Protein-RNA networks revealed through covalent RNA marks

Christopher P Lapointe1, Daniel Wilinski1,2, Harriet A J Saunders1 & Marvin Wickens1

Protein-RNA networks are ubiquitous and central in biological control. We present an approach termed RNA Tagging that enables the user to identify protein–RNA interactions in vivo by analyzing purified cellular RNA, without protein purification or cross-linking. An RNA-binding protein of interest is fused to an enzyme that adds uridines to the end of RNA. RNA targets bound by the chimeric protein in vivo are covalently marked with uridines and subsequently identified from extracted RNA via high-throughput sequencing. We used this approach to identify hundreds of RNAs bound by a Saccharomyces cerevisiae PUF protein, Puf3p. The results showed that although RNA-binding proteins productively bind specific RNAs to control their function, they also ‘sample’ RNAs without exerting a regulatory effect. We used the method to uncover hundreds of new and likely regulated targets for a protein without canonical RNA-binding domains, Bfr1p. RNA Tagging is well suited to detect and analyze protein-RNA networks in vivo.

Proteins bind to and regulate RNAs, governing RNA processing, transport, translation and decay. A single protein can bind and control hundreds of RNAs, and a single RNA molecule can bind many proteins. These protein–RNA networks are essential, and their misregulation can lead to defects in cell function and to human disease. Global mapping of protein–RNA interactions across the proteome and transcriptome is thus a central goal.

Over the past decade, powerful RNA immunoprecipitation (RIP)-based approaches have made it possible to identify RNAs bound by a specific protein1. In RIP, RNA-binding proteins are immunopurified from cell lysates, and associated RNAs are identified by microarray or deep sequencing2–3. UV cross-linking before immunoprecipitation (CLIP) covalently links interacting proteins and RNAs, which facilitates their purification4–7. CLIP also makes use of partial RNase digestion of bound RNA to determine global binding sites for particular proteins5–7.

Despite their utility and strength, RIP and CLIP approaches have limitations. Protein–RNA complexes must be purified from cell lysates using antibodies directed to endogenous or epitope-tagged proteins. RIP, which requires native conditions, is susceptible to nonphysiological interactions in vitro8–10.

In CLIP, UV cross-linking is relatively inefficient or requires nucleotide analogs to enhance efficiency6,11,12. CLIP also requires numerous enzymatic steps. Moreover, because transient interactions are permanently captured by cross-linking, biologically meaningful interactions are difficult to distinguish from those that are not10.

We sought a method for identifying global protein–RNA interactions in vivo in which interactions were unambiguous and occurred inside the cell. The approach we report here, termed RNA Tagging, is independent of protein purification, cross-linking or radioactive-labeling steps. We used the approach to identify RNAs bound by two S. cerevisiae proteins, Puf3p and Bfr1p.

RESULTS
The RNA Tagging approach
To detect and probe protein–RNA interactions in vivo, we developed RNA Tagging. The key principle of the method is that binding of a protein to an RNA in vivo leaves a covalent mark on the RNA, which is subsequently detected in vitro. In its simplest application, an RNA-binding protein (RBP) is fused to the Caenorhabditis elegans poly(U) polymerase PUP-2 (Fig. 1a). This enzyme lacks RNA-binding domains and therefore does not uridylate RNA efficiently on its own13,14. As a result, the chimeric protein covalently ‘tags’ only the RNAs to which the RBP binds. Tagged RNAs bearing varied numbers of uridines (the ‘U-tag’) are identified from the pool of total RNA by targeted or high-throughput sequencing assays, facilitated by a reverse-transcription step that is selective for uridyalted RNAs (Fig. 1b).

Targeted detection of RNA Tagging
We first implemented RNA Tagging in S. cerevisiae and focused on the PUF protein Puf3p. This protein recognizes a well-defined sequence in hundreds of mRNA targets important for mitochondrial functions15–21. To create the RNA Tagging chimera, termed PUF3-PUP (where “PUP” stands for poly(U) polymerase), we inserted the pup-2 open reading frame downstream of PUF3 at its native locus in the S. cerevisiae genome.

We initially examined tagging of two known targets of Puf3p: HSP10 and COX17 mRNA15,17. We grew strains that expressed

---

1Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin, USA. 2Present address: Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, USA. Correspondence should be addressed to M.W. (wickens@biochem.wisc.edu).

RECEIVED 14 JULY; ACCEPTED 5 OCTOBER; PUBLISHED ONLINE 2 NOVEMBER 2015; DOI:10.1038/NMETH.3651
wild-type PUF3-PUP or a mutant PUF3-PUP chimera with a catalytically inactive PUP to mid-log phase and lysed cells under denaturing conditions. We next performed parallel RT-PCR assays on HSP10 and COX17 mRNA (Supplementary Fig. 1a). PUF3-PUP deposited U-tags on both mRNAs (Supplementary Fig. 1b,c).

A primer selective for uridylated RNAs (U-select primer) yielded prominent PCR products only in cells that expressed the wild-type chimeric protein. As controls, a primer selective for polyadenylated RNAs detected the mRNAs in all samples, and the mutant chimera failed to tag HSP10. The presence of the U-tag on HSP10 mRNA was confirmed by directed sequencing (Supplementary Fig. 1d). Similarly, a PUF5-PUP2 chimera added uridines to endogenous, wild-type PHD1 mRNA, a known target, but not to the same mRNA with mutant binding elements, which was confirmed by deep sequencing as described below (Supplementary Fig. 1e,f).

**Transcriptome-wide RNA Tagging**

To implement RNA Tagging transcriptome-wide, we grew yeast strains that expressed PUF3-PUP to mid-log phase and isolated RNA (Fig. 1a). We then enriched mRNAs and added 3’-terminal G and 1 nucleotides to serve as a 3’ adaptor (G-I tailing).23,24 Inosines were included to reduce the stability of potential G-quadruplexes.24 Next, we reverse-transcribed the G-I-tailed RNA using the U-select primer, synthesized the second strand of DNA, PCR-amplified the dsDNA and size-selected the PCR products using solid-phase reversible immobilization beads. DNA libraries were paired-end sequenced on an Illumina HiSeq 2500 instrument.

We identified tagged RNAs using a computational approach. We used the first sequencing read (read 1) to assign reads to particular genes, and we used the second sequencing read (read 2) to identify the 3’-terminal nucleotides (Fig. 1c,d). RNAs with U-tags, termed tagged RNAs, were defined as RNAs that ended in at least eight adenosines not encoded in the genome (the poly(A) tail) followed by at least one uridine not encoded in the genome or the U-select primer. To ensure that U-tags of various lengths were accurately detected, we sequenced synthetic DNA libraries with known numbers of uridines. The libraries contained the adaptor sequences, a poly(A)12 tail and variable-length U-tags (Supplementary Fig. 2). The synthetic U-tags were accurately measured and readily distinguished (Fig. 1e).

