exhibit higher RNP reactivity in RNP_S1+U1 than RNP_U1. Three sites – 41, 62, 124 – exhibit RNP reactivity in all three samples. This suggests that SRSF1-mediated U1 snRNP stabilization to *AdML* is a global phenomenon involving cooperation as well as competition between SRSF1 and U1 snRNP at different sites. The protein contacts are also mapped onto the SHAPE-derived secondary structure model of *AdML* pre-mRNA (Supplementary Figure 6B, C, D).

Finally, we compared the RNP-reactivity in RNP_S1+U1 and the fold-changes in SHAPE reactivity in SHAPE_S1 (Figure 6E). At least 1000-fold enhancement of SHAPE reactivity was observed in SHAPE_S1 within two-nucleotides flanking 12 sites that exhibited RNP reactivity in RNP_S1+U1 but not in RNP_S1 (40, 54, 55, 59, 85, 97, 122, 133, 139, 179, 180, 189). Among these 12 sites, 11 sites exhibit stable lysine contact in RNP_S1+U1 (except for 97) i.e., they do not exhibit >1000-fold increase in SHAPE reactivity in SHAPE_S1+U1. These results corroborate the fact that remodeling of *AdML* by SRSF1 enhances the flexibility of various nucleotides across the pre-mRNA for enhanced contacts with U1 snRNP.

Overall, SHAPE- and RNP-reactivity together provide a clearer mechanistic insight into the recruitment of U1 snRNP to *AdML* by remodeling of the pre-mRNA 3D structural scaffold by SRSF1.

## DISCUSSION

How the splicing machinery distinguishes between the highly degenerate splice signal sequences from background sequences have remained a mystery for decades. In the current study we examined the early steps of U1 snRNP recruitment to *β-globin* pre-mRNA by EMSA, SHAPE, RNP-MaP, and pull-down assay in the presence of different SR proteins which promote splicing of the test substrates. Our study provides the following insights regarding the regulation of splicing substrate definition by SR proteins.

**Multicopy cooperative binding of SR proteins:** A minimum number of copies of different pre-mRNA-specific SR protein molecules must bind both the exonic and the intronic motifs in appropriate combinations for an effective pre-mRNA structural remodeling. This is in contrast to the assembled early spliceosome where only one or two ESEs are occupied by a specific SR protein (Jobbins et al., 2018). Multiple molecules of individual SR proteins bind the pre-mRNA cooperatively, which compensates for the weak affinity of SR proteins for their cognate binding sites (about 0.5 µM for SRSF1 (Cho et al., 2011a) and 0.3 µM for SRSF2 (Zhang et al., 2015)). The cooperative binding occurs likely through a dynamic RNA-protein-protein-RNA interactions and/or cycles of binding and structural remodeling for binding to occluded/weaker sites. We previously observed that recruitment and assembly of these SR:pre-mRNA complexes is often regulated by single-stranded segments immediately upstream of the 5′SS, which regulate splicing across the transcriptome (Saha et al., 2020). Our current work suggests this segment could be critical for SR protein-mediated U1 snRNP recruitment. Nonetheless, the number of SR protein molecules that could bind each pre-mRNA molecule appeared to be much higher than what could be predicted by counting the occurrence of *in vitro* selected binding sites or identified ESEs due to cooperative nature of the binding. This does not mean that the SR protein binding is completely non-specific but rather that the repertoire of known strong and weak binding sites is incomplete (Anczuków et al., 2015).

**The diversity of pre-mRNA 3D structural scaffolds and its remodeling by SR proteins:** With a human adenovirus 2 pre-mRNA, we recently demonstrated that the 3D structural scaffold of a pre-mRNA could be an integral component of the mammalian splicing code beyond the splice signal sequences, splicing regulatory sequences, and secondary structure regulating the exposure/occlusion/proximity of the last two (Saha et al., 2021). We also showed that the functionality of this 3D structural scaffold in recruiting early spliceosomal components could be modulated by individual SR proteins. In the current work, we reaffirmed the role of the 3D structural scaffold with a mammalian splicing substrate human *HBA* IVS I, provided a glimpse of its diversity, and mechanistically correlated it with the widespread functionalities of the SR proteins (Pandit et al., 2013) and the single-stranded segments immediately upstream of the 5′SS (Saha et al., 2020).

We observed that U1 snRNP could recognize the remodeled 3D structural scaffold of *β-globin* comprised of the 5′ exon and the intron. During co-transcriptional recruitment of U1 snRNP, it is reported to remain bound to both the 5′SS and the elongating RNA polymerase II until after the transcription of the 3′ exon starts (Leader et al., 2021). Thus, recognition of the remodeled splicing-conducive structural scaffold of the 5′ exon and the intron by U1 snRNP could be a plausible widespread mechanism *in vivo*.

Mutation in both 5′ and 3′ splice signals do not strongly inhibit U1 snRNP binding to full-length *β-globin*. This suggests that *β-globin* has a loose global 3D structure that is not easily disrupted by splice signal mutations. Additionally, since binding of U1 snRNP to protein-free WT and Δ5′SS *β-globin* increases the structural similarity of the two variants within the intron and the 3′ exon, we surmise that *β-globin* is assembled into a compact 3D structural scaffold upon binding of U1 snRNP. In contrast, *AdML* pre-mRNA fails to recruit U1 snRNP upon mutation of any of the four major splice signals (Saha et al., 2021). This suggests that the 3D structural scaffold of *AdML* is compact and strong and is easily disrupted by mutation in any of the splice

signals. This also suggests that a mutation in the splice signal in pre-mRNAs like *AdML* would have inhibited spliceosome assembly at an earlier stage compared to pre-mRNAs like *β-globin*.

The question remains if the concept of a remodeled pre-mRNA 3D structural scaffold for recruitment of early spliceosomal components could be valid for introns that are several kilobases long. A recent report suggests that the kinetics of removal of the same intron in the same cells varies greatly and the introns could be progressively removed by recursive splicing across the transcriptome (Wan et al., 2021). The selection of the relatively short recursive introns could be mediated by recognizable remodeled 3D structural scaffolds. Additionally, the variable rate of intron removal could be due to RNA structure-mediated RNA-protein interaction dynamics (Çetin et al., 2020; Saldi et al., 2021).

**SR protein displacement by U1 snRNP:** The recruited U1 snRNP contacts nucleotides across the pre-mRNA, potentially substituting the SR protein molecules at certain positions. The remaining bound SR protein molecules are displaced from the pre-mRNA by additional stoichiometric level of U1 snRNP particles. The displacement likely occurs due to the ability of U1 snRNP components to weakly interact with various segments of the pre-mRNA. These RNA:U1 snRNP interactions Individually are not sufficiently strong to promote binding of the participating U1 snRNP particles to the pre-mRNA. However, these interactions of an excess of U1 snRNP particles could displace weakly bound SR proteins from across the pre-mRNA. Since U1 snRNP is present in vast excess as compared to other U snRNPs *in vivo* (Burge et al., 1999) and *in vitro* RNA-protein interactions are faithful representations of their *in vivo* functionalities (Taliaferro et al., 2016), this is likely an essential novel function of U1 snRNP for spliceosome assembly. Thus, U1 snRNP and SR proteins exhibit cooperative as well as competitive relationships for global stable engagement of U1 snRNP.

U2AF65 binding to splicing-defective splice signal mutants of *β-globin* led to formation of large aggregates in a U1 snRNP stoichiometry-dependent manner. We also observed that U1

snRNP binding to protein-free WT and Δ5′SS mutant of *β-globin* involved very similar global contact except for segments around the 5′SS. Based on these results, we hypothesize that U1 snRNP binding to a pre-mRNA remodeled by SR proteins and subsequent SR protein displacement (as opposed to U1 snRNP binding to a protein-free pre-mRNA) leads to a splicing-conducive 3D structure only for a *bona fide* splicing substrate. Given that U1 snRNP amply binds non-coding RNAs (Engreitz et al., 2014), this mechanism could be important for impeding wasteful spliceosome assembly on non-coding RNAs.

**U1 snRNP:SR protein interactions:** This dynamic process greatly enhances the affinity of U1 snRNP for the pre-mRNA. Currently, it is not entirely clear whether an interaction between the pre-mRNA-bound U1 snRNP and the pre-mRNA-bound SR protein is important for U1 snRNP stabilization at this stage. We have observed that SRSF1 that can strongly interact with U1 snRNP in a pre-mRNA-independent manner alone could stabilize U1 snRNP on WT *β-globin* (this work) and WT *AdML* (Saha et al., 2021). However, SRSF5-RBD but not SRSF1 stabilizes U1 snRNP on a 3′SS mutant of *AdML* that splices using a cryptic 3′SS 6-nt downstream of the authentic 3′SS with WT-like efficiency (Saha et al., 2021). Interestingly, SRSF5-RBD does not exhibit a detectable pre-mRNA-independent interaction with U1 snRNP. SRSF2-RBD, which can interact with U1 snRNP cannot stabilize *β-globin*:U1 snRNP interactions although *β-globin* contains SRSF2-specific ESEs and is responsive to SRSF2-dependent splicing activation (Schaal and Maniatis, 1999). This suggests that the effectiveness of an SR protein in activating splicing is also regulated by the splicing-conducive 3D structural scaffold.

Additionally, the pre-mRNA-independent interactions between certain SR proteins and U1 snRNP would lead to formation of binary complexes upon the displacement of the SR proteins by U1 snRNP. This could disrupt the functions of free U1 snRNP particles *in vivo* upon elevation of expression of these SR proteins. Further research is required for functional analysis of this phenomenon.

**Potential mechanism of splicing inhibition by SR proteins:** Since there is a stoichiometric relationship between the bound SR protein molecules and U1 snRNP particles required to displace them, an increase in SR protein levels without an increase in U1 snRNP level could occlude the pre-mRNA and repress splicing (Akerman et al., 2015; Busch and Hertel, 2012).

