Dramatically reduced spliceosome in *Cyanidioschyzon merolae*

Martha R. Stark[a], Elizabeth A. Dunn[b], William S. C. Dunn[a], Cameron J. Grisdale[c], Anthony R. Daniele[a], Matthew R. G. Halstead[a], Naomi M. Fast[c], and Stephen D. Rader[a,b,1]

[a]Department of Chemistry, University of Northern British Columbia, Prince George, BC, V2N 4Z9 Canada; and Departments of [b]Biochemistry and Molecular Biology, and [c]Botany, University of British Columbia, Vancouver, BC, V6T 1Z4 Canada

Edited by Joan A. Steitz, Howard Hughes Medical Institute, Yale University, New Haven, CT, and approved February 9, 2015 (received for review September 1, 2014)

The human spliceosome is a large ribonucleoprotein complex that catalyzes pre-mRNA splicing. It consists of five snRNAs and more than 200 proteins. Because of this complexity, much work has focused on the *Saccharomyces cerevisiae* spliceosome, viewed as a highly simplified system with fewer than half as many splicing factors as humans. Nevertheless, it has been difficult to ascribe a mechanistic function to individual splicing factors or even to discern which are critical for catalyzing the splicing reaction. We have identified and characterized the splicing machinery from the red alga *Cyanidioschyzon merolae*, which has been reported to harbor only 26 intron-containing genes. The U2, U4, U5, and U6 snRNAs contain expected conserved sequences and have the ability to adopt secondary structures and form intermolecular base-pairing interactions, as in other organisms. *C. merolae* has a highly reduced set of 43 identifiable core splicing proteins, compared with ∼90 in budding yeast and ∼140 in humans. Strikingly, we have been unable to find a U1 snRNA candidate or any predicted U1-associated proteins, suggesting that splicing in *C. merolae* may occur without the U1 small nuclear ribonucleoprotein particle. In addition, based on mapping the identified proteins onto the known splicing cycle, we propose that there is far less compositional variability during splicing in *C. merolae* than in other organisms. The observed reduction in splicing factors is consistent with the elimination of spliceosomal components that play a peripheral or modulatory role in splicing, presumably retaining those with a more central role in organization and catalysis.

pre-mRNA splicing | spliceosome core | U1 snRNP | genome reduction | splicing mechanism

Pre-mRNA splicing occurs by two transesterification reactions that are catalyzed by the spliceosome, a large macromolecular assembly of five snRNAs and more than 200 proteins in humans (1). These components are thought to assemble onto each new pre-mRNA transcript in an ordered fashion through the recognition and binding of three highly conserved sequences in the transcript: the 5′ splice site, the branch site, and the 3′ splice site (2, 3). Some of these interactions occur via direct RNA/RNA base pairing between the transcript and snRNAs; for example, both U1 and U6 snRNAs base pair to the 5′ splice site of the pre-mRNA transcript, and, similarly, U2 snRNA base pairs to the branch site (3).

Given the complexity of the human spliceosome, it is of considerable interest to find a more tractable splicing system with fewer components to study the core processes of splicing (assembly, catalysis, and fidelity). The *Saccharomyces cerevisiae* (yeast) spliceosome has been proposed as a simplified model system, because it contains only about 100 proteins (4). Indeed, substantial progress in understanding the spliceosome has been made by studying yeast splicing (3, 5). Nevertheless, the yeast spliceosome is still a highly complex system in which to investigate the role of individual proteins, let alone attempt to develop a completely defined splicing system for more incisive experiments. The publication of the *Cyanidioschyzon merolae* genome sequence revealed that it has only 27 introns (6), indicating that it might be a simpler system in which to investigate splicing.

C. merolae is an acidophilic, unicellular red alga that grows at temperatures of up to 56 °C (6). At 16.5 million base pairs, its genome is similar in size to that of *S. cerevisiae* and contains a comparable number of genes; however only one tenth as many introns were annotated in *C. merolae*: 26 intron-containing genes, 0.5% of the genome (6). The small number of introns in *C. merolae* raises the questions of whether the full complexity of the canonical splicing machinery has been maintained or whether *C. merolae* also harbors a reduced set of splicing factors.

We have undertaken a comprehensive bioinformatic survey of the *C. merolae* splicing machinery, identifying four snRNAs (U2, U4, U5, and U6) and 69 splicing proteins, 43 of which are predicted to be associated with small nuclear ribonucleoprotein (snRNPs) or part of complexes that associate directly with the spliceosome. Surprisingly, we were not able to identify any candidates for the U1 snRNA or U1-associated proteins, leading us to conclude that *C. merolae* does not contain a U1 snRNP. The profile of splicing proteins retained in *C. merolae* provides a means of assessing the contribution of specific proteins to the splicing reaction, at least insofar as the missing ones clearly are not always essential for splicing. The U2 snRNP is the most complex particle, with 10 proteins present in addition to the Sm proteins. Many U5-associated proteins also are retained, whereas the Prp19/CDC5L complex (NTC), step-specific proteins, and tri-snRNP–specific proteins are largely predicted to be absent. Indeed, the *C. merolae* spliceosome is notable for the almost complete absence of proteins that join or leave the spliceosome after B complex formation, suggesting that the initial core contains all the components required to carry out the splicing reaction. The novelty of a splicing system

Significance

The spliceosome—the molecular particle responsible for removing interrupting sequences from eukaryotic messenger RNA—is one of the most complex cellular machines. Consisting of five snRNAs and over 200 proteins in humans, its numerous changes in composition and shape during splicing have made it difficult to study. We have characterized an algal spliceosome that is much smaller, with only 43 identifiable core proteins, the majority of which are essential for viability in other organisms. We propose that this highly reduced spliceosome has retained only the most critical splicing factors. *Cyanidioschyzon merolae* therefore provides a powerful system to examine the spliceosome's catalytic core, enabling future advances in understanding the splicing mechanism and spliceosomal organization that are challenging in more complex systems.

Author contributions: M.R.S., N.M.F., and S.D.R. designed research; M.R.S., E.A.D., W.S.C.D., C.J.G., A.R.D., M.R.G.H., and S.D.R. performed research; C.J.G. contributed new reagents/analytic tools; M.R.S., E.A.D., W.S.C.D., C.J.G., A.R.D., N.M.F., and S.D.R. analyzed data; and M.R.S., E.A.D., C.J.G., N.M.F., and S.D.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: rader@unbc.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416879112/-/DCSupplemental.

reduced to ~40 proteins suggests that *C. merolae* may be a more tractable system for studying the central processes of pre-mRNA splicing.

Results

***C. merolae* Transcripts Are Spliced.** Twenty-seven introns were identified computationally in the *C. merolae* genome (figure S2 in ref. 6). To ascertain whether the intron-containing transcripts are in fact spliced, we performed RT-PCR on DNase-treated total RNA. Each primer pair, except for those for gene CMK245C, resulted in two PCR products, a longer intron-containing amplicon and a shorter amplicon corresponding to spliced mRNA (Fig. 1). The CMK245C transcript appears to be completely spliced. Two junctions (CMQ117C and CMO094C) were confirmed by sequencing reverse transcriptase products, and the remaining introns were observed to splice at their expected junctions based on our sequencing of background mRNAs in the RNA immunoprecipitation sequencing (RIP-seq) Illumina library (see below and *SI Materials and Methods*).

Identification of snRNAs. Given that intron-containing transcripts are spliced in *C. merolae*, we sought to identify components of the splicing apparatus, beginning with the snRNAs. BLAST searches were unsuccessful, presumably because of the small size of snRNAs (generally ~150 nt), so we turned to Infernal (7) to take advantage of conserved secondary structure elements in the search. U2 and U4 snRNA candidates were identified using Infernal single-covariance model searches, and U5 and U6 candidates were identified using multiple covariance models. Unexpectedly, we found no candidates for U1 snRNA despite repeated searches with a variety of parameters.

To test expression of the predicted snRNAs, we performed Northern analysis on total RNA extracted from *C. merolae*. Fig. 2*A* demonstrates that each of the identified snRNAs was expressed and that their sizes corresponded to those predicted bioinformatically, except for U5, which was substantially larger than predicted.

We determined the snRNA ends by a combination of sequencing and alignment with known snRNAs. The ends were similar to our predictions, confirming the unexpectedly large U5, which had 111- and 168-nt extensions on either side of the conserved central core (see Table S1 for snRNA accession numbers and sequences). Sequence alignments between the *C. merolae* candidates and snRNAs from other organisms are shown in Fig. S1, demonstrating sequence conservation in those regions that also are conserved among other organisms. For example, the branch site-binding region of U2 is conserved in *C. merolae* and is complementary to annotated branch sites in the genome (underlined in Fig. S1) (3). Similarly, the U6 ACAGAGA sequence is present—albeit with a change to ACUGAGA—as is U5 loop I (8, 9). In addition, U2, U4, and U5 each have recognizable Sm protein-binding sites, whereas U6 has a binding site for the like Sm (LSm) proteins (underlined in Fig. S1) (3).

Differences between *C. merolae* snRNAs and those of other species include the extended U5, slightly longer U4 and U6, and a shorter U2 snNRA. U4 has an 18-nt insertion, relative to the human sequence, starting at nucleotide 77 (Fig. S1). This insertion falls in the central domain of U4, in the middle of the hypothetical U4/U6 stem III, which is evolutionarily conserved, but for which there is no experimental support (10, 11). In addition, the 5′ stem-loop of U6 is ~10 bp longer than that in other species (Fig. S1). Finally, U2 snRNA lacks ~60 nucleotides from the 3′ end, after the Sm-binding site, a region that generally forms one or more stem-loops in other organisms (12). No *C. merolae* snRNAs have stem-loops predicted downstream of their Sm site.

U4 and U6 snRNAs in other organisms are known to form an extended base-pairing interaction, allowing them to comigrate on a nondenaturing gel. To test whether this interaction also occurs in *C. merolae*, we compared cold phenol-extracted total RNA (Fig. 2*B*, lanes 1 and 3) with the same samples incubated for 5 min at 70 °C to disrupt base pairing (Fig. 2*B*, lanes 2 and 4). Heat treatment resulted in the disappearance of the low-mobility band detected by both U4 (lane 1) and U6 (lane 3) probes and the appearance of higher-mobility bands detected by U4 (lane 2) and U6 (lane 4) probes. Control experiments with *S. cerevisiae* snRNAs (lanes 5–8) demonstrate the analogous interaction. Our results are consistent with the identification of these snRNAs as U4 and U6 and with the conservation of their ability to form extensive base-pairing interactions (Fig. 3 and discussed below).

snRNA Secondary Structure. Conservation of the secondary structure is a useful criterion supporting homology. To determine whether the candidate snRNAs in *C. merolae* were capable of adopting secondary structures similar to those of their *S. cerevisiae* counterparts, we manually folded them into the experimentally determined secondary structures of homologous snRNAs. As shown in Fig. 3, the individual snRNAs are capable of forming structures similar to those formed by their homologs. U2 can adopt the two conformations proposed to be involved in progression through the splicing cycle (Fig. 3*A*) (13). The insertion in U4 has the potential to form an extended stem that interrupts the central domain (Fig. 3*B*) as well as a 5′ kissing loop present in *S. cerevisiae*. The central core of the *C. merolae* U5 sequence also is consistent with proposed secondary structures from other organisms (Fig. 3*C*) (14). The secondary structures of U5's 5′ and 3′ extensions are unique to *C. merolae* and therefore cannot be compared with other U5s, although the 3′ end can adopt a characteristic stem-loop adjacent to the predicted Sm site. The 5′ and 3′ extensions are highly complementary to one another, and we propose that they form long-range stems, based on sfold analysis (15). Finally, U6 snRNA can form the 3′ internal stem loop known to be present in the active spliceosome as well as the conserved 5′ stem-loop (Fig. 3*D*) (16, 17).

Notably, these putative *C. merolae* snRNAs have the potential to form the same intermolecular base-pairing interactions as their counterparts. For example, as in *S. cerevisiae*, U4 and U6 in *C. merolae* have complementary sequences in stems I, II, and III, (Fig. S2*A*) (8). Similarly, U2 and U6 can base pair to each other

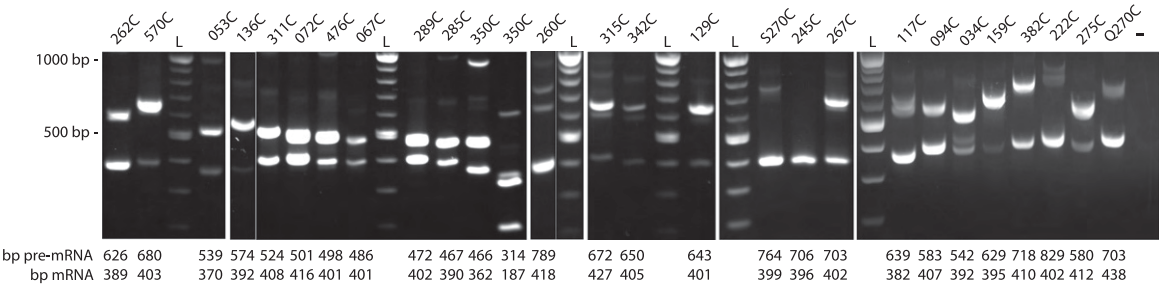

Fig. 1. Amplification of pre-mRNA and mRNA by RT-PCR for all 26 *C. merolae* intron-containing genes. Gene names are listed above each lane. "L" indicates the 100-bp DNA ladder, "–" indicates the negative control with no reverse transcriptase. Expected amplicon sizes for pre-mRNA and mRNA spliced products are given below. The 500- and 1,000-bp ladder bands are indicated on the left.

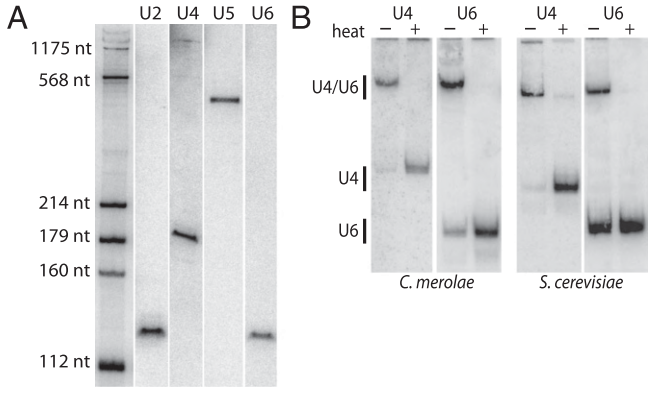

Fig. 2. Analysis of *C. merolae* snRNA expression and base pairing. (*A*) Denaturing Northern analysis of snRNA candidates. *C. merolae* total RNA was probed for each snRNA, as indicated. The leftmost lane contains *S. cerevisiae* total RNA probed for all five snRNAs as size markers, with sizes indicated on the left. (*B*) Nondenaturing Northern analysis of U4/U6 base-pairing status. Total RNA from *C. merolae* and *S. cerevisiae* was probed for U4 or U6, as labeled, with alternate lanes either heat treated (+) or not (–). Band identity is shown at left.

and to the corresponding regions of the substrate transcript in either the three-helix form (Fig. S2*B*) (18) or the four-helix form (Fig. S2*C*) (19). Finally, U5 has the potential to form interactions with *C. merolae* exons similar to those seen in other organisms (Fig. S2*D*) (9). Overall, the *C. merolae* snRNAs appear to be highly conserved in sequence and base-pairing interactions.

snRNA Stability. *C. merolae* is a moderate thermophile, raising the possibility that its snRNAs have evolved greater intrinsic stability to avoid denaturing under these conditions, as has been shown in a number of other cases (20). To test this possibility, we compared the calculated stability of each *C. merolae* snRNA stem with the homologous stems in *S. cerevisiae* and human snRNAs. Each snRNA has at least one stem that is substantially more stable in *C. merolae* than in yeast or humans. U2's stem IIb (Fig. 3*A*) is 30–50 kJ/mol more stable than the orthologous stems in yeast and humans, and U2 stem IIc is 40–50 kJ/mol more stable than the corresponding structures (Table S2). Similarly, the U6 5′ stem-loop has a predicted ΔG of −128 kJ/mol, compared with only −49 and −51 kJ/mol for human and yeast, respectively. U5's stem Ia and variable stem loop (VSL) are both substantially more stable than their orthologous stems, and U4 stem II also is more stable than that in yeast and humans. In contrast, many of the other snRNA stems have comparable, and sometimes less, predicted stability in *C. merolae* than in yeast or humans (Table S2), indicating that greater stability is not a universal feature of *C. merolae* snRNA. Although the intramolecular stems show a pattern of increased stability in certain cases, intermolecular stems between U4/U6, U2/U6, and U2/intron do not differ substantially in stability between *C. merolae* and the other organisms (Table S2).

***C. merolae* Splicing Proteins.** Despite the conservation of the sequence and secondary structure of U2, U4, U5, and U6, we were unable to identify a *C. merolae* U1 snRNA. In other organisms, U1 snRNA is associated with a number of snRNP-specific proteins in addition to the Sm proteins, which are common to all snRNAs except U6. These proteins include U1-70K (found in all organisms tested), U1-A, U1-C, and a variety of other proteins (1, 4).

To determine whether U1-associated proteins are present in *C. merolae*, which in turn would support the existence of the U1 snRNA, we performed BLAST searches (21) using Reciprocal Best Hit methodology (22) to identify homologs of all known human and yeast splicing proteins listed in the online Spliceosome Database and in recent papers (4, 23, 24). The candidate homologs were assigned to particles, splicing steps, or other categories according to their human homolog (1). Based on our criteria (*Materials and Methods*), we were able to identify 10 U2-associated proteins along with the Commitment Complex proteins Msl5, Mud2, and Sub2; four U5-associated proteins; two U4/U6-associated proteins; three from the NTC and related proteins; seven LSm and seven Sm proteins; and a variety of individual proteins found at various steps of splicing (Table 1 and Tables S3–S5). We did not find an LSm8 homolog, suggesting that *C. merolae* has only the LSm1–7 complex (25). Including tangentially associated splicing proteins, such as Dbr1 and Fal1, we predict a total of 69 splicing proteins in *C. merolae,* 43 of which we consider to be core proteins associated directly with snRNPs or the spliceosome.

Although many proteins either were clearly present (often already correctly annotated in the *C. merolae* genome database) or had no hits that were even remotely related, a number were ambiguous, often because of the presence of common motifs [e.g., RNA-recognition motifs (RRMs), tetratricopeptide repeats (TPRs), DEAD-boxes, and others]. We included the marginal candidates Mud2, Prp21, hnRNP H3, and Yju2 on the basis of additional analysis (*SI Materials and Methods* and Table S4). Conversely, we ruled out candidates for Clf1, hnRNP C, hnRNP M, and Prp2 because of the lack of gene-specific features. We did not find any components of the minor spliceosome (26). Given the difficulties inherent in identifying distant homologs bioinformatically, we acknowledge that this list is unlikely to be complete or final. Therefore it will be important to confirm these predictions experimentally.

Notably, we were unable to identify any U1-associated proteins, even when we extended our search to include homologs from additional organisms (*Materials and Methods* and Tables S3 and S5). As a further test for the presence of U1, we asked whether other proteins whose function is related to U1's role in splicing were present in *C. merolae*. One such protein is Prp28, the DExD/H-box ATP hydrolase that catalyzes the dissociation of U1 from the 5′ splice site, allowing U6 to bind in its place (27). In the absence of U1, Prp28 could be rendered redundant. Consistent with this possibility, all Prp28 candidates were more similar to a different ATPase (Table 1 and Table S5). We were able to identify all of the other splicing-associated ATPases except Prp2, demonstrating that ATPases are not generally difficult to identify when present. The simplest explanation for the predicted absence of U1 snRNA, all its associated proteins, and Prp28 is that splicing in *C. merolae* proceeds via a U1-independent mechanism.

RNA Immunoprecipitation. The unexpected absence of U1 snRNA and any U1-associated proteins in our bioinformatic searches suggested that a bioinformatics-independent strategy might be necessary to find U1. We therefore took advantage of the conserved, hypermethylated cap structure—a trimethylguanosine (TMG)—found on snRNAs in other organisms. We immunoprecipitated TMG-containing RNAs from total *C. merolae* RNA using an anti-TMG antibody and sequenced the resulting pool. Northern analysis of the immunoprecipitated RNAs demonstrated 80–90% depletion of all snRNAs in the supernatant compared with total RNA (Fig. 4, lanes 1, 2, 4, and 5). The snRNAs except for U6 were recovered in 65–90% yield in the eluate (Fig. 4, lanes 3 and 6, and table at right), yielding an enrichment of 500- to 1,000-fold relative to unselected RNAs (the total RNA decreased from 1.1 mg in the input to 1.8 μg in the eluate). Because U6 does not have a TMG cap, it is immunoprecipitated only via its association with U4. Silver staining revealed the presence of at least 10 bands smaller than U5 that appeared enriched in the eluate relative to the supernatant, aside from those comigrating with the snRNAs, as well as a large number of larger bands (Fig. 4, lanes 7 and 8). Bands visible in the supernatant are likely to be rRNAs, some fraction of which bound nonspecifically to the resin and eluted with the snRNAs. The entire pool of immunoprecipitated RNA was used as input for a modified Illumina TruSeq library preparation, from which we obtained paired-end reads of 100 base pairs.

After eliminating rRNAs and mRNAs (i.e., transcripts annotated in the *C. merolae* genome as protein-coding), we were left

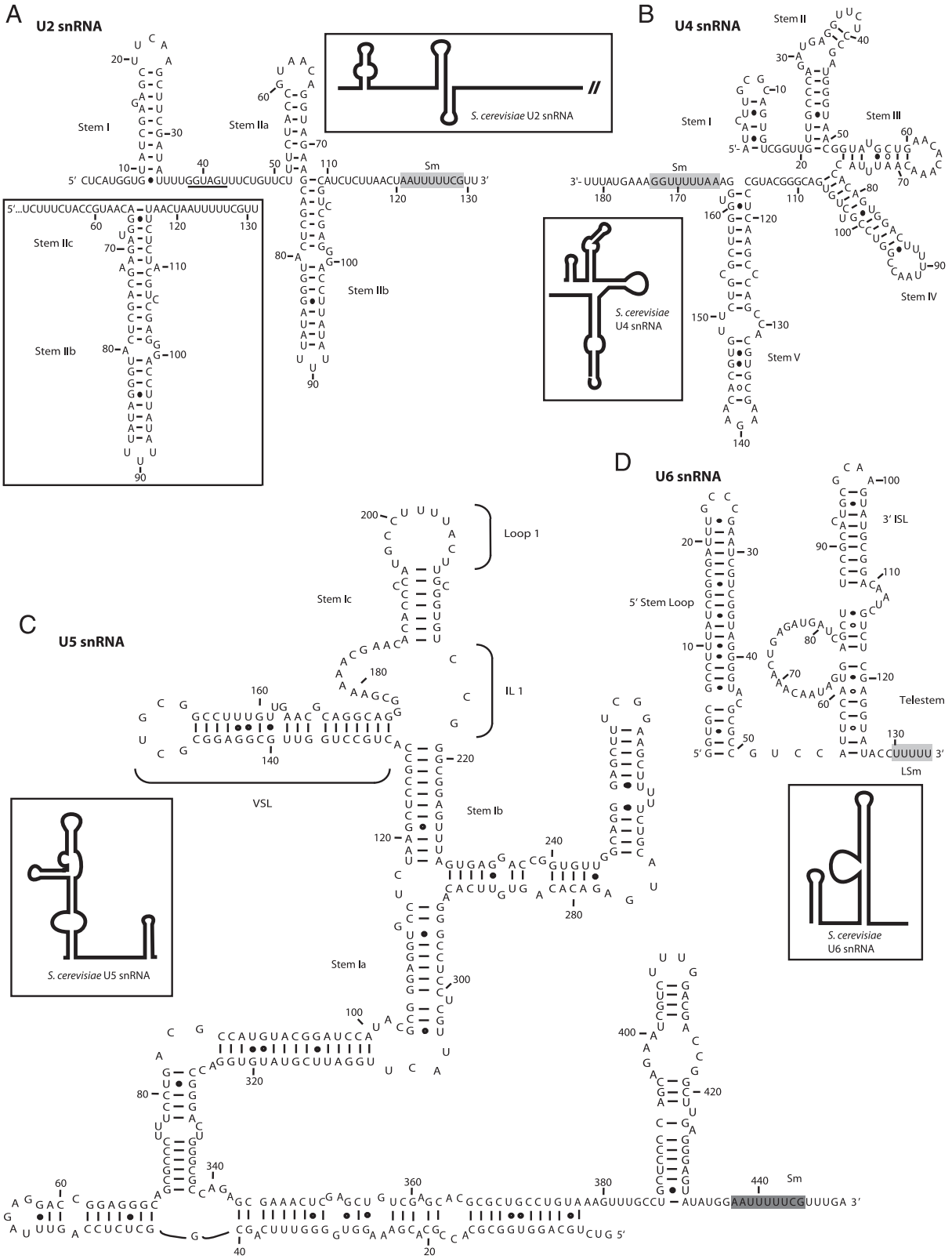

Fig. 3. Predicted secondary structures of *C. merolae* snRNAs (*A–D*) with *S. cerevisiae* structures depicted schematically for comparison. The alternative U2 toggle structure (13), in which stem IIc replaces stem IIa, is shown in the *Inset* in *A*. The conserved core of U5 (*C*) extends from nucleotides 112–282. *S. cerevisiae* secondary structure models are based on refs. 12 and 74 for U2 (*A*), on ref. 75 for U4 (*B*), on ref. 76 for the conserved portion of U5 (*C*), and on ref. 17 for U6 (*D*). The U2 branch site-binding region is underlined, and the Sm- and LSm-binding sites are in gray boxes.

Table 1. *C. merolae* splicing proteins

Particle or step	Name *S. cerevisiae* (*H. sapiens*)	Accession no.	E value*	Identity, %	DEG†	CSRC‡
Sm	SmB/B′	CMK022C	3E-35	32	•	•
	SmD1	CMF084C	3E-15	32	•	•
	SmD2	CMN302C	1E-17	38	•	•
	SmD3	CMM065C	1E-14	37	•	•
	SmE	CMM109C	2E-15	35	•	
		CMH215C				
	SmF	CMQ171C	2E-20	50	•	•
	SmG	CMO342C	2E-11	39	•	•
U2	Prp9 (SF3a3)	CMQ406C	3E-45	22	•	•
	Prp11 (SF3a2)	CMH102C	5E-11	51	•	
		CMN095C				
	Prp21 (SF3a1)	CMJ300C	9E-14	34	•	•
	Hsh155 (SF3b1)	CMB002C	1E-73	25	•	•
	Cus1 (SF3b2)	CMT357C	4E-30	31	•	•
	Rse1 (SF3b3)	CML103C	2E-24	27	•	•
	Hsh49 (SF3b4)	CME063C	1E-13	42	•	
	Rds3 (SF3b6)	CMS014C	3E-37	31	•	•
	Prp5 (hPRP5)	CMR433C	4E-66	32	•	
U2 related	Prp43 (hPRP43)	CMM048C	0	48	•	
	Mud2 (U2AF65)	CMS438C	7E-63	22	•	
U5	Prp8 (220K)	CMH168C	0	34	•	•
	Brr2 (200K)	CML192C	1E-119	35	•	•
	Snu114 (116K)	CMK208C	1E-42	35	•	•
	Dib1 (15K)	CMN033C	2E-78	34	•	
		CMS018C				
U4/U6	Prp3 (90K)	CMT170C	7E-11	21	•	
	Snu13 (15.5K)	CMP335C	5E-52	57	•	
U6§	Lsm1	CMT394C	1E-19	44		
	Lsm2	CMB130C	6E-23	41	•	
	Lsm3	CMT262C	3E-20	44	•	
	Lsm4	CMG061C	1E-27	44	•	
		CMT545C				
	Lsm5	CMP159C	5E-18	38	•	
	Lsm6	CMP138C	1E-28	31	•	
	Lsm7	CMP206C	1E-11	41	•	
Cap binding	Sto1 (CBP80)	CMJ189C	5E-92	18		•
	Cbc2 (CBP20)	CMQ282C	1E-43	53		•
Complex A	Msl5/BBP/SF1 (RBM10)	CMI292C	2E-39	34	•	
	Sub2 (hUAP56)	CME073C	0	59	•	
NTC	Cef1 (CDC5L)	CMR098C	4E-33	34	•	•
	Prp46 (PRL1)	CMR305C	2E-38	28	•	•
	Bud31 (G10)	CMG014C	4E-19	35		•
Complex B	Prp38 (hPRP38)	CMJ144C	1E-57	27	•	
Complex B^{act}	Yju2 (CCDC94)	CMN267C	3E-5	25	•	
Second step	Prp16 (hPRP16)	CMQ385C	0	35	•	•
	Prp22 (hPRP22)	CMG044C	1E-147	46	•	•
EJC/TREX	Fal1 (eIF4A3)	CMK028C	0	78	•	•
	Yra1 (Aly/THOC4)	CMH135C	3E-10	37		
	n/a (THOC2)	CMG046C	1E-41	21		
SR	Rsp31¶	CMO009C	3E-24	34		
	n/a (SRSF2)	CML202C	3E-12	30		
hnRNP	n/a (hnRNP H3)	CMF163C	3E-15	38		
Miscellaneous	Dbr1 (hDBR1)	CMK205C	2E-67	41		
	n/a (SRPK1)	CMK182C	1E-61	41		
	n/a (DDX3X)	CMT173C	1E-177	59		
	Dbp2 (p68/DDX5)	CMR479C	0	54		
	n/a (PABP1)	CMJ286C	3E-76	43		
	n/a (DHX36/RHAU)	CMC171C	3E-42	34		
	n/a (PPP1CA)	CME079C	0	75		
	n/a (RBBP6/PACT)	CMQ079C	1E-29	32		
	n/a (Quaking)	CMA075C	9E-41	55		
	n/a (RACK1)	CMI283C	7E-150	27		
	n/a (TOE1)	CMK240C	7E-20	23		
	Rts2 (HsKin17)	CMG137C	9E-17	33	•	

Table 1. Cont.

