





Exon-independent recruitment of SRSF1 is mediated by U1 snRNP stem-loop 3

Andrew M Jobbins^{1,†,‡,§} , Sébastien Campagne^{2,†,¶} , Robert Weinmeister^{1,3,†}, Christian M Lucas¹, Alison R Gosliga^{1,††}, Antoine Clery², Li Chen¹, Lucy P Eperon¹, Mark J Hodson¹, Andrew J Hudson³, Frédéric H T Allain^{2,*}  & Ian C Eperon^{1,**} 

Abstract

SRSF1 protein and U1 snRNPs are closely connected splicing factors. They both stimulate exon inclusion, SRSF1 by binding to exonic splicing enhancer sequences (ESEs) and U1 snRNPs by binding to the downstream 5' splice site (SS), and both factors affect 5' SS selection. The binding of U1 snRNPs initiates spliceosome assembly, but SR proteins such as SRSF1 can in some cases substitute for it. The mechanistic basis of this relationship is poorly understood. We show here by single-molecule methods that a single molecule of SRSF1 can be recruited by a U1 snRNP. This reaction is independent of exon sequences and separate from the U1-independent process of binding to an ESE. Structural analysis and cross-linking data show that SRSF1 contacts U1 snRNA stem-loop 3, which is required for splicing. We suggest that the recruitment of SRSF1 to a U1 snRNP at a 5' SS is the basis for exon definition by U1 snRNP and might be one of the principal functions of U1 snRNPs in the core reactions of splicing in mammals.

Keywords exon definition; RNA splicing; RNA–protein interaction; SRSF1; U1 snRNP

Subject Categories Chromatin, Transcription & Genomics; RNA Biology

DOI 10.15252/emj.2021107640 | Received 4 January 2021 | Revised 4 October 2021 | Accepted 7 October 2021 | Published online 15 November 2021

The EMBO Journal (2022) 41: e107640

Introduction

SRSF1 is one of the most-studied regulators of alternative splicing. It is the archetypal member of the family of SR proteins, proteins that have one or two RNA recognition motif (RRM)-type RNA-binding

domains and a C-terminal RS domain rich in arginine-serine dipeptides that can be phosphorylated extensively. The SR proteins have roles in a wide variety of processes, including transcription (Ji *et al*, 2013), nuclear export, translation and nonsense-mediated decay (Long & Caceres, 2009; Maslon *et al*, 2014; Muller-McNicoll *et al*, 2016), but it is for their roles in alternative splicing that they are best known. SRSF1 in particular activates or represses the inclusion of hundreds of exons (Pandit *et al*, 2013; Anczukow *et al*, 2015; Bradley *et al*, 2015), and this activity is thought to be the primary reason why it is both essential (Wang *et al*, 1996; Longman *et al*, 2000; Lin *et al*, 2005b) and an oncoprotein (Das & Krainer, 2014). It also plays an essential but poorly understood role in the core reactions of splicing.

Three key observations made in 1993 underpin the prevailing model for the actions of SR proteins in splicing. SR proteins were found to bind purine-rich enhancer sequences (Lavigneur *et al*, 1993; Sun *et al*, 1993), to interact with the core splicing components U2AF35 and U1-70K (Wu & Maniatis, 1993) and, for SRSF1 in particular, to stabilize the binding of U1 snRNP complexes at 5' splice sites (5'SSs) (Eperon *et al*, 1993). The resultant model for activation of splicing by SRSF1 is that it binds exonic splicing enhancer sequences (ESEs) (Graveley & Maniatis, 1998; Sanford *et al*, 2009; Clery *et al*, 2013; Pandit *et al*, 2013; Ray *et al*, 2013; Anczukow *et al*, 2015) and then recruits limiting splicing factors such as U1 snRNPs or U2-associated proteins to 5' or 3' splice sites, respectively, by direct protein–protein interactions that stabilize the association of the splicing factor with the pre-mRNA (Eperon *et al*, 1993; Lavigneur *et al*, 1993; Wu & Maniatis, 1993; Amrein *et al*, 1994; Kohtz *et al*, 1994; Staknis & Reed, 1994; Jamison *et al*, 1995; Tarn & Steitz, 1995; Wang *et al*, 1995; Cao & Garcia-Blanco, 1998; Graveley *et al*, 2001; Martins de Araujo *et al*, 2009; Cho *et al*, 2011; Smith *et al*, 2014; Akerman *et al*, 2015). This model is commonly depicted in cartoon representations of

1 Leicester Institute of Structural & Chemical Biology and Department of Molecular & Cell Biology, University of Leicester, Leicester, UK

2 Institute of Biochemistry, ETH Zürich, Switzerland

3 Leicester Institute of Structural & Chemical Biology and Department of Chemistry, University of Leicester, Leicester, UK

*Corresponding author. Tel: +41 0 44 633 39 40; E-mail: allain@bc.biol.ethz.ch

**Corresponding author. Tel: +44 0 116 229 7012; E-mail: eci@leicester.ac.uk

†These authors contributed equally to this work

‡Present address: MRC London Institute of Medical Sciences, London, UK

§Present address: Institute of Clinical Sciences, Imperial College London, London, UK

¶Present address: Inserm U1212, CNRS UMR5320, ARNA Laboratory, Bordeaux Cedex, France

††Present address: Institut für Industrielle Genetik, Abt.(eilung) Systembiologie, Universität Stuttgart, Stuttgart, Germany

splicing (Will & Luhrmann, 2011; Lee & Rio, 2015; Wahl & Luhrmann, 2015). Moreover, this model is compatible with the two properties that led to the isolation of SRSF1 originally: its ability to modulate 5'SS selection and its ability to restore splicing activity to S100 cytoplasmic extracts (Ge & Manley, 1990; Krainer *et al*, 1990a, 1990b, 1991; Ge *et al*, 1991), since the latter could be accounted for by the ability to enhance the binding of scarce components and depended on exon sequences (Chandler *et al*, 1997; Mayeda *et al*, 1999). We inferred from recent single-molecule experiments that the activity of a mammalian ESE is limited by a low probability of transient occupancy by SRSF1, and, using chimeric RNA with non-RNA linkers, that the association of bound SRSF1 with 3' splice site factors involves direct protein–protein interactions mediated by 3D diffusion; this stabilizes the binding of the successful molecule of SRSF1 (Jobbins *et al*, 2018).

Several observations suggest that SRSF1 has other activities. First, SRSF1 was shown to stimulate splicing in S100 extracts of truncated pre-mRNA substrates that contained only 1–3 nt of the 5' exon and lacked a 3' exon (Hertel & Maniatis, 1999; Zhu & Krainer, 2000). Second, whereas the activation of splicing by SR proteins requires phosphorylation of their RS domains (Caceres & Krainer, 1993; Zuo & Manley, 1993; Mermoud *et al*, 1994; Roscigno & Garcia-Blanco, 1995; Cao *et al*, 1997; Xiao & Manley, 1997, 1998; Graveley & Maniatis, 1998; Zhu & Krainer, 2000; Cartegni & Krainer, 2003; Shen & Green, 2006), the catalytic reactions of splicing require dephosphorylation of the RS domains of SR proteins and specifically SRSF1 (Mermoud *et al*, 1992, 1994; Cao *et al*, 1997; Xiao & Manley, 1998). This suggests that SRSF1 has an additional function in splicing. Third, cross-linking has shown that the RS domain of SRSF1 contacts the pre-mRNA in mature spliceosomal complexes B and C (Shen *et al*, 2004). Attempts to distinguish the putative roles of SRSF1 in splicing from those in alternative splicing by identifying the domains responsible have foundered, possibly because both of the two original assays might have been assaying the recruitment activity; alternatively, multiple domains might be involved in interactions or the RS domain might be involved in autoregulation (Caceres & Krainer, 1993; Zuo & Manley, 1993; Wang & Manley, 1995; Caceres *et al*, 1997; Eperon *et al*, 2000; Zhu & Krainer, 2000; Lin *et al*, 2005b; Shaw *et al*, 2007; Cho *et al*, 2011; Serrano *et al*, 2016).

Establishing whether SRSF1 really is engaged in the constitutive reactions of splicing is made particularly difficult because the SR proteins are found at variable and non-stoichiometric levels in spliceosome preparations (Schmidt *et al*, 2014). This is not unexpected, given that SRSF1 can bind to a wide range of pre-mRNA sequences and might also interact by electrostatic interactions of its RS domain. Binding by multiple molecules that are involved in different interactions precludes the use of conventional methods for resolving different events. Here, we have unravelled this complexity by the use of single-molecule methods and reveal that SRSF1 is associated in a 1:1 complex with U1 snRNPs, interacts with stem-loop 3 of U1 snRNA and is recruited by U1 snRNPs to 5' splice sites. This suggests a function for U1 snRNPs in splicing reactions, which is otherwise still missing, and explains why 5' splice sites can act like ESEs in exon definition (Kreivi *et al*, 1991; Hwang & Cohen, 1997; Fernandez Alanis *et al*, 2012; Braun *et al*, 2018; Singh & Singh, 2019; Erkelenz *et al*, 2021).

Results

Detection of single molecules in complexes

To detect single complexes of SRSF1 with pre-mRNA, Cy5-labelled pre-mRNA was incubated in splicing conditions in nuclear extracts prepared from cells expressing functional monomeric mEGFP-SRSF1 (Table 1; Appendix Fig S1; (Sleeman *et al*, 1998; Ellis *et al*, 2008)). In most of the experiments shown here, the incubations were done in the presence of ATP and a 2'-O-methyl oligonucleotide complementary to U6 snRNA that caused spliceosome assembly to stall at complex A (Donmez *et al*, 2007). Nuclear extract (NE) 3 contained mCherry-U1A as well, which associates with pre-mRNA only as part of the U1 snRNP (Hodson *et al*, 2012). The reaction mixtures were diluted and then applied to coverslips (Fig 1). Complexes adsorb directly to the glass surface and are stable indefinitely. Total internal reflection fluorescence microscopy was used to locate surface-bound RNA molecules, and then, protein fluorescence was recorded until all the fluorophores had bleached. Colocalized spots containing Cy5 and mEGFP or mCherry were identified, and the number of steps in which the protein fluorophores bleached was recorded for each complex. Each step is caused by the bleaching of a single molecule. The number of complexes in which bleaching had taken place in 1, 2, etc. steps is plotted as a percentage of the number of RNA molecules detected (Fig 1), and the absolute number of complexes is given above each bar.

Stable binding of SRSF1 depends on U1 snRNP and 5' splice sites

We reported recently that most of the molecules of a pre-mRNA containing one intron with a single strong 5' splice site (GloC) were associated with just one molecule of mEGFP-SRSF1 in complex A conditions (Jobbins *et al*, 2018) (Fig 2A), despite the presence of a number of potential binding sites in the exon (Schaal & Maniatis, 1999). To test whether this surprisingly restricted binding was connected with splicing, a second pre-mRNA was tested that contained two introns, derived by duplication of the sequences in GloC (Hodson *et al*, 2012). The distribution expected if all SRSF1-associated pre-mRNA molecules were bound by either one or two molecules of mEGFP-SRSF1 in this extract (NE1) is shown in

Table 1. Concentrations and properties of expressed fluorescent proteins.

	NE1	NE2	NE3	
	mEGFP-SRSF1	mEGFP-SRSF1	mEGFP-SRSF1	mCherry-U1A
Concentration in extract (μM)	1.5	0.72	n.d.	n.d.
% tagged: endogenous	90:10	59:41	58:42	57:43
Molecules bleaching in 2 vs 1 steps, +ATP	32:129	9:118	18:138	15:124

Concentrations were calculated from Western blotting of nuclear extract alongside recombinant protein; the percentages labelled or unlabelled (endogenous) were also calculated by Western blotting; the numbers of molecules recorded as bleaching in one or two steps when not associated with exogenous RNA are shown. n.d., not determined.

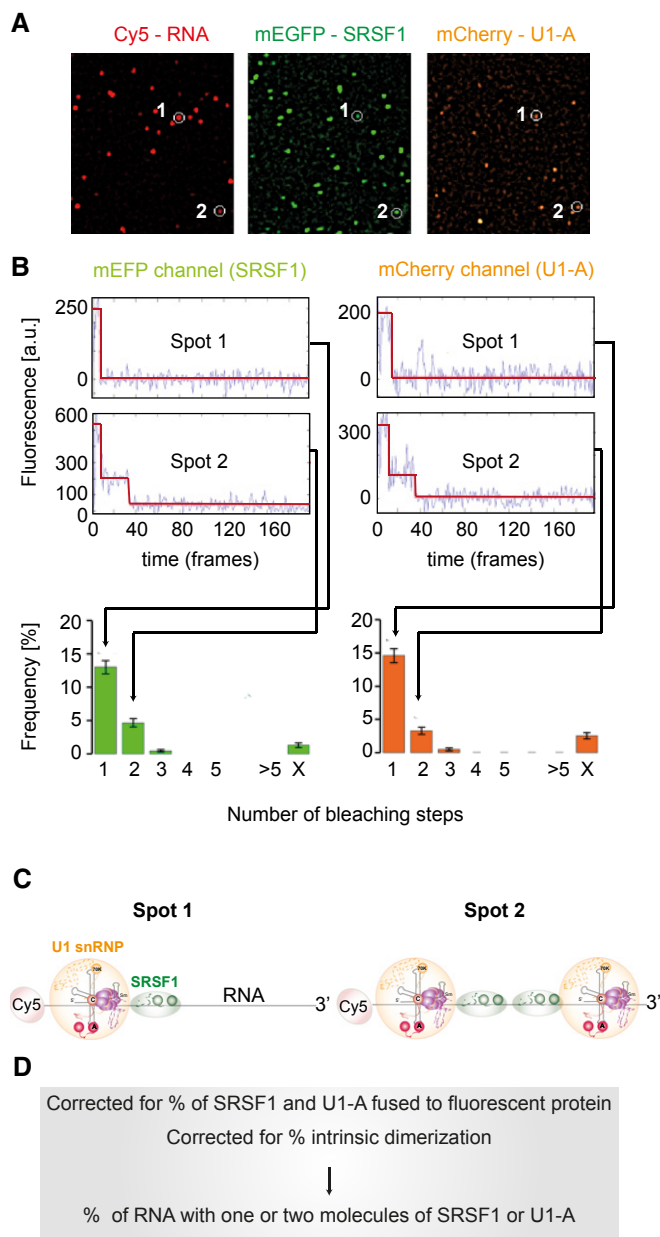


Figure 1. Representation of methods used for determining the frequencies with which 1, 2, etc., molecules of fluorescent protein are associated with pre-mRNA.

- A** After incubation of Cy5-labelled pre-mRNA in a nuclear extract expressing mEGFP-SRSF1 and mCherry-U1A, samples are diluted and applied to a coverslip. Fluorescent images are acquired at time frames of 20–100 ms from excitation successively at 633 nm (to detect Cy5), 561 nm (to detect mCherry) and 488 nm (to detect mEGFP), and at each stage, collection continues until all the molecules have bleached. Cy5-labelled RNA molecules (red) and mEGFP (green) and mCherry (orange) molecules are detected as spots in the composite images from each laser (Materials and Methods). Colocalized spots are identified (white circles); the other spots are not colocalized.
- B** Time courses of fluorescence are analysed to detect bleaching of the fluorophores in each spot. Two hypothetical molecular complexes are followed, one in which a molecule of RNA is associated with one fluorescent protein of each type (Spot 1) and the other in which it is associated with two fluorescent proteins of each type (Spot 2; either the result of binding to another site in the pre-mRNA or dimerization). Each spot in which mEGFP bleached in one step contributes to the frequency of the $N = 1$ class; each spot in which it bleached with two steps is added to the $N = 2$ class, etc. The proportions of RNA spots associated with mEGFP signals that bleach in one, two, etc., steps are shown as a frequency histogram.[§]
- C** Inferred compositions of Spot 1 and Spot 2. Note that the positions and mechanisms of binding are not revealed by this method. The 5' Cy5 fluorophore on the pre-mRNA is highlighted in red; the U1 snRNP is represented by an orange disc, containing a diagram of the U1 RNA secondary structure with solid spheres representing the structured domains of the Sm proteins and the U1-specific proteins, lines showing extended regions of the protein structures and dashed lines showing unstructured regions, since it was detected by mCherry-U1A fluorescence; SRSF1 is represented by a green ellipse, containing two darker green circles representing the two RNA-binding domains, since it was detected by mEGFP fluorescence.
- D** The frequency of complexes containing one, two, etc., molecules of SRSF1 and U1 snRNPs is calculated by taking into account the relative levels of the endogenous, untagged proteins and the frequency of intrinsic dimerization of each protein.