**Insights into the mechanisms of splicing regulation by other SRps:** Our data explain why the position of individual short degenerate binding motifs of SR proteins individually cannot explain the splicing outcome. It also explains why cooperative binding of these proteins involving homomeric interactions to repeating or 'multivalent' motifs is important. Inability of individual binding motifs to explain the splicing outcome and the cooperative binding are also reported for many other SRps (Banani et al., 2017; Cereda et al., 2014; Chou et al., 2000; Kato and McKnight, 2018; Li et al., 2012; Ule and Blencowe, 2019; Ule et al., 2003). Thus, we infer that the mechanistic knowledge presented in the current work could also be extended to a variety of other SRps.

**Translational potential of these findings:** Expression of the serine-arginine-rich (SR) proteins are tightly regulated in cells and changes in their expression, particularly overexpression, has been shown to transform cells through regulation of alternative splicing of tumor suppressor genes (Das and Krainer, 2014; Karni et al., 2007; Kędzierska and Piekiełko-Witkowska, 2017; Kim et al., 2016; Luo et al., 2017; Sokół et al., 2018; Sun et al., 2010; Wu et al., 2010; Yang et al., 2018). The changes in splicing caused by overexpression of individual SR proteins are often explained by concentration-dependent changes in SR protein binding to the RNA motifs. Therefore, currently the therapeutic intervention involves inactivation of the active SR proteins. Since SR proteins are activated upon phosphorylation of the RS domain and consequent exposure the RNA binding domain (RBD) containing the RRMs (Aubol et al., 2016; Cho et al., 2011b; Tacke et al., 1997), most of the inhibitors to counteract SR overexpression in tumors target the kinases that phosphorylate them (Bates et al., 2017). However, the targets of these

kinases include a wide variety of proteins and not limited to one or two specific SR proteins (Gou et al., 2020). Our work provides evidence for intermolecular interactions of SR proteins (Cléry et al., 2021) to be important for recruitment of early spliceosomal components, highlighting potential targets for SR protein-specific therapeutic intervention.

Overall, this work provides a greater understanding of concentration-dependent changes in functions of individual SR proteins and potentially other splicing regulatory proteins in splicing activation as well as repression.

**AUTHOR CONTRIBUTION**

KS designed and performed the experiments, analyzed the data, and wrote the paper. GG designed some experiments, wrote the paper, and supervised the project.

**ACKNOWLEDGEMENT**

The authors would like to acknowledge Dr Simpson Joseph for critical reading of the manuscript and Mike Minh Fernandez for some early work.

**FUNDING**

National Institute of General Medical Sciences [R01GM085490].

## REFERENCES


Akerman, M., Fregoso, O.I., Das, S., Ruse, C., Jensen, M.A., Pappin, D.J., Zhang, M.Q., and Krainer, A.R. (2015). Differential connectivity of splicing activators and repressors to the human spliceosome. Genome Biology *16*, 1-18.

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N.H., Raimer, A., Sun, S., Jensen, M.A., Hua, Y., *et al.* (2015). SRSF1-Regulated Alternative Splicing in Breast Cancer. Mol Cell *60*, 105-117.

Aubol, B.E., Wu, G., Keshwani, M.M., Movassat, M., Fattet, L., Hertel, K.J., Fu, X.D., and Adams, J.A. (2016). Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. Mol Cell *63*, 218-228.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nature reviews Molecular cell biology *18*, 285-298.

Baralle, M., and Baralle, F.E. (2018). The splicing code. Bio Systems *164*, 39-48.

Bates, D.O., Morris, J.C., Oltean, S., and Donaldson, L.F. (2017). Pharmacology of Modulators of Alternative Splicing. Pharmacological reviews *69*, 63-79.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the spliceosomes. Cold Spring Harbor Monograph Series *37*, 525-560.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdiscip Rev RNA *3*.

Cereda, M., Pozzoli, U., Rot, G., Juvan, P., Schweitzer, A., Clark, T., and Ule, J. (2014). RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. Genome Biol *15*, R20.

Çetin, B., Song, G.J., and O'Leary, S.E. (2020). Heterogeneous Dynamics of Protein-RNA Interactions across Transcriptome-Derived Messenger RNA Populations. Journal of the American Chemical Society *142*, 21249-21253.

Chandler, S.D., Mayeda, A., Yeakley, J.M., Krainer, A.R., and Fu, X.-D. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. Proc Natl Acad Sci USA *94*, 3596-3601.

Cho, S., Hoang, A., Chakrabarti, S., Huynh, N., Huang, D.B., and Ghosh, G. (2011a). The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs. Nucleic Acids Res *39*, 9413-9421.

Cho, S., Hoang, A., Sinha, R., Zhong, X.-Y., Fu, X.-D., Krainer, A.R., and Ghosh, G. (2011b). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc Natl Acad Sci USA *108*, 8233-8238.

Chou, M.Y., Underwood, J.G., Nikolic, J., Luu, M.H., and Black, D.L. (2000). Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Mol Cell *5*, 949-957.

Cléry, A., Krepl, M., Nguyen, C.K.X., Moursy, A., Jorjani, H., Katsantoni, M., Okoniewski, M., Mittal, N., Zavolan, M., Sponer, J., *et al.* (2021). Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon7 splicing. Nat Commun *12*, 428.

Das, S., and Krainer, A.R. (2014). Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. Molecular cancer research : MCR *12*, 1195-1204.

Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell *159*, 188-199.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Gou, L.T., Lim, D.H., Ma, W., Aubol, B.E., Hao, Y., Wang, X., Zhao, J., Liang, Z., Shao, C., Zhang, X.*, et al.* (2020). Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation. Cell *180*, 1212-1227.e1214.
Guth, S., Tange, T.Ø., Kellenberger, E., and Valcárcel, J. (2001). Dual function for U2AF35 in AG-dependent pre-mRNA splicing. Mol Cell Biol *21*, 7673-7681.
Holmes, D.L., and Stellwagen, N.C. (1991). Estimation of polyacrylamide gel pore size from Ferguson plots of normal and anomalously migrating DNA fragments. I. Gels containing 3% N,N'-methylenebisacrylamide. Electrophoresis *12*, 253-263.
Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J., and Moore, M.J. (2011). Ordered and dynamic assembly of single spliceosomes. Science *331*, 1289–1295.
Jobbins, A.M., Reichenbach, L.F., Lucas, C.M., Hudson, A.J., Burley, G.A., and Eperon, I.C. (2018). The mechanisms of a mammalian splicing enhancer. Nucleic Acids Res *46*, 2145-2158.
Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol *14*, 185-193.
Kato, M., and McKnight, S.L. (2018). A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. Annu Rev Biochem *87*, 351-390.
Kędzierska, H., and Piekiełko-Witkowska, A. (2017). Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. Cancer letters *396*, 53-65.
Kim, H.R., Lee, G.O., Choi, K.H., Kim, D.K., Ryu, J.S., Hwang, K.E., Na, K.J., Choi, C., Kuh, J.H., Chung, M.J.*, et al.* (2016). SRSF5: a novel marker for small-cell lung cancer and pleural metastatic cancer. Lung cancer (Amsterdam, Netherlands) *99*, 57-65.
Kondo, Y., Oubridge, C., Roon, A.M.v., and Nagai, K. (2015). Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. Elife *4*, DOI: 10.7554/eLife.04986.
Leader, Y., Lev Maor, G., Sorek, M., Shayevitch, R., Hussein, M., Hameiri, O., Tammer, L., Zonszain, J., Keydar, I., Hollander, D.*, et al.* (2021). The upstream 5' splice site remains associated to the transcription machinery during intron synthesis. Nat Commun *12*, 4545.
Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F.*, et al.* (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336-340.
Luo, C., Cheng, Y., Liu, Y., Chen, L., Liu, L., Wei, N., Xie, Z., Wu, W., and Feng, Y. (2017). SRSF2 Regulates Alternative Splicing to Drive Hepatocellular Carcinoma Development. Cancer research *77*, 1168-1178.
Mathews, D.H. (2014). RNA secondary structure analysis using RNAstructure. Curr Protoc Bioinformatics *46*, 12.16.11-12.16.25.
Oh, J.M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I.*, et al.* (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. Nat Struct Mol Biol *24*, 993-999.
Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Jr, M.A., and Fu, X.-D. (2013). Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Mol Cell *50*, 223-235.
Pineda, J.M.B., and Bradley, R.K. (2018). Most human introns are recognized via multiple and tissue-specific branchpoints. Genes Dev *32*, 577-591.
Roca, X., Krainer, A.R., and Eperon, I.C. (2013). Pick one, but be quick: 5′ splice sites and the problems of too many choices. Genes Dev *27*, 129-144

