Endogenous U2·U5·U6 snRNA complexes in *S. pombe* are intron lariat spliceosomes

WEIJUN CHEN,[1,2] HENNADY P. SHULHA,[2,3] AMI ASHAR-PATEL,[1,2] JING YAN,[1,2] KARIN M. GREEN,[4] CHARLES C. QUERY,[5] NICK RHIND,[2] ZHIPING WENG,[2,3] and MELISSA J. MOORE[1,2,6]

[1]Howard Hughes Medical Institute, [2]Department of Biochemistry and Molecular Pharmacology, [3]Bioinformatics and Integrative Biology, [4]UMMS Proteomics and Mass Spectrometry Facility, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA
[5]Albert Einstein College of Medicine, Bronx, New York 10461, USA

ABSTRACT

Excision of introns from pre-mRNAs is mediated by the spliceosome, a multi-megadalton complex consisting of U1, U2, U4/U6, and U5 snRNPs plus scores of associated proteins. Spliceosome assembly and disassembly are highly dynamic processes involving multiple stable intermediates. In this study, we utilized a split TAP-tag approach for large-scale purification of an abundant endogenous U2·U5·U6 complex from *Schizosaccharomyces pombe*. RNAseq revealed this complex to largely contain excised introns, indicating that it is primarily ILS (intron lariat spliceosome) complexes. These endogenous ILS complexes are remarkably resistant to both high-salt and nuclease digestion. Mass spectrometry analysis identified 68, 45, and 43 proteins in low-salt-, high-salt-, and micrococcal nuclease-treated preps, respectively. The protein content of a *S. pombe* ILS complex strongly resembles that previously reported for human spliced product (P) and *Saccharomyces cerevisiae* ILS complexes assembled on single pre-mRNAs in vitro. However, the ATP-dependent RNA helicase Brr2 was either substoichiometric in low-salt preps or completely absent from high-salt and MNase preps. Because Brr2 facilitates spliceosome disassembly, its relative absence may explain why the ILS complex accumulates logarithmically growing cultures and the inability of *S. pombe* extracts to support in vitro splicing.

Keywords: intron lariat spliceosome; proteomic; RNAseq; spliceosome footprint; *S. pombe*

INTRODUCTION

Removal of introns from precursors to messenger RNAs (pre-mRNAs) is an essential step in eukaryotic gene expression. Introns are excised by the spliceosome, a multi-megadalton machine composed of five uridine-rich small nuclear RNAs (U1, U2, U4, U5, and U6 snRNAs) and scores of polypeptides. Many proteins stably associate with the snRNAs to form small ribonucleoprotein particles (snRNPs), while many others function as transiently interacting "splicing factors." The snRNPs and splicing factors are thought to assemble de novo on each new intron via a highly dynamic process involving multiple stable intermediates. Recruitment of a U1 snRNP to the 5′ end of the intron (5′ splice site; 5′SS) and U2 snRNP to the branch site (BS) near the 3′ end of the intron (3′ splice site; 3′SS) results in pre-spliceosome (complex A) formation. Addition of the U5·U4/U6 tri-snRNP forms the pre-catalytic spliceosome (complex B), which is then converted to the activated spliceosome (B^{act}) by a major structural rearrangement involving destabilization of U1 and U4 snRNPs and addition of the multiprotein NineTeen Complex (NTC). Further structural rearrangements result in the catalytically active complexes (B*, C1, and C2) in which the first (5′SS cleavage and lariat formation) and second (3′SS cleavage and exon ligation) chemical steps of splicing occur (Jurica and Moore 2003; Wahl et al. 2009; Hoskins and Moore 2012; Li et al. 2013). Spliced exon release from the spliced product complex (P) (Ilagan et al. 2013) results in formation of intron-lariat spliceosome complex (ILS) (Yoshimoto et al. 2009; Fourmann et al. 2013), containing the lariat intron, U2, U5, and U6 snRNAs and the NTC. ILS complex disassembly and subsequent reassembly of the U5·U4/U6 tri-snRNP completes the spliceosome cycle and readies the components for another round. Because each step in the spliceosome cycle involves the comings and goings of multiple components, each stable and isolatable complex has a signature RNA and protein composition (Jurica and Moore 2003; Wahl et al. 2009; Hoskins and Moore 2012).

[6]**Corresponding author**
E-mail melissa.moore@umassmed.edu

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.040980.113.

To date, most studies of spliceosome composition and structure have been performed on individual snRNPs or stable intermediate complexes assembled in vitro on a limited set of well-spliced pre-mRNAs (Will and Luhrmann 2011). The latter complexes are usually obtained by artificially stalling spliceosome assembly or disassembly at a particular step (Jurica and Moore 2002; Jurica 2008). When endogenous spliceosomes have been isolated, they often contain protein components representative of multiple assembly states (Stevens 2000; Ohi et al. 2002; Chen et al. 2007). One notable exception is a stable endogenous complex containing U2, U5, and U6 snRNAs observed in logarithmically growing *Schizosaccharomyces pombe* (Huang et al. 2002). Previous purification of this complex (a.k.a., *S. pombe* CDC5 complex) via a TAP-tag on CDC5 revealed it to contain numerous orthologs of proteins associated with human C complex (McDonald et al. 1999; Ohi et al. 2002; Ren et al. 2011) and to structurally resemble human C complexes upon single particle reconstruction of electron microscope images (Ohi et al. 2007).

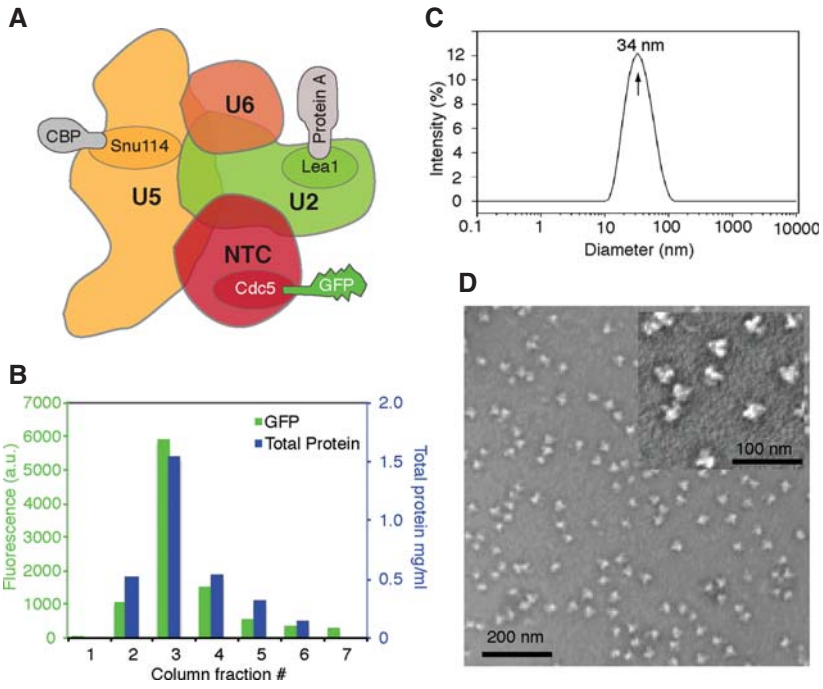

FIGURE 1. Purification of endogenous U2·U5·U6 complexes. (*A*) Scheme of the U2·U5·U6 complex showing fusion partners for GFP, protein A, and calmodulin binding peptide (CBP) tags. (*B*) Calmodulin bead elution profile of GFP fluorescence and total protein from a high-salt (HS) purification. (*C*) Dynamic light scattering profile of purified HS U2·U5·U6 complexes. (*D*) Negative stain EM images of purified HS U2·U5·U6 complexes at 66,000× and 132,000× (*inset*) magnification.

Here, we report large-scale isolation of these endogenous *S. pombe* U2·U5·U6 complexes using double-affinity purification via separate affinity tags affixed to two different proteins. Purifications were carried out under different stringencies to reveal both loosely and tightly bound factors. Comprehensive RNA and protein compositional analyses reveal this complex to consist largely of endogenous ILS complex. The absence of SF3A/B components coupled with the micrococcal nuclease cleavage pattern of U2 snRNA reveals new information about spliceosome conformational and compositional changes following spliced exon release. Finally, the relative absence of proteins required for spliceosome disassembly may explain why this complex accumulates in rapidly dividing *S. pombe* and why extracts from these cells fail to support in vitro splicing.

RESULTS

Split TAP-tag purification of endogenous *S. pombe* U2·U5·U6 complexes

S. pombe cell lysates contain significant amounts of a U2·U5·U6 snRNA complex indicative of post-activation spliceosomes (Huang et al. 2002; Ohi et al. 2002; Ren et al. 2011). To specifically purify this complex, we utilized a split TAP-tag approach, with protein A attached to U2 snRNP protein Lea1 (U2 A′ in human) and calmodulin binding peptide (CBP) attached to U5 snRNP protein Snu114 (U5 116K in human). An additional GFP tag on Cdc5 (an NTC component), Spp42 (U5 snRNP protein Prp8), or Brr2 (another U5 protein) allowed for real-time monitoring of later purification steps, as well as assessment of the extent to which the GFP-tagged species copurified with Lea1 and Snu114 (Fig. 1A). All three tags were introduced by homologous recombination into haploid cells, with tagged proteins replacing their untagged counterparts. Growth rates and cell morphologies of all triple-tagged strains were indistinguishable from the parental untagged strain (data not shown), indicating functionality of the fusion proteins.

