

Acknowledgements

We thank M. Björås, L. Eide, K. Baynton, K. I. Kristiansen, J. Myllyharju, K. Skarstad and J. Klavness for help and discussions, and L. Eide, K. Baynton and A. Klungland for critical reading of the manuscript. We are grateful to M. Björås for preparation of [³H]MNU-labelled DNA substrates, to L. Eide for the construction of the plasmid for expression of GST-AlkB, and to D. Daoudi for technical assistance. This work was supported by the Research Council of Norway and the Norwegian Cancer Society. E.S. also acknowledges support from the European Commission.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.Ø.F. (e-mail: pal.falnes@labmed.uio.no).

Comprehensive proteomic analysis of the human spliceosome

Zhaolan Zhou*†, Lawrence J. Licklider†‡, Steven P. Gygi†‡ & Robin Reed†

* Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

† Department of Cell Biology, ‡ Taplin Biological Mass Spectrometry Facility, Harvard Medical School, Boston, Massachusetts 02115, USA

The precise excision of introns from pre-messenger RNA is performed by the spliceosome, a macromolecular machine containing five small nuclear RNAs and numerous proteins. Much has been learned about the protein components of the spliceosome from analysis of individual purified small nuclear ribonucleoproteins¹ and salt-stable spliceosome ‘core’ particles^{2,3}. However, the complete set of proteins that constitutes intact functional spliceosomes has yet to be identified. Here we use maltose-binding protein affinity chromatography^{4,5} to isolate spliceosomes in highly purified and functional form. Using nanoscale microcapillary liquid chromatography tandem mass spectrometry⁶, we identify ~145 distinct spliceosomal proteins, making the spliceosome the most complex cellular machine so far characterized. Our spliceosomes comprise all previously known splicing factors and 58 newly identified components. The spliceosome contains at least 30 proteins with known or putative roles in gene expression steps other than splicing. This complexity may be required not only for splicing multi-intronic metazoan pre-messenger RNAs, but also for mediating the extensive coupling between splicing and other steps in gene expression.

Spliceosomes undergo multiple assembly stages and conformational changes during the splicing reaction^{7–9}. To obtain as complete a spliceosomal proteome as possible, we purified spliceosomes from a mixture of all stages of assembly, including spliceosomes that had undergone the first or second catalytic steps of the splicing reaction. In addition, we assembled spliceosomes on two distinct pre-messenger RNAs (pre-mRNAs; adenovirus major late, AdML, and Fushi tarazu, Ftz) so that the proteins common to both could be used as a criterion for identifying splicing-specific proteins.

Representative data for purification of AdML spliceosomes are shown in Fig. 1. AdML-M3 pre-mRNA contains three hairpins that bind to the MS2-MBP fusion protein used for affinity purification (Fig. 1a: MS2 is a bacteriophage coat protein; MBP is maltose-binding protein). AdML pre-mRNA, which lacks these hairpins, was used as a negative control. After adding the MS2-MBP fusion protein to the two pre-mRNAs, spliceosomes were assembled *in vitro*, isolated by gel filtration, affinity-selected by binding to

amylose resin, and eluted with maltose under salt conditions optimal for splicing (60 mM KCl)^{4,5}. The products of the first and second catalytic steps of splicing were detected in the gel filtration fraction for both AdML and AdML-M3 pre-mRNAs (Fig. 1b, lanes 2 and 5). In contrast, after binding to and elution from the amylose-affinity resin, only the splicing products of AdML-M3 spliceosomes were detected (lanes 3 and 6). We conclude that the spliceosomes are highly purified, as there is no detectable non-specific binding of AdML spliceosomes to the affinity resin. This conclusion is bolstered by the observation that spliceosomal small nuclear RNAs