Saha, K., England, W., Fernandez, M.M., Biswas, T., Spitale, R.C., and Ghosh, G. (2020). Structural disruption of exonic stem–loops immediately upstream of the intron regulates mammalian splicing Nucleic Acids Res *48*, 6294-6309.
Saha, K., Fernandez, M.M., Biswas, T., Joseph, S., and Ghosh, G. (2021). Discovery of a pre-mRNA structural scaffold as a contributor to the mammalian splicing code. Nucleic Acids Res *49*, 7103-7121.
Saldi, T., Riemondy, K., Erickson, B., and Bentley, D.L. (2021). Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. Mol Cell *81*, 1789-1801.e1785.
Schaal, T.D., and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol *19*, 261-273.
Screaton, G.R., Cáceres, J.F., Mayeda, A., Bell, M.V., Plebanski, M., Jackson, D.G., Bell, J.I., and Krainer, A.R. (1995). Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. EMBO J *14*, 4336-4349.
Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H.*, et al.* (2014). Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat Struct Mol Biol *21*, 997-1005.
Sokół, E., Kędzierska, H., Czubaty, A., Rybicka, B., Rodzik, K., Tański, Z., Bogusławska, J., and Piekiełko-Witkowska, A. (2018). microRNA-mediated regulation of splicing factors SRSF1, SRSF2 and hnRNP A1 in context of their alternatively spliced 3'UTRs. Experimental cell research *363*, 208-217.
Sun, S., Zhang, Z., Sinha, R., Karni, R., and Krainer, A.R. (2010). SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. Nat Struct Mol Biol *17*, 306-312.
Tacke, R., Chen, Y., and Manley, J.L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc Natl Acad Sci USA *94*, 1148-1153.
Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., Bazile, C., and Burge, C.B. (2016). RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. Mol Cell *64*, 294-306.
Tomezsko, P.J., Corbin, V.D.A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M.D., McIntosh, L., Papenfuss, A.T., Emery, A.*, et al.* (2020). Determination of RNA structural diversity and its role in HIV-1 RNA splicing. Nature *582*, 438-442.
Ule, J., and Blencowe, B.J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol Cell *76*, 329-345.
Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.
Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M.*, et al.* (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell *184*, 2878-2895.e2820.
Weidmann, C.A., Mustoe, A.M., Jariwala, P.B., Calabrese, J.M., and Weeks, K.M. (2021). Analysis of RNA-protein networks with RNP-MaP defines functional hubs on RNA. Nat Biotechnol *39*, 347-356.
Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*, pii: a003707.
Wu, H., Sun, S., Tu, K., Gao, Y., Xie, B., Krainer, A.R., and Zhu, J. (2010). A splicing-independent function of SF2/ASF in microRNA processing. Mol Cell *38*, 67-77.
Yang, S., Jia, R., and Bian, Z. (2018). SRSF5 functions as a novel oncogenic splicing factor and is upregulated by oncogene SRSF3 in oral squamous cell carcinoma. Biochimica et biophysica acta Molecular cell research *1865*, 1161-1172.

Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc Natl Acad Sci U S A *112*, E4726-4734.
Zhou, Z., and Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191-207.

## METHODS

### Cloning, protein expression, and *in vitro* reconstitution and purification of U1 snRNP

Different subunits of U1 snRNP and the SR proteins were expressed and purified as described before (Saha et al., 2021). Unless otherwise indicated, U1 snRNP was reconstituted and purified with all full-length proteins except U1-70k, which was truncated (1-215 amino acids) as described before (Saha et al., 2021)

### RNA constructs

*β-globin* and *AdML* RNA were as described before (Saha et al., 2020). 14-nt long *Ron* ESE (5′-UGGCGGAGGAAGCA-3′) sequence and *β-globin* 5′SS (5′-GGGCAGGUUGGUAU-3′) were described before (Cho et al., 2011a; Saha et al., 2021)..

### Electrophoretic mobility shift assay (EMSA)

EMSA with full-length pre-mRNA and 14-nt long RNAs were carried out with 4% (89:1) and 6% (80:1) native polyacrylamide gels, respectively, as described before (Saha et al., 2021). Antibody super-shift was carried out with anti-SNRPC antibody (Abcam, ab157116) and anti-SRSF1 antibody (Life Technologies, 32-4500). After the formation of complexes as described above, 0.25 µg antibody was added to the 15 µl reaction, incubated at 30 $^{o}$C for 5 min and resolved on polyacrylamide gels.

### *In vitro* selective 2′ hydroxyl acylation followed by primer extension by mutational profiling (SHAPE-MaP)

All SHAPE assays were carried out with samples treated with 4 mM N-Methylisatoic anhydride (NMIA, Millipore-Sigma). For SHAPE assay with *β-globin*, first 800 nM RNA was denatured at 95 °C for 3 min in 50 mM NaCl and 5 mM EDTA pH 8.0 and snap-cooled on ice. After neutralizing EDTA with $MgCl_2$, 100 nM (final concentration) RNA was mixed in 1 ml reaction containing 20 mM HEPES pH 7.5, 250 mM NaCl, 2 mM $MgCl_2$, 0.5 M urea, 0.3% polyvinyl

alcohol (P-8136, Sigma), and 1 mM 2-mercaptoethanol along with the indicated level of SR proteins or SR protein storage buffer and incubated at 30 °C for 5 min. Then the solution was transferred to a fresh tube containing 33.3 µl 120 mM NMIA or DMSO (vehicle), mixed immediately, and incubated at 30 °C for 15 min and then at room temperature for 15 min. Thereafter, 9M urea is added to the reaction up to the final concentration of 2M and then the reaction was treated with 0.5 mg/ml (final concentration) proteinase K (NEB) at 37 °C for 30 min. Next, RNA was ethanol-precipitated with 300 mM sodium acetate pH 5.2, the pellet was dissolved in 50 µl water, and then the RNA was purified by Monarch RNA cleanup kit (New England Biolabs).

For U1 snRNP binding to *β-globin*+SRSF1-RBD complex, a very high concentration of U1 snRNP had to be added to the reaction. Raising U1 snRNP to such a concentration was difficult. Additionally, the presence of such a high concentration of U1 snRNP in the reaction might affect overall SHAPE reactivity and hence, it was necessary to purify the complex. However, purifying this complex by amylose pull-down was inefficient due to weaker binding of U1 snRNP to *β-globin* compared to *AdML*. Therefore, it was not possible to carry out SHAPE reaction on the ternary complex of *β-globin*. For SHAPE assay with *AdML*, 500 nM refolded RNA with 3X MS2-binding loops at the 3′ end was first bound to 1500 nM MBP-MS2 protein for 15 min at 30 °C. Then the RNA was added to 250 µl reaction mixture with the same buffer composition as with *β-globin* at a final concentration of 100 nM, was mixed with 300 nM U1 snRNP or 500 nM SRSF1 (full-length SRSF1 with all serine residues in its RS domain replaced with glutamate) or both, and was incubated at 30 °C for 5 min. Then complexes were bound to 15 µl amylose agarose (NEB) by rotating the tube at 4 °C for 30 min. The resin was washed with chilled buffer (20 mM HEPES pH 7.5, 250 mM NaCl, 2 mM $MgCl_2$, 1 mM 2-mercaptoethanol) and then eluted in 500 µl of this buffer containing 20 mM maltose for rotating the tube at 4 °C for 2 hours. Then 250 µl solution was transferred to 8.33 µl 120 mM NMIA or

equal volume of DMSO (vehicle), mixed gently, and incubated at 16 °C for 1 hour. The solution was then treated with 0.5 mg/ml proteinase K in the presence of added 2 M urea at 37 °C for 30 min and the RNA was ethanol-precipitated followed by RNA cleanup with Monarch RNA cleanup kit. Preparation of denatured control, library preparation, deep sequencing, and data processing were carried out as described before (Saha et al., 2021). SHAPE-derived secondary structure models were obtained using RNAstructure software (Mathews, 2014).

SHAPE differentials were calculated by determining fold-changes in SHAPE reactivity of a pre-mRNA upon binding to a splicing factor compared to the protein-free pre-mRNA. Chemical reactions of flexible nucleotides with structure probing reagents are sensitive to local chemical environment and hence, fold-changes likely reflect a more meaningful information regarding structural alteration of the RNA. For calculating fold-changes, all nucleotides with zero or negative SHAPE reactivity values were considered to be 0.0001.

**Detection of RNA-protein interactions by RNP-MaP**

It was not possible to carry out RNP-MaP on *β-globin*+SRSF1-RBD+U1 snRNP complex because it was necessary to purify the complexes by amylose pull-down before carrying out RNP-MaP (also see the method for SHAPE). This is because the complexes are assembled in the presence of 0.5 M urea and U1 snRNP is stored in a buffer containing arginine-HCl and glutamate-KOH, which would interfere with the SDA crosslinking reagent used for RNP-MaP since SDA reacts with primary amines. Thus, purified complexes were needed for RNP-MaP analysis. Since we could only purify *AdML* complexes efficiently by amylose pull-down, we carried out RNP-MaP with *AdML* only. RNP-MaP for detection of protein interaction sites on an RNA was carried out as described before (Weidmann et al., 2021). *AdML* complexes were assembled and purified as in SHAPE. The complex was eluted by rotating the tubes for 2 hours at 4 °C in elution buffer containing 4 mM SDA (NHS-diazirine, succinimidyl 4,4′-azipentanoate,

Thermo Fisher Scientific, diluted from 100 mM stock in DMSO) or equivalent concentration of DMSO. Then the resin was removed by gentle centrifugation, the liquid was transferred to a 24-well plate on ice and was exposed to 3 J/cm$^2$ of 365 nm light in UVP 1000CL (Analytik Jena) crosslinker. The proteins were then digested in the presence of 1.5% SDS, 20 mM EDTA, and 0.5 mg/ml proteinase K for 2 hours at 37 °C and the RNA was purified by phenol:chloroform extraction and ethanol precipitation. The RNA pellet was dissolved in 50 µl water and was further purified by Monarch RNA cleanup kit. The reverse transcription reaction was carried out as described before (Weidmann et al., 2021). The library preparation and deep sequencing were carried out as in SHAPE. Analyses of RNP sites and RNP reactivities were carried out as described before (Weidmann et al., 2021).

