

New strategies reveal key features of protein-RNA networks

By

Christopher P. Lapointe

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: December 8, 2016

The dissertation is approved by the following members of the Final Oral Committee:

Marvin Wickens, Professor, Department of Biochemistry

Judith Kimble, Professor, Department of Biochemistry

Michael Sheets, Professor, Department of Biomolecular Chemistry

Catherine Fox, Professor, Department of Biomolecular Chemistry

Philip Anderson, Professor, Department of Genetics

New strategies reveal key features of protein-RNA networks

Christopher P. Lapointe

Under the supervision of Professor Marvin Wickens

At the University of Wisconsin-Madison

Thesis Summary

Protein-RNA interactions are ubiquitous. They determine how much protein is produced from an mRNA, and when and where that occurs. Single RNA-binding proteins often bind to many RNAs. Multiple proteins can simultaneously bind to a single RNA molecule. The particular combination of proteins bound dictates the fate of that mRNA. Humans possess approximately 1,600 RNA-binding proteins and 26,000 genes that produce mRNAs, which are alternatively spliced to produce many different unique mRNA products. Such networks of protein-RNA interactions, referred to as “protein-RNA networks”, underlie fundamental cellular processes and have well-established connections to disease. The challenges now are to understand how they are formed, how they function, and how they are balanced *in vivo*.

I employed an integrated approach of biochemistry, molecular biology, genomics, genetics, and bioinformatics to develop new strategies for the dissection of protein-RNA networks. I began with structure-function analyses of an RNA-modifying enzyme (Chapter 2), which enabled me to develop a method to identify protein-RNA interactions *in vivo*, which we termed “RNA Tagging” (Chapter 3). I used RNA Tagging to reveal that while proteins productively bind specific RNAs to control their function, they also “sample” RNAs by binding briefly to them without exerting a regulatory effect. I next integrated RNA

Tagging with a different approach to identify RNAs bound by a particular protein called HITS-CLIP. In this meta-analysis, I demonstrated that the two approaches are complementary, and together they defined a core set of mRNAs controlled by an RNA-binding protein (Chapter 4). These studies enabled collaborative multi-omic analyses, which revealed that *Saccharomyces cerevisiae* Puf3p is a key post-transcriptional repressor of mitochondrial biogenesis factors, particularly for the respiratory chain and mitochondrial translation. In parallel, I employed RNA Tagging to dissect a protein-RNA network controlled by three related RNA-binding proteins – Puf3p, Puf4p, and Puf5p – in yeast (Chapter 5). I revealed that the architecture of the network is intricately balanced via an interplay between binding affinity and relative abundance of the protein and RNA molecules *in vivo*.

Together, my thesis research includes the development of a novel method and new strategies to analyze protein-RNA interactions that occur inside living cells. These approaches have several advantages over those previously described, provide new insights into how proteins regulate RNAs, and are widely applicable to protein-RNA networks throughout biology. They provide new opportunities to probe the nature and dynamics of protein-RNA networks in living cells.

Acknowledgements

I was incredibly fortunate to have the help, support, and encouragement of many people throughout my graduate education. Without them, getting to this point would have been impossible.

To Camila, thank you for coming on this journey with me and starting our lives together in Madison. I am so lucky to have shared with you our at-home lab meetings, cycling and triathlon training, backpacking and hiking adventures, and exploration of Madison. Most of all, thank you for your love, support, and help, which has allowed me to grow as both a person and a scientist. Without you, I wouldn't be who I am today. I look forward to seeing what our future holds.

To my advisor, Marv Wickens, thank you for giving me the opportunity to train in your lab. You have been a tremendous advisor, much more than anything I could have asked for. You helped me think outside the box, and gave me the time and freedom to solve fun puzzles in RNA biology. I am forever in your debt.

To Judith Kimble, thank you for: providing a second lab-home for me during the last year, and pushing me to think critically about my data and experiments. I am also very grateful for your support, advice, and help as I applied for post-docs.

To Mike Sheets, thank you for: serving on my committee throughout my entire time in grad school; sharing your knowledge about frogs, RBPs, and translational control; and your support, advice, and help during my post-doc search.

To Phil Anderson & Scott Kennedy, thank you for: letting me attend your lab meetings while Marv was on sabbatical, your thoughts and feedback as I struggled to get RNA Tagging working, and for being on my committee.

To Catherine Fox, thank you for joining my committee during my last stages of grad school.

To all the current and former members of the Wickens lab, thank you for making science fun and making it easy to come to lab every day. I'm also very grateful for everyone's help, advice, training, suggestions, and ideas throughout my projects. To Daniel Wilinski, thank you for: surviving the nearly 6 years as my bay mate, humoring me on our countless trips for coffee to talk science, teaching me about life as an adult, spending more time with me than Bethany, and being an amazing friend. To Trish Hoang (honorary Wickens lab member), thank you for being fun and energetic every day, and for making radioactive work enjoyable. To Shruti Waghay, thank you for: spending your last few months in the lab as my bay mate so I wouldn't be lonely after both Daniel and Harriet abandoned me, giving me all the latest info on the best restaurants in Madison, and convincing me California is the place to be. MiniPrep! To Melanie Preston, thank you for: being my across-the-bay mate, letting me hang with Maple and Nina, and for putting up with my tapping when I zoned out listening to music. Oh, yea, and Promega Box! To Zak Campbell, thank you for: Friday, all things puns, Money!, and making lab a fun place to be every day. To Cary Valley, thank you for: teaching me all things soccer related, the "two-handed" model, and trusting me to watch Indy. To Douglas Porter, thank you for sharing your perspectives on genomics and RBPs, your selflessness, and your ability to tell a joke and make me laugh when I was least expecting it. To Harriet Saunders, thank

you for teaching me I knew nothing about tea. To Natascha Buter, thank you for keeping the lab organized and running smoothly, and letting me get to know Lana, Kip, and Finn. To Amy Cooke, thank you for: training me, being incredibly fun, and making me feel truly welcome in the Wickens lab. To Craig Stumpf, thank you for telling me about Greenbush donuts. To Hugo Medina and Brian Carrick, thank you for sharing your enthusiasm, energy, new ideas, and perspectives with me.

To Laura Vanderploeg, thank you for: always being willing to drop whatever you are working on to answer my questions, all your help with displaying my data, and showing me it's possible to make amazing figures.

To Marie Adams and the UW Sequencing Facility, thank you for all your help as I tried to develop RNA Tagging.

To the IPIB program, Media center, and Biochem IT department, thank you for all of your support and help throughout my training.

To my friends, thank you for: making grad school fun, game nights, beers at the terrace, cycling in the country, getting married in Madison, football games, camping, and of course, friendsgivings.

And last but definitely not least, thank you to my family, especially my parents, Rhonda and Don; my brother, Pat; Roberto and Grace. Thank you for always being there for me, and for helping me achieve my goals.

Table of Contents

Thesis Summary	i
Acknowledgements	iii
Table of Contents	vi

Chapter 1: RNA, meet protein; Protein, RNA.

Title page	1
Ribonucleic Acid	2
Proteins bind to RNA	3
Cytoplasmic control of mRNA by proteins	5
Methods to identify protein-RNA interactions	6
Ribonucleotidyl transferases	9
PUF proteins	12
Brief outline	18
Figures and Tables	19
References	43

Chapter 2: The nucleic acid binding domain and translational repression activity of a *Xenopus* terminal uridylyl transferase

Title page	50
Abstract	51
Introduction	52
Results	54
Discussion	61
Methods	65
Acknowledgements	70
Footnotes	71
Figures and Tables	72

References.....	89
-----------------	----

Chapter 3: Protein-RNA networks revealed through covalent RNA marks

Title page	99
Abstract.....	100
Introduction	101
Results	103
Discussion.....	114
Methods	116
Acknowledgements	131
Competing financial interests	132
Figures and Tables.....	133
References.....	174

Chapter 4: An integrated and multi-omic strategy reveals Puf3p-mediated regulation of the mitochondrial respiratory chain

Title page	181
Abstract.....	182
Introduction	183
Results	188
Discussion.....	199
Methods	205
Figures and Tables.....	213
References.....	233

Chapter 5: Principles that control the architecture of a protein-RNA network

Title page	238
Abstract.....	239
Introduction	240

Results	242
Discussion	254
Methods	257
Acknowledgements	262
Figures and Tables.....	263
References.....	295

Chapter 6: Perspectives

Title page	301
Broader Perspectives	306
Figures	316
References.....	318

Appendix 1: RNA regulatory networks diversified through curvature of the PUF protein scaffold

Title page	320
Abstract.....	321
Introduction	322
Results	324
Discussion.....	334
Methods	337
Acknowledgements	343
Figures and Tables.....	344
References.....	384

Chapter 1

RNA, meet Protein; Protein, RNA.

I wrote this chapter and assembled the figures.

