Molecular Cell
Article

Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition

Carlo Baejen,[1,3] Phillipp Torkler,[2,3] Saskia Gressel,[2] Katharina Essig,[2] Johannes Söding,[1,*] and Patrick Cramer[1,2,*]
[1]Max-Planck-Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
[2]Gene Center Munich and Department of Biochemistry, Center for Integrated Protein Science CIPSM, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 Munich, Germany
[3]Co-first author
*Correspondence: johannes.soeding@mpibpc.mpg.de (J.S.), patrick.cramer@mpibpc.mpg.de (P.C.)
http://dx.doi.org/10.1016/j.molcel.2014.08.005

SUMMARY

Biogenesis of eukaryotic messenger ribonucleoprotein complexes (mRNPs) involves the synthesis, splicing, and 3′ processing of pre-mRNA, and the assembly of mature mRNPs for nuclear export. We mapped 23 mRNP biogenesis factors onto the yeast transcriptome, providing 10^4–10^6 high-confidence RNA interaction sites per factor. The data reveal how mRNP biogenesis factors recognize pre-mRNA elements in vivo. They define conserved interactions between splicing factors and pre-mRNA introns, including the recognition of intron-exon junctions and the branchpoint. They also identify a unified arrangement of 3′ processing factors at pre-mRNA polyadenylation (pA) sites in yeast and human, which results from an A-U sequence bias at pA sites. Global data analysis indicates that 3′ processing factors have roles in splicing and RNA surveillance, and that they couple mRNP biogenesis events to restrict nuclear export to mature mRNPs.

INTRODUCTION

Biogenesis of eukaryotic mRNAs involves pre-mRNA synthesis by RNA polymerase (Pol) II and cotranscriptional RNA processing, which encompasses 5′ capping, intron splicing, and 3′ RNA cleavage and polyadenylation (3′ processing). The mature mRNA is packaged with RNA-binding proteins into messenger ribonucleoprotein particles (mRNPs) and exported to the cytoplasm where it directs protein synthesis. Factors for mRNP biogenesis are recruited cotranscriptionally by interactions with the C-terminal domain (CTD) of Pol II (Buratowski, 2009; Heidemann and Eick, 2012; Hsin and Manley, 2012; Perales and Bentley, 2009), and by interactions with the emerging pre-mRNA transcript (Chan et al., 2011; Darnell, 2013; Mandel et al., 2008; Müller-McNicoll and Neugebauer, 2013; Proudfoot, 2011; Wahl et al., 2009).

Mapping of mRNP biogenesis factors onto pre-mRNA and mature mRNA promises insights into RNA determinants for splicing, 3′ processing, and RNA export, and into the coupling between these processes. Biogenesis factors can in principle be mapped onto the transcriptome by in vivo protein-RNA crosslinking and immunoprecipitation (CLIP) (Ule et al., 2005). CLIP is based on UV light-induced crosslinking and identifies direct protein-RNA interaction sites after sequencing of the crosslinked RNA regions (Milek et al., 2012). CLIP-based methods could indeed provide transcriptome maps for several human 3′ processing factors (Martin et al., 2012) and mRNA-binding proteins in the yeast *Saccharomyces cerevisiae* (Tuck and Tollervey, 2013). However, mRNP biogenesis factors have not been systematically mapped onto pre-mRNA, likely due to difficulties in trapping short-lived RNAs in cells, and due to the complexity caused by different pre-mRNA species.

Here we present high-confidence transcriptome maps for 23 mRNP biogenesis factors in yeast, where pre-mRNA complexity is low because spliced protein-coding genes contain only single introns. These maps were obtained by photoactivatable-ribonucleoside-enhanced (PAR)-CLIP, which was developed in human cells (Hafner et al., 2010) and recently adopted to yeast (Creamer et al., 2011; Schulz et al., 2013). Compared to other CLIP methods, PAR-CLIP uses less invasive, low-energy UV light, which results in a specific U-to-C base transition at the crosslinked sites that facilitates their precise localization at low false-positive rates.

Our analysis includes factors implicated in 5′ cap binding, splicing, 3′ processing, and mRNA export. Six of these 23 factors, namely Gbp2, Hrp1/Nab4, Mex67, Nab2, Pab1, and Tho2, were recently mapped using a technique called CRAC (Tuck and Tollervey, 2013). This published study focused on the distribution of RBPs between mRNAs and noncoding (nc) transcripts, whereas we focus here on pre-mRNA recognition during mRNP biogenesis. We show that PAR-CLIP captures short-lived pre-mRNA intermediates, and provide insights into the in vivo RNA-binding preferences of mRNP biogenesis factors, the recognition of introns and 3′ processing sites in pre-mRNA, and the interdependence of different steps in mRNP biogenesis.

RESULTS AND DISCUSSION

Transcriptome Maps of mRNP Biogenesis Factors

To map mRNP biogenesis factors over cellular RNA at high resolution, we optimized the PAR-CLIP protocol and obtained high RNA labeling efficiencies with 4-thiouracile (4tU) in exponentially

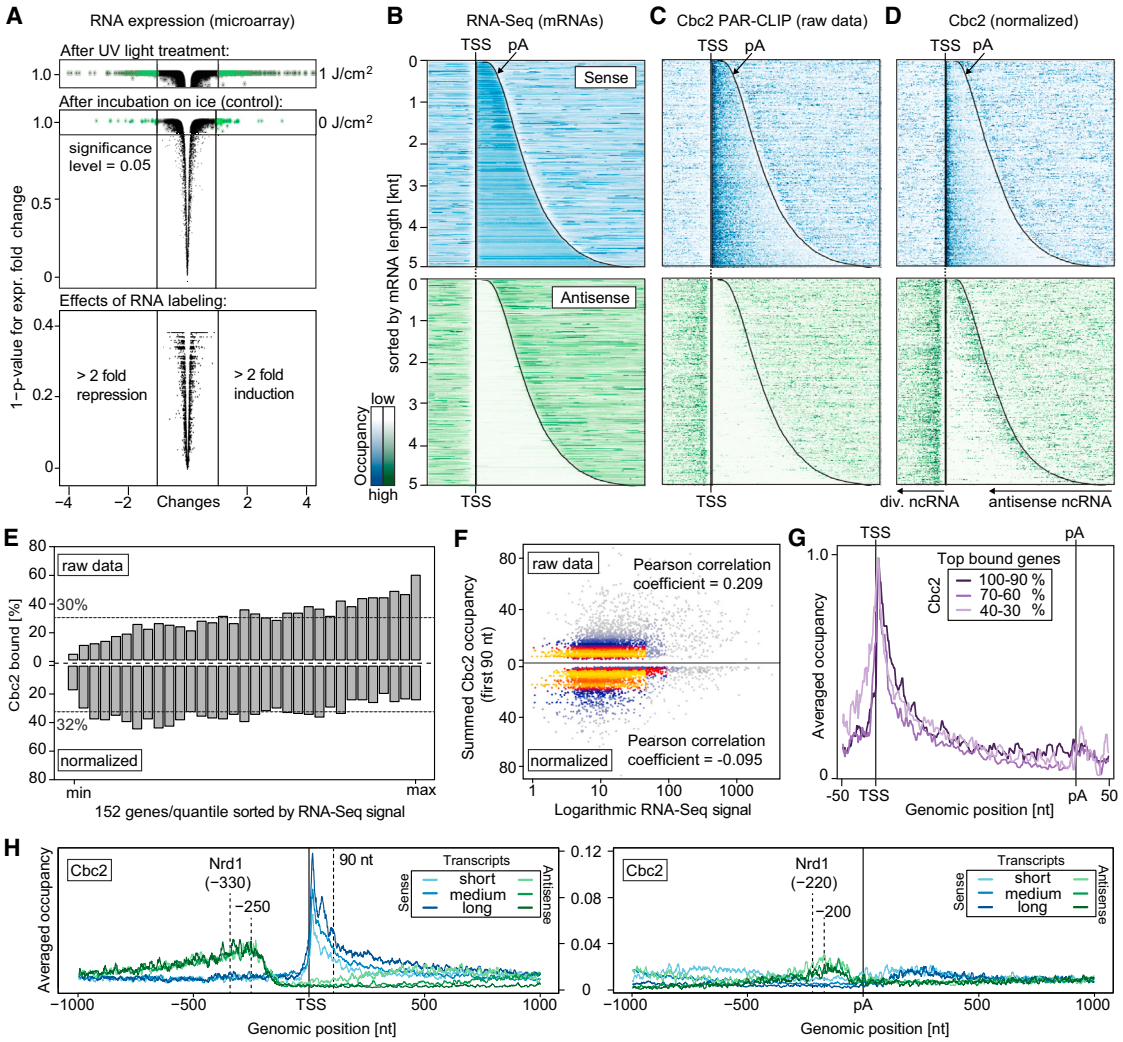

Figure 1. RNA Abundance-Normalized PAR-CLIP Estimates Factor Occupancies over the Yeast Transcriptome

(A) 4-thiouracil (4tU) labeling has only a very minor effect on cellular mRNA levels. Vulcano plots of expression fold changes for mRNAs measured by Affymetrix microarrays show that only few mRNAs significantly change their abundance due to RNA labeling, incubation on ice, and UV light exposure.

(B) Smoothed Cbc2 RNA-Seq data in sense (blue) and antisense (green) direction for all open reading frame-containing transcribed regions (ORF-Ts). ORF-Ts are sorted by length and aligned at their transcription start site (TSS).

