3928–3944 *Nucleic Acids Research, 2007, Vol. 35, No. 12* *Published online 30 May 2007*
doi:10.1093/nar/gkm347

Proteomic analysis of *in vivo*-assembled pre-mRNA splicing complexes expands the catalog of participating factors

Yen-I G. Chen[1], Roger E. Moore[2], Helen Y. Ge[2], Mary K. Young[2], Terry D. Lee[2] and Scott W. Stevens[1,3,4,*]

[1]Graduate program in Microbiology, [2]City of Hope Beckman Research Institute, Duarte, CA 91010, [3]Section of Molecular Genetics and Microbiology, University of Texas at Austin, 1 University, Station #A4800, Austin, TX 78712 and [4]Institute for Cellular and Molecular Biology, University of Texas at Austin, TX, USA

Received March 14, 2007; Revised April 19, 2007; Accepted April 20, 2007

ABSTRACT

Previous compositional studies of pre-mRNA processing complexes have been performed *in vitro* on synthetic pre-mRNAs containing a single intron. To provide a more comprehensive list of polypeptides associated with the pre-mRNA splicing apparatus, we have determined the composition of the bulk pre-mRNA processing machinery in living cells. We purified endogenous nuclear pre-mRNA processing complexes from human and chicken cells comprising the massive (>200S) supraspliceosomes (a.k.a. polyspliceosomes). As expected, RNA components include a heterogeneous mixture of pre-mRNAs and the five spliceosomal snRNAs. In addition to known pre-mRNA splicing factors, 5′ end binding factors, 3′ end processing factors, mRNA export factors, hnRNPs and other RNA binding proteins, the protein components identified by mass spectrometry include RNA adenosine deaminases and several novel factors. Intriguingly, our purified supraspliceosomes also contain a number of structural proteins, nucleoporins, chromatin remodeling factors and several novel proteins that were absent from splicing complexes assembled *in vitro*. These *in vivo* analyses bring the total number of factors associated with pre-mRNA to well over 300, and represent the most comprehensive analysis of the pre-mRNA processing machinery to date.

INTRODUCTION

Eukaryotic RNA polymerase II (RNA Pol II) transcripts are matured through a highly coordinated program of processing steps prior to export from the nucleus to the cytoplasm where they are translated into protein by the ribosome (1–5). These pre-mRNA processing events include 5′ end modification by 7-methyl-guanosine cap addition and binding of the nuclear cap binding complex, intron removal by the spliceosome, 3′ end cleavage and poly-adenosine tail addition, transcript-specific modifications such as adenosine deamination and binding of specific protein factors to regulate and promote mature mRNA export from the nucleus. Coordination of these events involves interaction between the machineries involved in each process. For example, the RNA Pol II transcription complex communicates and interacts extensively with the 5′ end capping, pre-mRNA splicing and 3′ end processing machineries (1).

While native pre-mRNAs contain multiple, often extremely large introns, *in vitro* pre-mRNA splicing reactions are carried out using synthetic pre-mRNA fragments containing a single, efficiently spliced intron of a size compatible with acrylamide gel electrophoresis analysis. Although the core pre-mRNA processing machinery will likely be very similar between different transcripts as well as for the multiple introns contained within a single transcript, the bulk pre-mRNA processing machinery purified from its native context is likely to contain a more comprehensive sample of the polypeptides required for or participating in the splicing of pre-mRNA in vertebrate cells.

Several groups have purified and characterized spliceosomes formed on model vertebrate pre-mRNAs *in vitro* (6–8) and shown that they contain a remarkably large number of associated polypeptides. Nevertheless, as these synthetic precursors have generally been modified from their natural state by internal deletions within the intron and truncations of the exons, the pattern of associated proteins is inevitably less complex than on the full-length, generally multi-intronic precursors that exist *in vivo*. In addition, the spliceosomes purified from *in vitro* reactions were assembled on pre-mRNAs derived from either the adenovirus major late or β-globin loci. Thus, it is likely

*To whom correspondence should be addressed. Tel: +1-512-232-9303; Fax: +1-512-232-3432; Email: scott.stevens@mail.utexas.edu

that there exist a number of factors that are required for or participate in pre-mRNA processing *in vivo*, yet are not present in previously purified splicing complexes because they are specific to one or more of the thousands of other pre-mRNAs present in metazoan cells. Finally, the pathway by which pre-mRNA processing complexes are assembled *in vitro* using salt-extracted nuclear fractions most likely bypasses many interactions relevant to this process *in vivo*. Thus spliceosomes purified following *in vivo* assembly are expected to contain additional components that reflect the native pathway, but are not required to effect model intron removal *in vitro*. Additionally, factors that assist in inter-spliceosome interactions in multi-intron substrates will be absent from mono-spliceosome purifications, and should be present in *in vivo*-purified complexes.

Consistent with this view, other investigators have shown that endogenous pre-mRNA is processed in extremely large ribonucleoprotein particles, called supraspliceosomes (9–11) or polyspliceosomes (12). Biochemical and structural analyses of these complexes have demonstrated the presence of RNA Pol II transcripts (13,14) and the pre-mRNA splicing machinery components (15,16) as well as functional interactions that mirror those in active splicing complexes assembled *in vitro* (12). The higher order particles formed *in vivo* partly reflect the presence of multiple introns, an average of eight per pre-mRNA (17) with some transcripts possessing as many as 147 introns [Nebulin (18)], that need to be faithfully removed prior to nuclear export. In Figure 1, we present a schematic model of the pre-mRNA processing pathway *in vivo* that encompasses the concept of the supra/polyspliceosome. Whether the individual 'spliceosome' moieties are formed via stepwise snRNP assembly on individual introns (2) or via pre-formed penta-snRNPs (19) in vertebrates is still a matter of considerable debate, although recent chromatin immunoprecipitation experiments in human cells provide support for the penta-snRNP model (20).

With the goal of expanding our understanding of pre-mRNA splicing as it occurs in intact cells, we have purified the endogenous pre-mRNA processing machines from HeLa cells and from chicken DT40 pre-B cells (21) on a preparative scale and have defined their RNA and polypeptide compositions. We have chosen the chicken DT40 system to compare with the HeLa system for a number of reasons. First, we have shown that working with this rapidly growing cell type, which possesses high rates of homologous recombination, allows for downstream experimental flexibility in epitope-tagging of other genes (22). The evolutionary distance between human and chicken will also allow us to assess the evolutionary conservation of the machinery as well as validating novel co-purifying factors. We show that these pre-mRNA processing complexes contain spliced and unspliced mRNAs, all five spliceosomal snRNAs and polypeptides involved in all aspects of pre-mRNA processing from transcription to nuclear export.

Although our strategy may not be sensitive enough to identify very low abundance pre-mRNA-specific factors, it has allowed us to probe more deeply into the general pre-mRNA processing machinery present in vertebrate cells.

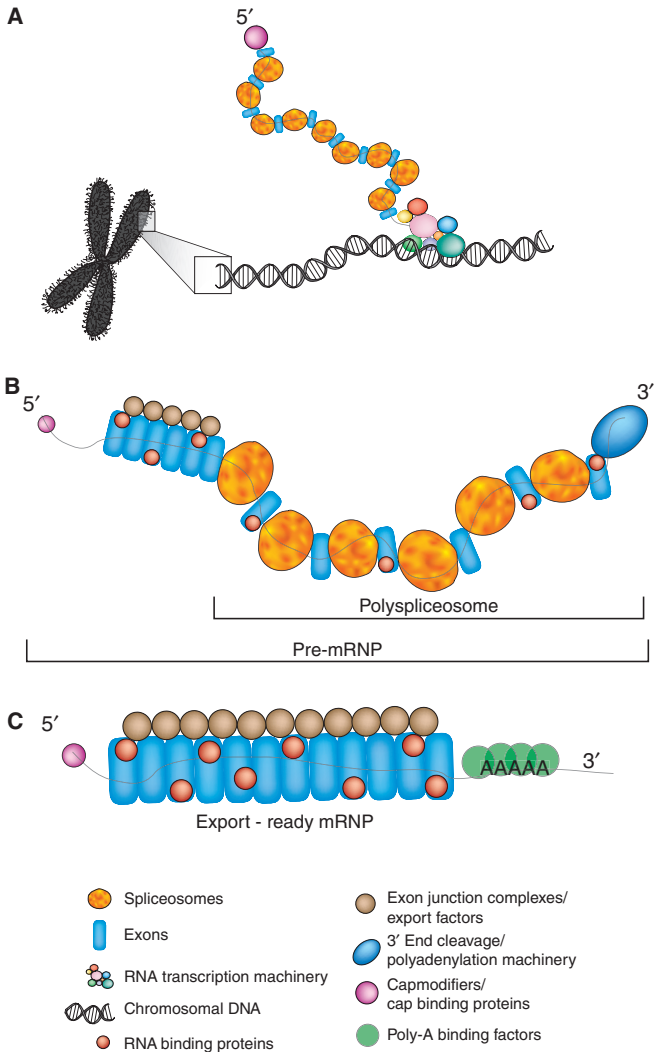

Figure 1. Model describing the role of vertebrate supraspliceosomes in gene expression. (**A**) Co-transcriptional assembly of spliceosomes, 5′ end modification machinery and other pre-mRNA binding factors on RNA polymerase II transcripts. (**B**) The released transcript is partially spliced and bound by numerous spliceosome moieties as well as the 5′ cap-binding complex and 3′ end processing factors. (**C**) The mature mRNA is associated in the nucleus with RNA binding proteins, 5′- and 3′-end stabilizing factors (the CBP heterodimer and poly(A)-binding protein), and proteins that promote export to the cytoplasm.

Indeed, when combined with the data from *in vitro*-assembled spliceosome characterization and known splicing factors not detected in any complexes previously purified, we show there are at least 305 polypeptides involved in or present during the processing of nuclear pre-mRNA.

MATERIALS AND METHODS

Purification of HeLa supraspliceosomes

Ten liters of HeLa cells (purchased from the National Cell Culture Center) were processed essentially as described (11). Briefly, cells were washed in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) and disrupted by mechanical breakage in a glass dounce

(20 strokes, pestle 'B') in a hypotonic solution (30 mM Tris–Cl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol) at 4°C. Nuclei were pelleted at $1000 \times g$ at 4°C for 5 min through the hypotonic buffer containing 25% glycerol. The nuclei were washed three times in the hypotonic buffer containing 0.5% Triton X-100 and once with detergent-free hypotonic buffer. Nuclei were re-suspended in a low-salt buffer (LS+; 10 mM Tris–Cl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.15 mM spermine, 0.05 mM spermidine) and sonicated twice for 20 s at the maximum microtip setting. The resulting nuclear debris was pelleted at $14\,000 \times g$ for 10 s, and the supernatant was layered onto a 15–45% glycerol gradient (11ml Beckman SW41) made isotonic to LS- buffer (LS buffer without polyamines) and sedimented at $40\,000 \times g$ for 90 min. Fractions (420 µl) were collected from the top. Protein and nucleic acid were separated by phenol/chloroform extraction and precipitation with acetone (23) (protein) or ethanol (nucleic acid). Fractions corresponding to the supraspliceosomes were pooled from six velocity gradients run in parallel fashion, diluted to ~8% glycerol with LS- buffer and incubated with 20 mg Y12 antibody which had been covalently attached to 1 g CnBr-sepharose (GE Biosciences) according to the manufacturer's instructions. After incubation with rotation for 2 h at 4°C, the sepharose matrix was washed with 200 ml LS- buffer by gravity flow in a column and supraspliceosome material was eluted with 0.2 M glycine. Protein and nucleic acids were separated by phenol/chloroform extraction as described above.