Appendix Fig S2. The observed distribution (Fig 2A) was intermediate between the two canonical distributions. Estimations of the proportions of the two classes suggest that approximately 55% of the SRSF1-pre-mRNA complexes contained one molecule of SRSF1 and 45% contained two molecules of SRSF1. In previous work, we showed that under these conditions about 40 and 60% of the same pre-mRNA was colocalized with one or two U1 snRNPs, respectively (Hodson *et al*, 2012), whereas only one U2 snRNP-containing complex was detected (Chen *et al*, 2017). The results in Fig 2A raise the possibility that SRSF1 was associated with U1 snRNPs in spliceosomal complex A.

The link to U1 snRNPs was tested by repeating the experiment with a pre-mRNA containing a mutant 5' splice site that

substantially reduces U1 snRNP association (Hodson *et al*, 2012). The fraction of pre-mRNA molecules associated with SRSF1 was halved in the mutant (Fig 2B). A very similar change was seen when the binding of U1 snRNP in NE2 was blocked by a complementary 2'-O-methyl oligonucleotide (Hodson *et al*, 2012) (Fig 2B). The residual distributions in these conditions appear to represent alternative modes of association with pre-mRNA molecules that are unable to form complex A (see Discussion). We conclude that U1 snRNP binding to 5'SSs directly or indirectly determines whether SRSF1 binds with a fixed stoichiometry.

The strong dependence on U1 snRNPs and the 5'SS was seen also when experiments were done after ATP depletion, in conditions used to form complex E, often taken as a model for the earliest interactions that form prior to complex A or splice site selection (Michaud & Reed, 1991, 1993; Jamison *et al*, 1992; Lim & Hertel, 2004; Kotlajich *et al*, 2009). Under these conditions, SRSF1 is substantially hypophosphorylated (Appendix Fig S1). However, compared with complex A conditions, there was an increase in the proportion of GloC pre-mRNA molecules associated with two molecules of mEGFP-SRSF1 and a noticeable increase in

[§]Correction added on the 4th of January 2022, after first online publication: the x-axis label has been corrected from "time (min)" to "time (frames)".

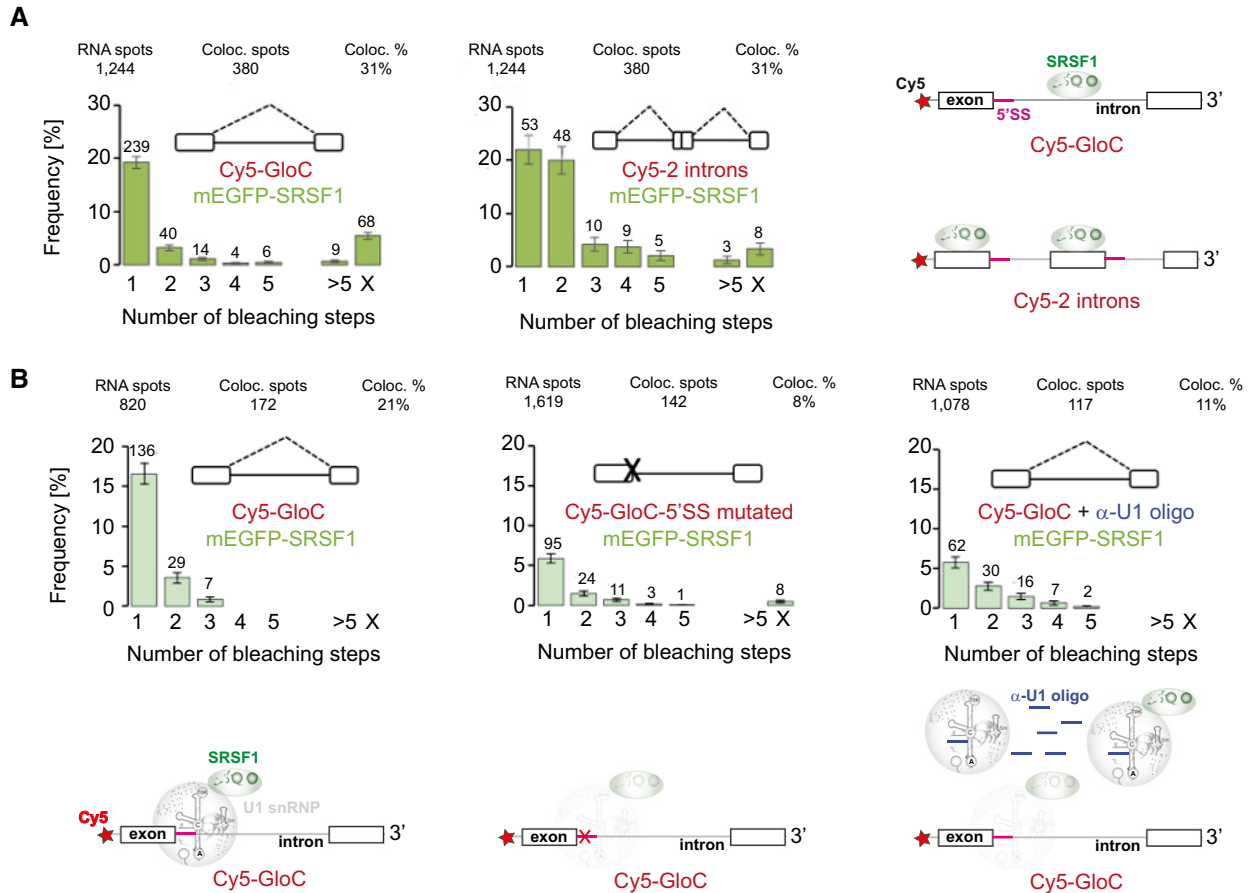


Figure 2. Dependence of stoichiometric SRSF1 association with pre-mRNA on 5' splice sites and U1 snRNA.

A Two-way colocalization of mEGFP-SRSF1 with Cy5-pre-mRNA containing one (left) or two (right) introns. In this experiment, nuclear extract (NE) 1 was used. The SRSF1 histograms are colour-coded in three shades of green to indicate the extract used in each experiment shown. The number above each bar indicates the number of complexes in which complete bleaching was achieved in 1, 2, 3, etc., steps, and hence the number of complexes in which there were 1, 2, 3, etc., molecules of mEGFP-SRSF1. > 5 refers to complexes where a higher number of bleaching steps were measured; X represents complexes where the number could not be determined. RNA spots refer to numbers of molecules of Cy5-labelled pre-mRNA detected in the fields acquired; Coloc. spots is the number of these associated with mEGFP-SRSF1, and Coloc. % is the resulting percentage. The error bars are the square root of the variance of the binomial probability that an RNA spot will be associated with the given number of protein bleaching steps. The pre-mRNA is represented above the histograms. For globin-derived pre-mRNA, the white boxes represent exons and the line in between is the intron. Potential splicing pathways are shown as dashed lines. The pre-mRNA with two introns was formed by duplication of GloC sequences. The cartoons on the right show the inferences drawn from the histograms. The pre-mRNA molecules are shown in grey, with the Cy5 label shown as a red star and 5' splice sites as purple segments. From the histogram on the left we infer that, in those complexes where a molecule of pre-mRNA was associated with SRSF1, only one molecule of SRSF1 (green) was present, while the histogram on the right shows that about half of the complexes of pre-mRNA containing two introns were associated with two molecules of SRSF1.

B Two-way colocalization of mEGFP-SRSF1 with Cy5-pre-mRNA showing dependence of SRSF1 binding on 5' splice sites and U1 snRNP. Complexes formed in NE2 on GloC pre-mRNA (left), a derivative with an inactivating mutation in the 5' splice site mutant (centre) and on GloC in the presence of an additional 2'-O-methyl oligonucleotide complementary to the 5' end of U1 snRNA (right). The last distribution is geometric (χ^2 , $P_{geo} = 0.72$). Error bars are as described in (A). The interpretative cartoons underneath are based on the observed dependence of SRSF1 binding on U1 snRNPs and represent the possibility that in many cases, SRSF1 binding might be mediated by U1 snRNPs. SRSF1 is in green as in Fig 1, and the U1 snRNPs are in grey since they were not detected directly in this experiment. Paler colours indicate reduced levels of a labelled protein in the complex.

larger complexes that are not U1-dependent (Appendix Fig S3). This is consistent with the co-existence of U1 snRNP-dependent recruitment and sequence-independent interactions made by the positively charged hypophosphorylated RS domains. Non-specific binding to the pre-mRNA by an unphosphorylated RS domain has been observed previously (Cho *et al*, 2011). The involvement of phosphorylation was confirmed by the inclusion of phosphatase inhibitors. In this case, the pattern of SRSF1 association

approached more closely that seen in complex A, and the complexes with very large numbers of molecules of SRSF1 were much reduced (Appendix Fig S3D). These results are very similar to those we observed previously with U2AF65 and U2 snRNP (Chen *et al*, 2017), and confirm that complex I, formed in the presence of phosphatase inhibitors, might be a better model for early splicing complexes in higher eukaryotes than complex E (Chen *et al*, 2017).

SRSF1 binds with U1 snRNP

The observation that SRSF1 recruitment is determined by U1 snRNP does not necessarily mean that both factors are present on the pre-mRNA concurrently. Specifically, U1 snRNP might recruit SRSF1 and then dissociate. To test this, colocalization studies were done with a nuclear extract (NE3) containing mEGFP-SRSF1 and mCherry-U1A, which only associates with pre-mRNA as part of the U1 snRNP (Hodson *et al*, 2012). In both cases, the labelled component was expressed at about the same level as the endogenous protein (Table 1; Appendix Fig S1). Both proteins were predominantly found as single molecules on the pre-mRNA (Fig 2B), although there was a higher proportion of complexes with mEGFP-SRSF1 than with mCherry-U1A containing two labelled proteins. When pre-mRNA complexes containing both protein fluorophores (three-way colocalization) were examined, the distribution of each was similar to that observed in the two-way colocalization (Fig 2B, lower panel; χ^2 , $P_{SRSF1\ 3\text{-way} = 2\text{-way}} = 0.93$; χ^2 , $P_{U1A\ 3\text{-way} = 2\text{-way}} = 0.76$) and, significantly, these complexes were found at about half the frequency of complexes that were selected for the presence of one protein without regard to the other. This result is exactly what would have been expected if every one of the complexes containing labelled SRSF1 contained U1 snRNP (either labelled or unlabelled), and vice versa.

SRSF1 is recruited to pre-mRNA without exons

The concurrent colocalization might be attributed to mutual stabilization resulting from the binding of SRSF1 to an ESE and interactions with U1 snRNP (Jobbins *et al*, 2018). This possibility was tested with a pre-mRNA in which the exons contained only approximately 2 nucleotides (Fig 3C). These transcripts, as with those in Fig 4 also, were labelled using 5' Cy5-maleimide, which was used previously for studies on spliceosome assembly (Ohrt *et al*, 2012). The substitution of Cy5 for the normal cap produces only relatively small (20–25%) reductions in the rates of splicing, and complexes accumulate within the times used for incubations (Appendix Fig S4).

The colocalization percentages showed that there was no reduction in the binding of either SRSF1 or U1 snRNPs, and, again, concurrent detection was observed in around one-half of the complexes. Strikingly, whereas the proportions of complexes with one or two molecules of mCherry-U1A are the same on this pre-mRNA as with GloC (χ^2 , $P_{U1A\ 2\text{-way}\ 2C = 2B} = 0.49$), there are proportionately fewer complexes containing two molecules of mEGFP-SRSF1 than was seen with GloC (χ^2 , $P_{SRSF1\ 2\text{-way}\ 2C = 2B} = 0.033$), and the distribution of complexes with one or two molecules of mEGFP-SRSF1 now matches that observed for mCherry-U1A (χ^2 , $P_{2C\ 2\text{-way}\ SRSF1 = U1A} = 0.64$). We infer that the complexes with GloC contain some pre-mRNAs in which SRSF1 binds weakly via the exon sequences, but that with the short exons, recruitment is associated primarily with U1 snRNP binding. We conclude that SRSF1 recruitment with U1 snRNP does not require exonic splicing enhancer sequences.

The association of U1 snRNP and SRSF1 does not require pre-mRNA

The strong link between U1 snRNP and SRSF1 association with the pre-mRNA suggested the possibility that interactions between the

two molecules might occur even prior to or in the absence of binding to pre-mRNA. Around 24% of mEGFP-SRSF1 molecules and 21% of mCherry-U1A molecules were associated with mCherry-U1A and mEGFP-SRSF1, respectively, in the absence of pre-mRNA (Fig 3D). Given the fraction of each protein that was labelled, we infer that around half the molecules in each case were present in a U1:SRSF1 heterodimer. Interestingly, as with the results with the 2 nt exon (Fig 3C), the distributions of complexes with one or two molecules of mEGFP-SRSF1 and mCherry-U1A are the same (χ^2 , $P_{2C\ SRSF1 = U1A} = 0.55$). When the experiments were repeated in the presence of ribonuclease, to remove residual pre-mRNA, the level of association was reduced but was still substantial (Appendix Fig S5). Previous analyses had shown that the U1-70K protein can interact directly with SRSF1, via either the RS or RRM domains (Kohtz *et al*, 1994; Xiao & Manley, 1997; Cho *et al*, 2011), and that they are in close proximity in cells even after ribonuclease treatment (Ellis *et al*, 2008). Our results are consistent with the possibility that a significant proportion of each protein is present in a heterodimer in functional splicing conditions.

Recruitment of SRSF1 via U1 snRNP or an ESE is separate processes

The recruitment of a single molecule of SRSF1 by U1 snRNP is at odds with the conventional picture in which SRSF1 binds independently to ESE sequences in a pre-mRNA and then stabilizes the binding of splicing factors. To test whether our observations have revealed a second and different mode of SRSF1 binding, we used a chimeric pre-mRNA (BGSMN2) in which the 5' part was derived from GloC and the 3' part from SMN2 exon7, which does not contain an ESE bound by SRSF1. The splicing of this was enhanced by the addition of a 3'-terminal SRSF1-dependent ESE (BGSMN2 + ESE-Ax4), which increases the association of U2AF35, U2AF65 and U2 snRNPs (Jobbins *et al*, 2018). In the absence of the ESE, the predominant complexes formed in complex A conditions contained one molecule of mEGFP-SRSF1, which was U1 snRNP-dependent (Fig 4A). The addition of the ESE caused an upsurge in molecules associated with two molecules of mEGFP-SRSF1, such that 100% occupancy by two molecules of SRSF1 could not be excluded (χ^2 , $P_{2\ SRSF1} = 0.26$). In this case, addition of the anti-U1 oligonucleotide substantially reduced the frequency of bleaching in two steps from approximate equality with bleaching in one step to about one-third of the value. This indicates a major reduction in complexes containing two molecules of SRSF1, but the proportion cannot be estimated because of the background associated with U1 sequestration (c.f. Fig 2B). We conclude that the ESE sequence acted independently from the U1 snRNP to recruit an additional molecule of SRSF1 and that U1-dependent recruitment is a separate phenomenon.

A downstream 5'SS recruits SRSF1 and, like an ESE, recruits U2 snRNP to the 3'SS

A poorly understood property of U1 snRNP is the ability of a 5'SS at the 3' end of an exon to stimulate splicing via an increase in U2AF65 binding to the upstream 3'SS (Grabowski *et al*, 1991; Kreivi *et al*, 1991; Hoffman & Grabowski, 1992; Cote *et al*, 1995; Izquierdo *et al*, 2005; Singh *et al*, 2007; Palacino *et al*, 2015). This is fundamental to exon definition models of mammalian splicing (Robberson *et al*,

1990; Sterner *et al*, 1996). Although the mechanisms involved have not been described, and U1 snRNPs do not appear to interact directly with U2AF, this property of U1 snRNP is similar to that of an ESE. The data described above suggest that the underlying mechanism might be the recruitment of SRSF1 by U1 snRNP.