***In vivo* DMS-MaP-seq**

*HeLa* cells grown in six-well plate were co-transfected with 5 nmol of U1 snRNA antisense morpholino oligonucleotide (Oh et al., 2017) and 1 µg *β-globin* minigene plasmid and were incubated for 18 hours. Cells were then treated with 2% dimethyl sulfate (DMS, Millipore Sigma) in 2 ml incomplete DMEM medium and incubated at 37 °C for 5 min. Then the medium was aspired out and cells were washed twice with 30% 2-mercaptoethanol in PBS. Total RNA was then extracted with Trizol (Thermo Fisher Scientific) with DNase treatment following manufacturer's instructions. Reverse transcription was carried out with TGIRT III reverse transcriptase (Ingex) as described previously (Tomezsko et al., 2020). Due to modification of the RNA, which are substituted by mutations during reverse transcription, the primers used for sequencing library preparation lost significant complementarity to the cDNA. The weak annealing prevented amplification of the 450-nt long *β-globin* pre-mRNA with end-to-end amplification. Therefore, we split the cDNA into two adapted fragments for deep sequencing - 1-228 nt and 109-442 nt. DMS reactivity was obtained using DREEM (Tomezsko et al., 2020). We used data for 1-180 nt from the 5′ segment and 181-442 nt from the 3′ segment.

## FIGURE LEGENDS

**Figure 1. SRSF1-RBD stabilizes the interactions between the global 3D structural scaffold of *β-globin* and U1 snRNP**

(A) *β-globin* forms ternary complexes with SRSF1-RBD and U1 snRNP, which migrate primarily as two major bands (marked with black arrows in lane 4); U1 snRNP also binds free *β-globin* (lanes 6-8); the ternary complex is super-shifted with αSRSF1 (lane 9, marked with an orange arrow) and αU1-C (lanes 18-19); slightly more smeary complexes of similar migration patterns were formed with *β-globin* Δ5′SS (lanes 10-17); red script indicates radiolabeled components; the position of the free probe is indicated with an exon-intron-exon schematic. (B) *βg-ΔEx2* (*β-globin* lacking the 3′ exon) formed U1 snRNP-dependent complexes in the presence of 60 nM SRSF1-RBD (the complexes are marked with arrows in lane 3); these U1 snRNP-dependent complexes were significantly weakened for *βg-ΔEx2* with 5′SS mutations (*βg-ΔEx2* Δ5′SS) and *βg-ΔEx2* with hybridization-mutation immediately upstream of the 5′SS (*βg-ΔEx2* EH3+4) (the position of the weakened complexes are marked with curved brackets in lanes 9 & 15); U1 snRNP shows no well-defined interaction with the free RNAs (lanes 5, 6, 11, 12, 17, 18); the position of the free probe is indicated with an exon-intron schematic. (C & D) Overlaid SHAPE reactivity of protein-free *β-globin* WT and its 5′SS mutant (Δ5) (C) and U1 snRNP-bound *β-globin* WT and the *β-globin* 5′SS-mutant (D); the segments showing a high correlation between moving averages (period = 6) of SHAPE reactivity are marked in each plot and the corresponding r and *p* values are indicated; exon-intron boundaries are demarcated with dotted vertical lines. (E) The ternary complexes assembled with U1 snRNP mixed with short *β-globin* 5′SS RNA (*βg*-5′SS) showing formation of compact complexes. (F) Mixing of *β-globin*:U1 snRNP binary complexes with short *β-globin* 5′SS RNA disintegrates the binary complexes releasing free probe. (G) Take-home message of this figure.

**Figure 2. Cooperative binding of SRSF1-RBD to *β-globin* is important for stabilization of U1 snRNP**

(A) EMSA showing gradual upshift of radiolabeled *β-globin* complexes upon titration with SRSF1-RBD. (B) EMSA showing formation of weak complexes with radiolabeled 14-nt long *Ron* ESE (a binding site for SRSF1). (C) Varied U1 snRNP stabilization efficiency of 22X, 18X, 14X, 10X, 6X, and 2X SRSF1-RBD bound to 25 nM *β-globin* (compare lanes 3, 5, 7, 9, 11, 13); smeary complexes are kinetically unstable complexes. (D) Take-home message of this figure.

**Figure 3. Multiple U1 snRNP particles are required for displacement of majority of SRSF1-RBD molecules and specifying the splicing substrate for U2AF65**

(A) Titration of SRSF1-RBD-saturated *β-globin* with U1 snRNP displaces SRSF1-RBD molecules forming U1 snRNP-dependent complexes (marked with arrows) (lanes 2-13). (B) Super-shift of complexes formed by titrating SRSF1-RBD-saturated *β-globin* with U1 snRNP with αSRSF1 showing exponential increase in the migration rate of the super-shifted complexes up to 80 nM U1 snRNP. (C) U1 snRNP assembled with Sm $B_{1-174}$ and without U1-A displaces SRSF1-RBD molecules from *β-globin* releasing free pre-mRNA (lanes 2, 8-12). (D) Titration of the quaternary complexes of *β-globin* WT and *β-globin* ΔPPT assembled with SRSF1-RBD, U1 snRNP, and U2AF65 with additional U1 snRNP distinctively changes their migration forming aggregate-like mutant complexes caught in the well; the ratios of the band-intensity of the free pre-mRNA and the quaternary complexes (marked with yellow arrows) are indicated in red script below the image; the mutated sequence in ΔPPT is shown. (E) Take-home message of this figure.

**Figure 4. SR protein:*β-globin* ratio dictates U1 snRNP stabilization efficiency and U1 snRNP level requirement for SR protein displacement**

(A) U1 snRNP stabilization efficiency on 25 nM *β-globin* with 10X, 8X, 6X, 3X, and 2X SRSF1-RBD (lanes 4-7), with 6X, 3X, and 2X SRSF1-RBD and 10X SRSF2-RE (lanes 8-10), with 6X,

3X, and 2X SRSF1-RBD and 7X SRSF2-RE (lanes 11-13), with 6X, 3X, and 2X SRSF1-RBD and 3X SRSF2-RE (lanes 14-16), with 6X and 3X SRSF1-RBD and 2X SRSF2-RE (lanes 17-18). (B) EMSA showing formation of stable U1 snRNP-dependent *β-globin* complexes with optimal level of SRSF1-RBD alone (lane 2), with suboptimal level of SRSF1-RBD compensated for by low levels of SRSF2-RE and/or SRSF5-RBD (compare lanes 3 with 4-9). (C) Efficiency of assembly of early spliceosomal complexes with U1 snRNP, U2AF65, SRSF1-RBD, and SRSF2-RE is highest for WT *β-globin*, slightly less for splicing-competent splice signal mutants (ΔBS and Δ3′SS), and least for splicing-defective splice signal mutants (Δ5′SS and ΔPPT). (D) Transfection-based splicing efficiency of *β-globin* WT and splice signal mutants; '*' indicates an mRNA produced using a cryptic 3′SS 26-nt downstream of the authentic 3′SS. (E) GST-pull down assay showing pre-mRNA-independent interactions between U1 snRNP and SRSF1-RBD as well as SRSF2-RBD. (F) Take-home message of this figure.

**Figure 5. Global structural remodeling of *β-globin* with different SR protein combinations**

(A) Fold changes in SHAPE reactivity upon binding of 15X SRSF1-RBD (15S1), 4X SRSF1-RBD + 12X SRSF2-RE (S1S2), 4X SRSF1-RBD + 8X SRSF5-RBD (S1S5), and 4X SRSF1-RBD + 6X SRSF2-RE + 4X SRSF5-RBD (S1S2S5) to *β-globin* shown for 20-150-nt (top), 150-300-nt (middle), 300-420-nt (bottom). (B, C, D) The central nucleotide of 5-nt long segments across the pre-mRNA within which at least one nucleotide exhibits >1000-fold enhancement of SHAPE reactivity was given a score of 1 in the optimal samples (15S1, S1S2, S1S5, S1S2S5) (Supplementary Figure 5A); the common 105 nucleotides scored 1 in all optimal samples are indicated with blue circles in B, C, D; in the sub-optimal samples (4S1, 15S2, 15S5), any of these 105 sites that did not exhibit the same property as the optimal samples were given a score of negative 1 and any site that did exhibit a similar property were given a score of positive 1; all other sites were given a score of zero in the sub-optimal samples; 33 sites in sample 4S1

(orange dotted line in B), 20 in 15S5 (grey dotted line in C), and 11 in 15S2 (yellow dotted line in D) scored negative 1.

**Figure 6. Global structural remodeling and protein contacts involved in U1 snRNP recruitment to *AdML***

(A) Fold changes in SHAPE reactivity of *AdML* upon binding of 5X SRSF1-RE (optimal sample, blue line) or 2X SRSF1-RE (sub-optimal sample, orange line); $\log_{10}$ of fold changes in SHAPE reactivity are plotted along *y*-axis showing both increase (i.e., $y > 1$) or decrease (i.e., $y < 1$); the segment immediately upstream of the 5′SS is marked with a rectangle above the plot. (B) Fold changes in SHAPE reactivity upon binding of 5X SRSF1-RE (blue line) or 5X SRSF1-RE + U1 snRNP (yellow line); fold changes in SHAPE reactivity are plotted along *y*-axis showing only the increase (i.e., $y > 1$). (C) Fold changes in SHAPE reactivity upon binding of U1 snRNP alone (grey line) or 5X SRSF1-RE + U1 snRNP (yellow line). (D) RNP reactivity for *AdML* bound to U1 snRNP (blue line), SRSF1-RE (orange line), SRSF1-RE + U1 snRNP (grey dotted line); *in vitro* selected SRSF1-binding sites are demarcated below the plot and exon-intron boundaries are indicated with dotted vertical lines. (E) Fold-changes in SHAPE reactivity of 5XSRSF1-RE+*AdML* sample (blue line) plotted along primary *y*-axis and RNP-reactivity of 5XSRSF1+U1 snRNP+*AdML* sample (orange line) plotted along secondary *y*-axis.