To maximize yields, cells were grown in 5× YES media and harvested in late log phase (OD600 = ~23). Following cell lysis, immunoglobulin G (IgG)-sepharose and calmodulin affinity steps were carried out as described (Materials and Methods). Physiological (low salt; LS) complexes were obtained from 400 mM KCl cell lysate and purified in 150 mM NaCl. Increasing lysate KCl to 700 mM significantly increased yields, but 1 M total salt resulted in very low yields (Supplemental Fig. S1A). Salt-stable core (high salt; HS) complexes were thus obtained from 700 mM KCl lysates and exposed to 700 mM NaCl during early purification stages to remove loosely bound factors. Treatment of these HS complexes with MNase (MN) during the IgG-

sepharose step removed solvent-accessible RNA sequences. In all three preps (LS, HS, and MN), total protein and CDC5-GFP fluorescence exhibited identical calmodulin elution profiles (Fig. 1B; Supplemental Fig. S1B). Generally, 6 L of culture yielded ~250 g packed cell pellet from which we were able to purify ~2 mgs of endogenous U2·U5·U6 snRNA complexes.

To assess complex integrity and hetereogeneity, HS complexes were analyzed by dynamic light scattering (DLS) and negative stain electron microscopy (EM). DLS revealed one major species with an estimated diameter of 34 nm and MW of ~2.5 MDa (Fig. 1C). Consistent with this, electron micrographs (Fig. 1D) revealed largely monodispersed (>95% monomer, <5% dimer) 30- to 35-nm particles of uniform size and reproducible shape similar to human C complex (Jurica et al. 2004) and *S. pombe* CDC5 complex (Ohi et al. 2007). Negative stain EM images of LS and MN particles yielded similar results (Supplemental Fig. S1C).

Protein composition of endogenous U2·U5·U6 complexes

SDS-PAGE revealed nearly identical protein banding patterns for the LS, HS, and MN preps (Fig. 2A, left). Quantitative analysis by liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) revealed 29 abundant proteins common to all three preps (Fig. 2A, right; Table 1). Consistent with other analyses of in vitro-assembled spliceosomes (Schmidt et al. 2010; Agafonov et al. 2011; Ilagan et al. 2013), the most abundant species was Prp19, having a three- to fourfold higher stoichiometry than any other polypeptide. In isolation, Prp19 is a known tetramer (Ohi et al. 2005). All other abundant species were present at an approximately 1:1 stoichiometry, with the exception of translation elongation factor EF1A, which was significantly reduced in the MN prep. Multiple substoichiometric proteins were also present, with a total of 68, 45, and 43 species being consistently quantifiable in the LS, HS, and MN preps, respectively (Table 1).

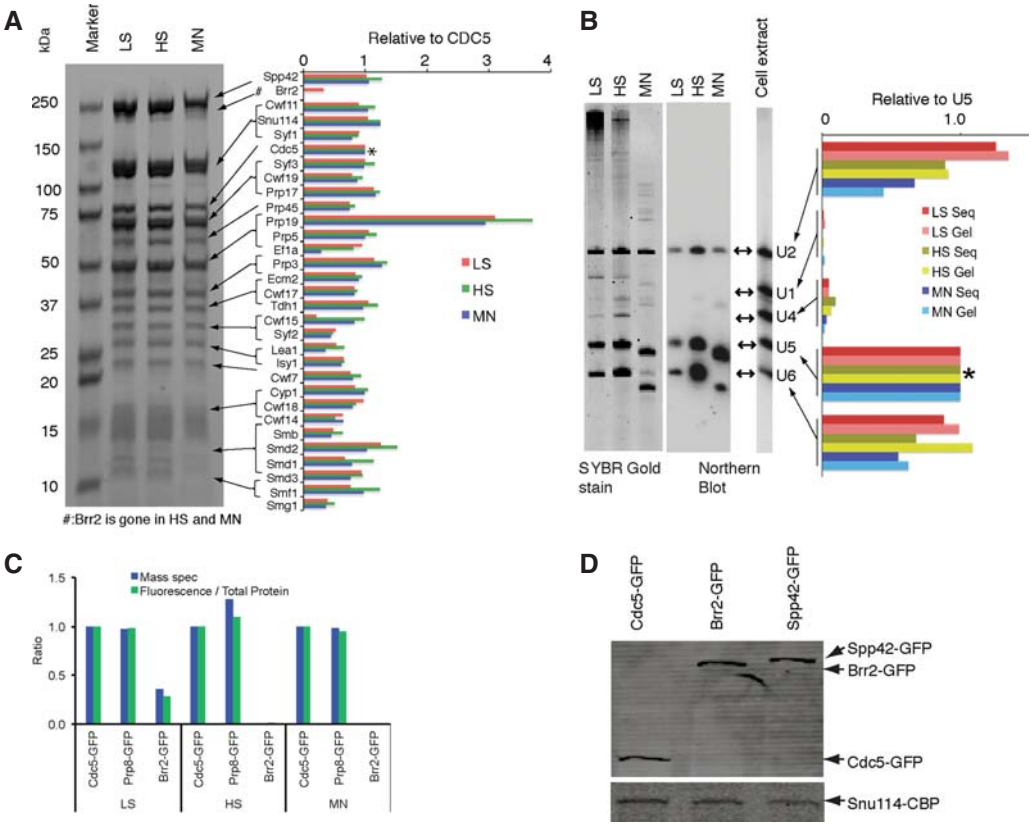

FIGURE 2. Protein and RNA compositions of purified U2·U5·U6 complexes. (*A*) Protein composition. *Left*: Coomassie blue-stained SDS-PAGE of total protein from LS, HS, and MN preps. *Right*: Abundances of major proteins relative to CDC5 (*) in each prep. Protein stoichiometries were estimated by quantitative mass spec (see Materials and Methods). (*B*) RNA composition. *Left*: SYBR Gold staining of LS, HS, and MN prep RNAs resolved by denaturing PAGE. *Center*: Northern blots using probes for U1, U2, U4, U5, and U6 snRNAs of same RNAs at *left*, plus RNA purified from total cell extract. *Right*: Relative snRNA stoichiometries to U5 (*) in LS, HS, and MN preps as determined by SYBR Gold staining (Gel) and RNAseq (Seq). (*C*) Endogenous *S. pombe* U2·U5·U6 complexes contain substoichiometric or undetectable amounts of Brr2 (related to *A* and *C*). Bar graph of amounts of Spp42 (Prp8) and Brr2 relative to CDC5 in LS, HS, and MN preps. Fluorescence signals were compared by normalizing to total protein. (*D*) Western blots showing expression levels of GFP-tagged proteins in cell extracts, with the expression levels of CBP-tagged on Snu114 in cell extracts as control.