Ribonucleic acid (RNA)

RNA is vital and multi-functional. Messenger RNA (mRNA) serves as an essential intermediary between DNA and protein during gene expression (1). Through alternative splicing, mRNA allows single genes to encode different protein variants, greatly diversifying the cellular genome (2). Rather than encode proteins, so-called “non-coding” RNAs perform critical structural, catalytic, and regulatory roles (3-7). Non-coding RNAs are essential to the ribosome, the spliceosome, and telomerase (8-10). Transfer RNAs (tRNAs) decode mRNA in the ribosome, which allows proper production of protein (11). Other non-coding RNAs regulate gene expression at both the DNA and RNA level; they can control the amount of RNA transcribed from DNA at particular genes and target particular mRNAs for destruction (12).

While important for nearly all cells, proper regulation of RNA is especially important during early development and in the nervous system. During early embryogenesis, transcription is typically silent and gene expression is primarily controlled at the post-transcriptional level (13-15). Maternal mRNAs are translated and degraded through highly-coordinated regulatory events, controlled by both proteins and non-coding RNAs, which dictate many of the earliest stages of embryogenesis. Similar regulatory mechanisms are present in the nervous system. Repressed mRNAs that encode particular proteins are localized to distal regions of neurons, and they are activated in response to stimuli and translated locally, which impacts nervous system development and synaptic plasticity (16, 17). Furthermore, some mammalian cells lack nuclei and rely on post-transcriptional control to respond to stimuli and regulate gene expression (18).

RNAs are highly regulated by proteins. RNAs are synthesized by RNA polymerases (19, 20), and they work in conjunction with many other proteins and enzymes to create functional RNAs. Complexes of proteins recognize features in DNA and nascent RNA to initiate and terminate transcription (21, 22). Nascent mRNAs are co-transcriptionally processed to add a 5' cap, spliced to join exons and remove introns, cleaved to create 3' termini, and extended to add a polyadenosine (poly(A)) tail (**Fig. 1**) (23, 24). Similarly, tRNAs, rRNAs, and a variety of small, non-coding RNAs (such as microRNAs) are also processed into mature RNAs. For example, microRNAs (miRNAs) are typically synthesized as multi-kilobase primary RNA transcripts that must be sequentially and precisely cleaved by two separate protein complexes into ~22 nucleotide long mature miRNAs (25).

While many different types of RNAs play important roles in cells and organisms, the focus of my thesis research has largely been on the control of mRNAs by proteins. In the next few sections, I will cover key concepts of how proteins bind to mRNAs, how proteins regulate mRNAs, and how to identify which RNAs are bound by a protein.

Proteins bind to mRNA

Protein-RNA interactions are ubiquitous (26, 27). Single proteins often bind to hundreds of distinct RNAs in the cell, and a single RNA molecule often binds to many proteins at once. In humans, there are estimated to be about 1,500 proteins that bind RNA, which we refer to as RNA-binding proteins (RBPs) (27). With around 20,000 protein-coding genes that may be alternatively spliced to create unique mRNA isoforms (28), there are likely hundreds of thousands of unique mRNAs present in an organism. Given

that the particular combination of proteins bound to an mRNA determines its fate, cells and organisms must rely on principles to guide the correct combination proteins onto the correct RNAs at precisely the right time.

RNA-binding proteins often utilize conserved domains to bind mRNA. The first conserved RNA-binding domain was discovered in the 1980's as a protein component of heterogeneous nuclear ribonucleoprotein particles (hnRNPs), which are required for the processing of nascent RNAs in the nucleus (29). Originally termed the RNP domain, today it is more commonly known as the RNA recognition motif (RRM) and is among the most common protein domains (30). Considerable progress has been made in the last several decades to identify and characterize other RNA-binding domains, which include: K-homology (KH), zinc finger (ZF), pentatricopeptide repeat (PPR), Pumilio/FBF (PUF), cold-shock (CSD), RGG box, double-stranded RNA-binding (dsRBD), and Piwi/Argonaute/Zwille (PAZ) domains (31-33). Other low complexity amino acid segments rich in basic amino acids can also bind RNA (32). A survey of the more than 1,500 known RNA-binding proteins in humans nicely illustrated that they most often possess multiple RNA-binding domains in tandem (27). Intriguingly, however, some RNA-binding proteins lack conserved and/or known RNA-binding domains, suggesting much remains to be uncovered.

Proteins bind to mRNA by recognizing its phosphate backbone, particular RNA sequences and/or secondary structures, or a combination thereof (33). In the simplest example of molecular complementarity, positively charged amino acids make contact with the negatively charged phosphate backbone of RNA. Amino acids also make base-specific contacts with mRNA and is exemplified by proteins with PUF domains (34). Three

amino acids in each PUF domain make base-specific contacts with RNA nucleotides. When combined with the tendency to have multiple RNA-binding domains in tandem, proteins can possess extensive sequence specificity. Proteins may also recognize specific RNA secondary structures. For example, dsRBDs found in *Staufen* recognize double-stranded RNA (35). Individual proteins can also use multiple modes of RNA binding in parallel to achieve the required specificity for their biological function (27, 33).

Cytoplasmic control of mRNA by proteins

RNA-binding proteins symphonically collaborate to control nearly all aspects of mRNA (24, 36). Many regulatory events occur in the nucleus, but many critical regulatory events also occur in the cytoplasm. Proteins bind to mRNA in the cytoplasm to control its subcellular localization, stability, and translation. The cytoplasmic control of mRNA will be the focus of the next several paragraphs.

mRNAs possess two untranslated regions, at their 5' and 3' ends, that serve as regulatory “hotspots”. The 5' untranslated region (UTR) has an integral role in the regulation of translation (37). For example, the 5' cap structure is recognized and bound by eukaryotic initiation factor 4E (eIF4E), and the RNA helicase eIF4A subsequently helps clear RNA secondary structures near the cap to facilitate recruitment of the small subunit of the ribosome (38). In parallel, the 3' UTR is a multi-functional regulatory hotspot. Many RBPs bind to specific sequences and structures in 3' UTRs to control the stability and/or translation of the RNA (**Fig. 2**) (13, 39-41). RBPs specifically bind to sequences in 3' UTRs of mRNAs to promote removal of the poly(A) tail, which results in translational repression and destabilization of the mRNA (42, 43). RBPs may also bind to 3' UTRs or

promote lengthening of poly(A) tails to increase translation, which is especially critical during the earliest stages of embryogenesis (42, 43).

RNA-binding proteins regulate RNA through intrinsic enzymatic activities or through the recruitment of protein partners (33). Some proteins possess both RNA-binding and catalytic domains, such as those found in Argonaute proteins (44, 45). However, many proteins only possess RNA-binding domains and must recruit other proteins to elicit a regulatory response (33). For example, RBPs can recruit a suite of proteins called the CCR4-NOT complex to mRNAs to shorten or remove the poly(A) tail, which decreases mRNA translation and stability (46). Separation of RNA-binding and regulatory activities into distinct units allows greater regulatory potential, as single RBPs may elicit different regulatory effects dependent on the cellular context. For example, nutrient conditions alter post-translational modifications on particular yeast RNA-binding proteins and dictates the regulatory response (47-49), presumably through recruitment of different protein partners in each condition.

Methods to identify protein-RNA interactions

A fundamental tenet in the analysis of RBPs is to identify the RNAs they bind *in vivo*. Such information often yields tremendous insight into the biological role of individual proteins, especially since RBPs often bind functionally-related mRNAs (50, 51). In the next several paragraphs, I will review two prevailing approaches to identify protein-RNA interactions *in vivo* – “RNA immunoprecipitation (RIP)” and “UV-crosslinking immunoprecipitation (CLIP)”. Both are based off the same core principle: co-purify RBP-RNA complexes from cell, tissue, or whole animal extracts. While they have proven

incredibly powerful in the dissection of protein-RNA networks, RIP and CLIP possess key weaknesses that warrant attention.

RIP is a powerful tool to identify protein-RNA interactions, but it has several limitations. RIP traditionally utilizes native immunopurifications to isolate RBP-RNA complexes from a cell lysate (**Fig. 3A**). Associated RNAs are subsequently identified by microarray or deep sequencing analysis (52, 53). Importantly, RIP followed by microarray analysis (RIP-chip) studies identified that RNAs co-purified by proteins were often functionally related and involved in similar biological processes (51, 52, 54). Despite its clear utility, RIP has limitations (**Fig. 4**) (55). Native purifications limit the stringency of wash steps designed to remove non-specific, weak interactions with the protein, antibody, or beads. Thus, RIP is susceptible to high background, which can be counteracted by performing many time-consuming and costly biological replicates. RIP may also detect non-physiological interactions that occur solely in cell lysate *in vitro* (56-58). Furthermore, the binding site of a protein on the RNA it binds must be inferred via bioinformatic approaches and confirmed independently.