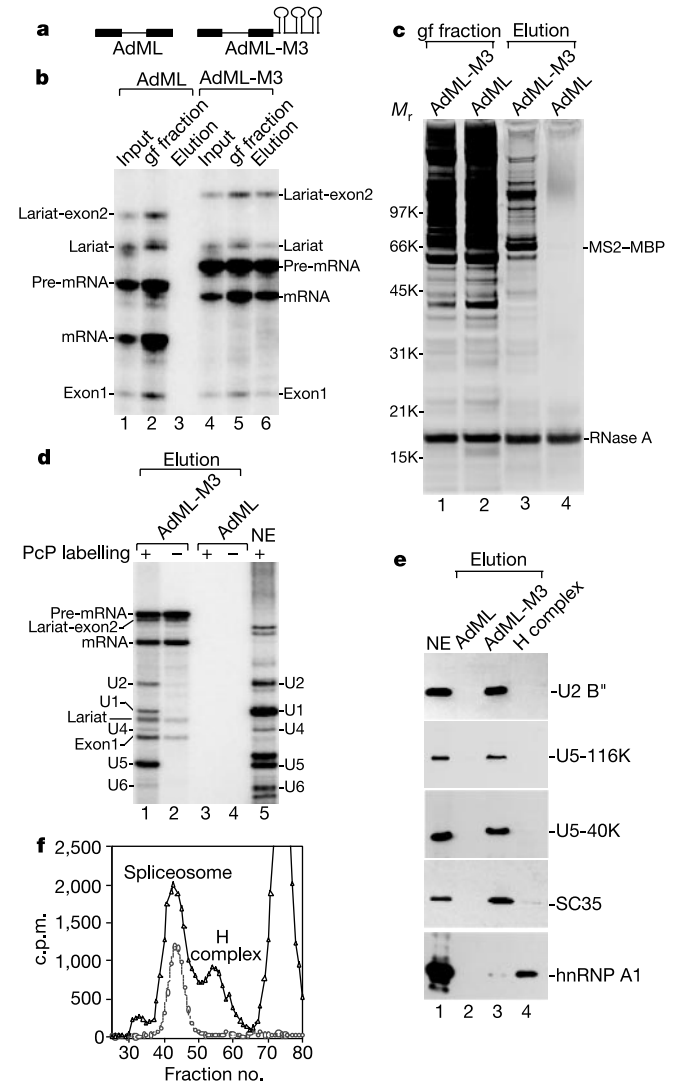


Figure 1 Isolation of spliceosomes. **a**, Schematic of pre-mRNAs. **b**, Total RNAs from the initial splicing reaction (input), gel filtration fraction (gf fraction), or final elution were separated on a 15% denaturing polyacrylamide gel. Unspliced pre-mRNA, splicing intermediates (ariat-exon2 and exon1) and products (ariat and mRNA) are indicated. **c**, Proteins from an RNase-A-treated aliquot of the gel filtration fraction or final elution were separated on 4–12% SDS-PAGE and stained with silver. The protein markers and positions of MS2-MBP and RNase A are indicated. **d**, Analysis of snRNAs. Total RNAs were extracted from the nuclear extract (NE), AdML, or AdML-M3 elution, ³²P-end-labelled (+) or not (-), and fractionated on an 8% polyacrylamide gel. Splicing intermediates, products and snRNAs are indicated. **e**, Western blot of nuclear extract (NE) and the final elutions from AdML, AdML-M3 or H complexes using the indicated antibodies. **f**, Spliceosomes were separated on a Sephacryl-S500 gel filtration column before (triangle) and after (circles) MBP-affinity purification. The positions of the spliceosome and H complex are indicated. The peak in fractions 68–80 contains degraded RNAs and free proteins.

(snRNAs; U1, U2, U4, U5 and U6) are detected in the AdML-M3 but not the AdML elution (Fig. 1d, lanes 1 and 3). Likewise, analysis of silver-stained proteins by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) revealed a large number of proteins specifically associated with the AdML-M3 spliceosome, whereas only the MS2-MBP fusion protein and RNase A are detected in the AdML control (Fig. 1c, lanes 3 and 4).