**RNA Tagging identified global Puf3p targets**

Analysis of the PUF3-PUP tagging strain yielded a set of tagged RNAs. Of the approximately ten million reads, about 50% aligned uniquely to a single location in the yeast genome (‘uniquely mapped’). We detected just over 1 million tagged RNAs, which corresponded to one to more than ten nucleotides in length, and U-tags of all lengths were enriched approximately 500- to 1,800-fold in the PUF3-PUP strain relative to a control strain (Fig. 2a). The number of uridines in the PUF3-PUP strain was highly reproducible ($\rho = 0.93$, $P = 0$) (Fig. 2b). The number of TRPM detected for each Puf3p target was moderately correlated with the mean U-tag length ($\rho = 0.5$, $P = 0$) and was not correlated with RNA abundance (Supplementary Fig. 3a,b). The number of uridines in the U-tag was weakly and inversely correlated with RNA abundance (Fig. 3a).
Supplementary Fig. 3c

Figure 2 | RNA Tagging identified transcriptome-wide Puf3p targets. (a) Enrichment of tagged RNAs detected across different U-tag lengths in PUF3-PUP yeast relative to a control yeast strain (BY4742). (b) Scatter plot of tagged RNAs detected in the PUF3-PUP strain versus in the control strain (BY4742). (c) Plot of the number of tagged RNAs detected for the 476 Puf3p targets in two biological replicates. Spearman’s correlation coefficient (ρ) is indicated (ρ = 0.93, P = 0, n = 476). (d) Proportional Venn diagram depicting the overlap between Puf3p targets identified by RNA Tagging and those identified by other approaches. (e) Selected GO-term enrichments (1/P value) of Puf3p targets identified by RNA Tagging, RIP-chip15 and PAR-CLIP25. For simplicity, only three biological process terms are shown (a full list of enriched terms is provided in Supplementary Data 1). (f) Enriched sequence motifs, determined by MEME, in the 3′ UTRs of Puf3p targets identified by RNA Tagging and RIP-chip15, and in the PAR-CLIP peaks25. Percentages indicate the fraction of 3′ UTRs in each set that contributed to the motif.

(ρ = −0.37, P = 0) (Supplementary Fig. 3c). The set of RNA Tagging targets significantly overlapped with those identified by RIP-chip15 and photoactivatable ribonucleoside–enhanced CLIP (PAR-CLIP)25 (hypergeometric tests, all P < 2.2 × 10−16 (Fig. 2d)). Furthermore, Gene Ontology (GO) analyses revealed that Puf3p targets were greatly enriched for mitochondrial functions, similar to the previously identified targets (Fig. 2e).

Puf3p targets identified by RNA Tagging were highly enriched for Puf3p-binding elements. Using the unbiased algorithm Multiple Em for Motif Elicitation (MEME)26, we determined that Puf3p targets identified by RNA Tagging were highly enriched for Puf3p-binding elements in their 3′ UTRs (Fig. 2f). Puf3p tagged approximately 70% (170 of 246) of mRNAs with the consensus sequence CHUGUAHAUA in their 3′ UTRs, which represents the highest-affinity Puf3p-binding elements16. The binding element present in targets identified by RNA Tagging was similar to the one identified in the RIP-chip targets, whereas the PAR-CLIP targets yielded a more degenerate element (Fig. 2f and Supplementary Fig. 4).

The above data demonstrate that RNA Tagging globally identified protein–RNA interactions in vivo. The approach reproducibly identified more than 400 mRNAs bound by Puf3p in the cell, and these were highly enriched for the expected mitochondrial functions and Puf3p-binding elements.

RNA Tagging and binding affinity

We hypothesized that RNA Tagging might reveal the relative affinity of Puf3p for its different targets in the cell. For example, high-affinity targets would have relatively long interactions with PUF3-PUP, providing ample time for long U-tags to be added to the RNA. In contrast, low-affinity targets would have relatively brief interactions with PUF3-PUP, resulting in shorter U-tags.

To test this hypothesis, we used a two-dimensional ranking of Puf3p targets uniquely enabled by the RNA Tagging approach. We considered two attributes of the targets: the number of tagged RNAs detected, and the number of uridines added. On the basis of these two parameters, we hierarchically clustered Puf3p targets by the number of tagged RNAs detected at increasing U-tag lengths. Clustering results were visualized in a heat map, with the highest ranked target at the top (Fig. 3a). As expected, target rank was strongly correlated with TRPM (ρ = −0.91, P = 0) and U-tag length (ρ = −0.75, P = 0) (Supplementary Fig. 5a,b). Target rank was largely uncorrelated with RNA abundance (Supplementary Fig. 5c).

Puf3p targets are a continuum, but to facilitate downstream analyses, we separated them into three distinct groups, referred to as classes. Puf3p target classes were defined using the dendrogram from the clustering analysis and sequential statistical analyses (Online Methods). Class A Puf3p targets, which consisted of the highest-ranked genes, had the most TRPM detected and the longest U-tags (Fig. 3a). They possessed nearly perfect Puf3p-binding elements in their 3′ UTRs (Fig. 3a), dramatically exemplified by the cytosine enrichment at the −2 position, which enhances Puf3p binding in vitro and PUF3-dependent regulation in vivo16,27. In contrast, class C was the lowest-ranked group, and these targets had the fewest TRPM and shortest U-tags. Class C targets contained degenerate binding elements in their 3′ UTRs (Fig. 3a) and were expressed more highly than class A or B targets (Supplementary Fig. 6a). They also lacked enriched Puf3p-binding elements in their 5′ UTRs or open reading frames, which agrees well with the propensity of PUF proteins to bind 3′ UTRs15,22,28,29. The average position of the binding elements in the 3′ UTRs of targets was nearly identical across classes (Supplementary Fig. 6b,c). Similarly, the number of Puf3p targets identified by RNA Tagging, RIP-chip, and PAR-CLIP was highest for mitochondrial functions and Puf3p-binding elements.
of tagged RNAs and the number of uridines detected on target RNAs were not correlated with the distance from the binding element to the 3′ terminus of the transcript (Supplementary Fig. 6d,e).