Recently, several UV-crosslinking immunoprecipitations followed by high-throughput sequencing (CLIP-seq) approaches were developed to simultaneously identify transcriptome-wide targets and RNA-binding sites of individual proteins (**Fig. 3B**) (59-64). In CLIP-seq, proteins are UV-crosslinked to associated RNAs in intact cells. UV-crosslinking has several advantages: 1) proteins are covalently bonded to the associated RNA, ensuring only direct protein-RNA interactions are captured; 2) covalent linkage enables stringent wash and partial RNase foot-printing steps, which greatly reduce background interaction and reveal protein binding sites, respectively; and 3) proteins are

crosslinked to the RNA in intact cells and tissues, thus ensuring only *in vivo* interactions are detected. The crosslinked protein-RNA complexes are subsequently immunopurified, stringently washed, subjected to a partial RNase digestion, and removed from the protein. Recovered RNA fragments are then ligated to DNA adapters, reverse transcribed, PCR amplified, and finally high-throughput sequenced. While complicated, CLIP-seq approaches have yielded great insight into how proteins regulate mRNAs, including how they are alternatively spliced and polyadenylated in the brain (59, 60). CLIP-seq approaches have become the gold standard for dissecting protein-RNA interactions *in vivo*.

Despite their power, CLIP-seq approaches are marred by several technical limitations (**Fig. 4**). Live cells must be irradiated with UV light, requiring relevant tissues or cells to be extracted from live animals and thus removed from their native context. UV crosslinking is also inefficient, as only 1-5% of protein-RNA complexes become crosslinked (65). Large amounts of starting material are therefore required or, as in photoactivatable ribonucleoside enhanced CLIP (PAR-CLIP), nucleotide analogs with enhanced crosslinking efficiency must be introduced into cells *ex vivo* (61). Protein-RNA complexes are then immunopurified from cell lysates, which requires an antibody for the RNA-binding protein or an epitope tagged version.

In parallel, both CLIP-seq and RIP-seq approaches face a key challenge in the interpretation of their large data sets. Global approaches often identify hundreds to thousands of RNAs that interact with the protein of interest. Out of the sea of interactors, a key challenge is to identify biologically relevant interactions (*i.e.* those that lead to detectable regulation). Regulatory interactions are often obfuscated by transient

interactions with unclear relevance, which are permanently captured by UV crosslinking. This problem quickly compounds, since a common focus is now on RNA regulatory networks formed by multiple proteins, particularly in specific tissues and cell types. Parsing the networks is thus becoming increasingly complex and requires new strategies to highlight biologically relevant interactions.

Given the above limitations, **a major goal of my thesis research was to develop a new method to identify protein-RNA interactions *in vivo***. We desired an approach that would unambiguously detect protein-RNA interactions – independent of protein purification, UV-crosslinking, or radioactive-labeling steps. We hypothesized that the fusion of an RNA-modifying enzyme to an RNA-binding protein of interest would covalently mark, or “tag”, the RNAs bound by that protein (**Fig. 5**). The “tagged” RNAs would subsequently be identified from the pool of total RNA *in vitro*, ideally using high-throughput sequencing to enable transcriptome-wide experiments. Our envisioned approach, which we termed “RNA Tagging”, met all of our desired criteria. We also hypothesized that an RNA Tagging approach would highlight RNAs that are regulated by the RBP from those that are bound transiently. The fused RNA-modifying enzyme would likely mimic regulation mediated by the RBP of interest. A higher frequency of modification on an RNA may reflect a higher degree of regulation by the RBP, and vice versa. However, previous attempts in the lab to develop RNA Tagging were unsuccessful, as suitable and effective “tagging enzymes” remained elusive. Thus, we sought and pursued a new class of candidate enzymes.

Ribonucleotidyl transferases

Enzymes called RNA polymerases catalyze the synthesis of RNA. RNA polymerases play essential roles across all kingdoms of life (20), as they synthesize every RNA molecule produced by a cell. RNA polymerases typically synthesize RNA from a template strand of RNA or DNA, using base complementarity to correctly guide incorporation of new nucleotides into nascent RNA (19). For example, RNA polymerase II synthesizes mRNAs by copying a DNA strand from the relevant gene (19, 20).

A subclass of RNA polymerases adds ribonucleotides to the 3' end of RNA molecules independent of a nucleic acid template. The canonical member of this family of ribonucleotidyl transferases (rNTRs) is poly(A) polymerase, which adds poly(A) tails to the end of all mRNAs in the nucleus (66, 67). Other “non-canonical” rNTRs function in the cytoplasm and add a variety of nucleotides, including adenosines and uridines (68, 69). The nucleotide specificity of non-canonical rNTRs must be determined experimentally, especially since a single enzyme can switch specificity from adenosine to uridine (70). Non-canonical rNTRs have emerging roles in several biological processes, including early development and the nervous system (71). For example, the non-canonical rNTR GLD-2 is a cytoplasmic poly(A) polymerase that activates translation of stored mRNAs, particularly during early development (72-74).

Non-canonical rNTRs that add uridines to the 3' end of RNA are called terminal uridylyl transferase (TUTs) or, interchangeably, poly(U) polymerases (PUPs) (75-77). TUTs/PUPs are conserved throughout Eukarya and add uridines to a wide-array of RNA substrates, including mRNAs, miRNAs, and snRNAs (68). All TUTs/PUPs possess two core protein domains: the nucleotidyl transferase and poly(A) polymerase-associated domains, which combine to form the active site of the enzymes (78-80). Some

TUTs/PUPs also possess “accessory domains”, such as zinc fingers, RNA recognition motifs, and other potential RNA-binding domains (66, 68).

TUTs/PUPs are broadly separated into two groups: those with and those without accessory domains. TUTs/PUPs with accessory domains are epitomized by TUT4 and TUT7 proteins. These enzymes are conserved throughout metazoans (68) and have been implicated in the regulation of early development and cancer, primarily through regulation of let-7 miRNAs (81-86). They possess the core nucleotidyl transferase and poly(A) polymerase-associated domains, housed in their C-terminal halves (**Fig. 6A**). They also contain a degenerate copy of both domains in their N-terminal halves, which have several key amino acid substitutions that are predicted to eliminate catalytic activity. TUT7 and TUT4 enzymes also possess four conserved zinc finger domains and a conserved yet uncharacterized region of basic amino acids. The role of accessory domains and how they affect the function of individual TUTs/PUPs is largely unstudied. TUTs/PUPs without accessory domains, such as *Caenorhabditis elegans* PUP-2, possess only the core catalytic domains (**Fig. 6B**) (75). The enzymes vary in size, and the amino acids outside the core domains have unknown functions. TUTs/PUPs without accessory domains are also conserved, although they appear absent in humans and other mammals (68).

TUTs/PUPs are ideal candidate tagging enzymes for RNA Tagging (Fig. 7). TUTs/PUPs add 3' terminal uridines – a detectable, covalent modification – and single enzymes act on a variety of RNA substrates. Furthermore, some systems, such as *S. cerevisiae*, lack their own TUTs/PUPs and the nuclease that removes the residues. **However, there is very little known about the biochemical properties of TUTs/PUPs and how they add uridines to RNA.** Catalytic residues have been identified, but the role

of accessory domains is completely unexamined. For example, are the degenerate nucleotidyl transferase and poly(A)-polymerase domains required for U-addition? How do the zinc fingers affect U-addition? Do any domains bind RNA? What's the role of the conserved region of basic amino acids? Answers to these questions will be key to the development of RNA Tagging.

PUF proteins

Relatively few protein-RNA networks have been mapped despite their ubiquity and importance. To uncover principles that broadly govern protein-RNA networks, I have studied a highly conserved family of RBPs called PUF proteins (87). I primarily focused on three broad questions:

- 1) Which RNAs are bound by particular PUF proteins?**
- 2) Which PUF-RNA interactions lead to regulation?**
- 3) How are networks comprised of multiple PUF proteins coordinated?**

PUF proteins are important mRNA regulators. PUF proteins and the PUF RNA-binding domain were named for the two founding members of the family: *D. melanogaster* Pumilio and *C. elegans* EBF (88, 89). PUF proteins are conserved across Eukarya, and have been studied in fungi, plants, trypanosomes, fruit flies, nematodes, planarians, and mammals (87, 90, 91). They have critical roles in stem cells, early development, gametogenesis, embryogenesis, and the nervous system, particularly for synaptic plasticity (92).

In the next several paragraphs, I will describe some of what has been learned about PUF proteins, with a particular focus on their biological functions and how they selectively bind particular RNAs.

Conserved regulation

One extensively characterized PUF protein is Pumilio from *D. melanogaster*. *Drosophila* with mutations in *pumilio* have defects in anterior/posterior patterning, germline development, neuron excitability, presynaptic growth, and dendrite morphogenesis (90). Several phenotypes are at least partially explained by the misregulation of particular mRNAs. For example, *hunchback* mRNA is normally expressed throughout the anterior/posterior axis of drosophila embryos, but Hunchback protein is only expressed in the anterior region of the embryo, thereby creating a gradient of Hunchback protein along the anterior/posterior axis (93). *Pumilio* helps create the gradient of Hunchback protein by binding to specific sequences in the 3' UTR of *hunchback* mRNA and repressing its translation in the posterior of embryos (94-96).