Purification of chicken supraspliceosomes

Six liters of SmD3-TAP DT40 cells (22) were grown in Dulbecco's modified Eagle media supplemented with 5% chicken serum and 2.5% Fetalplex (Gemini Bio-Products) to a density of 7.5×10^5 cells/ml for TAP purification. Cells were harvested by centrifugation (1000 x *g* for 5 min), washed twice with ice-cold PBS, allowed to swell in 10 ml of TM buffer (10 mM Tris–Cl pH 7.5, 3 mM $MgCl_2$) with 0.2 mM PMSF, 1 µg/ml leupeptin, and 1 µg/ml pepstatin for 10 min ice, and lysed with 25 strokes of a Dounce homogenizer at 4°C. The nuclei were pelleted and washed twice with 10 ml of TM buffer containing 0.1% NP40, re-suspended in 5 ml of low salt buffer (30 mM Tris–Cl, 125 mM KCl, 5 mM $MgCl_2$, 0.5% Triton-X100), and sonicated at the maximum output, twice for 20 s on ice with 1 min in ice between sonications. The sonicated mixture was centrifuged at 14 000 x *g* for 1 min and the supernatant was used for TAP purification. TAP-tagged protein material for SmD3-TAP DT40 cells was affinity purified by the TAP procedure (24). The TEV eluate was layered onto glycerol gradients and fractionated as described above for the human supraspliceosomes.

Immunoprecipitations

Polyclonal antisera directed against the carboxyl-terminal 15 amino acids of KIAA0332 and NP_035897 (NCBI accession numbers) were produced by Genemed Synthesis and the IgG fraction was partially purified by ammonium sulfate precipitation at 50% saturation. Antiserum or non-immune serum was incubated for 1 h at 4°C with the sample(s) of interest prior to addition of 50 µl Protein-A agarose beads. This mixture was incubated one further hour with rotation at 4°C prior to washing with 4 × 15 ml IPP150. Proteins and nucleic acids were released from the matrix by incubation in IPP150 at 100°C for 5 min. The supernatant was collected and phenol extracted as described above to harvest, separate and precipitate the nucleic acids and proteins.

Northern blot analysis

Nucleic acids were transferred to Brightstar membranes (Ambion) and hybridized with snRNA probes consisting of antisense chicken snRNAs transcribed with α^{32}P-GTP using T7 RNA polymerase (U5) or SP6 RNA polymerase (U1, U2, U4, U6 snRNAs) from plasmids containing cDNA versions of the chicken snRNAs.

Western blot analysis

Polypeptides were resolved in 10% polyacrylamide gels (25), transferred to nitrocellulose membranes (Biorad) and blotted with antiserum as described in the text. The secondary antibody used was horseradish peroxidase-conjugated goat anti-rabbit IgG (Rockland) and the signal was detected by enhanced chemiluminescence (Perkin Elmer).

Mass spectrometry peptide identification

Pooled supraspliceosome protein fractions were separated by polyacrylamide gel electrophoresis and stained with Coomassie Blue G-250 (26). Discrete gel slices were dissected from the top of the gel lane to the bottom and all regions were subjected to trypsin digestion. Mass spectrometry and database searching was performed as previously described (27–29).

RESULTS AND DISCUSSION

Purification of endogenous human pre-mRNA processing complexes

We discovered that in gently sonicated nuclei treated with low salt (11), the majority of the snRNA, as judged by visual inspection of ethidium bromide stained gels (Figure 2A), was engaged in very large (>80S) ribonucleoprotein complexes that closely resemble supraspliceosomes in sedimentation values and other properties (9,10,30). These particles may also be related to the polyspliceosomes described in salt-extracted nuclei, which sedimented as complexes slightly smaller than our supra/polyspliceosome, likely reflecting salt-induced factor loss during nuclear extraction (12).

For purification of the human supraspliceosomes, we gently sonicated nuclei in a buffered low-salt solution and purified the material in the ~200S region (Figures 2A and 2B) as previously described (11). This material was immunopurified using the antibody Y12 on a solid matrix. Remarkably, this treatment nearly quantitatively retained the detectable material from this region of the glycerol gradient as judged by coomassie gel staining,

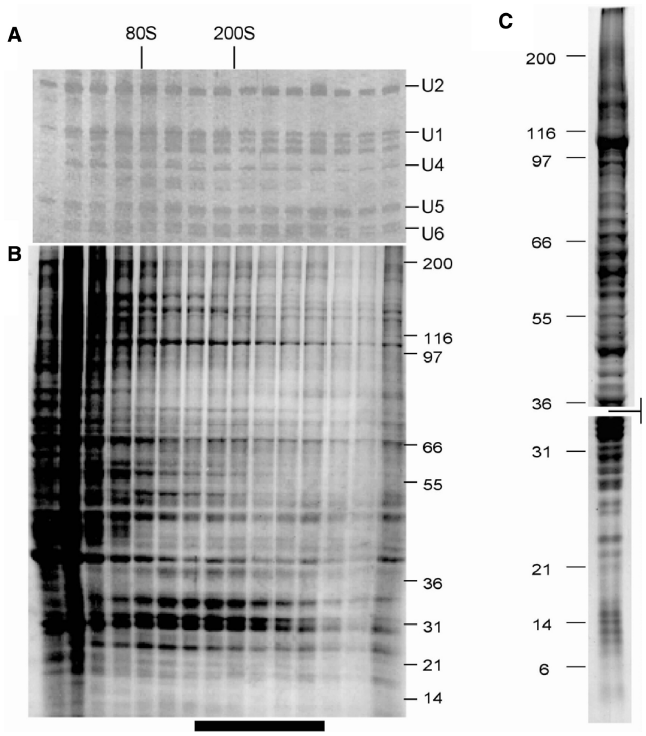

Figure 2. Human supraspliceosome-associated polypeptides and snRNAs. RNA (**A**) and protein (**B**) were extracted from preparative glycerol gradient fractions and electrophoretically resolved through urea-PAGE (A) or SDS-PAGE (B) gels stained with silver (RNA) or coomassie blue (protein). Bar below B represents the fractions of the material pooled for immunopurification with Y12 antibody. (**C**) Affinity-purified supraspliceosomal proteins run under two SDS-PAGE conditions to resolve either large or small polypeptides. Gels were aligned to show all polypeptides in the affinity-purified fractions and are delineated by the marking between them. The entire gel lanes shown from the two gels in (C) were dissected and each gel slice was subjected to mass spectrometry protein identification. The proteins identified are reported under the Hs PS column in Tables 1–3 and in Supplemental Table S1.

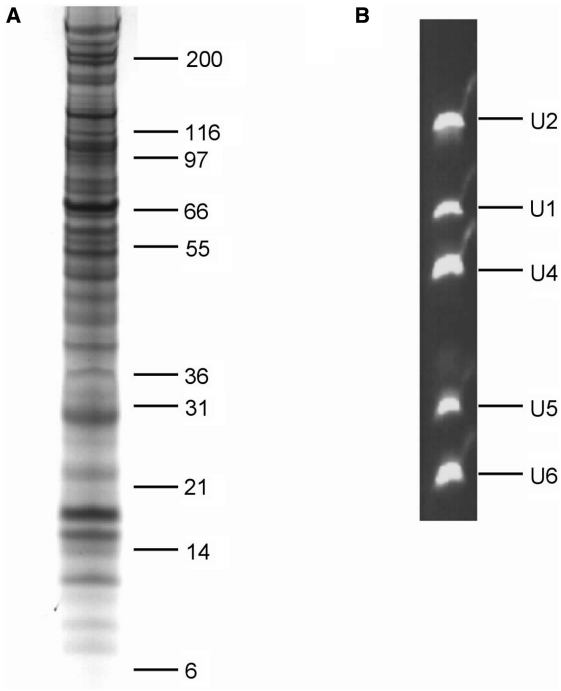

Figure 3. Chicken supraspliceosome-associated polypeptides and snRNAs. Fractions corresponding to the CLEP-tag purified, glycerol gradient-sedimented chicken supraspliceosomes were separated into protein (**A**) and RNA (**B**) fractions and electrophoresed through SDS-PAGE (A) or urea-PAGE (B) gels and stained with coomassie blue (protein) or ethidium bromide (RNA). The identities of the snRNAs are indicated on the right of panel B. The entire gel lane from (A) was dissected and each gel slice was subjected to mass spectrometry for protein identification. The proteins identified are reported under the Gg PS column in Tables 1–3 and in Supplemental Table S1.

even after extensive washing, indicating that the majority of the nuclear contents of this size are Sm-antigen-containing complexes. The lack of polyribosomes in the rapidly sedimenting material (as judged by the absence of 5S and 5.8S rRNAs and ribosomal proteins by mass spectrometry) indicates that the nuclei we prepare are not contaminated with cytoplasm. The material purified on a preparative scale was separated by SDS-PAGE gels of two compositions to provide optimal resolution of the large number of polypeptides present (Figure 2C). To demonstrate the specificity of these purifications, gradient-separated supraspliceosomes were subjected to affinity chromatography using identical beads and identical washing and elution conditions, but lacking the Y12 antibody. In Figure S1, we show that from the mock purification, there is no detectable coomassie-stained material in the resulting protein gel (panel A) and no snRNAs present (panel B).

Purification of endogenous chicken pre-mRNA processing complexes

Using our recently developed CLEP tagging procedure (22), we tagged the SmD3 polypeptide in chicken DT40 cells by introducing a TAP tag (24) at the native genomic locus. For each experiment, 61 of SmD3-TAP-DT40 cells were harvested and processed as described for purification of supraspliceosomes from HeLa cells. Affinity chromatography was performed according to the TAP procedure (24) and the TEV eluate was sedimented through a glycerol gradient. The material corresponding to the supraspliceosomes was isolated; proteins and nucleic acids are shown in Figures 3A and 3B, respectively. In Figure S1, we show the proteins (panel C) and RNA (panel D) resulting from an identical affinity purification procedure performed using extracts from untagged DT40 cells. The absence of proteins, beyond the contaminating TEV protease, and the absence of snRNAs indicates that the purification is specific and the proteins identified by mass spectrometry are likely to be *bona fide* supraspliceosome components. Additional confidence is provided in that there is size-selection as well as one or two steps of affinity chromatography.