**Figure 7. Proposed model of U1 snRNP and U2AF65 recruitment to the full-length pre-mRNA**

Irrespective of whether the global 3D structural scaffold of the pre-mRNA is capable of recruiting U1 snRNP, different pre-mRNA-specific-SRps like SR proteins bind the pre-mRNA cooperatively and structurally remodel the pre-mRNA making it more recognizable by U1 snRNP. Multiple copies of U1 snRNP particles displace the majority of the RBP molecules from the pre-mRNA with concomitant recruitment of U1 snRNP to the remodeled pre-mRNA.

Interactions of U1 snRNP with the pre-mRNA as well as the RBP are indicated with stars - a solid yellow star indicates top surface and a hollow star bottom surface. U1 snRNP-mediated SRp displacement makes the complex a more suitable substrate for U2AF65.

**SUPPLEMENTARY FIGURE LEGENDS**

**Supplementary Figure 1. SHAPE-derived secondary structure models of *β-globin* variants**

(A & B) SHAPE-derived secondary structure models of *β-globin* WT (A) and its 5′SS mutant (B); the sequences mutated in *β-globin* Δ5′SS are shown; nucleotides are color-coded according to their SHAPE reactivity as indicated in the associated table. (C) Transfection-based splicing assay showing splicing efficiency of WT *β-globin* and its 5′SS mutant. (D) Displacement of loosely bound SRSF1-RBD molecules from the ternary complex with a short SRSF1-RBD binding RNA (*Ron* ESE) improves resolution of the ternary complexes on gel. (E) Short *β-globin* 5′SS RNA behaves like an ESE for binding of SRSF1-RBD (lane 2), displacement of SRSF1-RBD from *β-globin* 5′SS by U1 snRNP (lanes 3-8), and inability of U1 snRNP of strong interaction with short *β-globin* 5′SS RNA (lanes 9-14).

**Supplementary Figure 2. Multicopy SRSF1-RBD binding and U1 snRNP recruitment to *β-globin***

(A) Titration of 100 nM *β-globin* traced with radiolabeled *β-globin* with SRSF1-RBD showing binding of multiple copies of SRSF1-RBD to each *β-globin* molecule. (B) Cooperative binding of SRSF1-RBD to 10 pM radiolabeled *AdML* pre-mRNA; radiolabeled components are shown in red script. (C) Titration of *β-globin* bound to 25X SRSF1-RBD with U1 snRNP showing about 60X U1 snRNP is required to resolve the U1 snRNP-dependent complexes on gel.

**Supplementary Figure 3. Displacement of SRSF1-RBD from RNA and inefficient U2AF65 binding to protein-free *β-globin***

(A) Titration of *Ron* ESE bound to SRSF1-RBD with U1 snRNP showing displacement of SRSF1-RBD molecule from the RNA releasing free probe. (B) U2AF65, at 25 nM concentration, does not exhibit a detectable binding to full-length protein-free *β-globin*.

**Supplementary Figure 4. Importance of appropriate SR protein combinations in appropriate stoichiometry for early spliceosome assembly**

(A) SR protein constructs used in this study: RNA binding domain (RBD) of SRSF1 (1-203 a.a.) (SRSF1-RBD) or full-length SRSF1 with fully phosphorylated mimetic RS domain of SRSF1 (204-246 a.a.) with all serine residues replaced with glutamate (SRSF1-RE), RBD of SRSF2 (1-127 a.a.) in chimera with fully phosphorylated mimetic RS domain of SRSF1 (204-246 a.a.), RBD of SRSF5 (1-184 a.a.). (B) EMSA showing titration of 10 pM *β-globin* with low concentrations of SRSF2-RE (lanes 1-9) and SRSF5-RBD (lanes 10-18) and 25 nM *β-globin* with high concentrations of SRSF5-RBD (lanes 19-25). (C) EMSA showing U1 snRNP stabilization on 25 nM *β-globin* with 15X SRSF1-RBD, 4X SRSF1-RBD + 12X SRSF2-RE, 4X SRSF1-RBD + 8X SRSF5-RBD, 4X SRSF1-RBD + 6X SRSF2-RE + 4X SRSF5-RBD on *β-globin*. (D) Examination of stability of the SR-stabilized U1 snRNP:*β-globin* complexes with short *β-globin* 5′SS RNA. (E) Displacement of SRSF5-RBD and SRSF2-RE from SR protein-saturated *β-globin* by U1 snRNP does not produce similar compact complexes as shown in C. (F) The ‘quaternary complexes’ containing *β-globin* variants, U2AF65, and U1 snRNP assembled with SRSF2-RE do not exhibit clear distinction between the WT and the mutant variants.

**Supplementary Figure 5. Global structural remodeling of *β-globin* for U1 snRNP recruitment**

(A) When any common nucleotide within a 5-nt long segment exhibited an enhancement of >1000-fold SHAPE reactivity in the optimal samples (15S1, S1S2, S1S5, S1S2S5), the central nucleotide was given a score of 1 in Figure 5B, C, D. (B) EMSA showing cooperative binding of

SRSF1-RBD to *β-globin* intron with six nucleotides extension toward 5′ exon (*βg-In*). (C) EMSA showing weak binding of SRSF5-RBD and SRSF2-RE to *βg-In* compared to full-length *β-globin*.

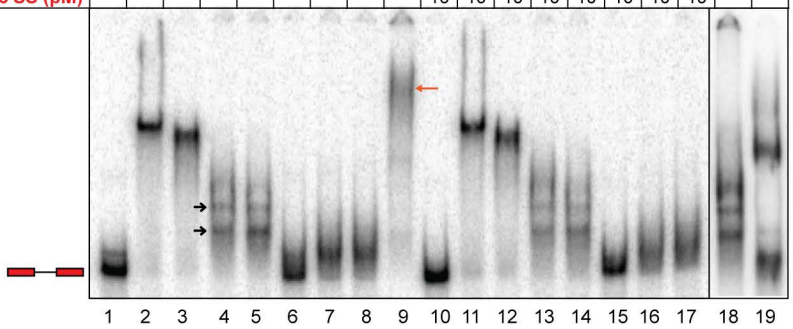
**Supplementary Figure 6. Global structural remodeling of a pre-mRNA and U1 snRNP recruitment**

(A) *In vivo* DMS reactivity of *β-globin* in *HeLa* cells with disrupted U1 snRNP mapped onto the SHAPE-derived secondary structure model by color-coding the nucleotides according to their DMS reactivity as indicated in the associated table; all G and U nucleotides exhibited zero reactivity. (B, C, D) RNP-MaP-derived protein-contacting nucleotides in U1 snRNP- (B), SRSF1- (C), or U1 snRNP+SRSF1- (D) bound *AdML* are marked in red in a previously determined (Saha et al., 2021) SHAPE-derived secondary structure model of *AdML*.

Figure 1

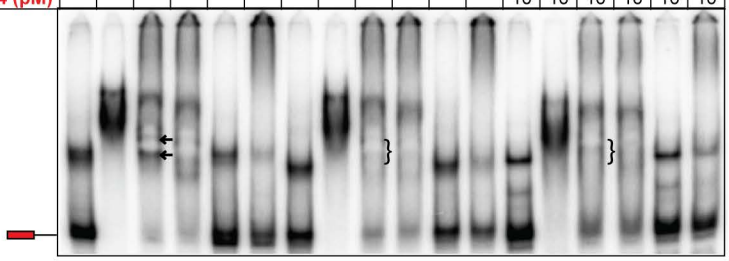
A

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SRSF1-RBD (nM) | | 60 | 60 | 60 | 60 | | | | 60 | | 60 | 60 | 60 | 60 | | | | 60 | 60 |
| U1 snRNP (nM) | | | 40 | 120 | 200 | 40 | 120 | 200 | 200 | | | 40 | 120 | 200 | 40 | 120 | 200 | 200 | 200 |
| αSRSF1 (μg) | | | | | | | | | 0.25 | | | | | | | | | | |
| αU1-C (μg) | | | | | | | | | | | | | | | | | | | 0.25 |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | | | | | | | | 10 | 10 |
| *β-globin* Δ5′SS (pM) | | | | | | | | | | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |

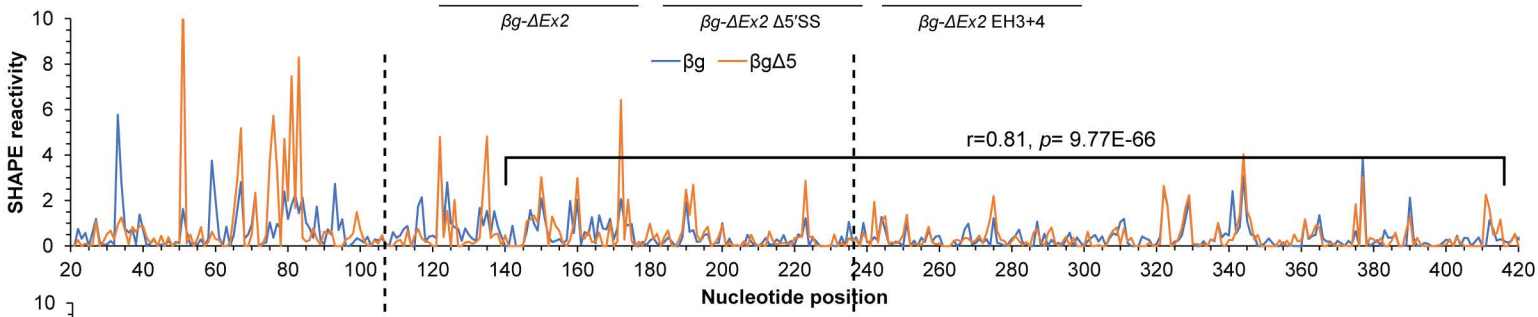



B

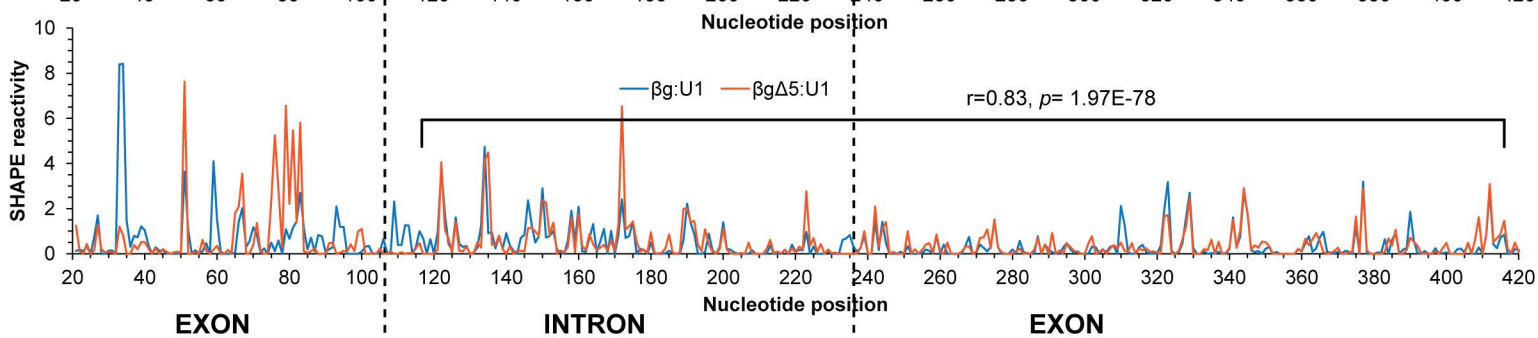
| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SRSF1-RBD (nM) | | 60 | 60 | 60 | | | | 60 | 60 | 60 | | | | 60 | 60 | 60 | | |
| U1 snRNP (nM) | | | 80 | 160 | 80 | 160 | | | 80 | 160 | 80 | 160 | | | 80 | 160 | 80 | 160 |
| *βg*ΔEx2 (pM) | 10 | 10 | 10 | 10 | 10 | 10 | | | | | | | | | | | | |
| *βg* ΔEx2 Δ5′SS (pM) | | | | | | | 10 | 10 | 10 | 10 | 10 | 10 | | | | | | |
| *βg* ΔEx2 EH3+4 (pM) | | | | | | | | | | | | | 10 | 10 | 10 | 10 | 10 | 10 |




C



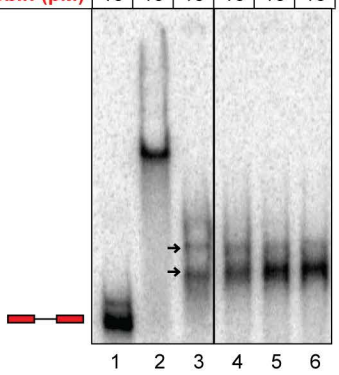

D




EXON INTRON EXON

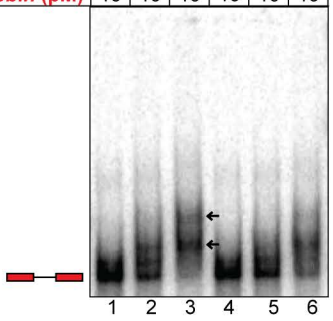
E

| | | | | | | |
|---|---|---|---|---|---|---|
| SRSF1-RBD (nM) | | 60 | 60 | 60 | 60 | 60 |
| U1 snRNP (nM) | | | 160 | 160 | 160 | 160 |
| βg-5′SS (μM) | | | | 0.4 | 0.8 | 1.2 |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 |




F

| | | | | | | |
|---|---|---|---|---|---|---|
| U1 snRNP (nM) | 40 | 80 | 200 | 40 | 80 | 200 |
| βg-5′SS (μM) | | | | 0.8 | 0.8 | 0.8 |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 |




G

pre-mRNA+U1 snRNP

⬇

weak global pre-mRNA:U1 snRNP interactions

pre-mRNA+SRSF1+U1 snRNP

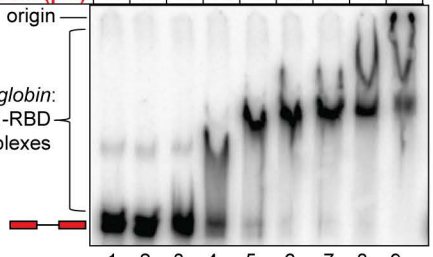
⬇

strong global pre-mRNA:U1 snRNP interaction

# Figure 2

**A**

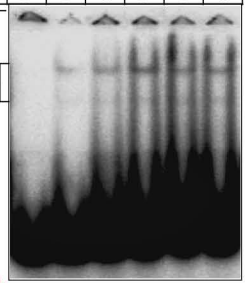
| SRSF1-RBD (nM) | | 2 | 10 | 20 | 40 | 50 | 60 | 70 | 120 |
|---|---|---|---|---|---|---|---|---|---|
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |




1 2 3 4 5 6 7 8 9

**B**

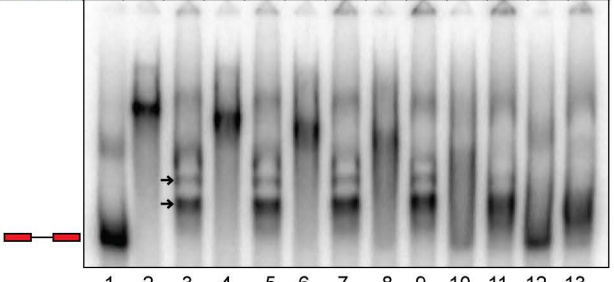
| SRSF1-RBD(nM) | | 20 | 40 | 80 | 160 | 240 |
|---|---|---|---|---|---|---|
| *Ron* ESE (nM) | 5 | 5 | 5 | 5 | 5 | 5 |




1 2 3 4 5 6

**C**

| SRSF1-RBD (nM) | | 550 | 550 | 450 | 450 | 350 | 350 | 250 | 250 | 150 | 150 | 50 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U1 snRNP(μM) | | | 1.5 | | 1.5 | | 1.5 | | 1.5 | | 1.5 | | 1.5 |
| *β-globin* (nM) | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |



1 2 3 4 5 6 7 8 9 10 11 12 13

**D**

*β-globin* + ≥10 SRSF1 molecules

↓

Cooperative binding

↓ U1 snRNP

Stable ternary complex

*β-globin* + <10 SRSF1 molecules

↓

Cooperative binding

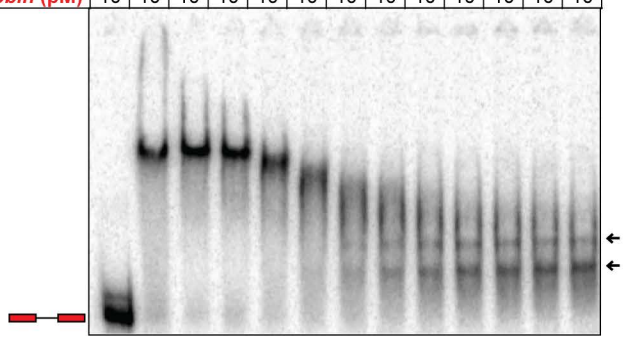
↓ U1 snRNP

Unstable ternary complex

# Figure 3

**A**

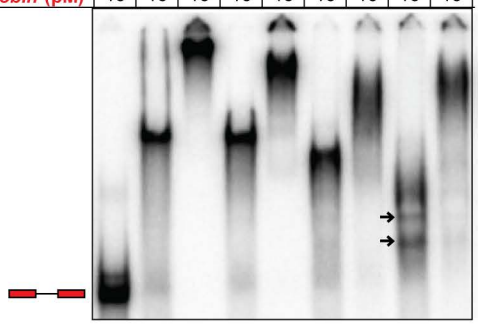
| SRSF1-RBD (nM) | | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U1 snRNP (nM) | | | 10 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 200 | 300 |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |



1 2 3 4 5 6 7 8 9 10 11 12 13

**B**

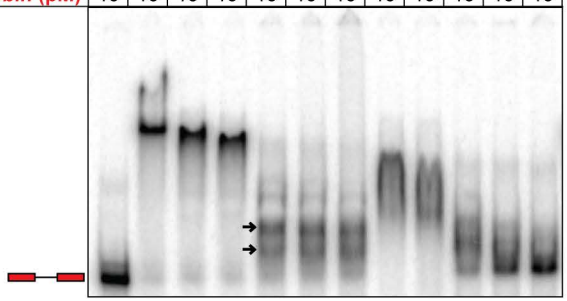
| SRSF1-RBD (nM) | | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
|---|---|---|---|---|---|---|---|---|---|
| U1 snRNP (nM) | | | | 40 | 40 | 80 | 80 | 160 | 160 |
| αSRSF1 (nM) | | | 0.25 | | 0.25 | | 0.25 | | 0.25 |
| αU1-C (nM) | | | | | | | | | |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |



1 2 3 4 5 6 7 8 9

**C**

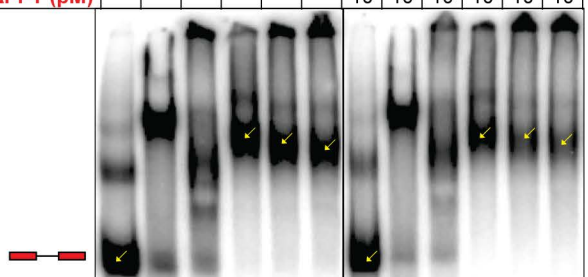
| SRSF1-RBD (nM) | | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U1 snRNP $B_{174}$ (nM) | | | 20 | 40 | 80 | 160 | 480 | | | | | |
| U1 snRNP $B_{174}$ ΔA (nM) | | | | | | | | 20 | 40 | 80 | 160 | 480 |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |



1 2 3 4 5 6 7 8 9 10 11 12

**D**

| SRSF1-RBD (nM) | | 60 | 60 | 60 | 60 | 60 | | 60 | 60 | 60 | 60 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U1 snRNP (nM) | | | 160 | 160 | 300 | 400 | | | 160 | 160 | 300 | 400 |
| U2AF65 (nM) | | | | 25 | 25 | 25 | | | | 25 | 25 | 25 |
| *β-globin* (pM) | 10 | 10 | 10 | 10 | 10 | 10 | | | | | | |
| *β-globin* ΔPPT (pM) | | | | | | | 10 | 10 | 10 | 10 | 10 | 10 |



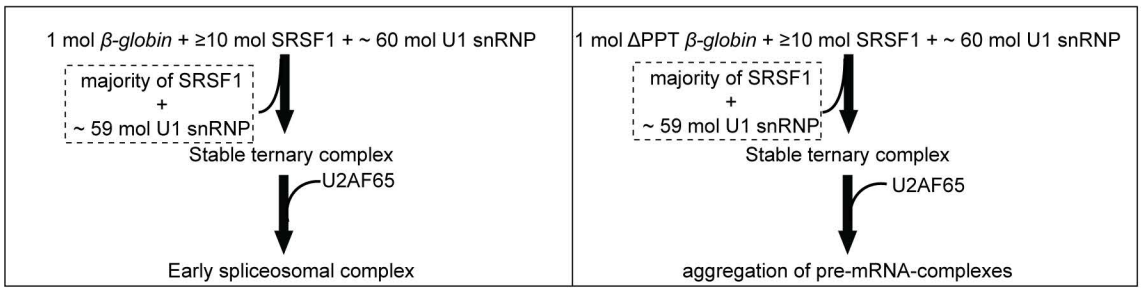
1 2 3 4 5 6 7 8 9 10 11 12

2.0
1.8
1.7

1.5
1.3
1.1

CCTATTGGTCTATTTTCCC>AAAAAAGGAAAAAAAAAAA

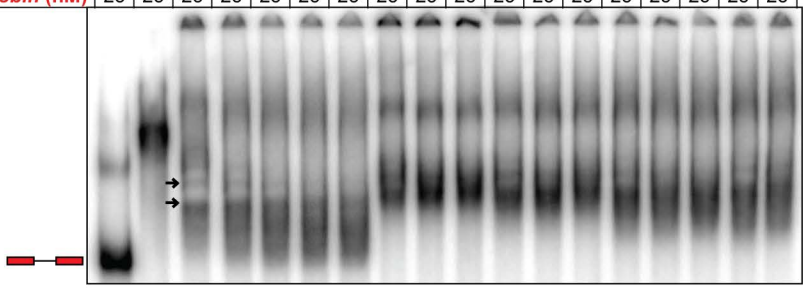
**E**



# Figure 4

**A**

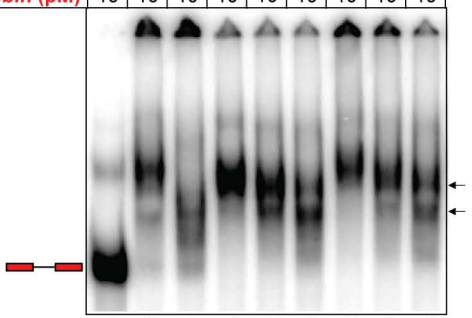
| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SRSF1-RBD (nM)** | | 550 | 250 | 200 | 150 | 75 | 50 | 150 | 75 | 50 | 150 | 75 | 50 | 150 | 75 | 50 | 150 | 75 |
| **SRSF2-RE (nM)** | | | | | | | | 250 | 250 | 250 | 175 | 175 | 175 | 75 | 75 | 75 | 50 | 50 |
| **U1 snRNP (nM)** | | | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| ***β-globin* (nM)** | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

**B**

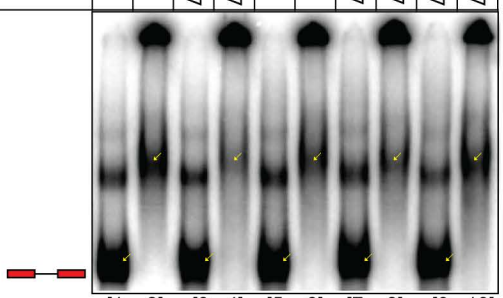
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **SRSF2-RE (nM)** | | | | 23 | 11 | 6 | | | |
| **U1 snRNP (nM)** | | 160 | 160 | 100 | 100 | 100 | 100 | 100 | 100 |
| **SRSF1-RBD (nM)** | | 60 | 10 | 20 | 20 | 20 | 20 | 20 | 20 |
| **SRSF5-RBD (nM)** | | | | | | | 32 | 16 | 8 |
| ***β-globin* (pM)** | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |



1 2 3 4 5 6 7 8 9

**C**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **SRSF1-RBD (nM)** | | 20 | | 20 | | 20 | | 20 | | 20 |
| **SRSF2-RE (nM)** | | 10 | | 10 | | 10 | | 10 | | 10 |
| **U1 snRNP (nM)** | | 400 | | 400 | | 400 | | 400 | | 400 |
| **U2AF65 (nM)** | | 25 | | 25 | | 25 | | 25 | | 25 |
| ***β-globin* (10 pM)** | WT | WT | Δ5′SS | Δ5′SS | ΔBS | ΔBS | ΔPPT | ΔPPT | Δ3′SS | Δ3′SS |



[1 2] [3 4] [5 6] [7 8] [9 10]

1.5 2.4 1.8 2.1 1.6

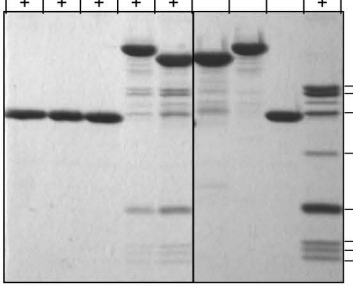
**D**

| ***β-globin* variants** | WT | Δ5′SS | ΔBS | ΔPPT | Δ3′SS |
|---|---|---|---|---|---|

ΔBS: CACTGAC>CGCTGGC
Δ3′SS: AG>CC

**E**

| | GST Pull-down | | | | | Input | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GST-SRSF1-RBD** | | | | + | | | + | | |
| **GST-SRSF2-RBD** | | | | | + | + | | | |
| **GST** | + | + | + | | | | | + | |
| **SRSF1-RBD** | | + | | | | | | | |
| **SRSF2-RBD** | | | + | | | | | | |
| **U1 snRNP** | + | + | + | + | + | | | | + |




**F**

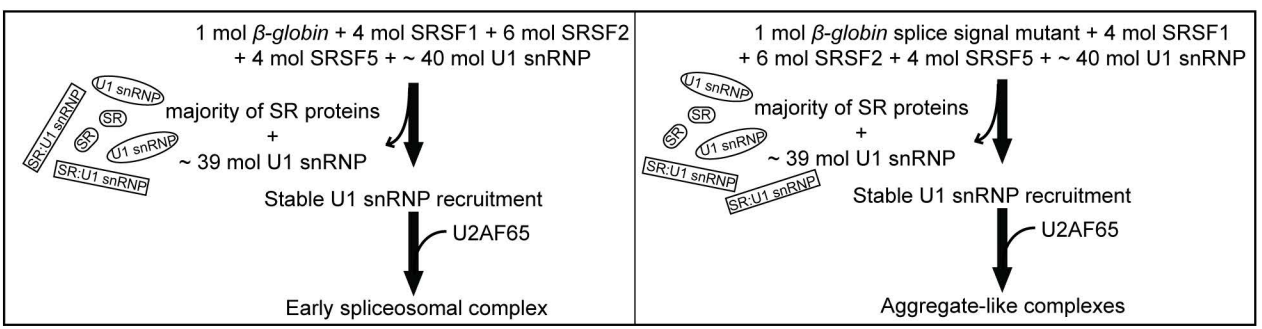
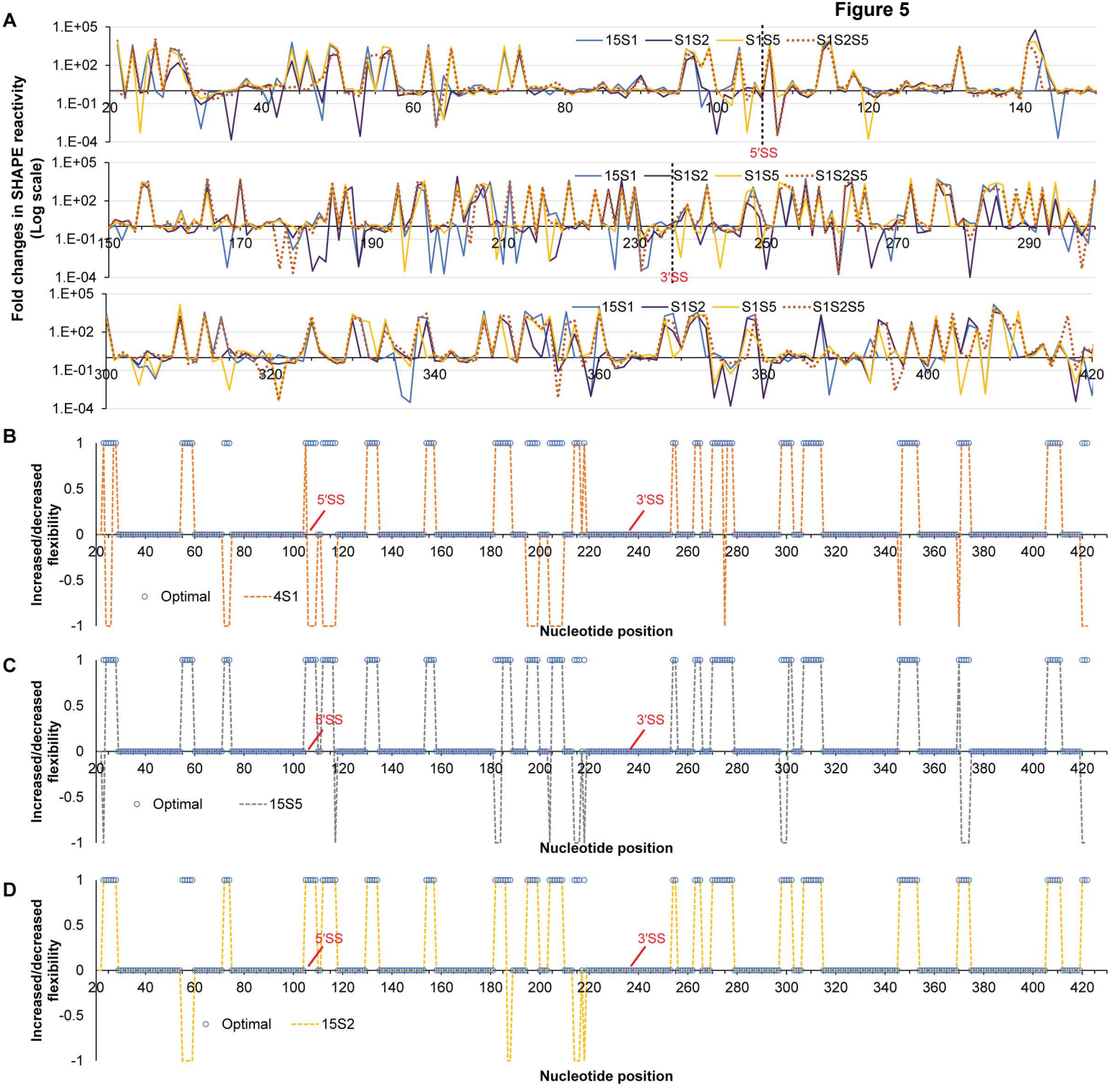