(C) Smoothed, raw Cbc2 RNA-binding strength as measured by the number of PAR-CLIP U-to-C transitions per U site in sense (blue) and antisense (green) direction for all ORF-Ts sorted by length and aligned at their TSS.

(D) Normalization of PAR-CLIP signals reduces noise. Relative Cbc2 occupancy estimated by dividing the number of U-to-C transitions for each U site by the RNA-Seq signal at the corresponding genomic position in sense (blue) and antisense (green) direction for all ORF-Ts.

(E) Normalization of PAR-CLIP signals facilitiates interpretation as occupancy profiles. Whereas raw PAR-CLIP binding strength (shown in C) strongly depends on mRNA level, normalized occupancies (shown in D) are independent of mRNA levels. The y axis shows the percentage of transcripts bound by Cbc2, where "bound" is defined as the sum of the first 90 nt of a transcript > = the mean of the sums of the first 90 nt of all ORF-Ts.

(F) Normalization abolishes the dependence of estimated occupancy on mRNA level. Pearson correlation between mRNA level and the PAR-CLIP binding strength in the first 90 nt of each ORF-T before (top) and after (bottom) RNA abundance normalization.

(G) Cbc2-binding profiles are independent of factor occupancy. Transcript-averaged Cbc2 occupancy for three mRNA level classes (100%–90%, 70%–60%, and 40%–30% expression quantile).

(H) Transcript-averaged Cbc2 occupancies in sense (blue) and antisense (green) directions, centered at the TSS (left) and the polyadenylation site (pA) (right), for short (0–1 kb), medium (1–2 kb), and long (2–5 kb) transcripts.

growing yeast cells (Experimental Procedures). We found conditions that led to very high reproducibility between biological replicates (see Figure S1A available online) and enabled high 4tU incorporation levels of ~2% (Andrus and Kuimelis, 2001) without significant changes in cellular mRNA abundance (Figures 1A and S1B). We also developed a computational pipeline for PAR-CLIP data analysis (P.T., C.B., A. Graf, S. Krebs, P.C, and J.S., unpublished data). This pipeline includes a statistical model for

Table 1. mRNP Biogenesis Factors Analyzed here by PAR-CLIP

Biogenesis Event	Factor/ Subunit	Complex	RNA-Binding Domain[a]	PAR-CLIP Crosslink Sites	False Discovery Rate (%)
Capping	Cbc2	CBC	RRM	98,034	0.178
Splicing	Luc7	U1 snRNP	ZF	93,261	1.035
	Mud1	U1 snRNP	RRM	99,384	1.918
	Nam8	U1 snRNP	RRM	151,813	1.675
	Snp1	U1 snRNP	RRM	25,493	0.447
	Ist3	U2 snRNP	RRM	66,003	3.184
	Mud2	BBP-U2AF65	RRM	801,430	1.769
	Msl5	BBP-U2AF65	ZN	476,370	1.961
3′ processing	Rna15	CFIA	RRM	582,756	3.463
	Mpe1	CPF	ZF	122,500	2.262
	Yth1	CPF (PFI)	ZF	59,049	3.432
	Cft2	CPF (CFII)	–	189,866	1.723
	Pab1	–	RRM	233,513	2.052
	Pub1	–	RRM	371,902	1.332
Export	Hpr1	THO/TREX	–	249,887	1.913
	Tho2	THO/TREX	–	400,965	1.064
	Sub2	TREX	–	228,620	1.085
	Mex67	TREX	–	288,579	1.010
	Yra1	Export adaptor	RRM	400,156	0.681
	Nab2	Export adaptor	ZF	283,606	2.413
	Npl3	Export adaptor	RRM	770,240	1.282
	Hrb1	SR-like	RRM	395,402	0.976
	Gbp2	SR-like	RRM	65,692	0.182

[a]RRM, RNA recognition motif; ZF, zinc finger domain.

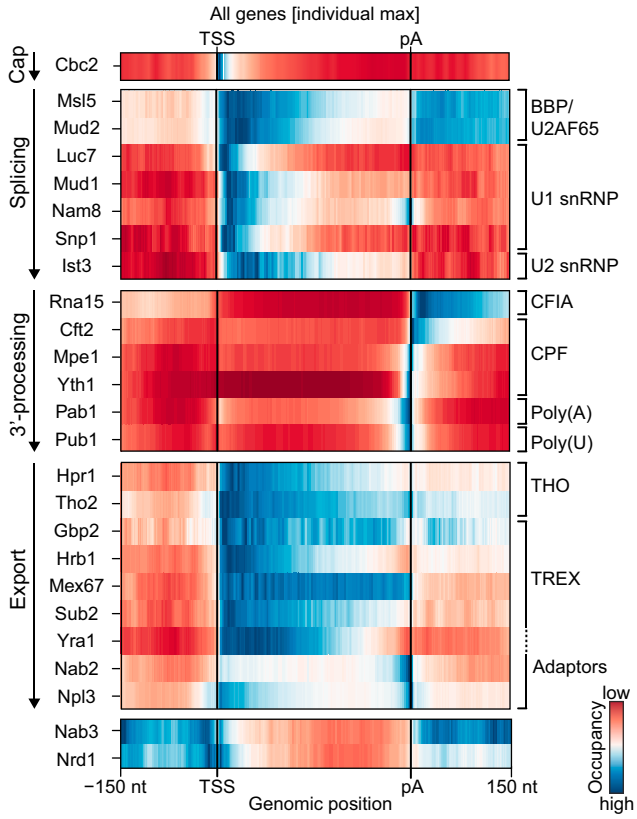

Figure 2. RNA-Binding Profiles for 23 mRNP Biogenesis Factors
Transcript-averaged occupancy profiles of mRNP biogenesis factors, scaled such that their TSSs and pA sites coincide. The color code shows the occupancy relative to the maximum occupancy per profile (dark blue).

crosslink site determination and an analysis of the sequence neighborhood of crosslinked sites with the motif discovery tool XXmotif (Hartmann et al., 2013). For each factor, we obtained between 25,000 and 800,000 high-confidence protein-RNA binding sites at a p value below 5×10^{-3}, which corresponds to false discovery rates between only 0.18% and 3.5% (Table 1; Experimental Procedures).

We applied the optimized protocol to 23 mRNP biogenesis factors that showed reproducible PAR-CLIP signals (Table 1). These included the Cbc2 subunit of the cap-binding complex (CBC) and components of the splicing machinery, namely the yeast homologs of the branchpoint (BP)-binding protein BBP (Msl5) and U2AF65 (Mud2), and subunits of the snRNPs U1 (Luc7, Mud1, Nam8/Mud15, Snp1) and U2 (Ist3/Snu17). Factors in the 3′ processing machinery included the Rna15 subunit of cleavage factor (CF) IA, and three subunits of the cleavage and polyadenylation factor (CPF), namely Mpe1, Yth1 (in the CPF subcomplex PFI), and Cft2/Ydh1 (CPF subcomplex CFII). We also included nine proteins implicated in mRNP export, in particular subunits of the THO/TREX complex (Hpr1, Tho2, Sub2), the export factor Mex67, and its putative mRNA adaptors Nab2, Npl3 (also known as Nop3 or Nab1), and Yra1/She11, and the SR-like factors Gbp2 and Hrb1. We also studied the poly(A)- and poly(U)-binding proteins Pab1 and Pub1 that regulate mRNP export and stability (Mangus et al., 2003). Biological replicates for a random selection of factors revealed a high reproducibility (Figures S1C–S1E). The obtained data map the protein-RNA interaction landscape underlying mRNP biogenesis (Figures 2, S2A, and S2B).

RNA Abundance Normalization Reveals Capped Transcripts

PAR-CLIP crosslinks for the CBC subunit Cbc2 clustered at the 5′ ends of mRNAs as expected, but often extended for several hundred nucleotides (nt) downstream (Figure 1C). We found that Cbc2 binding appeared more focused at mRNA 5′ ends after the data were corrected for RNA abundance (Figure 1D), as measured by RNA-Seq under the same experimental conditions (Figure 1B). We estimated relative occupancies of the crosslinked factors along mRNAs by dividing the frequency of U-to-C transitions by the RNA-Seq signal at this site (Experimental Procedures). The normalization reduced the transcript-to-transcript signal fluctuation, led to an even distribution of estimated occupancy signals over RNAs with different abundance (Figure 1E), and abolished a weak artificial correlation of PAR-CLIP signals with RNA levels (Figure 1F). As a result, the distribution of crosslinking sites over transcripts was independent of the number of observed crosslinks (Figures 1G and S1F), confirming

the high data quality. The normalization procedure thus prevents misinterpretation due to systematic overrepresentation of abundant transcripts. In the normalized data, strongest binding of Cbc2 was observed within the first ~90 nt downstream of the transcription start site (TSS) within the 5′ untranslated region (5′ UTR) of mRNAs (Figures 1G, 1H, and 2).