Particle or step	Name *S. cerevisiae* (*H. sapiens*)	Accession no.	E value*	Identity, %	DEG†	CSRC‡
	n/a (ERH)	CMR260C	2E-22	42		
	Mtr4 (SKIV2L2)	CMA072C	0	39	•	
	Rvb2 (TIP48)	CMT427C	0	56	•	
	Tef1 (eEF1A)	CMH226C	0	75		
	Rpg1 (eIF3A)	CMH060C	7E-40	28	•	
	Prt1 (EIF3B)	CMK285C	5E-111	31		
	n/a (RPSA)	CMT410C	9E-89	56		
	n/a (TUBA1B)	CMT504C	0	77		
	Tub2 (TUBB4B)	CMN263C	0	72		

*Best E value (among all species, forward and reverse BLASTs) and percent identity for best BLAST.
†Essential for viability in *S. cerevisiae, Schizosaccharomyces pombe, or Mus musculus* according to the Database of Essential Genes (30).
‡C complex proteins detected by mass spectrometry after treatment with 1 M salt (33).
§It is unclear whether the CmLSm proteins associate with CmU6 (*Discussion*).
¶Rsp31 is an *Arabidopsis thaliana* SR protein.

with 58 transcripts that were reproducibly immunoprecipitated and had a sequencing coverage of at least 500 reads (Table S6). U2, U4, and U5 were found among these 58 RNAs, as expected from Northern analysis. U6 was present, but at less than 500-fold coverage, because of its lack of a TMG cap. The remaining RNAs were examined for the expected hallmarks of U1 snRNA: complementarity to known 5′ splice sites in *C. merolae* (GUAAGU) (6), presence of an Sm-binding site, and binding sites for conserved U1 proteins U1-A and U1-70K (28). Eleven RNAs had a potential Sm-binding site and complementarity to the 5′ splice site, and five of these also contained one canonical U1 protein-binding site, but sequence alignments to U1 sequences from the Rfam database (29) failed to reveal significant sequence similarity. Given the enrichment of U2, U4, and U5 in this experiment, it is unlikely that we would have missed U1 if it were present.

Discussion

The pre-mRNA splicing machinery is highly conserved across eukaryotes, with the number of identified splicing proteins ranging from ∼100 in *S. cerevisiae* to well over 200 in humans (4, 23). Our computational and biochemical analysis of the splicing factors present in the acidophilic alga *C. merolae* demonstrates a dramatically smaller set of splicing machinery than has been found in other organisms. With only ∼40 predicted core splicing-associated proteins and four snRNAs, *C. merolae* appears to have been subject to strong selective pressure to reduce its spliceosomal complexity, along with its complement of introns. Strikingly, our multiple, independent search methods have not yielded any evidence for the presence of U1 snRNA or its associated proteins. The best explanation for these observations is that, to our knowledge, *C. merolae* is the first known eukaryote to splice introns in the absence of U1 snRNP.

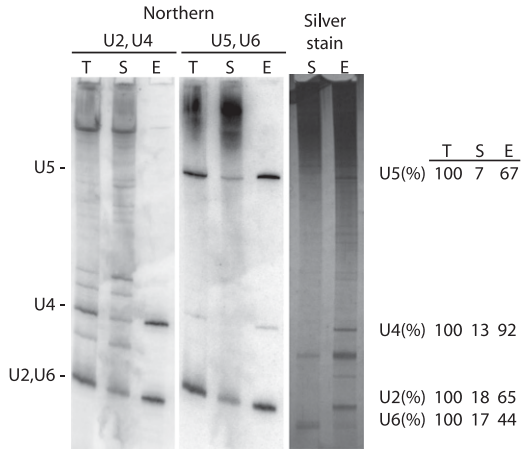

Fig. 4. Immunoprecipitation of RNA using anti-TMG antibodies. Denaturing Northern analysis of total RNA (T) from *C. merolae* (lanes 1 and 4), supernatant (S) from immunoprecipitation (lanes 2 and 5), and immunoprecipitated eluates (E, lanes 3 and 6). Northern lanes were probed for *C. merolae* snRNAs as indicated above. Bands are labeled at left. Band intensities on the Northern blot were measured and normalized to the fraction loaded on the gel, yielding the values shown at right. Lane 7 shows silver stain analysis of immunoprecipitated supernatant, and lane 8 shows immunoprecipitated eluate.

Conservation of Splicing Factors. Although *C. merolae* lives in an extreme environment, the splicing factors it has retained are not dramatically different from those characterized in other organisms. Its snRNAs, for example, are of similar length and sequence, except for U5, and appear to adopt the same conformations and interactions as in yeast and humans. Many of the retained proteins also are highly conserved. This conservation in the face of strong pressure to eliminate splicing components raises the possibility that the eliminated proteins are not involved in key catalytic or assembly events of the splicing reaction. To test this idea, we asked whether *C. merolae*'s splicing components are enriched in proteins that are essential for viability in other organisms, reasoning that core splicing components would be more likely to be essential. Of ∼100 splicing-related proteins in yeast, 65 are essential (i.e., a knockout of the gene is inviable in rich medium at 30 °C). Thirty-nine of the 43 core splicing proteins in *C. merolae* are essential in at least one organism according to the Database of Essential Genes (30), as indicated in Table 1. Of the nonessential proteins, the cap-binding complex (Sto1 and Cbc2) appears to be less important in budding yeast than in humans, where it may be essential (31). Intriguingly, Bud31, although not required for growth at 30 °C, becomes essential in yeast under heat stress (32), suggesting why it may be retained in *C. merolae*, which grows at up to 56 °C.

Bessonov et al. (33) have reported a biochemical strategy to identify the core of the human spliceosome, in which they purified the C complex and treated it with 1 M salt to determine which components are most stably associated. Of the 54 proteins detected in this C complex salt-resistant core (CSRC), 23 were found in *C. merolae* (Table 1), and all had been assigned to the C complex, supporting our classification of these proteins. Indeed, these 23 proteins form nearly the entire complement of the *C. merolae* C complex (Fig. 5), consistent with the view that the most functionally important splicing proteins are overrepresented in this organism. (Of the three exceptions, Hsh49, Prp11, and SmE, the third may not have been detected because of its small size, and Hsh49 was barely detected in any complex; *SI Materials and Methods*). The remaining 31 proteins from the CSRC do not have identifiable *C. merolae* homologs. The 18 core proteins from *C. merolae* that were not found in the CSRC were primarily early

splicing factors, such as Commitment Complex and U4 snRNP proteins, LSms, and components of the A and B complexes (Table 1).

To understand better the conservation of its spliceosomal components, we assigned *C. merolae* proteins to particles, steps, or other categories by placing them in the same groupings that have been determined for human splicing factors (Fig. 5) (1, 23). The most striking difference between the predicted complement of splicing factors in *C. merolae* and other organisms is the near absence of proteins that in yeast and humans enter and leave the spliceosome during the splicing reaction. In other words, the *C. merolae* spliceosome appears to have much less change in composition during splicing. For example, in the transition from the yeast B complex to B^{act}, 12 proteins join the spliceosome, and 35 dissociate (4). In *C. merolae*, however, only Yju2 is predicted to join the spliceosome. The U4 snRNP (U4 snRNA and its associated proteins, seven Sms, Snu13, and Prp3) and four other proteins (Dib1 from the U5 snRNP, Prp38 from the tri-snRNP, and Prp5 and Prp43 from the U2 snRNP) are predicted to dissociate (Fig. 5). In addition, the seven LSm proteins of the U6 snRNP, if present, probably would dissociate upon release of U4 (see discussion below and ref. 34). Similarly, in the transition from B^{act} to the C complex, nine proteins join and two proteins leave in yeast, whereas in *C. merolae* the addition of only two proteins—Fal1 and Prp22—is predicted, and only Yju2 is predicted to leave (Fig. 5). In contrast, in humans there is an exchange of more than 50 proteins at both the B-to-B^{act} and the B^{act}-to-C transitions (1, 35).

NTC Complex. The NTC complex is highly conserved in all organisms studied and plays an important role in the transition from the B to B^{act} complex (24). Surprisingly, of 19 NTC proteins known from other organisms, only three were detected in *C. merolae*: Cef1/CDC5L, Prp46, and Bud31. The first two form part of the salt-resistant core of this complex in humans (36), whereas Bud31 has been classified as part of Sf3b as well as the NTC complex (37). Bud31 has been shown to be required for efficient progression to the first step of splicing (32). The paucity of NTC proteins in *C. merolae* is particularly unexpected, given that this complex has been implicated in a wide range of processes outside of splicing, including transcription and mRNA export (38).

DEAD Box Proteins. Of the eight splicing-associated DExD/H-box ATPases in yeast, two are predicted to be missing in *C. merolae*, Prp2 and Prp28. Prp2 is required in budding yeast to displace the Sf3 complex from the branch site region of the transcript to allow the first chemical reaction of splicing (39). Prp2 is at the center of a network of interactions, directly contacting the Sf3 protein Ysf3 (39) as well as Cwc22 and Spp2. Cwc22 is essential for Prp2 function: In its absence, hydrolysis of ATP by Prp2 simply results in the release of Prp2 without promoting the first chemical step (40). Spp2 interacts directly with Prp2 and is released upon ATP hydrolysis (41). The correlated absence of all three of these proteins supports our conclusion that Prp2 is missing in *C. merolae*. With the absence of these proteins, the only first-step protein remaining in *C. merolae* is Yju2, which interacts directly with U2 snRNA (42) and is required for progression to the first chemical step of splicing (43). However, there is a weak *C. merolae* candidate for Prp2, CME166C. Although it is annotated as a probable Prp43 ortholog, there is a better match for Prp43 (CMM048C), so CME166C remains an unassigned helicase with features of splicing proteins. If future biochemical work demonstrates that CME166C is, in fact, orthologous to Prp2, it would suggest that Prp2's function in dissociating proteins is incidental to another role, presumably involving RNA changes, as appears to be the case for Prp22 and Prp43 (see below).

The absence of Prp28, which is required for the exchange of U6 for U1 at the 5′ splice site (27), is not surprising given the apparent absence of U1 snRNA—another example of the correlated absence of functionally related proteins. The six remaining ATPases (Prp5, Prp16, Prp22, Prp43, Brr2, and Sub2) and one GTPase (Snu114) all have *C. merolae* homologs, consistent with the critical role of NTP-driven conformational changes in splicing as well as their role in increasing splicing fidelity (3).

The *C. merolae* LSm Complex Lacks LSm8. It is not yet clear whether the LSm complex that we have identified in *C. merolae* functions

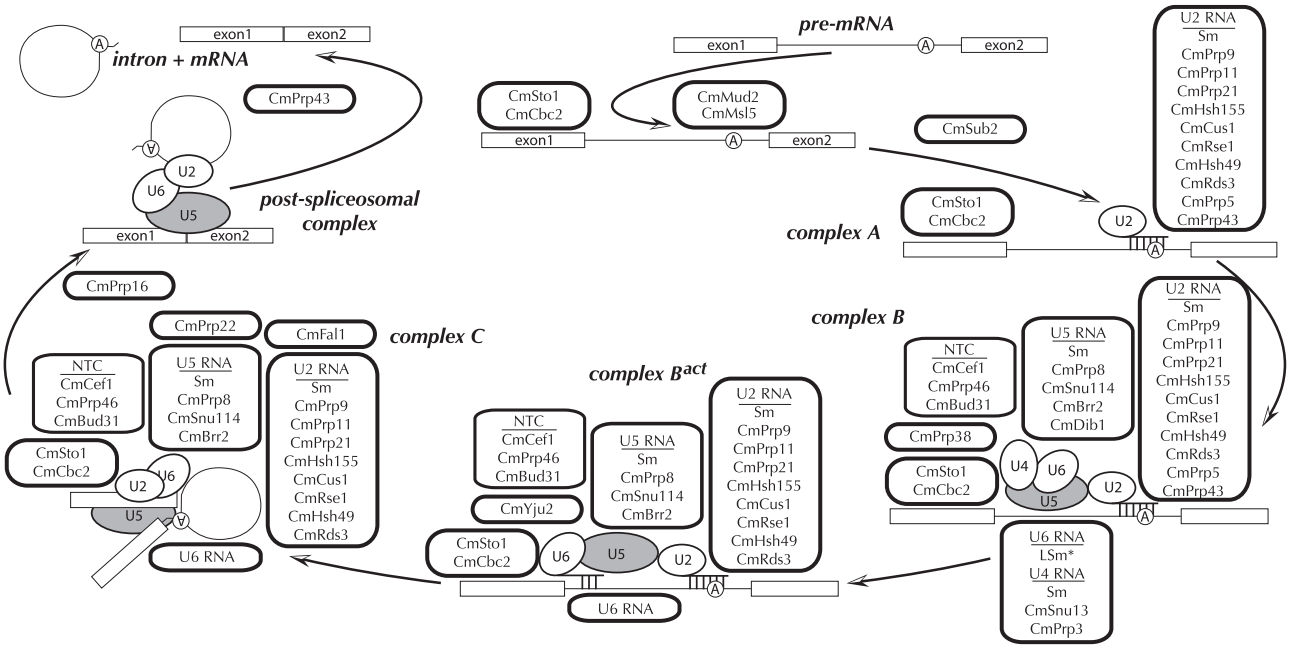

Fig. 5. Summary of *C. merolae* splicing proteins mapped onto the splicing reaction. Splicing steps begin with pre-mRNA and proceed clockwise through complexes A, B, B^{act}, and C and the postspliceosomal complex. The asterisk indicates that the U6-association status of the LSm proteins has not yet been determined. An additional 26 noncore splicing factors are not shown (Table 1).

in splicing. In eukaryotes, two separate LSm complexes exist: the cytoplasmic LSm1–7 complex that binds the polyadenylated 3′ end of mRNAs and is involved in mRNA degradation and the LSm2–8 complex that functions in splicing in the nucleus where it binds the 3′ end of U6 snRNA (44). Although these complexes share six of seven proteins, they differ in the presence of either LSm1 or LSm8. Our bioinformatic search for the LSm proteins suggests that we have identified the LSm1–7 complex and that an eighth LSm homolog is not present. It is possible that this complex functions in both mRNA degradation and splicing; however, it is not immediately clear how a single complex could be involved in such disparate functions. Nevertheless, we feel it is likely that these proteins associate with the U6 snRNA; otherwise, U6 would have no associated proteins, because Prp24 appears to be absent in *C. merolae*.

U1-Independent Splicing in *C. merolae*. The apparent absence of U1 in *C. merolae* raises the question of how the 5′ splice site is recognized. However, there is ample precedent for U1-independent splicing, both in artificial contexts (45) and with naturally occurring transcripts (46, 47). There are several mechanisms by which U1-independent splicing can occur. For example, recent single-molecule experiments demonstrated that the pre-mRNA transcript could be recognized by U2 snRNP before U1 (48). Overexpression of SR proteins could compensate for depletion of U1 in HeLa cells (45); however, we have detected only two candidate SR proteins, CmRsp31 and CmSRSF2, in *C. merolae*. Extending the base-pairing interaction between U6 and the 5′ splice site also increased the splicing efficiency of U1-independent splicing (49). Notably, *C. merolae* U6 has extended complementarity to 5′ splice sites, with six of seven positions capable of forming standard Watson–Crick base pairs, making it comparable in strength to the canonical U1 5′ splice site interactions (Fig. S2*E*). In other organisms U6 forms only three base pairs with the 5′ splice site. Therefore it is plausible that *C. merolae* can dispense with U1 by relying only on a U6 5′ splice site interaction. Other U1-independent interactions with the 5′ splice site have been reported, for instance by Prp8 (50) and U5 snRNA (51). Intriguingly, the 5′ end of U5 snRNA (5′ GUCUGC) is complementary to all annotated *C. merolae* 5′ splice sites (Fig. S2*E*), raising the possibility that initial recognition of introns in *C. merolae* occurs via U5 snRNA.

In addition to recognition of the 5′ splice site, the predicted absence of U1 also raises questions about other roles for U1 that could be missing in *C. merolae*. U1 has been shown to play a role in increasing transcription from intron-containing genes (52) or regulating alternative splicing (53). Furthermore, a number of studies have demonstrated roles for U1 outside of pre-mRNA splicing. These include regulating the use of cleavage and polyadenylation sites (54) and conferring proper polarity to bidirectional transcription start sites (55). Whether any of these processes are affected in *C. merolae* remains to be determined.

Commitment Complex. One of the earliest steps in splicing is the formation of the Commitment Complex, in which the Msl5/Mud2/U2AF1 heterotrimer stabilizes U1 snRNP association with the 5′ splice site by binding to intron features and the U1 snRNP protein Prp40 (3). We found apparent *C. merolae* orthologs of Msl5 and Mud2, but not of U2AF1. Neuvéglise et al. (56) showed that, although U2AF1 does bind the 3′ splice site, which is conserved in *C. merolae* (6), it is required only for transcripts with branch site-to-3′ splice site distances less than 15 nt. This observation may explain its absence in *C. merolae*, in which this distance averages 29 nt. In contrast, human Mud2 (U2AF2), which binds the polypyrimidine tract, was found to be critical for excision of all introns tested (57). Because there is no polypyrimidine tract in *C. merolae* (6), we suggest that CmMud2 could bind the branch site directly, as has been reported in *S. cerevisiae* (58), which also has less prominent polypyrimidine tracts. In the absence of U1, some other mechanism may be required to ensure splicing of bona fide introns (i.e., those introns containing both a 5′ splice site and a branch site), perhaps involving a bridging interaction mediated by the extended U5 snRNA.

Biogenesis and Disassembly Factors. Another category of splicing factors that appear to be nearly absent in *C. merolae* is snRNP biogenesis proteins, the factors involved in initial assembly of snRNPs. The U2 protein Cus2, Aar2 from the cytoplasmic form of U5, Snu40 from the nuclear U5, Sad1 from the tri-snRNP, and the SMN protein required for correct Sm ring assembly in humans were not detected in *C. merolae* (59–61). SMN may be rendered unnecessary by the absence of stem-loops 3′ of the Sm binding site in *C. merolae* snRNAs. *C. merolae* also appears to lack a homolog of Prp24, a protein involved in promoting base-pair formation between U4 and U6 snRNAs in yeast and humans (62). It is possible that the extra domain in U4 compensates in some way for the absence of Prp24, allowing U4 and U6 to form the base-paired di-snRNP without the assistance of a trans-acting factor.

Spliceosome disassembly minimally requires release of the mRNA product and the lariat intron, in which Prp22 and Prp43 have been implicated respectively, but also has been shown to involve protein dissociation and snRNP disassembly. Fourmann et al. (63) recently showed that Prp22 activity is associated with the disappearance of the RES complex (Bud13, Pml1, and Ist3), Cwc21, and Cwc22 from the postcatalytic spliceosome. We have not found any of these proteins in *C. merolae*, suggesting that Prp22-catalyzed protein dissociation is only incidental to its role in mRNA release.

Similarly, the spliceosome disassembly factors Ntr1 and Ntr2 also appear to be missing. This complex, along with the ATPase Prp43, binds to the spliceosome via interactions between Ntr2 and Brr2 and promotes disassembly of the postsplicing particle into its component parts—the U2 snRNP, U5 snRNP, U6 snRNP, and NTC—as it releases the lariat intron (64). The absence of so many assembly and disassembly factors in *C. merolae* raises the possibility that its spliceosome functions as a preassembled holoenzyme that does not proceed through the stepwise formation and dissociation seen in other organisms (2, 65). There is some precedent for the idea that what appears to be dissociation by one assay may represent weakened interaction, but not complete dissociation, under gentler conditions. For example, oligonucleotide-directed selection of U4 snRNA copurifies excised lariat intron, suggesting that the U4 dissociation before the chemical steps of splicing seen on gels may be caused by the stringency of the assay (66).

Noncore Proteins. Twenty proteins classified by Agafonov (1) or Hegele (23) as miscellaneous splicing proteins have apparent homologs in *C. merolae*. A number of these proteins were identified on the basis of physical interactions rather than functional assays, leaving open the possibility that some of them are not actually involved in splicing. For example, Fal1 (eIF4A3) is an ATPase that contacts the mRNA (67) and loads exon junction complex (EJC) components Y14 and Magoh (68). It is recruited to the spliceosome by Cwc22 (69), so, given the predicted absence of the latter, it seems likely that Fal1 has been retained for roles outside of splicing. Others are kinases, phosphatases, or ubiquitin ligases that may be involved in regulating splicing or are simply RNA-binding proteins. Surprisingly, the alternative splicing factor Quaking appears to have an ortholog, even though *C. merolae* has only one known gene (CMR350C) with more than one intron. Sequencing of CMR350C splice junctions revealed singly-spliced transcripts that may be intermediates of the fully spliced form but no exon-skipped variants that would be most indicative of alternative splicing. We have not included these proteins as part of our list of core-splicing machinery, and it is likely that some of these have other cellular roles.

RNA Stability. One prediction concerning RNAs from thermophilic organisms is that they might be more resistant to thermal denaturation, and hence more intrinsically stable, than RNAs from mesophiles. Our results (Table S2) suggest that this

prediction is true to a limited extent in *C. merolae*, in that approximately only one stem in each snRNA has substantially more stability in *C. merolae* than in the orthologous region from yeast or human snRNAs, but the majority of stems have comparable stabilities in these three organisms. One factor that might restrict a stem's ability to evolve greater thermal stability would be a requirement for it to unwind readily as it undergoes conformational rearrangements. Such a requirement would allow us to distinguish "structural" stems, those that are required for snRNP stability, perhaps as protein-binding sites, from "functional" stems, i.e., those that must change conformation or binding interactions during the splicing reaction. Using this framework, we suggest that U2 stems IIb and IIc, U4 stem II, U6 stem I, and U5 stems Ia and VSL are structural, and the remaining stems are predicted to be functional. The prediction of structural stems is consistent with data showing that the U6 5′ stem-loop is strongly protected from hydroxyl radicals in the snRNP particle (17). However, the observation that U2 stems IIb and IIc are more stable suggests that the reality may be more nuanced than this straightforward classification into structural and functional stems allows, because stem II is thought to toggle between two conformations during the splicing reaction.

Our primarily bioinformatic analysis suggests that most peripheral or regulatory splicing proteins have been eliminated in *C. merolae*, leaving a spliceosome enriched in catalytically essential components. We note that these results remain to be confirmed biochemically. This highly simplified splicing machinery should provide a powerful system in which to study key features of the pre-mRNA splicing mechanism. The apparent absence of biogenesis proteins raises the possibility that the entire complement of splicing proteins and particles might be expressed recombinantly, allowing the generation of a completely defined splicing system and providing assembled particles for biophysical studies.

Materials and Methods

For detailed methods, please refer to *SI Materials and Methods*.

***C. merolae* RNA Preparation.** The 10D strain of *C. merolae* (NIES-1332), obtained from the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/), was cultured as described (70). *C. merolae* cultures (50–500 mL) were harvested in log phase at an OD_{750} 1.8–2.0 and were lysed by sonication in the presence of 1% SDS. RNA was acid phenol/chloroform extracted and EtOH precipitated. Where appropriate, total RNA was denatured by heating for 3 min at 65 °C.

***C. merolae* Splicing.** All 26 genes predicted to contain introns were tested for splicing via RT-PCR analysis. Total RNA was treated with DNase, and RT-PCR reactions were carried out using the appropriate primers with reverse transcriptase and Taq DNA polymerase. The negative control reaction contained no reverse transcriptase. Reaction products were run on agarose gels and visualized on a Chemi Imager (Alpha Innotec). The exon junctions from CMQ117C and CMO094C were confirmed by sequencing reverse transcriptase products, and the remaining introns were observed to splice at their expected junctions based on our sequencing of background mRNAs in the RIP-seq Illumina library (*SI Materials and Methods*).

Bioinformatic Analysis. We used Infernal v1.0 (7) to search for snRNAs in the *C. merolae* genome. Sequences were aligned with Clustal Omega v1.1.0 (71), and U4 and U5 secondary structures were modeled using sfold v2.2 (15). RNA stem stabilities were calculated with mfold (72). Splicing protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) website and were used in a Reciprocal Best Hit strategy (22) with an E value threshold up to 100 to maximize our chances of finding homologs. A *C. merolae* gene was considered a clear homolog if, in searches with sequences from two or more species, the top hit in the *C. merolae* database retrieved the initial search protein and the E value was smaller than 10^{-10} (Tables S3 and S4).

snRNA Sequencing. The 5′ and 3′ ends of three of the snRNAs were sequenced by circularizing total RNA and specifically amplifying the snRNA junctions by RT-PCR, followed by cloning into pUC19 and sequencing. The 5′ end of U6 was determined by primer extension sequencing; the 3′ end was predicted by alignment to U6 snRNAs from other organisms.

Denaturing and Native Northern Analysis. To determine candidate snRNA expression and length, *C. merolae* total RNA (and *S. cerevisiae* total RNA as a control) was electrophoresed through a 6% denaturing polyacrylamide gel, transferred to a nylon membrane, and probed for *C. merolae* U2, U4, U5, and U6 snRNAs as well as all five of the *S. cerevisiae* snRNAs. The base-pairing status of U4 and U6 snRNAs extracted from *C. merolae* was analyzed by native Northern and compared with U4/U6 from *S. cerevisiae*.

RNA Immunoprecipitation. Anti-TMG antibodies (200 μg /mL; K121; Santa Cruz Biotechnology) were bound to protein G Sepharose before the addition of total *C. merolae* RNA. RNA was eluted with proteinase K. The supernatant (flow-through) and eluates were electrophoresed on a 6% denaturing polyacrylamide gel. The gel was cut in two, and one half was transferred to a nylon membrane and probed first for U2 and U4 and then for U5 and U6 snRNAs. The other half of the gel was silver stained. The remainder of the eluted RNA was used for sequencing.

Illumina Sequencing. Two independent samples of anti-TMG isolated RNA were used as input for two TruSeq RNA library preparations (Illumina). Reads first were mapped to *C. merolae* rDNA sequences to filter out rDNA contamination and then were mapped to the *C. merolae* genome. A pileup file was created using the SAMtools 0.1.18 mpileup option (73), and a custom Python script was used to identify contiguous stretches of expressed regions with greater than 500× coverage that did not overlap annotated genes. This process resulted in 82 sequences from the first experiment and 87 from the second, with 58 (~70%) common to the two experiments. These sequences were analyzed for features of U1 snRNA.

ACKNOWLEDGMENTS. We thank Camellia Presley, Sepehr Masoumi-Alamouti, and Everett Versteeg for early bioinformatic work on this project; Cassandra Fayowski for assistance with *C. merolae* splicing analysis; Alex Aravind (University of Northern British Columbia, UNBC) for computational resources and assistance; and Andrew MacMillan for encouragement and thoughtful comments on the manuscript. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant 298521 and UNBC Office of Research awards (to S.D.R.) and by an NSERC Postgraduate Scholarship award (to E.A.D.). N.M.F.'s research is supported by an NSERC Discovery Grant and a grant to the Centre for Microbial Diversity and Evolution (CMDE) from the Tula Foundation. C.J.G. was supported by an NSERC doctoral scholarship and a CMDE postdoctoral fellowship from the Tula Foundation.