C. elegans have twelve known PUF proteins implicated in a diverse array of biological functions (87). The first PUF proteins discovered in *C. elegans* are paralogs called fem-3 binding factor 1 and 2 (FBF-1 and FBF-2, respectively) and are collectively referred to as FBF. In part, FBF helps control the transition from spermatogenesis to oogenesis in the hermaphroditic germline by binding to a specific sequence in the 3' UTR of *fem-3* mRNA (89, 97). FBF is also required for germline stem cell maintenance (98) and synaptic plasticity in response to odors (99). Similarly, several other *C. elegans* PUF

proteins help control the nematode germline through regulation of meiotic progression, gametogenesis, and embryogenesis (100).

The regulation of stem cells and gametogenesis by PUF proteins is conserved throughout Eukarya. Like FBF, mammalian PUF proteins PUM1 and PUM2 have roles in stem cells and the nervous system (101, 102). Furthermore, a PUF protein in planarians has also been linked to the maintenance of stem cells (91), and two *S. cerevisiae* PUF proteins are required for proper regulation of mating-type switching (103).

Conserved structure and RNA-binding specificity

Canonical PUF proteins possess eight tandem PUF RNA-binding domains, referred to as “PUF repeats” (34, 87). The eight PUF repeats are typically found in the C-terminal halves of the proteins (**Fig. 8A**) (34). Crystal structures of PUF proteins revealed that PUF repeats are primarily composed of three alpha-helices that form a compact bundle, separated by variable length loops (**Fig. 8B**) (104-110). Together, the eight PUF repeats form a distinctive, half-moon-like crescent, which binds to RNA in an anti-parallel manner on its inner, concave surface (**Fig. 9A**). In the simplest examples, each PUF repeat binds to a single base (104, 105, 107). Most PUF proteins use a conserved mechanism, described in more detail below, to bind to specific RNA sequences, called PUF-binding elements (PBEs). Classic PBEs are eight nucleotides in length and conform to the consensus 5'-UGUAHAUA-3', with the UGUA tetranucleotide motif forming its core (34, 87).

Each PUF repeat possesses three amino acids, called the tripartite recognition motif (TRM), that most often have specificity for a particular RNA nucleotide (111, 112).

Tripartite recognition motifs extend outward from the RNA-binding surface of PUF proteins (**Fig. 9B**) (104-109). PUF proteins and in particular their TRMs use a series of hydrogen bonds and van der Waals interactions to bind specific bases, thereby providing PUF proteins with the specificity for PBEs. TRMs in repeats 5-8 and 1-2 appear to have the highest degree of specificity, while those in repeats 3-4 exhibit the greatest variability in binding preferences (111, 112). Intriguingly, tripartite recognition motifs are highly modular and can be synthetically designed to bind particular sequences (111, 113).

Yeast PUF proteins

The yeast *S. cerevisiae* has three canonical PUF proteins: Puf3p, Puf4p, and Puf5p (87). Each protein has the characteristic eight PUF repeats in their C-terminal half, with a divergent N-terminal half that possesses low-complexity regions (**Fig. 10**). Much has been uncovered about how these PUF proteins selectively bind specific RNA sequences, which RNAs they bind *in vivo*, and how they regulate bound RNAs.

The three canonical yeast PUF proteins utilize a similar and highly conserved mechanism to bind specific RNA bases. Much like PUFs from other organisms, yeast PUF proteins make base specific contacts with the RNA substrate through the tripartite-recognition motifs present in each PUF repeat and preferentially bind to a 5' UGUA followed by a downstream UA (107, 109, 110, 112). Puf3p binds in perhaps the simplest way – each PUF repeat contacts a single RNA base (**Fig. 11A-B**) (107). Puf3p also has an additional level specificity for sequences with a cytosine one or two nucleotides upstream of the UGUA, which is housed in a pocket directly adjacent to the PUF repeats (**Fig. 11C**) (107).

Puf4p and Puf5p bind 9 and 10 nucleotide long binding elements. Crystal structures of Puf4p bound to an RNA substrate revealed that Puf4p, much like FBF, binds a nine nucleotide long binding element (**Fig. 12A-B**) (109, 110). The 9 bases are accommodated by the 8 PUF repeats through the rotation of a base away from the PUF RNA-binding surface (**Fig. 12C**) (109). Non-structural studies with Puf5p revealed that it also binds a 9 nucleotide long sequence with high affinity, in addition to its canonical 10 nucleotide long element (109). Thus, it was hypothesized that both proteins use a “two-handed” binding mechanism, where the critical interactions are between the proteins and the 5' UGUA and the 3' UA (109).

Yeast canonical PUF proteins bind to hundreds of mRNAs. In a landmark study, RIP-chip was used to demonstrate that Puf3p, Puf4p, and Puf5p bound to about two hundred mRNAs each, which were largely distinct (54). The mRNAs were highly enriched for their expected PUF-binding elements; binding elements with 8, 9, and 10 nucleotides were present in Puf3p, Puf4p, and Puf5p bound mRNAs, respectively. The mRNAs were also differentially enriched for biological functions. Puf3p-bound mRNAs were highly enriched for nuclear-encoded mRNAs important for mitochondrial function, Puf4p-bound mRNAs were enriched for ribosome biogenesis related functions, and Puf5p-bound mRNAs enriched for the regulation of chromatin.

In fermentation growth conditions, canonical PUF proteins are often negative regulators of mRNA and translation (92, 114). Puf3p has been studied extensively, particularly in relation to its regulation of a single mRNA, *COX17*. Puf3p binds to *COX17* mRNA through two 8-nucleotide long binding elements in its 3' UTR (54, 115, 116). Once bound, Puf3p recruits the CCR4-NOT and PAN2-3 deadenylase complexes to shorten

the poly(A) tail of *COX17* mRNA and thus destabilize it (47, 115, 117, 118). In parallel, Puf3p is required for the localization of many nuclear-encoded mRNAs to the outer surface of mitochondria (119). Similar to Puf3p, Puf4p and Puf5p interact with CCR4-NOT to destabilize bound mRNAs (49, 103, 120, 121). In parallel, Puf5p also represses mRNA translation (122, 123). Puf5p-mediated repression of *HO* mRNA requires EAP1, which binds to eukaryotic translation initiation factor 4E, thus linking 3' UTR binding events to 5' UTR-linked regulation (123).

Many questions remain about yeast PUF proteins. For example, how does Puf3p regulate mitochondrial function at the molecular level? Puf3p binds to many mRNAs for nuclear-encoded mitochondrial proteins and a few studies suggest that Puf3p represses mitochondrial biogenesis (48, 54, 124-126). Yet, particular mitochondrial pathways regulated by Puf3p remain unknown. Furthermore, is the primary role of Puf3p in the regulation of mitochondria, or is it a pleiotropic regulator, as suggested by a recent study that found Puf3p binds a large cohort of mRNAs (> 1,000) with little effect on their regulation (127)? In parallel, how does Puf5p, which only has 8 PUF repeats, bind to variable length sequence elements with high affinity *in vitro*, and does it do so *in vivo*? Do Puf4p and Puf5p share many mRNA targets, since they bind to the same RNA sequence *in vitro*, or are their mRNA targets really distinct as suggested in prior studies? If they do, how are targets chosen by each protein and how is the broader PUF regulatory network balanced?

Brief outline

Throughout the rest of my thesis, I will present experiments, data, and interpretations that represent the vast majority of my graduate research. I will conclude by sharing my thoughts on a few future directions for the specific projects I led, where I believe the field is heading, and major challenges the field faces in the near future. Overall, my primary research goals were to:

1. Analyze how TUTs/PUPs add uridines to RNA (Chapter 2);
2. Develop RNA Tagging, a simplified and unambiguous approach to identify protein-RNA interactions transcriptome-wide (Chapter 3);
3. Develop strategies to highlight protein-RNA interactions that lead to regulation in large data sets (Chapter 4);
4. Analyze how an RNA regulatory network composed of multiple proteins is coordinated and balanced (Chapter 5).

Figure 1.

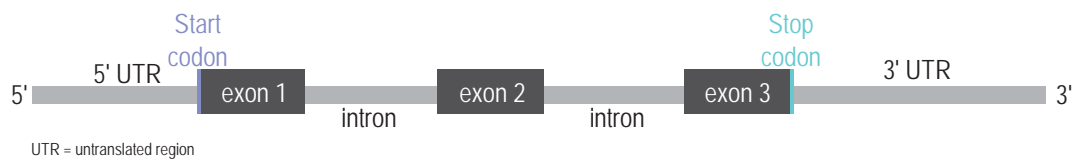
A**B**

Figure 1. Anatomy of an mRNA. A) Schematic of a nascent mRNA. UTR, untranslated region. B) Schematic of a fully processed, mature mRNA.