The normalization also enhanced Cbc2 signals on ncRNA transcripts (Figures 1C and 1D, green panels), facilitating the detection of capped ncRNAs (Figures 1H and S2). Widespread Cbc2 binding was observed at the 5′ end of divergent ncRNA transcripts that emerged from bidirectional promoters antisense to mRNAs. Cbc2 sites were found from ~120 nt upstream of the TSS of the sense transcript, with the peak of Cbc2 crosslinking at ~250 nt (Figure 1H, left panel). This is consistent with the presence of two distinct Pol II initiation complexes for sense and divergent transcription from bidirectional promoters (Rhee and Pugh, 2012), and indicates that divergent transcripts are capped before they associate with the Nrd1 complex that triggers their degradation (Jensen et al., 2013; Mischo and Proudfoot, 2013; Schulz et al., 2013). Cbc2 also crosslinked to antisense RNA 100–300 nt upstream of the polyadenylation (pA) site, identifying capped antisense ncRNAs at the 3′ ends of many genes (Figure 1H, right panel). We also identified Cbc2-binding sites in cryptic unstable transcripts (CUTs) and stable untranslated transcripts (SUTs) (Wery et al., 2011), with stronger signals for CUTs (Figures S3A and S3B).

The Cbc2 data enabled comparison with the recent CRAC-based mapping of Cbc1, the other subunit of CBC (Tuck and Tollervey, 2013). Both Cbc1 and Cbc2 showed RNA interactions at the 5′ ends of transcripts, cross-validating the studies (a detailed comparison also for other factors is found in Figure S2C). However, the PAR-CLIP protocol and normalization procedure used here apparently led to more focused signals at RNA 5′ ends (Figures 1G and 1H; see Figures S2C and S2D for comparison of other factors) and enhanced signals for short-lived RNAs and RNAs with low abundance (Figures 2 and S3A–S3D), prompting us to use it for an investigation of factors involved in the recognition of short-lived pre-mRNA elements.

Conserved Recognition of Pre-mRNA Introns

Intron recognition is the initial step in pre-mRNA splicing and was extensively studied in vitro (Will and Lührmann, 2011). It begins with binding of BBP to the BP and binding of U2AF65 to a pyrimidine-rich region between the BP and 3′ splice site (3′ SS), and continues with binding of the U1 snRNP to the 5′ SS. The resulting complex E is then remodeled, and U2 snRNA displaces BBP by base pairing with the BP region, positioning U2 snRNP near the 3′ SS and giving rise to complex A (Figure 3G). The protein-RNA interactions underlying intron recognition may be largely conserved between yeast and human but have not been systematically analyzed in vivo.

Although introns are rapidly degraded in vivo, our protocol could capture intron sequences bound by splicing factors involved in intron recognition (Figures 3A, 3B, and S4A). Crosslinking signals for the BBP homolog Msl5 and the U2AF65 homolog Mud2 spanned entire introns and showed peaks near the 5′ SS and the 3′ SS, respectively (Figure 3D). The BP motif UACUAAC was detected around Mud2- and Msl5-bound sites in intron-containing genes (Figures 3A and S4A) and is generally located within ~50 nt upstream of the 3′ SS (Figure 3E). When we averaged crosslink density after aligning introns at the BP, Msl5 displayed a peak on the BP (Figure 3F), consistent with binding of yeast Msl5 to the BP in vivo. Mud2 and Ist3 peaked 15 nt and 27 nt downstream of the BP, respectively (Figure 3F). Thus we could resolve binding of the U2AF65 homolog Mud2 to a pyrimidine/U-rich region that was defined in the human system (Mackereth et al., 2011, Zarnack et al., 2013). These results agree with in vitro-derived functions of the Msl5-Mud2 complex in BP recognition (Berglund et al., 1997), and in bridging between the BP and U1 snRNP at the 5′ SS (Abovich and Rosbash, 1997). Msl5 and Mud2 also crosslinked to intron-less RNAs (Figures 2 and S2A), consistent with scanning of RNAs for U-rich regions by the U2AF65-BBP complex.

Crosslinks of U1 snRNP subunits peaked ~17 nt downstream of the 5′ SS (Figure 3D). Motif searches around crosslinking peaks (±25 nt) detected the consensus 5′ SS sequence GUAUGU in Luc7, Mud1, Nam8, and Snp1 data (Figures 3A and S4A). As expected, crosslink sites of U1 snRNP subunits were not significantly enriched around the BP (Figure S4B). The U2 subunit Ist3 crosslinked mainly ~10 nt upstream of the 3′ SS (Figure 3D). These results agree with the in vitro-derived binding of U1 and U2 snRNPs near the 5′ SS and the 3′ SS, respectively (Will and Lührmann, 2011). The splice site RNA motifs were apparently responsible for recruitment of U1 and U2 snRNPs, because their subunits generally did not crosslink to intron-less RNAs (Figure S2A). To investigate the order of factor binding to introns, we calculated a "splicing index" (Figures 3C, S4C, and S4D; Experimental Procedures) (Schneider et al., 2012). All splicing factors obtained negative splicing indices, demonstrating preferential binding to unspliced RNA. The strongest preference for unspliced over spliced RNA was obtained for Mud2, the weakest for Ist3. Thus our in vivo data support the two-state model of intron recognition derived from in vitro studies (Figure 3G).

Unified Recognition of Pre-mRNA Polyadenylation Sites

In human cells, recognition of the pA site involves several RNA sequence elements that are bound by the cleavage and polyadenylation specificity factor (CPSF) complex (Chan et al., 2011; Mandel et al., 2008; Proudfoot, 2011). The CPSF subunit CPSF-160 recognizes the pA signal (PAS) sequence AAUAAA upstream of the pA site. Subunits CPSF-100 and CPSF-30 bind neighboring U-rich sequences, and subunit CPSF-73 cleaves the RNA (Chan et al., 2011; Mandel et al., 2008). Homologous subunits are found in the yeast CPSF counterpart CPF, which also contains additional proteins, such as Mpe1 (Vo et al., 2001).

After extensive trials we could map CPF subunits Cft2/Ydh1 (CPSF-100), Yth1 (CPSF-30), and Mpe1 onto pre-mRNA (Figure 4A). Cft2 crosslinked to regions flanking the pA site, consistent with binding near the cleavage site in vitro (Dichtl and Keller, 2001). Yth1 showed a peak ~17 nt upstream of the pA site, consistent with in vitro results (Barabino et al., 2000), and with localization of its human counterpart CPSF-30 in vivo (Martin et al., 2012). Mpe1 gave rise to a peak ~6 nt upstream the pA site, explaining why it is an essential factor required for

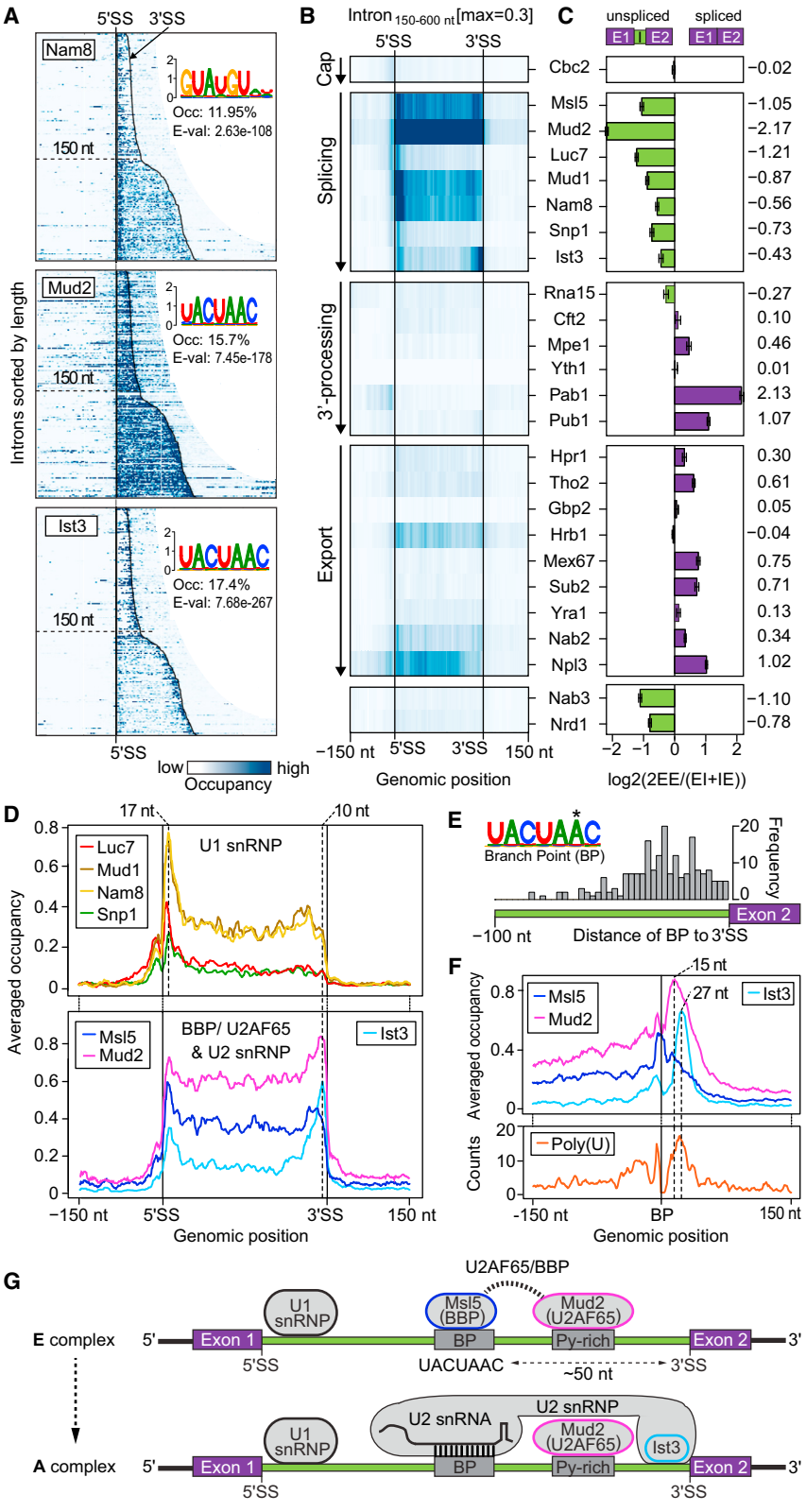

Figure 3. Conserved Recognition of Pre-mRNA Introns In Vivo

(A) Normalized and smoothed occupancy profiles of U1 subunit Nam8, Mud2 (human U2AF65), and U2 subunit Ist3 around introns of up to 600 nt length. Introns were sorted by length and aligned at their 5′ splice site (5′ SS).

