nature structural & molecular biology

A U1 snRNP–specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange

Byung Ran So, Lili Wan, Zhenxi Zhang, Pilong Li, Eric Babiash, Jingqi Duan, Ihab Younis & Gideon Dreyfuss

Despite equal snRNP stoichiometry in spliceosomes, U1 snRNP (U1) is typically the most abundant vertebrate snRNP. Mechanisms regulating U1 overabundance and snRNP repertoire are unknown. In Sm-core assembly, a key snRNP-biogenesis step mediated by the SMN complex, the snRNA-specific RNA-binding protein (RBP) Gemin5 delivers pre-snRNAs, which join SMN–Gemin2–recruited Sm proteins. We show that the human U1-specific RBP U1-70K can bridge pre-U1 to SMN–Gemin2–Sm, in a Gemin5-independent manner, thus establishing an additional and U1-exclusive Sm core–assembly pathway. U1-70K hijacks SMN–Gemin2–Sm, enhancing Sm-core assembly on U1s and inhibiting that on other snRNAs, thereby promoting U1 overabundance and regulating snRNP repertoire. SMN–Gemin2's ability to facilitate transactions between different RBPs and RNAs explains its multi-RBP valency and the myriad transcriptome perturbations associated with SMN deficiency in neurodegenerative spinal muscular atrophy. We propose that SMN–Gemin2 is a versatile hub for RNP exchange that functions broadly in RNA metabolism.

In eukaryotes, U1 small nuclear ribonucleoprotein (snRNP) is necessary for precursor (pre)-mRNA splicing[1,2] and telescripting, suppression of premature cleavage, and polyadenylation that ensures full-length transcription[3–7]. U1's function in telescripting probably explains why human cells have a higher abundance of U1 compared to the other spliceosomal snRNPs (U2, U4, U5 and U6)[8]. All the spliceosomal small nuclear RNAs (snRNAs) that acquire Sm cores (U1, U2, U4, U5 and the minor spliceosomes U11, U12 and U4atac) have an snRNA-defining motif termed the snRNP code, which consists of the Sm site ($AU_{5-6}G$), around which the Sm core forms, and an adjacent 3′-terminal stem-loop[9–11]. In Sm cores, Sm heterodimers Sm D1D2 (D1D2) and Sm BD3 (BD3) and the heterotrimer Sm FEG (FEG) subunits are arranged in the order D1, D2, F, E, G, D3 and B, and each Sm protein interacts with a single nucleotide of the Sm site[12–17]. The snRNP code (~50 nt) is recognized by the RBP Gemin5 (refs. 18–20). Sm-core assembly is mediated by the survival of motor neurons (SMN)–Gemin complex, a large oligomeric complex (30–70S) comprising SMN, Gemins2–8 and Unrip[9,21,22]. The SMN–Gemin complex is ubiquitously expressed in eukaryotes (except *Saccharomyces*, which have a Gemin2 ortholog but no detectable SMN) and is found in both the nucleus and in the cytoplasm. SMN and its interacting protein Gemin2, the two most highly conserved components of the SMN–Gemin complex, bind Sm proteins[23–25]. Pre-snRNAs, including pre-U1 snRNA, are delivered to SMN–Gemin2 by Gemin5, where they encounter the Sm proteins and acquire Sm cores[20]. Sm-core assembly is a prerequisite for snRNP maturation—including hypermethylation of the 5′-monomethyl guanosine cap (to form a trimethylguanosine) and 3′-end processing—and for snRNP function[26,27]. U1 in vertebrates differs from other snRNAs in two respects: first, it has a divergent snRNP code, owing to its noncanonical Sm site (AUUUGUG or AUUUCUG), which has weaker binding affinity to Gemin5 (ref. 19), and second, its Sm-core assembly strongly depends on stem-loop 1 (SL1), a U1-specific structure outside the snRNP code[28–30]. Interestingly, earlier studies have shown that SL1 alone can inhibit Sm-core assembly[30]; however, the role of SL1 and U1's divergent snRNP code in Sm-core assembly remains unknown. Here, we investigated the roles of these U1 elements in Sm-core assembly, seeking to understand how U1 overabundance is achieved and more generally what factors contribute to cells' unequal snRNP repertoire.

RESULTS

U1-specific U1-70K protein bridges U1 snRNA to the SMN complex

To understand the role of U1's SL1 and snRNP code in Sm-core assembly, we compared the binding of SMN-complex components (SMN, Gemin2–8 and Unrip) to biotinylated pre-U1, mature U1, U1A3 (a U1 SL1 mutant defective in U1 snRNP–specific U1-70K binding)[28], U1ΔSm (a U1 Sm-site mutant that cannot assemble the Sm core), super-U1 (a chimeric U1 in which the Sm site was replaced by a canonical Sm site, AUUUUUG), U4 snRNA (U4), a canonical snRNP code–containing snRNA, and U4ΔSm. The RNAs were incubated in human (HeLa) cell extracts in which SMN, Gemin5 or U1-70K had been knocked down by RNA interference (RNAi) (**Fig. 1a**). Although U1-70K has not previously been suggested to have a role in Sm-core assembly, we tested its effect because it is the only protein known to bind SL1 (refs. 28,29). For most experiments, we used pre-U1, which is 50 nt longer at the 3′ end than U1, because it represents the U1 substrate for Sm-core assembly in cells[20]. For U4, whose precursor is only 6 nt longer than mature U4, there was little difference between the precursor and mature forms (data not shown).

Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. Correspondence should be addressed to G.D. (gdreyfuss@hhmi.upenn.edu).