TABLE 1. Proteins detected in LS, HS, and MN preps

Proteins						LS		HS		MN	
#	*S. pombe*	*S. cerevisiae*	Human	Accession number[a]	MW (kDa)	# Peptides	RS[b]	# Peptides	RS[b]	# Peptides	RS[b]
1	Prp19	Prp19	hPRP19	SPAC29A4.08c	54	23	2.95	19	2.99	20	2.94
2	Prp3	Prp3	hPRP3	SPAC3A12.11c	44	17	1.07	18	1.23	16	1.28
3	Smd2	Smd2	Smd2	SPAC2C4.03c	13	7	1.03	6	1.10	6	1.03
4	Syf3	Syf3	hSYF3	SPBC31F10.11c	81	40	1.00	39	1.03	39	0.98
5	Prp5	Prp46	PRL1	SPBP22H7.07	52	25	1.00	22	1.00	23	1.00
6	Cdc5	Cef1	Cdc5L	SPAC644.12	87	50	1.00	50	1.00	49	1.00
7	Cwf15	Cwc15	CCAP2	SPBC337.06c	30	11	0.99	9	1.00	12	0.97
8	Prp17	Prp17	hPRP17	SPBC6B1.10	63	31	0.98	27	0.97	33	1.17
9	Snu114	Snu114	U5-116K	SPBC215.12	111	49	0.98	48	0.96	51	1.25
10	Spp42	Prp8	U5-220K	SPAC4F8.12c	275	76	0.97	83	0.94	88	1.06
11	Ef1a	Tef1	Ef1-α1	SPCC794.09c	50	10	0.95	8	0.93	8	0.29
12	Smd3	Smd3	Smd3	SPBC19C2.14	11	8	0.94	6	0.92	6	0.76
13	Cwf18		CCDC12	SPCP1E11.07c	17	11	0.92	9	0.86	10	0.80
14	Tdh1	Tdh1		SPBC32F12.11	36	15	0.23	17	0.86	18	0.77
15	Syf1	Syf1	hSYF1	SPBC211.02c	93	39	0.85	36	0.84	40	0.78
16	Cwf11		KIAA0560	SPBC646.02	148	32	0.84	35	0.81	34	1.04
17	Ecm2	Ecm2	RBM22	SPCC550.02c	40	28	0.80	23	0.77	25	0.90
18	Cwf17		U5-40K	SPBC1289.11	37	21	0.77	18	0.77	19	0.82
19	Cwf19		CWF19L2	SPAC30D11.09	74	45	0.76	46	0.76	44	0.87
20	Cyp1		PPIL1	SPAC57A10.03	17	9	0.74	7	0.75	8	0.99
21	Smf1	Smf1	Smf1	SPBC3E7.14	9	7	0.73	6	0.72	5	0.97
22	Cwf7		Spf27	SPBC28F2.04c	21	16	0.71	17	0.71	16	0.80
23	Smd1	Smd1	Smd1	SPAC27D7.07c	13	7	0.71	6	0.68	5	0.79
24	Prp45	Prp45	SKIP	SPCC188.11	63	32	0.67	28	0.67	27	0.74
25	Isy1	Isy1	hISY1	SPBC32F12.05c	26	14	0.65	10	0.66	12	0.62
26	Cwf14	Dub31	G10	SPBC24C6.11	17	8	0.63	6	0.53	7	0.64
27	Lea1	Lea1	U2A′	SPBC1861.08c	27	13	0.58	10	0.53	10	0.36
28	Syf2	Syf2	Syf2	SPBC3E7.13c	28	15	0.49	14	0.51	11	0.44
29	Smb	Smb1	Smb	SPAC26A3.08	15	9	0.43	8	0.42	7	0.44
30	Smg1	Smg1	Smg1	SPBC4B4.05	9	6	0.36	4	0.23	3	0.37
31	Cwf16	Yju2	CCDC130	SPBC18H10.10c	34	9	0.13	8	0.20	9	0.21
32	Gpd3			SPBC354.12	36	6	0.14	4	0.10	4	0.12
33	Ssa2	Ssa1	CCAP1	SPCC1739.13	70	9	0.24	6	0.10	7	0.11
34	Pabp	Pab1	Pabp1	SPAC57A7.04c	72	13	0.14	9	0.08	0	
35	Cps8	Act1	Actg1	SPBC32H8.12c	42	10	0.19	8	0.06	8	0.14
36	Bis1			SPCC364.02c	43	10	0.15	9	0.02	12	0.17
37	Rpl13	Rpl13	Rpl13	SPAC664.05	24	6	0.40	1		0	
38	Sde2		C1orf55	SPAC31G5.18c	29	4	0.39	0		0	
39	Brr2	Brr2	U5-200K	SPAC9.03c	249	4	0.36	0		0	
40	Rps18-1	Rps18a	Rps18	SPBC16D10.11c	17	2		0		0	
41	Rpl8-1	Rpl2	Rpl8	SPAC1F7.13c	27	6	0.34	1		1	
42	Rps11-2	Rps11a	Rps11	SPAC144.11	18	3	0.27	0		0	
43	Rpl4-1	Rpl4	Rpl4	SPBP8B7.03c	40	6	0.25	2		4	0.09
44	Prp43	Prp43	hPrp43	SPBC16H5.10c	84	4	0.25	0		0	
45	Rps1-2	Rps1a	Rps1	SPAC22H12.04c	29	4	0.25	0		0	
46	Ntr1 like	Ntr1	Tfip11	SPAC1486.03c	92	4	0.24	0		0	
47	Orphan			SPAC17A2.08c	31	4	0.18	0		0	
48	Rpl25a	Rpl25a	Rpl25	SPBC106.18	16	2		0		0	
49	Rpl12-1	Rpl12a	Rpl12	SPCC16C4.13c	18	5	0.14	1		2	
50	Rpl19-2	Rpl19	Rpl19	SPCC1682.14	23	5	0.12	1		2	
51	Sap145	Cus1	Sf3b145	SPAC22F8.10c	69	3	0.12	0		0	
52	Mcp60	Hsp60	Hspd1	SPAC12G12.04	62	5	0.11	1		0	
53	Rps2	Rps2	Rps2	SPCC576.08c	28	3	0.11	0		1	
54	Rpl3	Rpl3a	Rpl3	SPAC17A5.03	44	2		0		1	
55	Rpl20	Rpl20a	Rpl20	SPAC26A3.04	21	2		0		1	
56	Rps7	Rsp7	Rps7	SPAC18G6.14c	22	6	0.10	2		0	
57	prp22	Prp22	hPrp22	SPAC10F6.02c	131	4	0.10	0		0	
58	Mug161	Drn1	Cwf19L2	SPAC1F3.09	63	2		0		0	

(*continued*)

TABLE 1. *Continued*

Proteins						LS		HS		MN	
#	*S. pombe*	*S. cerevisiae*	Human	Accession number[a]	MW (kDa)	# Peptides	RS[b]	# Peptides	RS[b]	# Peptides	RS[b]
59	Rpl23-1	Rpl23a	Rpl23	SPAC3G9.03	15	2		0		0	
60	Sme1	Sme1	Sme	SPBC11G11.06c	10	5	0.08	1		2	
61	Cip1	Pin4		SPBC16A3.18	53	2		0		0	
62	Rps8-2	Rps8b	Rps8	SPAC521.05	23	2		0		0	
63	Sap155	Hsh155	Sf3b155	SPAC27F1.09c	135	2		0		0	
64	Sum3	Ded1	Ddx3	SPCC1795.11	70	1		0		0	
65	Cip2			SPAC12G12.03	62	2		0		0	
66	Rps15-2	Rps15a	Rps15	SPAC1071.07c	18	2		0		0	
67	Rps3	Rps3	Rps3	SPBC16G5.14c	28	1		0		0	
68	Sap130	Prp12	Sf3b130	SPAPJ698.03c	135	2		1		0	

[a]The accession numbers were taken from www.pombase.org.
[b](RS) Relative stoichiometry to CDC5 protein.

The additional polypeptides in the LS prep were predominantly ribosomal proteins.

With the exception of EF1A and Tdh1, all abundant polypeptides were proteins previously known to associate with spliceosomes. Most highly represented were the Sm snRNP core proteins, the NTC, the NTC-related complex, and U5 snRNP. Together, the latter three species comprise the 35S U5 complex (Makarov 2002; Makarova et al. 2004). The only abundant U2-specific protein was Lea1 (U2 A′), to which the protein A tag was attached. Notably, no substituents of SF3A or SF3B (U2 snRNP-associated complexes required for early spliceosome assembly steps) were detectable in either the HS or MN prep; this was consistent with the predominant purified species being a post-activation complex, as SF3A and B are destabilized coincident with spliceosome activation (Bessonov et al. 2008; Lardelli et al. 2010). Also notably absent from the HS prep were the second step factors Slu7, Prp22, and Prp18, all of which are abundant components of *Saccharomyces cerevisiae* C complex purified under LS-like conditions (Fabrizio et al. 2009). Slu7 was found in human P complex (Ilagan et al. 2013), and *S. cerevisiae* ILS complex (Fourmann et al. 2013) but was undetectable in our preparations (Fig. 5, below). The only known second step factor present in stoichiometric amounts in the HS and MN preps was Prp17, as has also been observed in human salt-stable C complex core preparations (Bessonov et al. 2008), human P complex (Ilagan et al. 2013), and *S. cerevisiae* ILS complex (Fourmann et al. 2013).

Surprisingly, the U5-specific protein Brr2 was threefold less abundant than other U5 proteins (e.g., Spp42, Snu114, and Cwf17) in the LS prep and undetectable in the HS and MN preps (Table 1; Fig. 2A). Consistent with this, Brr2-GFP fluorescence was threefold reduced relative to either CDC5-GFP or Spp42-GFP in LS preps and was completely absent from HS preps (Fig. 2C). Western blotting revealed that this was not due to poor Brr2-GFP expression, as its levels in cell extracts were comparable to those of CDC5-GFP and Spp42-GFP (Fig. 2D). Thus, Brr2 is not stably associated with endogenous *S. pombe* U2·U5·U6 complexes. In contrast, Brr2 is an abundant component of both *S. cerevisiae* and human C complexes (Bessonov et al. 2008; Fabrizio et al. 2009) as well as human P complex (Ilagan et al. 2013) and *S. cerevisiae* ILS complex (Fourmann et al. 2013). Thus, with regard to their protein content, endogenous *S. pombe* U2·U5·U6 complexes are similar, but not identical, to human C and P complexes and *S. cerevisiae* C and ILS complexes.

RNA composition of purified complexes

RNAs in the LS, HS, and MN preps were analyzed both by denaturing PAGE and high-throughput sequencing (RNAseq) (see Materials and Methods). SYBR Gold staining of gels revealed three major 100- to 200-nt bands in the LS and HS preps (Fig. 2B, left). Northern blotting confirmed that these were U2, U5, and U6 snRNAs (Fig. 2B, center). SYBR Gold staining and RNAseq revealed three snRNAs to be present at comparable ratios (Fig. 2B, right). The next most abundant small RNA species were U4 and U1, with relative stoichiometries of 0.05 and 0.01, respectively. Other small RNAs (e.g., snoRNAs and tRNAs) collectively constituted <0.5% of the total. Thus, the bulk of our purified complexes constituted post-activation spliceosomes (i.e., subsequent to U1 and U4 departure).