RNA content of the supraspliceosomes

RNAs corresponding to the supraspliceosome fractions from human and chicken cells are shown in Figures 2B and 3B, respectively. Identities of the RNAs were confirmed by northern blotting (data not shown). The presence of all five spliceosomal snRNAs in this material

indicated that it contained a mixture of pre-mRNA splicing complexes in varying stages of assembly and activity, as both U1 and U4 snRNAs have been shown to be released from the spliceosome before the first catalytic step of the splicing reaction *in vitro* (31–33). Alternatively, the presence of both U1 and U4 may reflect functional differences between our preparations and the spliceosomes assembled *in vitro*; for example, it is possible that the U1 and U4 snRNAs do not completely dissociate in conjunction with catalytic activation *in vivo*, but are only destabilized and maintained locally. The presence of all five snRNAs in roughly equivalent amounts also lends experimental evidence to the participation of the penta-snRNP in these functional complexes (19).

Mass spectrometry analysis of supraspliceosome-associated polypeptides

Polyacrylamide gel lanes from the entire human and chicken supraspliceosome fraction were dissected and the material analyzed by tandem mass spectrometry (27–29). Though a wide variety of polypeptides were identified, it is notable that we detected very little background contamination of factors known to be unrelated to gene expression. In Tables 1–4, we categorize the identified polypeptides according to function. Remarkably, we detected 222 distinct polypeptides in the chicken supraspliceosomes, and 177 distinct polypeptides in the HeLa supraspliceosomes. These numbers are significantly higher than the number of polypeptides detected in any one of the three previously published spliceosome purifications (6–8).

The polypeptides identified by mass spectrometry were validated by analyzing the percent-coverage for each protein (Supplemental Table 1). Although there is a distribution of coverage for the identified proteins, we note that while many snRNP-associated factors had a large percentage of their sequence identified, some snRNP-associated proteins had <10% coverage. Differences in coverage may reflect differences in abundance, a paucity of appropriately-sized trypsin fragments or to peculiarities in the mass spectrometry detection of a particular peptide. The low-coverage of some of the novel polypeptides that may reasonably be implied to function in pre-mRNA processing (i.e. contain RNA binding motifs) may reflect their association with a smaller subset of pre-mRNAs than a general RNA binding protein such as an hnRNP. In the case of one novel factor, ZFR, the percent-coverage was low (3.4% for chicken, 8% for human) but its association with the splicing machinery was verified independently (see below).

Known snRNP-associated polypeptides

By mass spectrometry, we identified nearly all of the known pre-mRNA splicing snRNP-associated polypeptides (Table 1). We were initially surprised by the apparent absence in our preparations of a subset of snRNP-associated proteins found in most or all of the previously purified spliceosomes. However, upon closer inspection, we observed that for each polypeptide not represented in our mass spectrometry results, the inability to be detected correlated with the presence of an abundant hnRNP protein of similar molecular weight. The coverage of the major spliceosomal snRNP proteins was more complete for the chicken supraspliceosomes. Indeed, the CLEP tagging and purification procedure was sensitive enough to detect the presence of two minor AT–AC spliceosome components, the U11/U12–65K and U11/U12–48K polypeptides (34) in the chicken fractions, whereas no AT–AC specific splicing components were detected in the human complexes. The ability to detect all of the Sm proteins, but not all of the LSM proteins may reflect the difficulty in detecting all of these proteins in splicing complexes as shown previously (35,36), the 5-fold abundance differences of the two classes of proteins or perhaps due to the LSM proteins leaving the spliceosome during the process of pre-mRNA splicing (37).

By western blotting of chicken cell nuclear extracts with the Y12 monoclonal antibody, we determined that the Sm antigens in chicken cells were not reactive with the Y12 antibody (data not shown), eliminating the possibility of a direct comparison between Y12-immunopurified material from chicken and human cells.

snRNP biogenesis factors

Previous spliceosome purifications did not yield polypeptides known to be involved in snRNP biogenesis. We note in Table 1 that there are three of these present in the chicken supraspliceosomes (SIP1, SMNrp30 and Coilin). These factors, which are involved in the *de novo* assembly of snRNPs, are contained in Cajal bodies (CBs), nuclear organelles enriched for pre-mRNA splicing factors (38). We hypothesize that the CLEP tagging procedure may allow purification of a subset of snRNPs in the process of being re-targeted to the CBs. We did not detect this class of polypeptides in the HeLa supraspliceosomes. Other factors present in the chicken supraspliceosome but not those from HeLa cells were cyclophillins and chromatin remodeling proteins, which may be related to procedural differences in the purification methods.

Known spliceosome associated proteins (SAPs)

In the second half of Table 1, we compile a list of the 59 SAPs identified in one or more of the spliceosome purifications. The *in vivo*-assembled complexes contained pre-mRNA-interacting factors such as U2AF (39,40), PTB (41,42), and the cap binding complex proteins (43), which were present in some but not all of the *in vitro*-assembled spliceosomes. We detected the majority of PRP19-complex (NTC) related components as well, including homologues of Prp19p (44,45), Syf1p (45), Syf2p (45), Syf3/Clf1p (45,46), Isy1p (47), SKIP/Prp45p (48,49) and CDC5/Cef1p (50,51), which were also detected *in vitro*, and BCAS2/SPF27, which was only detected in supraspliceosomes assembled *in vivo*. Interestingly, our preparations included a number of polypeptides that are snRNP-associated in yeast but have not been identified in purified metazoan snRNPs or spliceosomes. Among these factors are putative orthologues of yeast Prp38p (52), Prp39p (53), Prp40p (54), Aar2p (55) and Luc7p (56).

Table 1. Comparisons of snRNP, snRNP biogenesis and known spliceosome associated proteins (SAPs) profiles from supraspliceosomes purified from human or chicken cells and mono-spliceosomes purified from three *in vitro*-assembled preparations

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
U1 snRNP							
ENSG00000104852	SNRP70	U1–70K	•	•	•	•	•
ENSG00000077312	SNRPA	U1A	•		•	•	•
ENSG00000124562	SNRPC	U1-C				•	•
U2 snRNP							
ENSGALG00000008038	SF3B1	SF3b155	•	•		•	•
ENSG00000087365	SF3B2	SF3b145	•	•	•	•	•
ENSGALG00000002531	SF3B3	SF3b130	•	•	•	•	•
ENSGALG00000000581	DDX42	SF3b125	•	•			
ENSGALG00000007937	SF3A1	SF3a120	•	•		•	•
ENSGALG00000021679	SF3A2	SF3a66	•	•	•	•	•
ENSGALG00000001540	SF3A3	SF3a60	•	•	•	•	•
ENSGALG00000013352	SF3B4	SF3b49	•		•	•	•
ENSGALG00000008729	SNRPB2	U2B″	•	•	•	•	•
ENSGALG00000007170	SNRPA1	U2A′	•	•	•	•	•
ENSGALG00000016501	–	SF3b14	•			•	
ENSGALG00000020000	SF3B5	SF3b10	•			•	
U2-snRNP associated							
ENSGALG00000014395	DHX15	PRP43/DDX15	•	•		•	•
ENSGALG00000002612	–	SR140	•	•		•	•
ENSGALG00000006332	RBM17	SPF45	•		•	•	•
ENSGALG00000008561	SMNDC1	SPF30	•		•	•	•
ENSGALG00000003824	CHERP	CHERP	•		•	•	
U5, U4/U6 & U4/U6•U5 snRNP							
ENSGALG00000002943	PRPF8	U5–220K	•	•		•	•
ENSGALG00000003477	ASCC3L1	U5–200K	•	•		•	•
ENSGALG00000000988	EFTUD2	U5–116K	•	•		•	•
ENSG00000175467	SART1	U4/U6•U5–110K	•	•		•	•
ENSGALG00000006001	C20ORF14	U5–102K	•	•		•	•
ENSG00000174243	DDX23	U5–100K	•	•	•	•	•
ENSGALG00000000465	PRPF3	U4/U6–90K	•	•	•	•	•
ENSG00000168883	USP39	U4/U6•U5–65K	•	•		•	•
ENSG00000105618	PRPF31	U4/U6•U5–61K	•	•		•	•
ENSGALG00000008857	PRPF4	U4/U6–60K	•	•	•	•	•
ENSGALG00000004874	PPIH	USA–CYP	•				•
ENSGALG00000000615	WDR57	U5–40K	•	•	•		
ENSGALG00000011931	NHP2L1	U4/U6•U5–15.5k	•			•	•
ENSGALG00000017396	TXNL4A	U5–15K	•				•
AT/AC							
ENSGALG00000005162	RNPC3	U11/U12–65K	•				
ENSGALG00000013005	C6ORF151	U11/U12–48K	•				
Sm/LSM							
ENSGALG00000007250	SNRPB	SmB/B'	•	•	•	•	•
ENSGALG00000011842	SNRPD1	SmD1	•	•		•	•
ENSG00000125743	SNRPD2	SmD2	•	•		•	•
ENSGALG00000006596	SNRPD3	SmD3	•	•		•	•
ENSGALG00000000137	SNRPE	SmE	•	•		•	•
ENSGALG00000011409	SNRPF	SmF	•	•		•	•
ENSG00000143977	SNRPG	SmG	•	•		•	•
ENSG00000111987	LSM2	LSM2	•			•	•
ENSG00000170860	LSM3	LSM3				•	•
ENSGALG00000003385	LSM4	LSM4	•			•	•
ENSG00000106355	LSM5	LSM5					
ENSGALG00000009985	LSM6	LSM6	•	•		•	•
ENSG00000130332	LSM7	LSM7				•	•
ENSGALP00000014820	LSM8	LSM8	•	•		•	
snRNP biogenesis							
ENSGALG00000010154	SIP1	SIP1	•				
ENSG00000119953	SMNDC1	SMNrp30	•				
ENSGALG00000003158	COIL	Coilin	•				
SAPs							
ENSGALG00000002514	SFPQ	PSF	•	•		•	

(Continued)