The rank of targets correlated well with their measured binding affinities in vitro. We compared the median RNA Tagging rank of targets with six specific binding elements to the in vitro binding affinities of purified Puf3p for those same sequences16 (Supplementary Fig. 7a). Median target rank correlated well with \( K_d \) \((r = 0.98, P = 0.0009; \rho = 0.94, P = 0.0048)\) (Fig. 3b). Similarly, \( K_d \) was correlated with TRPM and U-tag length (Supplementary Fig. 7b,c). Comparisons of \( K_d \) to RNA abundance and the distances from binding elements to 3′ termini or stop codons yielded no significant correlations. Randomized data also yielded no significant correlations for any of the above analyses.

These findings support the hypothesis that RNA Tagging reveals high- and low-affinity targets in vivo. This is demonstrated by the covariation of target rank (and hence class) with the quality of Puf3p-binding elements and with binding affinity measured in vitro.

RNA Tagging distinguished regulation from ‘sampling’

We next examined the relationship between affinity and in vivo regulation. Puf3p is required for localization of specific mRNAs to mitochondria \(18,19\) and regulates mitochondrial function \(20,21\). Puf3p also destabilizes some of its target mRNAs \(16,17,27,30,31\). We hypothesized that class A Puf3p targets, which were the best-detected RNA Tagging targets and bound with the highest affinities, would exhibit the greatest enrichment for mitochondrial association and PUF3-dependent stability, whereas class C targets would exhibit the least.

Puf3p-target classes correlated with localized translation at mitochondria. We mined published data that identified mRNAs \(18\) and proteins \(32\) localized to mitochondria. Class A Puf3p targets were significantly enriched for mRNAs and proteins localized...
to mitochondria (hypergeometric tests, all $P < 2.2 \times 10^{-16}$) (Fig. 3c). Enrichment steadily decreased from class A to class C targets. We also mined recently published data that identified mRNAs translated by ribosomes localized to the outer mitochondrial surface, captured through proximity-specific ribosome profiling35. Puf3p targets were significantly enriched for mRNAs translated at mitochondria (Kolmogorov-Smirnov tests, all $P < 2.2 \times 10^{-16}$) (Fig. 3d). Notably, classes A and B were highly enriched, whereas class C was weakly enriched. Trends were similar without the translation inhibitor cycloheximide, which confirmed that Puf3p targets are actively translated at mitochondria (Supplementary Fig. 8).

Puf3p-target classes also correlated with sensitivity to deletion of PUF3. We mined published microarray experiments that measured global changes in mRNA abundance and decay rate in wild-type and puf3Δ strains34. Puf3p targets identified by RNA Tagging were significantly more abundant and more stable in the puf3Δ strain relative to all other mRNAs (Kolmogorov-Smirnov tests, all $P < 2.2 \times 10^{-16}$) (Fig. 3e,f). Enrichments for both abundance and stability progressively decreased across Puf3p target classes, with class A targets exhibiting the greatest effects. Class C targets were hardly enriched for effects of PUF3 on either abundance or stability. All specific mRNAs previously shown to be stabilized in a puf3Δ strain were class A or class B targets, which independently corroborated our meta-analysis of the global experiments37 (Supplementary Fig. 9).

The correlation between Puf3p-target class and both known Puf3p biological function and binding affinity suggests that the highest-ranked Puf3p RNA Tagging targets are those that are bound and regulated in vivo. In contrast, the lowest-ranked targets, which have degenerate or perhaps less accessible binding elements, are bound very weakly. The fact that these RNAs (class C) were tagged indicates that they were bound, yet they were largely unregulated. We refer to this behavior as ‘sampling,’ which we define as binding of the protein to RNA sufficiently long to tag it but insufficiently long to exert its regulatory effect—likely too briefly to recruit effector proteins or allow them to act. On average, the RNAs that were sampled were more abundant, which might help drive their interaction in vivo.

**RNA Tagging identified global Bfr1p targets**

We next implemented RNA Tagging to analyze Bfr1p, which lacks canonical RNA-binding domains. Bfr1p is implicated in the secretory pathway35,36 and is localized to the endoplasmic reticulum (ER) under normal conditions37,38 and to P-bodies after stress39. Bfr1p was also found to be associated with more than 1,000 mRNAs by RIP-chip29. Intriguingly, its reported mRNA targets were not enriched for targets with a role in the secretory pathway.

RNA Tagging with BFR1-PUP identified more than 1,000 functionally enriched tagged RNAs. As with Puf3p, tagged RNAs were highly enriched over many U-tag lengths (Fig. 4a). In the BFR1-PUP strain, 1,296 mRNAs and two small nucleolar RNAs (snR11 and snR31) were detected at levels above background in three biological replicates and were termed Bfr1p targets (Online Methods) (Fig. 4b). TRPM values were reproducibly detected across replicates (all pairwise $p \geq 0.84$) (Fig. 4c). TRPM, U-tag length and RNA abundance were all largely uncorrelated (Supplementary Fig. 10). Approximately 30% of the targets had been previously identified by RIP-chip39, which represents a significant overlap (hypergeometric test, $P < 2.2 \times 10^{-16}$) (Fig. 4d). Unlike Puf3p targets, Bfr1p targets identified by RNA Tagging lacked a defined binding element.

As determined by GO analyses, RNA Tagging targets were much more functionally enriched than those identified by RIP-chip. RNA Tagging targets were greatly enriched for cytoplasmic translation and membrane-associated functions, whereas RIP-chip targets were at most weakly enriched (Fig. 4e). Deeper dissection showed that targets uniquely identified by RNA Tagging, as well as those identified by both RNA Tagging and RIP-chip, were similarly enriched for membrane-associated functions and the term “cytoplasmic translation,” which predominately encompasses...
Supplementary Fig. 12. The weak correlation between target rank and the number of uridines in the U-tag indicated that in this case, unlike that of Puf3p, target rank was driven by TRPM.

Ribosomal proteins (Supplementary Fig. 11). In contrast, mRNAs uniquely identified by RIP-chip were enriched for ribosome biogenesis and the processing of noncoding RNAs.

Bfr1p binds mRNAs translated at the ER

To more closely examine Bfr1p targets, we performed a two-dimensional analysis with Bfr1p targets as we did with Pu3p. Bfr1p targets were grouped into four classes, classes A–D, with class A again containing the highest-ranked targets (Fig. 5a). Target rank was strongly correlated with TRPM ($\rho = -0.87$, $P = 0$), whereas target rank was weakly correlated with the average number of uridines in the U-tag and RNA abundance (Supplementary Fig. 12). The weak correlation between target rank and the number of uridines in the U-tag indicated that in this case, unlike that of Pu3p, target rank was driven by TRPM.