To further examine the purity of the MBP-purified spliceosomes, we performed western blotting using antibodies that recognize representative splicing factors (Fig. 1e). As an additional control for specificity, we examined MBP-purified H complex, which contains heterogeneous nuclear ribonucleoproteins (hnRNPs) and assembles nonspecifically on RNAs in splicing extracts¹⁰. The U2 small nuclear ribonucleoprotein (snRNP) protein B^{''}, two U5 snRNP proteins (of relative molecular mass 116,000 and 40,000 (M_r 116K and 40K, respectively))¹, and SC35, a member of the SR family of splicing factors¹¹, were detected in the AdML-M3 elution, but not in the AdML or H complex elution (Fig. 1e). In contrast, hnRNP A1 is enriched in the H complex¹⁰. These data indicate that the MBP-affinity purification is specific for AdML-M3 pre-mRNA. The data also show that the MBP-purified spliceosomes are

associated with the spliceosomal snRNAs as well as with a complex set of proteins. The purified spliceosomes are not significantly contaminated with the H complex or with HeLa nuclear extract proteins. Finally, MBP-purified spliceosomes elute in the same gel filtration fractions before and after affinity purification (Fig. 1f). Together, these data show that the MBP-isolated spliceosomes are both highly purified and intact.

To determine whether the MBP-purified spliceosomes are functional, we isolated spliceosomes that had not yet undergone the first catalytic step of splicing, and used them in an *in vitro* complementation assay. The complementing extract was specifically depleted of the splicing factor SF3a⁵, which is a component of U2 snRNP and required for early spliceosome assembly^{5,8}. Thus, the depleted extracts do not support the assembly of the early spliceosome, but contain factors necessary to chase the pre-mRNA in a pre-assembled spliceosome to spliced product. To distinguish between complementation and *de novo* splicing, we used a mixture of MBP-purified spliceosomes and 'naked' pre-mRNA (lacking hairpins). As shown in Fig. 2a and b, splicing products were detected only from MBP-purified Ftz-M3 or AdML-M3 spliceosomes in the SF3a-depleted extract, but not from the corresponding naked