Figure 5


A
Fold changes in SHAPE reactivity (Log scale)
15S1
S1S2
S1S5
S1S2S5
5'SS
3'SS
B
Increased/decreased flexibility
Optimal
4S1
Nucleotide position
C
15S5
D
15S2

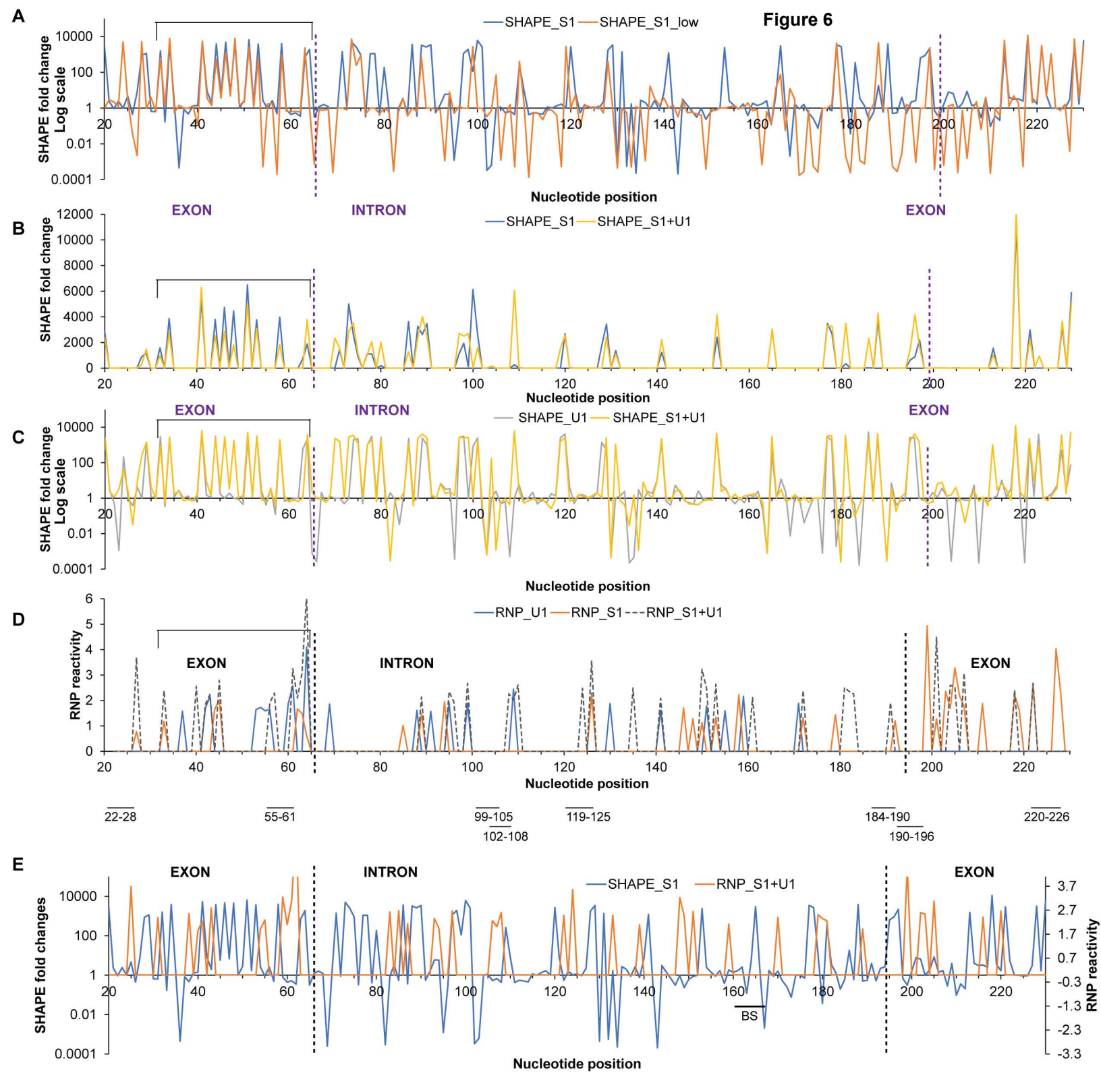
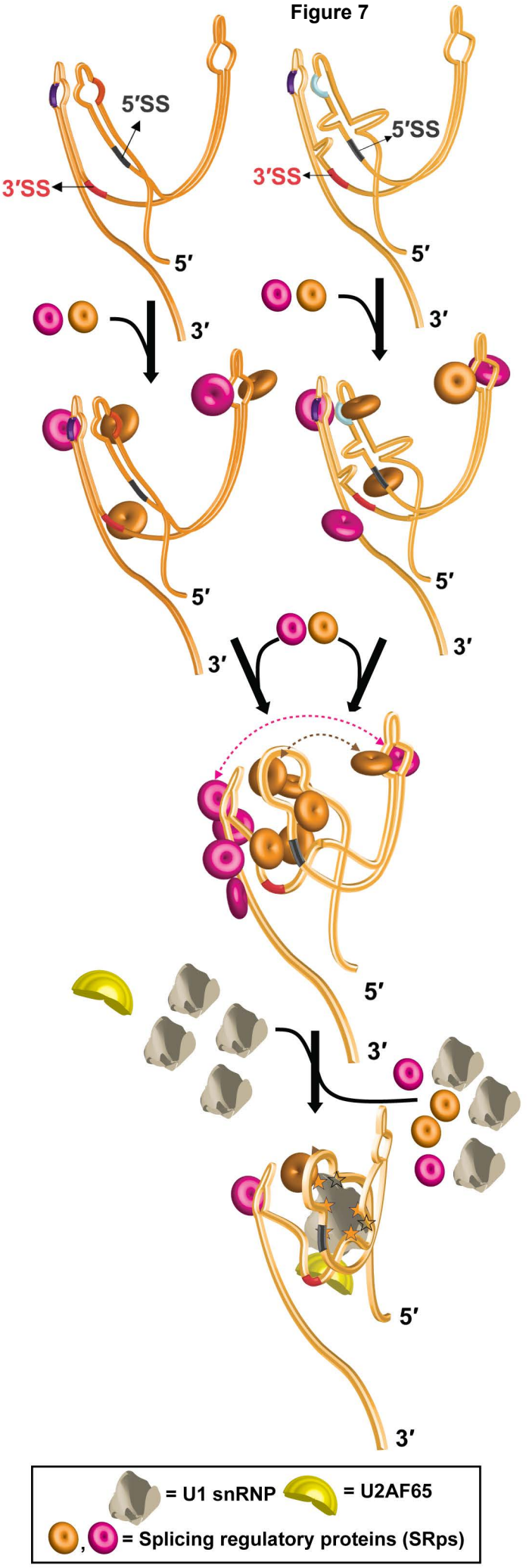


Figure 6
A
SHAPE_S1
SHAPE_S1_low
SHAPE fold change
Log scale
Nucleotide position
EXON
INTRON
EXON
B
SHAPE_S1
SHAPE_S1+U1
SHAPE fold change
Nucleotide position
C
SHAPE_U1
SHAPE_S1+U1
SHAPE fold change
Log scale
Nucleotide position
D
RNP_U1
RNP_S1
RNP_S1+U1
RNP reactivity
Nucleotide position
22-28
55-61
99-105
102-108
119-125
184-190
190-196
220-226
E
SHAPE_S1
RNP_S1+U1
SHAPE fold changes
RNP reactivity
BS
Nucleotide position

Figure 7


5′SS
3′SS
5′
3′
5′SS
3′SS
5′
3′
5′
3′
5′
3′
5′
3′
5′
3′
= U1 snRNP
= U2AF65
= Splicing regulatory proteins (SRps)