Received 16 December 2015; accepted 4 January 2016; published online 1 February 2016; doi:10.1038/nsmb.3167

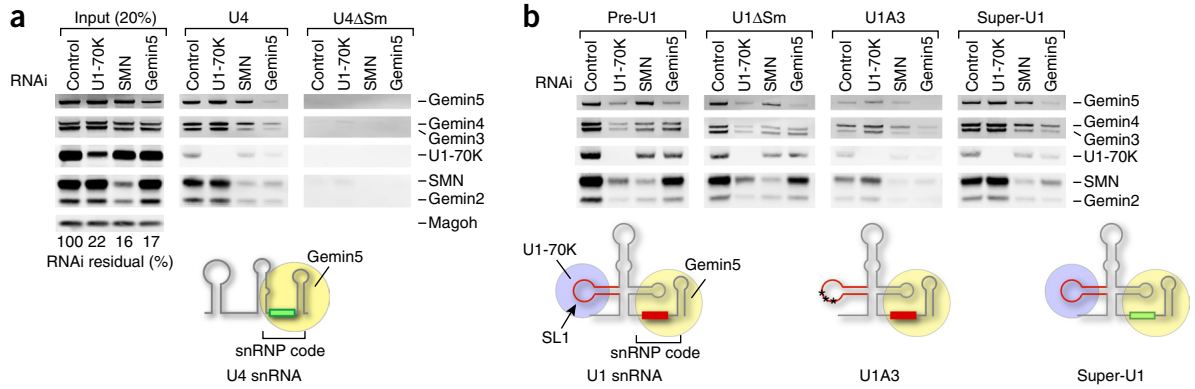

Figure 1 The U1 snRNP–specific stem-loop 1–binding protein U1-70K bridges pre-U1 or U1 snRNA to the SMN complex independently of Gemin5. (**a**) Western blot analysis of SMN-complex components bound to biotinylated U4 and U4ΔSm snRNAs in HeLa cells with control, U1-70K, SMN or Gemin5 short interfering RNA (siRNA) knockdown. The input lanes show 20% of each of the cell extracts used. The knockdown efficiencies relative to Magoh loading control are indicated as percentages of residual protein for each knockdown compared to control knockdown (set at 100%). A schematic illustration of U4 structure and its canonical Sm site in the snRNP code, which is the RNA sequence necessary and sufficient for Gemin5 binding, are indicated. (**b**) Western blot analysis as in **a** for biotinylated U1, including U1 precursor (pre-U1), U1ΔSm, SL1 mutant (U1A3) and super-U1. A schematic illustration of the U1 is shown, depicting SL1 for U1-70K binding and U1's divergent Sm site in the snRNP code for Gemin5 binding. The U1A3 has a triple mutation (U27G A29C U30C) in the SL1, which abolishes U1-70K binding. Super-U1 is a mutation that replaces U1's Sm site with a canonical Sm site. Uncropped scans of western blots are shown in **Supplementary Data Set 1**.

As expected, Gemin5 knockdown abolished U4's association with the SMN complex[18], particularly with SMN–Gemin2, whereas SMN knockdown had no effect on U4's binding to Gemin5 (**Fig. 1a**). U4ΔSm did not bind Gemin5 or any component of the SMN complex (**Fig. 1a**).

In contrast, the association of pre-U1, U1 (data not shown) or U1ΔSm with SMN–Gemin2 was only partially decreased by Gemin5 knockdown but was strongly decreased by U1-70K knockdown (**Fig. 1b**). However, U1-70K knockdown had no effect on U4's association with the SMN complex (**Fig. 1a**). U1A3 binding to the SMN complex was unaffected by U1-70K knockdown but was reduced by Gemin5 knockdown, thus suggesting that U1A3 associates with the SMN complex through Gemin5, albeit more weakly than U4, owing to U1's non-canonical Sm site (**Fig. 1b**). Binding of super-U1 to the SMN complex was unaffected by U1-70K knockdown but was strongly decreased by Gemin5 knockdown (**Fig. 1b**). Thus, U1-70K bridges pre-U1 and U1 to the SMN complex apparently independently of Gemin5, thus representing an alternative and more effective pathway to that of Gemin5, whereas the binding of U4 to the SMN complex strictly depends on Gemin5. These data demonstrate that whereas pre-U1 can use either U1-70K or Gemin5, U1-70K is normally the main adaptor to SMN–Gemin2–Sm proteins. These observations also suggest that U1's Sm site may have diverged to become a weaker Gemin5 binder because it allows U1 to use an alternate route to bind the SMN complex.

U1-70K enhances Sm-core assembly on U1 and inhibits that on other snRNAs

Next, we used a quantitative snRNP assembly assay[31,32] to determine the roles of U1-70K and Gemin5 in Sm-core assembly (**Fig. 2a** and **Supplementary Fig. 1**). As expected, Sm-core assembly on all tested RNAs depended on SMN–Gemin2 and required an Sm site. Gemin5 knockdown decreased Sm-core assembly on canonical snRNP code–containing snRNAs (U2, U4 and U5) to a greater extent than that on SL1-containing snRNAs. Strikingly, U1-70K knockdown decreased pre-U1 assembly (by 50%) but enhanced U2, U4 and U5 Sm-core assembly (up to 100%). Loss of U1-70K binding or enhancement of Gemin5 binding in U1A3 and super-U1, respectively, made them more similar to non-U1 snRNAs. These observations indicate a new role for U1-70K as a specific pre-U1 Sm core–assembly factor and an inhibitor of Sm-core assembly on other snRNAs.

U1-70K regulates cells' snRNP repertoire

We then studied the effect of U1-70K knockdown on U1 snRNP abundance in cells. U1-70K RNAi decreased the steady-state abundance

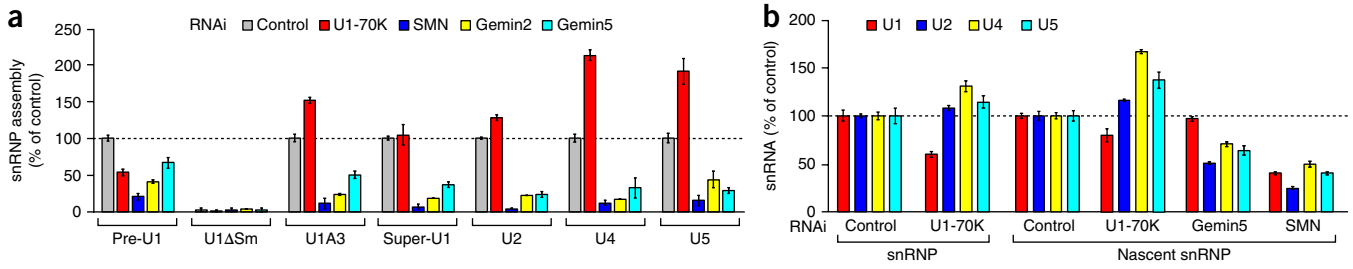

Figure 2 U1-70K enhances Sm-core assembly on U1 and inhibits that on other snRNAs *in vitro*, and regulates the snRNP repertoire in cells. (**a**) Quantitative *in vitro* Sm core–assembly activities on the indicated snRNAs in extracts from cells with control, U1-70K, SMN, Gemin2 or Gemin5 siRNA knockdown. The Sm core–assembly activities on each snRNA are compared to control RNAi extracts (100% activity), except those for U1ΔSm and U4ΔSm, whose relative activities are compared to pre-U1 and U4, respectively. Error bars, s.d. ($n = 3$ independent cell cultures). (**b**) Quantitative measurements of snRNAs from mature and nascent snRNPs in cells with control, U1-70K, SMN or Gemin5 knockdown, assessed by quantitative real-time PCR of the RNAs immunoprecipitated with anti-Sm (Y12) antibody. The relative amount of snRNAs compared to those under control RNAi are shown. Error bars, s.d. ($n = 3$ independent cell cultures).