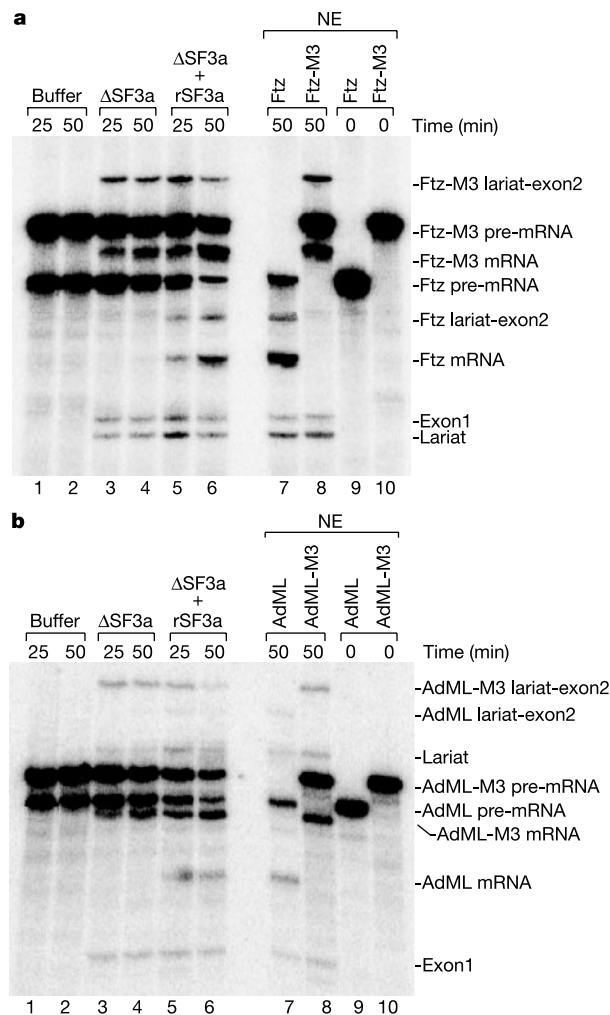


Figure 2 The MBP-purified spliceosomes are functional in an *in vitro* complementation assay. **a**, Purified early spliceosomes assembled on Ftz-M3 pre-mRNA were mixed with Ftz pre-mRNA lacking hairpins, and incubated with either buffer alone (lanes 1 and 2), with SF3a-depleted nuclear extract (lanes 3 and 4), or with SF3a-depleted extract plus recombinant SF3a protein (lanes 5 and 6) under splicing conditions for 25 min (lanes 1, 3 and 5) or 50 min (lanes 2, 4 and 6). Ftz pre-mRNA (lanes 7 and 9) or Ftz-M3 pre-mRNA

(lanes 8 and 10) were incubated under standard splicing conditions in nuclear extract for 0 or 50 min to serve as markers for the splicing products. **b**, Same as **a** except that AdML or AdML-M3 pre-mRNA was used. Total RNA was extracted and separated on an 8% (Ftz) or 16% (AdML) denaturing polyacrylamide gel. Splicing intermediates and products are indicated.

pre-mRNAs (lanes 3 and 4). We conclude that the MBP-purified spliceosomes assembled on both AdML-M3 and Ftz-M3 pre-mRNAs are functional.

The protein composition of the purified, intact and functional spliceosomes assembled on AdML-M3 and Ftz-M3 pre-mRNAs was then determined by liquid chromatography tandem mass spectrometry⁶ (LC-MS/MS). The AdML H complex was purified and sequenced for comparison. Proteins that are shared between AdML and Ftz spliceosomes are shown in Supplementary Tables I-III. (These tables contain the GenBank accession number for each protein, with live links to the protein sequence and the SWISS-PROT protein data bank. Other names for each protein are also listed.)

A total of ~145 distinct proteins were identified as shared between the two spliceosomes, including 88 known splicing factors/snRNP proteins/spliceosomal proteins (Supplementary Tables I and II). The proteins listed in Supplementary Table III were abundant in the H complex, and thus were not counted as spliceosome proteins. The proteins in Supplementary Table I are presented in functional groups, listing those that associate with the spliceosome earlier followed by those that associate later. Previous work¹ identified 43 distinct proteins in U1, U2, U4, U5 and U6 snRNPs. Of these, 41 were identified in our spliceosomes (Supplementary Table I; hLSm5 was not detected, and hLSm8 was detected only in the AdML spliceosome). The additional 47 proteins in Supplementary Table I correspond to previously known splicing factors and spliceosome components⁷⁻⁹. Among these are all known canonical members of the SR family of splicing factors¹¹⁻¹³, all known DExD box splicing factors⁷, all proteins found in the salt-stable spliceosome 'core'^{2,3}, all known second catalytic step splicing factors¹⁴, and known components of the spliced messenger ribonucleoprotein (mRNP) that promotes mRNA transport to the cytoplasm¹⁵ (Supplementary Table I). A number of factors encoded by genes implicated in genetic diseases, ranging from retinitis pigmentosa (U4/U6-61K^{16,17}; U4/U6-90K¹⁸; U5-220K¹⁹) to cancer (WTAP²⁰), are also associated with the spliceosomes (Supplementary Tables I and II).

We also found 58 proteins that have not been previously identified as spliceosomal proteins in humans. These are listed in Supplementary Table II, and are organized according to the motifs in each protein. Some of these proteins (36) have previously described names and/or functions. The remainder are uncharacterized, and have been designated 'functional spliceosome associated proteins' (fSAPs). Among the already named proteins are 5 cyclophilin-like proteins (Supplementary Table II). Recent studies revealed a role for one cyclophilin (U4/U6-20K) in splicing²¹, and our data raise the possibility that cyclophilins (prolyl-isomerases) may have a general function in mediating conformational changes in the spliceosome. Data from several observations suggest that most, if not all, of the 58 new proteins are specifically associated with spliceosomes. First, all 58 are common to spliceosomes assembled on two different pre-mRNAs. Second, these proteins are either not detected or not enriched in the purified H complex. Third, our data (Fig. 1) show that the spliceosomes are highly purified from HeLa nuclear extract contaminants. Fourth, the new proteins are present in the spliceosome with the same relative certainty as the known splicing proteins (based on the number of unique peptides identified; compare Supplementary Tables I and II). Fifth, a large number of the proteins have domains, such as RNA recognition motifs or DExD boxes, which are hallmarks of splicing proteins. Sixth, some of the proteins are known to co-localize with splicing factors in nuclear speckles and/or to localize in the nucleus²² (see Supplementary Table II). Last, many of the proteins (26) have known, or apparent, yeast counterparts that function in mRNA processing in *Saccharomyces cerevisiae*²³ (Supplementary Table II).