Table 1. Continued

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSGALG00000002060	FUS	TLS/FUS	•	•		•	•
ENSGALG00000012468	PRPF39	PRP39	•				
ENSGALG00000010501	SNW1	SKIP/PRP45	•	•	•	•	•
ENSGALG00000016704	CDC5L	CDC5	•	•	•		•
ENSGALG00000005012	RAB43	ISY1	•			•	•
ENSGALG00000008429	CRNKL1	CRN1	•	•	•	•	•
ENSGALG00000009257	PRLG1	Prp46/PRL1	•			•	•
ENSGALG00000015061	CDC40	CDC40/PRP17	•			•	•
ENSGALG00000002002	BCAS2	SPF27	•	•			
ENSGALG00000013919	PRPF19	PRP19	•	•	•	•	•
ENSGALG00000010627	PRPF38A	PRP38	•	•			
ENSGALG00000005507	NONO	p54nrb	•	•			
ENSGALG00000004555	RBM22	ECM2/RBM22	•			•	•
ENSGALG00000001247	SYF2	SYF2	•			•	•
ENSGALG00000000726	ELAVL1	ELAV/Hu	•	•		•	•
ENSGALG00000009001	–	CWC22	•			•	•
ENSGALG00000008149	EWSR1	EWSR1 (RBP)	•	•		•	
ENSGALG00000004705	BUD31	BUD31	•			•	•
ENSG00000076924	XAB2	SYF1	•	•		•	•
ENSGALG00000001962	PTBP1	PTB	•	•		•	
ENSGALG00000011857	LUC7L2	LUC7/CROP	•				•
ENSG00000196504	PRPF40A	PRP40	•	•			•
ENSGALG00000010167	PNN	Pinin	•	•			
ENSGALG00000012813	PRPF4B	Prp4K	•	•		•	•
ENSGALG00000000833	IK	RED	•	•		•	•
ENSGALG00000001500	–	SLU7	•			•	•
ENSG00000063244	U2AF2	U2AF65	•	•	•	•	•
ENSGALG00000016198	U2AF1	U2AF35	•	•	•	•	•
ENSG00000168066	SF1	SF1			•	•	•
ENSGALG00000001034	C20ORF4	AAR2	•				
ENSGALG00000005525	SFRS1	SF2/ASF	•	•		•	•
ENSG00000102241	HTATSF1	TAT-SF1					•
ENSGALG00000002087	NCBP1	CBC80	•	•		•	•
ENSGALG00000006843	NCBP2	CBC20	•	•		•	•
ENSG00000087087	–	ASR2B				•	•
ENSG00000100296	THOC5	KIAA0983	•	•		•	•
ENSG00000159086	C21ORF66	C21ORF66				•	•
ENSG00000126803	HSPA2	HSP70-2	•	•		•	•
ENSGALG00000006512	HSPA8	HSP71	•	•		•	•
ENSGALG00000009838	AQR	Aquarius	•	•		•	•
ENSGALG00000002014	SMU1	SMU1	•	•		•	•
ENSGALG00000005623	TFIP11	SPP382	•	•		•	•
ENSG00000137656	–	CWC26				•	•
ENSG00000126698	DNAJC8	SPF31			•	•	•
ENSG00000105705	SF4	SF4			•	•	
ENSG00000113649	TCERG1	CA150		•	•	•	•
ENSGALG00000004626	RBM5	E1B-AP5	•	•		•	
ENSG00000100056	DGCR14	DGCR14				•	•
ENSG00000105298	C19ORF29	C19ORF29				•	•
ENSG00000109536	FRG1	FRG1				•	
ENSG00000171824	EXOSC10	RRP6				•	
ENSG00000160799	CCDC12	CCDC12/CWF18				•	
ENSGALG00000011678	DNAJC13	DnaJ	•	•		•	
ENSG00000100813	ACIN1	Acinus		•		•	•
ENSG00000131051	RNPC2	HCC		•		•	•
ENSG00000084463	WBP11	WBP11		•		•	
ENSG00000196419	XRCC6	Ku70		•		•	

[a]Data available at: http://www.ensembl.org.
[b]HUGO Gene Nomenclature Committee designation, a dash represents a polypeptide which has not yet been assigned a systematic name.
[c]Common name.
[d]*Gallus gallus* supraspliceosome.
[e]*Homo sapiens* (HeLa) supraspliceosome.
[f]Neubauer *et al*. spliceosome data (8).
[g]Rappsilber *et al*. spliceosome data (6).
[h]Zhou *et al*. spliceosome data (7).
[f,g,h]Were compiled from Jurica and Moore (104).

Table 2. Polypeptides demonstrated or predicted by sequence homology to interact with the pre-mRNA, mRNA or the spliceosome and comparisons of those identified in the supraspliceosomes with those of spliceosomes formed *in vitro*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
RNA Helicase-like							
ENSGALG00000016461	DDX1	DDX1	•	•			
ENSGALG00000016231	DDX3X	DDX3	•	•			•
ENSGALG00000003532	DDX5	DDX5/p68	•		•	•	•
ENSGALG00000012247	DDX17	DDX17/p72	•	•		•	•
ENSGALG00000012147	DDX18	DDX18	•				
ENSG00000102786	DDX26	DDX26/HDB				•	
ENSGALG00000006974	DDX27	DDX27	•				
ENSGALG00000003030	DDX41	DDX41/ABSTRAKT	•			•	•
ENSG00000123136	DDX39	DDX39		•			
ENSG00000145833	DDX46	DDX46/PRP5		•		•	•
ENSGALG00000008530	DDX48	DDX48	•	•		•	•
ENSGALG00000004144	DDX50	DDX50/Gu-ß	•	•			
ENSGALG00000001186	DHX8	DHX8/PRP22	•	•		•	•
ENSG00000135829	DHX9	DHX9/HELICASEA	•	•		•	•
ENSG00000137333	DHX16	DHX16/PRP2				•	•
ENSGALG00000005027	DHX30	DHX30	•	•			
ENSGALG00000003658	DHX35	DHX35	•				
ENSG00000174953	DHX36	DHX36		•			
ENSGALG00000014709	SKIV2L2	SKIV2L2	•	•		•	•
ENSG00000198563	BAT1	UAP56	•	•			•
hnRNP							
ENSGALG00000006160	HNRNPA0	hnRNPA0	•	•			
ENSGALG00000011036	HNRNPA2B1	hnRNPA2/B1	•	•	•	•	•
ENSG00000135486	HNRPA1	hnRNPA1	•	•	•	•	•
ENSGALG00000009250	HNRNPA3	hnRNPA3	•	•		•	•
ENSGALG00000014381	HNRNPAB	hnRNPAB	•	•			
ENSG00000092199	HNRPC	hnRNPC1/C2	•	•	•	•	•
ENSG00000138668	HNRNPD	hnRNPD0/AUF1	•	•			
ENSGALG00000011184	HNRNPDL	hnRNPD0	•	•		•	•
ENSG00000169813	HNRPF	hnRNPF				•	
ENSGALG00000006457	RBMX	hnRNPG	•	•		•	
ENSGALG00000005955	HNRNPH1	hnRNPH1	•	•		•	
ENSGALG00000003947	HNRNPH3	hnRNPH3	•	•			
ENSGALG00000012591	HNRNPK	hnRNPK	•	•		•	•
ENSG00000104824	HNRNPL	hnRNPL	•	•		•	•
ENSGALG00000000377	HNRNPM	hnRNPM	•	•	•	•	
ENSGALG00000015830	SYNCRIP	hnRNPQ	•	•		•	•
ENSGALG00000000814	HNRNPR	hnRNPR	•	•		•	•
ENSGALG00000010671	HNRNPU	hnRNPU	•	•		•	
ENSGALG00000018665	–	hnRNP novel	•				
ENSG00000126457	HRMT1L2	HRMT1L2	•				
SR family							
ENSG00000133226	SRRM1	SRm160				•	•
ENSG00000167978	SRRM2	SRm300				•	•
ENSGALG00000005525	SFRS1	SF2p33	•				
ENSG00000161547	SFRS2	SC35				•	•
ENSGALG00000000533	SFRS3	SFRS3SRp20	•	•		•	•
ENSG00000116350	SFRS4	SRp75					•
ENSGALG00000009484	SFRS5	SRp40	•				•
ENSGALG00000000990	SFRS6	SRp55	•	•		•	•
ENSGALG00000013825	SFRS7	9G8	•	•		•	•
ENSGALG00000002487	SFRS8	SFRS8	•				
ENSG00000111786	SFRS9	SRp30		•		•	•
ENSGALG00000006531	SFRS10	SFRS10	•	•			•
ENSG00000116754	SFRS11	SRp54				•	•
ENSG00000153914	SFRS12	SFRS12		•			
ENSGALG00000004133	FUSIP1	SRrp35	•	•			
Cyclophillins							
ENSGALG00000013383	PPIE	CYP-E	•			•	•
ENSGALG00000004874	PPIH	USA-CYP	•				
ENSGALG00000014747	SDCCAG10	CYP16	•			•	
ENSG00000137168	PPIL1	PPIL1/CWF27	•			•	•
ENSG00000100023	PPIL2	PPIL2/CYP60	•			•	•
ENSG00000115934	PPIL3	PPIL3B				•	•
ENSG00000113593	PPWD1	PPWD1	•			•	•

(Continued)

Table 2. Continued

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
RBP							
ENSG00000102317	RBM3	RBM3		•			
ENSGALG00000018992	RBM4B	RBM4B/Lark	•				
ENSGALG00000004626	RBM5	RBM5	•			•	
ENSGALG00000007017	RBM7	RBM7	•			•	
ENSG00000131795	RBM8A	RBM8B	•	•			
ENSGALG00000013411	RBM14	RBM14	•	•			
ENSG00000162775	RBM15	RBM15		•		•	•
ENSGALG00000002372	RBM15B	RBM15B	•	•			
ENSG00000197676	RBM16	RBM16		•			
ENSGALG00000004555	RBM22	RBM22	•				
ENSG00000119707	RBM25	RBM25		•			
ENSG00000091009	RBM27	RBM27		•			
ENSGALG00000011038	CBX3	RNPS1	•				•
ENSG00000033030	ZCCHC8	ZFP8		•		•	
ENSGALG00000006113	ZNF326	ZFP326	•	•			
ENSG00000179950	–	PUF60		•		•	•
ENSG00000197381	ADARB1	ADAR1		•			
ENSG00000160710	ADAR	ADAR2		•			
ENSGALG00000010952	–	Requiem	•				
ENSGALG00000011570	ILF2	NFAT45	•	•		•	•
ENSG00000129351	ILF3	NFAT90	•	•		•	•
ENSG00000169564	PCBP1	PolyrCBP	•				
ENSGALG00000012225	CIRBP	CIRBP	•	•			
ENSG00000056097	ZFR	ZFR	•	•			
ENSGALG00000009427	TIAL1	TIA1	•				
ENSGALG00000001187	STRBP	STRBP	•				
ENSG00000136231	IGF2BP3	IMP3	•	•		•	•
ENSG00000060138	CSDA	CSDA				•	•
ENSG00000121774	KHDRBS1	SAM68		•			
ENSG00000126254	-	Novel RRM		•			
ENSG00000132773	TOE1	TOE1		•			
ENSG00000142864	SERBP1	SERBP1		•			
Export/transcription/NMD							
ENSGALG00000014915	THOC1	THO1/HPR1	•	•			
ENSGALG00000008507	THOC2	THO2	•	•		•	•
ENSG00000051596	THOC3	THO3/TEX1	•	•			
ENSGALG00000007237	THOC4	THO4/ALY	•	•		•	•
ENSGALG00000004571	PPP1CA	GLC7/PPP1CA	•				
ENSGALG00000002569	RAN	Ran	•	•			
ENSG00000119392	GLE1L	GLE1L		•			
ENSGALG00000007653	RAE1	GLE2/RAE1	•	•			
ENSGALG00000003220	RENT1	UPF1/RENT	•				
ENSG00000131795	RBM8A	Y14/RBM8A	•	•			•
ENSGALG00000002144	THRAP3	TRAP150	•	•			
ENSG00000172660	TAF15	TAF15/RBP56	•				
ENSG00000162231	NXF1	TAP		•			
ENSG00000065978	YBX1	YBX1		•		•	•
ENSGALG00000010689	MAGOH	Mago nashi	•	•			•
3′ end proc.							
ENSG00000071894	CPSF1	CPSF1				•	
ENSGALG00000010783	CPSF2	CPSF2	•				
ENSGALG00000016424	CPSF3	CPSF3	•	•			
ENSGALG00000004714	CPSF4	CPSF4	•				
ENSGALG00000003084	CPSF5	CPSF5		•		•	•
ENSG00000111605	CPSF6	CPSF6				•	•
ENSGALG00000011685	CSTF3	CSTF-77	•	•			
ENSGALG00000013943	FIP1L1	FIP1	•				
ENSG00000172239	PAIP1	PAIP1		•			
ENSG00000100836	PABPN1	PABPN1		•	•	•	•
ENSGALG00000003800	PABPC4	PABPC4		•			

[a]Data available at: http://www.ensembl.org.
[b]HUGO Gene Nomenclature Committee designation, a dash represents a polypeptide which has not yet been assigned a systematic name.
[c]Common name.
[d]*Gallus gallus* supraspliceosome;
[e]*Homo sapiens* (HeLa) supraspliceosome.
[f]Neubauer *et al*. spliceosome data (8).
[g]Rappsilber *et al*. spliceosome data (6).
[h]Zhou *et al*. spliceosome data (7).
[f,g,h]Were compiled from Jurica and Moore (104).