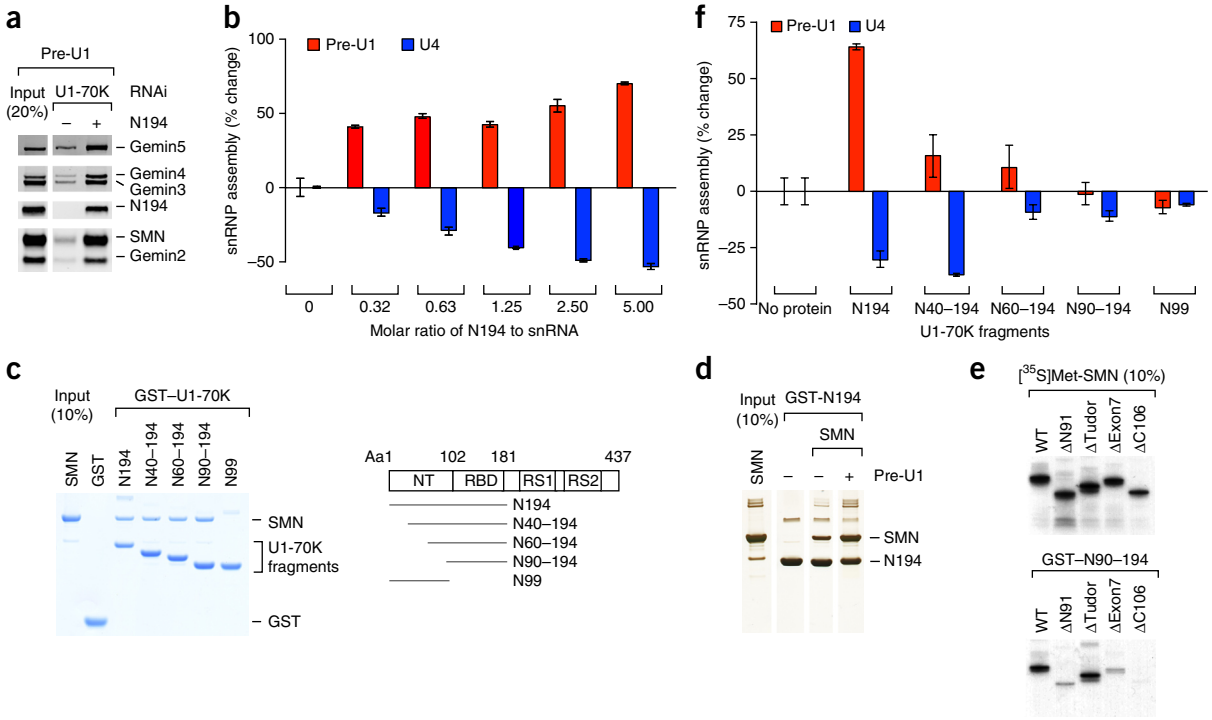

Figure 3 U1-70K bridges pre-U1 snRNA to SMN and mediates preferential Sm-core assembly on U1 snRNA rather than on other snRNAs. (**a**) Western blot of SMN-complex proteins bound to the biotinylated pre-U1 snRNA in U1-70K–knockdown cell extracts complemented with recombinant U1-70K N194. The input lane shows 20% of the cell extracts and the N194 used for binding. (**b**) *In vitro* Sm core–assembly activities on pre-U1 and U4 with increasing concentrations of N194 in U1-70K–knockdown extracts. The rescued snRNP-assembly capacities on snRNAs were determined as percentage changes from the U1-70K knockdown without N194. Error bars, s.d. ($n = 3$ independent cell cultures). (**c**) Binding of GST–U1-70K proteins to recombinant SMN protein. Schematic diagram of U1-70K and its deletion fragments with corresponding residue numbers are indicated. The input lane shows 10% of the SMN protein used for binding, and the gel was visualized by SimplyBlue staining. Aa, amino acids. (**d**) Binding of GST-N194 to SMN in the absence or presence of equimolar amounts of *in vitro*–transcribed pre-U1. The protein gel was visualized by silver staining. (**e**) Binding of GST–N90–194 protein to *in vitro*–translated [^{35}S]methionine-labeled wild-type (WT) SMN and its domain deletions. The input lanes show 10% of SMN's deletion domains (N91, amino acids 1–91; Tudor, amino acids 92–158; exon7, amino acids 279–294; and C106, amino acids 189–294), and the gel was visualized by autoradiography. (**f**) snRNP assembly measurements as in **b** with N194-deletion fragments with a 1.25 molar ratio of U1-70K proteins to snRNA. Error bars, s.d. ($n = 3$ independent cell cultures). Uncropped scans of blots, gels and autoradiographs are shown in **Supplementary Data Set 1**.

of U1 by 25–40% and increased that of other snRNPs by 10–30% (**Fig. 2b**, snRNP). To determine the effect on the rate of Sm-core assembly, we quantified the amount of snRNAs that assembled Sm cores in a 2-h pulse label with 4-thiouridine[33]. The results showed that U1-70K knockdown decreased U1 snRNP production and increased the formation of other snRNPs (**Fig. 2b**, nascent snRNP). U1-70K knockdown had no effect on SMN or Sm protein levels (**Supplementary Fig. 2**). In contrast to U1-70K knockdown, Gemin5 knockdown had no effect on U1 assembly but decreased that of other snRNPs. These findings are consistent with the *in vitro* Sm-core assembly results and demonstrate that U1-70K is required for generating the overabundance of U1 compared to the other snRNPs, and it regulates cells' snRNP repertoire.