Figure 2.

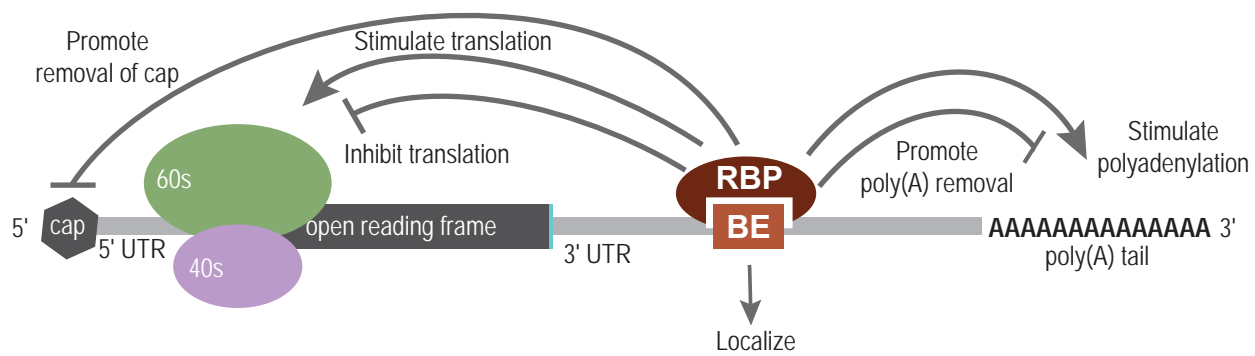


Figure 2. Examples of cytoplasmic control of mRNAs by RNA-binding proteins. RNA-binding proteins (RBP) often bind to particular sequences, termed binding elements (BE), located in the 3' UTR of mRNAs. Once bound, RBPs can control the translation, stability, and localization of the mRNA, often through the recruitment of other regulatory proteins to the mRNA. Thus, a single RBP can elicit different regulatory outcomes simply by recruiting different regulatory proteins.

Figure 3.

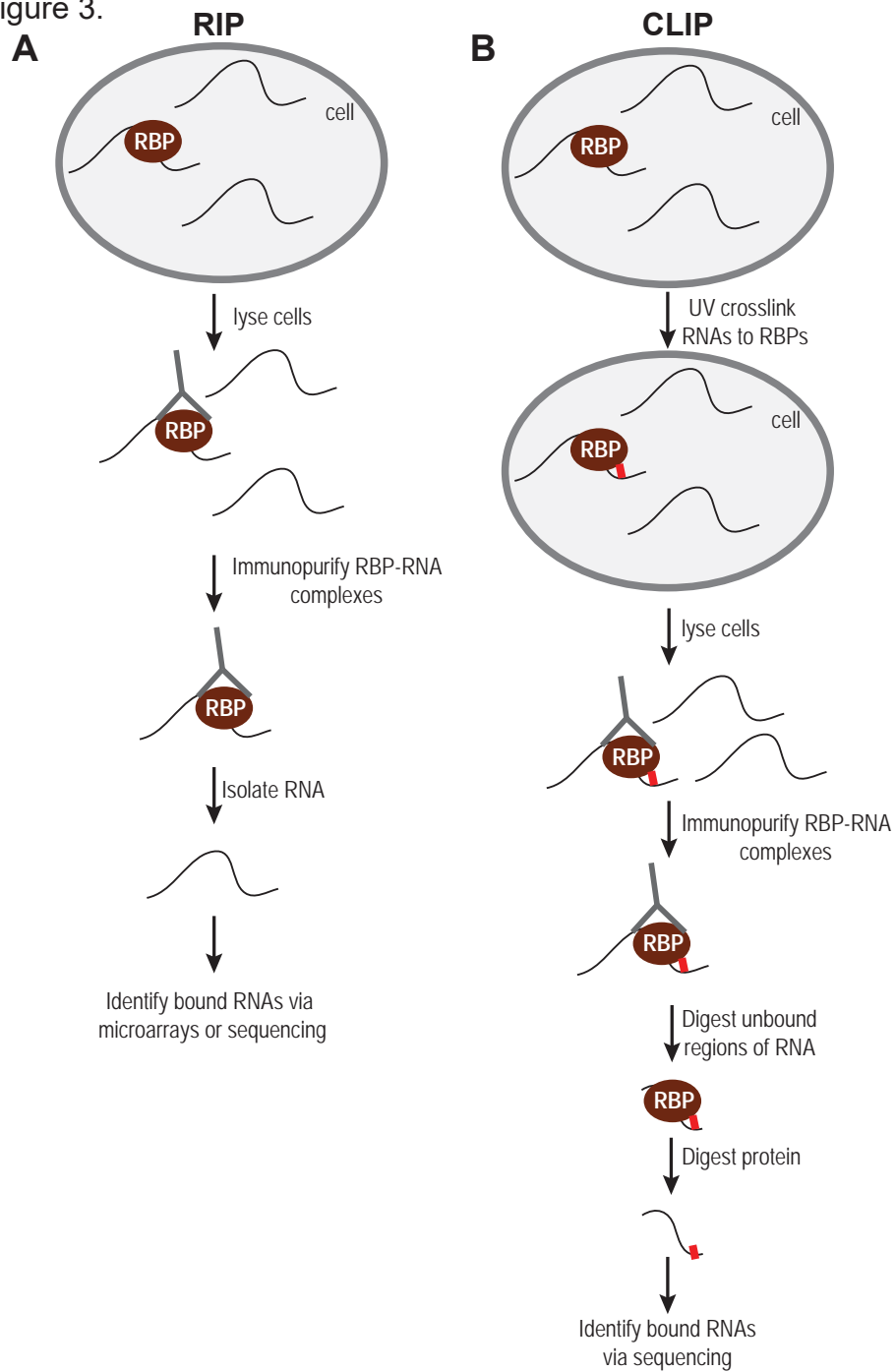


Figure 3. Schematic outlines of techniques to identify protein-RNA interactions. A) RNA immunoprecipitations (RIPs). Cells of interest are lysed and the RBP of interest is immunopurified from cell lysate *in vitro*. RNAs that co-purify with the RBP are then isolated and subsequently identified by RT-PCR, microarray, or high-throughput sequencing approaches. RBP, RNA-binding protein. B). UV-crosslinking immunoprecipitations (CLIP). Live cells are irradiated with UV light to crosslink RBP-RNA complexes. After cell lysis, RBP-RNA complexes are immunopurified from cell lysate *in vitro*, and the bound RNA is partially digested by RNase treatment to remove regions not protected by the RBP. The RBP is then digested via Proteinase K treatment and the bound RNA fragments are identified via high-throughput sequencing. RBP, RNA-binding protein.

Figure 4.

RIP-seq	CLIP-seq	
-	--	Simple
✓	--	Fast
✓	-	Independent of crosslinking
-	-	Independent of immunopurification
-	✓	Resistant to in vitro artifacts
-	--	Low input material
✓	✓ ++	Binding site determination
-	-	Distinguish biologically meaningful from stochastic and/or transient interactions
-	-	Amenable to tissue-specific studies in vivo

Figure 4. Chart of strengths and weaknesses of RIP-seq and CLIP-seq approaches.

Figure 5.

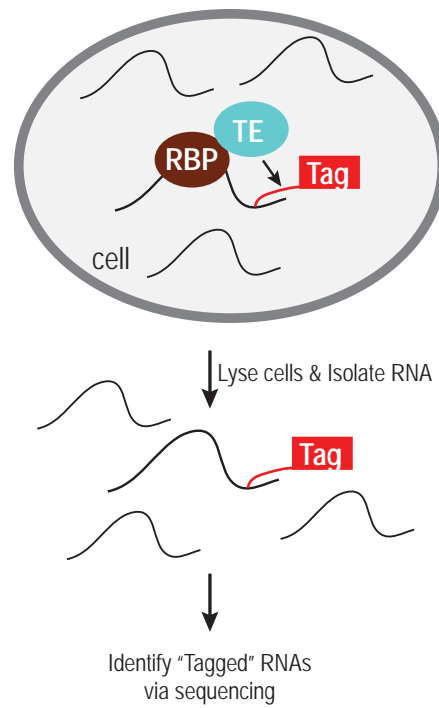


Figure 5. Schematic outline of the proposed “RNA Tagging” approach. Live cells will express a chimeric protein composed of the RBP of interest fused to a tagging enzyme (TE). The RBP-TE fusion protein will thus covalently “tag” the bound RNAs *in vivo*, and the tagged RNAs will be identified *in vitro* via high-throughput sequencing.

Figure 6.