The highest-ranked Bfr1p targets were the most enriched for membrane-related functions. By mining published data, we found that class A targets were significantly enriched for proteins that are secreted, predicted to have a transmembrane domain and localized to the ER (hypergeometric test, all $P < 2.2 \times 10^{-16}$) (Fig. 5b–d). Enrichments progressively decreased from class A to class D targets. Furthermore, class A Bfr1p targets were the least enriched for mRNAs that encode proteins localized to the nucleus, nucleolus and mitochondria (Supplementary Fig. 13). These enrichments progressively increased across classes to levels near those expected by random chance. Bfr1p targets were also highly enriched for mRNAs found in P-bodies (hypergeometric test, $P < 2.2 \times 10^{-16}$) (Fig. 5e). The enrichment progressively decreased from class A to class C targets, but then slightly increased for class D targets.

The localization of Bfr1p to the ER, its presence on polysomes and the enrichment of its best targets for membrane-related proteins suggested that many of its targets would be translated at the ER. To test this, we mined recently published data that identified ribosome-occupied mRNAs specifically localized at the ER, captured by a proximity-specific ribosome-profiling experiment.
Bfr1p targets were highly enriched for abundant, ER-localized mRNAs. In comparison with all mRNAs, Bfr1p targets were significantly enriched for ER-localized translation, in contrast to Bfr1p targets identified by RIP-chip (Kolmogorov-Smirnov tests, all \( P < 2.2 \times 10^{-16} \)) (Fig. 5f). The enrichment of ER-localized translation progressively decreased from class A to class D targets. Bfr1p targets were similarly enriched for both Sec-dependent and Sec-independent translocation events (Kolmogorov-Smirnov tests, all \( P < 2.2 \times 10^{-15} \)) (Supplementary Fig. 14). Bfr1p bound about 60% of the approximately 700 mRNAs enriched for ER-localized translation, and the Bfr1p-bound mRNAs were significantly more abundant than those not bound by Bfr1p (Fisher-Pitman permutation test, \( P < 10^{-6} \)) (Supplementary Fig. 15).

Our findings illustrate that Bfr1p preferentially binds mRNAs that encode ribosomal and membrane-associated proteins, many of which are translated at the ER. These data clarify seemingly contradictory reports of Bfr1p function in vivo (reviewed in more detail in the Discussion).

**DISCUSSION**

RNA Tagging identifies targets of RNA-binding proteins in vivo, relying solely on the covalent marks left on the RNA. The approach is facile, reproducible and sensitive. Furthermore, RNA Tagging distinguishes between productive and nonproductive binding events in vivo, as the number of uridines added by the poly(U) polymerase is likely to be a direct reflection of the duration of the protein’s binding to the RNA. In organisms with endogenous enzymes that add and remove uridines, endogenous uridylated mRNAs are sufficiently stable to be detected and can be accounted for computationally using the same approach as described here. RNA Tagging is adaptable to specific cell types and tissues of living animals, as it requires minimal starting material and only purified RNA.

RNA Tagging can provide insight into the biological roles of RNA-binding proteins. Bfr1p predominately tagged mRNAs that encode ribosomal and membrane-associated proteins, enrichment missed in earlier RIP-chip studies. Bfr1p is part of a large protein complex and is required for the localization of mRNAs to P-bodies and the bud tip. Thus, our findings and previous studies suggest that Bfr1p is an integral component of a trafficking complex that localizes mRNAs to specific locations in the cell, particularly the ER.

RNA Tagging should facilitate access to areas of RNA biology that until now have been difficult to examine. For example, it might be possible to detect RNAs that are both directly and indirectly associated with a protein of interest, aided by the use of a poly(U) polymerase with its own intrinsic but weak RNA-binding activity. Large protein complexes often contain critical factors that only indirectly associate with RNA, such as several eukaryotic translation-initiation factors or components of the CCR4-NOT complex. The dynamics of RNA-protein interactions may be analyzed through rapid induction of the tagging protein, providing snapshots of the interactions at a given time. The development of new tagging enzymes that deposit different marks would enable multiple proteins of interest to be probed simultaneously, providing insight into the exchanges of proteins on RNAs, how RNA-binding proteins collaborate to regulate RNA, and the encounters of single RNA molecules in the cell. It remains to be seen whether PUP fusions bound to elements in the 5′ UTR will tag efficiently; flexible protein linkers or PUPs that possess higher rates of catalysis may be useful for this purpose. Regardless, the versatility of RNA Tagging should enable approaches to unexplored problems in RNA biology in living cells.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** NCBI Sequence Read Archive accession: SRP063022.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank members of the Wickens lab for helpful comments and suggestions throughout the work, and for their thoughtful discussions of the manuscript. We appreciate discussions with S. Kennedy (Harvard University), P. Anderson (University of Wisconsin–Madison) and their labs, as well as discussions and efforts with E. Grayhack and E. Phizicky (University of Rochester) in early stages of the work. We thank J. Kimble and E. Sorokin (University of Wisconsin–Madison) for use of a computational server, and L. Vanderpoel of the Biochemistry Media Lab for help with the figures. We also thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility, particularly M. Adams and M. Sussman, for high-throughput sequencing facilities and services. The work was supported by the US National Institutes of Health (grant GM50942) and by Wharton and Biochemistry Scholar Fellowships to C.P.L.

**AUTHOR CONTRIBUTIONS**

C.P.L. and M.W. conceived of the method and designed the experiments. C.P.L. performed the experiments. C.P.L., D.W., H.A.J.S. and M.W. analyzed the data. C.P.L. and M.W. wrote the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
12. Fecko, C.J. et al. Comparison of femtosecond laser and continuous wave UV sources for protein-nucleic acid crosslinking. Photochem. Photobiol. 83, 1394–1404 (2007).

13. Lapointe, C.P. & Wickens, M. The nucleic acid-binding domain and translational repression activity of a Xenopus terminal uridylyl transferase. J. Biol. Chem. 288, 20723–20733 (2013).

14. Kim, B. et al. TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. EMBO J. 34, 1801–1815 (2015).

15. Gerber, A.P., Herschlag, D. & Brown, P.O. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol. 2, e79 (2004).

16. Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M. & Hall, T.M. A 5′-cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. Proc. Natl. Acad. Sci. USA 106, 20192–20197 (2009).

17. Olivas, W. & Parker, R. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J. 19, 6602–6611 (2000).

18. Saint-Georges, Y. et al. Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. PLoS One 3, e2293 (2008).

19. Gadir, N., Haim-Vilensky, L., Kraut-Cohen, J. & Gerst, J.E. Localization of mRNAs coding for mitochondrial proteins in the yeast Saccharomyces cerevisiae. RNA 17, 1551–1565 (2011).

20. Chatenay-Lapointe, M. & Shadel, G.S. Repression of mitochondrial translation, respiration and a metabolic cycle-regulated gene, SSU1, by the yeast Pumilio-family protein Puf3p. PLoS One 6, e20441 (2011).