(B) Transcript-averaged occupancy profiles of all factors around introns between 150 and 600 nt length.

(C) Splicing factors show high affinity for unspliced RNAs. Splicing indices (Experimental Procedures) indicate the binding preference for spliced versus unspliced RNAs for all factors.

(D) Intron-averaged factor occupancy profiles show binding of U1 snRNP near the 5′ splice site (5′ SS) and binding of the U2 snRNP and the commitment complex (BBP/U2AF65) over the entire intron with a peak at the 3′ splice site (3′ SS).

(E) The branchpoint (BP) lies within 50 nt upstream of the 3′ SS. Distance distribution of the branch point (BP) motif from the 3′ SS.

(F) Yeast Msl5 (human BBP) binds the BP in vivo, whereas Mud2 (U2AF65) and U2 snRNP (Ist3) bind downstream of the BP. Transcript-averaged occupancy profiles of Msl5, Mud2, and Ist3, centered at the BP (top), compared to the poly(U) distribution over the same region (bottom).

(G) Model of factors recognizing an intron during formation of E and A complexes.

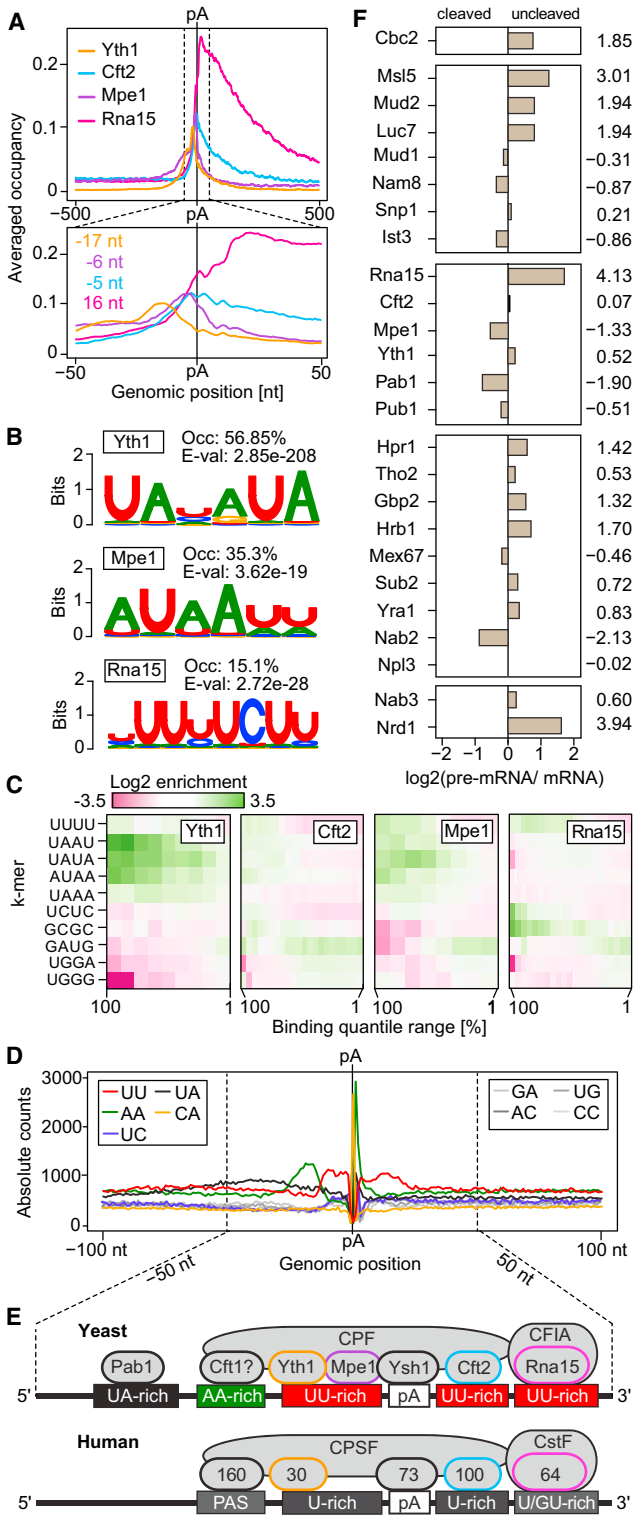

Figure 4. Unified Model for Polyadenylation Site Recognition In Vivo

(A) CFIA subunit Rna15 binds downstream of CPF complex and the pA site. Averaged occupancy profiles of Rna15 and CPF subunits Cft2, Mpe1, and Yth1, aligned at the pA site show that CPF binds at the pA site, whereas CFIA binds downstream in vivo.

(B) RNA motifs enriched in a window of ±25 nt around the crosslinked sites with fraction of occurrence and XXmotif E value.

(C) RNA 3′ processing factors have distinct tetramer-binding preferences. Shown are log-odd scores for enrichments of selected tetramers (y axis) for bins of binding sites ranging from 100% to 1% occupancy (x axis).

(D) Sequences areound the pA site exhibit an "A-U bias." Shown is the distribution of nucleotide composition around the pA site.

(E) Unified model for pA site recognition in *S. cerevisiae* and human by the two major, conserved 3′ processing complexes CPF (CPSF) and CFIA (CstF) bound on pre-mRNA.

(F) Processing indices measuring the tendency of the factors to bind to uncleaved pre-mRNA rather than cleaved RNA, computed as log2 odds ratios uncleaved versus cleaved RNA bound by the factor.

3′ processing (Vo et al., 2001). Although Cft1/Yhh1 (CPSF-160) and Ysh1 (CPSF-73) did not show PAR-CLIP signals, these data locate the yeast CPSF counterpart CPF at the pA site in vivo and define many of its subunit-RNA interactions.

Human CPSF is assisted by the CstF complex, which binds to pre-mRNA downstream of CPSF (Chan et al., 2011; Mandel et al., 2008). However, the yeast CstF counterpart CFIA is believed to bind upstream of the CPSF counterpart CPF (Chan et al., 2011; Mandel et al., 2008), and this model is based on in vitro evidence that the CFIA subunit Rna15 binds upstream of the pA site (Gross and Moore, 2001). In contrast, we observed very strong crosslinking of Rna15 downstream of the pA site in vivo, with a peak at ~16 nt (Figure 4A). These results agree with an alternative in vitro study (Dichtl and Keller, 2001), and with binding of the human Rna15 homolog CstF64 downstream of the pA site in vivo (Martin et al., 2012). The PAR-CLIP peak for Rna15 downstream of the pA site is also consistent with an occupancy peak observed in the same region by chromatin immunoprecipitation (Mayer et al., 2012). Thus CFIA is located downstream, rather than upstream, of the pA site and CPF, consistent with the position of the human CstF complex downstream of the pA site and downstream of the CPF counterpart CPSF. These results lead to a unified model for pA site recognition by the two conserved 3′ processing complexes bound to pre-mRNA (Figure 4E).

Definition and Decoration of mRNA 3′ Ends

To investigate how RNA sequence-binding preferences of processing factors define the pA site, we searched for sequence motifs around crosslinking peaks. Peaks for Yth1 and Mpe1 often contained the motifs UAUAUA ("efficiency element"; Guo et al., 1995; Graber et al., 2002) and AUAAUU, respectively, and Cft2, Mpe1, and Yth1 generally preferred RNA sites containing U/A-rich tetramer sequences (Figure 4C). Rna15 did not show enrichment for the "positioning element" AAUAAA (Figure 4C) that had been reported to bind in vitro (Gross and Moore, 2001). Instead, it preferentially bound regions in vivo that were enriched with the A-less tetrameric motifs UUUU and UCUC (Figure 4C), and a motif search yielded the sequence UUUUCUU (Figure 4B) that is very similar to the downstream U-rich element (Graber et al., 2002). More generally, pA site regions exhibit a bias in nucleotide composition. Whereas the immediate vicinity of the pA site showed an enrichment of UU and AA dinucleotides up- and downstream, respectively, flanking regions showed the inverse, a phenomenon we refer to as A-U bias (Figures 4D and S7).