Table 3. Structural, nucleoporin and novel polypeptides present in the supraspliceosomes and *in vitro*-assembled splicing complexes

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
Structural							
ENSGALG00000012533	MYH9	Myosin	•	•			
ENSGALG00000009126	TTN	Titin	•				
ENSGALG00000001381	ACTG1	Actin	•	•			
ENSGALG00000002478	MATR3	MATRIN3	•	•		•	•
ENSGALG00000008677	VIM	Vimentin	•	•			
ENSG00000117245	KIF17	Kinesin KIF17	•				
ENSGALG00000002197	NPM1	NUMATRIN	•	•		•	
ENSGALG00000014692	LMNB1	Lamin B	•	•			
ENSGALG00000013505	SYNE1	NuSpectrin	•	•			
ENSG00000140259	MFAP1	MFAP1				•	•
Chromatin modification							
ENSGALG00000000360	ARID1A	ARID1A-SWI/SNF	•				
ENSGALG00000013683	ARID1B	ARID1B-SWI/SNF	•				
ENSGALG00000010164	SMARCA2	SMARCA2	•				
ENSG00000127616	SMARCA4	Brahma/SMARCA4	•				
ENSGALG00000009913	SMARCA5	SMARCA5	•				
ENSGALG00000005983	SMARCB1	SMARCB1	•				
ENSGALG00000005048	SMARCC2	BRG1-SWI/SNF	•				
ENSGALG00000005048	SMARCC1	SMARCC1	•				
ENSGALP00000010010	SMARCD1	SMARCD1	•				
ENSGALG00000000363	SMARCD2	SMARCD2	•				
ENSGALG00000002100	SMARCE1	SMARCE1	•				
Nucleoporins							
ENSGALG00000005714	PKD1	PKD1 (NUP assoc)	•				
ENSGALG00000003830	NUP214	NUP214	•	•			
ENSGALG00000012720	NUP153	NUP153	•				
ENSGALG00000005078	NUP210	NUP210	•	•			
ENSG00000102900	NUP93	NUP93		•			
ENSG00000108559	NUP88	NUP88		•			
ENSG00000111581	NUP107	NUP107		•			
ENSG00000110713	NUP98	NUP98		•			
ENSG00000138750	NUP54	NUP54		•			
ENSG00000163002	NUP35	NUP35		•			
ENSG00000069248	NUP133	NUP133		•			
ENSG00000155561	NUP205	NUP205		•			
Novel or unknown to splicing							
ENSGALG00000011351	HSP90AA1	HSP90α	•	•			
ENSGALG00000010175	HSP90AA2	HSP90ß	•	•			
ENSGALG00000012726	HSP90B1	HSP108	•	•			
ENSGALG00000009967	LRPPRC	LRP130	•				
ENSGALG00000003693	MACF1	Macrophin	•				
ENSGALG00000007705	NCL	Nucleolin	•				
ENSGALG00000015933	C21ORF66	GCRBF	•				
ENSGALG00000008454	–	NOP58	•				
ENSGALG00000009061	ACTL6A	BAF53A	•				
ENSGALG00000007520	SSRP1	FACT80	•	•			
ENSGALG00000015821	CCT8	TCP1-theta	•				
ENSGALG00000014500	NOL5A	Nol5A/NOP56	•				
ENSGALG00000005624	MED12	TRAP230	•				
ENSGALG00000010973	TRA2A	TRA2α	•	•			
ENSGALG00000008372	XRN2	RAT1	•	•			
ENSG00000197157	–	SND1	•	•			
ENSGALG00000001948	SAFB2	SAFB/HSP27	•	•			
ENSGALG00000003177	BRD8	BRD8	•	•			
ENSGALG00000010699	–	FOG	•				
ENSGALG00000005177	C9ORF10	C9ORF10	•	•			
ENSGALG00000002653	–	ELG	•	•			
ENSGALG00000016949	WBP4	WBP4	•				
ENSGALG00000004133	FUSIP1	Fus IP	•	•			
ENSGALG00000017384	ERH	ERH	•	•			
ENSG00000079246	XRCC5	Ku80		•			
ENSG00000182562	ATAD3A	ATAD3A (AAA ATPase)	•	•			
ENSGALG00000001515	ATAD3B	ATAD3B (AAA ATPase)	•	•			
ENSG00000108588	CCDC47	CCDC47		•			

[a]Data available at: http://www.ensembl.org.
[b]HUGO Gene Nomenclature Committee designation, a dash represents a polypeptide which has not yet been assigned a systematic name.
[c]Common name.
[d]*Gallus gallus* supraspliceosome.
[e]*Homo sapiens* (HeLa) supraspliceosome.
[f]Neubauer *et al*. spliceosome data (8).
[g]Rappsilber *et al*. spliceosome data (6).
[h]Zhou *et al*. spliceosome data (7).
[f,g,h]Were compiled from Jurica and Moore (104).

Table 4. Known pre-mRNA processing proteins not present in any purified splicing complex

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]
ENSG00000095485	CWF19L1	CWF19
ENSG00000152404	CWF19L2	CWF19
ENSG00000165630	PRPF18	PRP18
ENSG00000140829	DHX38	PRP16
ENSG00000149532	–	CFI-59K
ENSG00000165494	PCF11	PCF11
ENSG00000172409	CLP1	CLP1
ENSG00000111880	HCAP1	HCE/CEG1
ENSG00000146007	ZMAT2	SNU23
ENSG00000108296	CCDC49	CWC25
ENSG00000101138	CSTF1	CSTF-50
ENSG00000101811	CSTF2	CSTF-64
ENSG00000161981	C16ORF33	U11/U12–25K
ENSG00000184209	–	U11/U12–35K

[a]Data available at: http://www.ensembl.org.
[b]HUGO Gene Nomenclature Committee designation, a dash represents a polypeptide which has not yet been assigned a systematic name.
[c]Common name.

Other polypeptides exclusively contained in the purified supraspliceosomes include, NONO/p54nrb, PNN/Pinin, CWC22, which have been previously implicated in splicing (57–60). Conversely, 13 polypeptides were exclusively associated with the *in vitro* purified spliceosomes, possibly reflecting specificity for the pre-mRNAs upon which these were assembled. Alternatively, the differences in composition may result from differences in the procedures. Most striking, however, is the absence of SF1/BBP from our supraspliceosome preparations. This may be due to the fact that SF1/BBP interacts very early with the pre-mRNA, remains associated for a short time and is replaced by U2 snRNP at the branchpoint sequence (61).

RNA helicase-like proteins

We identified a large number of DExH/D proteins in the preparations of supraspliceosome assembled *in vivo*. In addition to the RNA helicase-like polypeptides known to function in pre-mRNA splicing, such as DDX15/Prp43p (62–64), DHX8/Prp22p (65–67), DDX46/Prp5p (68–70), DDX5/p68 and DDX17/p72 (71), UAP56/Sub2p (72–74), DDX16/Prp2p (75,76) and the snRNP-associated helicases U5-200K/Brr2p (77–79) and U5-100K/Prp28p (80–82), we noted a number not previously implicated in pre-mRNA splicing and absent from the *in vitro* purified spliceosomes [Table 2 ('Gg' and 'Hs' to 'N', 'R' and 'Z')]. These include 13 additional polypeptides with sequence motifs indicative of DEAD, DEAH or Ski2p-like helicase family members, most of which were absent from spliceosomes assembled *in vitro*. Although we do not as yet have evidence that these proteins function in pre-mRNA splicing, the complete absence of DNA helicases in our preparations indicates a specificity (i.e. a specificity for RNA processing complexes and against chromatin components), which minimally suggests that they function in some aspect of RNA Pol II transcript processing.

Supraspliceosome-associated hnRNPs

Polypeptides termed hnRNPs are highly abundant nuclear proteins known to interact with hnRNA. We detected virtually all of the known hnRNP proteins in both human and chicken cells, as well as other hnRNP-like proteins annotated in genome databases (Table 2). We note that in the spliceosomes purified from *in vitro* extracts, some hnRNPs were identified; however, perhaps owing to specific binding of some hnRNPs to the bulk hnRNA and not the single transcript used in the *in vitro* spliceosome assembly reactions, a greater number of these polypeptides were detected in the supraspliceosomes.

SR proteins

Many splicing factors are rich in arginine and serine residues including long stretches of alternating dipeptides termed SR domains. These factors function at multiple steps in the pre-mRNA splicing pathway (83), and were constituents of the *in vitro*-purified splicing complexes. We detected 10 different SR family members in our purified supraspliceosomes from both chicken and human cells (Table 2), though not a complete set. One possible conclusion is that, due to the means by which these complexes were purified and analyzed, these SR proteins represent the major SR proteins functioning in these cells, and that those not detected in our preparations function in the splicing of a smaller subset of pre-mRNAs.

Cyclophillins

In addition to the known snRNP-associated USA–CYP (84,85), we detected five additional potential proline *cis-trans* isomerases co-purifying with spliceosomes from chicken, but interestingly, not from HeLa cells. Several of those from the chicken purification were also present in spliceosomes assembled *in vitro* (Table 3). As it is likely that these proteins also function in HeLa cells, their absence may represent operational differences in the ways in which the chicken and human cells were handled and the ways in which the complexes were purified and analyzed.

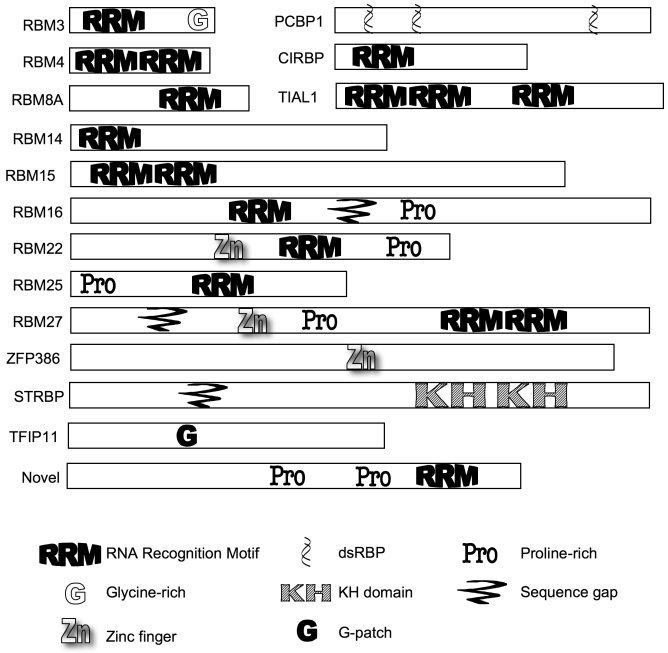

Figure 4. Novel spliceosome-associated polypeptides with predicted RNA binding motifs. Polypeptides from Table 2 with no known function in the pre-mRNA processing pathway are shown with graphical representations of the various RNA interaction or other noted motifs listed at the bottom of the Figure.