U1-70K domains bind SMN and mediate Sm-core assembly

Delineation of the domains of U1-70K showed that the C-terminal arginine- and serine-rich (RS) domain (amino acids 230–437), which is important for splicing[34], was not required for bridging pre-U1 to the SMN complex (data not shown). However, the N-terminal 194 amino acids of recombinant U1-70K (N194) restored the association of the SMN complex with pre-U1 in U1-70K–knockdown cell extracts (**Fig. 3a**). Furthermore, N194 bound SMN directly and enhanced Sm-core assembly on pre-U1 and inhibited that on U4 (**Fig. 3b**,**c**). The smallest N194 fragment that was sufficient to bind SMN was N90–194, which contains U1-70K's SL1-specific RNA-binding domain (RBD; amino acids 102–181)[29], whereas the N-terminal domain lacking the RBD, N99 (amino acids 1–99) did not bind SMN (**Fig. 3c**). Moreover, pre-U1 enhanced N194 binding to SMN (2.1 fold) (**Fig. 3d**). Thus, pre-U1 binding creates additional interactions between N194 and SMN, either through an RNA-mediated allosteric effect that increases the U1-70K interaction surface with SMN or through additional binding interactions between pre-U1 and SMN. The ability of the same type of RBD as that of U1-70K to engage in protein interactions through a surface other than the RNA-binding site has been described for other RBPs[35,36].

Binding of N90–194 to SMN was abrogated by the deletion of SMN's C-terminal YG box, an oligomerization domain upon which many of SMN's interactions and functions depend[37–39] (**Fig. 3e**). The deleted region included the portion absent in SMNΔ7, the major SMN isoform expressed in people with spinal muscular atrophy (SMA). SMNΔ7 lacks 16 exon7-encoded amino acids, which constitute much of the YG box. Deletion of SMN exons 1, 2a and 2b (ΔN91) also impaired U1-70K binding, thus suggesting that a peptide including an exon 2b-encoded domain that engages in homotypic interactions in oligomeric SMN[31,40] is also needed for binding. A more detailed definition of SMN's N90–194 binding site will require additional structural information about oligomeric SMN, whose structure is presently unknown.

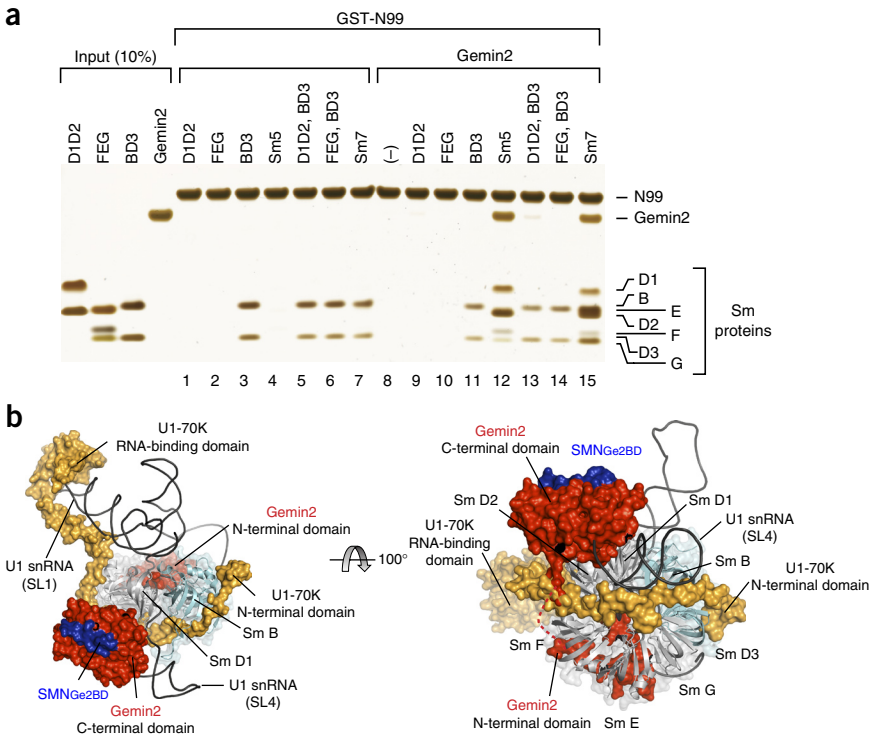

Figure 4 SMN–Gemin2 cooperates with U1-70K in Sm-protein recruitment. (**a**) *In vitro* binding of GST–U1-70K N99 to Gemin2 and Sm proteins. Each Sm-protein subunit, combinations of the subunits (lanes 1–7), Gemin2 only (lane 8) or Gemin2 with combinations of Sm-protein subunits (lanes 9–15) are indicated. The input panel shows 10% of the proteins used for binding, and the gel was visualized by silver staining. Uncropped images are shown in **Supplementary Data Set 1**. (**b**) Superimposed crystal structures of U1 snRNP[12,14,17] and SMN–Gemin2–Sm5 (ref. 25), showing cooperative roles in the Sm core–assembly intermediate of U1. Left, the Gemin2–Sm5–U1-70K intermediate contacts pre-U1 bound by U1-70K's SL1 RBD, places pre-U1's Sm site at the Sm5 inner ring and prevents binding of other RNAs. U1-70K's N-terminal domain binds directly to BD3, which facilitates Sm-protein recruitment and proper positioning for Sm-ring closure. For clarity, two other U1 snRNP specific proteins, U1A, which binds to SL2, and U1C, which binds to U1-70K and U1 snRNA, are not shown. Right, rotated view compared with that at left, showing that Gemin2 and U1-70K assist in Sm-protein recruitment. Gemin2 binds D1D2 through its C-terminal domain and FEG via its N-terminal domain. The loop connecting these two domains and shown in a dashed line is disordered, and its length (amino acids 70–82) is probably sufficient to avoid a clash with U1-70K's D2F binding region. Two patches (amino acids 10–31 and 39–59) of N-terminal U1-70K bind at the BD3 and D2F interface, respectively. The figures were prepared with PyMOL (http://www.pymol.org/).