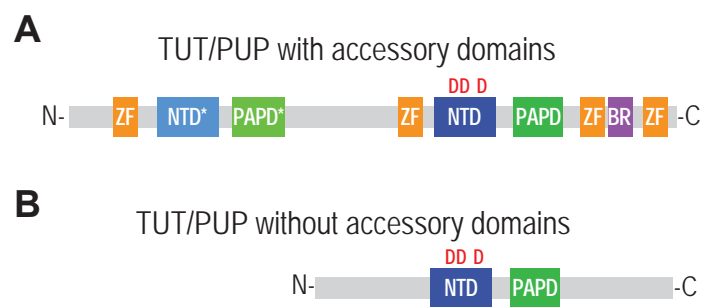


Figure 6. Schematic of the two TUT/PUP classes. A) Schematic of an example TUT/PUP with accessory domains. ZF, zinc finger domains; NTD*, cryptic nucleotidyl transferase domain; PAPD*, cryptic poly(A) polymerase-associated domain; NTD, active nucleotidyl transferase domain; PAPD, poly(A) polymerase-associated domain; BR, basic region. B) Schematic of an example TUT/PUP that lacks accessory domains. NTD, active nucleotidyl transferase domain; PAPD, poly(A) polymerase-associated domain.

Figure 7.

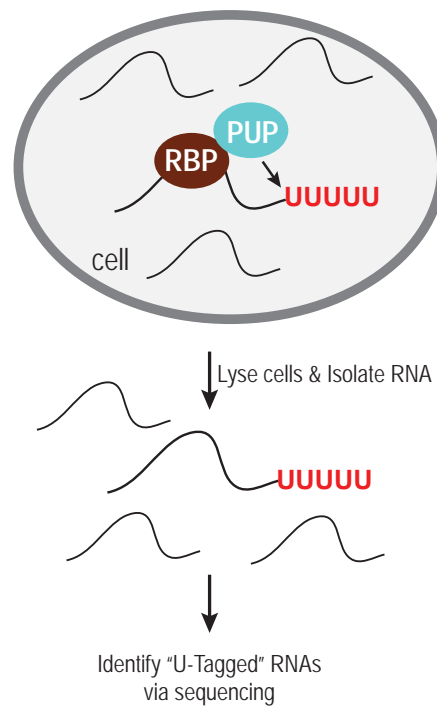

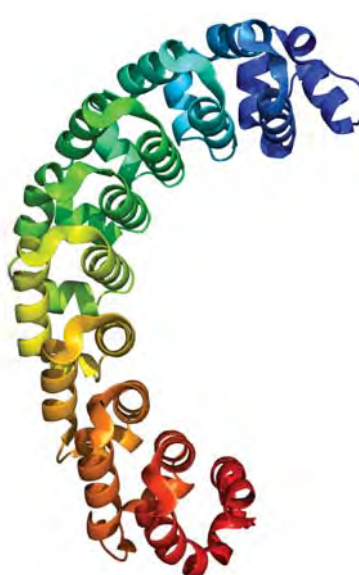


Figure 7. Schematic of the proposed RNA Tagging approach with PUPs. Live cells will express an RBP of interest fused to a PUP. RNAs bound by the RBP-PUP fusion protein will be covalently tagged with 3' uridines, termed the “U-tag”. Following cell lysis, U-tagged RNAs will be identified *in vitro* via high-throughput sequencing.

Figure 8.

A*D. melanogaster* Pumilio1  935  PUF domain*H. sapiens* Pum11  1186*C. elegans* FBF-21  632**B***D. melanogaster* Pumilio

PDB ID: 3H3D

H. sapiens Pum1

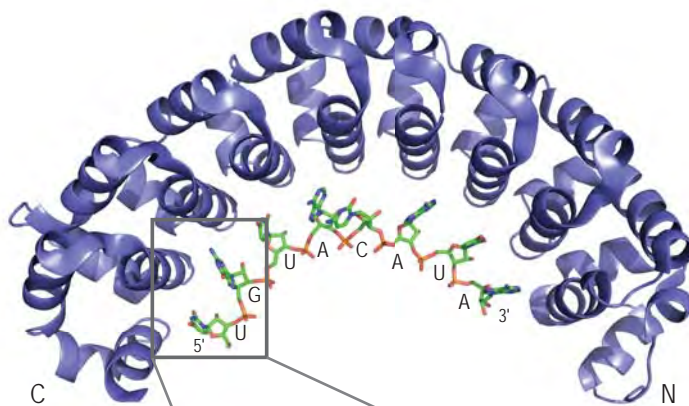
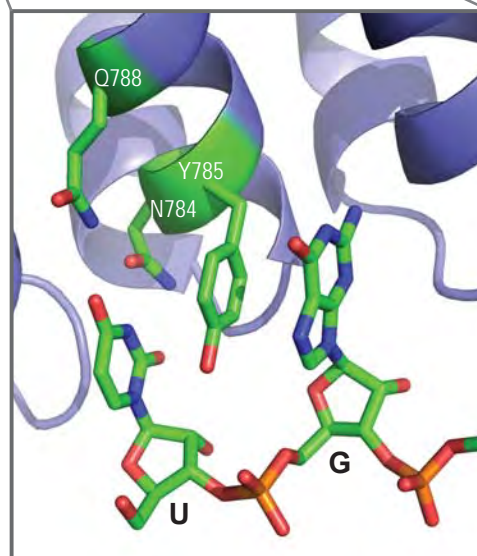
PDB ID: 1M8Z

C. elegans FBF-2

PDB ID: 3K64

Figure 8. Canonical PUF proteins. A) Schematic of canonical PUF proteins from *D. melanogaster*, *H. sapiens*, and *C. elegans*. Each protein has the stereotypical eight tandem PUF RNA-binding domains in their C-terminal halves. The N-terminal halves are divergent and typically possess low-complexity regions. B) Crystal structure of the RNA-binding domains of the three PUF proteins from panel A. Each PUF domain is comprised of three α -helices, and the PUF domains stack on each other to form a crescent shape. The N-termini are in blue and the C-termini are in red.

Figure 9.

A**B**

PDB ID: 5KLA

Figure 9. PUF protein bound to RNA. A) Crystal structure of the RNA-binding domain from *D. melanogaster* Pumilio in complex with RNA. The N- and C-termini of the protein are indicated. B) Zoomed-in image of a tripartite recognition motif binding to an RNA base. Two amino acids (N784, Q788) hydrogen bond with the base, termed “edge-on” interactions, while a third, typically hydrophobic or aromatic, amino acid (Y785) “stacks” on the RNA base, via pi-stacking interactions.

Figure 10.

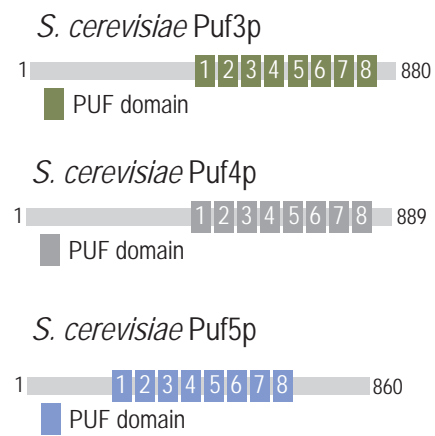


Figure 10. Schematics of *S. cerevisiae* Puf3p, Puf4p, and Puf5p.

Figure 11.