Other RNA binding proteins

There were 32 polypeptides identified among all of the splicing complex purifications possessing sequence homology to polypeptides believed to interact with RNA by virtue of containing RNA Recognition Motifs (RRM), double-stranded RNA binding domains (dsRBD) or other motifs implicated in RNA binding. Some were identified previously, such as the ELAV/Hu protein that binds AU-rich elements in both cytoplasmic (86) and nuclear RNAs (87), the U2AF-related PUF60 protein (88), the dsRBD-motif-containing NFAT45 and NFAT90 and RNA adenosine deaminases; these were previously shown to exist in large nuclear complexes (16) and believed to function in RNA Pol II transcript metabolism. Only a single predicted RNA binding protein was found in the *in vitro*-assembled splicing complexes but not in either the chicken or HeLa supraspliceosomes, while 22 were exclusively found in supraspliceosomes, but not in the *in vitro*-formed complexes. We believe this is most likely due to the use of a single pre-mRNA *in vitro*, while a broader spectrum of RNA binding proteins will be associated with bulk pre-mRNA. In Figure 4 we present a graphical representation of the 16 presumptive RNA binding proteins novel to our study and highlight the sequence motifs contained in each.

Non-spliceosomal pre-mRNA processing factors

Cap-binding proteins (CBC80 and CBC20) are present in both spliceosomes assembled *in vitro* and supraspliceosomes assembled *in vivo* (Table 1). In Table 2, we report the presence in our preparations of many 3′ end processing factors (CSPF, CSTF and poly-A binding proteins), a comprehensive set of proteins shown to be involved in mRNA export including the TREX complex (THO1/HPR1, THO2, THO3/TEX1, UAP56 and ALY) (89), export factors such as TAP (90,91), GLE1 (92), GLE2 (93) and GLC7 (94), and exon junction complex constituents including Y14 and Magoh. We also note that a single component of the nonsense-mediated decay (NMD) pathway (UPF1) (95) was identified in the chicken material. As NMD is likely to be active only in a very small subset of pre-mRNA processing complexes (96), we were surprised to observe even a single polypeptide implicated in this process.

Nuclear matrix and filament proteins

Recent data from several laboratories suggest a functional interaction between the structural proteins of the nuclear matrix and the gene expression machinery (97). Consistent with this model, we detected a number of nuclear matrix proteins in our endogenously formed pre-mRNA splicing complex preparations including actin, spectrin, matrin3, numatrin, lamin B and a matrix associated protein MAP1. Although we cannot confirm the functional relevance of these associations, we note that a few structural proteins also co-purified with *in vitro*-assembled spliceosomes. We also note that a number of hnRNPs and other known splicing factors such as Prp19p (98) were initially termed nuclear matrix-associated proteins, indicating an intimate relationship between the pre-mRNA processing machinery and the nuclear matrix. Indeed, it is an attractive hypothesis that pre-mRNA and mRNA are trafficked to the nuclear pore via the nuclear matrix.

Nuclear pore complex proteins

A substantial number of nucleoporins (NUPs) are present in the purified supraspliceosome complexes from human cells but not in spliceosomes assembled *in vitro*, which may perhaps be due to our use of sonication to release complexes from the purified nuclei versus salt extraction for preparation of splicing extracts. In the chicken supraspliceosomes, we detected a smaller set of NUPs, which may be due to their release by the detergent NP40 present during purification of these complexes. NP40 was absent during the purification of the human complexes, which likely maintained the integrity of hydrophobic interactions believed to stabilize the interaction of export complexes with NUPs.

Polypeptides novel to pre-mRNA splicing–SWI/SNF proteins and associated factors

A recent report from Muchardt and colleagues (99) has demonstrated that the SWI/SNF component Brahma/SMARCA4 (Brm) associates with the splicing apparatus and its presence favors the inclusion of alternatively spliced exons. The Prp4 kinase, which is present in both the human and chicken supraspliceosomes, has been reported to phosphorylate both Brm and the splicing factor U5–102K/hPrp6 (100) providing further evidence that it functions in Pol II transcript maturation. In the purified chicken supraspliceosomes, Brahma/SMARCA4,

and a number of other SWI/SNF-related polypeptides were identified (Table 3), all with high degrees of confidence given the depth of the peptide identification. As our mass spectrometry data neither included structural proteins of chromatin- such as histones, nor the DNA replication machinery or other DNA binding proteins, Brahma and other polypeptides with chromatin-related functions must specifically associated with the pre-mRNA processing complexes.

Novel polypeptides present in native supraspliceosomes

Our mass spectrometry peptide data revealed several novel and intriguing polypeptides in the supraspliceosome complexes (Table 3). We found chicken homologs of the yeast splicing factors Prp38p, Prp39p and Aar2p, previously unannotated in purified pre-mRNA splicing complexes. The other polypeptides of interest in the endogenous splicing complexes include the 5′ to 3′ exonuclease XRN2/Rat1p (101–103), which has been implicated in linking transcription termination with polyadenylation. XRN2/Rat1p was found in both the human and chicken preparations, as were two uncharacterized AAA ATPases, ATAD3A and ATAD3B. The identities of 25 additional novel polypeptides are reported in Table 3.

Co-immunoprecipitation of spliceosomal components by antiserum directed against a novel polypeptide

To demonstrate the authenticity and functional relevance of a novel polypeptide that co-purified with endogenous spliceosomes, we generated antiserum against ZFR (Table 2) and used it to specifically immunopurify ZFR-associated components. In Figure 5A, we show that the ZFR polypeptide is present in very high molecular weight complexes that co-migrate with supraspliceosomal material. In Figure 5B, we show that the anti-ZFR antiserum, but not the pre-immune serum or the Protein-A beads, immunoprecipitates the U1, U2, U4, U5 and U6 snRNAs. As a positive control, we showed that antiserum directed against the known spliceosomal protein SR140, prepared and analyzed under identical conditions, also immunoprecipitated all of the snRNAs. We also tested for the presence of another pre-mRNA splicing factor, hPrp43 (DHX15), in the material immunopurified with anti-ZFR; Figure 5C shows that the specific antiserum, but not the pre-immune serum or the Protein-A beads, immunoprecipitates hPrp43p/DDX15. This demonstrates that the novel spliceosome-associated factor ZFR is indeed associated with spliceosomal snRNAs and other spliceosomal proteins.

We note that there are snRNA stoichiometry differences between the two immunoprecipitations. Although we cannot be certain of exactly why this is, it may be that the ZFR polypeptide is associated with a number of different complexes, as is almost certain from its distribution in the glycerol gradient. As the immunoprecipitations were performed using nuclear extracts, and not size-selected complexes, the snRNA representation reflects the ZFR-associated material from the entire nucleus and not solely from the polyspliceosomes.

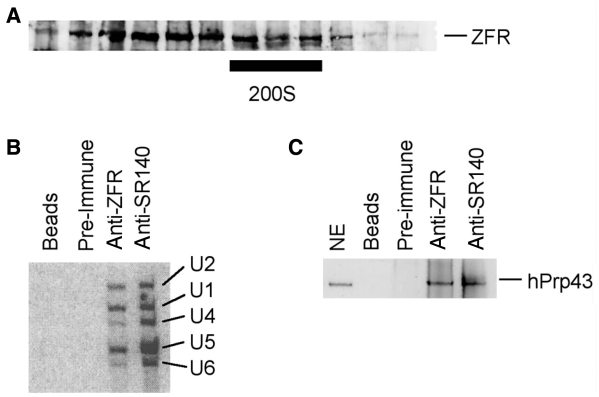

Figure 5. The novel Zn finger protein ZFR is a *bona fide* spliceosomal component. (**A**) ZFR sediments with the 200S particle in glycerol gradients. HeLa nuclear extract was subjected to glycerol velocity gradient sedimentation analysis as in Figure 2. Proteins from the indicated fractions were electrophoresed through SDS-PAGE gels and subjected to western blot analysis using anti-ZFR antiserum. The bar below the gel denotes the 200S region. (**B**) ZFR is specifically associated with spliceosomal snRNAs. Equal amounts of HeLa nuclear extract were incubated with protein-A beads (beads), pre-immune serum and protein-A beads (pre-immune), anti-ZFR antiserum and protein-A beads (anti-ZFR) or anti-SR140 antiserum and protein-A beads (anti-SR140) according to the Materials and Methods. Recovered nucleic acids were subjected to northern blot analysis and probed with antisense probes to human snRNAs (identities noted to the right of the Figure). (**C**) ZFR is specifically associated with complexes containing spliceosomal proteins. Immunoprecipitation conditions and lanes are as described in (B). Proteins were subjected to western blot analysis using hPrp43 antiserum.

In this work, we report the composition of the endogenous pre-mRNA processing machinery from human and chicken cells and provide a comparison between native supraspliceosome complexes and spliceosomes assembled on a model single-intron substrate *in vitro* from salt-extracted nuclear material. These supraspliceosomes have recently been shown to be functional in add-back experiments using micrococcal nuclease-treated extracts for *in vitro* splicing by Sperling and colleagues (21), further enhancing the functional relevance of our findings. In addition to confirming the set of factors known to interact with Pol II transcripts during splicing, we discovered an extensive array of novel factors by purifying supraspliceosomes from two types of vertebrate cells. Many of these have been implicated in pre-mRNA maturation including a subset of the SWI/SNF chromatin remodeling complex proteins, recently shown to influence alternative splicing patterns and to co-purify with pre-mRNA splicing factors. The novel polypeptides discovered in the endogenous complexes will provide a rich source of new proteins to investigate, ultimately enhancing our understanding of this incredibly complex macromolecular machine.

Comparison with *in vitro*-assembled spliceosomes

In Tables 1–3 we present a comparison of the polypeptides present in the supraspliceosomes purified in this work with those purified in previous *in vitro* spliceosome preparations. The core machinery (snRNPs, SAPs,

SR proteins, etc.) is well represented in the material derived from all of the purification schemes. However, a great many other factors are present in all of the preparations as well, highlighting the amazing number of proteins required to remove even the single intron used in the *in vitro* experiments. The methods used for the purification of all of the complexes represented in Tables 1–3 were operationally distinct and the *in vivo*-assembled spliceosomes contained a larger number of proteins than did the spliceosome preparations formed *in vitro*.