1. Agafonov DE, et al. (2011) Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* 31(13):2667–2682.
2. Cheng SC, Abelson J (1987) Spliceosome assembly in yeast. *Genes Dev* 1(9):1014–1027.
3. Dunn EA, Rader SD (2014) Pre-mRNA splicing and the spliceosome: Assembly, catalysis, and fidelity. *Fungal RNA Biology* (Springer, Cham, Switzerland).
4. Fabrizio P, et al. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* 36(4):593–608.
5. Hossain MA, Johnson TL (2014) Using yeast genetics to study splicing mechanisms. *Methods Mol Biol* 1126:285–298.
6. Matsuzaki M, et al. (2004) Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature* 428(6983):653–657.
7. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: Inference of RNA alignments. *Bioinformatics* 25(10):1335–1337.
8. Brow DA, Guthrie C (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334(6179):213–218.
9. Newman AJ, Norman C (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* 68(4):743–754.
10. Brow DA, Vidaver RM (1995) An element in human U6 RNA destabilizes the U4/U6 spliceosomal RNA complex. *RNA* 1(2):122–131.
11. Mougin A, Gottschalk A, Fabrizio P, Lührmann R, Branlant C (2002) Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J Mol Biol* 317(5):631–649.
12. Reddy R, Henning D, Epstein P, Busch H (1981) Primary and secondary structure of U2 snRNA. *Nucleic Acids Res* 9(21):5645–5658.
13. Hilliker AK, Mefford MA, Staley JP (2007) U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev* 21(7):821–834.
14. Dix I, Russell CS, O'Keefe RT, Newman AJ, Beggs JD (1998) Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae. *RNA* 4(10):1239–1250.
15. Ding Y, Chan CY, Lawrence CE (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32(web server issue):W135-41.
16. Rhode BM, Hartmuth K, Westhof E, Lührmann R (2006) Proximity of conserved U6 and U2 snRNA elements to the 5′ splice site region in activated spliceosomes. *EMBO J* 25(11):2475–2486.

17. Karaduman R, Fabrizio P, Hartmuth K, Urlaub H, Lührmann R (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J Mol Biol* 356(5): 1248–1262.
18. Madhani HD, Guthrie C (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71(5):803–817.
19. Sun JS, Manley JL (1995) A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev* 9(7):843–854.
20. Fang XW, et al. (2001) The thermodynamic origin of the stability of a thermophilic ribozyme. *Proc Natl Acad Sci USA* 98(8):4355–4360.
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
22. Ward N, Moreno-Hagelsieb G (2014) Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS ONE* 9(7):e101850.
23. Hegele A, et al. (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* 45(4):567–580.
24. Chen H-C, Cheng S-C (2012) Functional roles of protein splicing factors. *Biosci Rep* 32(4):345–359.
25. Tharun S, et al. (2000) Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* 404(6777):515–518.
26. Tarn WY, Steitz JA (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* 84(5):801–811.
27. Staley JP, Guthrie C (1999) An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3(1):55–64.
28. Hamm J, Kazmaier M, Mattaj IW (1987) In vitro assembly of U1 snRNPs. *EMBO J* 6(11): 3479–3485.
29. Burge SW, et al. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* 41(Database issue):D226–D232.
30. Zhang R, Ou H-Y, Zhang C-T (2004) DEG: A database of essential genes. *Nucleic Acids Res* 32(Database issue):D271–D272.
31. Gonatopoulos-Pournatzis T, Cowling VH (2014) Cap-binding complex (CBC). *Biochem J* 457(2):231–242.
32. Saha D, Khandelia P, O'Keefe RT, Vijayraghavan U (2012) Saccharomyces cerevisiae NineTeen complex (NTC)-associated factor Bud31/Ycr063w assembles on precatalytic spliceosomes and improves first and second step pre-mRNA splicing efficiency. *J Biol Chem* 287(8):5390–5399.
33. Bessonov S, Anokhina M, Will CL, Urlaub H, Lührmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452(7189):846–850.
34. Chan S-P, Kao D-I, Tsai W-Y, Cheng S-C (2003) The Prp19p-associated complex in spliceosome activation. *Science* 302(5643):279–282.
35. Bessonov S, et al. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* 16(12):2384–2403.
36. Grote M, et al. (2010) Molecular architecture of the human Prp19/CDC5L complex. *Mol Cell Biol* 30(9):2105–2119.
37. Wang Q, He J, Lynn B, Rymond BC (2005) Interactions of the yeast SF3b splicing factor. *Mol Cell Biol* 25(24):10745–10754.
38. Chanarat S, Sträßer K (2013) Splicing and beyond: The many faces of the Prp19 complex. *Biochim Biophys Acta* 1833(10):2126–2134.
39. Lardelli RM, Thompson JX, Yates JR, 3rd, Stevens SW (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* 16(3):516–528.
40. Yeh T-C, et al. (2011) Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome to escape from a futile pathway. *Mol Cell Biol* 31(1):43–53.
41. Roy J, Kim K, Maddock JR, Anthony JG, Woolford JL, Jr (1995) The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* 1(4):375–390.
42. Chiang T-W, Cheng S-C (2013) A weak spliceosome-binding domain of Yju2 functions in the first step and bypasses Prp16 in the second step of splicing. *Mol Cell Biol* 33(9): 1746–1755.
43. Liu Y-C, Chen H-C, Wu N-Y, Cheng S-C (2007) A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing. *Mol Cell Biol* 27(15):5403–5413.
44. Achsel T, et al. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* 18(20):5789–5802.
45. Crispino JD, Blencowe BJ, Sharp PA (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* 265(5180):1866–1869.
46. Crispino JD, Mermoud JE, Lamond AI, Sharp PA (1996) Cis-acting elements distinct from the 5′ splice site promote U1-independent pre-mRNA splicing. *RNA* 2(7): 664–673.
47. Fukumura K, Taniguchi I, Sakamoto H, Ohno M, Inoue K (2009) U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res* 37(6):1907–1914.
48. Shcherbakova I, et al. (2013) Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. *Cell Reports* 5(1):151–165.
49. Crispino JD, Sharp PA (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev* 9(18):2314–2323.
50. Maroney PA, Romfo CM, Nilsen TW (2000) Functional recognition of 5′ splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell* 6(2):317–328.
51. Wyatt JR, Sontheimer EJ, Steitz JA (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes Dev* 6(12B):2542–2553.
52. Kwek KY, et al. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat Struct Biol* 9(11):800–805.
53. Fukumura K, Inoue K (2009) Role and mechanism of U1-independent pre-mRNA splicing in the regulation of alternative splicing. *RNA Biol* 6(4):395–398.
54. Kaida D, et al. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468(7324):664–668.
55. Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499(7458):360–363.
56. Neuvéglise C, Marck C, Gaillardin C (2011) The intronome of budding yeasts. *C R Biol* 334(8-9):662–670.
57. Guth S, Martínez C, Gaur RK, Valcárcel J (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol Cell Biol* 19(12):8263–8271.
58. 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(7):843–854.
59. Yan D, et al. (1998) CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. *Mol Cell Biol* 18(9): 5000–5009.
60. Weber G, et al. (2013) Structural basis for dual roles of Aar2p in U5 snRNP assembly. *Genes Dev* 27(5):525–540.
61. Battle DJ, et al. (2006) The Gemin5 protein of the SMN complex identifies snRNAs. *Mol Cell* 23(2):273–279.
62. Raghunathan PL, Guthrie C (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* 279(5352):857–860.
63. Fourmann J-B, et al. (2013) Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev* 27(4):413–428.
64. Tsai R-T, et al. (2005) Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev* 19(24):2991–3003.
65. Stevens SW, et al. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* 9(1):31–44.
66. Blencowe BJ, Sproat BS, Ryder U, Barabino S, Lamond AI (1989) Antisense probing of the human U4/U6 snRNP with biotinylated 2′-OMe RNA oligonucleotides. *Cell* 59(3): 531–539.
67. Shibuya T, Tange TØ, Sonenberg N, Moore MJ (2004) eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* 11(4):346–351.
68. Zhang Z, Krainer AR (2007) Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. *Proc Natl Acad Sci USA* 104(28):11574–11579.
69. Barbosa I, et al. (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* 19(10):983–990.
70. Minoda A, Sakagami R, Yagisawa F, Kuroiwa T, Tanaka K (2004) Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, Cyanidioschyzon merolae 10D. *Plant Cell Physiol* 45(6): 667–671.
71. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
72. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415.
73. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
74. Brennwald P, Porter G, Wise JA (1988) U2 small nuclear RNA is remarkably conserved between Schizosaccharomyces pombe and mammals. *Mol Cell Biol* 8(12):5575–5580.
75. Myslinski E, Branlant C (1991) A phylogenetic study of U4 snRNA reveals the existence of an evolutionarily conserved secondary structure corresponding to 'free' U4 snRNA. *Biochimie* 73(1):17–28.
76. Frank DN, Roiha H, Guthrie C (1994) Architecture of the U5 small nuclear RNA. *Mol Cell Biol* 14(3):2180–2190.

Supporting Information

Stark et al. 10.1073/pnas.1416879112

SI Materials and Methods

***C. merolae* Culturing.** The 10D strain of *C. merolae* (NIES-1332), obtained from the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/), was cultured in modified Allen's autotrophic medium with double the amount of trace elements (MA2) (1). Cultures were grown in graduated glass cylinders at 45 °C in a CO_2 incubator, with 2% CO_2 bubbled directly into the cylinders via an aquarium air pump. Cells were grown under continuous light using a Verilux Happy Light, which provided 90 μmol photons·m^{-2}·s^{-1}. Doubling time was 12–15 h.

Total RNA Preparation. *C. merolae* cultures (50–500 mL) were harvested in log phase at an OD_{750} 1.8–2.0. Cell cultures were centrifuged for 5 min at 3,000 × *g* in 50-mL conical tubes in a Beckman Coulter Allegra X-12R centrifuge with an SX4750 rotor. Cells were resuspended in 5 mL cold phenol RNA lysis buffer [0.5 M NaCl, 0.2 M Tris·HCl (pH 7.5), 0.01 M EDTA, 1% SDS] per 50 mL of culture and were sonicated two times for 15 s to shear the genomic DNA. Total RNA was acid phenol/chloroform (pH 4.5) (Ambion) extracted two times in 15- or 50-mL conical tubes (centrifuged at 3,370 × *g* for 30 min for the first extraction and for 15 min for the second), followed by two chloroform extractions (5 min at 3,370 × *g*). RNA was EtOH precipitated and then stored in 70% EtOH at −20 °C until use, at which point it was resuspended in dH_2O. When appropriate, total RNA was denatured by heating 3 min at 65 °C.

***C. merolae* Splicing.** All 26 genes predicted to contain introns (table S3 of ref. 2) were screened for splicing through RT-PCR analysis of their expressed RNA. Oligonucleotide pairs (Invitrogen) were designed upstream and downstream of the proposed introns:

Oligo	Gene	Sequence
oSDR812	CMT570C	CGCGTGGCAACCGTGCCGCATCAAG
oSDR813	CMT570C	GCACGGGACGAGCCCAATCCTG
oSDR814	CMC053C	CGAGGAAAGTCCAGACAAGAGTTGG
oSDR815	CMC053C	CTGATTGCATGAGAGAATCACGCGG
oSDR816	CMD067C	GCACACGCTCGTCATAGTCCAGACC
oSDR817	CMD067C	GCGGTTCCCGTTATCGTCTGGTGGG
oSDR818	CME034C	GAAGCATTCAACCTGTTCGATCGCG
oSDR819	CME034C	CGAGCAAAATCTTCCCATTGGACG
oSDR820	CMF072C	CATCCGATGCCACAGGGCGCGAGAC
oSDR821	CMF072C	GGCAAGGACTACGCACGCCCGAAC
oSDR822	CMF136C	GCACGCGTTTCCATTCAGAGACTCC
oSDR823	CMF136C	CATGAGAGAGAGTGAACGGCTGCC
oSDR824	CMJ129C	CCTCCTCGACGCCGTCGCCATGATG
oSDR825	CMJ129C	GGTACTGCAACCGCCTCGCGCAAC
oSDR826	CMQ270C	GCAGCAGGTGACGGAACTACGACGG
oSDR827	CMQ270C	GCTGCTCACCTTCTTTTCCGCTG
oSDR828	CMR350C	GAGCACAACGGGGTCAGCTGCAC
oSDR829	CMR350C	CTCACTGGTTGCTTCCTGAACGAGC
oSDR830	CMR350C	GCCGATCGCCAATATTAGCCGATG
oSDR831	CMR350C	CGTCTTTGGCAGAGTCATCGATCCC
oSDR832	CMS262C	GCGATTGCTGAAGCCGCTGAGG
oSDR833	CMS262C	GGCACCAAGTTTCGCTGTACGCTTG
oSDR834	CMS315C	GTGCCGTACTCGTGAACACCGACG
oSDR835	CMS315C	GCAACGCCGAGGTCGCTCGTAC
oSDR836	CMT275C	CGGATCTTGCCGGAAAGGTTCGAG
oSDR837	CMT275C	CCACCTCAAATGGTATCTTGCCCG

Cont.

Oligo	Gene	Sequence
oSDR838	CMT476C	GTGGACACAGTGGTGCCTCCAGG
oSDR839	CMT476C	GTCTGAGCTGCAAGGCCCGTTGC
oSDR846	CMK260C	CGCGCTTCACCCCCAGTGTAATGCG
oSDR847	CMK260C	CGGCAGTGCGTCCGCAGCTCATG
oSDR848	CMN285C	GACAACGCCGTGTCCATCCACGACG
oSDR849	CMN285C	GAGCGTCGCTTGGCGCTTATCGC
oSDR850	CMO094C	CTCCCTGGAGCTTCCGCAACCAAAG
oSDR851	CMO094C	CAAGGAGCAGCGCTCTCGGGAG
oSDR852	CMO267C	GAGCACCTCCGAGCAGCCAAGGTTG
oSDR853	CMO267C	CCCGCAACATATACCGGCTGACGC
oSDR854	CMQ117C	GCCGAAGAACAAGGGCAAAGGAGG
oSDR855	CMQ117C	CTCGACTTCAGCTCCATCCCCAG
oSDR856	CMQ382C	CGGCCGAGTCGCGAGTAACCATTC
oSDR857	CMQ382C	GAGGTCAAAGAGTCGCACCGCGCC
oSDR858	CMR289C	GGCAACTGCCGCATGTGCTTGGTTG
oSDR859	CMR289C	GATCTCGGTCGATTCGCCACGTCC
oSDR860	CMS311C	CGTGTACGGCTTCGCACTTGCG
oSDR861	CMS311C	CTCGCGACCCATCATCTTACGCAAC
oSDR862	CMS342C	GACAGCACGAGCCGCTGGGCAG
oSDR863	CMS342C	GAGAATCTGACGTACGGTCTGGCG
oSDR864	CMT222C	GCTCCTTAGCGAGATCGAGAAGCTC
oSDR865	CMT222C	CTCGGCGCGATTCGCTTCCAC
oSDR875	CMS270C	CGGTGCGAGCAAACTCGCATGC
oSDR876	CMS270C	GCCCGCGAGGCATGGACAAG
oSDR877	CMO267C	CGAGCTGTCTCTCCACACTGGTCAG
oSDR878	CMO267C	CGTATCTCATACCCGGTGACAGCG
oSDR879	CMK245C	CGCACCATCCTGGTACAGGAGTCG
oSDR880	CMK245C	CCTCGAAGTGGACCTGCGCAAC

Total RNA was treated with 2 U of Turbo DNase (Ambion), and RT-PCR reactions were carried out using the appropriate primers with AffinityScript reverse transcriptase (Agilent) and Taq DNA polymerase (New England Biolabs), following the manufacturers' instructions. The negative control reaction contained no reverse transcriptase. The reaction products were run on 1% agarose gels containing ethidium bromide and were visualized on a Chemi Imager (Alpha Innotec). The resulting images were uniformly adjusted for contrast in a linear fashion. Two of the spliced products from the RT-PCR reactions (CMO094C and CMQ117C) were gel purified using an Omega gel extraction kit and were sequenced to determine the splice junctions. In addition, splicing at canonical junctions was observed for all introns (see *Illumina sequencing*, below).

Bioinformatic Analysis. The 99.98% complete *C. merolae* genome was downloaded from the *C. merolae* genome project website (merolae.biol.s.u-tokyo.ac.jp/) and formed our search area. We downloaded seed training datasets for the five snRNAs, U1, U2, U4, U5, and U6, from the Rfam database version 9.1 (3). The Infernal program version 1.0 (4) was used with each seed dataset to search for the corresponding snRNA in the *C. merolae* genome on a 1-GHz Linux machine. The program was run initially using the default settings of a single covariance model based on the entire seed dataset. The program then was instructed to divide the seed dataset into clusters of 60% or greater sequence identity and was rerun using multiple covariance models wherein each model was constructed from a single cluster to increase search sensitivity.

The set of sequences returned by Infernal for each snRNA was refined by considering only sequences with an E-value less than 0.5 and a Bit Score greater than 15. The set of possible candidate sequences was reduced further by excluding sequences that were at odds with regions of high or invariant conservation among well-characterized snRNAs. The candidate *C. merolae* snRNAs were chosen from their remaining respective sequence sets through individual examination of each candidate's ability to form a conserved secondary structure and the extensive intermolecular base-pairing interactions known to exist between snRNAs. No U1 snRNA candidates that met these criteria were found. SnRNA secondary structures were modeled manually when the sequences were sufficiently similar to yeast and human homologs. The 18-nt insertion in *C. merolae* U4 snRNA, relative to the human sequence, was modeled into a stem-loop based on structure output from Sfold (sfold.wadsworth.org) (5–7). The secondary structure of the 5′ and 3′ extensions of U5 snRNA was modeled using sfold, with the constraint that the Sm-binding site be single stranded.

Sequence alignments were prepared using Clustal Omega v1.1.0 (8) as implemented on the Mobyle@Pasteur website (mobyle.pasteur.fr/cgi-bin/portal.py) with the default settings. The Sm and LSm sites subsequently were aligned manually.

To assess the stability of the base-paired stems in *C. merolae* snRNAs, we used mfold (9) to calculate the stability of each stem as a hairpin. For stems that do not terminate in a loop, we added a tetraloop to join the two sides of the stem. Relative free energies for each hairpin are reported in kilojoules per mole.

A comprehensive list of splicing proteins was compiled by combining information from the Spliceosome Database (www.spliceosomedb.ucsc.edu/), published surveys (10–12), and the *Arabidopsis* database ASRG (13). Splicing protein sequences were retrieved from the NCBI website (www.ncbi.nlm.nih.gov/) and were used for protein BLAST searches (14) using the BLOSUM62 matrix and an initial expect threshold of 1E-04 with the program located at merolae.biol.s.u-tokyo.ac.jp. For searches that retrieved no hits, the expect threshold was increased until a hit was obtained, as reported in Tables S3 and S4. Searches were conducted with at least two homologs of a given protein, usually from *S. cerevisiae* and humans but also from *Arabidopsis* or *S. pombe* when an *S. cerevisiae* homolog did not exist. U1 proteins U1-A, U1-C, and U1-70k were sought with up to 12 homologs each (from *S. cerevisiae, Homo sapiens, Trypanosoma brucei, Dictyostelium discoideum, Chlamydomonas reinhardtii, Xenopus laevis, Bos taurus, Drosophila melanogaster, Arabidopsis thaliana, M. musculus, Candida orthopsilosis, Danio rerio, Caenorhabditis elegans*, and *Candida albicans*). The remaining budding yeast and human U1 proteins were sought with one to four homologs, depending on how many experimentally confirmed homologs we could locate in the literature. Protein sequences from candidate *C. merolae* orthologs were searched back against the organism from which the original sequence was taken (the reciprocal search), using default parameters of the NCBI BLAST tool. A *C. merolae* gene was considered a clear homolog if, in both searches, the top hit in the *C. merolae* database retrieved the initial search protein [i.e., they were reciprocal best hits (RBHs)] (15), and the E value was smaller than 10^{-10}. Additionally, in cases in which the top hit did not correspond to the query protein, all candidate *C. merolae* proteins with an expect threshold of less than 1E-02 or that appeared in the top 10 hits, whichever was less restrictive, were also searched against the original organism. The splicing proteins found in this way (non-RBH proteins) are reported in Table S4. For ambiguous cases, we analyzed domain structure using CDART from NCBI and related tools (16) and sequence alignments using NCBI's COBALT (17), which allowed us to distinguish false positives (containing shared domains) from orthologs (having gene-specific regions in common). Table 1 contains only the proteins found in *C. merolae*. Table S5 also lists those that were not. Tables S3 and S4 contain the entire bioinformatic analysis.

C. merolae proteins were annotated as components of the CSRC on the basis of their ortholog having been detected by Bessonov et al. (18). Because of the ambiguity in the enrichment of some proteins in the CSRC relative to dissociated fractions (e.g., SmE, which is unlikely to be missing if the other Sm proteins are present), we annotated any protein as belonging to the CSRC if its peptides were detected in the CSRC fractions, a more inclusive classification than that used by Bessonov et al.

snRNA Sequencing. To determine the ends of the snRNAs precisely, we circularized total RNA (19), specifically amplified the junction-containing regions of the snRNAs by RT-PCR, and cloned the products into vectors for sequencing. Before circularization, RNA was decapped. Five micrograms of total RNA was treated with 20 U tobacco acid pyrophosphatase (Epicentre) in 1× TAP buffer (Epicentre) in a total volume of 50 μL for 1 h at 37 °C. RNA was EtOH precipitated, resuspended in 10 μL of 10 mM Tris·HCl (pH 7.5), and heated at 70 °C for 5 min before being added directly to the chilled ligation reaction mix [1× RNA ligase buffer (Ambion), 10% DMSO, 25 U Optizyme RNase inhibitor (Fisher), 80 U T4 RNA ligase (Ambion) in a total volume of 100 μL]. Ligation was carried out at 4 °C for 26 h followed by EtOH precipitation. One microgram of circularized RNA was reverse transcribed using SuperScript III (Invitrogen) at 55 °C with gene-specific primers for each snRNA, according to the manufacturer's instructions. One tenth of each RT reaction was used in a standard PCR. PCR products were gel purified, and nested PCR was performed to ensure specificity of products. Nested PCR products were precipitated, digested with restriction enzymes, cloned into pUC19, and sequenced.

We were unsuccessful in circularizing U6 snRNA, probably because of the presence of a cyclic phosphate at its 3′ end (20). Therefore the 5′ end of U6 was determined by RNA sequencing. Forty micrograms of total RNA was hybridized to 400 fmol primer [5′ end-labeled using γ–^{32}P ATP (3,000 Ci/mmol) (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs)] according to the manufacturer's instructions in 100 mM NaCl, 10 mM Tris·HCl (pH 7.5), and 10 mM EDTA (pH 8) by heating for 5 min at 65 °C and then cooling over 15 min to 50 °C. The reaction was divided into four tubes, one for each ddNTP, and the primer was extended with 50 U AffinityScript reverse transcriptase (Agilent) in the presence of the appropriate nucleotide mix (for example, 100 μM each dATP, dTTP, dCTP + 25 μM dGTP + 200 μM ddGTP) in 50 mM Tris·HCl (pH 8.3) and 3 mM MgCl for 45 min at 50 °C. Reaction products were EtOH precipitated and electrophoresed through an 8% polyacrylamide (19:1) 7-M urea gel in 1× TBE (90 mM Tris-borate, 2 mM EDTA) at 20 mA. The gel was transferred to Whatman paper, exposed to a phosphor imager screen overnight at −80 °C, and then scanned on a Cyclone Phosphor Imager (Packard Instruments). The 3′ end of U6 was determined by sequence alignment of the LSm protein-binding site (21).

The following oligonucleotides were used for RT-PCR and primer extension:

U2: CTGCTTCTACCTGTTACGGTAGAAAG (oSDR1017) RT+PCR

GCAGATGAATTCGGGATATTTTATATTCCAGGGAGCCTG (oSDR1019) PCR+nested

GCAGATGGATCCAGAAACTACCAAAATATCGAAGCTTGAAG (oSDR1018) nested

U4: AAATTGTTTGTGTTCAGCATACCGTT (oSDR597) RT+PCR

GCAGATGAATTCCGGTCCGTCTGTGAC (oSDR1016) PCR+nested

GCAGATGGATCCCATCGGAGAACCTCATCTG (oSDR1015) nested

U5: GGACACCGCAAGTAAAAGGCATGG (oSDR768) RT+PCR

GCAGATGAATTCCGGTGTTGGCAGGG (oSDR1021) PCR+nested

GCAGATGGATCCCTGCCTGCGTTCAAC (oSDR1020) nested

U6: CTCAGTTTGTTATCCATGGAATGGACGGC (oSDR748) PE

Denaturing and Native Northern Analysis. To determine candidate expression and the overall length of each snRNA, 20 μg per lane of denatured *C. merolae* total RNA was electrophoresed through a 6% polyacrylamide (19:1)7-M urea gel in 1× TBE at 400 V and then was transferred to a Hybond N+ membrane (GE Healthcare) using a Panther Semidry Electroblotter HEP-3 (Owl) for 20 min at 2.5 mA/cm^2. Twenty micrograms of *S. cerevisiae* total RNA was run on the same gel. RNAs were crosslinked to the membrane in a UV Stratalinker 1800 (Stratagene) with 120,000 J of UV radiation. The membrane was cut into five strips before being probed for *C. merolae* U2, U4, U5, and U6 snRNAs and for all five *S. cerevisiae* snRNAs, which served as size markers for the algal snRNAs. Oligonucleotide probes were 5′ end-labeled as above. Northern blots were prehybridized in Rapid-Hyb Buffer (GE Healthcare) at 45–50 °C for 30 min, followed by the addition of the probe and hybridization for 2 h. Blots were washed, imaged on a phosphor imager screen overnight, and visualized with a Cyclone Phosphor Imager and OptiQuant software. The resulting images were uniformly adjusted for contrast.

The base-pairing status of U4 and U6 snRNAs extracted from *C. merolae* was analyzed by nondenaturing Northern and compared with U4/U6 from *S. cerevisiae*. Ten micrograms per lane of total RNA from *C. merolae* or *S. cerevisiae* was incubated on ice for 10 min or denatured at 70 °C for 5 min followed by 5 min on ice. RNA was electrophoresed through a 9% nondenaturing polyacrylamide (29:1) gel in 1× TBE at 300 V at 4 °C and then was transferred to nylon membrane and hybridized, as above, except hybridization was at 37 °C.

The following oligonucleotides were used to probe blots for *C. merolae* snRNAs:

U2: CAGAAACTACCAAAATATCGAAGCTTGAAGCTC (oSDR745)

U4: ATAGAACGTGAAATACTTTCCAAAAATTTCC (oSDR596)+(oSDR597) above

U5: (oSDR768) above

U6: AAAAAGGTATACCTCGAGACGATTGTC (oSDR598)+(oSDR748) above

The following oligonucleotides were used to probe blots for *S. cerevisiae* snRNAs:

U1: CAGTAGGACTTCTTGATC (U1-19G)

U2: CAGATACTACACTTG (U2-L15)

GCGTTGGACATAAACGGCTCGG (oSDR527)

U4: AGGTATTCCAAAAATTCCCTAC (U4-14b)

U5S and U5L: AAGTTCCAAAAAATATGGCAAGC (U5-7SmWTNR)

U6: TTGTTTCAAATTGACC (oSDR467)

RNA Immunoprecipitation. RNAs containing a TMG cap and RNAs base-paired to these capped RNAs were immunoprecipitated with anti-TMG antibodies. Two hundred fifty microliters of anti-TMG antibodies (200 μg/mL; K121; Santa Cruz Biotechnology) were bound to 250 μL protein G Sepharose (GE Life Sciences) in 1 mL IPP$_{150}$ buffer [20 mM Hepes (pH 7.9), 1.5 mM $MgCl_2$, 0.05% Nonidet P-40, 150 mM NaCl] plus 10 μg/mL *Escherichia coli* tRNA for 1 h at 4 °C in a head-over-tail mixer. Anti–TMG-bound Sepharose was washed with 3× 1.75 mL IPP$_{150}$ buffer (centrifuged at 10,000 × *g* for 10 s in an Eppendorf centrifuge) before 1 mg total *C. merolae* RNA (in 150 μL dH_2O) and 350 μL IPP$_{150}$ buffer were added. RNA was bound to the Sepharose for 2 h at 4 °C in a head-over-tail mixer, divided into two tubes, and washed with 3× 2 mL IPP$_{150}$ buffer. Then the bound RNA was eluted with 2× 400 μL elution buffer [10 mM Tris·HCl (pH 8.0), 10 mM EDTA, 0.5% SDS, 10 μg/mL proteinase K] for 15 min at 55 °C. The supernatant and eluates were acid phenol:chloroform extracted and EtOH precipitated overnight at −20 °C with 1/10 volume 2 M NaCl and 30 μg glycogen. The precipitated RNA was resuspended in dH_2O, and the concentration was determined using a Qubit fluorometer (Invitrogen). For the Northern analysis 1/50th of the total RNA, 1/30th of the supernatant, and 1/30th of the eluate were electrophoresed on a 6% denaturing gel. For silver staining 1/3,000th of the supernatant and 1/30th of the eluate were run on the same gel. The gel was cut in two, and one half was transferred to a nylon membrane and probed with γ−^{32}P ATP-labeled oligos specific for U2 and U4 snRNAs, as described above. After visualization, the blot was stripped and then reprobed with oligos specific for U5 and U6. Band intensities from the Northern blot were measured (OptiQuant software) and normalized to the fraction loaded on the gel. The estimate of snRNA enrichment was based on the 1,000-fold decrease in RNA (1 mg in the input, compared with ∼1 μg in the eluate) and the nearly quantitative recovery of the snRNAs. The other half of the gel was silver stained using the GE PlusOne DNA Silver Staining Kit according to the manufacturer's instructions. The remainder of the eluted RNA was used for sequencing (see below).