In the MN sample, both U5 and U6 snRNAs were quantitatively converted to faster migrating species, but their relative stoichiometries remained largely unchanged. RNAseq revealed that MNase removed 4 nt from the 5′ end of U5 and 5 nt from the 3′ end of U6 (Fig. 3A,B). Although migration of U2 was apparently unaffected by MNase treatment, its relative stoichiometry was reduced by nearly half. Analysis of the 5′ and 3′ termini of U2-mapping RNAseq tags in the MN library indicated two major cleavage sites: (1) within the GUAGUA branch recognition sequence; and (2) at the base of the penultimate stem–loop (Fig. 3B). The region between these two cleavage sites (within which our Northern probe

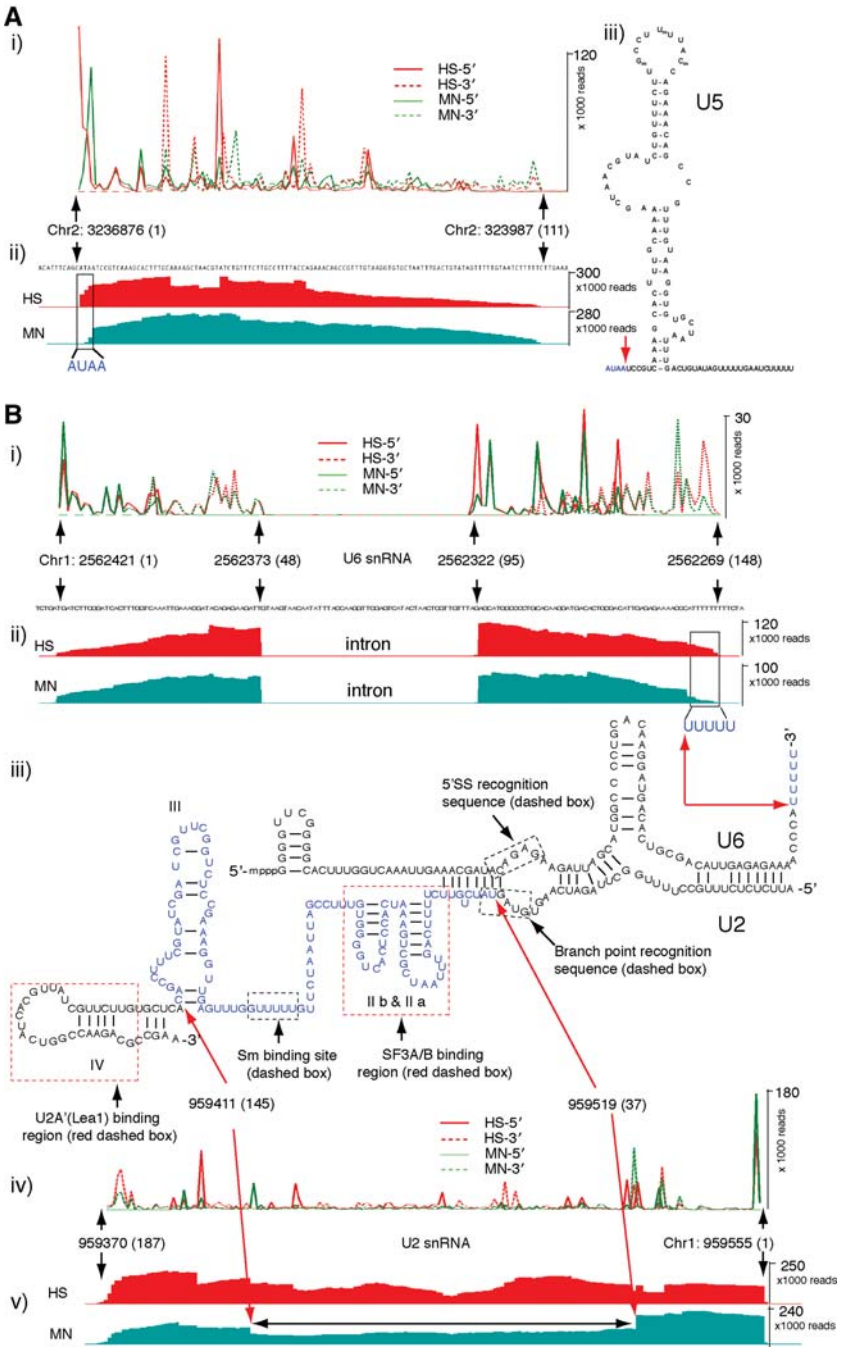

FIGURE 3. MNase cleavage site in U5, U6/U2 snRNAs. (*A*) MNase cleavage site in U5 snRNA. (i) Distribution of 5′ and 3′ tag ends in HS and MN libraries on U5 snRNA. (ii) UCSC screen shot showing full length *S. pombe* U5 snRNA. (iii) U5 secondary structure and MNase cleavage site (red arrow). (*B*) MNase cleavage sites in U6/U2 snRNAs. (i) Distribution of 5′ and 3′ tag ends in HS and MN libraries on U6 snRNA. (ii) UCSC screen shot showing full length *S. pombe* U6 snRNA. (iii) U6 secondary structure (5′–3′) and MNase cleavage site (red arrow) and U2 secondary structure (3′–5′) and MNase cleavage site (red arrow). (iv) UCSC screen shot showing full length *S. pombe* U2 snRNA. (v) Distribution of 3′ and 5′ tag ends in HS and MN libraries on U2 snRNA.

was located) was twofold underrepresented in the RNAseq data compared to the 5′ and 3′ ends. Additional Northern blotting with probes complementary to other regions in U2 confirmed the existence of numerous fragments in the MN sample (Supplemental Fig. S2). Thus, our purified complexes either contain two distinct U2 conformers (one fully protected from nuclease digestion and another in which the region between the branch recognition sequence and the penultimate stem–loop is nuclease-accessible), or this region of U2 snRNA is kinetically less accessible to MNase digestion than are the 5′ and 3′ ends of U5 and U6 snRNA, respectively.

Genome-wide analysis of RNAseq reads

In addition to U2, U5, and U6 snRNAs, heterogeneous higher molecular weight species were clearly present in all three samples (Fig. 2B, left). To identify these, we performed genome-wide analysis of our RNAseq tags using a stringent mapping algorithm (see Materials and Methods; see Supplemental Table S1A for data statistics). Of reads mapping to the *S. pombe* genome, 98%, 98%, and 99% (LS, HS, and MN libraries, respectively) mapped both within and sense to annotated genes. Further, of the nongenome mapping reads, 94%–99% mapped sense to exon-exon junctions in the predicted transcriptome. Thus, almost all of our mappable reads corresponded to the annotated *S. pombe* transcriptome (Rhind et al. 2011).

Known *S. pombe* genes can be divided into seven broad classes: CDS, rRNAs, snRNAs, snoRNAs, tRNAs, long terminal repeat (LTR)-containing transcripts, and other miscellaneous RNAs. The vast majority (≥99.5%) of tags mapping to known genes (including introns) mapped to the snRNA, rRNA, or CDS class, but the relative abundances among these classes varied among the three preps (Fig. 4A). Consistent with higher ribosomal protein abundance in the LS prep (Table 1), 40% of LS tags were from rRNA, compared to only 10% and 20% of HS and MN tags. Fewer tags in the MN library mapped to the CDS class (6%) than in the LS (17%) and HS (14%) libraries. This was also expected, as MNase should eliminate any sequences not directly protected by the spliceosome (see below).

There was excellent correlation between tag abundances mapping to individual CDSs across all three preps ($r > 0.8$),