What is perhaps most remarkable about our results is the fact that, despite the operationally distinct purification strategies, the basal pre-mRNA processing machinery required to effect the removal of a single intron *in vitro* is not vastly different than that purified from complex mixtures of all of the pre-mRNAs in a vertebrate or human nucleus. The major differences in composition between the previous purifications and the one described herein involve (i) polypeptides predicted by sequence homology to interact with the pre-mRNA (ii) the depth of coverage for polypeptides involved in export and 3′ end processing and (iii) polypeptides that may require that the pre-mRNA in these complexes follow the path of RNA Pol II transcription and nuclear trafficking, such as the SWI/SNF complexes, structural proteins and NUPs.

Pre-mRNA processing factors not present in any spliceosome purification

To complete the catalog of polypeptides that participate in the nuclear pre-mRNA processing pathway, we have compiled a list of factors know to function in processing of RNA Pol II transcripts, but not present in any of the five spliceosome preparations listed in Tables 1–3. In Table 4, we outline this relatively short list of 14 factors. Table 4 does not include factors implicated in yeast splicing, but for which no identifiable human or vertebrate homologue in the genomic databases. By adding together all of the polypeptides listed in Tables 1–4, we arrive at an estimate of at least 305 for the number of proteins that co-purify with endogenous pre-mRNA splicing complexes. This is by far the largest cataloguing of factors potentially required for or participating in this process to date.

Two possible classes of polypeptides may exist that are not detected in our preparations. First are those that are underrepresented because they may interact with only a small number of pre-mRNAs, such as intron- or exon-specific binding proteins. Other classes of proteins which may participate in pre-mRNA splicing but as absent from our analyses might include tissue-type developmental stage-specific factors which would not be present in our bulk supraspliceosome preparations due to the use of only two cell types. To date, such factors have generally been elucidated via genetic or molecular strategies focused on individual pre-mRNAs. However, with the introduction of the CLEP tagging technology to more cell types (22), it may be possible to rapidly enumerate factors that function in regulated or alternative splicing.

Although we cannot completely eliminate the possibility that there may be contaminants present in our preparations, the basic strategy we adopted is validated by the presence proteins such as Brahma that were not detected in the previously characterized *in vitro*-assembled spliceosomes yet have clearly been implicated in splicing through other means. Our analyses in aggregate greatly expand our knowledge of the protein factors that function both in basal and regulated splicing in vertebrate cells.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Jo Ann Wise for critically reviewing the manuscript, contributions of the other members of the Stevens Laboratory, Joan Steitz for the gift of Y12 antibodies, Cindy Will and Reinhard Lührmann for the hPrp43 antiserum, Phil Tucker and William Kuziel for thoughtful discussions, Hank Bose and the Bose laboratory members for advice on the DT40 system and use of their facilities and Gwen Gage for graphic art expertise. This work was supported by grants from the Welch Foundation (F-1564 to SWS), the National Science Foundation (MCB-0448556 to SWS), the American Cancer Society (RSG-05-137-01-MCB to SWS) and from the NIH (NIH P30 CA33572 to TDL). Funding to pay the Open Access publication charges for this article was provided by NSF (MCB-0448556 to SWS).

Conflict of interest statement. None declared.

REFERENCES

1. Maniatis,T. and Reed,R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499–506.
2. Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R. F. and Atkins,J. F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.
3. Shatkin,A.J. and Manley,J.L. (2000) The ends on the affair: Capping and polyadenylation. *Nat. Struct. Biol.*, **7**, 838–842.
4. Reed,R. and Magni,K. (2001) A new view of mRNA export: Separating the wheat from the chaff. *Nat. Cell Biol.*, **3**, E201–E204.
5. Tange,T.O., Nott,A. and Moore,M.J. (2004) The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.*, **16**, 279–284.
6. Rappsilber,J., Ryder,U., Lamond,A.I. and Mann,M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.*, **12**, 1231–1245.
7. Zhou,Z., Licklider,L.J., Gygi,S.P. and Reed,R. (2002) Comprehensive analysis of the human spliceosome. *Nature*, **419**, 182–185.
8. Neubauer,G., King,A., Rappsilber,J., Calvio,C., Watson,M., Ajuh,P., Sleeman,J., Lamond,A. and Mann,M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.*, **20**, 46–50.
9. Azubel,M., Wolf,S.G., Sperling,J. and Sperling,R. (2004) Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol. Cell*, **15**, 833–839.
10. Muller,S., Wolpensinger,B., Angenitzki,M., Engel,A., Sperling,J. and Sperling,R. (1998) A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. *J. Mol. Biol.*, **283**, 383–394.

11. Sperling,R., Sperling,J., Levine,A.D., Spann,P., Stark,G.R. and Kornberg,R.D. (1985) Abundant nuclear ribonucleoprotein form of CAD RNA. *Mol. Cell. Biol.*, **5**, 569–575.
12. Wassarman,D.A. and Steitz,J.A. (1993) A base-pairing interaction between U2 and U6 small nuclear RNAs occurs in >150s complexes in HeLa-cell extracts - implications for the spliceosome assembly pathway. *Proc. Natl Acad. Sci., USA*, **90**, 7139–7143.
13. Sperling,R., Koster,A.J., MelamedBessudo,C., Rubinstein,A., Angenitzki,M., BerkovitchYellin,Z. and Sperling,J. (1997) Three-dimensional image reconstruction of large nuclear RNP (lnRNP) particles by automated electron tomography. *J. Mol. Biol.*, **267**, 570–583.
14. Spann,P., Feinerman,M., Sperling,J. and Sperling,R. (1989) Isolation and visualization of large compact ribonucleoprotein-particles of specific nuclear RNAs. *Proc. Natl Acad. Sci. USA*, **86**, 466–470.
15. Yitzhaki,S., Miriami,E., Sperling,R. and Sperling,J. (1996) Phosphorylated Ser/Arg-rich proteins: Limiting factors in the assembly of 200S large nuclear ribonucleoprotein particles. *Proc. Natl Acad. Sci. USA*, **93**, 8830–8835.
16. Raitskin,O., Cho,D.S.C., Sperling,J., Nishikura,K. and Sperling,R. (2001) RNA editing activity is associated with splicing factors in InRNP particles: The nuclear pre-mRNA processing machinery. *Proc. Natl Acad. Sci. USA*, **98**, 6571–6576.
17. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
18. Sakharkar,M.K., Chow,V.T.K. and Kangueane,P. (2004) Distributions of exon and introns in the human genome. *In Silico Biol.*, **4**, 0032.
19. Stevens,S.W., Ryan,D.E., Ge,H.Y., Moore,R.E., Young,M.K., Lee,T.D. and Abelson,J. (2002) Composition and functional characterization of the yeast spliceosomal Penta-snRNP. *Mol. Cell*, **9**, 31–44.
20. Listerman,I., Sapra,A.K. and Neugebauer,K.M. (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.*, **13**, 815–822.
21. Buerstedde,J.M. and Takeda,S. (1991) Increased Ratio of Targeted to Random Integration after Transfection of Chicken B-Cell Lines. *Cell*, **67**, 179–188.
22. Chen,Y.-I.G., Maika,S.D. and Stevens,S.W. (2006) Epitope tagging of proteins at the native chromosomal loci of genes in mice and in cultured vertebrate cells. *J. Mol. Biol.*, **361**, 412–419.
23. Stevens,S.W. and Abelson,J. (2002) *Methods Enzymol.*, **351**, 200–220.
24. Puig,O., Caspary,F., Rigaut,G., Rutz,B., Bouveret,E., Bragado-Nilsson,E., Wilm,M. and Séraphin,B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.
25. Graham,D.R.M., Garnham,C.P., Fu,Q., Robbins,J. and Van Eyk,J.E. (2005) Improvements in two-dimensional gel electrophoresis by utilizing a low cost "in-house" neutral pH sodium dodecyl sulfate-polyacrylamide gel electrophoresis system. *Proteomics*, **5**, 2309–2314.
26. Neuhoff,V., Arold,N., Taube,D. and Ehrhardt,W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, **9**, 255–262.
27. Davis,M.T. and Lee,T.D. (1998) Rapid protein identification using a microscale electrospray LC/MS system on an ion trap mass spectrometer. *J. Am. Soc. Mass Sepctrom.*, **9**, 194–201.
28. Moore,R.E., Young,M.K. and Lee,T.D. (2000) Method for screening peptide fragment ion mass spectra prior to database searching. *J. Amer. Soc. Mass Spectrom.*, **11**, 422–426.
29. Eng,J.K., McCormack,A.L. and Yates,J.R. (1994) An approach to correlate tandem mass-spectral data of peptides with amino-acid sequences in a protein database. *J. Amer. Soc. Mass Spectrom.*, **5**, 876–989.
30. Azubel,M., Habib,N., Sperling,R. and Sperling,J. (2006) Native spliceosomes assemble with Pre-mRNA to form supraspliceosomes. *J. Mol. Biol.*, **356**, 955–966.
31. Yean,S.L. and Lin,R.J. (1991) U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction. *Mol. Cell. Biol.*, **11**, 5571–5577.
32. Lamond,A.I., Konarska,M.M., Grabowski,P.J. and Sharp,P.A. (1988) Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc. Natl Acad. Sci. USA*, **85**, 411–415.
33. Konarska,M.M. and Sharp,P.A. (1986) Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, **46**, 845–855.
34. Will,C.L., Schneider,C., Hossbach,M., Urlaub,H., Rauhut,R., Elbashir,S., Tuschl,T. and Luhrmann,R. (2004) The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA*, **10**, 929–941.
35. Gottschalk,A., Neubauer,G., Banroques,J., Mann,M., Lührmann,R. and Fabrizio,P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6•U5] tri-snRNP. *EMBO J.*, **18**, 4535–4548.
36. Stevens,S.W. and Abelson,J. (1999) Purification of the yeast U4/U6•U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl Acad. Sci. USA*, **96**, 7226–7231.
37. Chan,S.P., Kao,D.I., Tsai,W.Y. and Cheng,S.C. (2003) The Prp19p-associated complex in spliceosome activation. *Science*, **302**, 279–282.
38. Gall,J.G., Bellini,M., Wu,Z. and Murphy,C. (1999) Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol. Biol. Cell*, **10**, 4385–4402.
39. Kramer,A. and Utans,U. (1991) 3 protein factors (SF1, SF3 and U2AF) function in pre-splicing complex-formation in addition to snRNPs. *EMBO J.*, **10**, 1503–1509.
40. Bennett,M., Michaud,S., Kingston,J. and Reed,R. (1992) Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev.*, **6**, 1986–2000.
41. Gil,A., Sharp,P.A., Jamison,S.F. and Garciablanco,M.A. (1991) Characterization Of cDNAs Encoding The Polypyrimidine Tract-Binding Protein. *Genes Dev.*, **5**, 1224–1236.
42. Patton,J.G., Mayer,S.A., Tempst,P. and Nadalginard,B. (1991) Characterization And Molecular-Cloning Of Polypyrimidine Tract-Binding Protein - A Component Of A Complex Necessary For Pre-Messenger-RNA Splicing. *Genes Dev.*, **5**, 1237–1251.
43. Lewis,J.D., Izaurralde,E., Jarmolowski,A., McGuigan,C. and Mattaj,I.W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.*, **10**, 1683–1698.
44. Cheng,S.C., Tarn,W.Y., Tsao,T.Y. and Abelson,J. (1993) Prp19 - a Novel Spliceosomal Component. *Mol. Cell. Biol.*, **13**, 1876–1882.
45. Russell,C.S., Ben-Yehuda,S., Dix,I., Kupiec,M. and Beggs,J.D. (2000) Functional analyses of interacting factors involved in both pre-mRNA splicing and cell cycle progression in Saccharomyces cerevisiae. *RNA*, **6**, 1565–1572.
46. Chung,S.Y., McLean,M.R. and Rymond,B.C. (1999) Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA*, **5**, 1042–1054.
47. Dix,I., Russell,C., Ben Yehuda,S., Kupiec,M. and Beggs,J.D. (1999) The identification and characterization of a novel splicing protein, Isy1p, of *Saccharomyces cerevisiae*. *RNA*, **5**, 360–368.
48. Figueroa,J.D. and Hayman,M.J. (2004) The human Ski-interacting protein functionally substitutes for the yeast PRP45 gene. *Biochem. Biophys. Res. Comm.*, **319**, 1105–1109.
49. Albers,M., Diment,A., Muraru,M., Russell,C.S. and Beggs,J.D. (2003) Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *RNA*, **9**, 138–150.
50. Tsai,W.Y., Chow,Y.T., Chen,H.R., Huang,K.T., Hong,R.I., Jan,S.P., Kuo,N.Y., Tsao,T.Y., Chen,C.H. *et al.* (1999) Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. *J. Biol. Chem.*, **274**, 9455–9462.
51. Ben-Yehuda,S., Dix,I., Russell,C.S., McGarvey,M., Beggs,J.D. and Kupiec,M. (2000) Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae. *Genetics*, **156**, 1503–1517.
52. Blanton,S., Srinivasan,A. and Rymond,B.C. (1992) Prp38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 snRNA levels. *Mol. Cell. Biol.*, **12**, 3939–3947.