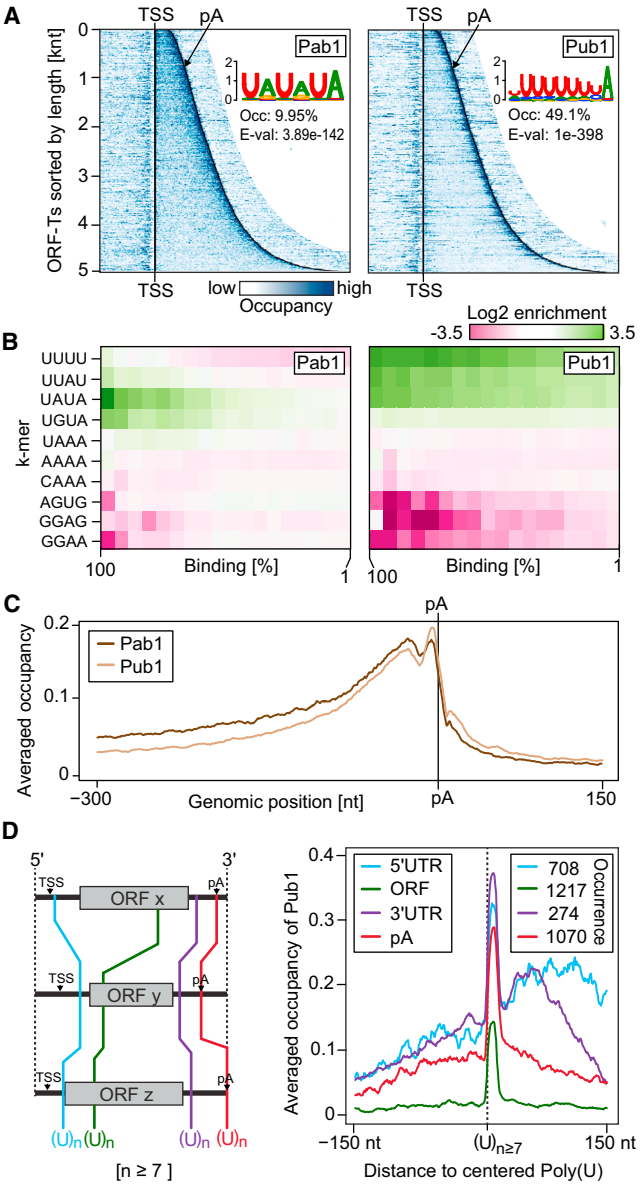

Figure 5. Pab1 and Pub1 Bind UA-and U-Rich Sequences at mRNA 3′ Ends

(A) Normalized and smoothed occupancy profiles of the poly(A)-binding protein Pab1 and the poly(U)-binding protein Pub1 derived from PAR-CLIP data in sense direction for all ORF-Ts. ORF-Ts were sorted by length and aligned at their transcription start site (TSS). The motifs were enriched around binding sites (±25 bp).

(B) Pab1 and Pub1 bind to U/A-rich sequences. Log2 enrichment of selected tetramer motifs around Pab1 (left) and Pub1 (right) binding sites compared to unbound sequence regions, analyzed within 18 equal-sized bins of occupancy quantiles between 100% and 1% site occupancy (x axis).

(C) Averaged occupancy profiles of Pab1, Pub1, and Yth1 derived from PAR-CLIP data in sense direction for all ORF-Ts, centered at the pA site of all ORF-Ts.

(D) Pub1 preferentially binds poly(U)n ≥ 7 tracts near the pA site. Average occupancy profiles of Pub1 around Poly(U)n ≥ 7 tracts within the 5′ UTR, ORF, or 3′ UTR, or near pA sites.

The A-U bias apparently directs binding of CPF subunits upstream and around pA sites and binding of Rna15 downstream of pA sites, due to corresponding sequence preferences of these factors (Figure 4C). In some yeast mRNAs, the A-rich upstream region contains a positioning element (Gross and Moore, 2001) that may bind Cft1 and may correspond to the human polyadenylation signal (Guo and Sherman, 1995), and an UA-rich efficiency element (Guo et al., 1995) that may bind CFIB/Hrp1 (Kessler et al., 1997). These two elements are, however, dispensable for RNA cleavage in vitro (Dichtl and Keller, 2001), consistent with our view that the A-U bias, rather than specific sequence elements, underlies pA site recognition. A similar A-U bias was observed around human pA sites (Martin et al., 2012) and befits the conserved arrangement of 3′ processing factors revealed here.

Additional data showed that the AU/UU-rich regions upstream of pA sites bind Pab1 and Pub1 (Figure 5). Both factors gave rise to crosslinking near the 3′ end of mRNAs (Figures 5A and 5C). Pab1 bound upstream of the pA site to the UAUAUA "efficiency element" motif (Figures 5A and 5B) as described (Riordan et al., 2011; Tuck and Tollervey, 2013), and showed some depletion at the Yth1 site (Figures 4E and 5C). Pub1 occupied both UA-rich regions in the 3′ UTR (Duttagupta et al., 2005; Vasudevan and Peltz, 2001) and poly(U) tracts (Figure 5B) but also bound upstream of the open reading frame (ORF) in the 5′ UTR (Figure 5D) as described (Cui et al., 1995; Ruiz-Echevarría et al., 1998; Ruiz-Echevarría and Peltz, 2000). Pub1 and Pab1 were generally depleted from the translated ORF (Figures 5C and 5D), indicating that these factors are displaced during translation in the cytoplasm. Taken together, these data may be explained as follows. The two major 3′ processing complexes CPF and CFIA bind to pre-mRNA regions with an A-U bias and pA site, causing RNA cleavage and polyadenylation, and subsequent release of 3′ processing factors, which enables complete decoration of the mRNA 3′ end with Pab1 and Pub1.

Transcription-Coupled mRNP Export

In our current view, mRNA export begins with the recruitment of the THO/TREX complex during Pol II elongation (Luna et al., 2012; Strässer et al., 2002). Mature mRNA is then exported from the nucleus by the heterodimeric export factor Mex67-Mtr2 (Grüter et al., 1998; Segref et al., 1997). Mex67 uses mRNA adaptor proteins such as Nab2, Npl3, and Yra1 (Hackmann et al., 2011; Iglesias et al., 2010; Rodríguez-Navarro and Hurt, 2011; Stewart, 2010). PAR-CLIP analysis revealed similar distributions of the THO subunits Tho2 and Hpr1 over mRNAs (Figures 6A and S5A) and no mRNA preferences, indicating that the THO complex is a general factor associated with Pol II transcripts. Tho2 gave stronger signals, consistent with its role in THO complex recruitment (Chávez et al., 2000; Gewartowski et al., 2012). Mex67 bound RNA in vivo (Figure 6B), explaining how it remains bound to mRNA after release of adaptor proteins. Mex67 did not show preferences for RNA motifs, consistent with its function as a general export factor, and consistent with data obtained by CRAC (Tuck and Tollervey, 2013; Figure S2C).

The export adaptors Nab2, Npl3, and Yra1 showed different crosslinking patterns, indicating specific, nonredundant functions (Figure 6C). The number of mRNAs bound by two or three

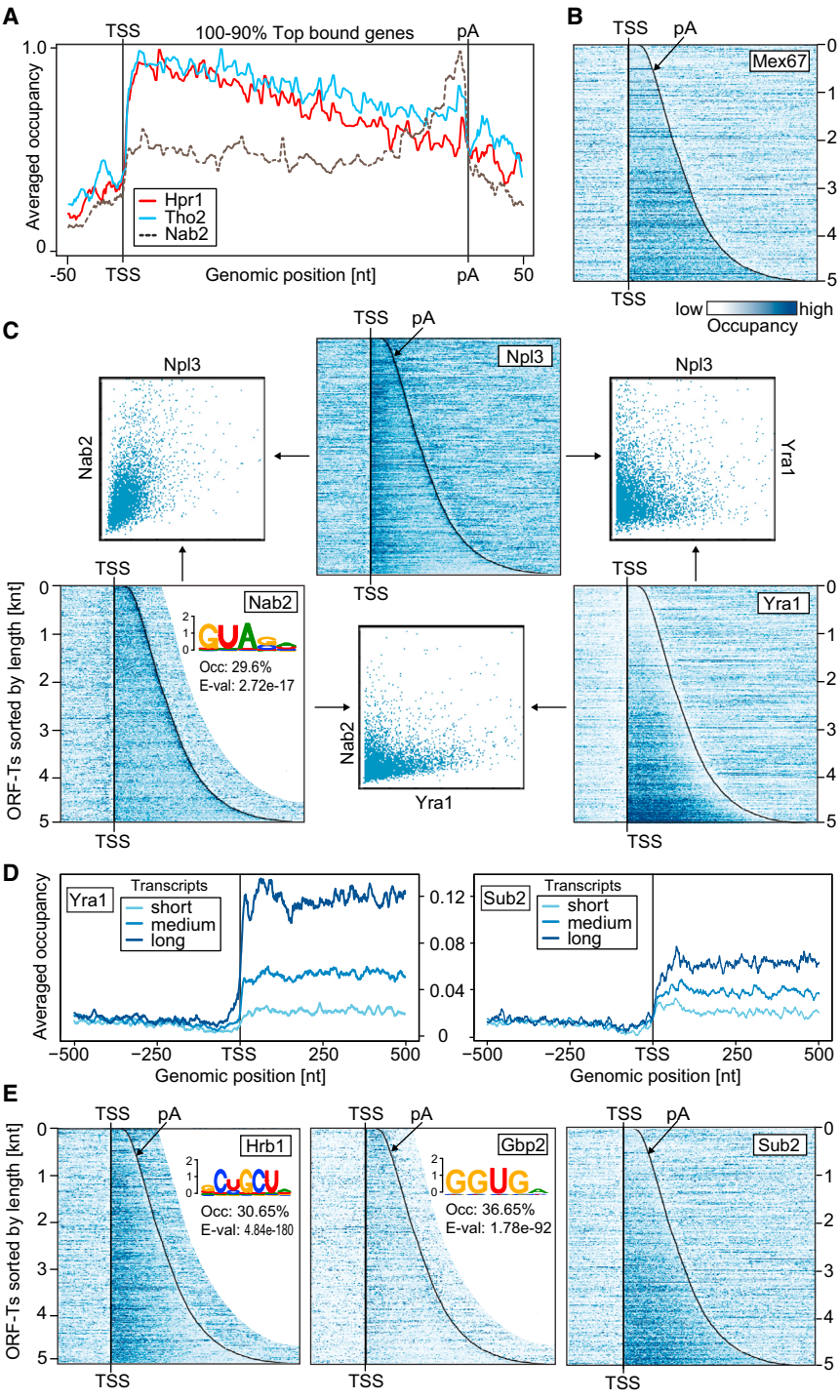

Figure 6. Export Adaptors Differ in Their mRNA-Binding Preference

(A) Normalized and smoothed average occupancy profiles for Nab2 and the THO complex subunits Tho2 and Hpr1, derived from PAR-CLIP experiments for top-bound ORF-Ts (100%–90%).