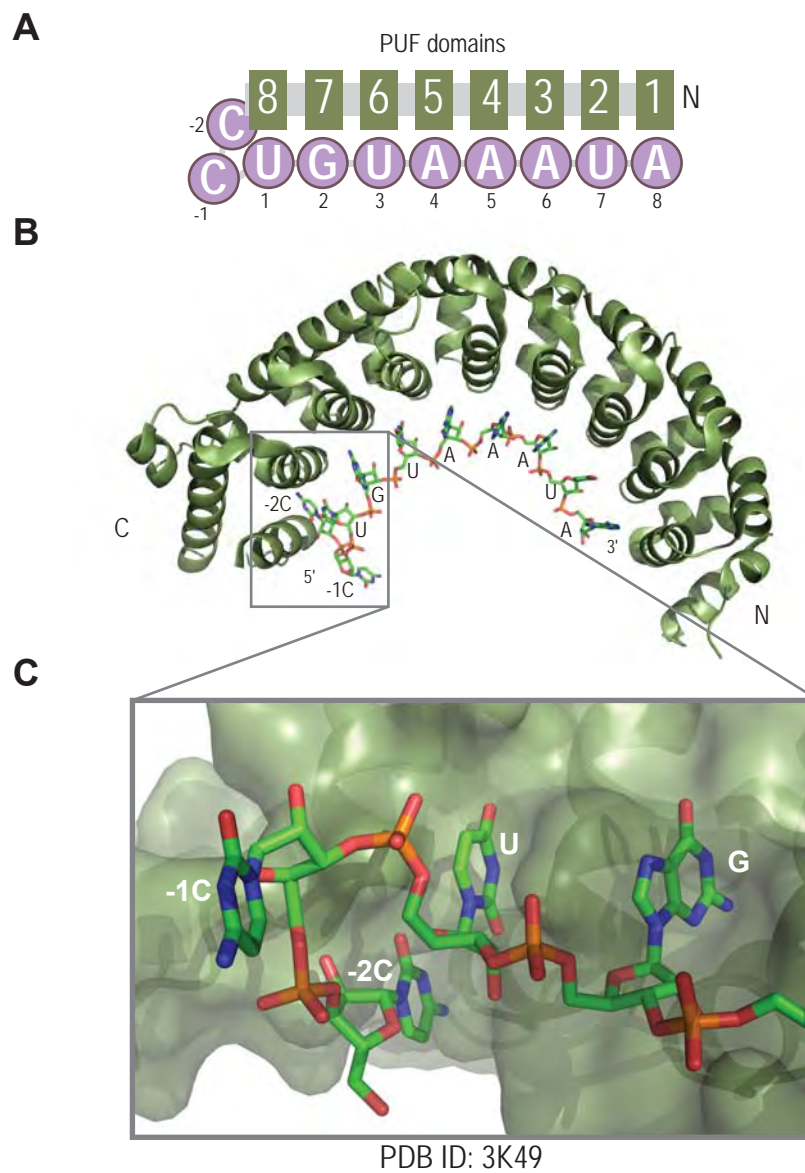
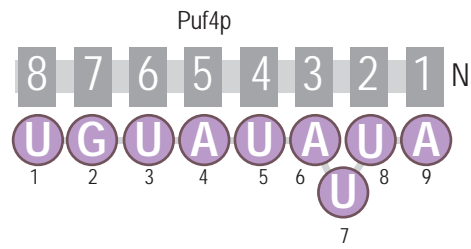
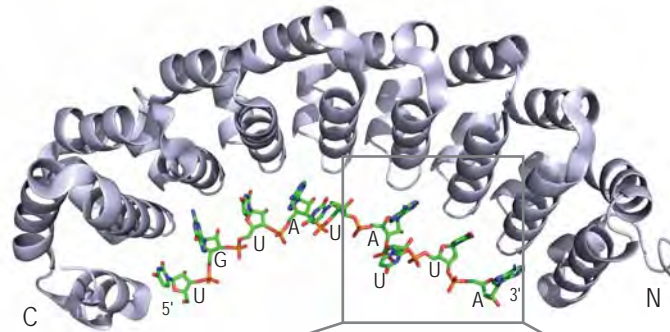
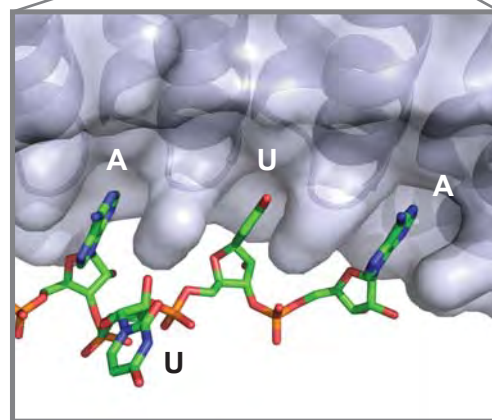


Figure 11. Structure of *S. cerevisiae* Puf3p bound to RNA. A) Cartoon that depicts Puf3p bound to RNA. Each PUF repeat in Puf3p contacts one RNA base. In parallel, Puf3p has a high-affinity pocket for cytosine residues (the “-2C pocket”) upstream of the first U base in its binding sequence. B) Crystal structure of Puf3p bound to its cognate RNA binding sequence. Termini of the protein and RNA, and the identity of the RNA bases are indicated. C) Zoomed-in image of the -2C pocket in Puf3p.

Figure 12.

A**B****C**

PDB ID: 3BX2

Figure 12. Structure of *S. cerevisiae* Puf4p bound to RNA. A) Cartoon that depicts Puf4p bound to RNA. Puf4p accommodates a 9 nucleotide long binding element by excluding an RNA base (the “flipped-out” base) from the surface of the protein. B) Crystal structure of Puf4p bound to a cognate RNA binding sequence. Termini of the protein and RNA, and the identity of the RNA bases are indicated. C) Zoomed-in image of the flipped-out base.

REFERENCES

1. F. Crick, Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
2. D. L. Black, Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336 (2003).
3. M. Esteller, Non-coding RNAs in human disease. *Nature Reviews Genetics* **12**, 861-874 (2011).
4. A. Fatica, I. Bozzoni, Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics* **15**, 7-21 (2014).
5. E. Khurana *et al.*, Role of non-coding sequence variants in cancer. *Nature Reviews Genetics* **17**, 93-108 (2016).
6. S. R. Eddy, Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* **2**, 919-929 (2001).
7. J. S. Mattick, I. V. Makunin, Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**, R17-29 (2006).
8. H. F. Noller, Ribosomal RNA and translation. *Annu Rev Biochem* **60**, 191-227 (1991).
9. C. L. Will, R. Luhrmann, Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**, (2011).
10. K. Collins, The biogenesis and regulation of telomerase holoenzymes. *Nat Rev Mol Cell Biol* **7**, 484-494 (2006).
11. T. A. Steitz, A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* **9**, 242-253 (2008).
12. K. V. Morris, J. S. Mattick, The rise of regulatory RNA. *Nature Reviews Genetics* **15**, 423-437 (2014).
13. S. Kuersten, E. B. Goodwin, The power of the 3' UTR: translational control and development. *Nat Rev Genet* **4**, 626-637 (2003).
14. G. Stefani, F. J. Slack, Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* **9**, 219-230 (2008).
15. A. Pauli, J. L. Rinn, A. F. Schier, Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* **12**, 136-149 (2011).
16. H. Jung, C. E. Holt, Local translation of mRNAs in neural development. *Wiley Interdiscip Rev RNA* **2**, 153-165 (2011).
17. H. Jung, B. C. Yoon, C. E. Holt, Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat Rev Neurosci* **13**, 308-324 (2012).
18. G. A. Zimmerman, A. S. Weyrich, Signal-dependent protein synthesis by activated platelets: new pathways to altered phenotype and function. *Arterioscler Thromb Vasc Biol* **28**, s17-24 (2008).
19. P. Cramer *et al.*, Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37**, 337-352 (2008).
20. F. Werner, D. Grohmann, Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**, 85-98 (2011).
21. S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-143 (2015).

22. O. Porrua, D. Libri, Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**, 190-202 (2015).
23. M. J. Moore, N. J. Proudfoot, Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700 (2009).
24. G. Singh, G. Pratt, G. W. Yeo, M. J. Moore, The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* **84**, 325-354 (2015).
25. M. Ha, V. N. Kim, Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509-524 (2014).
26. D. D. Licatalosi, R. B. Darnell, RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11**, 75-87 (2010).
27. S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845 (2014).
28. I. Ezkurdia *et al.*, Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* **23**, 5866-5878 (2014).
29. G. Dreyfuss, M. S. Swanson, S. Pinol-Roma, Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* **13**, 86-91 (1988).
30. C. Maris, C. Dominguez, F. H. Allain, The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272**, 2118-2131 (2005).
31. T. Glisovic, J. L. Bachorik, J. Yong, G. Dreyfuss, RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **582**, 1977-1986 (2008).
32. C. G. Burd, G. Dreyfuss, Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615-621 (1994).
33. B. M. Lunde, C. Moore, G. Varani, RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8**, 479-490 (2007).
34. T. M. Hall, De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr Opin Struct Biol* **36**, 116-121 (2016).
35. M. L. Gleghorn, L. E. Maquat, 'Black sheep' that don't leave the double-stranded RNA-binding domain fold. *Trends Biochem Sci* **39**, 328-340 (2014).
36. M. Muller-McNicoll, K. M. Neugebauer, How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics* **14**, 275-287 (2013).
37. A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413-1416 (2016).
38. C. E. Aitken, J. R. Lorsch, A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol* **19**, 568-576 (2012).
39. B. Mazumder, V. Seshadri, P. L. Fox, Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci* **28**, 91-98 (2003).
40. C. H. de Moor, H. Meijer, S. Lissenden, Mechanisms of translational control by the 3' UTR in development and differentiation. *Semin Cell Dev Biol* **16**, 49-58 (2005).
41. M. Wickens, P. Anderson, R. J. Jackson, Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev* **7**, 220-232 (1997).
42. D. R. Schoenberg, L. E. Maquat, Regulation of cytoplasmic mRNA decay. *Nature Reviews Genetics* **13**, 246-259 (2012).