To test this hypothesis, a consensus 5'SS sequence was attached to the 3' end of both GloC and BGSMN2 pre-mRNAs. This increased splicing of BGSMN2 more than the addition of two repeats of a strong ESE (BGSMN2-ESEx2; Appendix Fig S6) (Jobbins *et al*, 2018). There was a significant increase in the proportion of

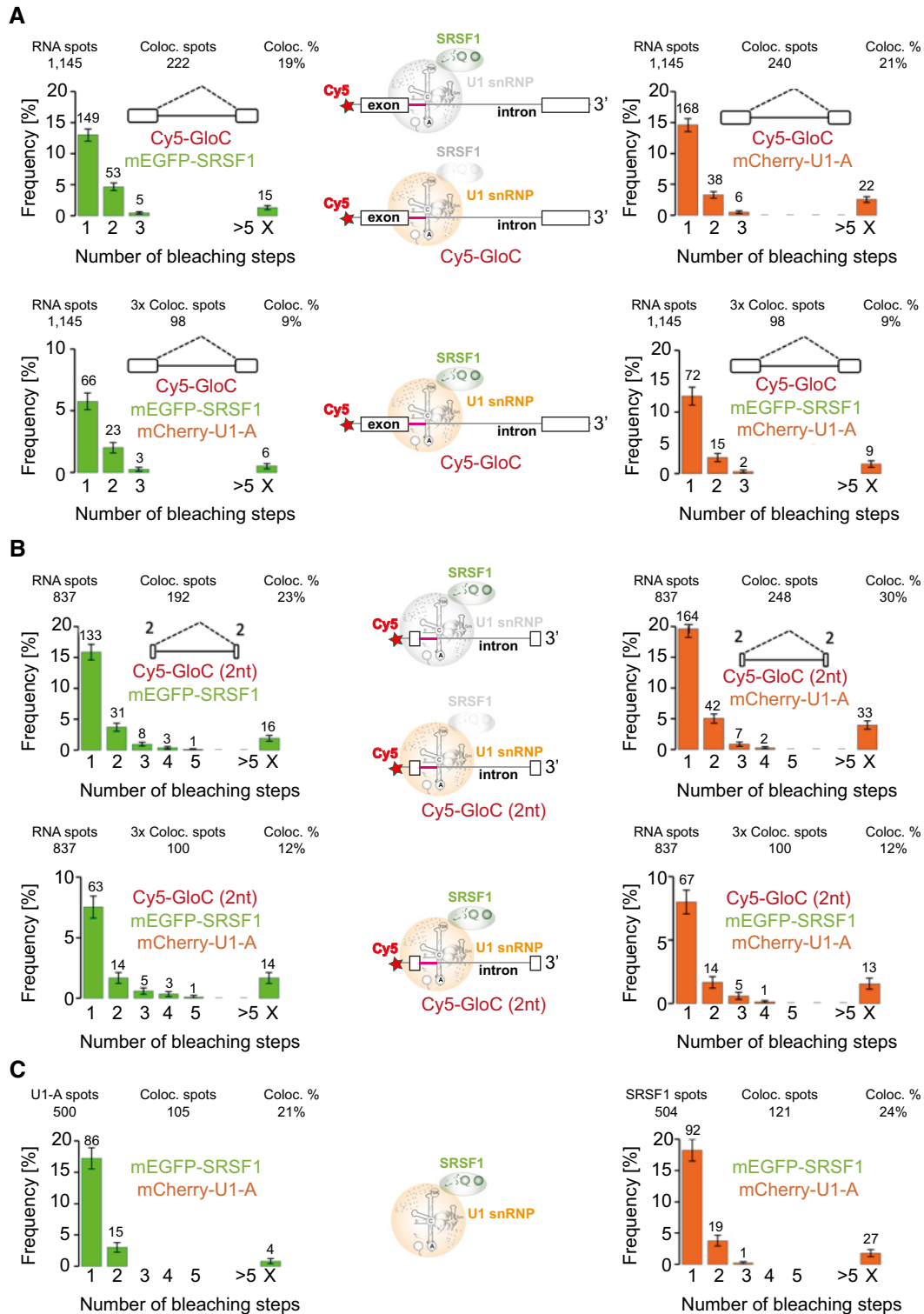


Figure 3.

Figure 3. SRSF1 and U1 snRNPs are present concurrently in individual pre-mRNA complexes even in the absence of exons.

- A Complexes formed on GloC in NE3, containing labelled mEGFP-SRSF1 and mCherry-U1A, in which the distribution of mEGFP-SRSF1 (left) and mCherry-U1A (right) in two-way colocalization with pre-mRNA is shown; below, the respective distributions of mEGFP-SRSF1 and mCherry-U1A on 3-way colocalized complexes. The error bars in the histograms are the square root of the variance of the binomial probability that an RNA spot will be associated with the given number of protein bleaching steps. The interpretative cartoons in the centre are based on the model in Fig 2B. The top cartoon, based on the upper-left histogram, shows the majority of SRSF1-pre-mRNA complexes contain a single molecule of SRSF1, while U1 snRNPs are in grey to indicate that the data relate only to SRSF1; the lower cartoon, based on the upper-right histogram, shows that most U1/pre-mRNA complexes contain only one U1 snRNP, while SRSF1 is in grey to indicate that the data relate only to U1 snRNPs; the third cartoon, based on the lower two histograms, shows that complexes shown to contain both U1 snRNPs and SRSF1 contain only a single molecule of each component.
- B Complexes formed in NE3, as in (A) but with GloC pre-mRNA containing exons of only 2 nts. The error bars and interpretative cartoons are as described for panel (A).
- C Frequencies of mEGFP-SRSF1 and mCherry-U1A in two-way colocalized complexes formed in the absence of pre-mRNA. Error bars are as described in (A). The inferred complex is shown in the cartoon.

pre-mRNA-associated complexes containing two U1 snRNPs (Fig 4B), which was estimated for GloC at 92% (χ^2 , $P_{2,U1} = 0.095$) and for BGSMN2 at 100% (χ^2 , $P_{2,U1} = 0.39$). In accordance with the hypothesis, there was also a corresponding increase in the proportion of complexes containing two or more molecules of SRSF1 (Fig 4B), estimated at 67% for GloC.

A characteristic property of exon definition is the recruitment of components to the upstream 3'SS. To test whether the appended 5'SS was involved in exon definition reactions, its effect on the recruitment of 3'SS components was analysed by single-molecule methods with nuclear extracts from cells expressing either mEGFP-U2B'' or both mEGFP-U2AF65 and mCherry-U2AF65 (Chen *et al*, 2017) and compared with contemporaneous results with the same extracts on BGSMN2 and a derivative carrying four repeats of an SRSF1-dependent enhancer (Jobbins *et al*, 2018). For all three components, the 3'-terminal 5'SS resulted in an increase in the proportion of transcripts bound, the increase being greatest for U2 snRNP and least striking for U2AF65 (Table 2; Appendix Fig S7). This is consistent with previous results showing that U2 snRNP recruitment is affected more strongly than U2AF35 and U2AF65 by the difference in sequence between SMN1 and SMN2 (Chen *et al*, 2017) or the presence of *cis*- or *trans*-acting enhancers (Martins de Araujo *et al*, 2009; Smith *et al*, 2014; Jobbins *et al*, 2018). Importantly, only a single molecule of each component was bound, showing that the 3'-terminal U1 snRNP/SRSF1 complex did not recruit extraneous or non-specific components. We conclude that the 3'-terminal 5'SS mediates increases in the recruitment of single molecules of U2AF and U2 snRNPs equivalent to those produced by four repeats of a strong enhancer, showing that it is functional in the key reaction of exon definition, and that U1 snRNP binding to it recruits a molecule of SRSF1.

SRSF1 binds directly to U1 snRNP on stem-loop 3

The results so far have shown that SRSF1 forms a heterodimer with U1 snRNP and is recruited by it to 5'SSs. To test whether the interaction between U1 snRNP and SRSF1 could be direct, we titrated ^{13}C ILV-labelled SRSF1 Δ RS with *in vitro* reconstituted U1 snRNP using NMR spectroscopy (Fig 5A), as we did previously for other splicing factors (Campagne *et al*, 2019, 2021; Jutzi *et al*, 2020). SRSF1 Δ RS lacks the RS domain (amino acids 198–248) and is more soluble than the full-length protein. It is fully active in 5'SS selection and can restore splicing of pre-mRNAs with strong 3'SS in S100 extracts (Eperon *et al*, 2000; Zhu & Krainer, 2000). Single-molecule analysis in complex A conditions showed that the predominant complexes

contained a single molecule of SRSF1 Δ RS on GloC RNA and two on BGSMN2+ESE-Ax4, respectively (Appendix Fig S8). Thus, the truncated protein maintains at least some of the U1 snRNP association properties of full-length SRSF1. Surprisingly, the methyl groups of SRSF1 Δ RS experienced chemical shift perturbations (CSP) upon addition of *in vitro* reconstituted U1 snRNP. These CSP saturated at ratio 1:1 (SRSF1 Δ RS:U1snRNP) and were very similar to the ones observed previously when SRSF1 Δ RS was titrated with its single-stranded RNA (ssRNA) target sequence containing a CA and a GGA-binding site for RRM1 and RRM2, respectively (Fig 5B). This suggests that SRSF1 Δ RS can bind U1 snRNA in a similar fashion. The analysis of the structure of U1 snRNP (Pomeranz Krummel *et al*, 2009; Kondo *et al*, 2015) and the sequence of the U1 snRNA revealed that stem-loop 3, which is solvent accessible and free of any U1-specific protein, contains the CA and GGA motifs with the CA being in the loop and the GGA motif in the adjacent double-stranded region (Clery *et al*, 2013, 2021). The possibility that stem-loop 3 would be the target site for SRSF1 Δ RS was tested by using U1 snRNA stem-loop 3 alone. This reproduced the same methyl CSP (Fig 5C), indicating that SRSF1 Δ RS binds U1 snRNP stem-loop 3 with RRM1 bound to the CA motif at the 5' side of the loop and RRM2 binding the GGA sequence in the 3' side of stem (Fig 5D and E).

This interaction was further investigated in the context of the SRSF1 Δ RS-SL3 complex, which has a smaller molecular weight and allows structural studies in solution. Using ^{15}N -labelled protein, we monitored the binding of SRSF1 Δ RS on the U1 snRNA SL3 (Fig 6A). SRSF1 Δ RS experienced amide CSP on both RNA-binding domains, similar to the ones observed when the protein was titrated with ssRNA. The chemical shifts of the H5-H6 of pyrimidines were also monitored upon addition of SRSF1 and CSP occurred mainly on the pyrimidines of the double-stranded region (U95-C100) and on the C101 from the loop (Fig 6B). In contrast, the binding of FUS RRM on SL3 induced CSP on U105 and U107 from the 3' part of the loop (Jutzi *et al*, 2020). Altogether, the data suggest that RRM1 binds to the CA motif at the 5' side of the loop and RRM2 to the GGA sequence in the 3' side of the stem. Using isothermal titration calorimetry, we determined a dissociation constant of $10.9 \pm 2.8 \mu\text{M}$ at 27°C that is roughly 200-times weaker than the affinity of SRSF1 Δ RS for its ssRNA target (Appendix Fig S9; (Clery *et al*, 2021)). Since the GGA motif is embedded in the upper part of SL3 and RRM2 is a ssRNA binder, it seems likely that SL3 has to be melted in order to be bound by RRM2. To validate this hypothesis, we monitored the imino proton signals of SL3 at 40°C upon addition of SRSF1 Δ RS using a highly sensitive 1D ^1H SOFAST based on an imino-selective

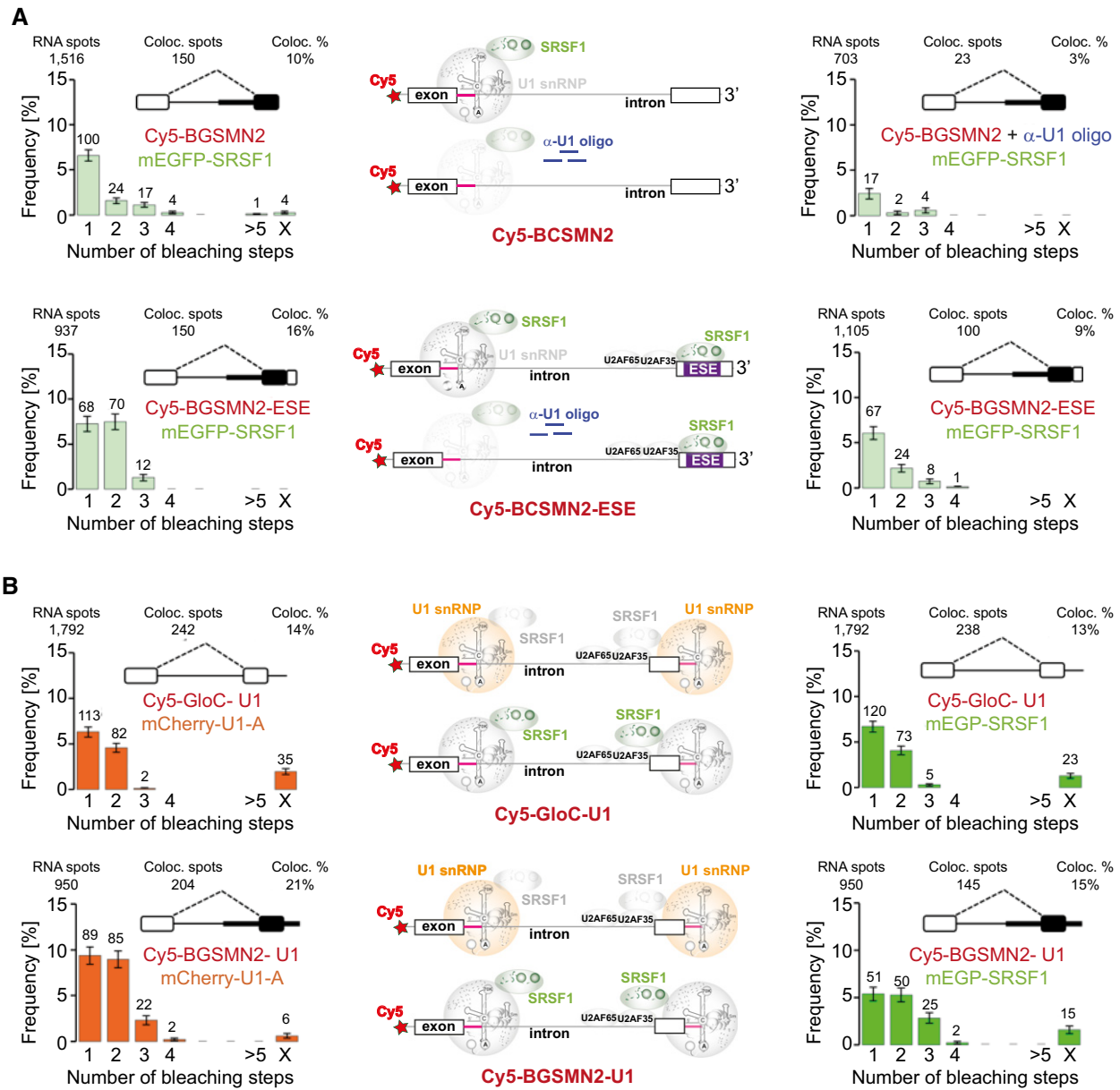


Figure 4. A 3'-terminal ESE and a 3'-terminal 5' SS both enable recruitment of an additional and independent molecule of SRSF1.