21. Garcia-Rodriguez, L.J., Gay, A.C. & Pon, L.A. Puf3p, a Pumilio family RNA binding protein, localizes to mitochondria and regulates mitochondrial biogenesis and motility in budding yeast. J. Cell Biol. 176, 197–207 (2007).

22. Wilinski, D. et al. RNA regulatory networks diversified through curvature of the PUF protein scaffold. Nat. Commun. 6, 8213 (2015).

23. Kusov, Y.Y., Shatirshvili, G., Dzagurov, G. & Gauss-Muller, V. A new G-tailing method for the determination of the poly(A) tail length applied to hepatitis A virus RNA. Nucleic Acids Res. 29, E57–7 (2001).

24. Lane, A.N., Chaîres, J.B., Gray, R.D. & Trent, J.O. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 36, 5482–5515 (2008).

25. Freeberg, M.A. et al. Pervasive and dynamic protein binding sites of the mRNA transcriptome in Saccharomyces cerevisiae. Genome Biol. 14, R13 (2013).

26. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Syst. Mol. Biol. 11, 133–127 (2010).

27. Miller, J.E. & Reese, J.C. Ccr4-Not complex: the control freak of eukaryotic mRNA decay. Nat. Rev. Mol. Cell Biol. 12, 153–166 (2011).

28. Wickens, M., Bernstein, D.S., Kimble, J. & Parker, R. A PUF family portrait: extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol. 2, e79 (2004).

29. Houshamandi, S.S. & Olivas, W.M. Yeast Puf3 mutants reveal the complexity of Puf-RNA binding and identify a loop required for regulation of mRNA decay. RNA 11, 1655–1666 (2005).

30. Huh, W.K. et al. Global analysis of protein localization in budding yeast. Nature 425, 686–691 (2003).

31. Williams, C.C., Jan, C.H. & Weissman, J.S. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science 346, 748–751 (2014).

32. Sun, M. et al. Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell 52, 52–62 (2013).

33. Jackson, C.L. & Kepes, F. BFR1, a multicopy suppressor of brefeldin A-induced lethality, is implicated in secretion and nuclear segregation in Saccharomyces cerevisiae. Genetics 137, 423–437 (1994).

34. Trautwein, M., Dengel, J., Schirle, M. & Spang, A. The polyamide-associated proteins Scp160 and Bfr1 prevent P body formation under normal growth conditions. J. Cell Sci. 127, 1992–2004 (2014).

35. Lang, B.D., Li, A., Black-Brewster, H.D. & Fridovich-Keil, J.L. The brefeldin A resistance protein Bfr1p is a component of polyribosome-associated mRNA complexes in yeast. Nucleic Acids Res. 29, 2567–2574 (2001).

36. Simpson, C.E., Lui, J., Kershaw, C.J., Sims, P.F. & Ashe, M.P. mRNA localization to P-bodies in yeast is bi-phasic with many mRNAs captured in a late Bfr1p-dependent wave. J. Cell Sci. 127, 1254–1262 (2014).

37. Ast, T., Cohen, G. & Schuldiner, M. A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. Cell 152, 1134–1145 (2013).

38. Mitchell, S.F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. Nat. Struct. Mol. Biol. 20, 127–133 (2013).

39. Jan, C.H., Williams, C.C. & Weissman, J.S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science 346, 1257521 (2015).

40. Munoz-Tello, P., Rajapp, L., Coquille, S. & Thore, S. Polyuridylation in eukaryotes: a 3′-end modification regulating RNA life. Biomed. Res. Int. 2015, 968127 (2015).

41. Norbury, C.J. Cytoplasmic RNA: a case of the tail wagging the dog. Nat. Rev. Mol. Cell Biol. 14, 643–653 (2013).

42. Chang, H., Lim, J., Ha, M. & Kim, V.N. TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. Mol. Cell 53, 1044–1052 (2014).

43. Newman, M.A., Mani, V. & Hammond, S.M. Deep sequencing of microRNA precursors reveals extensive 3′ end modification. RNA 17, 1795–1803 (2011).

44. Jackson, R.J., Hellen, C.U. & Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. 11, 113–127 (2010).

45. Miller, J.E. & Reese, J.C. Ccr4-Not complex: the control freak of eukaryotic cells. Crit. Rev. Biochem. Mol. Biol. 47, 315–333 (2012).
Online Methods

Yeast strains. All S. cerevisiae strains were constructed in BY4742 yeast (MAT\(\alpha\); his3A1; leu2Δ0; lys2Δ0; ura3Δ0). To construct RNA Tagging chimeras, we inserted the DNA sequence for the open reading frame of C. elegans pup-2 followed by a stop codon and the URA3 marker, including its native promoter and terminator sequences, in-frame at the 3′ end of PUF3 and BFR1 using standard yeast transformation techniques. The BFR1-PUF2 strains also contained a 3-HE epitope tag on the C terminus of the fusion protein. Catalytically inactive PUP2 strains (PUP2mut strains) had Asp185Ala and Asp187Ala substitutions in the PHD1 protein. For wild-type and mutant PHD1 strains, the endogenous 3′ UTR of PHD1 was replaced with URA3 using standard yeast transformation techniques. Next, single colonies were transformed with DNA that encoded an RGSH\(_6\) epitope tag fused to the C terminus of Phd1p and either wild-type or mutant PHD1 3′ UTRs, the latter of which had substitutions that disrupted known PuF5p-binding elements (UGUAGUUA to ACAAGUUA, and UGUAAACAUU to ACAAAACAUUA). Cells were selected on plates containing 5-FOA. Integration of the epitope tag and 3′ UTRs at the endogenous PHD1 locus was confirmed by sequencing. The pup-2 open reading frame and a 3-HE epitope tag were then inserted in-frame at the 3′ end of PUF3 as above in both the wild-type and mutant PHD1 strains.