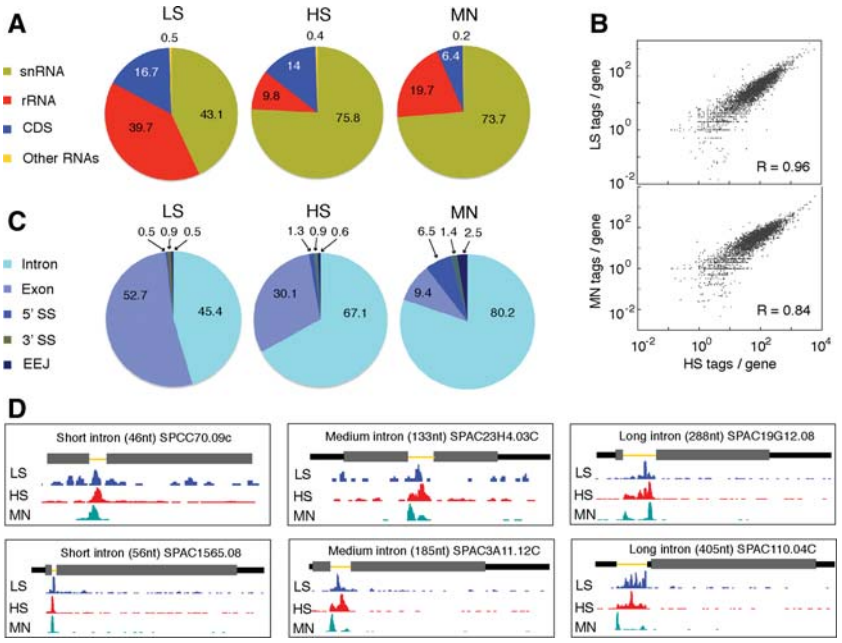

FIGURE 4. RNAseq data for LS, HS, and MN endogenous U2·U5·U6 complex preps. (*A*) Pie charts showing class distributions of transcript mapping reads for LS (*left*), HS (*middle*), and MN (*right*) libraries. The "Other RNAs" class includes snoRNAs, tRNAs, misc-RNAs, and long terminal repeats (LTRs). (*B*) Log_2 plots of tag abundance per CDS gene reveal excellent correlations between LS, MN, and HS libraries. (*C*) Pie charts showing distributions of intron, exon, 5′SS, 3′SS, and exon-exon junction mapping reads among the LS (*left*), HS (*middle*), and MN (*right*) libraries. (*D*) Two examples each of genes with a short (*left*: 46 and 56 nt), medium length (*middle*: 133 and 185 nt), or long (*right*: 288 and 405 nt) intron showing coverage of LS, HS, and MN tags along the gene. Thick gray rectangles: coding regions; medium black lines: UTRs; thin yellow lines: introns. Consistent with the pie chart in *C*, the LS library contains the most exon-mapping reads, whereas MN reads are almost exclusively confined to introns. Note the single MN peak on short introns and double MN peaks on medium length and long introns.

indicating the high quality of our data (Fig. 4B). Out of the 2345 annotated intron-containing CDS genes in *S. pombe*, only 73 had no uniquely mapping reads in the MN data set and <5 reads in the HS data set. Close reexamination of the *S. pombe* transcriptome RNAseq data (Rhind et al. 2011) revealed that 70 of these are misannotated introns, while the remaining three are poorly expressed in late log phase (W Chen, H Shulha, N Rhind, Z Weng, M Moore, in prep.). Thus, our data cover 99.9% of intron-containing CDS genes.

Endogenous U2·U5·U6 complexes contain predominantly excised introns

Among CDS mapping tags, the fraction mapping to introns and exons of single-intron genes varied among the three preps in expected ways (Fig. 4C; Supplemental Table S1C). The vast majority of MN tags were intronic, consistent with the idea that only regions protected by spliceosomes survive MNase digestion. In the LS and HS preps, both of which initially contained intact RNAs, the percentage of total tags mapping to exons was higher in the LS prep than in the HS prep. This suggests that exonic regions were preferentially dissociated upon high-salt treatment. All of these data features were readily apparent on individual genes (Fig. 4D).

To more rigorously examine the possibility that exonic regions were less stably bound to our purified complexes than intronic regions, we calculated intron and exon tag densities for the most abundant single-intron CDS genes represented in the LS and HS preps (Table 2). In both preps, 5′ and 3′ exon tag densities were highly similar, indicating no preferential dissociation of the 5′ exon as might be expected for splicing intermediate-containing complexes. Further, the higher intron:exon ratio in the HS prep (12.7–14.4) than the LS prep (6.5–7.4) indicates that exonic sequences were prone to dissociation under high-salt conditions. Thus, the predominant CDS RNA species (93% for the HS prep) consisted of introns no longer associated with exons. This indicates that endogenous *S. pombe* U2·U5·U6 snRNA complexes are predominantly ILS complexes.

DISCUSSION

In this paper, we describe large-scale purification and comprehensive proteomics and ribonomics characterization of an abundant endogenous splicing complex containing U2, U5, and U6 snRNAs from *S. pombe* (Fig. 1A). RNAseq revealed that these endogenous complexes contain predominantly spliced intron products representing 99.9% of intron-containing CDS genes. Consistent with this, the set of proteins stably associated with this endogenous U2·U5·U6 complex strongly resembles the proteins associated with intron lariat spliceosomes assembled in vitro on a single intron from *S. cerevisiae* (Fourmann et al. 2013). Thus, spliceosome disassembly from intron lariats is likely rate-limiting for pre-mRNA splicing during logarithmic

TABLE 2. Intron/exon tag densities and 5′SS, 3′SS numbers

Items	LS	HS
Average intron tag density/gene[a]	0.390	0.663
Average 5′ exon tag density/gene[a]	0.053	0.052
Average 3′ exon tag density/gene[a]	0.060	0.046
Intron tag density:5′ exon tag density	7.4	12.7
Intron tag density:3′ exon tag density	6.5	14.4
ILS complex/all purified complexes (%)	88	93

Data from the top 10% of single intron genes with signal.
[a]Per million mappable reads.

growth in *S. pombe* cells. A potential explanation for slow ILS complex disassembly is the relative absence of the ATPase Brr2, which is responsible for facilitating structural transitions involving U2 and U6 snRNAs.

Protein compositional similarities and differences between C, P, and ILS complexes

In comparison to the stable complexes observable during spliceosome assembly and activation, the complexes involved in spliceosome disassembly have been much less studied. One paper described isolation of two post-splicing complexes, Intron Large (IL) and Intron Small (IS), from in vitro splicing reactions in HeLa nuclear extract containing a chicken c-crystallin splicing substrate with three MS2 binding sites in the intron (Yoshimoto et al. 2009). Purified IL complex contained U2, U5, and U6 snRNAs and a number of bound proteins, whereas IS complex consisted largely of the naked intron. More recently, two other studies also employing the MS2 purification approach from in vitro splicing reactions reported much more comprehensive proteomics analyses of two late-stage splicing complexes: (1) HeLa P complex stalled on a splicing substrate derived from AdML pre-mRNA containing a very short (13-nt) 3′ exon that does not allow for efficient spliced exon release (Ilagan et al. 2013); and (2) *S. cerevisiae* ILS complex assembled on a splicing substrate derived from the Actin7 (Act7) pre-mRNA (Fourmann et al. 2013).

As expected, the protein composition of endogenous *S. pombe* ILS complexes closely resembles that of in vitro-assembled *S. cerevisiae* ILS complex (Fig. 5). Proteins shared between human P, *S. cerevisiae* ILS, and endogenous *S. pombe* ILS complexes include (with notable exceptions as discussed below) most U5, NTC, and NTC-related proteins. Taking into account the expected twofold and three- to fourfold stoichiometries of the Sm proteins and Prp19, respectively, and the snRNAs and protected intronic regions, the calculated MW of *S. pombe* ILS complexes is 2.1 MDa (MN) to 3.5 MDa (LS). This is very similar to human C complex (estimated MW 2.6 MDa), consistent with previous EM reconstructions showing strong structural similarities between these two complexes (Jurica et al. 2004; Ohi et al. 2007).

Nonetheless, numerous proteins stably associated with human C complex are missing from all three post-splicing complexes characterized to date. Several are displaced upon spliced exon dissociation by Prp22 (Fourmann et al. 2013). These include the second step factors Slu7 and Prp18 (Zhang and Schwer 1997), the RES complex (Lst3, Bud13, Pml1) that functions in nuclear retention of spliceosomes (Dziembowski et al. 2004), and NTC components Cwc21 and Cwc22. Cwc21 interacts directly with Prp8 and influences the fidelity of splicing (Grainger et al. 2009). Cwc22 was recently shown to serve as the binding platform for exon junction complex (EJC) assembly on the spliced exons (Barbosa et al. 2012; Steckelberg et al. 2012). As expected for complexes predominated by lariat introns, we detected no EJC proteins in any of our endogenous ILS complex preparations.

U2 snRNP proteins and U2 snRNA conformation

In comparison to human P complex and *S. cerevisiae* ILS complex, a number of proteins are significantly underrepresented in or completely absent from *S. pombe* ILS complexes. Of particular note are all specific components of 17S U2 snRNP (Brosi et al. 1993a; Krämer et al. 1999) except for U2A′ (Lea1), to which one of our affinity tags was attached. One missing protein is U2B″ (Msl1), an RNA recognition motif (RRM)-containing protein that binds the 3′-most structural element in U2 snRNA (stem–loop IV) as a heterodimer with U2A′ (Scherly et al. 1990; Williams and Hall 2011). The complex of U2 snRNA, the Sm proteins, and the U2A′:U2B″ heterodimer comprise 12S U2 snRNP (Boelens et al. 1991). Although Msl1 is essential for vegetative growth in *S. pombe* (Kim et al. 2010), it is inessential in *S. cerevisiae* (Tang et al. 1996). Together with Lea1, Msl1 has been implicated in pre-spliceosome formation and exhibits synthetic lethality with Mud2, a splicing factor involved in initial recruitment of U2 snRNP to the branch site (Abovich et al. 1994; Tang et al. 1996; Caspary 1998). Taken together, these data suggest that Msl1 functions primarily in early spliceosome assembly stages but is not required for overall spliceosome integrity. Although we observed no growth defects in our tagged *S. pombe* strains compared to the parental strain, it is possible that the presence of the protein A tag on Lea1 destabilized Msl1 association with U2 snRNA.