53. Lockhart,S.R. and Rymond,B.C. (1994) Commitment of yeast pre-messenger-RNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. *Mol. Cell. Biol.*, **14**, 3623–3633.
54. Kao,H.Y. and Siliciano,P.G. (1996) Identification of PRP40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.*, **16**, 960–967.
55. Gottschalk,A., Kastner,B., Luhrmann,R. and Fabrizio,P. (2001) The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA*, **7**, 1554–1565.
56. Fortes,P., Bilbao-Cortes,D., Fornerod,M., Rigaut,G., Raymond,W., Séraphin,B. and Mattaj,I.W. (1999) Luc7p, a novel yeast U1 snRNP protein with role in 5′ splice site recognition. *Genes Dev.*, **13**, 2425–2438.
57. Kameoka,S., Duque,P. and Konarska,M.M. (2004) P54(nrb) associates with the 5′ splice site within large transcription/splicing complexes. *EMBO J.*, **23**, 1782–1791.
58. Wang,P., Lou,P.J., Leu,S. and Ouyang,P. (2002) Modulation of alternative pre-mRNA splicing in vivo by pinin. *Biochem. Biophys. Res. Comm.*, **294**, 448–455.
59. Sakashita,E., Tatsumi,S, Werner,D., Endo,H. and Mayeda,A. (2004) Human RNPS1 and its associated factors: A versatile alternative pre-mRNA splicing regulator in vivo. *Mol. Cell. Biol.*, **24**, 1174–1187.
60. Ohi,M.D., Link,A.J., Ren,L.P., Jennings,J.L., McDonald,W.H. and Gould,K.L. (2002) Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol. Cell. Biol.*, **22**, 2011–2024.
61. Berglund,J.A., Abovich,N. and Rosbash,M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.*, **12**, 858–867.
62. Arenas,J.E. and Abelson,J.N. (1997) Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc. Natl Acad. Sci. USA*, **94**, 11798–11802.
63. Martin,A., Schneider,S. and Schwer,B. (2002) Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J. Biol. Chem.*, **277**, 17743–17750.
64. Fouraux,M.A., Kolkman,M.J.M., van der Heijden,A., de Jong,A.S., van Venrooij,W.J. and Pruijn,G.J.M. (2002) The human La (SS-B) autoantigen interacts with DDX15/hPrp43, a putative DEAH-box RNA helicase. *RNA*, **8**, 1428–1443.
65. Company,M., Arenas,J. and Abelson,J. (1991) Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature*, **349**, 487–493.
66. Wagner,J.D.O., Jankowsky,E., Company,M., Pyle,A.M. and Abelson,J.N. (1998) The DEAH-box protein PRP22 is an ATPase that mediates ATP- dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J.*, **17**, 2926–2937.
67. Schwer,B. and Gross,C.H. (1998) Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.*, **17**, 2086–2094.
68. Dalbadie-McFarland,G. and Abelson,J. (1990) PRP5: a helicase-like protein required for mRNA splicing in yeast. *Proc. Natl Acad. Sci. USA*, **87**, 4236–4240.
69. O'Day,C.L., Dalbadie-McFarland,G. and Abelson,J. (1996) The *Saccharomyces cerevisiae* Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J. Biol. Chem.*, **271**, 33261–33267.
70. Will,C.L., Urlaub,H., Achsel,T., Gentzel,M., Wilm,M. and Luhrmann,R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.*, **21**, 4978–4988.
71. Lin,C.R., Yang,L.Q., Yang,J.J., Huang,Y.L. and Liu,Z.R. (2005) ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol. Cell. Biol.*, **25**, 7484–7493.
72. Kistler,A.L. and Guthrie,C. (2001) Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. *Genes Dev.*, **15**, 42–49.
73. Zhang,M. and Green,M.R. (2001) Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev.*, **15**, 30–35.
74. Libri,D., Graziani,N., Saguez,C. and Boulay,J. (2001) Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev.*, **15**, 36–41.
75. King,D.S. and Beggs,J.D. (1990) Interactions of PRP2 protein with pre-mRNA splicing complexes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 6559–6564.
76. Teigelkamp,S., McGarvey,M., Plumpton,M. and Beggs,J.D. (1994) The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J.*, **13**, 888–897.
77. Noble,S.M. and Guthrie,C. (1996) Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics*, **143**, 67–80.
78. Raghunathan,P.L. and Guthrie,C. (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.*, **8**, 847–855.
79. Lauber,J., Fabrizio,P., Teigelkamp,S., Lane,W.S., Hartmann,E. and Lührmann,R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J.*, **15**, 4001–4015.
80. Strauss,E.J. and Guthrie,C. (1991) A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes Dev.*, **5**, 629–641.
81. Strauss,E.J. and Guthrie,C. (1994) PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing *in vitro*. *Nucleic Acids Res.*, **22**, 3187–3193.
82. Teigelkamp,S., Mundt,C., Achsel,T., Will,C.L. and Lührmann,R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain- containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA*, **3**, 1313–1326.
83. Bourgeois,C.F., Lejeune,F. and Stevenin,J. (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog. Nucleic Acids Res. Mol. Biol.*, **78**, 37–88.
84. Horowitz,D.S., Kobayashi,R. and Krainer,A.R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA*, **3**, 1374–1387.
85. Teigelkamp,S., Achsel,T., Mundt,C., Gothel,S.F., Cronshagen,U., Lane,W.S., Marahiel,M. and Lührmann,R. (1998) The 20kD protein of human [U4/U6•U5] tri-snRNPs is a novel cyclophyllin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA*, **4**, 127–141.
86. Myer,V.E., Fan,X.H.C. and Steitz,J.A. (1997) Identification of HuR as a protein implicated in AUUUA-mediated mRNA decay. *EMBO J.*, **16**, 2130–2139.
87. Zhu,H., Hasman,R.A., Barron,V.A., Luo,G.B. and Lou,H. (2006) A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators. *Mol. Biol. Cell*, **17**, 5105–5114.
88. Page-McCaw,P.S., Amonlirdviman,K. and Sharp,P.A. (1999) PUF60: a novel U2AF65-related splicing activity. *RNA*, **5**, 1548–1560.
89. Strasser,K., Masuda,S., Mason,P., Pfannstiel,J., Oppizzi,M., Rodriguez-Navarro,S., Rondon,A.G., Andres,A.K., Struhl,K. *et al.* (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, **417**, 304–308.
90. Kang,Y.B. and Cullen,B.R. (1999) The human Tap protein is a nuclear mRNA export factor that contains novel RNA-binding and nucleocytoplasmic transport sequences. *Genes Dev.*, **13**, 1126–1139.
91. Gruter,P., Tabernero,C., von Kobbe,C., Schmitt,C., Saavedra,C., Bachi,A., Wilm,M., Felber,B.K. and Izaurralde,E. (1998) TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell*, **1**, 649–659.
92. Murphy,R. and Wente,S.R. (1996) An RNA-export mediator with an essential nuclear export signal. *Nature*, **383**, 357–360.
93. Blevins,M.B., Smith,A.M., Phillips,E.M. and Powers,M.A. (2003) Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP. *J. Biol. Chem.*, **278**, 20979–20988.
94. Gilbert,W. and Guthrie,C. (2004) The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. *Mol. Cell*, **13**, 201–212.

95. Mitchell,P. and Tollervey,D. (2003) An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3′-> 5′ degradation. *Mol. Cell*, **11**, 1405–1413.
96. Le Hir,H., Nott,A. and Moore,M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends In Biochem. Sci.*, **28**, 215–220.
97. de Lanerolle,P., Johnson,T. and Hofmann,W.A. (2005) Actin and myosin I in the nucleus: what next? *Nat. Struct. Mol. Biol.*, **12**, 742–746.
98. Gotzmann,J., Gerner,C., Meissner,M., Holzmann,K., Grimm,R., Mikulits,W. and Sauermann,G. (2000) hNMP 200: A novel human common nuclear matrix protein combining structural and regulatory functions. *Exp. Cell Res.*, **261**, 166–179.
99. Batsche,E., Yaniv,M. and Muchardt,C. (2006) The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.*, **13**, 22–29.
100. Dellaire,G., Makarov,E.M., Cowger,J.J.M., Longman,D., Sutherland,H.G.E., Luhrmann,R., Torchia,J. and Bickmore,W.A. (2002) Mammalian PRP4 kinase copurifies and interacts with components of both the U5 snRNP and the N-CoR deacetylase complexes. *Mol. Cell. Biol.*, **22**, 5141–5156.
101. Amberg,D.C., Goldstein,A.L. and Cole,C.N. (1992) Isolation And Characterization Of Rat1 - An Essential Gene Of Saccharomyces-Cerevisiae Required For The Efficient Nucleocytoplasmic Trafficking Of Messenger-RNA. *Genes Dev.*, **6**, 1173–1189.
102. Johnson,A.W. (1997) Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol. Cell. Biol.*, **17**, 6122–6130.
103. Kim,M., Krogan,N.J., Vasiljeva,L., Rando,O.J., Nedea,E., Greenblatt,J.F. and Buratowski,S. (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, **432**, 517–522.
104. Jurica,M.S. and Moore,M.J. (2003) Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.