Illumina Sequencing. Two independent samples of anti-TMG isolated RNA were used as input for two TruSeq RNA library preparations (Illumina). Poly-A selection and RNA fragmentation, the initial steps of library preparation, were omitted. All other steps of library preparation were performed according to the manufacturer's protocol. Library quality control, pooling, and paired-end sequencing were performed by the Biodiversity Research Centre sequencing facility at the University of British Columbia,Vancouver, Canada on a HiSEq. 2**000**. Libraries were sequenced on separate sequence runs to avoid bias associated with a given flow cell or lane therein, producing a total of 122.7 million 100-bp read pairs. Bowtie version 0.12.7 was used for read mapping (22). Reads first were mapped to *C. merolae* rDNA sequences to filter out rDNA contamination. rDNA-filtered reads were mapped to the *C. merolae* genome, a pileup file was created using SAMtools 0.1.18 mpileup option (23), and a custom Python script was used to identify contiguous stretches of expressed regions with greater than 500× coverage that did not overlap with annotated genes. This process resulted in 82 sequences from the first experiment and 87 from the second, with 58 (∼70%) common to the two experiments. The sequences from these regions were used for downstream analysis of snRNA presence/absence. High-throughput sequencing (HTseq) was used to obtain raw expression counts of known snRNAs (www-huber.embl.de/users/anders/HTSeq/).

Background levels of mRNAs also were recovered during immunoprecipitation. From these we observed splicing of the 27 introns at their expected junctions, as well as the singly-spliced forms of CMR350C, which contains two introns. Because of low-level representation, these background data do not rule out the possibility of alternative splicing.

Analysis of RIP-SEQ Sequences. Consensus sequences used to determine whether any of the sequences from the RIP-Seq experiments corresponded to a possible U1 snRNA were as follows:

i) Sm binding site $AU_{3-7}G/C$ or $AU_{3-7}NUG$

The traditional consensus site has a G following the Us, but because *C. merolae* U2 and U5 Sm-binding sites are both AAUUUUUCG, we searched for either nucleotide at this position.

The 5′ splice sites are 5′-GUAAGU (22), GUAGGU (2), GUAAGC (2), GCAAGU (1), with the number in parentheses indicating the number of *C. merolae* introns with this sequence.

ii) The complement to 5′ splice site therefore is (C/U)(A/G)UN(U/C)(G/A)-5′

iii) U1-70K binding site GAUCA

iv) U1A binding site UUGCAC

Sequences containing a site complementary to the 5′ splice site upstream of a possible Sm-binding site, with or without an intervening U1 protein-binding site, were aligned with a 100-sequence U1 seed from the rfam database (3) (green alga *Chlamydomonas reinhardtii* U1 snRNA was used as the reference sequence) using Clustal Omega (8).

1. Minoda A, Sakagami R, Yagisawa F, Kuroiwa T, Tanaka K (2004) Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, Cyanidioschyzon merolae 10D. *Plant Cell Physiol* 45(6): 667–671.
2. Matsuzaki M, et al. (2004) Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature* 428(6983):653–657.
3. Burge SW, et al. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* 41(Database issue):D226–D232.
4. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: Inference of RNA alignments. *Bioinformatics* 25(10):1335–1337.
5. Ding Y, Lawrence CE (2001) Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res* 29(5):1034–1046.
6. Ding Y, Chan CY, Lawrence CE (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32(web server issue):W135-41.
7. Ding Y, Chan CY, Lawrence CE (2005) RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* 11(8):1157–1166.
8. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
9. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415.
10. Agafonov DE, et al. (2011) Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* 31(13):2667–2682.
11. Hegele A, et al. (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* 45(4):567–580.
12. Fabrizio P, et al. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* 36(4):593–608.
13. Wang B-B, Brendel V (2004) The ASRG database: Identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol* 5(12):R102.
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
15. Ward N, Moreno-Hagelsieb G (2014) Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS ONE* 9(7):e101850.
16. Marchler-Bauer A, et al. (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39(Database issue):D225–D229.
17. Papadopoulos JS, Agarwala R (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073–1079.
18. Bessonov S, Anokhina M, Will CL, Urlaub H, Lührmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452(7189):846–850.
19. Mandl CW, Heinz FX, Puchhammer-Stöckl E, Kunz C (1991) Sequencing the termini of capped viral RNA by 5′-3′ ligation and PCR. *Biotechniques* 10(4):484–486, 486.
20. Lund E, Dahlberg JE (1992) Cyclic 2′,3′-phosphates and nontemplated nucleotides at the 3′ end of spliceosomal U6 small nuclear RNA's. *Science* 255(5042):327–330.
21. Achsel T, et al. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* 18(20):5789–5802.
22. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
23. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

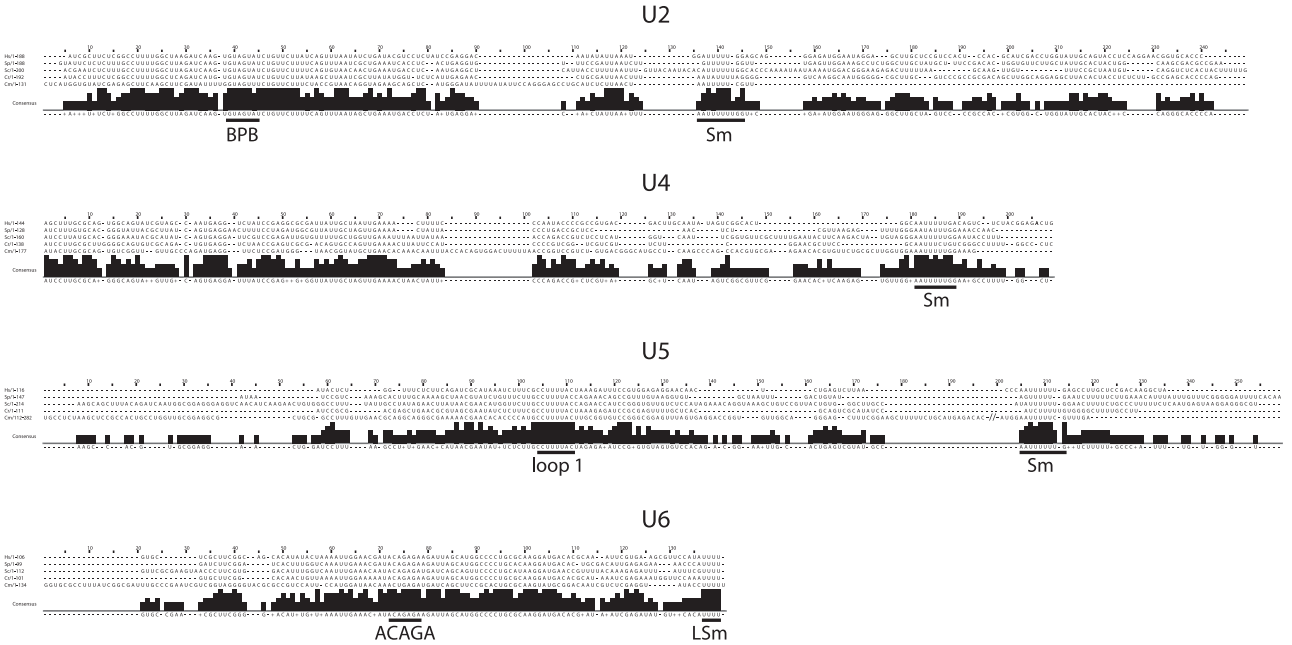

Fig. S1. Sequence alignment of *C. merolae* snRNAs and their homologs. snRNAs are aligned with homologs as indicated. Cr, *Chlamydomonas reinhardtii*; Cm, *Cyanidioschyzon merolae*; Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; BPB, branchpoint-binding region of U2; loop 1, invariant loop in U5; ACAGA, conserved motif in U6. Sm- and LSm-binding sites are as indicated. The bar graph below the sequence alignment shows the number of sequences in which that nucleotide is identical, and the consensus sequence is noted below the bar graph. Nucleotide position is indicated above the sequence alignment. Note that Sc U2 and Cm U5 are truncated. // in the Cm U5 sequence indicates missing nucleotides between the conserved core and the Sm-binding site.

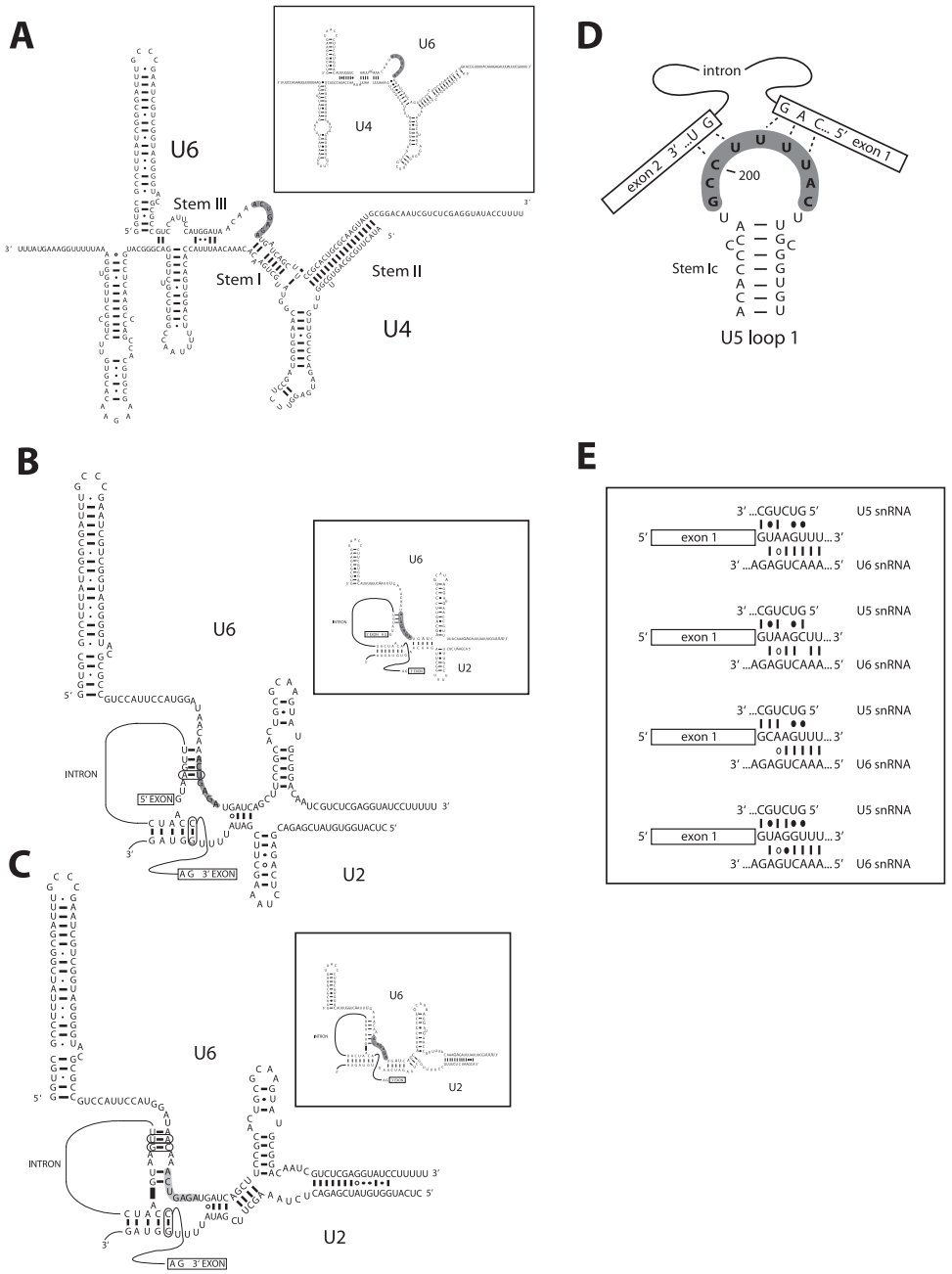

Fig. S2. Intermolecular interactions involving *C. merolae* snRNAs. (*A*) Base pairing between U4 and U6 based on Brow and Guthrie (1). (*B* and *C*) Base pairing between U2, U6, and the intron in the three-helix form (*B*) and the four-helix form (*C*), based on refs. 2 and 3, respectively. (*D*) Proposed interactions between U5 loop 1 and the exons (4). (*E*) Proposed model for interaction between snRNAs and all sequence variants of 5′ splice sites in *C. merolae*. The 5′ splice site of the pre-mRNA is indicated (*Middle*); the complementary region of U5 is shown above, and that of U6 is shown below. Vertical lines indicate Watson–Crick base pairs, closed circles denote G–U pairs, and the open circle denotes a possible G–A base pair. Interaction with U5 and U6 is proposed to occur sequentially.

1. Brow DA, Guthrie C (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334(6179):213–218.
2. Madhani HD, Guthrie C (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71(5):803–817.
3. Sun JS, Manley JL (1995) A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev* 9(7):843–854.
4. Newman AJ, Norman C (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* 68(4):743–754.

Other Supporting Information Files

Table S1 (PDF)
Table S2 (PDF)
Table S3 (PDF)
Table S4 (PDF)
Table S5 (PDF)
Table S6 (PDF)

Table S1. *C. merolae* snRNA sequences

snRNA	Chromosome accession number	FPKM*	chromosome/ location	sequence
U2	AP006493	690	11/762867-762997	CUCAUGGUGUAUCGAGAGCUUCAAGCUUCG AUAUUUUGGUAGUUUCUGUUCUUUCUACCG UAACAGGUAGAAGCAGCUCAUGGGAUAUUU UAUAUUCCAGGGAGCCUGCAUCUCUUAACU AAUUUUUCGUU
U4	AP006487	36095	5/222390-222566	AUACUUGCGCAGUGUCGGUUGUUGCCCAGA UGAGGUUCUCCGAUGGGUAACGGUAUGCU GAACACAAACAAUUUACCACAGUGGACUUU UUAACCGGUCCGUCUGUGACGGGCAUGCC UCAAGCCCAGCCACGUGCGAAGAACACGUG UUCUGCGCUUGGUGGAAAUUUUUGGAAAG
U5	AP006499	7046	17/771392-771841	GUCUGCAGGUGGCGCACCGCACGAAAGGU GGGGUUUCAGCGGCUCUCCAGUUUAGAGG ACCGGAGGGCAGCGCCCUUUCCUGACGCCA UGUACGGAUCCAUACGCGGGAGGUGCCUC UAAGCUCCGCCACUGCCUGGUUGCGGAGG CGCUGCGGCCUUUGUUGAACGCAGGCAGG GCGAAAAACGAACACACCCCAUGCCUUUUA CUUGCGGUGUCCGGGCGGAGUUUAGUGAG GACCGGUGUUGGCAGGGAGCUUUCGGAAG CUUUUUCUGCAUGAGACACAGUGUUCACAG GGCCUCCUCGUUACUUGGAUUCGUAUGUG GACCGGGACUGGGCGCCAGAGCGAAACUC GAGCUGUCGAGCACGCGCUGCCUGUAAAGU UUGCCUGCUCCCCAGCAGAAUCGUCUUUGG ACGACCGGCUUGAGGGAGUAUAUGGAAUUU UUCGUUUGA
U6	AP006501	168	19/483497-483364	GGUGCGCCUUUAUCGGCGAUUUGCCCGAA UCGUCGGUAGGGGUACGCGCCGUCCAUUC CAUGGAUAACAAACUGAGAUGAUCAGCUUC CGCACUGCGCAAGUAUGCGGACAAUCGUCU CGAGGUAUACCUUUUU

*FPKM fragments/Kb mapped

Table S2. snRNA Stem Stabilities

	C. merolae			*H. sapiens*			*S. cerevisiae*		
snRNA	**nt**	**%GC**	**ΔG (kJ/mol)**	**nt**	**%GC**	**ΔG (kJ/mol)**	**nt**	**%GC**	**ΔG (kJ/mol)**
U2									
Stem I	26	48	-38.5	20	63	-10.9	20	50	-28.4
Stem IIa	30	43	-35.6	20	33	-18.4	20	29	-25.5
Stem IIb	37	53	**-66.9**	17	67	-34.3	18	54	-14.2
Stem IIc	52	47	**-84.5**	16	25	-28.0*	52	32	-43.9
U4									
Stem I	15	27	18.0	15	55	7.1†	15	55	NFP
Stem II	31	59	**-51.5**	33	61	-36.8	34	47	-26.8
Stem III	24	31	7.1†	28	28	5	28	62	-27.2
Stem IV	31	61	-54.8	-	-	-	-	-	-
Stem V	47	64	-59.8†	33	64	-48.9	52	42	-75.3
Stem VI	-	-	-	14	60	-32.6	-	-	-
U5									
Stem Ia	32	63	-96.6*	8	50	-109*	18	44	-47.7*
Stem Ib	-	-	-	23	45	-54.0*	25	56	-61.1*
VSL	45	64	**-102.9**	-	-	-	34	46	-43.9
Stem Ic	27	59	-22.6	27	13	8.8	27	38	29.7
3' Stem Loop	44	59	-78.6	20	57	-43.5	28	52	-49.0
U6									
5' Stem Loop	50	61	**-128.9**	19	73	-44.8	25	60	-50.6
3' ISL	23	61	-56.5	22	50	-18.8	29	41	-38.1
U4/U6									
Stem I	20	40	-28.9*	16	31	-30.1*	18	50	-35.9*
Stem II	35	57	-116.7*	28	71	-88.7*	34	47	-93.7*
Stem III	-	-	-	18	57	-50.1	-	-	-
U2/U6									
Step 1 Intron I	8	25	-1.7*	10	20	-7.1*	6	33	4.2*
Step 1 Intron II	12	50	-29.3*	12	50	-25.9*	14	29	-25.9*
Step 1 Helix	8	38	-2.9†*	8	38	0.0*	10	40	-16.3*
Step 2 Intron 1	14	29	6.7*	10	20	-7.1*	6	33	4.2*
Step 2 Helix I	18	44	-15.9†*	16	44	-12.1*	18	44	-25.5
Step 2 Helix II	24	46	-10.9†	18	67	-41.0*	26	31	-56.5†*

* tetraloop added to cap stem for stability calculation
† A-G or A-C base pairs substituted with A-U for calculation
bold indicates notable stability of a Cm stem

Table S3. Protein search results for Reciprocal Best Hit (RBH) sequences

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
U1 snRNP					
	Snp1 (Sc)	Probable hnRNP A (CMR392C)	4E-7	Hrp1	6E-24
	Snp1 (Hs)	Similar to Cleavage Stimulation Factor (CMF108C)	2E-7	Similar to Cleavage Stimulation Factor, 3' pre-RNA, Subunit2, 64 kDa	2E-34
	Snp1 (Cr)	Similar to Gbp1p (CMM078C)	2E-6	Gbp1	2E-39
	Snp1 (Tb)	Probable hnRNP A (CMR392C)	4E-5	Similar to Polyadenylate Binding Protein 3	5E-15
	Snp1 (Xl)	Similar to Cleavage Stimulation Factor (CMF108C)	4E-7	Cleavage Stimulation Factor, 3' pre-RNA Subunit 2, 64 kDa	1E-35
	Snp1 (At)	Similar to Cleavage Stimulation Factor (CMF108C)	6E-7	Cleavage Stimulation Factor 64	4E-33
	Snp1 (Dm)	Similar to Cleavage Stimulation Factor (CMF108C)	8E-7	Cleavage Stimulation Factor 64 kDa Subunit	7E-30
	Snp1 (Mm)	Similar to Cleavage Stimulation Factor (CMF108C)	2E-7	Cleavage Stimulation Factor, 3' pre-RNA Subunit 2	2E-35
	Snp1 (Dd)	Similar to hnRNP (CMT598C)	1E-3	RNA-binding Region RNP-1 Domain-containing Protein	2E-7
	Snp1 (Co)	Probable hnRNP A (CMR392C)	8E-7	Hypothetical Protein CORT_0C04790	2E-29
	Snp1 (Bt)	Similar to Cleavage Stimulation Factor (CMF108C)	2E-7	Cleavage Stimulation Factor Subunit 2-like Protein	1E-35
	Snp1 (Sp)	Similar to TIA1 Cytotoxic Granule-Associated RNA-Binding Protein (CMS187C)	4E-7	RRM-containing Cyclophilin Regulating Transcription Rct1	4E-10
	Mud1 (Sc)	Hypothetical Protein (CMT606C)	7	Gyp6	1
	Mud1 (Hs)	Plastid Division Protein FtsZ2-1 (CMS004C)	3	Putative Tubulin Beta Chain-like Protein	3E-1

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Mud1 (Dm)	Probable hnRNP A (CMR392C)	5E-3	hnRNP 48.1	1E-33
	Mud1 (Ce)	Probable hnRNP A (CMR392C)	2E-4	Msi1	2E-26
	Mud1 (Sp)	Polyadenylate Binding Protein (CMJ286C)	5E-5	Polyadenylate Binding Protein	5E-129
	Mud1 (Xl)	eIF-3 Subunit G (CMH159C)	4E-4	eIF-3 Subunit G-A	3E-49
	Yhc1 (Sc)	Hypothetical Protein (CMT606C)	7	Gyp6	1
	Yhc1 (Hs)	Plastid Division Protein FtsZ2-1 (CMS004C)	3	Putative Tubulin Beta Chain-like Protein	3E-1
	Yhc1 (Sp)	Myo-inositol 1-phosphate Synthase (CMR036C)	1	Proteasome Component Ecm29	2E-1
	Yhc1 (Ca)	Similar to Oxidoreductase (CMQ274C)	2	Similar to Short Chain Alcohol Dehydrogenase	7E-6
	Yhc1 (Xl)	Plastid Division Protein FtsZ2-1 (CMS004C)	3	StAR-related Lipid Transfer Domain Containing 3	1
	Yhc1 (Bt)	Plastid Division Protein FtsZ2-1 (CMS004C)	3	Canniboid Receptor 2	8E-1
	Yhc1 (Dr)	Plastid Division Protein FtsZ2-1 (CMS004C)	3	Zinc Finger Protein 668	3
	Yhc1 (Mm)	Plastid Division Protein FtsZ2-1 (CMS004C)	3	Envelope Glycoprotein	6E+1
	Yhc1 (At)	Similar to SF3A Subunit 3 (CMQ406C)	2	Prp9p	1E-6
	Prp40 (Sc)	ORF515 (CMV241C)	6E-2	Mdn1	3E+3
	Prp40 (Hs)	Probable Eukaryotic Translation Initiation Factor eIF-5B (CML150C)	6E-2	eIF-5B	0
	Prp40 (Dm)	Similar to CCR4-NOT Transcription Complex, Subunit 4 (CMM335C)	2	CNOT4 Homolog	5E-29
	Prp40 (Xl)	Hypothetical Protein (CMG159C)	2E-1	Synaptosomal-associated Protein 47 kDa	2E-1
	Prp40 (At)	Hypothetical Protein (CMI068C)	5E-2	Putative Pectinesterase Inhibitor 28	1E-3

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Prp40 (Mm)	Probable Eukaryotic Translation Initiation Factor eIF-5B (CML150C)	1E-1	eIF-5B Protein	0
	Prp40 (Dd)	Similar to Vacuolar Sorting Protein/Ubiquitin Receptor VPS23 (CMK136C)	2E-1	Tumor Susceptibility Gene 101 Protein	1E-16
	Snu71 (Sc)	DNA Topoisomerase 1 (CMI252C)	1	DNA Topoisomerase 3	1E-25
	Snu71 (Hs)	Similar to Cleavage Stimulation Factor (CMF108C)	1E-6	Similar to Cleavage Stimulation Factor, 3' pre-RNA, Subunit2, 64kD	2E-34
	Snu71 (Mm)	Similar to Cleavage Stimulation Factor (CMF108C)	5E-6	Cleavage Stimulation Factor, 3' pre-RNA Subunit 2	2E-35
	Snu71 (At)	Similar to Cleavage Stimulation Factor (CMF108C)	8E-4	Cleavage Stimulation Factor 64	4E-33
	Snu71 (Xl)	Similar to Cleavage Stimulation Factor (CMF108C)	1E-6	Cleavage Stimulation Factor, 3' pre-RNA Subunit 2, 64 kDa	1E-35
	Snu71 (Dm)	Similar to Cleavage Stimulation Factor (CMF108C)	6E-3	Cleavage Stimulation Factor 64 kDa Subunit	7E-30
	Snu71 (Dd)	Similar to Cleavage Stimulation Factor (CMF108C)	1E-2	Cleavage Stimulation Factor 54 kDa Subunit	1E-18
	Prp39 (Sc)	Dynamin-related Protein Drp5 (CMN262C)	1E-1	Vps1	6E-105
	Prp39 (Hs)	Similar to psbB mRNA Maturation Factor Mbb1 (CMQ412C)	2E-18	UDP-N-Acetyl-glucosamine	4E-64
	Prp39 (Mm)	Similar to psbB mRNA Maturation Factor Mbb1 (CMT510C)	5E-15	UDP-N-Acetyl-glucosamine	2E-75
	Prp39 (At)	Similar to psbB mRNA Maturation Factor Mbb1 (CMT510C)	2E-20	UDP-N-Acetyl-glucosamine	5E-69
	Prp39 (Ce)	Similar to psbB mRNA Maturation Factor Mbb1 (CMT510C)	3E-12	Protein OGT-1	3E-70

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Prp39 (Dm)	Similar to psbB mRNA Maturation Mbb1 (CMT510C)	5E-12	Super Sex Combs	2E-71
	Prp39 (Dr)	Similar to psbB mRNA Maturation Factor (CMT510C)	2E-17	O-linked N-acetyl-glucosamine Transferase	2E-77
	Prp42 (Sc)	Similar to psbB mRNA Maturation Factor (CML157C)	1E-3	Cyc8	4E-7
	Prp42 (Ca)	Hypothetical Protein, Conserved (CMS420C)	3E-2	Pre-mRNA 3'-end Processing Factor RNA14 Protein	3E-35
	Prp42 (Co)	Stress-induced Phosphoprotein STI1 (CMR299C)	9E-1	STI1 Protein	1E-116
	Nam8 (Sc)	Polyadenylate Binding Protein (CMJ286C)	9E-43	Pab1	1E-86
	Nam8 (Cr)	Polyadenylate Binding Protein (CMJ286C)	2E-24	PAB Protein RB47	2E-66
	Nam8 (At)	Polyadenylate Binding Protein (CMJ286C)	1E-57	PAB8	3E-125
	Nam8 (Dd)	Polyadenylate Binding Protein (CMJ286C)	8E-10	RNA-binding Domain-containing Protein	8E-68
	Nam8 (Ca)	Polyadenylate-binding Protein (CMJ286C)	3E-8	Polyadenylate Binding Protein	1E-66
	Snu56 (Sc)	Small GTP-binding Protein Arf1 (CMQ074C)	1	Arf family GTPase Arf2	2E-98
	Luc7 (Sc)	C-type Cytochrome Biogenesis Protein Ccs1 (CMV075C)	3E-1	Mig3	7E-1
	Luc7 (Hs)	Hypothetical Protein (CMT366C)	1	Interleukin 17-C Receptor	6
	Luc7 (Mm)	Hypothetical Protein (CMT366C)	1	Cytosolic Carboxypeptidase 6	4E-1
	Luc7 (At)	Hypothetical Protein (CMH110C)	5E-1	TRAF-like Protein	4E-2
	Luc7 (Xl)	Similar to Kinetoplast-associated Protein (CMJ046C)	5E-2	Serine/Threonine-protein Kinase MRCK Alpha-like	8E-5
	Luc7 (Dm)	Ubiquitin-protein Ligase E3 (CMR077C)	2E-1	SD03277 Protein	2E-123