(B) Normalized and smoothed average occupancy profiles for the general export factor Mex67, derived from PAR-CLIP experiments for all ORF-Ts, sorted by length and aligned at their TSS.

(C) The export adaptors Yra1, Npl3, and Nab2 have distinct mRNA-binding preferences. Pairwise correlation scatterplots for occupancies of Yra1, Npl3, and Nab2 on ORF-Ts.

(D) Transcript-averaged Yra1 and Sub2 occupancies in sense directions, centered at the TSS, for short (0–1 kb), medium (1–2 kb), and long (2–5 kb) transcripts.

(E) Occupancy profiles for SR-like proteins Hrb1 and Gbp2 and for Sub2, derived from PAR-CLIP experiments for all ORF-Ts, sorted by length and aligned at their TSS.

export adaptors was limited (Figure 6C), showing that these factors exhibit mRNA preferences, as suggested by purification of mRNAs associated with Yra1 (Hieronymus and Silver, 2003). Yra1 occupancy decreased before the pA site, whereas Npl3 also showed crosslinking at 3′ ends, consistent with its influence on pA site choice (Bucheli et al., 2007; Deka et al., 2008). Whereas Nab2 preferentially bound short mRNAs (Figure 6C), Yra1 and Sub2 preferred long mRNAs (Figure 6D). Nab2 crosslinking density was also stronger at the 3′ ends of ORF-Ts as described (Figure 6A) (Tuck and Tollervey, 2013), consistent with its known influence on 3′ processing (Anderson et al., 1993; Green et al., 2002; Hector et al., 2002; Tuck and Tollervey, 2013). Nab2 sites were enriched for the motif GUAG (Figure 6C) as described (Riordan et al., 2011). Thus components of the mRNA export machinery show preferences for RNAs with specific sequences and lengths.

Global Analysis Links Splicing to 3′ Processing

We now subjected all PAR-CLIP data to a global analysis (Figure 7). In addition to the splicing index (Figures 3C, S4C, and S4D), we introduced a "processing index" (Experimental Procedures) that estimates whether factors preferentially bind uncleaved or cleaved RNA (Figures 4F and S5B). A plot of splicing versus processing indices (Figure 7A) indicates how the composition of protein-RNA complexes is remodeled during mRNP biogenesis (Figure 7B). We further calculated for each pair of factors the Pearson correlation coefficient of the total weighted occupancies over transcripts (Figure 7C; Experimental Procedures). This estimates the extent to which factors co-occupy the same transcripts. We further determined the extent to which two factors colocalize in a window of 25 nt around binding sites (Figure 7D; Experimental Procedures). Finally, we computed for each pair of factors the Pearson

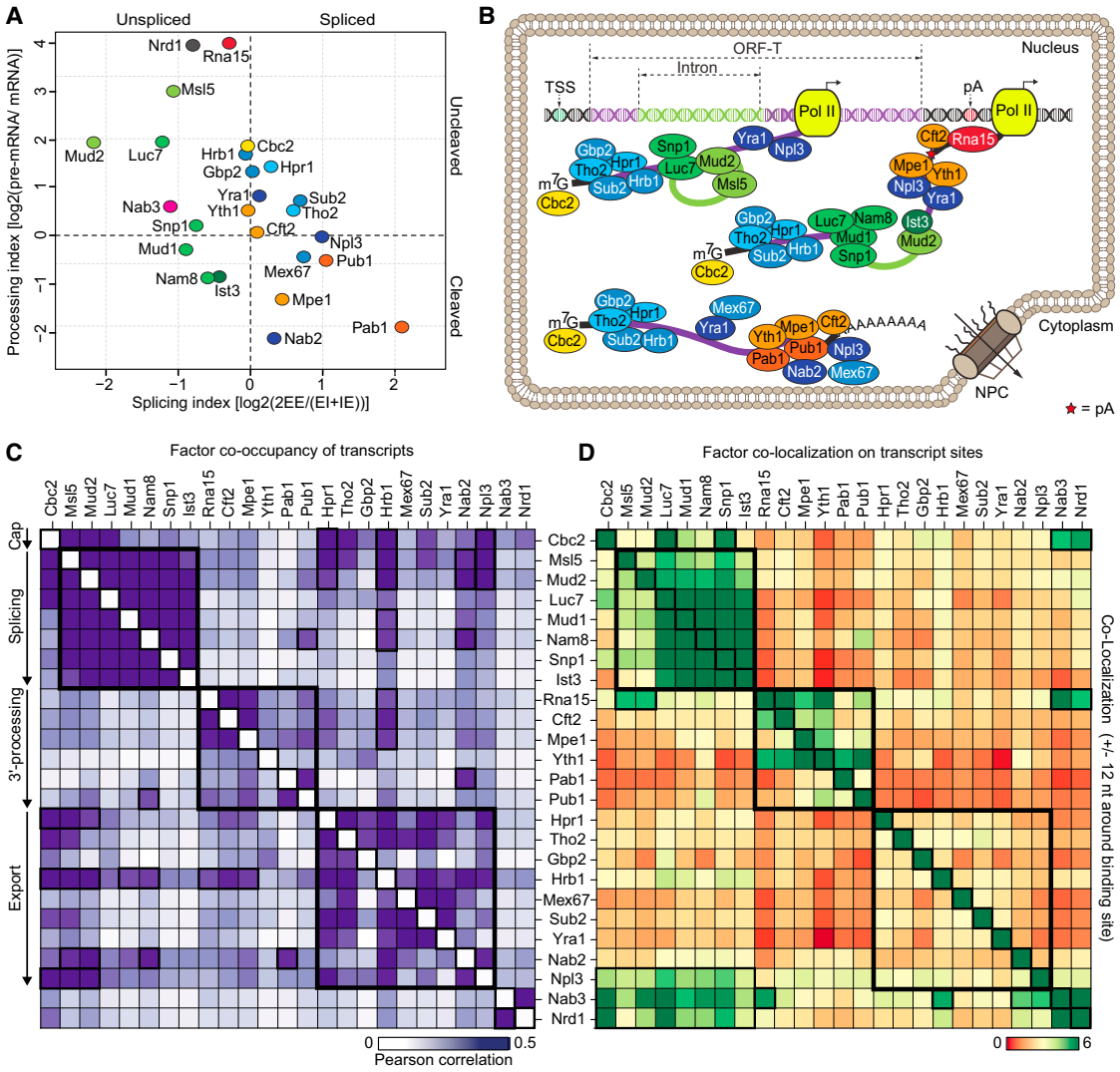

Figure 7. Global Analysis Reveals Links between Splicing, 3′ Processing, and Export

(A) "Splicing index" and "processing index" for all analyzed factors (yellow, capping; orange/red, 3′ processing; green, splicing; blue, export; black/pink, RNA surveillance). A splicing index of 0 (1) indicates binding only to unspliced (spliced) mRNA. Similarly, a processing index of 0 (1) signifies binding only to uncleaved pre-mRNA (cleaved mRNA). The splicing index is averaged over all intron-containing ORF-Ts; the processing index is averaged over all ORF-Ts.

(B) Model for mRNP biogenesis resulting from PAR-CLIP-based occupancy measurements.

(C) Factor co-occupancy of transcripts. Pairwise Pearson correlation coefficient of the total weighted occupancies over entire transcripts for all factors.

(D) Factor colocalization on transcript sites. Colocalization for each pair of factors in a window of 25 nt around the centered binding sites (column).

correlations between their averaged occupancy profiles, to measure the shape similarity of binding profiles (Figure S6; Experimental Procedures).

This global analysis provided evidence for an ancient link between splicing and 3′ processing. Splicing factors fell into two groups when sorted by their processing index (Figures 4F and 7A). The splicing factors Mud2, Msl5, Snp1, and Luc7 preferentially bound uncleaved RNA, whereas other splicing factors preferred cleaved RNA (Figure 7B). Mud2 and Msl5 profiles were correlated with those of 3′ processing factors Rna15 and Cft2, and Nam8 correlated with Mpe1, Pab1, and Pub1 (Figure S6). Also, Mud2, Msl5, and Nam8 crosslinked near the pA site (Figure 2), although this could also reflect splicing-independent functions of these factors at pA sites. Nam8 tended to colocalize with Pub1, whereas Mud2 and Msl5 tended to co-occupy transcripts with Hpr1, Hrb1, Nab2, and Npl3, and they colocalized with Rna15 (Figures 7C and 7D). Indeed, Rna15 preferentially bound unspliced mRNAs (Figures 3C and 7A), but also showed the lowest processing index (Figures 4F and 7A), confirming its early binding to pre-mRNA (Guo and Sherman, 1995; Leeper et al., 2010).