The remainder of 17S U2 snRNP is comprised of the nine SF3A and B proteins (SF3A: SF3a60, SF3a66, SF3a120; SF3B: SF3b14b, SF3b14a, SF3b49, SF3b130, SF3b145, and SF3b155) (Behrens et al. 1993; Hodges and Beggs 1994; Bessonov et al. 2008). During U2 snRNP assembly in vitro, SF3B proteins bind first and then recruit the SF3A proteins (Behrens et al. 1993; Brosi et al. 1993a; Krämer et al. 1999). SF3B components bind U2 snRNA stem–loop IIa, whereas SF3A complex binds stem–loops IIa and IIb (Yan and Ares 1996; Lin and Xu 2012). SF3A and B also protect the region around the GUAGUA branch recognition sequence from chemical modification (Behrens et al. 1993; Yan and Ares 1996).

Both SF3A and SF3B are required for early spliceosome assembly (Brosi et al. 1993a,b). At that stage, SF3B components make extensive interactions with the pre-mRNA to either side of the branch site and are important for U2 snRNP recruitment and retention (Brosi et al. 1993a; Gozani et al. 1996). Three studies had suggested that SF3B's interaction with the rest of the spliceosome is destabilized upon catalytic activation of the human and *S. cerevisiae* spliceosomes (Bessonov et al. 2008, 2010; Lardelli et al. 2010). More recently, however, Coltri et al. (2011) identified SF3B proteins in human C complex; they proposed that SF3A and SF3B release their hold on U2 snRNA during spliceosome activation but maintain their grip on the intron, where they help to hold the

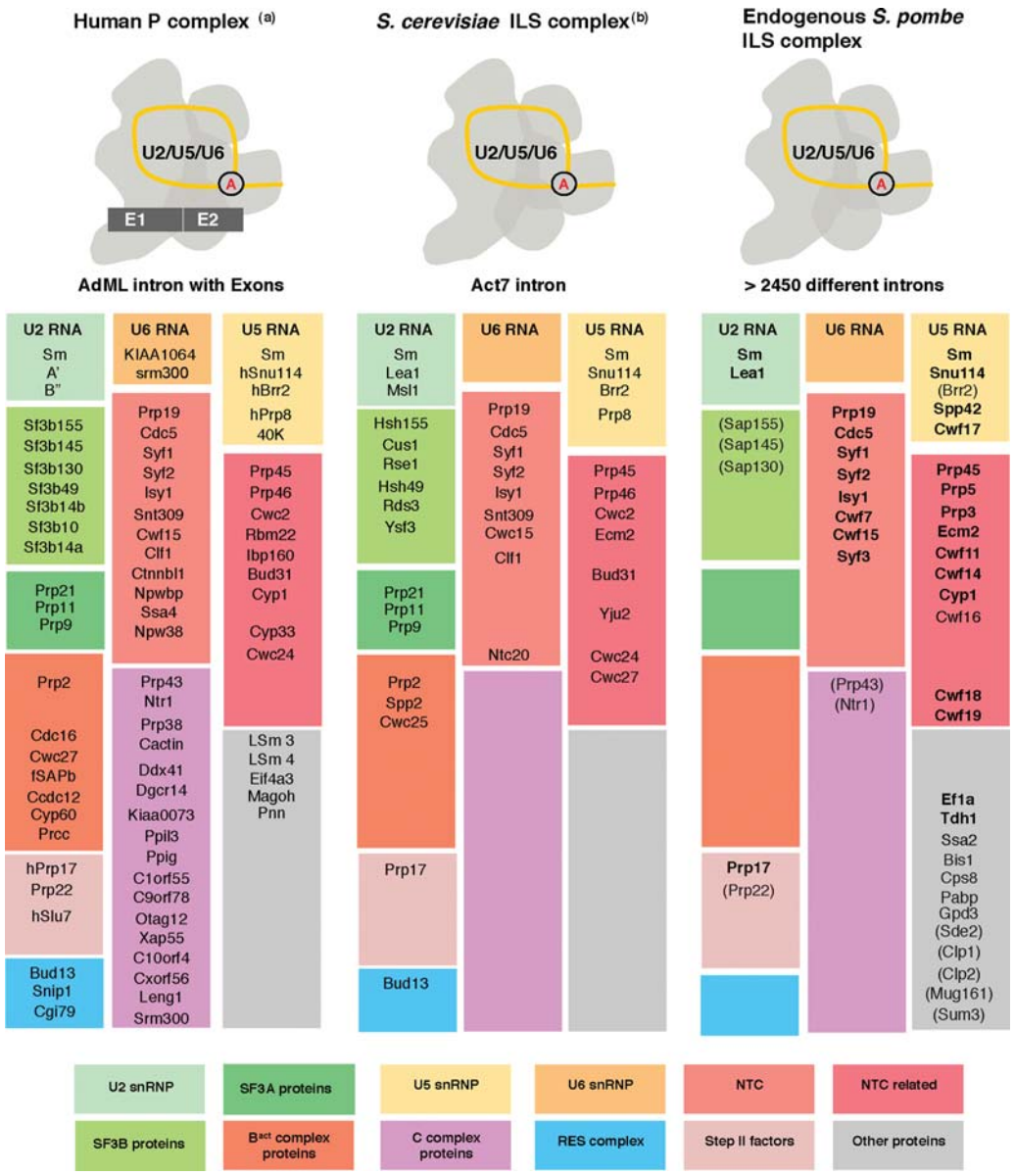

FIGURE 5. Proteins and RNAs composition of endogenous *S. pombe* ILS complex, *S. cerevisiae* ILS complex, and human P complex. Protein and snRNA compositions of human P complex[(a)] (*left*; containing AdML spliced intron without releasing exons) (Ilagan et al. 2013), *S. cerevisiae* ILS complex[(b)] (*middle*; only containing single Act7 spliced inton) (Fourmann et al. 2013), and endogenous *S. pombe* U2·U5·U6 complexes (*right*; containing more than 2450 spliced introns). Proteins and snRNAs are grouped and color-coded according to complex association or function as indicated at the *bottom*. More abundant proteins are indicated in bold text; empty colored rectangles indicate the absence of proteins in that structural or functional category.

splicing intermediates in place. Consistent with this, SF3b145 and SF3b130 were recently detected in human P complex (Ilagan et al. 2013), and most of the SF3A/B proteins were identified in *S. cerevisiae* ILS complex (Fourmann et al. 2013).

Our data indicate that SF3A components are completely absent from endogenous *S. pombe* ILS complexes and that the SF3B proteins are only loosely bound—they were substoichiometric in LS complexes and completely absent from HS and MN complexes (Fig. 2B; Table 1). Consistent with this, our RNAseq results revealed two major MNase cleavage sites (↓) in U2 snRNA: one within the GUAG$^{\downarrow}$UA branch recognition sequence and another at the 5′ end of stem–loop IV (Fig. 3B). Protection and retention of stem–loop IV is expected because this is the binding site of Lea1 (see above). Unexpected from previous data, however, is nuclease accessibility of the GUAGUA branch recognition sequence. A recent detailed study of snRNA nucleotide accessibility to chemical modification revealed only minimal modification of the GUAGUA sequence in pre- and post-activation spliceosomes (human B, B^{act}, and C complex) (Anokhina et al. 2013)

compared to downstream nucleotides (i.e., 3′ from the GUAGUA region). Our data thus suggest that after spliced exon release, the GUAGUA region becomes more solvent accessible, perhaps due to SF3 protein release.

Spliceosome disassembly: a slow step in the *S. pombe* splicing cycle?

Despite sustained efforts by several laboratories, no conditions have yet been found that allow complete spliceosome assembly and catalysis in *S. pombe* extracts. Whereas *S. cerevisiae* and human extracts contain significant amounts of U4/5/6 tri-snRNP ready to engage in new spliceosome assembly, crude *S. pombe* lysates are predominated by ILS complex (Huang et al. 2002). Therefore, spliceosome disassembly seems to be a slow step in the *S. pombe* splicing cycle. ILS complex disassembly requires the abundant DEAH ATPase, Prp43. In *S. cerevisiae*, Prp43 is recruited to post-second step spliceosomes by the NTC-related proteins Ntr1 and Ntr2 (Boon et al. 2006; Tanaka et al. 2007; Tsai et al. 2007). Ntr1 bridges Prp43 to Ntr2, which, in turn, interacts with the U5 snRNP protein and DEIH box ATPase Brr2 (Chen et al. 2013). Although *S. pombe* contains no clear Ntr2 homolog, it is of note that Ntr1, Prp43, and Brr2 were all significantly substoichiometric in our purified complexes (Table 1).

During spliceosome acitivation, Brr2 functions to unwind the U4/U6 helices to enable release of U4 snRNA and formation of the catalytic core in which U6 is extensively base-paired with U2 (Laggerbauer et al. 1998). Brr2′s ATPase activity is also required for spliceosome disassembly (Small et al. 2006; Hahn and Beggs 2010), and its ATPase activity in vitro is stimulated by hybridized U2 and U6 snRNAs (Xu et al. 1996). Thus, Brr2 likely functions to disrupt base-pairing interactions between U2 and U6 at this stage (Small et al. 2006). Because of its stable association with U5 snRNP as well as all major spliceosomal complexes in the human and *S. cerevisiae* systems, Brr2 was not previously thought to be a dynamically interacting splicing factor (Hahn and Beggs 2010). It was, therefore, surprising to us that Brr2 was significantly substoichiometric in (LS preps) or completely absent from (HS and MN preps) our purified endogenous *S. pombe* U2·U5·U6 complexes. Perhaps this unstable interaction of Brr2 with U2·U5·U6 complexes explains both ILS accumulation in logarithmically growing yeast and the inability of *S. pombe* extracts to support splicing in vitro (Huang et al. 2002).