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Luc7 (Dd)	Hypothetical Protein (CMT069C)	9E-2	WD40-like Domain-containing Protein	2E-1
	Urn1 (Sc)	Hypothetical Protein (CMP346C)	1E+1	Dgr2	6E-1
	Npl3 (Sc)	Sister-Chromatid Cohesion Complex Cohesin, Subunit SMC3 (CML027C)	3E-1	Cohesin Subunit Smc3	1E-164
	Npl3 (Ca)	Polyadenylate-binding Protein (CMJ286C)	3E-8	Polyadenylate Binding Protein	1E-66
	Npl3 (Sp)	Probable Serine Rich Pumilio Family RNA Binding Domain Protein (CMR037C)	2E-14	RNA-binding Protein	7E-74
Sm Proteins					
	B/B' (Sc)	Similar to Sm Protein B (CMK022C)	4E-6	SmB1 Protein	2E-11
	B/B' (Hs)	†Similar to Sm Protein B (CMK022C)	4E-5	snRNP Polypeptide B	3E-35
	D1 (Sc)	Sm Protein D1 (CMF084C)	7E-11	SmD1 Protein	3E-15
	D1 (Hs)	Sm Protein D1 (CMF084C)	2E-7	snRNP Sm D1	7E-11
	D2 (Sc)	Similar to Sm Protein D2 (CMN302C)	3E-14	SmD2 Protein	1E-17
	D2 (Hs)	Similar to Sm Protein D2 (CMN302C)	2E-13	snRNP D2	6E-17
	D3 (Sc)	Sm Protein D3 (CMM065C)	1E-9	SmD3 Protein	1E-13
	D3 (Hs)	Sm Protein D3 (CMM065C)	1E-10	snRNP Sm D3	1E-14
	E (Sc)	Similar to Sm Protein E (CMM109C/CMH215C)	1E-8	SmE1 Protein	1E-13
	E (Hs)	Similar to Sm Protein E (CMM109C/CMH215C)	5E-11	snRNP E	2E-15
	F (Sc)	Sm Protein F (CMQ171C)	2E-13	SmX3 Protein (SmF Protein)	2E-20
	F (Hs)	Sm Protein F (CMQ171C)	1E-13	snRNP F	2E-19

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	G (Sc)	Similar to Sm Protein G (CMO342C)	2E-6	SmX2 Protein (SmG Protein)	2E-10
	G (Hs)	Similar to Sm Protein G (CMO342C)	5E-7	snRNP G	2E-11
U2 snRNP					
	Prp9 (Sc)	Similar to SF3A Subunit 3 (CMQ406C)	5E-15	Prp9	1E-44
	Prp9 (Hs)	Similar to SF3A Subunit 3 (CMQ406C)	8E-15	SF3A Subunit 3 (Prp9)	3E-45
	Prp11 (Sc)	Similar to SF3A Subunit 2 (CMH102C/CMN095C)	2E-9	Prp11	7E-10
	Prp11 (Hs)	Similar to SF3A Subunit 2 (CMH102C/CMN095C)	5E-11	SF3A Subunit 2 (Prp11)	7E-10
	Prp21 (Sc)	Hypothetical Protein (CMJ300C)	2E-4	Prp21	6E-9
	Prp21 (Hs)	†Hypothetical Protein (CMJ300C)	3E-7	Splicing Factor 3A Subunit 1 (Prp21)	1E-12
	Prp21 (Mm)	†Hypothetical Protein (CMJ300C)	3E-7	Splicing Factor 3A Subunit 1 (Prp21)	9E-14
	Prp21 (Dm)	†Hypothetical Protein (CMJ300C)	9E-9	Splicing Factor 3A Subunit 1 (Prp21)	7E-13
	Prp21 (Ce)	†Hypothetical Protein (CMJ300C)	2E-7	PRP21	2E-11
	Prp21 (Sp)	Hypothetical Protein (CMJ300C)	6E-11	Splicing Factor Sap114 (Prp21)	5E-12
	Lea1 (Sc)	Similar to Protein Phosphatase 1 (CMM185C)	5E-3	Sds22	1E-44
	Lea1 (Hs)	Similar to Protein Phophatase-1 (CMM185C)	1E-6	Protein Phosphatase-1	1E-54
	Msl1 (Sc)	Probable RNA Binding Protein Mrd1P (CMO246C)	2E-2	Mrd1	8E-79
	Msl1 (Hs)	Hypothetical Protein, Conserved (CMA050C)	2E-1	Protein Lin-54 Homolog	5E-18
	Rse1 (Sc)	Similar to SF3b Subunit 3 (CML103C)	5E-4	Rse1	8E-4
	Rse1 (Hs)	Similar to SF3b Subunit 3 (CML103C)	2E-20	SF3B Subunit 3 (Rse1)	2E-24

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Rse1 (Ca)	Similar to SF3b Subunit 3 (CML103C)	9E-9	Potential U2 snRNP Complex SF3b Component	3E-11
	Rse1 (Mm)	Similar to SF3b Subunit 3 (CML103C)	2E-20	SF3B Subunit 3 (Rse1)	5E-23
	Ysf3 (Sc)	Hypothetical Protein, Conserved (CMI258C)	1E-1	Erc1	6E-38
	Ysf3 (Hs)	Probable Protease (CMQ089C)	2	Mitochondrial Processing Peptidase	6E-14
	SF3B14A (Hs)	Similar to RNA-Binding Protein with RRM (CMI122C)	6E-4	MK167 FHA Domain Nucleolar Phosphoprotein	1E-21
	Hsh49 (Sc)	†Similar to Splicing Factor 3B Subunit 4 (CME063C)	6E-10	Hsh49	2E-9
	Hsh49 (Hs)	†Similar to Splicing Factor 3B Subunit 4 (CME063C)	6E-10	SF3B Subunit 4 (Hsh49)	1E-13
	Rds3 (Sc)	Hypothetical Protein, Conserved (CMS014C)	7E-15	Rds3	3E-37
	Rds3 (Hs)	Hypothetical Protein, Conserved (CMS014C)	1E-12	PHF5A (Rds3)	3E-34
	Hsh155 (Sc)	Probable Splicing Factor 3B Subunit 1 (CMB002C)	3E-56	Hsh155	1E-61
	Hsh155 (Hs)	Probable Splicing Factor 3B Subunit 1 (CMB002C)	3E-66	SF3B Subunit 1 (Hsh155)	1E-73
	Cus1 (Sc)	Similar to Splicing Factor 3B Subunit 2 (CMT357C)	1E-23	Cus1	3E-23
	Cus1 (Hs)	Similar to Splicing Factor 3B Subunit 2 (CMT357C)	2E-22	SF3B Subunit 2 (Cus1)	4E-30
	Prp5 (Sc)	†Probable DEAD Box RNA Helicase (CMR433C)	3E-47	Prp5	1E-49
	Prp5 (Hs)	p68 RNA Helicase (CMR479C)	1E-125	DDX5 RNA Helicase	0
	Prp5 (Sp)	†Probable DEAD Box RNA Helicase (CMR433C)	8E-66	prp11 (Prp5)	4E-66
	Prp5 (Ce)	†Probable DEAD Box RNA Helicase (CMR433C)	4E-54	F53H1.1 (Prp5)	3E-62
U2 Related					
	Prp43 (Sc)	Pre-mRNA Splicing Factor ATP-dependent RNA Helicase PRP43 (CMM048C)	0	Prp43	0

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Prp43 (Hs)	Pre-mRNA Splicing Factor ATP-dependent RNA Helicase PRP43 (CMM048C)	0	Prp43	0
	Mud2 (Sc)	UDP-Glucose 4-Epimerase (CMA041C)	6E-1	UDP-Glucose 4-Epimerase	4E-126
	Mud2 (Hs)	Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	5E-13	Splicing Factor U2AF 65 kDa Subunit (Mud2)	2E-52
	Mud2 (Dm)	Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	8E-11	U2AF50 Protein (Mud2)	7E-63
	Mud2 (Sp)	†Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	3E-2	U2 Auxiliary Factor Small Chain	5E-29
	Mud2 (Ce)	†Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	1E-3	UAF-1 (Mud2)	2E-32
	Cus2 (Sc)	Similar to Nucleolar RNA-Binding Protein (CMO334C)	4E-2	Nop12	4E-17
	Cus2 (Hs)	Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	3E-4	Splicing Factor U2AF 65 kDa Subunit	2E-52
	U2AF1 (Hs)	Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	9E-3	Splicing Factor U2AF 65 kDa Subunit	2E-52
	Puf60 (Hs)	Polyadenylate Binding Protein (CMJ286C)	1E-11	Polyadenylate Binding Protein-1	8E-69
	Smndc1 (Hs)	eIF-3 Subunit G (CMH159C)	4E-1	eIF-3 Subunit 4	4E-48
	Rbm17 (Hs)	Similar to pre-mRNA Splicing Factor U2AF Large Chain (CMS438C)	6E-3	Splicing Factor U2AF 65 kDa Subunit	2E-52
	U2 SURP (Hs)	Similar to TIA1 Cytotoxic Granule-Associated RNA-Binding Protein (CMS187C)	4E-7	TIA1 Protein	1E-12
	DDX42 (Hs)	p68 RNA Helicase (CMR479C)	1E-100	DDX5 RNA Helicase	0
	CHERP (Hs)	Hypothetical Protein (CMN293C)	2E-1	Pin2/TERF1 Interaction Telomerase	5E-8

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
U5 snRNP					
	Prp8 (Sc)	Pre-mRNA Splicing Factor Prp8 (CMH168C)	0	Prp8	0
	Prp8 (Hs)	Pre-mRNA Splicing Factor Prp8 (CMH168C)	0	Prp8p	0
	Brr2 (Sc)	U5 Small Nuclear Ribonucleoprotein 200 kDa (CML192C)	1E-98	Brr2	1E-117
	Brr2 (Hs)	U5 Small Nuclear Ribonucleoprotein 200 kDa (CML192C)	1E-101	U5 snRNP 200 kDa Protein (Brr2p)	1E-119
	Prp6 (Sc)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	9E-6	Cyc8	4E-7
	Prp6 (Hs)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	1E-19	Crooked Neck-like 1 Protein	2E-13
	Prp6 (Mn)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	5E-19	PsbB mRNA Maturation Factor Mbb1	1E-24
	Prp6 (Co)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	4E-10	Prp6p	8E-9
	Prp6 (At)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	7E-20	Protein High Chlorophyll Fluorescent 107	8E-25
	Prp6 (Cr)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	5E-10	PsbB Maturation Factor	1E-27
	Prp6 (Bd)	Similar to psbB mRNA Maturation Factor Mbb1 (CML157C)	6E-18	Cytochrome Oxidase Subunit I	2
	Dib1 (Sc)	Thioredoxin-like U5 snRNP component Dim1 (CMN033C/CMS018C)	4E-37	Dib1	3E-72
	Dib1 (Hs)	Thioredoxin-like U5 snRNP component Dim1 (CMN033C/CMS018C)	3E-24	Dim1p (Dib1)	2E-78
	Snu114 (Sc)	†Similar to U5 snRNP-specific Protein (CMK208C)	7E-40	Snu114	7E-42

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Snu114 (Hs)	†Similar to U5 snRNP-specific Protein (CMK208C)	2E-43	EFTUD2p (Snu114)	1E-42
	Prp28 (Sc)	ATP-Dependent RNA Helicase (CMT173C)	1E-60	Ded1	6E-180
	Prp28 (Hs)	p68 RNA Helicase (CMR479C)	3E-68	DDX5 RNA Helicase	0
	Prp28 (Cr)	p68 RNA Helicase (CMR479C)	4E-80	p68	0
	Lin1 (Sc)	eIF2-alpha Kinase GCN2 (CMN123C)	1	Protein Kinase GC2	4E-29
	Lin1 (Hs)	Hypothetical Protein, Conserved (CMS418C)	3	C6orf136	4E-3
	snRNP40 (Hs)	Unknown WD-repeat Protein WDS (CMC158C)	5E-30	WD Repeat Containing Protein	1E-138
	Aar2 (Hs)	Hypothetical Protein (CMJ012C)	4	RUNX1p	4
U4/U6 snRNP					
	Prp31 (Sc)	Box C/D snoRNP Component Nop58 (CMT605C)	7E-15	Nop58	8E-129
	Prp31 (Hs)	Box C/D snoRNP Component Nop58 (CMT605C)	6E-19	Nop58p	8E-129
	Prp31 (Cr)	Box C/D snoRNP Component Nop58 (CMT605C)	3E-21	Nucleolar Protein, Component C/D snoRNPs	9E-156
	Prp31 (Tb)	Box C/D snoRNP Component Nop58 (CMT605C)	6E-13	Nucleolar RNA Binding Protein	6E-118
	Prp31 (Dd)	Box C/D snoRNP Component Nop58 (CMT605C)	1E-17	Nop5 Family Protein	2E-124
	Prp3 (Sc)	Hypothetical Protein (CMT170C)	1E-9	Khr1	3E-1
	Prp3 (Hs)	Hypothetical Protein (CMT170C)	7E-11	PRP3	3E-7
	Prp3 (Cr)	Hypothetical Protein (CMT170C)	2E-2	Hypothetical Protein 175482 (Prp3)	5E-5

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Prp3 (At)	Hypothetical Protein (CMT170C)	8E-9	Prp3	6E-3
	Snu13 (Sc)	Box C/D snoRNP and U4 snRNP Component Snu13p (CMP335C)	2E-32	Snu13	5E-52
	Snu13 (Hs)	Box C/D snoRNP and U4 snRNP Component Snu13p (CMP335C)	3E-28	Snu13p	1E-49
	Prp4 (Sc)	Unknown WD repeat protein WDS (CMC158C)	1E-74	Rsa4	4E-73
	Prp4 (Hs)	Unknown WD repeat protein WDS (CMC158C)	4E-83	Dynein assembly factor with WDR domains	7E-86
	PPIH (Hs)	Cyclophilin (CMO300C)	1E-47	Cyclophilin D	7E-70
	Prp24 (Sc)	Polyadenylate Binding Protein (CMJ286C)	1E-11	PABP1	8E-69
	Prp24 (Hs)	Probable hnRNP protein A (CMR392C)	2E-9	Daz-associated Protein 1	4E-31
Tri-snRNP					
	Snu66 (Sc)	Hypothetical Protein (CMJ064C)	1E-1	Snu66	1E+1
	Snu66 (Hs)	NADH Dehydrogenase I (Complex I) (CMM030C)	3	NADH-Ubiquinone Oxidoreductase B22 Subunit Homolog	1E-10
	Snu66 (Ca)	ORF515 (CMV241C)	2	YL1 Nuclear Protein	1E+2
	Snu66 (Rn)	Excinuclease ABC Subunit B (CMG133C)	5E-1	eIF 4A-III	1E-4
	Snu66 (Bt)	Hypothetical Protein (CMM056C)	2	Myosin Heavy Chain 3	9E-1
	Snu66 (Pd)	Lumina Binding Protein BiP (CMT579C)	1	Hsp70	0
	Spp381 (Sc)	Hypothetical Protein (CMR327C)	3E-2	Set1	1E+1
	Sad1 (Sc)	Unknown Zinc-finger Protein (CMK309C)	5E-2	Etp1	8E-58
	Sad1 (Hs)	Ubiquitin-specific Protease UBP15 (CMH074C)	5E-7	UBP15p	1E-73

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
LSm Proteins					
	LSm1 (Sc)	Similar to Sm-like Protein LSm1 (CMT394C)	2E-15	LSm1	2E-17
	Lsm1 (Hs)	Similar to Sm-like Protein LSm1 (CMT394C)	2E-15	LSm1p	1E-19
	LSm2 (Sc)	snRNP Core Sm-like SmX5 (CMB130C)	2E-15	LSm2	5E-20
	Lsm2 (Hs)	snRNP Core Sm-like Protein SmX5 (CMB130C)	2E-15	LSm2p	6E-23
	LSm3 (Sc)	U6 snRNA-associated Sm-like Protein LSm3 (CMT262C)	3E-17	LSm3	3E-20
	LSm3 (Hs)	U6 snRNA-associated Sm-like Protein LSm3 (CMT262C)	3E-17	LSm3p	3E-17
	LSm4 (Sc)	Similar to U6 snRNA-associated Sm-like Protein (CMG061C/CMT545C)	7E-11	LSm4	1E-17
	LSm4 (Hs)	Similar to U6 snRNA-associated Sm-like Protein (CMG061C/CMT545C)	9E-7	LSm4p	1E-27
	LSm5 (Sc)	U6 snRNA-associated Sm-like Protein LSm5 (CMP159C)	7E-11	LSm5	2E-16
	LSm5 (Hs)	U6 snRNA-associated Sm-like Protein LSm5 (CMP159C)	4E-14	LSm5p	5E-18
	LSm6 (Sc)	U6 snRNA-associated Sm-like Protein LSm6 (CMP138C)	1E-6	LSm6	1E-10
	LSm6 (Hs)	U6 snRNA-associated Sm-like Protein LSm6 (CMP138C)	5E-24	LSm6p	1E-28
	Lsm7 (Sc)	†U6 snRNA-associated Sm-like Protein LSm7 (CMP206C)	1E-6	LSm7	1E-11
	LSm7 (Hs)	Similar to U6 snRNA-associated Sm-like Protein LSm7 (CMP206C)	3E-5	LSm7p	5E-5
	Lsm8 (Sc)	Similar to Sm-like Protein LSm1 (CMT394C)	4E-4	LSm1	2E-17

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	LSm8 (Hs)	Similar to Sm-like Protein LSm1 (CMT394C)	3E-10	LSm1p	1E-19
Cap Binding Complex					
	Sto1 (Sc)	Probable mRNA Binding Cap Protein 80 (CMJ189C)	5E-32	Gcr3 (Sto1)	8E-65
	Sto1 (Hs)	Probable mRNA Binding Cap Protein 80 (CMJ189C)	3E-46	Nuclear Cap Protein Subunit 1, 80 kDa Variant (Sto1)	5E-92
	Cbc2 (Sc)	Similar to Nuclear Cap-binding Protein; CBP20 (CMQ282C)	2E-29	Cbc2	3E-36
	Cbc2 (Hs)	Similar to Nuclear Cap-binding Protein; CBP20 (CMQ282C)	3E-30	Nuclear Cap Protein Subunit 2 (Cbc2)	1E-43
Recruited At A Complex					
	Msl5 (Sc)	Branchpoint Bridging Protein Msl5p (CMI292C)	1E-16	Msl5	5E-35
	Msl5 (Hs)	Branchpoint Bridging Protein Msl5p (CMI292C)	6E-26	SF1 (Msl5)	2E-39
	Sub2 (Sc)	ATP Dependent RNA Helicase p47 (CME073C)	1E-145	Sub2	0
	Sub2 (Hs)	ATP Dependent RNA Helicase p47 (CME073C)	1E-150	RNA Helicase DDX39B (Sub2)	0
	Prp40 (Sc)	ORF515 (CMV241C)	6E-2	Mdn1	3E+3
	Prp40 (Hs)	Probable Eukaryotic Translation Initiation Factor eIF-5B (CML150C)	6E-2	eIF-5B Protein	0
	Snu71 (Sc)	DNA Topoisomerase 1 (CMI252C)	1	DNA Topoisomerase 3	1E-25
	Snu71 (Hs)	Similar to Cleavage Stimulation Factor (CMF108C)	1E-6	Similar to Cleavage Stimulation Factor, 3' pre-RNA, Subunit2, 64kD	2E-34
	Thrap3 (Hs)	Similar to SR Family Splicing Factor SC35 (CML202C)	7E-6	SRSF2	3E-12

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Ccar1 (Hs)	Hypothetical Protein, Conserved (CMM216C)	2	E1B-55 kDa	4E-7
	SUGP1 (Hs)	Hypothetical Protein (CMN293C)	3E-2	Pin2/TERF1 Interaction Telomerase	5E-8
	RBM5 (Hs)	Similar to TIA1 Cytotoxic Granule-Associated RNA-Binding Protein (CMS187C)	4E-4	TIA1 Protein	1E-12
	RBM10 (Hs)	Hypothetical protein (CMN293C)	6E-4	Pin2/TERF1 Interaction Telomerase	5E-8
NTC					
	Prp19 (Sc)	Ubiquitin conjugation factor E4 B (CMS041C)	3E-2	Ufd2	1E-80
	Prp19 (Hs)	Similar to Vegetatible Incompatibility Protein HET-E-1 (CMN161C)	2E-25	Dynein Assembly Factor	8E-31
	Prp19 (Cr)	Receptor for Activated Protein Kinase C (CMI283C)	4E-60	RACK1	4E-72
	Prp19 (At)	Mda1 (CMR185C)	1E-17	Unamed Protein Product	2E-20
	Cef1 (Sc)	Cell Division Control Protein 5 (CMR098C)	3E-13	Cef1	3E-28
	Cef1 (Hs)	Cell Division Control Protein 5 (CMR098C)	1E-15	CDC5 (Cef1)	9E-15
	Cef1 (Sp)	Cell Division Control Protein 5 (CMR098C)	2E-17	Cdc5 (Cef1)	4E-33
	Prp46 (Sc)	Similar to Pleiotropic Regulator 1 (CMR305C)	1E-26	Prp46	2E-29
	Prp46 (Hs)	Similar to Pleiotropic Regulator 1 (CMR305C)	1E-34	PRL1 (Prp46)	2E-38
	Cwc15 (Sc)	DegP Protease (CMM292C)	3E-1	Pin2/TERF1 Interaction Telomerase	1E+1
	Cwc15 (Hs)	Arginine--TRNA ligase (CMN177C)	2E+1	Arginyl-tRNA Synthetase	3E-153
	Snt309 (Sc)	DNA Polymerase Epsilon, catalytic subunit (CMQ098C)	2	DNA Pol2p	0

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Snt309 (Hs)	Similar to Myosin Heavy Chain (CMO091C)	3E-3	Desmoplakin	2E-2
	Snt309 (At)	SMC3 (CML027C)	3E-2	SMC3	4E-140
	Snt309 (Dm)	Hypothetical Protein (CMJ247C)	1E-1	Tubulin Tyrosine Ligase-like 3A	7
	Ssa4 (Sc)	Heat Shock Protein Hsp70 (CMP145C)	0	Hsp70 Family ATPase SSA3	0
	Ssa4 (Hs)	Heat Shock Protein Hsp70 (CMP145C)	0	Heat Shock Cognate 71 kDa Protein	0
	CTNNBL1 (Hs)	Adaptor-related protein complex 3, delta subunit (CMQ028C)	8E-1	AP3D1 Protein	3E-54
	WBP11 (Hs)	Arogenate/prephenate Dehyrogenase (CMS326C)	6E-1	WW domain-Binding Protein 11	1E+1
	PQBP1 (Hs)	Dynamin-related Protein Involved in Mitochondrial Division (CME019C)	6E-2	Dynamin-1-like Protein	0
	AD002 (Hs)	Arginine-TRNA ligase (CMN177C)	2E+1	Arginyl-tRNA Synthetase	3E-153
NTC Related					
	Clf1 (Sc)	Similar to psbB mRNA Maturation Factor (CML157C)	2E-33	Cyc8	4E-47
	Clf1 (Hs)	Similar to psbB Maturation Factor Mbb1 (CMT510C)	3E-39	UDP-N-Acetyl-Glucosamine	2E-75
	Ecm2 (Sc)	Similar to Retromer Component VPS5 (CME083C)	4E-1	Snx4	7E-3
	Ecm2 (Hs)	Similar to Protein Phosphatase 1, Regulatory Subunit (CMM185C)	4E-6	Protein Phosphatase 1 Regulatory Subunit 7	7E-50
	Cwc2 (Sc)	Probable RNA Binding Protein Mrd1p (CMO246C)	4E-2	Mrd1	8E-79
	Cwc2 (Hs)	Similar to TIA1 Cytotoxic Granule-Associated RNA-Binding Protein (CMS187C)	2E-2	TIA1 Protein	1E-12

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Bud31 (Sc)	Similar to G10 Protein (CMG014C)	1E-13	Bud31	3E-18
	Bud31 (Hs)	Similar to G10 Protein (CMG014C)	2E-15	Bud31p	4E-19
	Prp45 (Sc)	DNA mismatch repair protein MutS (CML164C)	1E-2	Mismatch repair ATPase MSH2	8E-128
	Prp45 (Hs)	DNA mismatch repair protein MutS (CML164C)	1E-2	DNA mismatch repair protein Msh2	6E-122
	Isy1 (Sc)	L-arabinose permease (CMK066C)	2E-1	Gal2	6E-85
	Isy1 (Hs)	Similar to mitochondrial magnesium transporter Mrs2p (CMM153C)	4E-1	Magnesium transporter MRS2 homolog	1E-59
	Syf1 (Sc)	Hypothetical protein (CML093C)	9E-3	Cyc8	2E-41
	Syf1 (Hs)	Stress-induced phosphoprotein STI1 (CMR299C)	2E-26	Stress induced phosphoprotein 1a	8E-69
	PPIL1 (Hs)	Cyclophilin (CMO300C)	3E-24	Cyclophilin D	7E-70
	AQR (Hs)	Hypothetical Protein, Conserved (CMK133C)	1E-11	DNA Replication ATP-dependent Helicase/Nuclease	2E-141
	PPIE (Hs)	Cyclophilin (CMO300C)	7E-54	Cyclophilin D	7E-70
Recruited at B Complex					
	Spp382 (Sc)	Similar to Putative Vesicle Transport Protein (CMF111C)	4E-1	Sec22	3E-43
	Spp382 (Hs)	Hypothetical Protein (CMN293C)	3E-1	Pin2/TERF1 Interaction Telomerase	5E-8
	Snu23 (Sc)	Hypothetical Protein (CMA025C)	1	Gcn5	6E-2
	Snu23 (Hs)	Similar to Splicing Factor 3a Subunit 2 (CMN095C)	1E-1	Splicing Factor 3a, Subunit 2, 66kDa	7E-10
	Prp38 (Sc)	Hypothetical Protein (CMJ144C)	1E-18	Prp38	1E-30

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Prp38 (Hs)	Hypothetical Protein (CMJ144C)	2E-22	Pre-mRNA Splicing Factor 38A (Prp38)	1E-57
	Prp38 (At)	Hypothetical Protein (CMJ144C)	6E-22	Prp38p	4E-56
	Prp38 (Dm)	Hypothetical Protein (CMJ144C)	1E-22	Prp38p	9E-55
	Smu1 (Hs)	Unknown WD-repeat Protein WDS (CMC158C)	2E-21	WD Repeat Containing Protein	1E-138
	MFAP1 (Hs)	Unknown Heatshock Protein (CMO059C)	4	DnaJ Homolog Subfamily C Member 21	1E-62
	IK (Hs)	Hypothetical Protein (CMM016C)	9E-1	Hypothetical Protein pp8997 Variant	1
	WBP4 (Hs)	ORF515 (CMV241C)	1	Nestin	9E+1
	PrfF4B (Hs)	Protein Kinase YAK1 (CMH056C)	7E-37	Tyrosine Phosphorylation-Regulated Kinase	1E-77
Recruited at B^{act} Complex					
	Cwc27 (Sc)	Hypothetical Protein (CMQ392C)	7E-3	Putative Peptidylprolyl Isomerase CWC27	1E-2
	Cwc27 (Hs)	Cyclophilin (CMO300C)	1E-47	Cyclophilin D	7E-70
	Cwc27 (Bt)	Cyclophilin (CMO300C)	5E-17	Peptidylprolyl Isomerase A	3E-68
	Cwc27 (Dr)	Cyclophilin (CMO300C)	2E-17	Peptidylprolyl Isomerase A	3E-72
	Prp2 (Sc)	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP43 (CMM048C)	1E-117	Prp43	0
	Prp2 (Hs)	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP43 (CMM048C)	1E-143	Prp43p	0
	Prp2 (At)	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP43 (CMM048C)	1E-161	Prp43	0

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Cwc22 (Sc)	Hypothetical Protein, Conserved (CMF179C)	2	Cpr5	6E-1
	Cwc22 (Hs)	Similar to Inner Centromere Protein (CMS331C)	2	Radixin	6E+1
	Spp2 (Sc)	Similar to Cyclin M2 (CMI026C)	2	Mam3	4E-43
	Spp2 (Hs)	Hypothetical protein (CMJ157C)	1E-2	G Patch Domain-Containing Protein 2	2E-3
	Cwc24 (Sc)	Similar to Ring Finger Protein (CME171C)	2E-17	Ubiquitin Protein Ligase Peoxin 10	1E-15
	Cwc24 (Hs)	Photoregulatory Zinc-Finger Protein COP1 (CMK039C)	2E-5	Ring Finger and WD-Repeat Protein	1E-96
	Cwc25 (Sc)	N-formylgutamate Deformylase (CMT029C)	4	Prp16	4
	Cwc25 (Hs)	Similar to Programmed Cell death Protein 5 (TFAR19) (CMF042C)	1E-1	PDCD Protein 5	1E-16
	Yju2 (Sc)	Hypothetical protein (CMN267C)	3E-5	Yju2	5E-5
	Yju2 (Hs)	Hypothetical Protein (CMN267C)	6E-3	CCDC94 (Yju2)	4E-3
	Yju2 (At)	Hypothetical Protein (CMN267C)	4E-3	mRNA Splicing Protein Yju2	2E-4
	Yju2 (Pb)	Hypothetical Protein (CMN267C)	1E-2	Related to Yju2-essential Nuclear Protein	2E-4
	Yju2 (Co)	Hypothetical Protein (CMN267C)	4E-3	Yju2p	2E-4
	Yju2 (Rn)	Hypothetical Protein (CMN267C)	5E-3	CCDC94 Protein (Yju2)	2E-3
	Yju2 (Dd)	Hypothetical Protein (CMN267C)	9E-2	CCDC94 (Yju2)	2E-3
	Yju2 (Dr)	Hypothetical Protein (CMN267C)	9E-3	CCDC94 (Yju2)	2E-3
	Znf830 (Hs)	Hypothetical Protein (CMF075C)	7E-3	Splicing factor 3a, subunit 2, 66kDa	7E-10
	CCDC12 (Hs)	Hypothetical Protein (CMQ091C)	3E-1	Keratin 23p	1