These results indicate that the machineries for splicing and 3′ processing interact in yeast, as inferred by genetics (Chanfreau et al., 1996), although it is currently believed that such an interaction is restricted to mammalian cells (Martinson, 2011; Proudfoot, 2011; Shi et al., 2009). Indeed, 3′ processing may

assist in splicing, but not the other way around, because unspliced and spliced transcripts recruit 3′ processing factors to a similar extent (data not shown). This model is consistent with the known stimulation of splicing by 3′ processing in human cells (Kyburz et al., 2006).

Transcript Surveillance and Fate

The global analysis also elucidated how nuclear export is restricted to mature mRNPs. First, export factors preferred spliced over unspliced mRNA, and generally did not bind uncleaved RNA (Figures 3C and 7A). The highest splicing index was found for Pab1, which binds mature mRNA (Brune et al., 2005), whereas the lowest splicing index was found for Mud2, which is expected to initiate intron recognition (Will and Lührmann, 2011). Second, binding profiles for export factors except Nab2 differed from those of 3′ processing factors (Figure S6), reflecting that export factors select 3′ processed mRNAs. Third, the SR proteins Gbp2 and Hrb1 (Windgassen and Krebber, 2003) overlapped with THO/TREX subunits, and Hrb1 tended to bind the same transcripts as the Mud2-Msl5 complex (Figure 7C). This is consistent with a role of Gbp2 and Hrb1 in restricting mRNA export to spliced transcripts (Hackmann et al., 2014). Gbp2 and Hrb1 showed distinct RNA-binding motifs (Figure 6E), and Hrb1 colocalized with splicing factors Luc7 and Snp1 (Figure 7D), consistent with a role in splicing (Kress et al., 2008; Shen and Green, 2006; Will and Lührmann, 2011).

A subset of 3′ processing factors also showed occupancy profiles that were similar to those of RNA surveillance factors Nrd1 and Nab3 (Figure S6). Rna15 colocalized with Nrd1 and Nab3 on transcripts (Figure 7D) and crosslinked to aberrant divergent ncRNAs (Figure S2). This indicates that some 3′ processing factors are part of the RNA surveillance machinery that terminates and degrades aberrant RNAs, as predicted by genetics (Mischo and Proudfoot, 2013). Nrd1 and Nab3 colocalized with Cbc2 (Figure 7D) and preferentially bound uncleaved pre-mRNA, in accordance with their role in triggering early termination of transcription. Nrd1 and Nab3 colocalized with splicing factors on introns (Figures 3C and 7D), likely because of their preferential binding to noncoding RNA regions (Schulz et al., 2013). These observations are consistent with a general nuclear RNA surveillance pathway and suggest that a transient surveillance/3′ processing complex takes a decision during RNA synthesis of whether a transcript is subjected to degradation or to polyadenylation and nuclear export.

Conclusions

Here we report high-confidence transcriptome maps for 23 protein factors involved in mRNP biogenesis in the eukaryotic model system *S. cerevisiae*. We demonstrate that PAR-CLIP efficiently captures short-lived unspliced and uncleaved pre-mRNAs. This allowed mapping of splicing factors onto introns and of 3′ processing factors within regions downstream of the pA site, which are rapidly removed and degraded in cells. The distribution of factors over various pre-mRNA species that result from events during mRNP biogenesis enabled integration of the data into a model for mRNP biogenesis based on factor occupancy.

The three most notable insights from our data include (1) the observation of intron recognition by the Mud2-Msl5 (human U2AF65-BBP) and the snRNPs U1 and U2 in vivo, (2) a unified, conserved arrangement of the two major 3′ processing complexes CPF and CFIA (human CPSF and CstF) around the pA site, and (3) links of the 3′ processing machinery to RNA splicing and nuclear RNA surveillance. An analysis of the RNA sequences underlying the crosslinked sites recovered known splicing motifs and revealed a conserved "A-U bias" at the pA site. It also defined eight specific RNA motifs bound by biogenesis factors, of which three were new, and showed that most factors exhibited binding preferences for certain RNA tetrameric motifs.

Our results support the emerging concept that RNA-binding factors, in contrast to DNA-binding factors, generally show binding preferences, rather than specificities, and exhibit site promiscuity. To achieve high target specificity, multiple interactions of RNA-binding subunits within a functional complex are required and/or additional protein interactions of factors must occur, such as binding to the Pol II CTD. Synergistic factor binding is evident within the machineries for splicing and 3′ processing. It explains how sites in pre-mRNA can be located with confidence despite little sequence conservation and a scarcity of motifs in RNA. It also explains how mRNA, which is restricted in its sequence due to its coding nature, can evolve to specifically bind multifactor complexes.

Finally, global analysis of our data revealed that processes involved in mRNP biogenesis are more tightly coupled than generally thought. An ancient link between 3′ processing and splicing apparently coordinates both processes and generates mature mRNPs that are selected for nuclear export. In particular, we observed direct RNA interactions of splicing factors at the pA site and a differential distribution of splicing factors on pre-mRNAs before and after their 3′ cleavage. How 3′ processing may influence spliceosome dynamics and how the composition of protein-RNA complexes is remodeled several times during mRNA biogenesis may be analyzed in the future.

EXPERIMENTAL PROCEDURES

TAP-Tag Validation by Western Blot Analysis

Saccharomyces cerevisiae BY4741 strains containing C-terminally tandem affinity purification (TAP)-tagged genes (Open Biosystems) were tested by western blotting for expression of the correct tagged protein. Cells were lysed and diluted lysate was run on a precast NuPAGE Bis-Tris gel (Invitrogen). Following SDS-PAGE, samples were blotted onto a PVDF membrane (Bio-Rad), and the membrane was probed with a primary antibody against the TAP tag (PAP; Sigma-Aldrich). Antibody detection was performed using Pierce enhanced chemiluminescence (ECL) western blotting substrate (Thermo Scientific) and Amersham Hyperfilm ECL (GE Healthcare).

Yeast Growth and RNA Labeling with 4-Thiouracil

S. cerevisiae cells expressing the TAP-tagged protein were grown at 30°C from OD_{600}:0.1 to OD_{600}:0.5 in CSM minimal medium (Formedium) supplemented with 10 mg/l uracil, 100 μM 4-thiouracil (4tU), and 2% glucose. After reaching OD_{600}:0.5, another 900 μM 4tU was added, and cells were grown further for 4 hr (OD_{600}:1.3–1.6).

PAR-CLIP of Yeast Proteins

PAR-CLIP and data acquisition were performed as described (Schulz et al., 2013) with minor modifications. Briefly, 4tU-labeled yeast cells were UV-irradiated with an energy dose of 12 J/cm^2 at 365 nm. Cells were lysed by bead beating, and the lysate was sonicated. Immunoprecipitation was performed on a rotating wheel for 2 hr at 4°C with rabbit IgG-conjugated protein G

magnetic Dynabeads (Invitrogen). Crosslinked RNA was partially digested with RNase T1 and prepared for cDNA library preparation. Following adaptor ligation, RNA was recovered by Proteinase K digestion for 2 hr at 55°C and subsequent acidic phenol/chloroform extraction. Reverse transcription was done for 1 hr at 44°C using SuperScript III RTase (Invitrogen). PCR amplification was performed using the NEXTflex barcode primer kit (Bio Scientific). Generated cDNA was purified, size-selected, and quantified using a 2100 Bioanalyzer (Agilent Technologies). Samples were sequenced on an Illumina machine (GAx or HiSeq 1500).

RNA-Seq for Global RNA Abundance Normalization

Yeast cells were treated as for PAR-CLIP (above) using the identical labeling conditions and a UV light (365 nm) energy dose of 1 J/cm^2. After bead beating, total RNA was isolated by acid phenol/chloroform extraction using Roti-Phenol (Carl Roth), and purified and concentrated using the RNA Clean Concentrator-5 (Zymo Research). Purified RNA was depleted of ribosomal RNAs using Ribo-Zero rRNA removal kit (Epicenter). The resulting rRNA-depleted RNA was used for multiplexed RNA-Seq library preparation using the NuGEN Encore Complete RNA-Seq Library Systems. Libraries were qualified on an Agilent Bioanalyzer 2100 (Agilent Technologies) and sequenced on an Illumina MiSeq machine.

Sequencing Data Quality Control and Mapping

We wrote a script for pre-processing of sequencing data obtained from the Illumina GaIIx or Illumina HiSeq machine. The script calls widely used NGS-software and custom scripts with parameter settings adapted to PAR-CLIP analysis. Adaptor sequences are first trimmed from the raw sequencing files. The quality filter then discards all reads containing unidentified nucleotides (N), Phreds scores below 30, reads shorter than 15 nt, or reads that are flagged by Illumina's internal chastity filter. Quality-trimmed reads are aligned to the *S. cerevisiae* genome (***sacCer3***, version 64.1.1) using the short read aligner Bowtie (version 0.12.7) (Langmead et al., 2009) with a maximum of one mismatch and taking unique matches only (options: -q -p 4 -S -sam -nohead -v 1 -e 70 -l 28 -y -a -m 1 -best -strata -phred33 -quals). The resulting SAM files are then converted into BAM and PileUp files using SAMTools (Li et al., 2009). Detailed descriptions of computational analyses are found in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO series accession number GSE59676. RNA-Seq data for global RNA abundance normalization were deposited under the SRA accession number SAMN02741679.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and Supplemental Experimental Procedures and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2014.08.005.