The splicing machinery is highly conserved from *S. cerevisiae* to

mammals. Among the ~145 human spliceosome-associated proteins, ~90 known or putative yeast homologues have been identified^{8,23} (Supplementary Tables I and II). The potentially greater complexity of the human spliceosome is not unexpected in light of the vastly greater complexity of splicing in metazoans compared to yeast. Indeed, most metazoan pre-mRNAs contain multiple introns, the introns are typically thousands of nucleotides, and the splicing signals are weakly conserved. Superimposed on this complexity is the high frequency of alternative splicing, which is in turn further complicated owing to regulation²⁴. Thus, many of the metazoan-specific proteins may play roles in the accurate recognition and joining of exons. The best-characterized example of such proteins is the SR family of splicing factors, which comprise at least 10 proteins in the human spliceosome, but is entirely lacking in yeast¹¹⁻¹³. SR proteins are required for both recognition of splice sites and alternative splicing, neither of which is relevant in yeast^{11-13,24}.

Splicing is thought to be coupled to several steps in gene expression, including transcription, polyadenylation and mRNA export²⁵⁻²⁷. A significant fraction of the spliceosome proteome, at least 30 proteins, are either known or candidate proteins for coupling splicing to other steps in gene expression (names shown in bold font in Supplementary Tables I and II). For example, the transcription cofactor TAT-SF1 is in the spliceosome (Supplementary Table II). This protein interacts with snRNPs, and is thought to reciprocally activate transcription elongation and splicing²⁸. Additional transcription factors (for example, CA150, XAB2 and SKIP) as well as polyadenylation factors (CF I) are also in the spliceosome (Supplementary Tables I and II), and may function in coupling. We also detect Aly and UAP56 in the spliceosome, proteins that are known to couple splicing to mRNA export¹⁵ (Supplementary Table I). Aly, CA150 and SKIP were previously detected in salt-stable spliceosome cores, indicating that at least some of the coupling proteins are tightly bound to the spliceosome².

Both Aly and UAP56 are components of the conserved TREX complex, which contains proteins that function in transcription elongation in yeast²⁹. Our data reveal that every protein associated with the TREX complex is also in the functional spliceosome (Supplementary Table II), suggesting that transcription, splicing and export may all be coupled via this complex. In metazoans, most pre-mRNAs contain small exons and enormous introns. Thus, the association of the TREX complex with the spliceosome may explain how mRNA export factors are co-transcriptionally loaded onto exons, which are destined for export, and excluded from introns, which are retained in the nucleus.

Although it is possible that more spliceosomal proteins will be found, our data suggest that the purified functional spliceosomes contain virtually the complete human splicing proteome. This proteome includes all of the previously known splicing factors, snRNP proteins and spliceosomal proteins, as well as a large set of newly identified proteins. The observation that the spliceosome is associated with numerous proteins that function in coupling splicing to other steps in gene expression provides compelling evidence for the emerging concept of an extensively coupled network of gene expression machines²⁵⁻²⁷. □