A Effects of 3'-terminal ESE repeats. Top left, two-way colocalization of complexes formed in NE2 on BGSMN2, a chimeric globin/SMN2 pre-mRNA. The heavy black line and box indicate the intron and exon sequences of SMN2 exon 7. Top right, the effects of inclusion of a 2'-O-methyl oligonucleotide complementary to the 5' end of U1 snRNA. The interpretative cartoons (top centre) show (upper) that most SRSF1/pre-mRNA complexes contain a single molecule of SRSF1 and (lower) that the proportion of SRSF1-associated pre-mRNA molecules was reduced sharply by occlusion of the 5' end of U1 snRNA by the oligonucleotides. Below, the corresponding experiments after addition of four tandem repeats of a strong ESE to the 3' terminus (Jobbins et al, 2018). Bottom left, most SRSF1/pre-mRNA complexes contain two molecules of SRSF1. In the interpretative cartoon in the centre (upper), this is shown by association of one SRSF1 with the U1 snRNP and the other with the ESEs (blue). We have shown previously that the activity of ESEs is limited by their occupancy, consistent with transient binding by SRSF1 and stabilization of one molecule by direct interactions based on diffusional encounters with splice site-associated factors, since their effects are not blocked by the incorporation of intervening non-RNA linkers (Jobbins et al, 2018). Bottom right, double occupancy of the pre-mRNA was greatly reduced by occlusion of the 5' end of U1 snRNA, suggesting that the U1-dependent but not ESE-dependent association of SRSF1 was lost (lower cartoon). The error bars in the histograms are the square root of the variance of the binomial probability that an RNA spot will be associated with the given number of protein bleaching steps.

B Effects of a 3'-terminal 5' SS. Two-way colocalizations of Cy5-pre-mRNA with mCherry-U1A (left) or mEGFP-SRSF1 (right) of complexes formed in NE3 on GloC pre-mRNA after addition of a 3'-terminal 5' SS (upper panel; c.f. Fig 3B) and on BGSMN2 with a 3'-terminal 5' SS (lower panel; c.f. Fig 4A and Jobbins et al, (Jobbins et al, 2018)). The interpretative cartoons in both cases show that the majority of pre-mRNA/U1 complexes and pre-mRNA/SRSF1 complexes (upper and lower members of each pair) contain two molecules of U1 and SRSF1, respectively. The presence of the components shaded in grey is inferred from the preceding experiments. Error bars are as described in (A).

Table 2. Effects of a downstream 5' SS or ESE sequences on recruitment of 3' SS components.

	mEGFP-U2AF65	mCherry-U2AF35	mEGFP-U2B''
BGSMN2	10	12	10
BGSMN2-ESEx4	16	25	25
BGSMN2-U1	14	20	27

Values show the percentage of pre-mRNA molecules in each single-molecule experiment colocalized with labelled U2AF and U2 snRNP, as indicated. The two components of U2AF were co-expressed and measurements made in the same experiment; mEGFP-U2B'' was expressed separately (Chen *et al*, 2017). BGSMN2-ESEx4 is BGSMN2 with a 3' extension containing four repeats of the enhancer used in BGSMN2-ESE (Fig 4A). The data for this and BGSMN2 are from Jobbins *et al*, (2018). The data for BGSMN2-U1 are shown in Appendix Fig S7. For each of the two extracts, the experiments with all three pre-mRNA substrates were done at the same time in the same nuclear extract.

excitation pulse ((Nikolaev *et al*, 2019); Fig 6C). The data showed that several imino signals in the upper part of the stem including U97 and G109 strongly decreased in intensity upon addition of SRSF1 Δ RS, suggesting that binding of the protein on SL3 induces the melting of the upper part of SL3. The competition between the formation of the base pairs and the binding of RRM2 on the GGA could explain why SRSF1 Δ RS has a low affinity for SL3 *in vitro*.

Functional relevance of the SRSF1-U1 snRNP interaction

The functional importance of these two binding sites in stem-loop 3 was tested using an assay in which the loss of splicing caused by 5' SS mutations was suppressed by expression of mutant U1 snRNAs carrying complementary nucleotide changes in their 5' ends (Zhuang & Weiner, 1986; Roca *et al*, 2012). The 5' SS of *SMN1* exon 7 was mutated at position +4 (A₊₄U) to impair splicing mediated by the endogenous U1 snRNP (Fig 6D). To rescue splicing, a plasmid was co-transfected that expressed U1snRNA Ψ_{+5A} with the endogenous promoter. However, the splicing was only partially rescued (Fig 6D). Since the bulged adenine in position -1 strongly weakens the 5' SS of *SMN* exon 7 (Campagne *et al*, 2019), the compensating C₋₉U mutation was introduced into the suppressor U1 snRNA, and the combination of both mutations rescued the splicing of *SMN1* exon 7. Mutating both putative binding sites of SRSF1 on stem-loop 3 resulted in a strong reduction of the activity of the suppressor U1 snRNA (Fig 6D), consistent with a functional role in splicing for these two SRSF1-binding sites in U1 snRNP stem-loop 3. Interestingly, the SL3 mutations strongly reduced the binding of SRSF1 *in vitro* while the binding of FUS remained unchanged (Appendix Fig S9). The importance of SL3 for splicing was shown previously in oocytes (Hamm *et al*, 1990), but we describe here the first evidence that it binds SRSF1, a central determinant of splicing activity, and propose a structural model for this interaction (Fig 6E).

SRSF1 cross-links preferentially with U1 stem-loop 3 in cells

To confirm that SRSF1 interacts specifically with SL3 *in vivo*, we analysed the distribution of cross-linking sites detected by eCLIP (Van Nostrand *et al*, 2020). To exclude the possibility that an

apparent preference for SL3 might simply reflect non-specific binding to the most accessible portions of the RNA, we compared this distribution with that of the other 146 RNA-binding proteins (Van Nostrand *et al*, 2020). The heat map in Fig 6F shows that cross-links to SRSF1 are concentrated in the loop of SL3 and to the 3' side of the stem, in perfect agreement with our NMR study and our splicing assay, and very few cross-links elsewhere. While other proteins show cross-links to the loop of SL3, relatively few have the cross-links on the 3' side of the stem (Appendix Fig S10). The only other regulatory proteins known at present to bind U1 snRNA are PTBP1 and FUS. PTBP1 shows most cross-links in the region of U1 SL4, its known binding site (Sharma *et al*, 2011), supporting the validity of this analysis. FUS binds SL3 at the UGU sequence in the 3' part of the loop (Jutzi *et al*, 2020), in between the CA and GGA motifs implicated here for SRSF1, consistent with the eCLIP data. The significance of the prominent SL4 cross-links is unknown but may originate from binding of the RS domain. The only proteins with cross-link distributions similar to SRSF1 are SRSF7 and SRSF9, in the adjacent tracks in Appendix Fig S10. We conclude that SRSF1 interacts with SL3 *in vitro* and *in vivo* and that this may be at least part of the reason why SL3 is essential for splicing.

Discussion

A new model for SRSF1 recruitment to pre-mRNA

The conventional model for the actions of SRSF1 on pre-mRNA splicing involves binding of the SR protein to exonic splicing enhancer sequences and the recruitment of U1 snRNP at the 5' SS and U2-associated factors at the 3' SS (Fig 7A). This is consistent with extensive data from transcriptome-wide analyses of binding sites by cross-linking, which have revealed that these are enriched upstream of 5' splice sites and, in some cases much more strongly, at the 5' end of exons just downstream of the 3' splice sites (Jamison *et al*, 1995; Sanford *et al*, 2009; Pandit *et al*, 2013; Anczukow *et al*, 2015; Bradley *et al*, 2015; Muller-McNicoll *et al*, 2016). The results we describe here show that there is another and completely different mode by which SRSF1 associates with pre-mRNA splicing complexes.

By using single-molecule methods that distinguish between complexes containing different numbers of molecules of SRSF1, we have been able to show that SRSF1 can associate in a defined stoichiometry with pre-mRNA in a process that is independent of exons but dependent on U1 snRNPs and 5' SSs. This occurs whether the 5' SS is able to participate in splicing or is at the 3' end of an exon and participates in exon definition (Fig 7B). Even in the absence of pre-mRNA, significant proportions of SRSF1 molecules and U1 snRNPs are associated in a stoichiometric complex, from which we infer that the binding of U1 snRNP to a 5' SS concomitantly recruits SRSF1. Our results are consistent with previous evidence showing that pure SRSF1 associates directly with U1 snRNPs (Kohtz *et al*, 1994; Xiao & Manley, 1997; Cho *et al*, 2011), forming stable complexes in the presence of a 5' SS (Jamison *et al*, 1995) and, moreover, evidence from immunoprecipitation that U1 snRNPs and SRSF1 are associated *in vivo* (Ellis *et al*, 2008; Chi *et al*, 2018; Huttlin *et al*, 2021), independently of RNA (Ellis *et al*, 2008). Further support for the

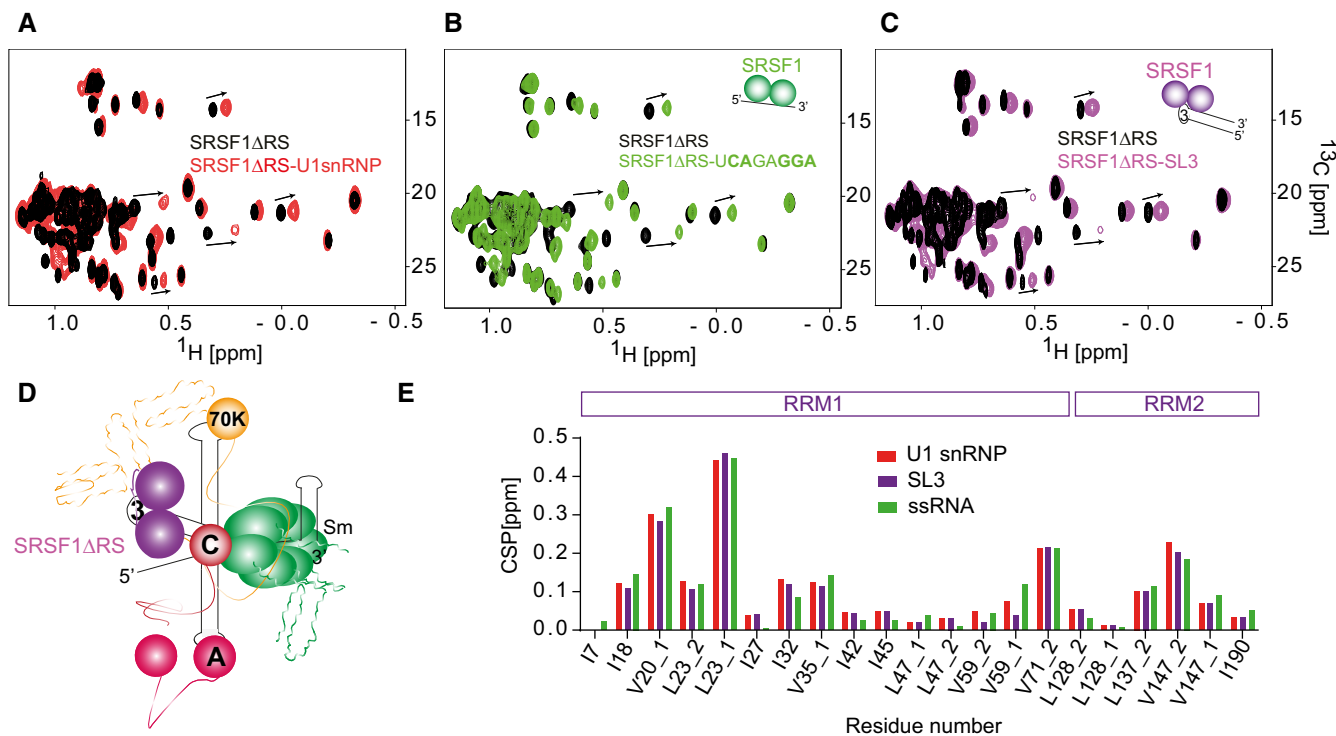


Figure 5. Direct interaction between SRSF1 and U1 snRNP.

- A Overlay of the 2D ^{13}C - ^1H HMQC of ^{13}C ILV-labelled SRSF1 Δ RS before and after addition of 1 molar equivalent of *in vitro* reconstituted U1 snRNP.
 B Overlay of the 2D ^{13}C - ^1H HMQC of ^{13}C ILV-labelled SRSF1 Δ RS before and after addition of 1 molar equivalent of the ssRNA target of SRSF1.
 C Overlay of the 2D ^{13}C - ^1H HMQC of ^{13}C ILV-labelled SRSF1 Δ RS before and after addition of 1 molar equivalent of U1 snRNA stem-loop 3.
 D Schematic representation of SRSF1 Δ RS bound on U1 snRNP.
 E Bar plot showing the chemical shift perturbations of the SRSF1 methyl groups when the protein was titrated with U1 snRNP, ssRNA or U1 snRNA SL3.

existence of this interaction *in vivo* comes from FRET, which showed that the U1 70K protein is in close proximity to SRSF1, even in the absence of transcription (Ellis *et al*, 2008).

Implications of SRSF1 binding to U1 snRNA SL3

The interaction of U1 snRNPs with SRSF1 has previously been attributed to protein–protein interactions between either the RS domain of U1-70K, the snRNP protein that binds stem-loop 1 of the U1 snRNA, and the phosphorylated RS domain of SRSF1 (Xiao & Manley, 1997) or between RRM domains of the same proteins (Cho *et al*, 2011). The interaction with SL3 was unexpected, although we have also shown recently that SL3 is bound by the FUS protein (Jutzi *et al*, 2020). Interestingly, SL3 is much longer in yeast U1 snRNPs and it has been proposed to act as a binding site for regulatory proteins (Li *et al*, 2017). The interaction of SRSF1 with bases in the stem of SL3 is not unprecedented, since the regulatory protein PTBP1 interacts with bases in the stem of U1 snRNA stem-loop 4 (Sharma *et al*, 2011). Given that the RNA-binding domains of SRSF1 interact only weakly with ESE sequences (Cho *et al*, 2011; Anczukow *et al*, 2015; Jobbins *et al*, 2018), it seems likely that the SL3 interaction is augmented by protein interactions. These probably involve the RS domain of SRSF1, since no protein interactions were detected by NMR spectroscopy with the SRSF1 Δ RS protein

(Fig 4A). In addition, the unwinding seen in the upper part of the stem suggests the possibility that a helicase is involved. Candidates for this activity might be UAP56 (DHX39B), which is present in the earliest complexes (Fleckner *et al*, 1997; Shen *et al*, 2008) and has also been found to interact with SL3 (Martelly *et al*, 2021), or DDX5, which is involved in the use of 5'SS and U1 snRNP binding or dissociation (Liu, 2002; Guil *et al*, 2003; Lin *et al*, 2005a; Kar *et al*, 2011). SRSF7 and SRSF9 show similar cross-linking distributions on U1 snRNA, which suggests that they may also interact with SL3 (Appendix Fig S10). Other proteins interact with SL3, but most of them do not interact with the 3' side of the stem and therefore may not compete with SRSF1. However, there are several that do seem to have the potential capability. It should be noted, though, that SRSF1 is present at a several-fold higher concentration than any potential competitors in HeLa cells (Hein *et al*, 2015). This does not mean that all functional competitors also compete for SL3; hnRNP A1, for example, is a functional competitor with SRSF1 (Mayeda & Krainer, 1992; Mayeda *et al*, 1993; Caceres *et al*, 1994; Eperon *et al*, 2000; Zhu *et al*, 2001) but it does not interact with the U1 snRNP (Eperon *et al*, 2000) and accordingly does not show specific binding on U1 snRNA. It seems probable that U1 snRNA is a target for a number of RNA-binding proteins and that their competition would be a major determinant of splicing patterns and efficiency in mammalian cells.