Unlike N194, N90–194 did not enhance SM-core assembly on pre-U1 or inhibit that on U4 (**Fig. 3f**), thus suggesting that peptides within U1-70K amino acids 1–89 are necessary for these activities. Recently determined X-ray crystal structures of mature U1 snRNP have shown that U1-70K amino acids 39–58 and 10–31 contact the Sm core's perimeter at two opposite sides, at D2F and BD3, respectively[12,14,17]. Deletion of amino acids 1–39 (N40–194) caused loss of the enhanced Sm-core assembly on pre-U1 but did not diminish the inhibition of Sm-core assembly on U4 (**Fig. 3f**). However, a further deletion of amino acids 40–59 (N60–194) caused loss of the inhibition of U4 Sm-core assembly. Thus, U1-70K's snRNA-selective effect on Sm-core assembly represents two separable activities: N40–194 is necessary and sufficient for inhibition of U4 Sm-core assembly, but enhancement of pre-U1 Sm-core assembly also requires amino acids 1–39. N99, which lacks the RBD, had no effect (**Fig. 3f**). This result suggests that the RBD, particularly when bound to pre-U1, blocks the access of other snRNAs to Sm proteins.

U1-70K and Gemin2 cooperate to recruit all Sm proteins

N99 directly bound BD3 but not D1D2 or FEG (**Fig. 4a**, lanes 1–7). In contrast, and as shown previously, Gemin2 did not bind BD3, nor did it bind D1D2 or FEG alone; however, it formed a stable complex with Sm pentamer D1D2–FEG (Sm5)[24,25]. SMN-bound Gemin2–Sm5 is a key intermediate in Sm-core assembly, and we therefore investigated whether Gemin2 and U1-70K could collaborate in binding Sm. Our results showed that N99 did not bind Gemin2 (**Fig. 4a**, lane 8); however, complexes containing N99, Gemin2 and Sm5 or all seven Sm proteins (Sm7) readily formed (**Fig. 4a**, lanes 12 and 15). The Sm7-containing complex would not be expected to form a stable Sm ring in the absence of an RNA's Sm site[15]. Thus, U1-70K and Gemin2 cooperate to recruit all of the Sm core's proteins, but neither alone can achieve this recruitment.

These findings are concordant with recent crystal structures of SMN–Gemin2–Sm5 and U1 snRNP[12,14,17,25]. Superimposition of the two structures supports the presence of an Sm-core assembly intermediate in which Sm5 is bound simultaneously by Gemin2 and U1-70K (**Fig. 4b**). It illustrates how U1-70K binds at the junction of

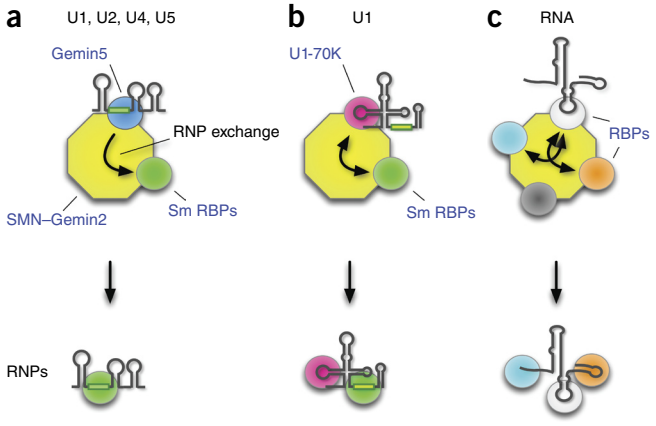

Figure 5 Schematic representation of SMN–Gemin2's function as a versatile hub for RNP exchange. The different-colored spheres represent the diverse RBPs that bind multivalent SMN–Gemin2. For simplicity, only the pre-snRNAs' key features relevant to this pathway are shown. (**a**) Gemin5, which recognizes the snRNP code common to all pre-snRNAs, is a 'drop-and-go' donor that does not remain with the fully assembled Sm core. (**b**) U1-70K, a pre-U1 stem-loop 1 (SL1)-binding protein associates with SMN–Gemin2 and remains part of the completed Sm core. (**c**) A representative and hypothetical RNP, exemplified by FUS/TLS, associates with the SMN–Gemin2, as described in the text.

D1D2 and FEG, thus positioning pre-U1's Sm site at the inner RNA-binding cleft, where N39's (amino acids 1–39) interactions with pre-U1 probably help to position pre-U1's Sm site in the hole of the Sm ring and recruit BD3 for Sm-core closure[12,14,17]. U1-70K's multiple interactions with SMN–Gemin2–Sm5, which may be initiated by pre-U1 binding, effectively hijack the Sm core–assembly module, thereby promoting U1 assembly and inhibiting that on other snRNAs. SMN performs a key role by acting as a scaffold for Gemin2, U1-70K and the C-terminal arginine- and glycine-rich (RG(G)) domains of Sms B, D3 and D1 (refs. 41–43), as described below.

DISCUSSION

Previous studies have suggested that the SMN complex is a dedicated Sm core–assembly device that operates with a single RBP, Gemin5, which binds a single RNA structure, the snRNP code, and joins the SMN–Gemin2 subunit[20,25]. However, our studies demonstrate that SMN–Gemin2 can receive RNAs from at least one additional and structurally unrelated RBP (U1-70K), which has a different RNA binding specificity (owing to SL1). In the case of Gemin5, there is a complete exchange between the RBP donor and the acceptor RBPs, Sm5. In contrast, U1-70K remains part of the mature RNP. Yet both paths use SMN–Gemin2, thus revealing that SMN–Gemin2 is a versatile platform for RBP-RNA transactions. The RNP-exchange function of SMN–Gemin2 leverages the donor's RNA specificity to produce a specific RNP, in this case the Sm core, that could not be produced by the Sm proteins on their own [44]. We refer to this function as RNP exchange and propose that the RNP-exchange function of SMN–Gemin2 provides a unifying theme that extends beyond Sm-core assembly and has a central role in RNA metabolism. Our concept of SMN–Gemin2 as a general hub for RNP exchange (represented in **Fig. 5**) is based on its ability to bind many RBPs and RNAs, thus increasing the opportunity for creating greater RNP diversity and explaining many previous observations.