Yeast growth and total RNA isolation. We grew all strains by inoculating 5 mL of yeast extract–peptone–dextrose plus adenine (YPAD) cultures with frozen yeast strains or freshly streaked colonies and incubating them at 30 °C and 180 r.p.m. until ~24 h, we seeded 25 mL of YPAD cultures with frozen yeast strains or freshly streaked colonies and incubating them at 30 °C and 180 r.p.m. After ~24 h, we seeded 25 mL of YPAD cultures at A\(_{600}\) ~ 0.0002 and grew them at 30 °C and 180 r.p.m. until A\(_{600}\) 0.5–0.8. Yeast were harvested by centrifugation for 10 min, 3,000 r.p.m. at 4 °C, and the pellets were washed once with 40 mL of ice-cold water. Cells were resuspended in 500 µL of RNA ISO buffer (0.2 M Tris-HCl, pH 7.5, 0.5 M NaCl, 0.01 M EDTA, 1% SDS). Then we added ~200 µL of acid-washed beads and 500 µL of phenol:chloroform:isoamyl alcohol (25:24:1) (PCA). Cells were lysed by vortexing for 20 s at room temperature followed by 20 s on ice ten times. Samples were then separated from the beads and split evenly into two tubes, and 375 µL of RNA ISO buffer and 375 µL of PCA were added to each tube. Samples were mixed by gentle shaking and were separated by centrifugation for 15 min at 15,000 r.p.m. at 4 °C. The aqueous layer was removed (~500 µL) and further extracted by two additional extractions (PCA followed by chloroform). After the extractions, the aqueous layer was removed and ~1 mL of 100% ethanol was added to the samples, which were gently mixed and incubated at ~50 °C for >1 h. Total RNA was pelleted by centrifugation for 30 min at 15,000 r.p.m. at 4 °C. Pellets were washed once with ~70% ethanol and resuspended in 43 µL of water. Separate tubes for each sample were then recombined and treated with 8 units of TURBO DNase (Life Technologies) for 1 h at 37 °C. Total RNA was purified using the GeneJet RNA Purification kit (Thermo Fisher Scientific) and eluted in 30 µL of water. RNA samples were stored at −80 °C until use.

Targeted RNA Tagging RT-PCR assays. Terminator treatment. To deplete rRNA, we treated 2 µg of total RNA with 2 units of Terminator enzyme (Epicentre) for 60 min at 30 °C. The reactions were subsequently purified using 1.8 volumes of room-temperature RNA Clean XP buffer (Agencourt) and the standard protocol. RNA-depleted RNA was eluted in 12 µL of water.

G-Tailing. Terminator-treated samples were G-1-tailed using 1,200 units of yeast poly(A) polymerase (Affymetrix), 0.5 mM GTP and 0.15 mM ITP and incubated at 37 °C for 90 min. Samples were diluted to 100 µL with water, and G-1–tailed RNA was extracted with two sequential organic extractions (PCA followed by chloroform). The final aqueous layer was removed, and 10 µL of 3 M sodium acetate, 1 µL of GlycoBlue (Life Technologies) and 600 µL of 100% ethanol were added to the samples. Samples were incubated at ~50 °C for >1 h. Samples were pelleted by centrifugation for 30 min at 15,000 r.p.m. at 4 °C. Pellets were washed once in ~70% ethanol and resuspended in 10 µL of water.

Selective reverse transcription. G-1–tailed samples were selectively reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) under nearly standard conditions. The G-1–tailed samples were split equally (typically 3 µL) across all RT reactions. 3 µL of samples were added to 1 µL of 1 mM U-select primer (G CTTGGACCCGAGAATTCCACCCCCCCCACAAA), 1 µL of 10 mM dNTP mix and 8 µL of water (13 µL total). Oligo-(dT) and −RT reactions used 1 µL of 1 mM oligo-(dT)\(_{42}\) (TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT) in place of the U-select primer. A master mix of 4 µL of 5× reaction buffer, 1 µL of 100 mM DTT and 1 µL of 40 U/µL RNase inhibitor per reaction was prepared separately. The primer–RNA mixes and the master mix were incubated at 65 °C for 5 min followed by 5 min at 50 °C in a thermocycler. With the primer–RNA mixes and the master mix still in the 50 °C thermocycler, RT enzyme was added to the master mix (except for −RT samples) and mixed thoroughly, and 7 µL of the resulting master mix was added to the primer–RNA mix. Samples were then incubated at 50 °C for 60 min followed by 5 min at 85 °C.

Polymerase chain reactions. 1 µL of cDNA straight from the RT reactions was PCR amplified using GoTaq Polymerase (Promega). The HSP10 specific forward primer was GACAGCATCCGGTTGTGTATT. The HSP10 specific reverse primer was TTTTCTTGTCTACATAATTGTGCCC. HSP10 primers and the U-select primer were used at final concentrations of ~1 µM and ~40 nM, respectively. The COX17 specific forward primer was ATGACTGAAACTGACAAGAAAC when used with the U-select primer. The internal COX17 primers were ACAAGAACAAGAAAACCCACGC and AAGATGCAATGTATCCCGCTC. All COX17 reactions were performed with final primer concentrations of ~40 nM. PCR parameters and steps were as follows: (1) 95 °C for 3 min, (2) 95 °C for 30 s, (3) 50 °C for 30 s, (4) 72 °C for 90 s, (5) repeat steps 2–4 24 times (HSP10) or 36 times (COX17), (6) 72 °C for 5 min, and (7) hold at 4 °C.

Cloning and Sanger sequencing. HSP10 PCR products were cloned using the TOPO-TA cloning kit (Life Technologies), standard reaction conditions and blue-white colony screening. Individual white colonies were grown in 5 mL of lysogeny broth (LB)-ampicillin media. Plasmids were isolated from saturated cultures with the GeneJET Plasmid Miniprep kit (Thermo Scientific (Fermentas)) and subsequently Sanger sequenced using standard reaction conditions.

Transcriptome-wide RNA Tagging library preparations. Poly(A) selection and rRNA depletion. Approximately 75 µg of
high-quality total RNA were poly(A) selected using the Dynabeads mRNA purification kit (Life Technologies) and the standard protocol. Samples were eluted in 28 μL of water. The poly(A)-selected RNA was then depleted of rRNA using the RibozeroGold (yeast) kit (Epicentre) and the standard protocol. Samples were eluted in 12 μL of water.

G-I tailing. Samples were G-I tailed as above, except for the following step: After the initial 90-min G-I–tailing reaction, an additional 1,200 units of yeast poly(A) polymerase was added to the reactions and incubated for an additional 30 min at 37 °C. G-I–tailed RNA was purified as above using PCA.

Selective reverse transcription and RNase H digestion. G-I–tailed samples were selectively reverse transcribed as above. cDNAs were digested with 1 μL of RNaseH (Invitrogen) for 20 min at 37 °C. cDNAs were purified using the GeneJet PCR purification kit (Thermo Fisher Scientific). cDNAs were eluted twice in 32 μL of water, giving a total of ~60 μL cDNA.