MATERIALS AND METHODS

S. pombe strain construction

All strains derived from yFS 104 (ura4-D18 leu1-32 h+) and yFS 105 (ura4-D18 leu1-32 h−). The *lea1* and *snu114* chromosomal loci fused at their C termini either to a TEV cleavage site followed by protein A or to the calmodulin binding peptide were constructed by PCR-based gene targeting (Bahler et al. 1998) using plasmid pFA6a-CTAP-MX6 (Tasto et al. 2001) as a PCR template. *cdc5*, *spp42* (*prp8*), and *brr2* chromosomal loci were individually 3′ end GFP-tagged by PCR-based gene targeting using pFA6a-GFP (F64L, S65T)-kanMX6 (Bahler et al. 1998) as a PCR template. Strain MJM01 (Cdc5-GFP, Lea1-protein A, Snu114-CBP, ura4-D18 leu1-32 h−) was generated by mating of appropriate parent strains, followed by sporulation and PCR validation of haploid derivatives. Strains MJM02 and MJM03 encoded Spp42-GFP and Brr2-GFP, respectively.

Complex purification

Purification of endogenous U2·U5·U6 complexes was carried out as previously described for TAP-tagged complexes (Puig et al. 2001) with the following modifications. For LS preps, cells from high-density cultures (5×YES; OD600 = 23–25) were washed in Buffer A (10 mM K-HEPES, pH 7.9, 40 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM PMSF, 2.0 mM benzamidine), repelleted, frozen as noodles in liquid nitrogen, and pulverized to a powder in a ball mill (PM 100, Retsch). Frozen cell powder was resuspended in Buffer B (Buffer A with 400 mM KCl) and the supernatant used for TAP purification in 150 mM NaCl as previously reported (Puig et al. 2001). Complexes were recovered from calmodulin resin by applying elution buffer containing 6 mM EGTA and collecting eight 400-µl fractions. HS preps were identical except that frozen cell powder was resuspended in Buffer C (Buffer A with 400 mM KCl and 300 mM NaCl), and the protein A binding and washing buffers contained 700 mM NaCl. MN preps were the same as HS preps except complexes bound to IgG beads were treated with MNase prior to TEV cleavage. For all preps, elution fractions were assessed by GFP fluorescence and Bradford assays. Detailed purification protocols, dynamic light scattering, and negative stain electron microscopy are described in Supplemental Materials and Methods.

Electron microscopy

For EM, purified complex (200 µg/mL) was applied to glow-discharged, carbon-coated grids and negatively stained with 2% uranyl acetate. Micrographs (66,000× magnification) were taken on a Philips CM12 microscope operating at 120 kV.

Mass spectrometry

Fifty micrograms total protein from each prep were digested with trypsin (Promega), and *S. cerevisiae* alcohol dehydrogenase (ADH) tryptic digest (P00330; Waters Inc.) was added to the digest as an internal standard (Silva et al. 2006). Each sample had 500 ng of tryptic peptides injected in triplicate, and nanoflow separation was performed on a nanoACQUITY Ultra Performance Liquid Chromatography (UPLC) system (Waters) equipped with a 75-µm × 150-mm analytical column packed with 1.7-µm bridged-ethyl hybrid (BEH) C_{18}, by application of a linear 60-min gradient from 3% to 90% acetonitrile (ACN) in 0.1% formic acid. A lock mass solution of 200 fmol/µL Glu-fibrinopeptide was delivered via the auxiliary solvent manager at 500 nL/min into the reference sprayer of the NanoLockSpray source. Analysis of nano-UPLC elutes were performed on a Q-Tof Premier mass spectrometer (Waters) operated in the V-mode, and alternating scans were used to detect precursor

ions and then fragment ions. The masses of the precursor ions were detected with a low collision energy (CE) of 5 eV, which was followed by a scan where the CE was ramped from 15 to 40 eV (MS^E scan). MS data were analyzed using ProteinLynxGlobalServer (PLGS) Identity^E Version 2.4 software (Waters) against the protein database Sanger *S. pombe* (www.sanger.ac.uk) which had been combined with an equal-sized database consisting of the randomized sequences for the estimation of false-positive (Li et al. 2009). For database search and protein acceptance, the following criteria were imposed: three fragment ions per peptide and seven fragment ions per protein with at least one unique peptide per protein; within 10 and 20 parts per million (ppm) of the theoretical masses for precursor and fragment ions, respectively; full tryptic specificity with one missed cleavage; variable modifications of oxidation (methionine), deamidation (asparagine, glutamine), as well as static modification of carbamidomethylation (cysteine). The database search returned results for each injection until 4% of the identified proteins were false (random sequence). After all of the data were processed, a further requirement was made that, for acceptance, a protein must be identified in at least two of three replicate injections of the same sample. Absolute quantitation of identified peptides was achieved with the ion chromatographic peak areas of the three most intense tryptic peptides for each protein relative to that of the signal intensity of the three most intense ions for the internal standard (ADH). For each accepted protein, the relative standard deviation of the femtomole amount from the three replicate runs was calculated (Silva et al. 2006).

Western blots

Cell extracts from strain MJM01 (Cdc5-GFP, Lea1-protein A, Snu114-CBP), MJM02 (Brr2-GFP, Lea1-protein A, Snu114-CBP), and MJM03 (Spp42-GFP, Lea1-protein A, Snu114-CBP) were used for Western blots. Total proteins were separated by 4%–12% SDS-PAGE and transferred to PVDF membrane. Western blots were carried out by using anti-GFP (SC-9996, Santa Cruz) with 800RD anti-rabbit IgG, anti-CBP (07-482, Millipore) with 680RD anti-mouse IgG, and detected on an Odyssey scanner (fluorescence).

RNA and Northern blots

Total RNA from purified complexes was resolved on an 8% polyacrylamide-8 M urea gel and the gel stained with SYBR Gold (Invitrogen). For Northern blotting, RNAs were transferred to Biotrans (+)^{TM} NYLON membranes by semidry electroblotting. snRNAs were detected with complementary, ^{32}P-labeled oligonucleotides; the U6 probe spanned its exon-exon junction.

Deep sequencing

RNAseq libraries were generated as previously described (Ingolia et al. 2009) with the following modifications. Total RNA from each prep was subjected to alkaline hydrolysis followed by incubation with T4 polynucleotide kinase (NEB) in the absence of ATP to remove 2′, 3′-cyclic and 3′ phosphate groups. Following denaturing gel electrophoresis, the 30- to 80-nt region was excised, and recovered RNAs were poly(A)-tailed and converted into libraries for deep sequencing on the Illumina GAII platform as previously described (Ingolia et al. 2009).

RNAseq data analysis

Raw 36-bp reads were searched for the poly(A) adapter sequence. If a run of six or more adenosines was detected, the sequence prior to the first A was excised and mapped to the *S. pombe* genome and transcriptome (Sanger Sept. 2011 version) with Bowtie (version 0.11.3). To avoid mapping to spurious locations in the genome, we only retained sequences that met the following criteria: (1) sequences 12- to 15-nt-long had to map perfectly to a single location in the genome or transcriptome; (2) sequences 16- to 19-nt-long had to map perfectly to the genome or transcriptome; (3) sequences ≥20 nt had to map to the genome or transcriptome with no more than one mismatch. If a sequence ≥16 nt mapped to multiple locations, each location was allocated an equal share. The UCSC genome browser (http://genome.ucsc.edu/) was used for making numerous figures. All other analyses were performed with custom-built PERL scripts.

Quantitative analyses employed five classes of CDS mapping reads: (1) exonic = reads mapping entirely inside an exon; (2) intronic = reads mapping entirely inside an intron; (3) 5′SS = reads straddling a 5′ splice site; (4) 3′SS = reads straddling a 3′ splice site; and (5) EEJ = reads straddling an exon–exon junction. For correlation with Pol II occupancy and mRNA abundance, we summed the total reads of all five types for each gene. Peak calling in intergenic regions was performed by extending each intergenic position with ≥11 tags in both directions until a position having no tag was encountered. Overlaps between peaks and annotated genes were performed in a strand-specific manner.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank D. Crawford and M. Lisbin for strain construction, J. Evans for mass spec analyses, M. Zapp and E. Kittler at the UMMS Deep Sequencing Core Laboratory for RNAseq, and A. Hoskins for critical reading of the manuscript. This work was supported by funding from HHMI, NIH RO1-GM53007 (M.J.M.) and NSF DBI-0850008 (Z.W.). M.J.M. is an HHMI Investigator.

Received June 27, 2013; accepted November 12, 2013.