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	PRCC (Hs)	Similar to Histidinol-Phosphate Aminotransferase (CMR123C)	2	Alanine Aminotransferase 2	3E-1
	GPATCH1 (Hs)	Similar to Transcription Elongation Factor TFIIS.h (CMC110C)	2E-1	KIAA0244	9E-12
	FRG1 (Hs)	Hypothetical Protein (CMK102C)	4	rRNA Processing Protein 36 Homolog	3E-8
Second Step Factors					
	Cdc40 (Sc)	Unknown WD-repeat Protein WDS (CMC158C)	5E-15	Rsa4	4E-36
	Cdc40 (Hs)	Unknown WD-repeat Protein WDS (CMC158C)	3E-27	WD Repeat Containing Protein	1E-138
	Prp18 (Sc)	Similar to Pfam GTP-CDC Domain Protein (CMT261C)	6E-1	Sey1	2E-19
	Prp18 (Hs)	ORF515 (CMV241C)	1	Nestin	9E+1
	Slu7 (Sc)	Similar to KRR1-interacting protein (CMT188C)	6E-3	Rgl1	3E-1
	Slu7 (Hs)	Hypothetical Protein, Conserved (CMK086C)	8	Cellular Nucleic Acid-binding Protein	6E-3
	Prp22 (Sc)	†Pre-mRNA Splicing Factor ATP-Dependent RNA Helicase Prp22 (CMG044C)	1E-110	Prp22	2E-135
	Prp22 (Hs)	†Pre-mRNA Splicing Factor ATP-Dependent RNA Helicase Prp22 (CMG044C)	1E-114	DHX8 Protein (Prp22)	1E-147
	Prp16 (Sc)	†Probable pre-mRNA Splicing Factor ATP-Dependent RNA Helicase Prp16 (CMQ385C)	4E-84	Prp16	5E-100
	Prp16 (Hs)	†Probable pre-mRNA Splicing Factor ATP-Dependent RNA Helicase Prp16 (CMQ385C)	2E-88	Prp16p	0

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
Recruited at C Complex					
	Syf2 (Sc)	Hypothetical protein (CMH172C)	3E-1	Kcc4	2E-1
	Syf2 (Hs)	Similar to Vacuolar Sorting Protein SNF7 (CMI044C)	1E-1	Charged Multivesicular Body Protein 4b	2E-6
	CXORF56 (Hs)	Similar to ATPase with Chaperone Activity (CML260C)	6	Cdc43p	4E+2
	DGR14 (Hs)	Clathrin Heavy Chain (CMF103C)	9E-1	Clathrin Heavy Chain 1	0
	C9ORF78 (Hs)	Similar to Mitochondrial Magnesium Transport Mrs2p (CMI181C)	9E-1	MRS2p	7E+1
	Cactin (Hs)	Similar to TATA Element Modulatory Factor (CMR155C)	2E-2	TATA Element Modulatory factor	2E-15
	Nosip (Hs)	Hypothetical Protein, Conserved (CMK221C)	8E-2	Golgin Subfamily A Member 6-like Protein 1-like	2
	Fam50A (Hs)	Probable WD-repeat Membrane Protein (CMG105C)	4	WD repeat-Containing Protein 36	2E-96
	C10ORF 55 (Hs)	Probable Dolichyl-Diphospholigosaccharide--Protein Glycosyltransferase (CMK108C)	4	Dolichyl-Diphospholigo-saccharide--Protein Glycosyltransferase	1E-50
	Leng1 (Hs)	Chloroplast Molecular Chaperone GrpE (CMT423C)	8E-1	GrpE Protein Homolog 2, Mitochondrial Precursor	5E-7
	FAM32A (Hs)	Unknown Heatshock Protein (CMO059C)	2	DnaJ Homolog Subfamily C Member 21	1E-62
	FRA10 AC1 (Hs)	Hypothetical protein, conserved (CMK092C)	6E-1	Zinc Finger Protein 624	8
	FAM50B (Hs)	Similar to Nucleolar Complex Nop3 (CMR049C)	1E-1	Nop3 Protein	9E-10

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	DDX41 (Hs)	ATP Dependent Helicase (CMT137C)	1E-143	Ded1p	6E-180
	PPIL3 (Hs)	Cyclophilin (CMO300C)	5E-29	Cyclophilin D	7E-70
	PPIG (Hs)	Cyclophilin (CMO300C)	5E-37	Cyclophilin D	7E-70
	Pdwd1 (Hs)	Cyclophilin (CMO300C)	4E-32	Cyclophilin D	7E-70
	DHX35 (Hs)	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP43 (CMM048C)	1E-158	Prp43	0
	Wdr83 (Hs)	Unknown WD-repeat Protein WDS (CMC158C)	5E-30	WD Repeat Containing Protein	1E-138
RES Complex					
	Bud13 (Sc)	Probable DNA Repair Protein RAD5 (CMR259C)	8E-1	Rad5	1E-7
	Bud13 (Hs)	Similar to Hedgehog Protein (CMH001C)	1E-2	Sonic Hedgehog Preproprotein	4E-11
	Ist3 (Sc)	Similar to TIA1 Cytotoxic Granule-Associated RNA-Binding Protein (CMS187C)	1E-7	Nam8	2E-10
	Ist3 (Hs)	Similar to Nucleolar RNA-Binding Protein (CMO334C)	5E-11	RNA-Binding Protein 34	2E-34
	Pml1 (Sc)	40S Ribosomal Protein S28 (CMO042C)	1	Ribosomal 40S Subunit S28B	5E-28
	Pml1 (Hs)	TATA-box Binding Protein-associated Factor 11 (CMH222C)	6E-1	Transcription Initiation Factor TFIID Subunit 11	9E-5
EJC/ TREX					
	Yra1 (Sc)	Unknown Transcriptional Coactivator (CMH135C)	1E-3	Hrb1p	2E-6
	Yra1 (Hs)	Unknown Transcriptional Coactivator (CMH135C)	3E-10	Transcriptional Coactivator ALY (Yra1)	1E-8
	Yra1 (At)	Unknown Transcriptional Coactivator (CMH135C)	2E-8	AT5G59950 (Yra1)	8E-10

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Yra1 (Dm)	Unknown Transcriptional Coactivator (CMH135C)	1E-8	THOC4 (Yra1)	8E-9
	THOC1 (Hs)	Hypothetical Protein, Conserved (CMS118C)	5E-1	Palmitoyltransferase ZDHHC2	5E-12
	THOC2 (Hs)	Hypothetical Protein (CMG046C)	1E-9	THOC2 Protein	1E-41
	THOC3 (Hs)	Unknown WD-repeat protein WDS (CMC158C)	3E-15	WDR5 Protein	9E-135
	THOC5 (Hs)	Hypothetical Protein (CMM239C)	1	Pin2/TERF1 Interaction Telomerase	5E-8
	THOC6 (Hs)	Unknown WD-40 Protein (CMO072C)	5E-7	Katanin p80	5E-20
	THOC7 (Hs)	Hypothetical Protein, conserved (CMD144C)	2	Pyroglutamyl-Peptidase 1-like Protein	8E-3
	RNPS1 (Hs)	Similar to TIA1 Cytotoxic granule-associated RNA-binding protein (CMS187C)	9E-5	TIA1p	1E-12
	Fal1 (Sc)	Eukaryotic Translation Initiation Factor eIF-4A (CMK028C)	1E-118	Fal1	2E-168
	Fal1 (Hs)	Eukaryotic Translation Initiation Factor eIF-4A (CMK028C)	1E-139	Eukaryotic Initiation Factor4a (Fal1)	0
	Nam7 (Sc)	Probable Sen1 (CMP054C)	5E-87	Putative RNA/DNA helicase Sen1	7E-115
	Nam7 (Hs)	Probable Sen1 (CMP054C)	2E-88	UPF1 (Nam7)	9E-104
	MAGOH (Hs)	Mitochondrial F-Type ATPase F1 Subunit Alpha, precursor (CMT434C)	2	ATP Synthase Subunit Alpha, Mitochondria Isoform A Precursor	0
	Rbm8A (Hs)	Similar to TIA1 Cytotoxic Granule-Associated RNA-Binding Protein (CMS187C)	4E-4	TIA1 Protein	1E-12
	Mex67 (Sc)	Probable Ran GTPase Activating Protein (CMG035C)	2	Rna1 Protein	1E-5

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Mex67 (Hs)	Similar to Proliferation Related Acidic Leucine Rich Protein PAL31 (CMO169C)	4	Acidic Leucine-rich Nuclear Phosphoprotein 32 Family Member E	8E-21
	SAP18 (Hs)	Similar to Cleavage and Polyadenylation Specificity Factor Subunit (CMS297C)	2	Cleavage and Polyadenylation Specificity Factor	5E-22
	Acinus (Hs)	Hypothetical protein, conserved (CMG016C)	6E-3	Beta-Myosin Heavy Chain	2E-2
SR Proteins					
	SRZ-21 (At)	Similar to hnRNP (CMT598C)	2E-5	ELAV-Like Protein 2	9E-9
	SRZ-22 (At)	Similar to hnRNP (CMT598C)	7E-6	ELAV-Like Protein 2	9E-9
	SR33 (At)	Probable RNA Binding Protein Mrd1p (CMO246C)	1E-9	SRSF2	3E-16
	SR45 (At)	Similar to Nuclear Cap-binding Protein; CBP20 (CMQ282C)	2E-5	Nuclear Cap Binding Protein, Subunit 20	1E-47
	RSP31 (At)	Similar to Splicing Factor RSp31 (CMO009C)	2E-21	Splicing Factor RSP31	3E-24
	SRSF1 (Hs)	Similar to Single-stranded Telomeric DNA-binding Protein GBp1 (hnRNP M) (CMM217C)	4E-40	PABPC4	5E-58
	SRSF2 (Hs)	Similar to SR Family Splicing Factor SC35 (CML202C)	3E-14	SRSF2 (SC35)	3E-12
	SRSF3 (Hs)	Similar to TIA1 Cytotoxic granule-associated RNA-binding protein (CMS187C)	4E-25	TIAp	1E-12
	SRSF4 (Hs)	Similar to Single-stranded Telomeric DNA-binding Protein GBp1 (hnRNP M) (CMM217C)	4E-40	PABPC4	5E-58
	SRSF5 (Hs)	Similar to Gbp1p (CMM078C)	2E-40	PABPC4	5E-58

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	SRSF6 (Hs)	Similar to Single-stranded Telomeric DNA-binding Protein GBp1 (hnRNP M) (CMM217C)	4E-40	PABPC4	5E-58
	SRSF7 (Hs)	Polyadenylate Binding Protein (CMJ286C)	3E-22	PABPC4	3E116
	SRSF8 (Hs)	Eukaryotic Translation Initiation Factor eIF-3 Subunit G (CMH159C)	1E-9	Eukaryotic Translation Initiation Factor eIF-3 Subunit G	2E-48
	SRSF9 (Hs)	Similar to Single-stranded Telomeric DNA-binding Protein GBp1 (hnRNP M) (CMM217C)	7E-39	PABPC4	5E-58
	SRSF10 (Hs)	Similar to SR Family Splicing Factor SC35 (CML202C)	1E-10	SRSF2	3E-12
	SRSF11 (Hs)	Similar to Nuclear Cap-binding Protein; CBP20 (CMQ282C)	2E-5	Nuclear Cap Binding Protein, Subunit 20	3E-36
	SRSF12 (Hs)	Similar to Cleavage Stimulation Factor (CMF108C)	4E-26	Similar to Cleavage Stimulation Factor	2E-49
	SFSWAP (Hs)	Hypothetical Protein (CMS099C)	2E-1	JARID1A Variant Protein	1
	TRA2A (Hs)	Similar to Nucleolar RNA-binding Protein (CMO334C)	8E-7	RNA-Binding protein 34	2E-35
	TRA2B (Hs)	Similar to hnRNP (CMT598C)	2E-5	ELAV-Like Protein 2	9E-9
	ARGLU1 (Hs)	ATP-binding Cassette, Sub-family B (MDR/TAP), Member 6 (CMT066C)	4E+1	ATP-binding Cassette, Sub-family B (MDR/TAP), Member 6	1E-150
	SRRM1 (Hs)	Hypothetical Protein (CMI068C)	2E-2	hCG30082 Protein	2E-8
	Cwc21 (Sc)	Probable DNA Repair Protein Rad5 (CMR259C)	6	Rad5	1E-7
	Cwc21 (Hs)	Similar to Syntaxin Binding Protein (CMI070C)	1E-2	Syntaxin-binding Protien 1	4E-18
hnRNP Proteins					

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	hnRNP PUL1 (Hs)	DEAD/H Box Protein (CMS248C)	2E-24	ATP-Dependent RNA Helicase DDX1	1E-159
	hnRNP PUL2 (Hs)	DEAD/H Box Protein (CMS248C)	2E-23	ATP-Dependent RNA Helicase DDX1	1E-159
	FUS (Hs)	Probable RNA Binding Protein Mrd1p (CMO246C)	6E-4	Mrd1 Protein	8E-79
	hnRNP A0 (Hs)	Probable hnRNP A (CMR392C)	2E-17	Daz-associated Protein 1	4E-31
	hnRNP A1 (Hs)	Probable hnRNP A (CMR392C)	3E-23	Daz-associated Protein 1	4E-31
	hnRNP A3 (Hs)	Probable hnRNP A (CMR392C)	2E-22	Daz-associated Protein 1	4E-31
	PCBP1 (Hs)	Polyribonucleotide Nucleotidyltransferase (CMH146C)	7E-3	Polyribonucleotide Nucleotidyltrans-ferase 1	4E-121
	PCBP2 (Hs)	Polyribonucleotide Nucleotidyltransferase (CMH146C)	1E-2	Polyribonucleotide Nucleotidyltrans-ferase 1	4E-121
	hnRNP AB (Hs)	Probable hnRNP A (CMR392C)	3E-23	Daz-associated Protein 1	4E-31
	hnRNP C (Hs)	Polyadenylate Binding Protein (CMJ286C)	2E-23	PABPC4	3E-116
	hnRNP C (Mm)	Polyadenylate Binding Protein (CMJ286C)	1E-21	Pabpc4	2E-111
	hnRNP D (Hs)	Probable hnRNP A (CMR392C)	9E-9	Daz-associated Protein 1	4E-31
	hnRNP F (Hs)	Similar to Heterogeneous Nuclear Ribonucleoprotein H3, Isoform a (CMF163C)	4E-22	hnRNP H3	3E-15
	hnRNP H1 (Hs)	Similar to Heterogeneous Nuclear Ribonucleoprotein H3, Isoform a (CMF163C)	2E-13	hnRNP H3	3E-15
	hnRNP H2 (Hs)	Similar to Heterogeneous Nuclear Ribonucleoprotein H3, Isoform a (CMF163C)	5E-6	hnRNP H3	3E-15
	hnRNP H3 (Hs)	Similar to Heterogeneous Nuclear Ribonucleoprotein H3, Isoform a (CMF163C)	3E-14	hnRNP H3	3E-15
	hnRNP K (Hs)	Phosphatidylinositol 4-kinase (CMI125C)	4	Phosphatidylinositol 4-kinase	5E-83

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	hnRNP L (Hs)	Similar to splicing factor RSp31 (CMO009C)	1E-1	RNA-binding protein 4	5E-7
	hnRNP M (Hs)	Similar to Gbp1 Protein (CMM078C)	7E-52	PABPC4	5E-58
	hnRNP R (Hs)	Polyadenylate Binding Protein (CMJ286C)	3E-9	Polyadenylate Binding Protein-1	8E-69
	hnRNP U (Hs)	DEAD/H Box Protein (CMS248C)	2E-12	ATP-Dependent RNA helicase DDX1	1E-159
	hnRNP LL (Hs)	Similar to splicing factor RSp31 (CMO009C)	7	RNA-binding protein 4	5E-7
	SYNCRIP (Hs)	Probable hnRNP A (CMR392C)	7E-11	Daz-associated Protein 1	4E-31
	RBMXL2 (Hs)	Probable hnRNP A (CMR392C)	1E-12	Daz-associated Protein 1	4E-31
	RBM X (Hs)	Probable hnRNP A (CMR392C)	3E-13	Daz-associated Protein 1	4E-31
	RALY (Hs)	Polyadenylate Binding Protein (CMJ286C)	1E-21	PABPC4	3E-116
	RALY L (Hs)	Polyadenylate Binding Protein (CMJ286C)	6E-23	PABPC4	3E-116
	hnRNP CL1 (Hs)	Polyadenylate Binding Protein (CMJ286C)	2E-22	PABPC4	3E-116
	hnRNP A2B1 (Hs)	Probable hnRNP A (CMR392C)	1E-25	Daz-associated Protein 1	4E-31
	hnRNPQ (Hs)	Polyadenylate Binding Protein (CMJ286C)	3E-11	Polyadenylate Binding Protein-1	8E-69
	PTBP1 (Hs)	Unknown Transcriptional Coactivator (CMH135C)	4E-2	Transcriptional Coactivator ALY	1E-8
	PTBP2 (Hs)	Probable RNA Binding Protein Mrd1p (CMO246C)	7E-3	Mrd1 Protein	8E-79
	UBP1 (At)	Polyadenylate-binding Protein (CMJ286C)	3E-12	PAB1 Protein	6E-71
	UBA (At)	Probable hnRNP A (CMR392C)	2E-8	Daz-associated Protein 1	4E-31
	RBP45 (At)	Polyadenylate-binding protein (CMJ286C)	8E-19	PAB1 Protein	6E-71
	RBP47 (At)	Polyadenylate-binding Protein (CMJ286C)	8E-19	PAB1 Protein	6E-71

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Ath1 (At)	Unknown Homeobox protein (CMR176C)	2E-15	BEL-1 like Homeobox	1E-19
Misc.					
	Ntr2 (Sc)	MGDG Synthase (CMI271C)	5	Trm5	8E-1
	Ntc20 (Sc)	4-amino-4-Deoxychorismate (ADC) Synthase (CMO324C)	3	Abz1	9E-65
	Cwc23 (Sc)	DnaJ (Hsp40) Homolog, Subfamily A (CML030C)	2E-6	Ydj1	2E-44
	Cwc23 (Hs)	DnaJ (Hsp40) Homolog, Subfamily A (CML030C)	4E-9	DnaJ Subfamily B Member	5E-74
	Dbr1 (Sc)	RNA Lariat Debranching Enzyme (CMK205C)	7E-39	Dbr1	3E-31
	Dbr1 (Hs)	RNA Lariat Debranching Enzyme (CMK205C)	2E-59	Lariat Debranching Protein (Dbr1)	2E-67
	YBX1 (Hs)	Unknown DUF21 Containing Protein (CMC089C)	5E-1	Metal Transporter CNNM2	5E-33
	SRRT (Hs)	Hypothetical Protein (CMO350C)	5E-1	Arsenite Resistant Protein ASR2	3
	SRRT (Mm)	Hypothetical Protein (CMO350C)	5E-1	Serrate RNA Effector Molecule Homolog	3E+1
	SRRT (Dr)	Hypothetical Protein (CMO350C)	2	Upstream-binding Protein 1	6
	SRRT (Xl)	Hypothetical Protein (CML318C)	8E-1	Protein Kinase Bub1	2
	RBM7 (Hs)	Similar to Cleavage Stimulation Factor (CMF108C)	5E-8	Similar to cleavage stimulation factor, 3' pre-RNA	2E-34
	ELAVL (Hs)	Polyadenylate-binding Protein (CMJ286)	1E-19	PAB1 Protein	6E-71
	PABP1 (Hs)	Polyadenylate-binding Protein (CMJ286)	3E-76	PAB1 Protein	6E-71
	ILF2 (Hs)	Phosphoserine Phosphatase (PSP) (CMI086C)	1	Phosphoserine Phosphatase Variant	8E-17
	ZC3H18 (Hs)	Similar to Proliferation Related Acidic Leucine Rich Protein PAL31 (CMO169C)	1E-1	Acidic Leucine-rich Nuclear Phosphoprotein 32 Family Member E	8E-21

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	DDX3X (Hs)	ATP Dependent RNA Helicase (CMT173C)	1E-140	DDX3X	9E-177
	Bub3 (Hs)	Probable mRNA-associated protein MRNP 41, RAE1 Homolog (CMI077C)	9E-50	RNA Export 1-like Protein	9E-81
	S164 (Hs)	Similar to Cleavage Stimulation Factor (CMF108C)	5E-6	Similar to cleavage stimulation factor, 3' pre-RNA	2E-34
	RBM23 (Hs)	Probable hnRNP A (CMR392C)	7E-11	Daz-associated Protein 1	4E-31
	RBM39 (Hs)	Probable hnRNP A (CMR392C)	2E-12	Daz-associated Protein 1	4E-31
	p72 (Hs)	p68 RNA Helicase (CMR479C)	1E-139	DDX5 RNA Helicase	0
	Dbp2 (Sc)	p68 RNA Helicase (CMR479C)	1E-135	Dbp2	4E-129
	Dbp2 (Hs)	p68 RNA Helicase (CMR479C)	1E-144	DDX5 RNA Helicase (Dbp2)	0
	DnaJC8 (Hs)	Unknown Heatshock Protein (CMO059C)	4E-8	DnaJC21	9E-63
	KIAA1967 (Hs)	Similar to Uroporphyrin-III C Methyltransferase (CMB055C)	3	NAD kinase	7
	NFAR (Hs)	Similar to dsRNA-Specific Ribonuclease (CMS219C)	3	Putative Ribonuclease III	7E-13
	ZNF207 (Hs)	Unknown Zinc Finger Protein (CMP283C)	7E-1	Znf48	2E-20
	PPIL4 (Hs)	Probable Cyclophilin B; CYPB (CMR272C)	1E-16	Cyclophiln B Protein	5E-64
	HCNGP (Hs)	Hypothetical Protein (CMK134C)	2	C17orf68p	2
	DHX9 (Hs)	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP43 (CMM048C)	1E-66	Prp43	0
	PPP1CA (Hs)	Serine/Threonine Protein Phosphatase PP1 Catalytic Subunit (CME079C)	1E-146	PP1CA Protein	0
	GCFC (Hs)	Hypothetical Protein, Conserved (CMQ312C)	1E-1	RNA 3'-terminal Phosphate Cyclase-like Protein	2E-7

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	BAG2 (Hs)	Hypothetical Protein, Conserved (CMK068C)	7E-1	5'-3' Exoribonuclease 2	1E-24
	RBM42 (Hs)	Polyadenylate-binding Protein (CMJ286C)	7E-11	PAB1 Protein	6E-71
	CIRP (Hs)	Probable hnRNP A (CMR392C)	2E-13	Daz-associated Protein 1	4E-31
	NIPP1 (Hs)	Hypothetical Protein (CMT135C)	6	UDP-N-acetyl-alpha-D-galactosamine Polypeptide N-Acetylgalactosamiyl-transferase-like 5	5
	SMN (Hs)	Hypothetical Protein (CMO120C)	5	Olfactory receptor 51G1	4
	PICIn (Hs)	26S Proteasome Regulatory Subunit RPN3 (CML133C)	8E-1	26S proteasome non-ATPase regulatory subunit3	3E-76
	Mep50 (Hs)	Probable Polyadenlyation Factor I Subunit 2 (CMQ221C)	3E-15	pre-mRNA 3' end Processing Protein WDR33	2E-133
	PRMT5 (Hs)	26S Proteasome Regulatory Subunit RPN3 (CML133C)	8E-1	26S Proteasome non-ATPase Regulatory Subunit3	3E-76
	SEC31L2 (Hs)	Vesicle Coat Complex Subunit Sec31 (CMR203C)	3E-48	SEC31A Protein	6E-34
	CCDC55 (Hs)	Hypothetical Protein, Conserved (CMK157C)	5E-2	Histone-lysine N-methyltransferase 2C	4E-11
	RBBP6 (Hs)	Unknown Zinc-Finger protein Mpe1p (CMQ079C)	1E-25	RBBP6	1E-29
	AGGF1 (Hs)	Similar to transcriptional activator FHA1 (CMH027C)	5E-2	Forkhead box protein M1 isoform 1	8E-4
	CUGBP1 (Hs)	Probable heterogeneous nuclear RNP protein A (CMR392C)	3E-10	Daz-associated protein 1	4E-31
	Fox2 (Hs)	Probable RNA Binding Protein Mrd1P (CMO246C)	8E-6	Mrd1P	8E-79
	Quaking (Hs)	Similar to QKI (CMA075C)	4E-35	Quaking Homolog	9E-41

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Sam-68 (Hs)	Similar to QKI (CMA075C)	9E-19	Quaking Homolog	9E-41
	Sim-2 (Hs)	Hypothetical protein (CMS184C)	8E-1	Znf664 Protein	9
	DDX57 (Hs)	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP43 (CMM048C)	1E-43	Prp43	0
	RACK1 (Hs)	Receptor for Activated Protein Kinase C (CMI283C)	1E-123	Guanine Nucleotide-Binding Protein Subunit beta-2-like 1 (RACK1)	7E-150
	Matrin3 (Hs)	Similar to transcription initiation protein SPT5 (CMD151C)	1E-2	SPT5p	1E -30
	DBPA (Hs)	Hypothetical protein (CME138C)	8	hCG2010820	5
	TOE1 (Hs)	Hypothetical Protein, Conserved (CMK240C)	7E-15	TOE1 Protein	7E-20
	RBM4 (Hs)	Polyadenylate-binding Protein (CMJ286C)	1E-7	PAB1 Protein	6E-71
	JUP (Hs)	Hypothetical Protein (CMP113C)	2E-2	Chondroitin sulfate N-acetylgalactosaminyl-transferase 2	2
	CCDC130 (Hs)	Similar to GTPase Activating Protein (CMJ230C)	1E-2	Gap Domain of Znf289	2E-33
	NKAP (Hs)	Hypothetical Protein (CMP252C)	3E-1	Myelin transcription factor 1-like protein isoform X1	1
	TTC14 (Hs)	Photosystem I Assembly Related protein (CMV197C)	4E-04	UDP-N-acetylglucosamine--Peptide N-acetylgucosaminyl-transferase 110kDa	1E-05
	CDK10 (Hs)	Cyclin Dependent Kinase, C-type (CMQ195C)	2E-66	CDK12 Protein	6E-9
	TFIP11 (Hs)	Hypothetical protein (CMN293C)	2	Pin2/TERF1 Interaction Telomerase	5E-8
	ELG (Hs)	Similar to Ubiquitin-specific Processing Protease Ubp2p (CMC070C)	7E-2	USP28p	8E-21

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Pinin (Hs)	Kinesin Related Protein (CMO070C)	8E-2	Kinesin-like Protein KIF15	1E-126
	NRIP2 (Hs)	Similar to putative v-snare binding protein (CMR127C)	7E-2	DDI1 Homolog	4E-40
	UBL5 (Hs)	Similar to Peptidyl-prolyl cis-trans Isomerase ESSP1 (CMH210C)	1E-16	Pin1 Peptidyl-prolyl cis-trans Isomerase	3E-17
	Rts2 (Sc)	Hypothetical Protein, Conserved (CMG137C)	8E-9	Rts2	9E-8
	Rts2 (Hs)	Hypothetical Protein, Conserved (CMG137C)	4E-16	DNA-RNA Binding Protein Kin17 (Rts2)	9E-17
	SNRNP27 (Hs)	Hypothetical Protein (CMS203C)	7E+1	Dap1-alpha	7
	HSPB1 (Hs)	Hypothetical Protein (CMP273C)	1	CCDC68	4E-1
	PPIL2 (Hs)	Cyclophilin (CMO300C)	1E-47	Cyclophilin D	7E-70
	ERH (Hs)	Similar to Enhancer of Rudimentaryp1 homolog (CMR206C)	2E-22	Enhancer of Rudimentary Homolog	8E-26
	Mtr4 (Sc)	Probable Nuclear Exosomal RNA Helicase MTR4 (CMA072C)	1E-178	Mtr4	0
	Mtr4 (Hs)	Probable Nuclear Exosomal RNA Helicase MTR4 (CMA072C)	0	SKIV2L2 (Mtr4)	0
	DHX36 (Hs)	†Similar to ATP-dependent RNA Helicase A (CMC171C)	5E-30	DHX36 Protein	3E-42
	Rvb2 (Sc)	RuvB-like DNA/RNA Helicase Reptin (CMT427C)	1E-135	Rvb2	0
	Rvb2 (Hs)	RuvB-like DNA/RNA Helicase Reptin (CMT427C)	1E-151	RuvB-like 2 (Rvb2)	0
	Tef1 (Sc)	eEF-1a (CMH226C)	0	Tef1	0
	Tef1 (Hs)	eEF-1a (CMH226C)	0	eEF-1a (Tef1)	0
	Rpg1 (Sc)	Eukaryotic Translation Factor eIF-3 A (CMH060C)	3E-27	Rpg1	2E-30