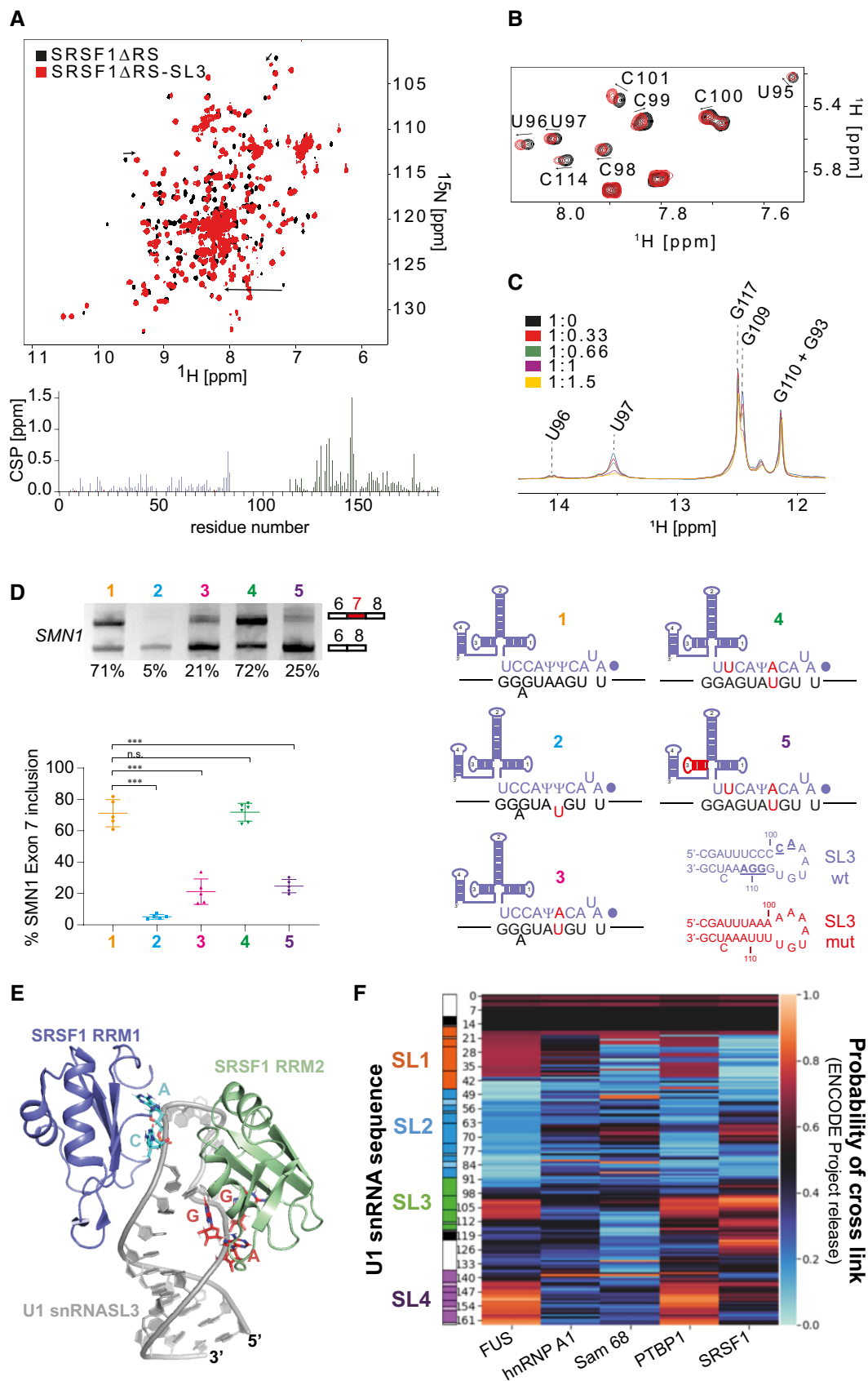


Figure 6.

Figure 6. SRSF1ΔRS binds U1 stem-loop 3 SL3 at nucleotides that are essential for the function of U1 snRNA.

- A Overlay of the 2D ^1H - ^{15}N HSQC spectra of SRSF1ΔRS before (black) and after addition of one equimolar amount of U1 snRNA SL3 (red). Bar plot showing the chemical shift perturbations (CSP) as a function of the residue number. Data corresponding to RRM1 and RRM2 are coloured in blue and green, respectively.
- B Overlay of the 2D ^1H - ^1H TOCSY spectra of U1 snRNA SL3 before (black) and after (red) addition of SRSF1ΔRS.
- C Overlay of the 1D ^1H SOFAST imino recorded upon successive addition of SRSF1ΔRS. The spectra are coloured according to the SL3:SRSF1ΔRS ratio.
- D Suppression of a 5'SS mutant by a complementary U1 snRNA is reduced by mutations of the nucleotides in U1 SL3 predicted to be bound by SRSF1. The diagram shows base-pairing between the 5' end of U1 snRNA and SMN1 exon 7. (1) Wild-type sequences, with the dot representing the 2,2,7-trimethylguanosine cap at the 5' end of U1 snRNA and Ψ representing pseudouridine; a bulged A is shown. (2) A mutated nucleotide in the SMN1 exon 7 minigene that impairs base-pairing with U1 snRNA is shown in red. (3) A compensating mutation was made in the gene encoding U1 snRNA. (4) An additional mutation was made in the U1 snRNA gene to improve base-pairing with the SMN1 exon 7 5'SS. (5) The U1 gene containing two mutations that allow efficient suppression of the mutated exon 7 5'SS was mutated further in SL3, as shown by the sequences of SL3 below, to alter those nucleotides that interact with SRSF1 (underlined in the SL3 wt sequence) without affecting the base-pairing in SL3. Minigenes expressing the wild-type and mutated sequences of SMN1 exon 7 and U1 snRNA were transfected into HEK293T cells, and the level of inclusion of SMN1 exon 7 was detected by RT-PCR. The products were separated by agarose gel electrophoresis and detected by ethidium bromide. The scatter plot shows the percentage of exon 7 inclusion for the 5 conditions. The data points are shown as well as the mean and the standard deviation. One-way ANOVA test was used to probe the significance of the data, and *** indicates that $P < 0.005$ ($n = 5$ biological replicates) while n.s. stands for non-significant.
- E Structural model of the SRSF1ΔRS-SL3 complex.
- F Heat map showing the distribution of ENCODE cross-links of selected proteins along U1 snRNA. The scale bar on the right-hand side shows the probability that the cross-links are enriched at a given position. The sequence of U1 snRNA, showing the stem-loops, is aligned on the l.h.s.

Alternative modes of SRSF1 association explain the functional equivalence of ESEs and 5'SS in exon definition

The SRSF1 molecule recruited by U1 snRNPs might be expected to have roles in either definition of the upstream exon or in splicing of the downstream intron. The effects on exon definition of a downstream 5'SS on short internal exons (Hwang & Cohen, 1996a; Fox-Walsh *et al*, 2005; Erkelenz *et al*, 2021) are similar to those of an ESE (see above). Indeed, there is a negative correlation between the strength of 5'SS sequences and the density of ESEs in the flanking exon sequences (Kralovicova & Vorechovsky, 2007; Caceres & Hurst, 2013). Previously, the combination of ESEs and splice site strength would have been interpreted as a means to ensure a minimum probability of U1 snRNP binding; in the light of our finding that the U1 snRNP recruits SRSF1, we might speculate instead that ESEs and U1 snRNP binding are also alternative ways to recruit SRSF1 for exon definition. This possibility is strengthened by the absence of any evidence for direct contacts by U1 snRNP in cross-exon interactions in mammals. While U1 snRNP does produce an increase in U2AF65 binding to the upstream 3'SS (Grabowski *et al*, 1991; Kreivi *et al*, 1991; Hoffman & Grabowski, 1992; Cote *et al*, 1995; Izquierdo *et al*, 2005; Singh *et al*, 2007; Palacino *et al*, 2015), the intermediate interactions across the exon have not been identified. Our findings raise the possibility that these interactions are mediated by U1-associated SRSF1 that interacts directly with U2AF (Wu & Maniatis, 1993) or U2 snRNP (Akerman *et al*, 2015), exactly as would an ESE-bound SRSF1 molecule (Fig 6). Direct interactions are also consistent with our finding that only one SRSF1 is recruited in complex A and there is no evidence of cooperative association of additional molecules of SRSF1 across the exon. Wu and Maniatis proposed such a bridging role for SRSF1 in cross-intron interactions (Wu & Maniatis, 1993), but ESE-based models have predominated in discussions of exon definition.

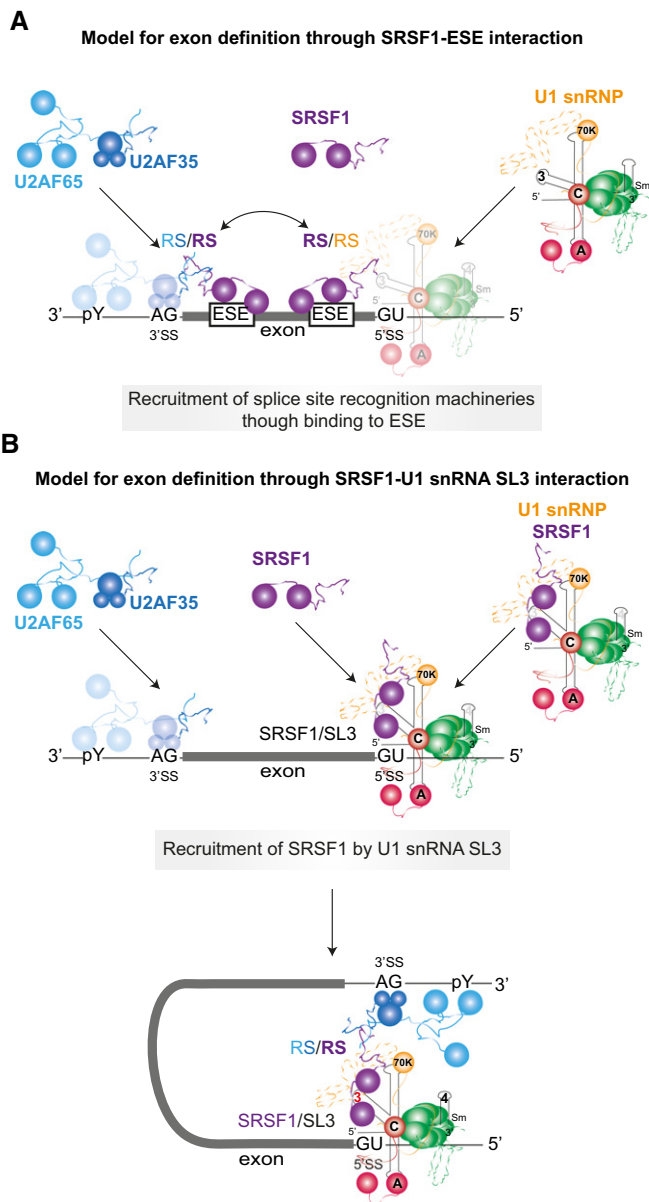
The uncertain role of U1 snRNP in splicing a downstream intron

A fundamental structural similarity between exon-bridging and intron-bridging splicing complexes has been proposed in yeast, where Prp40 contacts the U1 snRNP and the branchpoint-binding protein (Abovich & Rosbash, 1997; Li *et al*, 2019). Whether, by

analogy, the recruitment of SRSF1 by U1 snRNPs in mammals is important for splicing of the downstream intron is unclear. Surprisingly, it is not yet known whether the U1 snRNP itself has any direct role. The U1 snRNP was the first splicing factor discovered, and it has well-characterized roles in the selection of 5' splice sites (Roca *et al*, 2013). Its status as a splicing reaction component was originally in doubt because of its weak association with splicing complexes (Konarska & Sharp, 1986), but it is now seen as a stoichiometric component (Will & Luhrmann, 2011; Roca *et al*, 2013) until it is displaced during the formation of complex B (Charonton *et al*, 2019). It has been suggested that it forms cross-intron interactions via association with the proteins SF3a (Sharma *et al*, 2014), Prp5 (Xu *et al*, 2004; Shao *et al*, 2012) or Prp40 (Abovich & Rosbash, 1997), although none of these have been shown to bind as a single molecule and span the intron. The possibility that the U1 snRNP plays a direct role in the splicing reactions is weakened by three observations. First, it has been shown to enable splicing even when binding some nucleotides away from the 5' splice site (Hwang & Cohen, 1996b; Brackenridge *et al*, 2003). The exact position of the 5' splice site is determined by U6 snRNA base-pairing (Hwang & Cohen, 1996b; Brackenridge *et al*, 2003; Hang *et al*, 2015). Second, the U4/U6.U5 tri-snRNP can bind directly to a 5' splice site, independently of U1 snRNP and even in competition with it (Konforti *et al*, 1993; Konforti & Konarska, 1994; Maroney *et al*, 2000). Interestingly, the recruitment of the tri-snRNP is enhanced by SR proteins (Roscigno & Garcia-Blanco, 1995). Third, the splicing of some introns has been found to be independent of the U1 snRNP (Crispino *et al*, 1996; Fukumura *et al*, 2009; Raponi *et al*, 2009). Remarkably, some introns that are U1-dependent have been shown *in vitro* to splice without U1 snRNP if the concentration of SR proteins is increased (Crispino *et al*, 1994; Tarn & Steitz, 1994). It can be inferred that the U1 snRNP might not play a direct role in all splicing reactions but could also act upstream of SRSF1 in the reaction pathway.

SRSF1, recruited by U1 snRNP, might play similar roles in exon definition and splicing of the downstream intron

Evidence for a direct role of SRSF1 in splicing is limited, but not inconsequential. As noted above, SR proteins can substitute for U1



snRNP or recruit the tri-snRNP, and the catalytic reactions of splicing require dephosphorylation of SRSF1. When SRSF1 was added to an S100 cytoplasmic extract to restore splicing, it was found that its RS domain contacted the 5'SS in complexes B and C and that it was required for U6 snRNA base-pairing to the 5'SS (Shen & Green, 2004). The addition of SRSF1 was not required either when an RS domain was tethered by MS2 near to the 5'SS (Shen & Green, 2004) or when the base-pairing between U6 snRNA and the 5'SS had been improved (Shen & Green, 2007). Dephosphorylation of SRSF1 is required after assembly of complex B (Cao *et al*, 1997) which, like the cross-linking results (Shen & Green, 2004), is consistent with the presence of SRF1 in the mature spliceosome. Most strikingly, SRSF1 RRM2 is visible as a discrete component and structure in pre-B^{act} spliceosomal complexes, which form after the U1 snRNP has been displaced (Townsend *et al*, 2020). These observations suggest that

Figure 7. Mechanisms by which SRSF1 and U1 snRNPs collaborate in pre-mRNA splicing.

- A** The ESE-dependent activity of SRSF1. SRSF1 probably binds transiently to the ESE and is in a dynamic equilibrium with the free protein until it collides by 3D diffusion with a U2AF/U2-pre-mRNA complex (Jobbins *et al*, 2018). Where binding of the 3' splice site components would otherwise be weak, these interactions enhance the stability of binding of both SRSF1 and the 3' complex. Similarly, the binding of SRSF1 to an ESE may enhance U1 snRNP recruitment.
- B** Recruitment of SRSF1 by interaction with U1 SL3 (this paper) may enable cross-exon interactions (exon definition) and cross-intron interactions. SRSF1 is not recruited to the pre-mRNA by its interaction with an ESE but by its interaction with the U1 snRNA, which thereby brings SRSF1 to the 5' splice site. This recruitment of SRSF1 may enable it to enhance the use of an upstream 3'SS in the same way as it does when bound to an ESE. SRSF1 recruited by a U1 snRNP may also mediate cross-intron interactions with the same partners but at the downstream 3'SS during splicing reactions (Wu & Maniatis, 1993). In addition, in both cross-intron and cross-exon configurations, U1 snRNP-bound SRSF1 may recruit the tri-snRNP to the intron definition or exon definition A complexes (Roscigno & Garcia-Blanco, 1995; Schneider *et al*, 2010). Competition by other proteins for SL3 might be the basis for many examples of regulation.

SRSF1, recruited by a U1 snRNP to the 5'SS, which we demonstrate here, might play similar roles in constitutive splicing reactions and exon definition. First, interactions with U2AF would establish connections across the intron or exon, respectively. In addition, SRSF1 might in both cases recruit the tri-snRNP. This would explain its persistence in splicing complexes B or even C, after the departure of U1 snRNPs. Moreover, exon definition complexes also contain tri-snRNPs and can assemble complex B directly (Schneider *et al*, 2010). If binding to SL3 and SL4 is the focus of competition among splicing regulatory proteins, then it is possible that much regulation occurs at the point of tri-snRNP recruitment. Our finding that SRSF1 binds U1 snRNA SL3 and is recruited by U1 snRNPs supports the possibility that SRSF1 is playing much wider roles in pre-mRNA splicing than just the binding of exonic splicing enhancers.