Methods

Plasmids encoding AdML-M3 or Ftz-M3 pre-mRNAs and the purification of spliceosomes were described previously^{4,5}. For peptide identification from database sequences using tandem mass spectrometry, the proteins in the MBP-purified complexes were fractionated from the gel origin to a distance of 3 cm into a 10% SDS-PAGE gel and stained with Coomassie blue. The gel lane was divided horizontally into ~10 slices which were individually processed to yield in-gel trypsin digestion products. The resulting peptide mixtures were separated and analysed in an automated system by nanoscale LC-MS/MS on an LCQ-DECA ion trap mass spectrometer (Thermo Finnigan) using the vented-column approach⁶. The data were searched against the non-redundant human database from NCBI with the Sequest algorithm³⁰. Peptide matches were validated by previously established criteria, including manual inspection when necessary⁶. More than 65,000 sequencing attempts (tandem mass spectra) were acquired, resulting in the identification of 8,400 unique peptides. The proteins identified are presented in

Supplementary Tables I–III. Tubulin, keratins and ribosomal proteins were found in all complexes, probably as contaminants.

Received 9 April; accepted 12 July 2002; doi:10.1038/nature01031.

1. Will, C. L. & Luhrmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301 (2001).
2. Neubauer, G. *et al.* Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet.* **20**, 46–50 (1998).
3. Bennett, M., Michaud, S., Kingston, J. & Reed, R. Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev.* **6**, 1986–2000 (1992).
4. Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401–405 (2000).
5. Das, R., Zhou, Z. & Reed, R. Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell* **5**, 779–787 (2000).
6. Licklider, L. J., Thoreen, C. C., Peng, J. & Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography–tandem mass spectrometry with a vented column. *Anal. Chem.* **74**, 3076–3083 (2002).
7. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326 (1998).
8. Burge, C. B., Tuschl, T. H. & Sharp, P. A. *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, New York, 1999).
9. Hastings, M. L. & Krainer, A. R. Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13**, 302–309 (2001).
10. Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. & Reed, R. Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol. Cell Biol.* **12**, 3165–3175 (1992).
11. Fu, X. D. The superfamily of arginine/serine-rich splicing factors. *RNA* **1**, 663–680 (1995).
12. Tacke, R. & Manley, J. L. Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* **11**, 358–362 (1999).
13. Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–1211 (2000).
14. Zhou, Z. & Reed, R. Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *EMBO J.* **17**, 2095–2106 (1998).
15. Reed, R. & Hurt, E. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523–531 (2002).
16. Vithana, E. N. *et al.* A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8**, 375–381 (2001).
17. Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H. P. & Luhrmann, R. Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6•U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.* **21**, 1148–1157 (2002).
18. Chakarova, C. F. *et al.* Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **11**, 87–92 (2002).

19. McKie, A. B. *et al.* Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.* **10**, 1555–1562 (2001).
20. Little, N. A., Hastie, N. D. & Davies, R. C. Identification of WTAP, a novel Wilms' tumour 1-associating protein. *Hum. Mol. Genet.* **9**, 2231–2239 (2000).
21. Horowitz, D. S., Lee, E. J., Mabon, S. A. & Misteli, T. A cyclophilin functions in pre-mRNA splicing. *EMBO J.* **21**, 470–480 (2002).
22. Andersen, J. S. *et al.* Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1–11 (2002).
23. Stevens, S. W. *et al.* Composition and functional characterization of the yeast spliceosomal pentanRNP. *Mol. Cell* **9**, 31–44 (2002).
24. Cartegni, L., Chew, S. L. & Krainer, A. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nature Rev. Genet.* **3**, 285–298 (2002).
25. Bentley, D. Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* **11**, 347–351 (1999).
26. Hirose, Y. & Manley, J. L. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**, 1415–1429 (2000).
27. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
28. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929–933 (2001).
29. Strasser, K. *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304–308 (2002).
30. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank N. Dorman, E. Ibrahim and K. Magni for comments on the manuscript; C. Thoreen for database support; R. Das for SF3a-depleted nuclear extract; and R. Luhrmann and G. Dreyfuss for antibodies. HeLa cells were obtained from the National Cell Culture Center. This work was supported by the NIH (S.P.G. and R.R.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.R. (e-mail: rreed@hms.harvard.edu).