Particle or Complex	Query Protein (Species)*	Best Hit (Accession Number)	E-value	Reciprocal Best Hit	RBH E-value
	Rpg1 (Hs)	Eukaryotic Translation Factor eIF-3 A (CMH060C)	2E-39	EIF3A (Rpg1)	7E-40
	Prt1 (Sc)	Eukaryotic Translation Factor eIF-3 B (CMK285C)	2E-67	Prt1	3E-74
	Prt1 (Hs)	Eukaryotic Translation Factor eIF-3 B (CMK285C)	7E-98	EIF3B (Prt1)	5E-111
	RPSA (Hs)	40S Ribosomal Protein SA (CMT410C)	1E-66	RPSA	9E-89
	TUBA1B (Hs)	Alpha Tubulin (CMT504C)	0	Alpha Tubulin	0
	Tub2 (Sc)	Beta Tubulin (CMN263C)	0	Tub2	0
	Tub2 (Hs)	Beta Tubulin (CMN263C)	0	TUBB4B (Tub2)	0
	SRPK1 (Hs)	Serine/arginine-rich Protein Specific Kinase (CMK182C)	9E-56	SRPK1 Protein	1E-61
	SRPK2 (Hs)	Serine/arginine-rich Protein Specific Kinase (CMK182C)	9E-52	SRPK1 Protein	1E-61
	MGC2039 8 (Hs)	Similar to Splicing Factor 3a subunit 2, Partial (CMN095C)	7E-3	Splicing factor 3a, subunit 2, 66kDa	7E-10
	C19orf43 (Hs)	Acetylornithine Transaminase (CMQ064C)	3	Ornithine Aminotransferase Mutant Y85i	2E-72
	CDK11A (Hs)	Cyclin Dependent Kinase, A-type (CME119C)	5E-75	CDK3 Protein	2E-135

*species abreviations, common name and eukaryotic lineage

Ag	*Ashbya gossypii*	Yeast	Fungi
An	*Aspergillus nidulans*	None	Fungi
At	*Arabidopsis thaliana*	Thale cress	Land plants
Bd	*Bactrocera dorsalis*	Oriental fruit fly	Arthropods
Bt	*Bos Taurus*	Bovine	Ungulates
Ca	*Candida albicans*	Yeast	Fungi
Ce	*Caenorhabditis elegans*	Nematode worm	Nematoda
Co	*Candida orthopsilosis*	Yeast	Fungi
Cr	*Chlamydomonas reinhardtii*	Green alga	Unicellular alga
Dd	*Dictyostelium discoideum*	Slime mold	Myxogastrids
Dh	*Debrayomyces hansenii*	Yeast	Fungi
Dm	*Drosophila melanogaster*	Fruit fly	Arthropods
Dr	*Danio rerio*	Zebrafish	Teleostei
Hs	*Homo sapiens*	Human	Primates
Mm	*Mus musculus*	House mouse	Rodents

Mn	*Morus notabilis*	Mulberries	Land plants
Mo	*Magnaporthe oryzae*	Rice blast fungus	Fungi
Mt	*Medicago truncatula*	Barrel medic	Land plants
Pb	*Psudeozyma brasiliensis*	Yeast	Fungi
Pd	*Penicillium digitatum*	Green mold	Fungi
Rn	*Rattus norvegicus*	Brown rat	Rodents
Sc	*Saccharomyces cerevisiae*	Baker's yeast	Fungi
Sp	*Schizosaccharomyces pombe*	Fission yeast	Fungi
Tb	*Trypanosoma brucei*	Sleeping sickness parasite	Euglenozoa
Xl	*Xenopus laevis*	African frog	Amphibians

†proteins that were not the best hit in the forward BLAST search; see Table S4

Table S4. Protein search results for non-RBH sequences

Query Protein (Species) *	Top Hit (Accession Number)	Top Hit E-value	Target Protein Rank (Accession Number)	Target Protein E-value	RBH of Top Hit (E-value)	RBH of Target Protein (E-value)
Prp21 (Hs)	Ubiquitin with Short C-terminal Extension (CMK296C)	7E-27	Sixth (CMJ300C)	3E-07	Ubiqitin-60S Ribosomal Protein (2E-32)	Prp21 (Second) (1E-12)
Prp21 (Mm)	Ubiquitin with Short C-terminal Extension (CMK296C)	7E-27	Seventh (CMJ300C)	3E-07	Ubiqitin-60S Ribosomal Protein (1E-32)	Prp21 (9E-14)
Prp21 (Dm)	Ubiquitin with Short C-terminal Extension (CMK296C)	3E-25	Fifth (CMJ300C)	9E-09	Ribosomal Protein L40 (4E-33)	Prp21 (7E-13)
Prp21 (Ce)	Ubiquitin with Short C-terminal Extension (CMK296C)	4E-28	Seventh (CMJ300C)	2E-07	UBQ-2 Protein (3E-32)	Prp21 (2E-11)
Hsh49 (Sc)	Polyadenylate Binding Protein (CMJ286C)	6E-17	Third (CME063C)	6E-10	PAB1 (1E-69)	Hsh49 (2E-9)
Hsh49 (Hs)	Polyadenylate Binding Protein (CMJ286C)	3E-29	Third (CME063C)	6E-10	PAB Protein (2E-88)	Hsh49 (1E-13)
Prp5 (Sc)	p68 RNA Helicase (CMR479C)	3E-82	Fifth (CMR433C)	3E-47	Dbp2 (5E-178)	Prp5 (1E-49)
Prp5 (Sp)	p68 RNA Helicase (CMR479C)	1E-118	Third (CMR433C)	8E-66	Dbp2 (0)	Prp5 (4E-66)
Prp5 (Ce)	p68 RNA Helicase (CMR479C)	4E-115	Fourth (CMR433C)	4E-54	F58E10.3 (1E-153)	Prp5 (3E-62)
Mud2 (Sp)	Probable RNA Binding Protein Mrd1p (CMO246C)	2E-08	Thirteenth (CMS438C)	3E-02	Mrd1 Protein (1E-67)	Mud2 (Fourth) (1E-23)
Mud2 (Ce)	Similar to Cleavage Stimulation Factor (CMF108C)	3E-08	Eighth (CMS438C)	1E-03	CPF-2 Protein (2E-28)	Mud2 (2E-32)
Snu114 (Sc)	Eukaryotic Translation Elongation Factor 2 (CMS428C)	5E-90	Second (CMK208C)	7E-40	Elongation Factor 2 (0)	Snu114 (7E-42)
Snu114 (Hs)	Eukaryotic Translation Elongation Factor 2 (CMS428C)	0E+00	Second (CMK208C)	2E-43	Elongation Factor 2 (0)	Snu114 (1E-42)
Prp22 (Sc)	Prp43 Protein (CMM048C)	1E-149	Second (CMG044C)	1E-110	Prp43 (0)	Prp22 (2E-135)

Query Protein (Species) *	Top Hit (Accession Number)	Top Hit E-value	Target Protein Rank (Accession Number)	Target Protein E-value	RBH of Top Hit (E-value)	RBH of Target Protein (E-value)
Prp22 (Hs)	Prp43 Protein (CMM048C)	1E-158	Second (CMG044C)	1E-114	Prp43 (0)	Prp22 (1E-147)
Prp16 (Sc)	Prp43 Protein (CMM048C)	1E-138	Fourth (CMQ385C)	4E-84	Prp43 (0)	Prp16 (5E-100)
Prp16 (Hs)	Prp43 Protein (CMM048C)	1E-146	Fourth (CMQ385C)	2E-88	Prp43 (0)	Prp16 (0)
Lsm7 (Sc)	Lsm3 Protein (CMT262C)	6E-07	Second (CMP206C)	1E-06	Lsm3 (6E-7)	Lsm7 (1E-11)
Sm B/B' (Hs)	Lsm3 Protein (CMT262C)	2E-05	Second (CMK022C)	4E-05	Lsm3 (3E-17)	Sm B/B' (3E-35)
DHX36 (Hs)	Prp22 Protein (CMG044C)	4E-60	Fourth (CMC171C)	5E-30	DHX8 (1E-147)	DHX36 (3E-42)

*species abbreviation (Sc) *Saccharomyces cerevisiae*, (Hs) *Homo sapiens*, (Dm) *Drosophila melanogaster*, (Sp) *Schizosaccharomyces pombe*, (Ce) *Caenorhabditis elegans*, (Mm) *Mus musculus*

TABLE S5 - *C.merolae* Splicing Proteins Present and Not Found

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
U1 not found	Mud1 (U1-A), Snp1 (U1-70K), Yhc1 (U1-C), Prp39, Snu71, Prp40, Prp42, Nam8, Snu56, Luc7, Urn1, Npl3					
Sm	SmB/B'	CMK022C	3E-35	32	•	•
	SmD1	CMF084C	3E-15	32	•	•
	SmD2	CMN302C	1E-17	38	•	•
	SmD3	CMM065C	1E-14	37	•	•
	SmE§	CMM109C CMH215C	2E-15	36	•	
	SmF	CMQ171C	2E-20	50	•	•
	SmG	CMO342C	2E-11	39	•	•
U2	Prp9 (SF3a3)	CMQ406C	3E-45	22	•	•
	Prp11 (SF3a2) §	CMH102C CMN095C	5E-11	51	•	
	Prp21 (SF3a1)	CMJ300C	9E-14	34	•	•
	Hsh155 (SF3b1)	CMB002C	1E-73	25	•	•

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
	Cus1 (SF3b2)	CMT357C	4E-30	31	•	•
	Rse1 (SF3b3)	CML103C	2E-24	27	•	•
	Hsh49 (SF3b4)	CME063C	1E-13	42	•	
	Rds3 (SF3b6)	CMS014C	3E-37	31	•	•
	Prp5 (hPRP5)	CMR433C	4E-66	32	•	
U2 related	Prp43 (hPRP43)	CMM048C	0	48	•	
	Mud2 (U2AF65)	CMS438C	7E-63	22	•	
U2 not found	Lea1, Msl1, Ysf3, SF3b14a, Cus2, U2AF1, Puf60, Smndc1, Rbm17, U2SURP, CHERP, DDX42					
U5	Prp8 (220K)	CMH168C	0	34	•	•
	Brr2 (200K)	CML192C	1E-119	35	•	•
	Snu114 (116K)	CMK208C	1E-42	35	•	•
	Dib1 (15K) §	CMN033C CMS018C	2E-78	34	•	

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
U5 not found	Prp28, Lin1, snRNP40, Aar2, Prp6					
U4/U6	Prp3 (90K)	CMT170C	7E-11	21	•	
	Snu13 (15.5K)	CMP335C	5E-52	57	•	
U4/U6 not found	Prp31, Prp4, PPIH					
tri-snRNP not found	Snu66, Spp381, Sad1					
U6¶	Lsm1	CMT394C	1E-19	44		
	Lsm2	CMB130C	6E-23	41	•	
	Lsm3	CMT262C	3E-20	44	•	
	Lsm4 §	CMG061C CMT545C	1E-27	44	•	
	Lsm5	CMP159C	5E-18	38	•	
	Lsm6	CMP138C	1E-28	31	•	
	Lsm7	CMP206C	1E-11	41	•	

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
U6 not found	Prp24, Lsm8					
Cap Binding	Sto1 (CBP80)	CMJ189C	5E-92	18		•
	Cbc2 (CBP20)	CMQ282C	1E-43	53		•
Complex A	Msl5/BBP/SF1 (RBM10)	CMI292C	2E-39	34	•	
	Sub2 (hUAP56)	CME073C	0	59	•	
Complex A not found	Prp40, Snu71,Thrap3, Ccar1, SUGP1, RBM5, RBM10, YBX1					
NTC	Cef1 (CDC5L)	CMR098C	4E-33	34	•	•
	Prp46 (PRL1)	CMR305C	2E-38	28	•	•
	Bud31 (G10)	CMG014C	4E-19	35		•
NTC not found	Prp19, Cwc15, Snt309, CTNNBL1, WBP11, PQBP1, Clf1, Prp45, Isy1, Syf1, Ecm2, Cwc2, PPIL1, AQR, PPIE					
Complex B	Prp38 (hPRP38)	CMJ144C	1E-57	27	•	
Complex B not found	Spp382, Snu23, Smu1, MFAP1, IK, WBP4, PRPF4B					

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
Complex B^{act}	Yju2 (CCDC94)	CMN267C	3E-5	25	•	
Complex B^{act} not found	Cwc27, Prp2, Cwc22, Spp2, Cwc24, Cwc25, Znf830, CCDC12, PRCC, GPATCH1, FRG1, PPIL2					
Second Step	Prp16 (hPRP16)	CMQ385C	0	35	•	•
	Prp22 (hPRP22)	CMG044C	1E-147	46	•	•
Second Step not found	Prp18, Slu7, Cdc40/Prp17					
Complex C not found	Syf2, DDX41, CXorf56, DGCR14, C9orf78, PPIL3, PPWD1, DHX35, Cactin, Nosip, WDR83, FAM50A, PPIG, C1orf55, LENG1, FAM32A, FRA10AC1, FAM50B, CDK10, hnRNP C					
RES not found	Bud13, Ist3, Pml1					
EJC/TREX	Fal1 (eIF4A3)	CMK028C	0	78	•	•
	Yra1 (Aly/THOC4)	CMH135C	3E-10	37		
	n/a (THOC2)	CMG046C	1E-41	21		
EJC/ TREX not found	MAGOH, RBM8A, SAP18, Acinus, THOC1, THOC3, THOC5, THOC6, THOC7, RNPS1, Mex67, SAP18					
SR	Rsp31[#]	CMO009C	3E-24	34		

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
	n/a (SRSF2)	CML202C	3E-12	30		
SR proteins not found	SRZ-21, SRZ-22, SR33, SR45, SRSF1, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF8, SRSF9, SRSF10, SRSF11, SRSF12, SFSWAP, TRA2A, TRA2B, ARGLU1					
SR related not found	SRRM1, Cwc21					
hnRNP	n/a (hnRNP H3)	CMF163C	3E-15	38		
hnRNP not found	hnRNP PUL1, hnRNP PUL2, FUS, hnRNPA0, hnRNP A1, hnRNP A3, PCBP1, PCBP2, hnRNP AB, hnRNP C, hnRNP D, hnRNP F, hnRNP H1, hnRNP H2, hnRNP K, hnRNP L, hnRNP M, hnRNP R, hnRNP U, hnRNP LL, SYNCRIP, RBMXL2, RBMX, RALY, RALY L, hnRNP CL1, hnRNPA2B1, hnRNP Q, PTBP1, PTBP2, UBP1, UBA, RBP45, RBP47, Ath1					
Miscellaneous	Dbr1 (hDBR1)	CMK205C	2E-67	41		
	n/a (SRPK1)	CMK182C	1E-61	41		
	n/a (DDX3X)	CMT173C	1E-177	59		
	Dbp2 (p68/DDX5)	CMR479C	0	54		
	n/a (PABP1)	CMJ286C	3E-76	43		
	n/a (DHX36/RHAU)	CMC171C	3E-42	34		
	n/a (PPP1CA)	CME079C	0	75		

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
	n/a (RBBP6/PACT)	CMQ079C	1E-29	32		
	n/a (Quaking)	CMA075C	9E-41	55		
	n/a (RACK1)	CMI283C	1E-150	27		
	n/a (TOE1)	CMK240C	7E-20	23		
	Rts2 (HsKin17)	CMG137C	9E-17	33	•	
	n/a (ERH)	CMR260C	2E-22	42		
	Mtr4 (SKIV2L2)	CMA072C	0	39	•	
	Rvb2 (TIP48)	CMT427C	0	56	•	
	Tef1 (eEF1A)	CMH226C	0	75		
	Rpg1 (eIF3A)	CMH060C	7E-40	28	•	
	Prt1 (EIF3B)	CMK285C	5E-111	31		
	n/a (RPSA)	CMT410C	9E-89	56		

Particle or Step	Name *S.cerevisiae* (*H.sapiens*)	*C. merolae* homologue	E value*	identity (%)	Essential †	1M salt core‡
	n/a (TUBA1B)	CMT504C	0	77		
	Tub2 (TUBB4B)	CMN263C	0	72		
Miscellaneous not found	Ntr2, Ntc20, Cwc23, SRRT, RBM7, ELAVL, ILF2, ZC3H18, Bub3, S164, Rbm23, RBM39, p72, DNAJC8, KIAA1967, NFAR, ZNF207, PPIL4, HCNGP, DHX9, GCFC, BAG2, RBM42, CIRP, NIP1, SMN, PICln, Mep50, PRMT5, SEC31L2, CCDC55, AGGF1, CUGBP1, Fox2, Sam-68, Sim-2, DDX57, Matrin3, DBPA, RBM4, JUP, CCDC130, NKAP, TTC14, TFIP11, ELG, Pinin, NRIP2, UBL5, SNRNP27, HSPB1, SRPK2, MGC20398, C19orf43, CDK11A					
Minor Spliceosome not found	20K, 25K, 31K, 35K, 48K, 59K, 65K, ZRSR2					

* Best E value (among all species, forward and reverse BLASTs) and percent identity for best BLAST
† Essential for viability in *S. cerevisiae, S. pombe, or M. musculus* according to the Database of Essential Genes (1)
‡ Indicates human orthologue was a C complex protein detected by mass spectrometry after treatment with 1 M salt (2)
§ These proteins are encoded by more than one gene: CMM109C and CMH215C encode SmE proteins that differ by a single amino acid. CMH102C and CMN095C encode Prp11 proteins that differ by a single amino acid. CMN033C and CMS018C encode Dib1 proteins that differ by 4 amino acids. CMG061C and CMT545C encode identical LSm4 proteins.
¶ It is unclear whether the CmLSm proteins associate with CmU6 (see Discussion)
# Rsp31 is an *A. thaliana* SR protein

1. Zhang R, Ou H-Y, Zhang C-T (2004) DEG: A database of essential genes. Nucleic Acids Res 32(Database issue):D271–D272.

2. Bessonov S, Anokhina M, Will CL, Urlaub H, Lu_hrmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. Nature 452(7189):846–850.

Table S6. Sequences from RNA immunoprecipitation

chromosome/ location	FPKM*	sequence
1/44485-44586	455	AAGCACCUGAACGGACGGACAGACAGACAGAUAGAUAGAGAGGCAGA GAGGCAGACAGACAGACAGACAGAGAGGCAGAGAGGCAGACAGACAG ACAGAGAG
1/227940-228014	446	GGCACCCUACUCAUUCGGAAAACUCUGCAUCAUAAAAGCUGUUUAGU GCAUUUCGUGCAAACGAGGCGUCUCUU
1/395766-395922	422	UUCCCAGGCGGUACCAACGACAGGUCCGUGUUCUUUUUUCCUAUCCG AUGGAAAGUCUGGAUGUUAUAGAGGCGAACAUUGGCGAGAAUGCUCC GGUUCCACGUUGGAAACGGCGCCAGCAGCGCCAACUCGCCGAGCAGC UCGCCAGAGCCGACUC
2/54849-55258	566	UCCUUCUCUCUUUCUCUCUCUACAUUCUCCUGUCCGCGCCCUUGGCC AAAACUCUGAAAAGAUAUCGCUCGCGUCACUCGUGUUCCUAUGUAGG GUAUAUAGAUACGCACUCACAGACGCCUCCGGGCCUUUCCUUGACGG GCCGCGUGCUACCGGAUUCAGCUUCUCCUCGUACACUACGAGCCUGU CAUCGUUUUCUGCUGCUGCCUCCGGCCCUCCGUUCGUUGCGGCCAA GGUCCCUGUCAUGGGCUUCCUUCCCCGUCGCUCGAACACCCCGAGU GAACCCACUUUCAGAGCUUCGCCGAAACAACUAGACCUCUCAGACGG GGGUGGUGUGUGUGGGGCUUUAAGAUACUUGCUCUUUAGUACACAG UCUGUGGCUAGUUUGUUUUGUUGUUUUUUUCUUUUGU
4/444044-444215	686	CCCCUCCCUCCAAGUUACAUUACCUAUGCUGAUUCUCGGAUAUUCUG GUCGUUUCAUACGUCCUAUCGUGUGUGCCUAGUACCUGCCUUUCCAU AUCCGCCUUUGGACCUUAAAUGCUGGAGCAGAUAGUUCAUCGAGCUA GAGCAGCUUUUCUAUACCCCAGAGGGCGUCG
4/462251-462389	773	ACAUAACGGUCUCCAGGACGUACUUGAUGUGUAAACCGCCCCUCUCU CUUCGCUUUCUGAAUCUCCGCCAUUAGUUCCGGAACACGAUUCUUGU UUUUCGUGAGCUCCAUUCCCAAUAGAGAGCGCUACCGGUCAACGC
4/486958-487010	671	UCGUUUGAGACCUAUCAAAGGAACAGAUCGGUGAUACGAAGAGAAGA GAAAGC
5/134499-134639	558	CCCUCUUUCCAAACUGCAGUGUGCUUUUCUGACCAUCUCAUGCGUCU CCUAUCGAAUCUGCUCUUUCAUCUCAUCAUCAAGAAAAUUCGCGGUC AAAGUCACCGGAAAAAGAGAAAACUCGCACUGGAGCUACCGCUCCAU
5/222357-222561 (**CmU4**)	36095	GAAAGUACACAUAUUUCUUACUUGGUGUUUUCCAUACUUGCGCAGUG UCGGUUGUUGCCCAGAUGAGGUUCUCCGAUGGGUAACGGUAUGCUG AACACAAACAAUUUACCACAGUGGACUUUUUAACCGGUCCGUCUGUG ACGGGCAUGCCUCAAGCCCAGCCACGUGCGAAGAACACGUGUUCUGC GCUUGGUGGAAAUUUUUG

chromosome/ location	FPKM*	sequence
5/524246-525974	652	AAUCGUACAACCUACGGCAGCAGCAACACCAACGCCGACGGCAACACC GACACCUUCGCCUACGCCAACUGCGACACCAACGCCUACGCCUACAC CGACACCGACACCGACACCGACGGCAACAAAAACGCCUACACCGACG GCAACACCGACACCGACACCGACACCGACGGCAACAAAAACGCCUACG CCUACGGCAACCGCGACACCGACACCAACACCUACGCCUACGCCAACA GCAACACCGACACCAACGCCUACGCCAACUGCGACACCAACGCCUAC GCCUACGCCUACGCCGACUGCAACUCCGACGCCAACACCAACUGCGA CACCAACUGCAACACCGACACCGACGGCAACACCAACUGCAACACCGA CACCGACACCGACACCGACACCGACACCAACGCCAACUGCGACACCAA CGCCUACGCCUACGCCAACUGCGACACCGACACCGACACCAACGCCU ACGCCAACUGCAACGCCGACACCAACACCUACGCCAACUGCGACACCA ACGCCUACGCCAACUGCAACGCCGACACCUACACCGACACCUACACCA ACUGCAACACCAACGCCUACGCCUACGGCAACCGCGACACCGACACC AACAUCUACGCCAACGCCUACGCCAACUGCAACACCAACGCCAACUGC AACACCAACGCCUACGCCUACGCCUACGCCGACGGCAACACCAACUG CGACACCAACUGCAACACCGACUGCAACUCCGACGCCUACGCCAACU GCAACACCGACACCGACACCGACACCGACACCGACACCGACACCGACA CCGACACCAACGCCUACGCCAACUGCAACCCCGACACCGACGCCGAC ACCAACACCAACGCCAACUGCGACACCAACGCCUACGCCUACGCCAAC UGCAACACCGACACCAACGCCUACGCCAACUGCGACACCAACGCCUAC GCCUACGCCUACGCCGACUGCAACUCCGACGCCAACACCAACUGCGA CACCAACUGCAACACCGACACCGACGGCAACACCAACUGCCACACCGA CACCGACACCGACACCGACACCGACACCGACACCAACGCCAACUGCGA CACCAACGCCUACGCCUACGCCAACUGCGACACCGACACCGACACCAA CGCCUACGCCAACUACAACGCCGACACCAACACCUACGCCAACUGCGA CACCAACGCCUACGCCAACUGCAACGCCGACACCUACACCGACACCUA CACCAACUGCAACACCAACGCCUACGCCUACGGCAACCGCGACACCG ACACCAACAUCUACGCCAACGCCUACGCCAACUGCAACACCAACGCCA ACUGUGACACCAACGCCUACGCCUACGCCUACGCCGACGGCAACACC AACUGCGACACCAACUGCAACACCGACUGCAACUCCGACGCCUACGC CAACUGCAACACCGACACCGACACCGACACCGACACCGACACCAACGC CAACUGCGACACCAACGCCUACGCCUACGCCAACUGCGACACCGACA CCGACACCAACGCCUACGCCAACUGCGACACCGACACCGACACCAAC GCCUACGCCAACUGCAACGCCGACACCAACACCUACGCCAACUGCAAC GCCGACACCUACACCGACACCUACACCAACUGCAACACCAACGCCUAC GCCUACGGCAACCGCGACACCUACACCGACACCUACACCAACUGCAAC ACCAACGCCUACGCCU
5/527205-527471	991	UGGGGAGCAGCGCUCGGCUCUGACAGCACGGAAGGUAUCAUGCCAUA UGCAGAAGGACUGUACCGCGUAUGGACAAGUGCAGCACGUAUUCUGA AAGAUAGUCGCGUGAAUGCUAUAGCUGUUGCCUGCUAAGAAUGCGCU GCGAAGCAAAGAAGGGGGAGAGCUGAUACUGGAGACCAACUCUGAGA UGCCUUGCAGACACUUUGCGACUUUUUUUACGAGUAUUAUUCGUAGU AGGAUUCUCCGAACCGAAAAAGAUCCAGGGGG
6/496493-496672	646	CCCGUUACGUAAACUCUUUAUAGAUAAGAAAUGGCGGUGCGCAGCGC UGCGCUUGUCGCGGCCAUACAACCUUCGAAAACGCUCCCAAGCGCUC GGUCCUGUCAUGUUCUCCGCACGCACUGAUUACAGCCCGUCUAGCCA GAUACGACGUAUGCGCUGCGCAAGUUGCUUGUUCGAAAG

chromosome/ location	FPKM*	sequence
7/316559-316631	741	UAGUGGAGCUGCACACGGGCGUACCUCUCACUUGUUUGAGUUAACUG UUUGAAAACAUAGCGCGUAAUUGAUC
8/318981-319140	1432	GGUUCCACAGCUGCGCACGGAUAUCUAUGGUUCCUGUCAGUUUCCCA GGCAUCGGUUGUCUGUCAGCGUGAACCUGGCCAUGGAGAACGACGA GCUUGGGCGGCAGCGCUUUUGGAGAGACUUCAUUGGCAGCUGUGAC UGGAUUUGGAUGAAUGGGCGG
8/736675-736914	615	UUUUAUGUGCGAAGAAGUACAGUAGCGCCAAUACUUCCGGAGAGCAA AAGCAGCAGCUGUCGGUUACUUGAGAGAGAAUGAGCCGAGCAUUGCA UACAUCUAUGCUUCGAAAUAGUAUAGUGGCUGGGGAGAUGUUGAGCG CUCCGCUCGAACAUACAUUCAGAUAUGAAAAUUGGCUGUUGGCACUG GGGCGUGGUUCUUGGAACGGUGCAGUGAGAGCGAUAAGGAUUUCAC AGUUGG
9/56925-57751	192	GCCCCGUUCAGCGGCACUGCCUGAAACCGACAGAAGGCACCGAAUGA GUACAGCAGCCGAAUCGAGGAGUUGUGGACAACCUAGAGUAGAUCGA GAGGGCCAUAGAACGUUCAUGGAACUCUUCCAGAAAAAAACACAGCG GCUGCUACCGAAUAGGAUCCGGCGAACACCUCCGCCCAACAAACUAA CAAACAAACAAAGAAACAGCUGGAGCGUCAAGCUCCCUGUCUGGUGC CGAUUUCUCCACAGCGCGUGUGGUGCACCGAUAGACUCCGGGCUGA CGCUGGUUCAUUCCUGGAAAAGCACCACGGGCUCUAUCAGGCGAAUG CAACCGCUCCGUGGCCAUAGAUCGGCGCCAAAAUGCUGAUAAACAGC GUCAGUGCCCGUUCAGAUACCUAGAAGACAUCCGGCAUGUCUGGACG GCAGCGCAGUGCCGCUGAAACCUUUCCCAGAGACCAGCAGCAAACGC GUCCCAUACGACGUCCCGAUAAAGCAGACCACGACCAGCGUACCGCG ACCACCUGCGCUGGCAUCAGGCCCGAGCGAUACGCACUGCGCUCAGA CGUGCGGUUGGUGCAAGCCCUGUCGCUUCACUUGCCAUAGGUACCG CAGCAUUUUCAUGGUCGAAGAUGAACGCCUCUGUCCAAACUCUAUGG CGUGCGCGUUGCGAGUCGACUGUCGGCGUCGGGGGCUCGGCUACUC CGAAGACGCUCAAUGUGUGUUCGCGUAGAGUGUCGGGUGCCUCGAG GUGCAGCGUUCGAGGCACCACGCUACGUCUUGCUCCGGUGCCGCGU GCCUACUUUACACGUAUCCACAACAACACACAG