Materials and Methods

Sequences

Most of the transcripts used were derivatives of rabbit β -globin exon 2-exon 3, with a shorter intron and truncated 3' exon, as described previously (Skordis *et al*, 2003; Hodson *et al*, 2012). The exons in the pre-mRNA with 2 nts in each exon comprised (5') GG and (3') GA. The globin-SMN2 exon 7 transcripts were also as described (Skordis *et al*, 2003; Smith *et al*, 2014), but transcription terminated at nucleotide 48 in SMN2 exon 7 so as to exclude the exonic part of the 5' splice site.

Oligonucleotide annealing

The transcription of RNA substrates and the subsequent annealing of the fluorescent oligonucleotide β -5'-Cy5 (Cy5-UAGACAACCA GCAGCC C-biotin, 2'-O-methyl/LNA) were done as described (Hodson *et al*, 2012). All labelled constructs were run on a 6% polyacrylamide gel and imaged using a Typhoon imager (GE Healthcare)

to ensure that the level of free fluorescent oligonucleotides was below 2%.

Labelling RNA at the 5' end

For Figs 2C and 3, the substrate RNAs were labelled using 5' Cy5-maleimide (Ohrt *et al*, 2012). RNA was transcribed in the presence of 10 mM guanosine-5'-*O*-monophosphorothioate (BioLog) instead of cap analogue. Following transcription, the mixture was incubated with RNase-free DNase and extracted thrice with phenol–chloroform. After purification of the RNA by gel filtration on S-300 (GE Healthcare), the RNA was precipitated with ethanol and incubated for 4 h with 0.5 mM cyanine5-maleimide (Lumiprobe 13080) at ambient temperature. The RNA was purified by high resolution denaturing polyacrylamide gel electrophoresis.

Splicing and analysis of complexes

Following transfection of HeLa cells with plasmids expressing mEGFP-SRSF1 (with or without the addition of mCherry-U1A), the cells were cultured for about 48 h and extracts were prepared. Aliquots of the extracts were made and stored at -80°C . The activities of the extracts in splicing and complex assembly were analysed as described previously (Hodson *et al*, 2012). Proteins were detected by Western blotting, using primary antibodies against GFP (Protein-Tech 66002) or SRSF1. Fluorescent secondary antibodies were used against mouse IgG (IRDye 680LT, LI-COR 926-68020) and rabbit IgG (IRDye 800CW, LI-COR 926-32211). The concentration standard was recombinant GFP. The ratios of labelled and endogenous proteins in the various extracts are shown in Table 1. The total concentrations of SRSF1 (labelled and endogenous) in NE1 and NE2 are around 1.7 and 1.2 μM , respectively, and the respective concentrations in the splicing reactions would be 0.8 and 0.6 μM . This is comparable with a previous estimate of 1.3 μM for the concentration of SRSF1 in a splicing reaction *in vitro* (Mayeda *et al*, 1993).

Sample preparation

For reactions containing pre-mRNA labelled with biotinylated oligonucleotides, coverslips (22 mm by 50 mm #1, Menzel-Gläser) were washed with distilled water, followed by purified water and sonicated in a water bath for 12 min and dried under a nitrogen stream. They were cleaned in an argon plasma (MiniFlecto-PC-MFC, Gala Instruments) five times for 5 min each with pure argon at 0.15 mbar and an applied power of 80 W. Double-sided tape was used to create a channel of width 5–10 mm parallel to the coverslide, which was covered by another coverslip, 21 mm by 26 mm #1, to form the sample chamber. The sample chamber was incubated with 20 $\mu\text{g}/\text{ml}$ biotin-BSA in PBS for 10 min, washed with buffer A (100 mM NaCl, 50 mM HEPES pH 7.5, 1 μM DTT, 20 units/ml RNase OUT [Invitrogen]), incubated with 10 $\mu\text{g}/\text{ml}$ streptavidin (Invitrogen) in PBS and again washed with buffer A.

All samples were prepared with 50% nuclear extract, 3.2 mM MgCl_2 , 50 mM monopotassium glutamate and 3 units RNase OUT (Invitrogen) for incubations without ATP, and for incubations in the presence of ATP 20 mM phosphocreatine, 1.5 mM ATP and 20 mM HEPES pH 7.5 were included. To halt splicing at complex A, either 1 μM of a 2'-*O*-methyl oligonucleotide complementary to U6 or 570

nM anacardic acid was added, as described (Hodson *et al*, 2012). U1 snRNA was inhibited by the inclusion of a complementary 2'-*O*-methyl oligonucleotide (GCCAGGUAAGUAU–biotin) at 3.3 μM . PhosSTOP (Roche) was included as a broad-spectrum phosphatase inhibitor at a concentration of 2 \times . Samples were pre-incubated for 15 min at 30°C before pre-mRNA was added at a final concentration of 50–60 nM and incubated for a further 15 min at 30°C . Samples were diluted using buffer A, loaded into the sample chamber and incubated for 5 min. The sample chamber was then flushed with 2 mM protocatechuic acid (Spectrum Chemicals), 90 nM protocatechuic acid 3,4-dioxygenase (Sigma), 1 mM DTT, 0.16 units/ μl RNaseOut and incubated for a further 5 min.

For reactions containing 5' end-labelled RNA, coverslips were soaked in 1 M KOH for four hours before washing in distilled water, sonication and plasma cleaning as above. Biotinylated BSA and streptavidin were omitted. Buffer A was replaced with 40 mM HEPES pH 7.5, 3.2 mM MgCl_2 , 50 mM monopotassium glutamate, 50 mM KCl, 0.1 mM EDTA and 0.5 mM DTT. Complexes were adsorbed directly onto the glass surface. Although the adsorbed complexes are stable, transient interactions might be lost during dilution and the period of a few minutes during which complexes diffuse to the slide surface after loading. Experiments to test whether the interaction between mEGFP-SRSF1 and mCherry-U1A was dependent on pre-mRNA were done by including 1 μl of an RNase A/T1 mixture (Thermo Scientific; RNase A at 2 mg/ml and RNase T1 at 5 units/ μl) in a 10 μl reaction that otherwise reproduced splicing reaction conditions.

Data acquisition

An objective-based total internal reflection fluorescence microscope was used for all single-molecule experiments. The laser power incident on the objective was approximately 4 mW at 488 nm and 3 mW at 561 nm. Only the central part of the approximately Gaussian-shaped intensity profile of the laser beam was used, together with the central 250 by 250 region of the 512 by 512 pixels CCD chip to ensure a homogenous illumination across the field of view ($\leq 50\%$). The internally reflected beam was directed onto a quadrant photodiode where its displacement was measured. It was fed back to the microscopic stage to create a feedback loop to ensure the positional stability of the sample in the nanometre range. The acquired data were saved in the Tagged Image File Format (TIFF) as well as an accompanying text file containing all the information about the execution of the individual acquisition. A typical acquisition involved the successive collection of 50 frames with excitation at 633 nm, to identify Cy5-pre-mRNA spots, followed by 250 frames at 561 nm (mCherry, where collected) and 250 frames at 488 nm (mEGFP). Frames were collected at a rate of 20 s^{-1} . All fluorophores bleached completely during the appropriate stage. Between 50 and 100 fields were acquired from each channel on the coverslip.

Data analysis

The data were analysed with a MATLAB program. Spots were detected by creating two composite images for each wavelength by computing the mean and the maximum value for each pixel across all images of a wavelength. Spots were identified by moving a square, 19 pixels by 19 pixels, across the composite images to find

areas where the maximum pixel intensity inside the box was 1.119 times higher than the mean pixel intensity. The size of the box was reduced until only one peak of intensity was within this box, and rejected if the size was ultimately smaller than 9 pixels by 9 pixels. The intensity around the remaining peak was fitted with a 1D Gaussian for each dimension, with $0x$ at the centre:

$$G(x) = B + \frac{1}{\sqrt{2\pi}} \times \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right].$$

A spot was considered valid, if the fit yielded values for $0.9 \leq \sigma \leq 3.5$, $0 \leq B \leq 0.4$ for Cy5, and $0.3 \leq \sigma \leq 4.7$, $0 \leq B \leq 0.53$ for mEGFP and if $|\mu| < \text{size of the box} - 2.5$. For each dimension, μ was taken as the true centre of the spot. A linear transformation in each dimension to correct for chromatic aberration yielded the “true” position of each spot (x_t, y_t):

$$x_t = x + \frac{(x - x_c)}{SF_x}, \quad y_t = y + \frac{(y - y_c)}{SF_y}.$$

The values for the parameters x_c , SF_x , y_c and SF_y were calibrated using oligonucleotides labelled with both ATTO 488 and Cy5. The coordinates of the molecular spots were measured with excitation at 488 and 633 nm, and then, the data were fitted by a model in which divergence of the 488 nm signal from that at 633 nm is proportional to distance from a centre at which no divergence occurs. This produces the coordinates of the centre (x_c, y_c) and the scaling factors (SF_x, SF_y).

Two spots of different wavelengths were considered colocalized if their corrected separation was below or equal to two pixels. The probability of random background colocalization was calculated for each image and was generally around 1–2%. This was considered to be so small that it was not subtracted from the colocalization values.

Background-corrected intensity traces for each spot were calculated from the raw data by subtracting the mean intensity of all the pixels that were five pixels away from the centre from the mean intensity of the pixels found in a central box of three pixels by three pixels. Step detection was done recursively by a Bayesian step detector (Ó Ruanaidh & Fitzgerald, 1996) to identify the most likely change points for the original trace and each following sub-trace to the left and right of the identified change-point, until the likelihood of the change-point is below a set threshold. These change points were used to estimate the intensity plateaus of the trace, thus allowing any re-activation of fluorophores to be discounted. The change points were checked by visual inspection and assignments accepted or corrected. The number of plateaus detected minus one was taken as the number of fluorophores present. The validity of the assignment of the step numbers to the traces was checked for an experiment in which a substantial number of RNA molecules were deemed to be associated with two molecules of mEGFP-SRSF1 (Fig 4B, GloC-U1) by measuring the total number of photon counts emitted by mEGFP for each complex, and comparing the results from complexes assigned as containing one or two molecules of mEGFP-SRSF1 (Appendix Fig S11). The median number of photon counts emitted from complexes containing one molecule of mEGFP-SRSF1 was 1,472, whereas the median for complexes containing two was

3,085. The distribution between the two classes was significantly different by a two-tailed Mann–Whitney test (SPSS); $P_{1-2} = < 0.001$. As a check, the values from the complexes containing one molecule were added to the same values rearranged in a random order. The pseudo-double complexes had a median of 3,802, and the distribution was not significantly different from that of the *bona fide* complexes assigned as having two molecules ($P_{2-(1+1)} = 0.2$).

Statistics

Each frequency histogram shows the result of measurements collated from around 50 fields acquired from each sample. The error bars are the square root of the variance of the binomial probability that an RNA spot will be associated with the given number of protein bleaching steps. The expected distributions for binding by a single molecule were taken to be the same as those seen for bleaching in one or two steps under the same conditions but in the absence of pre-mRNA (Table 1). The distributions expected if there were binding by two molecules on pre-mRNA with duplicated splice sites were calculated based on the observed levels of dimerization in the absence of pre-mRNA and the proportions of protein labelled, as described previously (Hodson *et al*, 2012; Chen *et al*, 2017). Chi-squared tests included frequency classes with expected values ≥ 5 . The results were significantly different from the expected distributions. This could result from either pre-mRNA associated with only one or no molecules of SRSF1 or the existence of non-fluorescent (misfolded or pre-bleached) mEGFP. To avoid over-estimating binding, it was assumed that all molecules were fluorescent (previously, with an EGFP-PTB assumed to occupy all available pre-mRNA fully we estimated that $> 90\%$ were active) and the probability of occupancy and the total number of accessible molecules of pre-mRNA were allowed to vary. This enabled an estimation of the fraction of pre-mRNA molecules bound by SRSF1 that contained two molecules of SRSF1 by minimization of chi-squared. However, the extra variables precluded the use of chi-squared tests of significance. An additional complication was the existence of the significant background from molecules that did not form complex A. Poisson and geometric distributions were fitted to this background data by minimizing the value of chi-squared, and tests of significance were done on those categories in which 4 or more pre-mRNA molecules were expected using N-2 and N-3 (respectively) degrees of freedom.

With a pre-mRNA containing two introns (Fig 1B), we estimated that about as many SRSF1-pre-mRNA complexes contained two molecules of SRSF1 as contained one. This was roughly consistent with our previous measurements of U1 snRNP binding to the same pre-mRNA. This calculation was based on the proportion of SRSF1 protein that was tagged by mEGFP and on the observed pattern of dimerization in the absence of exogenous RNA. This procedure might underestimate the level of complexes containing two molecules of SRSF1, since a proportion of mEGFP molecules might not be actively fluorescent. Inactive mEGFP might result from several factors: (i) a significant proportion of mEGFP molecules might have entered a brief dark state, known as blinking; (ii) bleaching might have occurred before fluorescence was recorded; and (iii) a significant proportion of mEGFP molecules might be misfolded. Blinking (i) would not affect the results, since the dark state lifetime of mEGFP is only 1–2 s (Dickson *et al*, 1997; Garcia-Parajo *et al*, 2000;

Vamosi *et al.*, 2016). Spots were detected in our experiments using both the mean and maximum fluorescence, meaning that the total yield from a fluorophore is unimportant, and measurements were made for 15–30 s, so that the probability of being in a dark state throughout is negligible. The second factor, bleaching prior to recording or so early in recording that no mean level of initial fluorescence was measured, undoubtedly happens but is likely to affect a relatively small proportion of molecules because fluorescence was recorded continuously during the switch from illumination at 633 nm (for Cy5 excitation) to 488 nm (for mEGFP). The third factor, misfolding, is harder to estimate. It is usually estimated, when the proteins cannot be purified, by assuming that the population of complexes is homogeneous. In such a case, the distribution is analysed as we have done here and the difference from the expected distribution is attributed to misfolding. It would not be justified in the present work because there are no grounds for the *a priori* assumption of homogeneity; indeed, the question at issue is whether more than one molecule of SRSF1 is bound. Thus, we considered it safer to assume that misfolding was negligible and that the population was heterogeneous. In the first assay of this type, misfolding was estimated at 20% (Ulbrich & Isacoff, 2007), although in this case the EGFP sequence was expressed on the C-terminal side of the fusion. Misfolding of mEGFP is particularly likely in this location, since folding would be affected by any misfolding of the upstream sequence (Waldo *et al.*, 1999; Wang & Chong, 2003). In contrast, it has been reported that the maturation of N-terminal GFP fusions expressed in HEK293 cells was almost complete (Liesche *et al.*, 2015). The fluorescent protein moieties in mEGFP-SRSF1 and mCherry-U1A are N-terminal and expressed in mammalian cells. The absence of significant levels of dark mEGFP is consistent with the results in Fig 4. In Fig 4A, the results for BGSMN2-ESE are consistent with full binding by two molecules of SRSF1 (c.f. Appendix Fig S2), and in Fig 4B, the results suggest that double occupancy of 92% of GloC-U1 pre-mRNA molecules by U1 snRNPs is associated with double occupancy of 67% by SRSF1. Our estimate that about half of the complexes of SRSF1 with pre-mRNA containing two introns contained two molecules of SRSF1 in Fig 2A is therefore a minimum but probably close to the correct value.

An alternative explanation for the binding of two molecules of mEGFP-SRSF1 might be that pre-mRNA substrates with ESEs or an additional U1 snRNP-binding site might be more prone to dimerization. Analysis of the data for GloC-U1 in Fig 3B showed that about 10% of the Cy5 spots contained two molecules of pre-mRNA. If these were excluded from the analysis, the results were the same (by a chi-squared test, $P_{inc = exc} = 0.91$), and the possibility was not taken into account further.

A less distinct form of binding produces a significant background of molecules bound by more than one or two molecules of SRSF1. This was conspicuous when U1 snRNA binding was blocked (Figs 2B and 4A). In some cases, the residual distribution was geometric, suggesting that it might be accounted for by cooperative interactions, but in other cases, the pattern was less clear. There are several reasons for thinking that this background was present to some extent in many of the experiments and represented a separate class of molecules from those bound primarily by a single SRSF1 molecule. First, if the U1-dependent association involved binding of an additional molecule of SRSF1 to RNA molecules already bound by SRSF1 according to the background distribution, then the net effects

would be a shift of the distributions from n to $n + 1$. On the contrary, the U1-dependent increase is restricted to the complexes with one or two bleaching steps (according to the number of sites or introns). Second, we have shown (Jobbins *et al.*, 2018) that additional copies of the ESE on the 3' end of BGSMN2 (c.f. Fig 4A) increased splicing, increased complex A formation and reduced the background distribution. Third, U2AF35, U2AF65 and U2 snRNPs show a similar super-stoichiometric distribution after sequestration of U1 snRNPs or ablation of 3'SS signals (Chen *et al.*, 2017). Thus, the distributions that we observe probably result from the superimposition of distributions arising from two mutually exclusive processes: specific recruitment of SRSF1 molecules via U1 snRNPs or ESEs, and the non-stoichiometric background complexes representing molecules unable to form complex A.