AUTHOR CONTRIBUTIONS

C.B. and P.C. designed the study. C.B. established protocols and planned experiments. C.B., S.G., and K.E. performed experiments. P.T., J.S., C.B., and P.C. designed data analysis methods. P.T. and J.S. developed the analysis package. P.T. carried out data analysis. P.C. and C.B. wrote the manuscript with input from all authors. P.C. and J.S. supervised the work.

ACKNOWLEDGMENTS

We would like to thank Helmut Blum, Stefan Krebs, Sylvia Mallok, Thomas Fröhlich, and Alexander Graf (LAFUGA) for sequencing, HPLC, and Galaxy maintenance. We thank Björn Schwalb for providing R scripts and Daniel Schulz and Christoph Engel for critically reading the manuscript. J.S. was supported by the Deutsche Forschungsgemeinschaft (SFB646, GRK1721, QBM), the Bundesministerium für Bildung und Forschung (BMBF, ebio), and the Bavarian Network for Molecular Biosystems (BioSysNet). P.C. was supported by the DFG (SFB646, SFB960, SFB1064, GRK1721, CIPSM, NIM, QBM), the Advanced Investigator Grant TRANSIT of the European Research Council, and the Jung and Volkswagen Foundations.

Received: May 5, 2014
Revised: July 8, 2014
Accepted: July 31, 2014
Published: September 4, 2014

REFERENCES

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403–412.

Anderson, J.T., Wilson, S.M., Datar, K.V., and Swanson, M.S. (1993). NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Mol. Cell. Biol. *13*, 2730–2741.

Andrus, A., and Kuimelis, R.G. (2001). Base composition analysis of nucleosides using HPLC. In Current Protocols in Nucleic Acid Chemistry, Serge L. Beaucage, ed. Current Protocols in Nucleic Acid Chemistry, Unit 10 16.

Barabino, S.M., Ohnacker, M., and Keller, W. (2000). Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J. *19*, 3778–3787.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781–787.

Brune, C., Munchel, S.E., Fischer, N., Podtelejnikov, A.V., and Weis, K. (2005). Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. RNA *11*, 517–531.

Bucheli, M.E., He, X., Kaplan, C.D., Moore, C.L., and Buratowski, S. (2007). Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. RNA *13*, 1756–1764.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Chan, S., Choi, E.A., and Shi, Y. (2011). Pre-mRNA 3′-end processing complex assembly and function. Wiley Interdiscip. Rev. RNA *2*, 321–335.

Chanfreau, G., Noble, S.M., and Guthrie, C. (1996). Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science *274*, 1511–1514.

Chávez, S., Beilharz, T., Rondón, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J. *19*, 5824–5834.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. *7*, e1002329.

Cui, Y., Hagan, K.W., Zhang, S., and Peltz, S.W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev. *9*, 423–436.

Darnell, J.E., Jr. (2013). Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. RNA *19*, 443–460.

Deka, P., Bucheli, M.E., Moore, C., Buratowski, S., and Varani, G. (2008). Structure of the yeast SR protein Npl3 and Interaction with mRNA 3′-end processing signals. J. Mol. Biol. *375*, 136–150.

Dichtl, B., and Keller, W. (2001). Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. EMBO J. *20*, 3197–3209.

Duttagupta, R., Tian, B., Wilusz, C.J., Khounh, D.T., Soteropoulos, P., Ouyang, M., Dougherty, J.P., and Peltz, S.W. (2005). Global analysis of Pub1p targets

reveals a coordinate control of gene expression through modulation of binding and stability. Mol. Cell. Biol. *25*, 5499–5513.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Gewartowski, K., Cuéllar, J., Dziembowski, A., and Valpuesta, J.M. (2012). The yeast THO complex forms a 5-subunit assembly that directly interacts with active chromatin. BioArchitecture *2*, 134–137.

Graber, J.H., McAllister, G.D., and Smith, T.F. (2002). Probabilistic prediction of Saccharomyces cerevisiae mRNA 3′-processing sites. Nucleic Acids Res. *30*, 1851–1858.

Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X., and Corbett, A.H. (2002). Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. J. Biol. Chem. *277*, 7752–7760.

Gross, S., and Moore, C.L. (2001). Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3′-end formation. Mol. Cell. Biol. *21*, 8045–8055.

Grüter, 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.

Guo, Z., and Sherman, F. (1995). 3′-end-forming signals of yeast mRNA. Mol. Cell. Biol. *15*, 5983–5990.

Guo, Z., Russo, P., Yun, D.F., Butler, J.S., and Sherman, F. (1995). Redundant 3′ end-forming signals for the yeast CYC1 mRNA. Proc. Natl. Acad. Sci. USA *92*, 4211–4214.

Hackmann, A., Gross, T., Baierlein, C., and Krebber, H. (2011). The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits. EMBO Rep. *12*, 1024–1031.

Hackmann, A., Wu, H., Schneider, U.M., Meyer, K., Jung, K., and Krebber, H. (2014). Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. Nat. Commun. *5*, 3123.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

Hartmann, H., Guthöhrlein, E.W., Siebert, M., Luehr, S., and Söding, J. (2013). P-value-based regulatory motif discovery using positional weight matrices. Genome Res. *23*, 181–194.

Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastia, L., and Swanson, M.S. (2002). Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. EMBO J. *21*, 1800–1810.

Heidemann, M., and Eick, D. (2012). Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code. RNA Biol. *9*, 1144–1146.

Hieronymus, H., and Silver, P.A. (2003). Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat. Genet. *33*, 155–161.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev. *24*, 1927–1938.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol. Cell *52*, 473–484.

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A., and Moore, C.L. (1997). Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. Genes Dev. *11*, 2545–2556.

Kress, T.L., Krogan, N.J., and Guthrie, C. (2008). A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol. Cell *32*, 727–734.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3′ end processing and splicing. Mol. Cell *23*, 195–205.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Leeper, T.C., Qu, X., Lu, C., Moore, C., and Varani, G. (2010). Novel protein-protein contacts facilitate mRNA 3′-processing signal recognition by Rna15 and Hrp1. J. Mol. Biol. *401*, 334–349.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Luna, R., Rondón, A.G., and Aguilera, A. (2012). New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. Biochim. Biophys. Acta *1819*, 514–520.

Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J., and Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature *475*, 408–411.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3′-end processing. Cell. Mol. Life Sci. *65*, 1099–1122.

Mangus, D.A., Evans, M.C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol. *4*, 223.

Martin, G., Gruber, A.R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. Cell Rep. *1*, 753–763.

Martinson, H.G. (2011). An active role for splicing in 3′-end formation. Wiley Interdiscip Rev RNA *2*, 459–470.

Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D.E., and Cramer, P. (2012). The spt5 C-terminal region recruits yeast 3′ RNA cleavage factor I. Mol. Cell. Biol. *32*, 1321–1331.

Milek, M., Wyler, E., and Landthaler, M. (2012). Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Semin. Cell Dev. Biol. *23*, 206–212.

Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim. Biophys. Acta *1829*, 174–185.

Müller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat. Rev. Genet. *14*, 275–287.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol. Cell *36*, 178–191.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev. *25*, 1770–1782.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Riordan, D.P., Herschlag, D., and Brown, P.O. (2011). Identification of RNA recognition elements in the Saccharomyces cerevisiae transcriptome. Nucleic Acids Res. *39*, 1501–1509.

Rodríguez-Navarro, S., and Hurt, E. (2011). Linking gene regulation to mRNA production and export. Curr. Opin. Cell Biol. *23*, 302–309.

Ruiz-Echevarría, M.J., and Peltz, S.W. (2000). The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell *101*, 741–751.

Ruiz-Echevarría, M.J., González, C.I., and Peltz, S.W. (1998). Identifying the right stop: determining how the surveillance complex recognizes and degrades an aberrant mRNA. EMBO J. *17*, 575–589.

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., and Tollervey, D. (2012). Transcriptome-wide analysis of exosome targets. Mol. Cell *48*, 422–433.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Lührmann, R., and Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. EMBO J. *16*, 3256–3271.

Shen, H., and Green, M.R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes Dev. *20*, 1755–1765.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. Mol. Cell *33*, 365–376.

Stewart, M. (2010). Nuclear export of mRNA. Trends Biochem. Sci. *35*, 609–617.

Strässer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A.G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304–308.

Tuck, A.C., and Tollervey, D. (2013). A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell *154*, 996–1009.

Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376–386.

Vasudevan, S., and Peltz, S.W. (2001). Regulated ARE-mediated mRNA decay in Saccharomyces cerevisiae. Mol. Cell *7*, 1191–1200.

Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Mol. Cell. Biol. *21*, 8346–8356.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Wery, M., Kwapisz, M., and Morillon, A. (2011). Noncoding RNAs in gene regulation. Wiley Interdiscip. Rev. Syst. Biol. Med. *3*, 728–738.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, 3.

Windgassen, M., and Krebber, H. (2003). Identification of Gbp2 as a novel poly(A)+ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. EMBO Rep. *4*, 278–283.

Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, N.M., and Ule, J. (2013). Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell *152*, 453–466.