chromosome/ location	FPKM*	sequence
9/786027-787088	626	UCGGAGAGUCUGCGGAUGCAGGCUUGUCUUCGUUGCACGACACGCU GGAGAAACUUGGCUGUUAGGAUGAAUCCUGGGAGCCGCCUCUCUCCU UCGUGAUCAUGCAUGGAAUGCGUUGUUGUCGGUUGGCAAGGCGGAU UUGUGCUUGACACCCACCAGUGGGUAGGCGCAGAAAGCAGUGGGCC CGUGUGCGCACCCGAUUCGCUGGACCUUCCGCCUAGAAGGUAGCGU GUGAGUUUGUUCGCCACUAAUGAGCGCUGAGGCUCGGUUAGCGGAC CAUCAGUUACCAGUGCUACACUCAGAGUGAGUGUCCGGCCUUCGGCU ACCACGCUGAUGCCCAGCUGCUGCUGUCCUUCAAGGAGCUCGCUGC GCUCCUGCUUCCAGCGCGGGCACUGUAAGCCAAAAAUAUGUCAGCGU CUCGGGGAUCUCAUCCCCACUCGCGUGAUCGAUGCAGGCGCACACCU UGCACGCAUGCGCUGGAGACACGUGUACGCACGGAGGCGCUGUGCG GGGCGCUGCACGCCCAACGCCAUCUGCGUGCGGGCGAGAUCGCAGC UGCCACGUGGCAUUUGCGCAUCAGUGCUGCUGAUGCCGAUACCUGC UUCCGUGCGCCACCGGGGAACCUCGCCGAGCAGCGUGAUCGGCCAC ACUGCCGCCGUGUGCGUAUGAAUGCGUGUGCGUGUGUGUGUGUGUU CAUUUCGCAACCCGUUUGUUGCGUGUGCAUUGUUGACGAACUGCGCA UUUGCCUAUCGCCUAGUGAAAGGGGCGUUUACGAGAGCACUGCAGG GUUUCGCGGGUGACACUCUUGUUGUUUUACGUAAAAGGCGUUAUUUU CCAAUGCUGUCGUUCGUUUCUGCGUUGUGGUGGCAUGCGCGACACG GUGUGGAUAUGCCGGGGUUCUUGCCCGGAGUGCGCUGCGGCCUCUC GUCUGGGCUUGUUCGGAACCGAAUGAUAUGGUCGGCUGCCACAGCG GCCAAGGGUGGGAAACCCAAAGAGGCUAGUCGCGAGCAGGCGAAGCG CCCCGAACGCAAGGACGACGACAGGGGCUUUCAAAAGGGUGGC
10/137841-138063	473	AGAGCAAAAUACAGUGUUUGAUCUCAAUACUAUUUCAACCCUCAAUUG GUUGUUCCCAAAAGCUUUUUCGCGGUUUACAGCACAAGCGUUCAAAG UGCGUUCGUUUUCGAGAAUAAAAAAGGAAAGUUAUGGACAAGAGAGU CGAGGCCUGCGGAGGCCACGUGUCACCAGGUUCGACAAAACACUGGA AAACAGUGCGGGGCCCAGAGCGACAGGAGAAGGG
10/295135-295175	942	CUUGAUGCUGUUACGGUCGACAUUCUUCGGCAGCUUGAGCG
11/33434-33942	253	GGGGAUUUCAGCGACUGAUGCAUCAGGAGCCGCGUGUUCGGCUACG CUUCGUAACCUAAGCUUCUAGAAACACAAAGUUAAGCGAUCAAGGCAA GGGCUAUUCUGCCACGGUGAGGCGUCUGCGCUGCAUGCCACCCCAA GUCAGAUCCGUCAUAUCGACUCCAGUAGGCGCGUAAGGUGCACGCAG CUCCGACGUGCACGCAUCGUCAUAGCUGCCGUCUACUUUCACGAUAA GGACAUGUUUCGUCUCAUGACCAGCUCGGUUGAUGGUACCGAAGUUG GACUGAAUGGGAAGCUCGCGCGAGGCCUUACGAGACGGUAUCAACGA GGGCCUUGCAGUCUGCUGCGUCACUUCUGCUGAGGGCGUCGAAGCU CGUACUGUUUGGACUCGAGCGCUUCCAGGCCUUGCGGGCGUAUUGC GAAGGCGAUUCGACGGAAAAGGCUUCGGGGGUGAGCACACCCCCAG GUCCACAUAAUCGUUACCUGAAGAGACCAUCCCGGGAGCGAGG
11/214250-214313	483	UCGGAUUUGUUAAGUUUACAAAUGAACGCACGUUGAACAUGUCUGCU ACAGACUGCGACGAAGC

chromosome/ location	FPKM*	sequence
11/384711-385964	778	GCUGUUUUGCCUGUGAGCUUGAAAUCCGUGAUCGCCUGUUCCACGA GAGAAACACAGACGACGGAGUUCUCUUCCAAAGGUGUAUGAAUGUCC ACACUGGCUGUUGCCACAACAGGCACGGUCUUGGCAGCUCGUGCUAC AGCCUCUGCUUCCUACAGCAUCUUCGCUUCAGUCACGAGUUCGAUUU GCGUGGAUAACAUGCCUCAUCGCUAUGCGCGGCGUCAAGCUGGACG GGAAAUUCUGCUGCCUCCACGACUAGGCGUGGAAACUGCGAUCCUGA AUCUGGUCAAAUCAGUGCUAAAAAAAGGUUAGUAAUCUGGUCCAAGG AACUGAAGCAGGUCGCGGGGACGCCACCGUUUUGAUGAGCGUACGC UGGCGACCAGUUGCAAGCGACUGGCAGGUCCAGGCACUACAGCUGG CCGCUGCUCGAAGAAUACGCCCAUCGACCGGGGCGCGUCUCACGACA CUGAGACAGUUGCUGGUAGCGACAUAACCGGUUCCGUCUCGAAUGGC CAACAUAGGGAGCACACUGCCAGAGUAAGCCCCAUUGUGUUCACGAG UCGCCGUCCAUGAAGGACGGUCGCUCUCGCUCGUUCCAUUGCUGAU UCAGUGCAGGCUGCCGCUGCCAAACAUGGGGCGCAAGCACCCCAACG CGGCAGCAGGUUCCAGCGCUGCACCAUCGACAAGCGACGUUCACUGU CGCGCAGCACUCCCGCUCUUAUUGGGGUUCUGUGGUUGCAGUCCCG CGAGCAGAAGUGUGCGCGACCUGAAACGCCCAGCUUUGCUUACCACC GACACUCUCUGACCUACUGUCCGGGGACUUUCACCCAGAGCCAAACG GCUGGCAUCGAAUCAGGGCGCACCUGUUCGCACCGAGCCUGGUUUA GGGGCUAAUCCAGGACCCUCCCGUAAACCGUACGUAGCGCUCUGAUA ACGAGGAACGCUUAGCAGAGUCUUCCCCGAGUUGAUCCGUUGCGUCC GGAGCAAUUCCAGUUUACUUACGUUCUGUUGGGCGCGAAAAAAGCAA CAUUAGCUUAUUGAGUGUGUGCUGGCGUCGGUAGCCAAUGUUCUCG UGGUGAAUACGUCCACGGACAAGGCGACUGCCACGGGCUGGUUGCG GAUCCGCGUAGACUUUCCUUUGGGAAAAGGAUCGUUUAAACGACACA GAGGUCCUUCACCCAAAAAAAAGCCCCGCUCGGCAGAUCGCACAACA UGCCGCAGCGUUUCUUCUAAUUCCGUCAACUAACCACGUCC
11/550921-551229	467	UGGAGUGGAGCGGGUUGUUGUCGGGAGCCGACAACCCGCAGGUCAG CGGAACAUCGCUUAGCGGCAAUCAUGAGGGAAUGGAGUGGAGAGACG UUGUUGUCGGGAGCCGACAACCCGCAGGUCAGCGGAACAUCGUUUAA CGGCAAUCAUGAGGGGAUGGAGUGGAGCGGGUUGUUGUCGGGAGCC GACAACCCGCAGGUCAGCGGAACAUCGCUUAGCGGCAAUCAUGAGGG AAUGGAGUGGAGAGACGUUGUUGUCGGGAGCCGACAACCCGCAGGU CAGCGGAACAUCGUUUAACGGCAAUCAUGA
11/762893-763006 (**CmU2**)	690	UUCGAUAUUUUGGUAGUUUCUGUUCUUUCUACCGUAACAGGUAGAAG CAGCUCAUGGGAUAUUUUAUAUUCCAGGGAGCCUGCAUCUCUUAACU AAUUUUUCGUUCUCUUUUCC
12/375914-376113	1462	CCUCCUUCUCAUCUUCAUCCUUCGUCACGUCAACGACAAAGUCGUUA UCCUCCACCUCCUGCGUCGCAAGAUCCGCCUCCUUCUCAUCUUCAUG GUUCGCCGAGGGCGGAACUGGCCGAUCUUUCUUCAACCGCCGAUAG UGUACACACAAGGCUACUCCAUCGAUUUCGAUCAUCCCAGUGAACAAC UUUUGCUUGUUC
12/377138-377293	913	GGACUCGGCACUAUUUUUUUCUGGCAGCCGGCGAACCUGGGCCUCC AGGAUCGUGCAAUUGCUUUUUCAUAGCGUUCCUUUUUUCGUUUUUUU AUUUUUGUAUUUCUUGAACGCUCUUUAAACCCUUCUGUUAUAUACUG CGAAAAAACAUUUGAC
12/543182-543223	630	ACGCUGAAACGCUCGCUUGAGUCGGUGGACGCGCUCGCGAUG

chromosome/ location	FPKM*	sequence
13/648388-648435	884	CAGGACGAUAAGUCCCCAGGGAAGAGAGUUGCAGAACGGUGCCUUUU
14/127504-127570	756	CCAUCAGAACGUUCCGUUGAACGCGAUCCGGUCAAUGCAGAACGGCU ACUCAUUUAGACCAGAGUUC
14/433866-434327	1784	UACUCGGUUGCAGCGAGCACAACGCUUAGCCACCAAAGCCGUAGAGC GUACGGCCCUGGCGCUUCAGCGCGUACACGACGUCCAUCGCGGUCA CCGUCUUGCGCCGGCCGUGCUCCGUAUACGUCACCGCGUCGCGGAU CACCGACUCCAGAAAUACCUUCAGAACGUUCCGCGUCUCCUCGUAGA UGAGGCCGGAGAUGCGCUUCACACCGCCGCGGCGCGCCAGACGCCG GAUAGCAGGCUUCGUAAUACCCUGGAUGUUGUCGCGCAGAAUCUUGC GGUGGCGCUUCGCACCGCCCUUCCCGAGACCCUUGCCGCCCUUGCC ACGACCUGACAUUGUUUCAAAAGAUAAGCUAGCUCACAAGCAAACAAA CACGCAGAAGAAGAGCAAACACAGGCUCAACCGUAGACGUGCGCGCG CGUAAACGAGAACGCAAACGCAAGCACACGCGCGAAGAGAAU
14/439359-439531	1791	ACCUUCUUCGCAACAGUCUUCCCUUUCAUAGCCAUGUCUCGGGUUCG CAGUCAGACACACUCGAGCGAUAGCGCUCACGGUUACAGAGCAACAA CAGCUGGGCACGCGCAAACACACGCGGGCAGGGACACAAAGACGCAC ACAGGCCAGGCACUUGUCGAGGCAGACCCGCG
14/721037-721187	874	UUCGCAAGUAUUGCUGCGAACGCGAGCAGAGCGCCUCAUUCGCAUAU GUGCCUGCUCUAGUCUGCGACACCGAGGAUCGGCACAGAUAUGGGAA AAGAUUCACGAUCGACUGGCCCGACUGCCAGACUCGCAAACGACAUG GCAUUCGACA
15/167096-167463	1865	AGUGGAAGAAGAAGCGUAUGCGCCGCUUAAAGCGAAAGCGACGAAGG AUGAGACAACGUAGUAAGUAAGGCUGCUCUUUUCGUCUCCAUAAAUG UUGGCAGCUUGGCUCGCAACGUGCACGACAGGCACUCGGGGCGCCA GUUCAUGAACGAGCUCGCGGCUGUGUUGCCGGUUUACGUGUUGGGG CGACUCUUUUUAGCUGAGAGUCGUCGUUGUCGUUGUCAUAAAUUGCG ACACAAAGUAUUUGCCGCCUCUUCUUCUCUAGACGGGUGAGUUGAAG CUACGAACUCGGAGGGCACAACUUUGACCUCGGAAACGAACCAGUUG GAACGGUAACCACUUUUAGCGGUAUUAUAAACAGGACGGAC
15/408883-409004	407	UGAAAACUGCGGGAAGCAACGUCUUUCUUUUGAAAGAGGUUUGUUGU UGCUGGUGGCGGUUGUGCAUCUUUAUAUCCUGGAUCGUGUCAGCGA UCUCACACGGUACAAAAGUUGAAGGGGUG
15/411466-411544	574	GCGUCGAUGACGAGACCGUUACAUGGUGCAGUUUUUAGCGUGCAAAU GCUCAGGAUAUGAUAUCCCGACGACUCCAAGC

chromosome/ location	FPKM*	sequence
15/446874-447523	683	AUCAUUAAGGUUGUACGUUGCGUCAACAGCCGACAGAGUUUUUCUAG GUACUUCGUAAACGAGUUCUAUGGCGCCGCAAGGAGACAAAGGUAAG GGAAAGGAGGCAGAGCAGUCCAAACCGAUGGGUGGUCCAGCGAUGG CCAAGCCUGCUGCGGAUAAGCCUGGGGAUAAGCCCGGACAAGGCAGG GCAGGUGGAAAGCCCAAGGGCAAGUGAGCACGUCCUCUGAAGGUUGU GGGUGUUCUACCGGAAUCGUAUGGUUGGUUAGUUGGUUGGUUAAUU GGUUAGUUGGUUGGUUGGUUGGCUAUCGUAUUGACAAGUGGGUGCU AACUUUAGCCGCCGGCGUGACACACGUUGUUUGGUAAGGUGUCGCA GUUCGCGAGCGAGGUCGCAGCGUCUGGACCGUCGACACGGCUGAGG GCGCGUCCUUGUCAGGUAAUCGUCCGCACUGGCGUGGUUGUCAGAG CACGGGUGCGCUGGAACGGAUGCUCGUGGUCGUCGUCGAUGCUCAG UAGAACAGCGGCGGCGCCGACUGCAUUCCGAUGAAUGGCAGGGUUU UCCGGCGAGGUCUUGGGGCAUCAUGAGCAUUGCCACCAUCGUCGUC GUCAUGUUGGUGCCGAUGUCGUCAUGGAUUCCAUGGUGGCCACGGG UG
15/643266-644249	2863	CCACUCUAAAUUUUCCCUAUUCAGCCUCAUCUCUGCUUUGCGUUCGA GCUCUCGGUCGAGAAAUACGAAAUAUCGCAGGAAAAGGCGAUAGUAU CUUUCCUUAUCUUUUUUCAACUCUGAUAUCUUGUCUGUUGGAUUCGC CAUCCCCAAGAUUGCACGGUGCGCUUGUAGAACCGCGCGGAUUACAG AAGCCUCAUUCGUGCUCUCGUUUGACCCCUUCUCUCCGGCGAAUUCG ACCUCAUCUUGAUGCUUACGAUCUUUCGUGAUGCCAUUCAUAAUCGC CCAUUUUCCAUUGCGCGAUAGGCCCGGGUAGAGUGCAAUCACGGCGC GUAUCAUGAAACGUUCGAGGCUCGUUGUCAAAUGGUUCUUCAGGUUC GUGAGGUACUUCAUCGCUCCGCAUGUGUAAAUAUCCCGGUGACUCUU GUACCGACGAACCCGUAAGAGGGAUGUUUUUCGUGGAAGGCAUGCAC UAGCUCAUUUUUCCUCCGCGCUUCCUCUGUACCAAGCAUCAGGUGAU GAAUAAAGGUCUUAUCGAAAAACCCGUCCGGGAACCUCGACCGUUUC CAUGUGCGUCACGUCGCGAUACAUCUCCCUCGCCAAAUACAUCAGUC CUGACGACGCCUUGACAAUACUCUUGGAGUACGAGUCGACACGUUUU UUGAUAGCGUCUCGCAACUUUUCGCUGUACGGAUCCUUGAUAUAUCC AAGAAGAAAGGUUUUCACCACCUUCUCGUCGUACAGCCGUCUUUCUU GCUGCCUUCGACGUAGGGCCUCCCCCUUUUGCUUCUUUUCCUCUUC CUUUUCUUCUGGAAUGGGCACUGUUUUUUGUUUUUUCUCCAGCAGCC GGCGACCCUGGGCCUCCAGGAUCGUGAACUUGCUUUUUCACAGCUUU CCUUGUUUUGUUUUUUUUUUGAAAAAGAACGCUCUUUGAACCCUUCU GUUAUAUACUGCAAAAAACACUAGACUAGACUUGUUUGAAGUUUG
15/679145-679299	304	CCCCCUAAGAUAUCCUGAGCGCAUCGUUCCGGUACAAUCGCUCGUUA CCAUCAGAUAUCGAGUGCGUUGCGUACUGCCAAACUCGAUACGAUGC GAGAUAUGUUUUUUCCCCAGAGAUGACAAUAGCUACCAGGCAAGCGU UGCGAGGGACUUUU
16/52555-52675	971	CCCGUCAUCCUGCUUCGAAGACCAGCUAAUGAAGAGUACCAGCAUCG UGAUGAAGACACCGAGCAGCACGAUGAGCAGACGUGAGUCGUCGAUU CCAAGCGGCUCACCGGUGCCGUCGGUG

chromosome/ location	FPKM*	sequence
17/63114-63525	359	UCGAAUUAGAAAUGUGCAUUCAUUCGCGAAUGAUGUUCGUCGUUCAU CGCAGUGUGGAACCGACCAUGCCUACUGCGACAUGUCCACGAUGCCA CCGAUCCGGGGCAACACCCAGGUUCUAGACGACAUAAAUGGGCAGGU GCAGCACGCUAUACCCCAGACCUGUUUCCACUCUGGUACAGUAGCCA UGAACCGGCACCCAAGGACGACGCUCAAGUGCUCCGUGCGUGAAUCG GAGAACCCAGCUAAGCUCCGGCGGCGAGCGAGACCAACGCUAGAUGG UUGAACGUGUGGUGGCGAUAUCGCCACAUCCCUCGUCUUUGGUGAG CCCUCAGAUCCAAUCCUCCUCGUUUUUCUCCACCUCUUUAAGCGCGA UUUUUGCUGUCCAGUUCCCUUUCCAGUCAGUGCUUAG
17/504446-504495	948	AAACGACCAGUUUCUCCCAGAGAACGUAUACAUCAUGGAUCCUAAAGG UA
17/771403-771879 (**CmU5**)	7046	GCGCACCGCACGAAAGGUGGGGUUUCAGCGGCUCUCCAGUUUAGAG GACCGGAGGGCAGCGCCCUUUCCUGACGCCAUGUACGGAUCCAUACG CGGGAGGUGCCUCUAAGCUCCGCCACUGCCUGGUUGCGGAGGCGCU GCGGCCUUUGUUGAACGCAGGCAGGGCGAAAAACGAACACACCCCAU GCCUUUUACUUGCGGUGUCCGGGCGGAGUUUAGUGAGGACCGGUGU UGGCAGGGAGCUUUCGGAAGCUUUUUCUGCAUGAGACACAGUGUUCA CAGGGCCUCCUCGUUACUUGGAUUCGUAUGUGGACCGGGACUGGGC GCCAGAGCGAAACUCGAGCUGUCGAGCACGCGCUGCCUGUAAAGUUU GCCUGCUCCCCAGCAGAAUCGUCUUUGGACGACCGGCUUGAGGGAG UAUAUGGAAUUUUUCGUUUGAUCCAGUAUUCACACUGUCCUCCGUCU CCCAUAACCAGC
17/907690-907898	804	GCUGUCGUCACAGGCUGCGUCGCUGCACUCGUACCGGUUUGCUGCA CAGGAUUUGACUCGUGCUUUGCACUUGAUUUCUUGGGGUGCGUUCC CUCGUCCGUUGAUUUAGAAGAGCUGUGAACACUGAAUCCACAGAGGU UCCCUAUUUUGACAGCGUCAGUCUCGAUACAACACCGUUGAGAAAUA AUCAAAGAGCAAACAAACGUUCG
17/1037451-1037563	483	GAGUUCCGAACGUCCUAUACACACCAUGAACUGGGGUGUACACUUCA GAGAGUCGAGAUGGAAACGUGCCAGAGGAGCCCUAAGGAUAUAAAAA AGUACGCAACGCUGACGGG
17/1158095-1158131	964	GGCUAGGACGCGGUUAUCUGAACUUGUCACCGUAUUC
18/120851-120913	893	CCAGGGACAUACCUUGCGCACAUUAGGGCGUCGGCGUGAUAACCUCU CGUCGACUGUCAAGUCGUCAGUUCAUUCGCAAAUACUUCCAGGUUCC AACCAGUCCUAUAGUAUCCCAAAAUCUAACCAGACGCUUCCAG
18/196352-196409	776	UCGUUGGAACCUGGAAGUAUUUGCGAAUGAACUGACGACUUGACAGU CGACGAGAGGU
18/447856-447983	553	CGCUUCCGCUGACGGGUACGCGGUGCGCUCGGAGCUCGGGUGCUCU GCGCGUUUGUUACGGAGUAUUGGGGUCAGCGCUGGACACGUUCAGG UACGGAGUAGGAAAAUAAGCAUGCAUAGGAUGGGGGG
18/465640-465768	927	CCCCAUCCUAUGCAUGCUUAUUUUCCUACUCCGUACCUGAACGUGUC CAGCGCUGACCCCAAUACUCCGUAACAAACGCGCAGAGCACCCGAGC UCCGAGCGCACCGCGUACCCGUCAGCGGAAGCGCC

chromosome/ location	FPKM*	sequence
18/508952-510214	1103	CCCGCCGCUUCUCUAUGCACCAGGCACUGGGCACAGACCCGGGGACU UUGCAAGCCGGUCGAGCGCCGCCUCCGCUGGGAAUAUAUACCCAUCU CGUGGUUGCCGCCCGCCCGCCCGGUCAACCGAGCGCGACGCGACAA CAGCACGCAGCUAUACAUUGUACCGAACAGACGCACGCACCCGUACA ACGCACCAUGCACAACCCGAGCUCCGCGCUCAGCGCCCGGCGAACAC CAUCGCCUCCAUCCGUACAGAUGAAUACCGCGACACGCCAGUGCCAA GGUCGCCCGGCGCCGCACGCGUCUCGAUGCAAGUGCGUGUCGCUGC UCCCGCCGCUCAAUGCACCGACGGCUGCAGUUGCGCAGCCGCACACG CGCACGCAGCAGGGAGCGCACACGCCCCGGACGCACGCGGCGCCGG ACUUGGCACAGGGCGCCGCUCCAAGUGGGCCGUGUCCUGUCUGCGG ACGGCCGCCAAGGCUCCCGUCAGCGCGGGGGUGCCAGGACCCCCGC CGUCCGGCGCCGUCAAUUCCCAAGAUGGGAAUCGCGGCUCUGCUGC GCGUCGACGAUCGCGCUGCCGCCGCCGGCGAUGGAUAGGCCCGCGG GCGCAAGUCCCUCUGGGAGGGAGUUUGCGCGCGCAGCGGCCUCUGC CAAACGCCGCGAGCAGCGCCCGCCGACGAGCACCCGAACAUGAACGC AGACCACGUCCCCACAGAUAUCUCAAUGUCACGAUCGUCAUCGCUGC CCCCCAGCGGGUACCAGCGGCACAUGGCUCGAACUGCGCUCGGCGC GGCUGUGUAGCGCGCGGCGCACAGCAGCACAGCAGACUCAUGCAGC GAAAGCAGGAUAACGGGACGGCGAGCGAGCCUCGGCGCAGGUCCUC CAGCCCGCACGCUUGCGCGCGACAGGACUUGGAGAAUGCCCCGCUU UGUGACUGCGCAAGCCACAACACUCGCGCCUAUCCACGUACCGAAUC GCGUUUCCACUUGCACUGCGCGCACCACACCUAUCAUGACCCCGCCC GAGACUUGUACCUGAUUGGCGCAUCGGCGCGUCCGCUUCCUACACG GGCGGUGCGACGCCCCUCGGACCGGACCCUCCGACGCGCGCUGUCG AUGUUCACGCGUCUGCCGCACGCGCGACGCGGUAGCUACCGCCAUG AACGGCGGCUCUGCGCCCGCGACGUCGCGUGCAACGAAGGAUGCGA UGACAAGCUCGCUUACAAUCCGCUGCGUCUACACGACGCUGCAGCCG UUACGAUUUG
18/1054762-1054825	1071	AGUUUUUCCGGAGAAGCUCUGAAAGUGGGCGAACGCGGGGUGCUUG GGCGACGGCGAAGGAAGC
18/1142041-1142176	461	GCGCAGAUCUCGCAGAUCUCGGCAGGGACCCCUGCCUGGACGCCCG UGAGAGUGGAUCACUUGUACCGAGGCGCGCGGUGAUAAAUUUCCAUC AUCAUCGCUGAGCAUGUGCCGGAUGAAGAUUCAUCACGGACGA
19/498151-498206	326	GUUCCGUUAACAACACAGCGUUGCCGCGGACACAACGCUCGAUCGUU GGUCUAUCC
19/1013895-1014118	869	UUUGGAAUCGCGUAUUGCGGGAUGGAUAUUACGAUUUAUGUAAGUGU UCAUCAGGAUGCUGAACUUCAAGUUCCAUGACAGAUACUUCCUUGCG GAUGUACUGGUCGUUUCGGACGUUCUAUCAUGAGUCUAGUGCCUGCA GUACGCAUACGCGUUAGCGCCGAGACCGCUGGAGCGGAUAUUUCAUA GAGCCCGAACGGCUUUGCCAUACGACUGGGGCGGGG
19/1062987-1063119	473	ACCGCGUAACUCGGCAUGCACUGGUUUGUGUAUACGAUUCCUUGGAU AAAGAACACCAGUAGCUGAAGUCUGCAUCAUGAAAGGAUCUGUGAAU UAGCAGUCCACCUUGUCGUUUUCUUUUUUGAGACCCGAG

chromosome/ location	FPKM*	sequence
20/27559-27706	490	AAACCAGUCACAAACUAUAUAUGAAAGAACAGCUAUCUUACAAGCCCCAUGCACUACGCCUGCCUAAGAGGUCUAGUUUUCCGGAGAAGCUCUGAAAGUGGGUUAACGCGGGGUGCUUGGGCGACGGCGAAGGAAGCCCACGACACGG
20/801125-801166	310	GUAACAUUCCACUGAAGACUCUUAUGGACCCCUAGCCUUCAC
20/1546265-1546379	1401	UCCCAGAUUCUGUGAAGUGAUGAACGGGAGAUGUCAGACUCUAUGUCUGAAGAGCGGUACGCCUUGUGAGAAGCAUGGAAGUAGUGCGUACGCAAAGCUUUCAUGAAAGGCC

*FPKM fragments/Kb mapped