Activity of SRSF1 fusions in alternative splicing

Plasmids expressing mEGFP-SRSF1, mCherry-SRSF1 or just the fluorescent proteins were transfected into HeLa cells. RNA was extracted after 48 h and analysed by reverse transcription and PCR. The PCR primers were complementary to sequences on either side of the SRSF1-responsive intron 4 of the endogenous SRSF1 gene (Sun *et al.*, 2010), and the products were 1,136 or 216 bp for intron-retained or spliced isoforms, respectively. After ethidium staining, the intensities of the bands were measured using a Typhoon imager.

Western blotting

GFP and U1A were detected using primary antibodies from Protein-Tech (66002) and Abcam (ab55751), respectively; SRSF1 was detected using a monoclonal antibody kindly provided by Dr A.R. Krainer (Cold Spring Harbor Laboratory) (Hanamura *et al.*, 1998). Fluorescent secondary antibodies were used against mouse IgG (IRDye 680LT, LI-COR 926-68020) and rabbit IgG (IRDye 800CW, LI-COR 926-32211). The recombinant GFP was kindly provided by Prof. C.R. Bagshaw (Leicester). For phosphatase treatment, nuclear extracts were incubated with shrimp alkaline phosphatase (NEB) at 1 unit/ μ l for 30 min at 37°C prior to gel electrophoresis.

Analysis of eCLIP data from ENCODE

The alignment of reads to the snRNAs was based on the eCLIP-seq Processing Pipeline v2.2 used in the ENCODE project (Consortium, 2012; Davis *et al.*, 2018). FASTA files were downloaded from the ENCODE database for all eCLIP experiments performed in the K562 and HepG2 cell lines. As these files were already demultiplexed, they were first processed using the Cutadapt steps provided in the pipeline (Martin, 2011). Alignment was performed using Bowtie2, with output for unaligned reads, and discordant or mixed pairs discarded, against snRNA sequences obtained from Ensembl (Langmead & Salzberg, 2012; Yates *et al.*, 2020). The output was then sorted and PCR duplicate removal performed using UMI-tools (Smith *et al.*, 2017). The resultant bam file was then processed using BEDTools bamToBed in paired-end mode, with the output processed using awk so that the span of the whole pair was reported in standard bed format (Quinlan & Hall, 2010). This was then mapped using Bedmap to provide read counts at each position on the snRNAs in BED format (Neph *et al.*, 2012).

Additional processing was performed using a custom Perl script to construct a BED file recording the 3' terminus of each read pair. As the eCLIP was performed without a poly-A pulldown, there was a substantial amount of noise in the percentage of total reads aligning to the snRNAs. To compensate for this, read counts were normalized at each position between the eCLIP experiments by taking the percentage of all pairs aligned to the specific snRNAs that were present at that position. For each experiment, the percentage at each position was then divided by the size-matched control, providing a value for enrichment at that position compared to the control. Processing was parallelized using GNU parallel (Tange, 2018).

The logarithm of the enrichment value was subtracted from the logarithm of the percentage of reads terminating at each position. The averages and standard deviations were calculated for all experiments at each position and for each experiment across all positions. As the values were normally distributed both position-wise and experiment-wise, the cumulative probabilities for the position-wise and experiment-wise distributions were multiplied and the square root taken. Following this, the data were collapsed for each protein across cell types, by taking the product of all the probabilities at each position and then raising the result to the power of either 0.5 or 0.25, depending on whether there were 2 or 4 experiments for that protein, respectively.

NMR spectroscopy analysis of the binding of SRSF1ΔRS to U1 snRNP

SRSF1ΔRS was produced in bacteria and purified as previously described (Anczukow *et al*, 2015). To ensure uniform ¹⁵N-labelling and partial deuteration, the protein was produced in M9 medium prepared with fresh D₂O and complemented with ¹⁵N-labelled ammonium chloride (1 g/l) and protonated glucose (2 g/l). One hour before induction of protein expression, 100 mg/l of alpha-ketobutyric acid (methyl-¹³C, 99%; 3,3-D₂, 98%, Cambridge Isotope Laboratory) and 60 mg/l of alpha-ketoisovaleric acid (¹³C₅, 98%; 3-D₁, 98%, Cambridge Isotope Laboratory) were added to the medium to ensure specific ¹³C-labelling of ILV methyl groups.

The reconstitution of U1 snRNP *in vitro* was performed as previously described (Campagne *et al*, 2019). The stem-loop 3 RNA and the mutant were produced by T7 run off using double-stranded DNA template (for SL3 wild type: 5'-GGCCAGTGAATTCTAATACGACTACTATAGCGATTTCCCAAATGTGGGAAACTCGC-3'; 5'-GCGAGTTTCCACATTTGGGAAATCGCTATAGTGAGTCGTATTAGAATTCAGTGGCC-3' and for SL3 mutant: 5'-GGCCAGTGAAT TCTAATACGACTACTATAGCGATTTAAAAAATGTGGGAAACTCGC-3'; 5'-GCGAGTTTCCACATTTTAAATCGCTATAGTGAGTCGTATTAGAATTCAGTGGCC-3'). After 4 h of incubation at 37°C, the transcription mixture was loaded into an HPLC system and the RNA was purified on an anion exchange column in denaturing conditions (85°C, 8 M urea). The RNA was extracted using butanol and refolded in water.

The binding of SRSF1ΔRS on U1 snRNP or U1 snRNA stem-loop 3 was monitored by NMR spectroscopy in the NMR buffer (NaPO₄ 10 mM pH7.0, L-arginine 50 mM, L-glutamate 50 mM, DTT 5 mM). Upon addition of U1 snRNP or U1 snRNA stem-loop 3, the chemical shifts of the ILV methyl and amide groups were monitored on a 2D ¹³C-¹H HMQC and 2D ¹⁵N-¹H HSQC, respectively. The NMR data

were recorded on a Bruker AVIII 700 Mhz at 313 K. Data were analysed using CARA (Keller, 2004).

The structural model presented in Fig 6E was generated using PyMol (Delano) and three following structures: the solution structure of FUS-SL3 (PDB ID 6SNJ), the solution structure of SRSF1 ψRRM bound to 5'-UGGAGGAC-3' (PDB ID 6M8D) and the solution structure of SRSF1 RRM1 bound to 5'-AACAAA-3' (PDB ID 6HPJ). Briefly, the lowest energy structures of the SRSF1 RRMs bound to CA and GGA motifs were aligned on the stem-loop 3 lowest energy structure from the FUS-SL3 ensemble. FUS RRM was hidden as well as the original CA and GGA motifs of SL3. The resulting image is presented in Fig 6E.

FUS RRM was produced as previously described (Loughlin *et al*, 2019) and titrated with SL3 wild-type and SL3 mutant at 30°C.

Isothermal titration calorimetry

ITC experiments were performed on a VP-ITC instrument (MicroCal), calibrated according to the manufacturer's instructions. Protein and RNA samples were dialysed against the NMR buffer (where the DTT was substituted by TCEP). Concentrations of proteins and RNAs were determined using optical density absorbance at 280 and 260 nm, respectively. 20 μM of each RNA was titrated with 180 μM of GB1-SRSF1ΔRS protein by 40 injections of 6 μl every 5 min at 27°C. Raw data were integrated, normalized for the molar concentration and analysed using the Origin 7.0383 software according to a 1:1 RNA:protein ratio-binding model.

U1 snRNA rescue assay on SMN1 exon7

The SMN1 minigene is encoded by the pCI-SMN1 plasmid (Lorson *et al*, 1999). It contains the following 5'SS GGA\GUAAGUCU at the 3'-end of exon 7. In order to make this splicing event independent of the endogenous U1 snRNP, the 5'SS was mutated to GGA\GUAUGUCU (mutation A + 4 U). To rescue the splicing, U1 snRNA containing the complementary mutation was expressed from a modified pG3-U1 snRNA (Raponi *et al*, 2009) containing the mutation U₅A. Since the rescue was only partial, we also closed the bulged created by the unpaired adenine in position -1 by adding the mutation C₁₀U on pG3-U1 U₅A. Using this plasmid, the splicing is driven by the exogenous U1 snRNA and the effects of mutations on stem-loop 3 were investigated. The mutations C₉₈₋₁₀₁A and G₁₀₈₋₁₁₀U were introduced into pG3-U1 snRNA U₅A C₁₀U, which disturb the CA and GGA motifs but maintain the base-pairing. All the mutations were introduced using a quick change protocol, and oligonucleotide sequences are available upon request.

One microgram of pCI-SMN1 and pG3-U1snRNA was transfected in HEK293T cells using the Lipofectamine 2000 Reagent (Life Technologies) according to the manufacturer's protocol. The cells were previously counted and seeded 24 h in advance at 400,000 cells/well. Cells were harvested after 40–44 h. Total RNA was extracted, and 1 μg was used for reverse transcription using Oligo(dT)15 primer and M-MLV Reverse Transcriptase (RNase H-; Promega). 10% of the resulting cDNA was then used for semiquantitative PCR using a vector-specific forward primer (pCI-fwd 5'-GGTGTCCACTCCCAGTTCAG-3') and a SMN1-specific reverse primer (SMN1rev 5'-GCCTCACACCGTGCTGG-3'). The PCR products were separated on a 4% polyacrylamide gel at 100 V for 1 h and stained with GelRed

(Biotium). Band integration was performed Image Studio Lite (LICOR), and the ratio between both isoforms for each condition was determined. The experiment was performed five times.

Data availability

The data supporting this publication are available from the University of Leicester's Figshare data repository <https://doi.org/10.25392/leicester.data.c.5521944>. The single-molecule image data in the Appendix are available upon request.

Expanded View for this article is available online.

Acknowledgements

We thank Dr A.R. Krainer for the anti-SRSF1 antibody. This work was supported by Leverhulme Trust grant RPG-2014-001 and BBSRC sLoLa award BB/T000627/1 to ICE and by an award from the SNF-NCCR RNA and Disease to F.H.T.A.

Author contributions

ICE, FHTA and AJH conceived and supervised the study; AMJ, RW and CML did single-molecule experiments; AJH and RW built the microscope; RW conceived, designed and wrote the MATLAB code; SC prepared *in vitro* reconstituted U1 snRNP and did the NMR experiments and mutant analyses; ARG did the bioinformatic analyses; AC prepared SRSF1 proteins; LC, LPE and MJH did additional experiments; and ICE and SC wrote the manuscript with inputs from AMJ, RW, ARG, AC and FHTA.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Abovich N, Rosbash M (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89: 403–412
- Akerman M, Fregoso OI, Das S, Ruse C, Jensen MA, Pappin DJ, Zhang MQ, Krainer AR (2015) Differential connectivity of splicing activators and repressors to the human spliceosome. *Genome Biol* 16: 119
- Amrein H, Hedley ML, Maniatis T (1994) The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by transformer 2. *Cell* 76: 735–746
- Anczuków O, Akerman M, Cléry A, Wu J, Shen C, Shirole N, Raimer A, Sun S, Jensen M, Hua Y *et al* (2015) SRSF1-regulated alternative splicing in breast Cancer. *Mol Cell* 60: 105–117
- Brackenridge S, Wilkie AO, Sreaton GR (2003) Efficient use of a 'dead-end' GA 5' splice site in the human fibroblast growth factor receptor genes. *EMBO J* 22: 1620–1631
- Bradley T, Cook ME, Blanchette M (2015) SR proteins control a complex network of RNA-processing events. *RNA* 21: 75–92
- Braun JE, Friedman LJ, Gelles J, Moore MJ (2018) Synergistic assembly of human pre-spliceosomes across introns and exons. *Elife* 7: e37751
- Caceres JF, Krainer AR (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J* 12: 4715–4726
- Caceres JF, Stamm S, Helfman DM, Krainer AR (1994) Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* 265: 1706–1709
- Caceres JF, Misteli T, Sreaton GR, Spector DL, Krainer AR (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* 138: 225–238
- Caceres EF, Hurst LD (2013) The evolution, impact and properties of exonic splice enhancers. *Genome Biol* 14: R143
- Campagne S, Boigner S, Rudisser S, Moursy A, Gillioz L, Knorlein A, Hall J, Ratni H, Clery A, Allain FH (2019) Structural basis of a small molecule targeting RNA for a specific splicing correction. *Nat Chem Biol* 15: 1191–1198
- Campagne S, de Vries T, Malard F, Afanasyev P, Dorn G, Dedic E, Kohlbrecher J, Boehringer D, Clery A, Allain FH (2021) An *in vitro* reconstituted U1 snRNP allows the study of the disordered regions of the particle and the interactions with proteins and ligands. *Nucleic Acids Res* 49: e63
- Cao W, Jamison SF, Garcia-Blanco MA (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. *RNA* 3: 1456–1467
- Cao W, Garcia-Blanco MA (1998) A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J Biol Chem* 273: 20629–20635
- Cartegni L, Krainer AR (2003) Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat Struct Biol* 10: 120–125
- Chandler SD, Mayeda A, Yeakley JM, Krainer AR, Fu XD (1997) RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc Natl Acad Sci USA* 94: 3596–3601
- Charenton C, Wilkinson ME, Nagai K (2019) Mechanism of 5' splice site transfer for human spliceosome activation. *Science* 364: 362–367
- Chen L, Weinmeister R, Kralovicova J, Eperon LP, Vorechovsky I, Hudson AJ, Eperon IC (2017) Stoichiometries of U2AF35, U2AF65 and U2 snRNP reveal new early spliceosome assembly pathways. *Nucleic Acids Res* 45: 2051–2067
- Chi B, O'Connell JD, Yamazaki T, Gangopadhyay J, Gygi SP, Reed R (2018) Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins. *Sci Rep* 8: 8755
- Cho S, Hoang A, Sinha R, Zhong XY, Fu XD, Krainer AR, Ghosh G (2011) 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
- Cléry A, Krepl M, Nguyen CKX, 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
- Clery A, Sinha R, Anczukow O, Corriero A, Moursy A, Daubner GM, Valcarcel J, Krainer AR, Allain FH (2013) Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc Natl Acad Sci USA* 110: E2802–E2811
- Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74
- Cote J, Beaudoin J, Tacke R, Chabot B (1995) The U1 small nuclear ribonucleoprotein/5' splice site interaction affects U2AF65 binding to the downstream 3' splice site. *J Biol Chem* 270: 4031–4036
- Crispino JD, Blencowe BJ, Sharp PA (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* 265: 1866–1869
- Crispino JD, Mermoud JE, Lamond AI, Sharp PA (1996) Cis-acting elements distinct from the 5' splice site promote U1-independent pre-messenger RNA splicing. *RNA* 2: 664–673