Second-strand synthesis. We added 60 μL of cDNA to 10 μL of 10× Klenow buffer (500 mM Tris–HCl, pH 7.5, 100 mM MgCl2, 10 mM DTT, 0.5 mg/mL BSA), 12 μL of water, 5 μL of 10 mM dNTPs, 10 μL of 10 μM second-strand synthesis primer (GTTCAGAGTTCTACAGTCCGA) and 3 μL of 5 U/μL Exo–Klenow DNA polymerase (Life Technologies). Reactions were incubated at 37 °C for 30 min and then purified twice using RNA Clean XP beads (Agencourt) at a 1:1 (bead:reaction) ratio. dsDNA was eluted in 50 μL of water.

Polymerase chain reactions. Samples were PCR amplified using GoTag polymerase (Promega). 5 μL of cDNA was added to 8.33 μL of 2× GoTagGreen master mix, 2 μL of water, 0.67 μL of 10 μM RP1 primer (AATGATACGGCGACCACCGAGATCTACACGTTCAAGCGAGTTCTACAGTCCGA) and 0.67 μL of 10 μM barcoded primer (CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGACTGGGTGGGGGGGGGTGGAATTCTCGGGTGCCAAGG) and the poly(A) tail sequence (AAAAAAAAAA) were removed from read 1 sequences using FASTA/Q Clipper (fastx_clipper -a sequence -l 15 -n -I -v input -o output -Q 34). Any read 1 sequences that were shorter than 15 nt after removal of either sequence were discarded. Samples were sequenced on an Illumina HiSeq 2500 instrument to obtain 50-bp paired-end read data sets. Throughout, the first sequencing read, which covered the 5′ end of the sequenced DNA fragment, is termed read 1, and the second sequencing read, which covered the 3′ end of the sequenced DNA fragment, is termed read 2. Raw data were deposited at the NCBI Sequence Read Archive (accession SRP063022).

FASTQ file manipulation and alignment. Read 1. All FASTQ processing (FASTX-toolkit, http://hannonlab.cshl.edu/fastx-toolkit/) and alignments to the yeast genome were done using local installations of the given software. The U-select primer sequence (TTTGGGGGGGGTGAATCTCGGGGTGCCAAGG) and the poly(A) tail sequence (AAAAA) were removed from read 1 sequences using FASTA/Q Clipper (fastx_clipper -a sequence -l 15 -n -I -v input -o output -Q 34). Any read 1 sequences that were shorter than 15 nt after removal of either sequence were discarded. Read 1 sequences were then aligned to the S. cerevisiae genome (version R64-1-1) using Bowtie49 with the following parameters: a seed length (-l) of 25 nt, no more than two mismatches (-n) and only a single reportable alignment (-m) in the genome (bowtie -t genome input output input output -Q 34). Reads that aligned to more than one location were discarded.

Read 2. The 5′ adaptor sequence (GATCGTGACTGTA GAACTCTGAAC) was removed from read 2 sequences using FASTA/Q Clipper and the same parameters as above. The last six nucleotides of the resulting read 2 sequences, which represented the random hexamer sequence from the second-strand synthesis, were then removed using FASTA/Q Trimmer (fastx_trimmer -t 6 -i input -o output -Q 34). The resulting read 2 sequences were reverse-complemented using FASTA/Q Reverse Complement (fastx_reverse_complement -i input -o output -Q 34), and any sequence corresponding to the U-select primer sequence was removed as above. Sequences with at least three adenosines following any number of uridines at their 3′ end (A-U tail sequences) were identified using regular expression searches in Perl. Read 2 sequences were aligned twice to the yeast genome: first without any A-U tail sequence, and then with any A-U tail sequence. This alignment process identified read 2 sequences with A-U tail sequences that were not encoded in the genome. Bowtie alignments were conducted essentially as above, except that the seed length was 20 nt and the −v alignment mode was used to exclude reads with three or more mismatches.

doi:10.1038/nmeth.3651
Definition of tagged RNAs. A tagged RNA was defined as a DNA fragment with a sequence that aligned uniquely to the yeast genome and contained at least eight adenosines followed by at least one uridine at the 3′ end that were not encoded by any adaptor sequence or the genome. Typically, read 1 identified the genomic location of a tagged RNA, whereas read 2 identified its A-U tail sequence. Read 2 also frequently determined the 3′ terminus of an RNA. The number of tagged RNAs per gene was calculated and normalized across samples (TRPM). Where indicated, TRPM enrichment was calculated as a ratio of TRPM obtained in strains with the relevant RBP-PUP chimera to that in strains without it.

Reproducible RNA Tagging targets. To be identified as a target, genes with tagged RNAs had to meet three criteria. First, the number of TRPM detected for a particular gene had to be at least tenfold greater than the number of TRPM detected for that gene in the non-tagging control sample. Second, the number of TRPM detected for a particular gene had to be greater than the error rate for falsely detecting tagged RNAs. A uridine was erroneously detected 3% of the time on a synthetic polyadenylated library without a U-tag (Supplementary Fig. 2b) (synthetic libraries are discussed above). Thus, the error rate was defined as the number of TRPM detected by error per gene (0.03 × (total number of replicates)). The total number of uridines detected for a particular gene had to be greater than the error rate (TRPM detection rate) × (number of uridines in the U-tag). If the number of TRPM detected for a particular gene had to be at least tenfold greater than the number of TRPM detected for that gene in the non-tagging control sample. Second, the number of TRPM detected for a particular gene had to be greater than the error rate for falsely detecting tagged RNAs. A uridine was erroneously detected 3% of the time on a synthetic polyadenylated library without a U-tag (Supplementary Fig. 2b) (synthetic libraries are discussed above). Thus, the error rate was defined as the number of TRPM detected by error per gene (0.03 × (total number of TRPM)/(total number of genes with TRPM)). Third, a gene had to have met both of the above criteria in all of the biological replicates. Supplementary Data 3 and 4 include comprehensive target lists of Puf3p and Bfr1p, respectively. Supplementary Data 5 shows sequencing results of a control strain (BY4742) without any tagging chimeras. Where indicated in figures, we uniformly transformed (+2) TRPM values to facilitate plotting on a log scale.

Hierarchical clustering. The TRPM values for each target were calculated across U-tag lengths of 1–10 uridines for each sample. TRPM values for biological replicates were then averaged (mean). Each U-tag length encompassed all TRPM with at least the indicated number of uridines. Prior to clustering, the data were sorted from most to least TRPM detected with at least 1 U in the U-tag. The data sets were log-transformed and hierarchically clustered using the Gene Cluster 3.0 software. Heat maps were generated in Matlab (version R2014a).