SMN and Gemin2, the most highly conserved components of the SMN complex, are ubiquitously expressed and are essential for cell viability across eukaryotes[21]. Several features of SMN–Gemin2 make it particularly suitable for carrying out RNP exchange. SMN is a highly oligomeric protein, and hence the SMN complex is a large particle (30–70S), that contains Gemins3–8 and Unrip, including the putative RNA or RNP ATPase Gemin3, whose precise functions are unknown but probably include assisting in RNP exchange. SMN has a relatively indiscriminate Tudor domain that binds peptides modified with methyl-arginine, a common post-translational modification in RG(G) domains, which are found in numerous RBPs[41–43]. Oligomeric SMN–Gemin2's pan-RBP-binding Tudor domains are thus a multivalent 'station' where diverse RBPs and their RNA cargos can congregate, thereby promoting RNP exchange. Indeed, a bewildering number of RBPs in addition to Sm and Lsm proteins bind the SMN–Gemin complex, including heterogeneous nuclear ribonucleoprotein (hnRNP) proteins (for example, hnRNPs A1, A2, Q, R, U and FUS/TLS), small nucleolar ribonucleoprotein proteins (for example, fibrillarin and GAR1), the SRP component SRP54 and the transcription regulator CARM1 (refs. 41,44–47). However, the diverse RBPs that interact with the SMN complex could not be explained in terms of Sm-core assembly. In addition, SMN can bind RNAs, albeit nonspecifically with low affinity[48], thus facilitating its potential to maintain transient association with RNA intermediates. Therefore, a common theme is that SMN–Gemin2 facilitates formation of diverse RNPs from a multitude of RBP donors and acceptors.

U1-70K's unexpected role in Sm-core assembly provides an exclusive path for pre-U1 to attain high U1 abundance, thereby allowing the U1 level to be regulated separately from those of the other snRNPs. More generally, U1-70K emerged as a key regulator of snRNP repertoire. Other factors, including the number of snRNA genes and their transcriptional activity, as well as the stability of snRNPs, undoubtedly also contribute to U1 abundance and snRNP repertoire. In humans, there are many (>100) U1 genes, pseudogenes and variants[49] and many genes encoding other snRNAs, but these genes remain poorly annotated, and little is known about their transcriptional activity. However, pre-snRNAs lacking Sm cores are rapidly degraded[50], thus suggesting that the Sm-core assembly by the SMN complex may be a rate-limiting step in snRNP biogenesis. This makes the SMN complex and the ability of U1-70K to program its Sm-core selectivity especially important factors in gene regulation.

SMN deficiency causes SMA, whose severity correlates with both the degree of SMN deficiency and the corresponding decrease in Sm-core assembly[32,51]. However, SMA is associated with a large number of widespread perturbations in RNA metabolism, including alterations in snRNP repertoire and pre-mRNA splicing[51–55]. Among these are specific alternative-splicing deficits that precede and can now explain important aspects of SMA's synaptopathology[55]. Although splicing changes are a plausible outcome of snRNP changes, other transcriptome perturbations in SMN-deficient cells, such as expression-level changes, are more readily explained as resulting directly from the loss of the RNP-exchange capacity of the SMN complex. For example, the SMN complex's multiple interactions with RBPs indeed reflect its role as a versatile hub for chaperoning RNPs involved in diverse transcriptional and post-transcriptional processes. This perspective offers important insights into the roles of RBP mutations in other human diseases. For example, the SMN complex has recently been reported to intersect with several major neurodegenerative diseases caused by specific mutations in RBPs, FUS/TLS and TDP-43, which cause amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD)[45,56–58]. These pathogenic mutants also elicit many transcription, splicing and microRNA changes. Notably, FUS/TLS (and TDP-43 via FUS/TLS) binds SMN's Tudor domain, as mediated by methylated arginines in its RG(G) domain, where a cluster of pathogenic mutations are located. Impairment of the general RNP-exchange function of the SMN–Gemin complex due to poisoning by mutant RBPs, small-molecule inhibitors or genetic SMN deficiency provides a potential alternative explanation for the myriad RNA-metabolism perturbations in ALS and SMA.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of our laboratory for helpful discussions and comments on the manuscript. This work was supported by the Association Française Contre les Myopathies (AFM) and by the US National Institutes of Health (R01 GM112923 to G.D.) G.D. is supported as an Investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