- Das S, Krainer AR (2014) Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol Cancer Res* 12: 1195–1204
- Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK et al (2018) The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46: D794–D801
- Dickson RM, Cubitt AB, Tsien RY, Moerner WE (1997) On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* 388: 355–358
- Donmez G, Hartmuth K, Kastner B, Will CL, Luhrmann R (2007) The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol Cell* 25: 399–411
- Ellis JD, Lleres D, Denegri M, Lamond AI, Caceres JF (2008) Spatial mapping of splicing factor complexes involved in exon and intron definition. *J Cell Biol* 181: 921–934
- Eperon IC, Ireland DC, Smith RA, Mayeda A, Krainer AR (1993) Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *EMBO J* 12: 3607–3617
- Eperon IC, Makarova OV, Mayeda A, Munroe SH, Caceres JF, Hayward DG, Krainer AR (2000) Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* 20: 8303–8318
- Erkelenz S, Poschmann G, Ptok J, Muller L, Schaal H (2021) Profiling of cis- and trans-acting factors supporting noncanonical splice site activation. *RNA Biol* 18: 118–130
- Fernandez Alanis E, Pinotti M, Dal Mas A, Balestra D, Cavallari N, Rogalska ME, Bernardi F, Pagani F (2012) An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum Mol Genet* 21: 2389–2398
- Fleckner J, Zhang M, Valcarcel J, Green MR (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* 11: 1864–1872
- Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci USA* 102: 16176–16181
- Fukumura K, Taniguchi I, Sakamoto H, Ohno M, Inoue K (2009) U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res* 37: 1907–1914
- Garcia-Parajo MF, Segers-Nolten GM, Veerman JA, Greve J, van Hulst NF (2000) Real-time light-driven dynamics of the fluorescence emission in single green fluorescent protein molecules. *Proc Natl Acad Sci USA* 97: 7237–7242
- Ge H, Manley JL (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell* 62: 25–34
- Ge H, Zuo P, Manley JL (1991) Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* 66: 373–382
- Grabowski PJ, Nasim FU, Kuo HC, Burch R (1991) Combinatorial splicing of exon pairs by two-site binding of U1 small nuclear ribonucleoprotein particle. *Mol Cell Biol* 11: 5919–5928
- Graveley BR, Maniatis T (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell* 1: 765–771
- Graveley BR, Hertel KJ, Maniatis T (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* 7: 806–818
- Guil S, Gattoni R, Carrascal M, Abian J, Stevenin J, Bach-Elias M (2003) Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol Cell Biol* 23: 2927–2941
- Hamm J, Dathan NA, Scherly D, Mattaj JW (1990) Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. *EMBO J* 9: 1237–1244
- Hanamura A, Caceres JF, Mayeda A, Franza Jr BR, Krainer AR (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* 4: 430–444
- Hang J, Wan R, Yan C, Shi Y (2015) Structural basis of pre-mRNA splicing. *Science* 349: 1191–1198
- Hein M, Hubner N, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak I, Weisswange I, Mansfeld J, Buchholz F et al (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163: 712–723
- Hertel KJ, Maniatis T (1999) Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proc Natl Acad Sci USA* 96: 2651–2655
- Hodson MJ, Hudson AJ, Cherny D, Eperon IC (2012) The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites. *Nucleic Acids Res* 40: 6850–6862
- Hoffman BE, Grabowski PJ (1992) U1 snRNP targets an essential splicing factor, U2AF65 to the 3' splice site by a network of interactions spanning the exon. *Genes Dev* 6: 2554–2568
- Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, Gygi MP, Thornock A, Zarraga G, Tam S et al (2021) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* 184: 3022–3040
- Hwang DY, Cohen JB (1996a) Base pairing at the 5' splice site with U1 small nuclear RNA promotes splicing of the upstream intron but may be dispensable for slicing of the downstream intron. *Mol Cell Biol* 16: 3012–3022
- Hwang DY, Cohen JB (1996b) U1 snRNA promotes the selection of nearby 5' splice sites by U6 snRNA in mammalian cells. *Genes Dev* 10: 338–350
- Hwang DY, Cohen JB (1997) U1 small nuclear RNA-promoted exon selection requires a minimal distance between the position of U1 binding and the 3' splice site across the exon. *Mol Cell Biol* 17: 7099–7107
- Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol Cell* 19: 475–484
- Jamison SF, Crow A, Garcia-Blanco MA (1992) The spliceosome assembly pathway in mammalian extracts. *Mol Cell Biol* 12: 4279–4287
- Jamison SF, Pasman Z, Wang J, Will C, Luhrmann R, Manley JL, Garcia-Blanco MA (1995) U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. *Nucleic Acids Res* 23: 3260–3267
- Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, Xiao R, Burge CB, Fu XD (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* 153: 855–868
- Jobbins AM, Reichenbach LF, Lucas CM, Hudson AJ, Burley GA, Eperon IC (2018) The mechanisms of a mammalian splicing enhancer. *Nucleic Acids Res* 46: 2145–2158
- Jutzi D, Campagne S, Schmidt R, Reber S, Mechttersheimer J, Gypas F, Schweingruber C, Colombo M, von Schroetter C, Loughlin FE et al (2020) Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis. *Nat Commun* 11: 6341
- Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S, Wu JY (2011) RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* 31: 1812–1821
- Keller RJ (2004) *The computer aided resonance assignment*, Goldau: Cantina Verlag
- Kohtz JD, Jamison SF, Will CL, Zuo P, Luhrmann R, Garcia-Blanco MA, Manley JL (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368: 119–124

- Konarska MM, Sharp PA (1986) Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* 46: 845–855
- Kondo Y, Oubridge C, van Roon AM, Nagai K (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife* 4: e04986
- Konforti BB, Koziolkiewicz MJ, Konarska MM (1993) Disruption of base-pairing between the 5' splice-site and the 5' End of U1 Snrna is required for spliceosome assembly. *Cell* 75: 863–873
- Konforti BB, Konarska MM (1994) U4/U5/U6 snRNP recognizes the 5' splice site in the absence of U2 snRNP. *Genes Dev* 8: 1962–1973
- Kotlajich MV, Crabb TL, Hertel KJ (2009) Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. *Mol Cell Biol* 29: 1072–1082
- Krainer AR, Conway GC, Kozak D (1990a) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* 62: 35–42
- Krainer AR, Conway GC, Kozak D (1990b) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev* 4: 1158–1171
- Krainer AR, Mayeda A, Kozak D, Binns G (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* 66: 383–394
- Kralovicova J, Vorechovsky I (2007) Global control of aberrant splice-site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. *Nucleic Acids Res* 35: 6399–6413
- Kreivi JP, Zefrivitz K, Akusjarvi G (1991) A U1 snRNA binding site improves the efficiency of *in vitro* pre-mRNA splicing. *Nucleic Acids Res* 19: 6956
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359
- Lavigueur A, La Branche H, Kornblihtt AR, Chabot B (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev* 7: 2405–2417
- Lee Y, Rio DC (2015) Mechanisms and regulation of alternative Pre-mRNA splicing. *Annu Rev Biochem* 84: 291–323
- Li X, Liu S, Jiang J, Zhang L, Espinosa S, Hill RC, Hansen KC, Zhou ZH, Zhao R (2017) CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat Commun* 8: 1035
- Li X, Liu S, Zhang L, Issaian A, Hill RC, Espinosa S, Shi S, Cui Y, Kappel K, Das R et al (2019) A unified mechanism for intron and exon definition and back-splicing. *Nature* 573: 375–380
- Liesche C, Grussmayer KS, Ludwig M, Worz S, Rohr K, Herten DP, Beaudouin J, Eils R (2015) Automated analysis of single-molecule photobleaching data by statistical modeling of spot populations. *Biophys J* 109: 2352–2362
- Lim SR, Hertel KJ (2004) Commitment to splice site pairing coincides with A complex formation. *Mol Cell* 15: 477–483
- Lin C, Yang L, Yang JJ, Huang Y, Liu ZR (2005a) ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol Cell Biol* 25: 7484–7493
- Lin SR, Xiao R, Sun PQ, Xu XD, Fu XD (2005b) Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation. *Mol Cell* 20: 413–425
- Liu ZR (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol Cell Biol* 22: 5443–5450
- Long JC, Caceres JF (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* 417: 15–27
- Longman D, Johnstone IL, Caceres JF (2000) Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *EMBO J* 19: 1625–1637
- Lorson CL, Hahnen E, Androphy EJ, Wirth B (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci USA* 96: 6307–6311
- Loughlin FE, Lukavsky PJ, Kazeeva T, Reber S, Hock E-M, Colombo M, Von Schroetter C, Pauli P, Cléry A, Mühlemann O et al (2019) The solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity. *Mol Cell* 73: 490–550.e496
- Maroney PA, Romfo CM, Nilsen TW (2000) Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell* 6: 317–328
- Martelly W, Fellows B, Kang P, Vashisht A, Wohlschlegel JA, Sharma S (2021) Synergistic roles for human U1 snRNA stem-loops in pre-mRNA splicing. *RNA Biol* <https://doi.org/10.1080/15476286.2021.1932360>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 10–12
- Martins de Araujo M, Bonnal S, Hastings ML, Krainer AR, Valcarcel J (2009) Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. *RNA* 15: 515–523
- Maslon MM, Heras SR, Bellora N, Eyraes E, Caceres JF (2014) The translational landscape of the splicing factor SRSF1 and its role in mitosis. *Elife* 3: e02028
- Mayeda A, Krainer AR (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68: 365–375
- Mayeda A, Helfman DM, Krainer AR (1993) Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol Cell Biol* 13: 2993–3001
- Mayeda A, Sreaton GR, Chandler SD, Fu XD, Krainer AR (1999) Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol Cell Biol* 19: 1853–1863
- Mermoud JE, Cohen P, Lamond AI (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res* 20: 5263–5269
- Mermoud JE, Cohen P, Lamond AI (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J* 13: 5679–5688
- Michaud S, Reed R (1991) An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev* 5: 2534–2546
- Michaud S, Reed R (1993) A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev* 7: 1008–1020
- Muller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* 30: 553–566
- Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S et al (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28: 1919–1920
- Nikolaev Y, Ripin N, Soste M, Picotti P, Iber D, Allain FH (2019) Systems NMR: single-sample quantification of RNA, proteins and metabolites for biomolecular network analysis. *Nat Methods* 16: 743–749
- Ó Ruanaidh JJK, Fitzgerald WJ (1996) *Numerical Bayesian methods applied to signal processing*. New York, NY: Springer
- Ohr T, Prior M, Dannenberg J, Odenwalder P, Dybkov O, Rasche N, Schmitzova J, Gregor I, Fabrizio P, Enderlein J et al (2012) Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* 18: 1244–1256

- Palacino J, Swalley SE, Song C, Cheung AK, Shu L, Zhang X, Van Hoosear M, Shin Y, Chin DN, Keller CG et al (2015) SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat Chem Biol* 11: 511–517
- Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares Jr M, Fu XD (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell* 50: 223–235
- Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458: 475–480
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842
- Raponi M, Buratti E, Dassie E, Upadhyaya M, Baralle D (2009) Low U1 snRNP dependence at the NF1 exon 29 donor splice site. *FEBS J* 276: 2060–2073
- Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A et al (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499: 172–177
- Robberson BL, Cote GJ, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10: 84–94
- Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR (2012) Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev* 26: 1098–1109
- Roca X, Krainer AR, Eperon IC (2013) Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev* 27: 129–144
- Roscigno RF, Garcia-Blanco MA (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* 1: 692–706
- Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* 19: 381–394
- Schaal TD, Maniatis T (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol* 19: 261–273
- Schmidt C, Gronborg M, Deckert J, Bessonov S, Conrad T, Luhrmann R, Urlaub H (2014) Mass spectrometry-based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labeling (SILAC), chemical labeling (iTRAQ), and label-free spectral count. *RNA* 20: 406–420
- Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, Luhrmann R (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell* 38: 223–235
- Serrano P, Aubol BE, Keshwani MM, Forli S, Ma CT, Dutta SK, Geralt M, Wuthrich K, Adams JA (2016) Directional phosphorylation and nuclear transport of the splicing factor srsf1 is regulated by an RNA recognition motif. *J Mol Biol* 428: 2430–2445
- Shao W, Kim HS, Cao Y, Xu YZ, Query CC (2012) A U1–U2 snRNP interaction network during intron definition. *Mol Cell Biol* 32: 470–478
- Sharma S, Maris C, Allain FH, Black DL (2011) U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Mol Cell* 41: 579–588
- Sharma S, Wongpalee SP, Vashisht A, Wohlschlegel JA, Black DL (2014) Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev* 28: 2518–2531
- Shaw SD, Chakrabarti S, Ghosh G, Krainer AR (2007) Deletion of the N-terminus of SF2/ASF permits RS-domain-independent pre-mRNA splicing. *PLoS One* 2: e854
- Shen H, Green MR (2004) A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell* 16: 363–373
- Shen H, Kan JL, Green MR (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell* 13: 367–376
- Shen H, Green MR (2006) RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev* 20: 1755–1765
- Shen H, Green MR (2007) RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat Struct Mol Biol* 14: 597–603
- Shen H, Zheng X, Shen J, Zhang L, Zhao R, Green MR (2008) Distinct activities of the DEXD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev* 22: 1796–1803
- Singh NN, Singh RN, Androphy EJ (2007) Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* 35: 371–389
- Singh RN, Singh NN (2019) A novel role of U1 snRNP: splice site selection from a distance. *Biochim Biophys Acta Gene Regul Mech* 1862: 634–642
- Skordis LA, Dunckley MG, Yue B, Eperon IC, Muntoni F (2003) Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc Natl Acad Sci USA* 100: 4114–4119
- Sleeman J, Lyon CE, Platani M, Kreivi JP, Lamond AI (1998) Dynamic interactions between splicing snRNPs, coiled bodies and nucleoli revealed using snRNP protein fusions to the green fluorescent protein. *Exp Cell Res* 243: 290–304
- Smith L, Dickinson R, Lucas C, Cousins A, Malygin A, Weldon C, Perrett A, Bottrill A, Searle M, Burley G et al (2014) A targeted oligonucleotide enhancer of SMN2 exon 7 splicing forms competing quadruplex and protein complexes in functional conditions. *Cell Rep* 9: 193–205
- Smith T, Heger A, Sudbery I (2017) UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res* 27: 491–499
- Staknis D, Reed R (1994) SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol Cell Biol* 14: 7670–7682
- Sterner DA, Carlo T, Berget SM (1996) Architectural limits on split genes. *Proc Natl Acad Sci USA* 93: 15081–15085
- Sun Q, Mayeda A, Hampson RK, Krainer AR, Rottman FM (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev* 7: 2598–2608
- Sun S, Zhang Z, Sinha R, Karni R, Krainer AR (2010) SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol* 17: 306–312
- Tange O (2018) GNU parallel 2018. <https://zenodo.org/record/1146014>
- Tarn WY, Steitz JA (1994) Sr proteins can compensate for the loss of U1 Snrnp functions in-vitro. *Genes Dev* 8: 2704–2717
- Tarn WY, Steitz JA (1995) Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. *Proc Natl Acad Sci USA* 92: 2504–2508
- Townsend C, Leelaram MN, Agafonov DE, Dybkov O, Will CL, Bertram K, Urlaub H, Kastner B, Stark H, Luhrmann R (2020) Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. *Science* 370: eabc3753
- Ulbricht MH, Isacoff EY (2007) Subunit counting in membrane-bound proteins. *Nat Methods* 4: 319–321

- Vamosi G, Mucke N, Muller G, Krieger JW, Curth U, Langowski J, Toth K (2016) EGFP oligomers as natural fluorescence and hydrodynamic standards. *Sci Rep* 6: 33022
- Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, Blue SM, Chen J-Y, Cody NAL, Dominguez D et al (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583: 711–719
- Wahl MC, Luhrmann R (2015) SnapShot: spliceosome dynamics II. *Cell* 162: 456–U254
- Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17: 691–695
- Wang J, Manley JL (1995) Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing *in vivo* in diverse ways. *RNA* 1: 335–346
- Wang Z, Hoffmann HM, Grabowski PJ (1995) Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA* 1: 21–35
- Wang J, Takagaki Y, Manley JL (1996) Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev* 10: 2588–2599
- Wang H, Chong S (2003) Visualization of coupled protein folding and binding in bacteria and purification of the heterodimeric complex. *Proc Natl Acad Sci USA* 100: 478–483
- Will CL, Luhrmann R (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3: a003707
- Wu JY, Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75: 1061–1070
- Xiao SH, Manley JL (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 11: 334–344
- Xiao SH, Manley JL (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J* 17: 6359–6367
- Xu YZ, Newnham CM, Kameoka S, Huang T, Konarska MM, Query CC (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J* 23: 376–385
- Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R et al (2020) Ensembl 2020. *Nucleic Acids Res* 48: D682–D688
- Zhu J, Krainer AR (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev* 14: 3166–3178
- Zhu J, Mayeda A, Krainer AR (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8: 1351–1361
- Zhuang Y, Weiner AM (1986) A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46: 827–835
- Zuo P, Manley JL (1993) Functional domains of the human splicing factor ASF/SF2. *EMBO J* 12: 4727–4737



License: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.