REFERENCES

Abovich N, Liao XC, Rosbash M. 1994. The yeast MUD2 protein: An interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev* **8:** 843–854.

Agafonov DE, Deckert J, Wolf E, Odenwalder P, Bessonov S, Will CL, Urlaub H, Luhrmann R. 2011. Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* **31:** 2667–2682.

Anokhina M, Bessonov S, Miao Z, Westhof E, Hartmuth K, Luhrmann R. 2013. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J* **32:** 2804–2818.

Bahler J, Wu JQ, Longtine MS, Shah NG, McKenzie A, Steever AB, Wach A, Philippsen P, Pringle JR. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14:** 943–951.

Barbosa I, Haque N, Fiorini F, Barrandon C, Tomasetto C, Blanchette M, Le Hir H. 2012. Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* **19:** 983–990.

Behrens SE, Tyc K, Kastner B, Reichelt J, Luhrmann R. 1993. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol* **13:** 307–319.

Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R. 2008. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452:** 846–850.

Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Luhrmann R. 2010. Characterization of purified human B^{act} spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16:** 2384–2403.

Boelens WW, Scherly DD, Beijer RPR, Jansen EJE, Dathan NAN, Mattaj IWI, van Venrooij WJW. 1991. A weak interaction between the U2A′ protein and U2 snRNA helps to stabilize their complex with the U2B″ protein. *Nucleic Acids Res* **19:** 455–460.

Boon KL, Auchynnikava T, Edwalds-Gilbert G, Barrass JD, Droop AP, Dez C, Beggs JD. 2006. Yeast ntr1/spp382 mediates prp43 function in postspliceosomes. *Mol Cell Biol* **26:** 6016–6023.

Brosi R, Gröning K, Behrens S, Luhrmann R, Krämer A. 1993a. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* **262:** 102–105.

Brosi R, Hauri HP, Krämer A. 1993b. Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem* **268:** 17640–17646.

Caspary F. 1998. The yeast U2A′/U2B″ complex is required for pre-spliceosome formation. *EMBO J* **17:** 6348–6358.

Chen YIG, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW. 2007. Proteomic analysis of *in vivo*-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res* **35:** 3928–3944.

Chen HC, Tseng CK, Tsai RT, Chung CS, Cheng SC. 2013. Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. *Mol Cell Biol* **33:** 514–525.

Coltri P, Effenberger K, Chalkley RJ, Burlingame AL, Jurica MS. 2011. Breaking up the C complex spliceosome shows stable association of proteins with the lariat intron intermediate. *PLoS One* **6:** e19061.

Dziembowski A, Ventura A-P, Rutz B, Caspary F, Faux C, Halgand F, Laprévote O, Séraphin B. 2004. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J* **23:** 4847–4856.

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Luhrmann R. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36:** 593–608.

Fourmann JB, Schmitzova J, Christian H, Urlaub H, Ficner R, Boon KL, Fabrizio P, Luhrmann R. 2013. Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev* **27:** 413–428.

Gozani O, Feld R, Reed R. 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* **10:** 233–243.

Grainger RJ, Barrass JD, Jacquier A, Rain JC, Beggs JD. 2009. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15:** 2161–2173.

Hahn D, Beggs JD. 2010. Brr2p RNA helicase with a split personality: Insights into structure and function. *Biochem Soc Trans* **38:** 1105–1109.

Hodges PE, Beggs JD. 1994. RNA splicing: U2 fulfils a commitment. *Curr Biol* **4:** 264–267.

Hoskins AA, Moore MJ. 2012. The spliceosome: A flexible, reversible macromolecular machine. *Trends Biochem Sci* **37:** 179–188.

Huang T, Vilardell J, Query CC. 2002. Pre-spliceosome formation in *S. pombe* requires a stable complex of SF1–$U2AF^{59}$–$U2AF^{23}$. *EMBO J* **21:** 5516–5526.

Ilagan JO, Chalkley RJ, Burlingame AL, Jurica MS. 2013. Rearrangements within human spliceosomes captured after exon ligation. *RNA* **19:** 400–412.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223.

Jurica MS. 2008. Searching for a wrench to throw into the splicing machine. *Nat Chem Biol* **4:** 3–6.

Jurica MS, Moore MJ. 2002. Capturing splicing complexes to study structure and mechanism. *Methods* **28:** 336–345.

Jurica MS, Moore MJ. 2003. Pre-mRNA splicing. *Mol Cell* **12:** 5–14.

Jurica MS, Sousa D, Moore MJ, Grigorieff N. 2004. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat Struct Mol Biol* **11:** 265–269.

Kim D-U, Hayles J, Kim D, Wood V, Park H-O, Won M, Yoo H-S, Duhig T, Nam M, Palmer G, et al. 2010. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* **28:** 617–623.

Krämer A, Grüter P, Gröning K, Kastner B. 1999. Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J Cell Biol* **145:** 1355–1368.

Laggerbauer B, Achsel T, Luhrmann R. 1998. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices *in vitro*. *Proc Natl Acad Sci* **95:** 4188–4192.

Lardelli RM, Thompson JX, Yates JR, Stevens SW. 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16:** 516–528.

Li G-Z, Vissers JPC, Silva JC, Golick D, Gorenstein MV, Geromanos SJ. 2009. Database searching and accounting of multiplexed precursor and production spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9:** 1696–1719.

Li X, Zhang W, Xu T, Ramsey J, Zhang L, Hill R, Hansen KC, Hesselberth JR, Zhao R. 2013. Comprehensive *in vivo* RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly. *Nucleic Acids Res* **41:** 3805–3818.

Lin P-C, Xu R-M. 2012. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J* **31:** 1579–1590.

Makarov EM. 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298:** 2205–2208.

Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, Luhrmann R. 2004. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *Nat Methods* **23:** 2381–2391.

McDonald WH, Ohi R, Smelkova N, Frendewey D, Gould KL. 1999. Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. *Mol Cell Biol* **19:** 5352–5362.

Ohi MD, Link AJ, Ren L, Jennings JL, McDonald WH, Gould KL. 2002. Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol* **22:** 2011–2024.

Ohi MD, Vander Kooi CW, Rosenberg JA, Ren L, Hirsch JP, Chazin WJ, Walz T, Gould KL. 2005. Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. *Mol Cell Biol* **25:** 451–460.

Ohi MD, Ren L, Wall JS, Gould KL, Walz T. 2007. Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proc Natl Acad Sci* **104:** 3195–3200.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Séraphin B. 2001. The tandem affinity purification (TAP) method: A general procedure of protein complex purification. *Methods* **24:** 218–229.

Ren L, McLean JR, Hazbun TR, Fields S, Vander Kooi C, Ohi MD, Gould KL. 2011. Systematic two-hybrid and comparative proteomic

analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. *PLoS One* **6:** e16719.

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, et al. 2011. Comparative functional genomics of the fission yeasts. *Science* **332:** 930–936.

Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW. 1990. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. *Nat Methods* **345:** 502–506.

Schmidt C, Lenz C, Grote M, Luhrmann R, Urlaub H. 2010. Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. *Anal Chem* **82:** 2784–2796.

Silva JC, Gorenstein MV, Li G-Z, Vissers JPC, Geromanos SJ. 2006. Absolute quantification of proteins by LCMS^{E}: A virtue of parallel MS acquisition. *Mol Cell Proteomics* **5:** 144–156.

Small EC, Leggett SR, Winans AA, Staley JP. 2006. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell* **23:** 389–399.

Steckelberg A-L, Boehm V, Gromadzka AM, Gehring NH. 2012. CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep* **2:** 454–461.

Stevens SW. 2000. Analysis of low-abundance ribonucleoprotein particles from yeast by affinity chromatography and mass spectrometry microsequencing. *Methods Enzymol* **318:** 385–398.

Tanaka N, Aronova A, Schwer B. 2007. Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev* **21:** 2312–2325.

Tang J, Abovich N, Rosbash M. 1996. Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B″ protein. *Mol Cell Biol* **16:** 2787–2795.

Tasto JJ, Carnahan RH, McDonald WH, Gould KL. 2001. Vectors and gene targeting modules for tandem affinity purification in *Schizosaccharomyces pombe*. *Yeast* **18:** 657–662.

Tsai RT, Tseng CK, Lee PJ, Chen HC, Fu RH, Chang KJ, Yeh FL, Cheng SC. 2007. Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. *Mol Cell Biol* **27:** 8027–8037.

Wahl MC, Will CL, Luhrmann R. 2009. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Will CL, Luhrmann R. 2011. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3:** a003707.

Williams SG, Hall KB. 2011. Human U2B″ protein binding to snRNA stemloops. *Biophys Chem* **159:** 82–89.

Xu D, Nouraini S, Field D, Tang S-J, Friesen JD. 1996. An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. *Nat Methods* **381:** 709–713.

Yan D, Ares M. 1996. Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. *Mol Cell Biol* **16:** 818–828.

Yoshimoto R, Kataoka N, Okawa K, Ohno M. 2009. Isolation and characterization of post-splicing lariat–intron complexes. *Nucleic Acids Res* **37:** 891–902.

Zhang X, Schwer B. 1997. Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res* **25:** 2146–2152.