B.R.S., L.W. and Z.Z. designed and performed experiments. B.R.S., L.W., Z.Z., P.L., E.B., J.D. and I.Y. contributed to data analysis. G.D. is responsible for the project's planning and experimental design. All authors contributed to writing the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A. & Steitz, J.A. The U1 small nuclear RNA-protein complex selectively binds a 5′ splice site *in vitro*. *Cell* **33**, 509–518 (1983).
2. Will, C.L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
3. Berg, M.G. *et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53–64 (2012).
4. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).
5. Vorlová, S. *et al.* Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* **43**, 927–939 (2011).
6. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
7. Ntini, E. *et al.* Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
8. Baserga, S.J. & Steitz, J.A. The diverse world of small ribonucleoproteins. *Cold Spring Harb. Monograph Arch.* **24**, 359–381 (1993).
9. Battle, D.J. *et al.* The SMN complex: an assembly machine for RNPs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 313–320 (2006).
10. Guthrie, C. & Patterson, B. Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387–419 (1988).
11. Yong, J., Golembe, T.J., Battle, D.J., Pellizzoni, L. & Dreyfuss, G. snRNAs contain specific SMN-binding domains that are essential for snRNP assembly. *Mol. Cell. Biol.* **24**, 2747–2756 (2004).
12. Kondo, Y., Oubridge, C., van Roon, A.M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife* **4**, e04986 (2015).
13. Leung, A.K., Nagai, K. & Li, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539 (2011).
14. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009).
15. Raker, V.A., Hartmuth, K., Kastner, B. & Lührmann, R. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol.* **19**, 6554–6565 (1999).
16. Urlaub, H., Raker, V.A., Kostka, S. & Lührmann, R. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **20**, 187–196 (2001).
17. Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M.C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172–4184 (2010).
18. Battle, D.J. *et al.* The Gemin5 protein of the SMN complex identifies snRNAs. *Mol. Cell* **23**, 273–279 (2006).
19. Lau, C.K., Bachorik, J.L. & Dreyfuss, G. Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. *Nat. Struct. Mol. Biol.* **16**, 486–491 (2009).
20. Yong, J., Kasim, M., Bachorik, J.L., Wan, L. & Dreyfuss, G. Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol. Cell* **38**, 551–562 (2010).
21. Cauchi, R.J. SMN and Gemins: 'we are family' … or are we?: insights into the partnership between Gemins and the spinal muscular atrophy disease protein SMN. *BioEssays* **32**, 1077–1089 (2010).
22. Fischer, U., Englbrecht, C. & Chari, A. Biogenesis of spliceosomal small nuclear ribonucleoproteins. *Wiley Interdiscip. Rev. RNA* **2**, 718–731 (2011).
23. Chari, A. *et al.* An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell* **135**, 497–509 (2008).
24. Grimm, C. *et al.* Structural basis of assembly chaperone-mediated snRNP formation. *Mol. Cell* **49**, 692–703 (2013).
25. Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146**, 384–395 (2011).
26. Will, C.L. & Lührmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301 (2001).
27. Matera, A.G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).
28. Hamm, J., Dathan, N.A., Scherly, D. & Mattaj, I.W. Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. *EMBO J.* **9**, 1237–1244 (1990).
29. Nelissen, R.L., Will, C.L., van Venrooij, W.J. & Lührmann, R. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* **13**, 4113–4125 (1994).
30. Yong, J., Pellizzoni, L. & Dreyfuss, G. Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J.* **21**, 1188–1196 (2002).
31. Wan, L., Ottinger, E., Cho, S. & Dreyfuss, G. Inactivation of the SMN complex by oxidative stress. *Mol. Cell* **31**, 244–254 (2008).
32. Wan, L. *et al.* The survival of motor neurons protein determines the capacity for snRNP assembly: biochemical deficiency in spinal muscular atrophy. *Mol. Cell. Biol.* **25**, 5543–5551 (2005).
33. Younis, I. *et al.* Minor introns are embedded molecular switches regulated by highly unstable U6ata U6atac snRNA. *eLife* **2**, e00780 (2013).
34. Kohtz, J.D. *et al.* Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119–124 (1994).
35. Cléry, A., Blatter, M. & Allain, F.H. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
36. Cho, S. *et al.* 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 (2011).
37. Burghes, A.H. & Beattie, C.E. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat. Rev. Neurosci.* **10**, 597–609 (2009).
38. Lorson, C.L. *et al.* SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nat. Genet.* **19**, 63–66 (1998).
39. Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775–1779 (2002).
40. Young, P.J. *et al.* The exon 2b region of the spinal muscular atrophy protein, SMN, is involved in self-association and SIP1 binding. *Hum. Mol. Genet.* **9**, 2869–2877 (2000).
41. Bedford, M.T. & Clarke, S.G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33**, 1–13 (2009).
42. Tripsianes, K. *et al.* Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat. Struct. Mol. Biol.* **18**, 1414–1420 (2011).
43. Liu, K. *et al.* Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One* **7**, e30375 (2012).
44. Yong, J., Wan, L. & Dreyfuss, G. Why do cells need an assembly machine for RNA-protein complexes? *Trends Cell Biol.* **14**, 226–232 (2004).
45. Ling, S.C., Polymenidou, M. & Cleveland, D.W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
46. Piazzon, N. *et al.* Implication of the SMN complex in the biogenesis and steady state level of the signal recognition particle. *Nucleic Acids Res.* **41**, 1255–1272 (2013).
47. Cheng, D., Côté, J., Shaaban, S. & Bedford, M.T. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol. Cell* **25**, 71–83 (2007).
48. Lorson, C.L. & Androphy, E.J. The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet.* **7**, 1269–1275 (1998).
49. O'Reilly, D. *et al.* Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.* **23**, 281–291 (2013).
50. Shukla, S. & Parker, R. Quality control of assembly-defective U1 snRNAs by decapping and 5′-to-3′ exonucleolytic digestion. *Proc. Natl. Acad. Sci. USA* **111**, E3277–E3286 (2014).
51. Gabanella, F. *et al.* Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. *PLoS One* **2**, e921 (2007).
52. Li, D.K., Tisdale, S., Lotti, F. & Pellizzoni, L. SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Semin. Cell Dev. Biol.* **32**, 22–29 (2014).
53. Tisdale, S. *et al.* SMN is essential for the biogenesis of U7 small nuclear ribonucleoprotein and 3′-end formation of histone mRNAs. *Cell Reports* **5**, 1187–1195 (2013).
54. Zhang, Z. *et al.* SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585–600 (2008).
55. Zhang, Z. *et al.* Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* **110**, 19348–19353 (2013).
56. Sun, S. *et al.* ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nat. Commun.* **6**, 6171 (2015).
57. Tsuiji, H. *et al.* Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO Mol. Med.* **5**, 221–234 (2013).
58. Yamazaki, T. *et al.* FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. *Cell Reports* **2**, 799–806 (2012).

ONLINE METHODS

Cell culture, RNA interference and preparation of cell extracts. HeLa PV cells were grown in DMEM supplemented with 10% FBS, L-glutamine, penicillin and streptomycin. Cell lines were tested for mycoplasma contamination. Transfection of control siRNA or siRNAs targeting U1-70K or the SMN-complex components (Dharmacon, GE healthcare) into HeLa cells was performed with Lipofectamine RNAiMax (Invitrogen), as described by the manufacturer. After 40–48 h transfection, cells were harvested and lysed to obtain cytoplasmic extracts for *in vitro* assays, as previously described[32]. Total protein concentrations of various extracts were determined by the Bradford protein assay (Bio-Rad). For nascent-snRNP measurements, knockdown cells were metabolically labeled with 200 μM of 4-thiouridine (4-shU) during the final 2 h of siRNA transfection[33].

***In vitro* transcription and labeling of RNAs.** All snRNAs used for *in vitro* assays were prepared by standard *in vitro* transcription with a MEGAshortscript T7 transcription kit (Ambion) in the presence of 0.65 mM biotin-16-UTP (Roche) and 2.6 mM UTP[32]. Transcribed RNAs were purified by electrophoresis on a 5 M urea 6% polyacrylamide gel, precipitated with ethanol and resuspended in RNase-free water. The concentrations of the labeled snRNAs were determined by UV absorbance at 260 nm.

***In vitro* transcription and translation.** *In vitro* transcription and translation reactions were performed with each plasmid in the presence of [^{35}S]methionine with T7 TNT Quick Coupled Transcription/Translation systems (Promega), as previously described[25].