Definition of target classes. To begin, classes were loosely defined to encompass groups of targets with similar TRPM and U-tag–length profiles. Boundaries between putative target classes were defined by the dendrogram from the clustering analysis. Statistical analyses (as outlined below) were conducted on each putative class, sequentially from the highest-ranked class to the lowest-ranked class, to determine whether it was distinct from directly adjacent putative classes. As an example, the enrichment of putative class A targets for a given observation (for example, RNAs with increased abundance in ΔPuf3) was compared to the enrichment in putative class B targets. If the enrichments of putative class A and class B targets were statistically indistinguishable, they were combined and the analysis was repeated with the next adjacent putative class (class C). If the enrichments of putative class A and class B targets were statistically different, putative class A targets were defined as actual class A targets, and the process was repeated with the remaining putative classes until only distinct classes remained.

Statistical analyses. All statistical analyses were done using RStudio (R version 3.1.2). Linear regression analyses were used to obtain R² values and the associated P values (summary(lm(y~x))). Shapiro-Wilk tests (shapiro.test(x)) were used to test normality as needed. Spearman’s (p) and Pearson’s (r) correlation coefficients and their associated P values were determined using the cor function from the hmisc package (rcorr(x, y, type = “spearman”) and rcorr(x, y, type = “pearson”), respectively). Hypergeometric distribution tests (phyper()) were used to determine whether the observed overlap between two data sets was significant. The total population size was defined as 6,607 genes, except in the following analyses: mRNA localization to mitochondria (6,256 genes), proteins with predicted transmembrane domains (TMHMM analyses, 6,713 genes), and yeast GFP localization (4,156 genes). Cumulative fraction plots were generated using the empirical cumulative distribution function (ecdf) (plot(ecdf(x), do.points = F, verticals = T, lty = 1, lwd = 3, …)). Two-sided Kolmogorov-Smirnov tests were performed using the ks.test function (ks.test(x,y)). For Supplementary Figures 6b,c and 15b, Fisher-Pitman permutation tests and permutations of the Wilcoxon–Mann–Whitney test were conducted using the “coin” package (pvalue(onesway_test(DV ~IV , distribution = approximate (B = 1000000))) and pvalue(wilcox_test(DV ~IV , distribution = approximate(B = 1000000))), respectively). Both tests behaved similarly for all comparisons. Where indicated, data were randomized 100,000 times using the “sample” function.

Venn diagrams. Proportional Venn diagrams were generated using Biovenn and then redrawn for publication.

MEME and directed motif searches. To be as inclusive as possible, we defined 3′ UTRs as the longest isoform for a particular gene previously observed or, if not previously defined, as 200 bases. MEME analyses were done on a local server using the following command: meme.bin input.txt -oc outputdirectory -dna -mod zoops -nmotifs 5 -minw 6 -maxw 15. The “maxsize” parameter was adjusted as needed. Enriched sequence motifs were identified in the 3′ UTRs of Puf3p targets and indicated subsets using MEME as described above. To determine the binding motif present in each class of Puf3p targets, we combined the binding elements present in each class, as determined using all of the Puf3p targets, to generate the motifs in Figure 3a. We conducted unbiased MEME analyses as described above on each of the classes to identify enriched motifs in the 5′ UTRs, open reading frames and 3′ UTRs, which confirmed the findings reported in Figure 3a. The RIP-chip motif was identified in the 3′ UTRs of the previously identified targets using MEME as above. The PAR-CLIP motif was previously identified but was shortened here for consistency. In all cases, motifs were prepared for publication using WebLogo 3 (ref. 52). The total number of genes with the C(AUC)UGUA(AUC)UAU consensus sequence in their 3′ UTR was determined using a Perl regular expression search on all 3′ UTR sequences. Genes with at least one occurrence of the motif were counted as positives.

DOI: 10.1038/nmeth.3651

© 2015 Nature America, Inc. All rights reserved.
Location of Puf3p-binding elements (PBEs) in 3′ UTRs. Many 3′ termini of mRNAs were detected in our data, especially when all RNAs that were detected with a poly(A) tail of at least eight adenosines (with or without a U-tag) were included. Using this information, we determined the most detected isoforms for particular mRNAs, the lengths of the 3′ UTRs and the position of the PBE relative to the stop codon and 3′ terminus. Genes with undetected 3′ termini and genes with negative or very large (>1,000 nt) distances to 3′ termini were excluded from the analyses. For Supplementary Figure 6d,e, the mean number of tagged RNAs, number of uridines added and distance from the PBE to the 3′ terminus for isoforms of 64 Puf3p targets (144 distinct mRNAs) detected by at least 31 reads (24,417 reads total) were calculated and compared. In these analyses, tagged RNAs with U-tags of more than six uridines were not analyzed because our deep sequencing did not yield 3′ termini for those mRNAs.

GO analyses. All GO analyses were completed using YeastMine from the Saccharomyces Genome Database (http://yeastmine.yeastgenome.org). All parameters were set to defaults (Holm-Bonferroni corrected). Puf3p and Bfr1p comprehensive GO-term data are available in Supplementary Data 1 and 2, respectively.

TMHMM prediction. To identify proteins with a predicted transmembrane domain, we downloaded the sequences of all proteins (6,713 proteins, including dubious proteins) from the Saccharomyces Genome Database. We then analyzed the sequences using the TMHMM 2.0 server53. Proteins with at least one predicted transmembrane domain were counted as positives.

RNA-seq. RNA isolation. Total RNA was isolated from S. cerevisiae (BY4742) cells by standard methods. We collected 50 ml of cells with A_{600} 0.5–0.8 by centrifugation at 3,200 r.p.m. at 4 °C, washed the cells once with cold water and snap-froze them in liquid N2. The tubes were vortexed for 30 s and then incubated on ice for 30 s; this process was repeated six times. The supernatant was removed, extracted with 1 mL of PCA and ethanol precipitated. RNA pellets were resuspended in 50 µL of water.

Library preparation. We used 2 µg of RNA as input. We depeled samples of rRNA using the Ribo-Zero Magnetic Gold Kit (Yeast) kit (Epicentre) and the standard protocol. Libraries were prepared using the TruSeq Stranded Total RNA kit (Illumina) and the standard protocol with 12 rounds of PCR. PCR samples were purified twice using RNA Clean XP beads and were eluted in 30 µL of water. Libraries were sequenced on an Illumina HiSeq 2000 to get 50-bp reads.

Data analysis. Mapped reads were assigned to genomic features by HTseq-count (htseq-count –s) (version 0.5.4p3). The mean number of fragments per kilobase of exon per million reads mapped (FPKM) of four biological replicates was calculated for each genomic feature (Supplementary Data 6).

49. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
50. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics 9, 488 (2008).
51. Xu, Z. et al. Bidirectional promoters generate pervasive transcription in yeast. Nature 457, 1033–1037 (2009).
52. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190 (2004).
53. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580 (2001).