Antibodies, western blotting and immunoprecipitation. The following antibodies were used in this study; anti-SMN (2B1 or 8/SMN from BD Transaction Laboratory), anti-Gemin2 (2E17 or 3F8 from Santa Cruz Biotechnology), anti-Gemin3 (12H12), anti-Gemin4 (64I1), anti-Gemin5 (10G11), anti-Sm (Y12), anti-glutathione S-transferase (26H1 from Cell Signaling Technology), anti–U1-70K (H111 from Synaptic Systems), anti-FXR1 (6BG10) and anti-Magoh (21B12). Validation of each primary antibody is provided on the manufacturers' websites or in refs. 20 and 31. Immunoprecipitation was performed with Y12 immobilized onto magnetic Dynabeads Protein G (Invitrogen) and incubated with total HeLa cell extracts prepared in RSB-100 buffer (10 mM Tris-HCl, pH 7.8, 100 mM NaCl, and 2.5 mM $MgCl_2$) containing 0.1% NP-40 and protease-inhibitor tablet (Roche) for 1.5 h at 4 °C. Supernatants were discarded, and the beads were washed five times with RSB-500 containing 0.1% NP-40. The beads were equilibrated with RSB-150 containing 0.02% NP-40 and collected for further experiments.

Plasmid construction and expression and purification of recombinant proteins. cDNA plasmids for human U1-70K were constructed in the pET42a vector (Novagen) with N-terminal GST and C-terminal His_6 tags. Procedures for recombinant-protein expression and purification were modified from those of a previous study[6]. Proteins were affinity purified with Ni-NTA beads and eluted with 500 mM imidazole in lysis buffer. Eluted proteins were further purified by glutathione beads (GE Healthcare), a heparin column and size-exclusion chromatography. Recombinant Gemin2 and Sm proteins were expressed in *Escherichia coli* and purified as previously described[25]. For efficient protein expression, Sm BD3 did not contain C-terminal RG(G) domains. The recombinant-protein concentrations were determined by the Bradford protein assay (Bio-Rad).

***In vitro* RNA pulldown.** Cytoplasmic extracts (2 mg/mL) from HeLa cells with knockdown of the individual indicated components were incubated with biotinylated snRNAs (10 nM) in reconstitution buffer (20 mM HEPES, pH 7.9, 50 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.25 mg/mL yeast tRNAs (Sigma), protease-inhibitor tablet (Roche) and 0.2 U/μL RNasin RNase inhibitor (Promega)) in 96-well plates (20 μL/well). Binding experiments were performed with cytoplasmic cell extracts at 4 °C without addition of ATP, to avoid Sm-core assembly, which would occlude the Sm site. All reactions were carried out with gentle mixing at 750 r.p.m. with a Thermomixer (Eppendorf) for 1 h, and RNA–protein complexes were captured with M-280 streptavidin Dynabeads (Invitrogen) in 100 μL of RSB-150 buffer containing 0.02% Triton X-100, protease-inhibitor tablet (Roche) and 0.2 U/μL RNase inhibitor for an additional hour. The beads were washed five times in RSB-200 buffer containing 0.02% Triton X-100, with a Kingfisher 96 magnetic particle processor (Thermo Fisher Scientific), as previously described[31,32]. Bound proteins on the beads were eluted by boiling in 10 μL of 1× sample buffer, resolved by SDS-PAGE and detected by western blotting. Uncropped scans of western blots are provided in **Supplementary Data Set 1**.

***In vitro* snRNP assembly.** A high-throughput and quantitative assay for *in vitro* Sm-core assembly was performed as previously described[31,32]. In brief, cytoplasmic extracts (2 mg/mL) from HeLa cells were incubated with biotinylated snRNAs (10 nM) in reconstitution buffer containing 2.5 mM ATP for 1 h at 30 °C. In complementation assays, 6.25–100 ng of recombinant U1-70K proteins was preincubated in cytoplasmic extracts with U1-70K knockdown for 20 min at 30 °C, and Sm core–assembly reactions were initiated by addition of biotinylated snRNAs. Assembled snRNPs were captured by Y12-immobilized Protein G magnetic beads in 100 μL of RSB-500 buffer containing 0.1% NP-40, protease-inhibitor tablet (Roche) and 0.2 U/μL RNase inhibitor for an additional hour at 30 °C. The beads were washed three times in RSB-500 buffer containing 0.1% NP-40 and 1 mg/mL heparin, then washed two times in the same buffer without heparin, with a magnetic particle processor. The Y12-immunoprecipiated snRNPs on the beads were resuspended in 120 μL of RSB-150 buffer containing 0.02% NP-40 and 0.08 μg/mL horseradish peroxidase–conjugated NeutrAvidin (Pierce) and gently mixed for 1 h at 30 °C. The beads were washed again five times with the same washing buffer (RSB-500 and 0.1% NP-40) and resuspended in 150 μL of SuperSignal ELISA Femto substrate (Pierce) mixture. Chemiluminescence signals were detected at 495 nm with a Wallac Victor2 plate reader (Perkin-Elmer).

***In vitro* protein binding.** Recombinant GST–U1-70K proteins (2.5 μg) were immobilized on glutathione-Sepharose beads (Invitrogen) in 500 μL RSB-200 buffer containing 0.02% Triton X-100. After unbound proteins were washed away, 5 μg of recombinant or *in vitro*–translated [^{35}S]methionine-labeled SMN proteins was incubated for 1 h at 4 °C and then washed with binding buffer four times. The bound proteins were eluted by boiling in 1× sample buffer, resolved by SDS-PAGE and visualized by SimplyBlue staining (Invitrogen), silver staining, or autoradiography. The band intensities were analyzed with Bio-Rad Quantity One software. Binding experiments were performed side by side with the same protein preparations and run on the same gel. Uncropped images of gels, autoradiographs or western blots are provided in **Supplementary Data Set 1**.

Reverse transcription and RT-qPCR measurement of snRNAs. Total snRNAs were isolated with TRIzol (Invitrogen), as specified by the manufacturer. Steady-state snRNPs from HeLa cells were isolated by Y12 antibody immunoprecipitation and protease K digestion followed by phenol-chloroform extraction and ethanol precipitation. Nascent snRNPs were obtained by the same Y12 immunoprecipitation from cells that were metabolically labeled with 4-shU[33]. After isolation of steady-state snRNPs, free thiol groups of uridine in snRNA were biotinylated with 0.2 mg/ml of biotin-HPDP (Pierce) and captured by Dynabeads MyOne Streptavidin beads (Invitrogen). cDNA generation and quantification of snRNA were performed as previously described[54].