43. J. Kong, P. Lasko, Translational control in cellular and developmental processes. *Nat Rev Genet* **13**, 383-394 (2012).
44. G. Hutvagner, M. J. Simard, Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**, 22-32 (2008).
45. G. Meister, Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics* **14**, 447-459 (2013).
46. M. A. Collart, The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **7**, 438-454 (2016).
47. M. A. Miller, J. Russo, A. D. Fischer, F. A. Lopez Leban, W. M. Olivas, Carbon source-dependent alteration of Puf3p activity mediates rapid changes in the stabilities of mRNAs involved in mitochondrial function. *Nucleic Acids Res* **42**, 3954-3970 (2014).
48. C. D. Lee, B. P. Tu, Glucose-Regulated Phosphorylation of the PUF Protein Puf3 Regulates the Translational Fate of Its Bound mRNAs and Association with RNA Granules. *Cell Rep* **11**, 1638-1650 (2015).
49. J. Russo, W. M. Olivas, Conditional regulation of Puf1p, Puf4p, and Puf5p activity alters YHB1 mRNA stability for a rapid response to toxic nitric oxide stress in yeast. *Mol Biol Cell* **26**, 1015-1029 (2015).
50. J. D. Keene, RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-543 (2007).
51. D. J. Hogan, D. P. Riordan, A. P. Gerber, D. Herschlag, P. O. Brown, Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology* **6**, e255 (2008).
52. S. A. Tenenbaum, C. C. Carson, P. J. Lager, J. D. Keene, Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A* **97**, 14085-14090 (2000).
53. J. Zhao *et al.*, Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**, 939-953 (2010).
54. A. P. Gerber, D. Herschlag, P. O. Brown, Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS biology* **2**, E79 (2004).
55. C. A. McHugh, P. Russell, M. Guttman, Methods for comprehensive experimental identification of RNA-protein interactions. *Genome biology* **15**, 203 (2014).
56. S. Mili, J. A. Steitz, Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692-1694 (2004).
57. K. J. Riley, J. A. Steitz, The "Observer Effect" in genome-wide surveys of protein-RNA interactions. *Mol Cell* **49**, 601-604 (2013).
58. K. J. Riley, T. A. Yario, J. A. Steitz, Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA* **18**, 1581-1585 (2012).
59. J. Ule *et al.*, CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215 (2003).
60. D. D. Licatalosi *et al.*, HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469 (2008).

61. M. Hafner *et al.*, Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).
62. J. Konig *et al.*, iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-915 (2010).
63. E. L. Van Nostrand *et al.*, Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-514 (2016).
64. B. J. Zarnegar *et al.*, irCLIP platform for efficient characterization of protein-RNA interactions. *Nat Methods* **13**, 489-492 (2016).
65. R. B. Darnell, HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* **1**, 266-286 (2010).
66. G. Martin, W. Keller, RNA-specific ribonucleotidyl transferases. *RNA* **13**, 1834-1849 (2007).
67. M. Edmonds, R. Abrams, Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *J Biol Chem* **235**, 1142-1149 (1960).
68. M. J. Schmidt, C. J. Norbury, Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *WIREs RNA* **1**, 142-151 (2010).
69. G. Martin, S. Doublet, W. Keller, Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. *Biochim Biophys Acta* **1779**, 206-216 (2008).
70. R. Trippe *et al.*, Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* **12**, 1494-1504 (2006).
71. C. J. Norbury, Cytoplasmic RNA: a case of the tail wagging the dog. *Nat Rev Mol Cell Biol* **14**, 643-653 (2013).
72. L. Wang, C. R. Eckmann, L. C. Kadyk, M. Wickens, J. Kimble, A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature* **419**, 312-316 (2002).
73. N. Suh, B. Jedamzik, C. R. Eckmann, M. Wickens, J. Kimble, The GLD-2 poly(A) polymerase activates *gld-1* mRNA in the *Caenorhabditis elegans* germ line. *Proc Natl Acad Sci U S A* **103**, 15108-15112 (2006).
74. J. E. Kwak, L. Wang, S. Ballantyne, J. Kimble, M. Wickens, Mammalian GLD-2 homologs are poly(A) polymerases. *Proc Natl Acad Sci U S A* **101**, 4407-4412 (2004).
75. J. E. Kwak, M. Wickens, A family of poly(U) polymerases. *RNA* **13**, 860-867 (2007).
76. M. Wickens, J. E. Kwak, Molecular biology. A tail tale for U. *Science* **319**, 1344-1345 (2008).
77. C. J. Norbury, 3' Uridylation and the regulation of RNA function in the cytoplasm. *Biochem Soc Trans* **38**, 1150-1153 (2010).
78. B. M. Lunde, I. Magler, A. Meinhart, Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res* **40**, 9815-9824 (2012).
79. P. Munoz-Tello, C. Gabus, S. Thore, Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure* **20**, 977-986 (2012).
80. L. A. Yates *et al.*, Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat Struct Mol Biol* **19**, 782-787 (2012).
81. M. J. Schmidt, S. West, C. J. Norbury, The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* **17**, 39-44 (2011).

82. I. Heo *et al.*, Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**, 276-284 (2008).
83. I. Heo *et al.*, TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696-708 (2009).
84. I. Heo *et al.*, Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. *Cell* **151**, 521-532 (2012).
85. J. P. Hagan, E. Piskounova, R. I. Gregory, Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* **16**, 1021-1025 (2009).
86. J. E. Thornton, H. M. Chang, E. Piskounova, R. I. Gregory, Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA* **18**, 1875-1885 (2012).
87. M. Wickens, D. S. Bernstein, J. Kimble, R. Parker, A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-157 (2002).
88. R. Lehmann, C. Nussleinvolhard, Involvement of the Pumilio Gene in the Transport of an Abdominal Signal in the Drosophila Embryo. *Nature* **329**, 167-170 (1987).
89. B. L. Zhang *et al.*, A conserved RNA-binding protein that regulates sexual fates in the C-elegans hermaphrodite germ line. *Nature* **390**, 477-484 (1997).
90. D. S. Spassov, R. Jurecic, The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life* **55**, 359-366 (2003).
91. A. Salvetti *et al.*, DjPum, a homologue of Drosophila Pumilio, is essential to planarian stem cell maintenance. *Development* **132**, 1863-1874 (2005).
92. M. A. Miller, W. M. Olivas, Roles of Puf proteins in mRNA degradation and translation. *Wiley Interdisciplinary Reviews-Rna* **2**, 471-492 (2011).
93. P. Lasko, mRNA Localization and Translational Control in Drosophila Oogenesis. *Csh Perspect Biol* **4**, (2012).
94. P. F. Cho *et al.*, Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. *Curr Biol* **16**, 2035-2041 (2006).
95. C. Wreden, A. C. Verrotti, J. A. Schisa, M. E. Lieberfarb, S. Strickland, Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. *Development* **124**, 3015-3023 (1997).
96. D. Chagnovich, R. Lehmann, Poly(A)-independent regulation of maternal hunchback translation in the Drosophila embryo. *P Natl Acad Sci USA* **98**, 11359-11364 (2001).
97. K. Friend *et al.*, A conserved PUF-Ago-eEF1A complex attenuates translation elongation. *Nat Struct Mol Biol* **19**, 176-183 (2012).
98. S. L. Crittenden *et al.*, A conserved RNA-binding protein controls germline stem cells in Caenorhabditis elegans. *Nature* **417**, 660-663 (2002).
99. J. A. Kaye, N. C. Rose, B. Goldsworthy, A. Goga, N. D. L'Etoile, A 3' UTR Pumilio-Binding Element Directs Translational Activation in Olfactory Sensory Neurons. *Neuron* **61**, 57-70 (2009).
100. E. Kaymak, L. M. Wee, S. P. Ryder, Structure and function of nematode RNA-binding proteins. *Curr Opin Struct Biol* **20**, 305-312 (2010).
101. J. P. Vessey *et al.*, Mammalian Pumilio 2 regulates dendrite morphogenesis and synaptic function. *P Natl Acad Sci USA* **107**, 3222-3227 (2010).

102. J. P. Vessey *et al.*, An Asymmetrically Localized Staufen2-Dependent RNA Complex Regulates Maintenance of Mammalian Neural Stem Cells. *Cell Stem Cell* **11**, 517-528 (2012).
103. B. A. Hook, A. C. Goldstrohm, D. J. Seay, M. Wickens, Two yeast PUF proteins negatively regulate a single mRNA. *J Biol Chem* **282**, 15430-15438 (2007).
104. X. Wang, P. D. Zamore, T. M. Hall, Crystal structure of a Pumilio homology domain. *Mol Cell* **7**, 855-865 (2001).
105. X. Q. Wang, J. McLachlan, P. D. Zamore, T. M. T. Hall, Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501-512 (2002).
106. Y. M. Wang, L. Opperman, M. Wickens, T. M. T. Hall, Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *P Natl Acad Sci USA* **106**, 20186-20191 (2009).
107. D. Zhu, C. R. Stumpf, J. M. Krahn, M. Wickens, T. M. Hall, A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc Natl Acad Sci U S A* **106**, 20192-20197 (2009).
108. C. Qiu *et al.*, Divergence of Pumilio/fem-3 mRNA Binding Factor (PUF) Protein Specificity through Variations in an RNA-binding Pocket. *Journal of Biological Chemistry* **287**, 6949-6957 (2012).
109. C. T. Valley *et al.*, Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc Natl Acad Sci U S A* **109**, 6054-6059 (2012).
110. M. T. Miller, J. J. Higgin, T. M. Hall, Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol* **15**, 397-402 (2008).
111. Z. T. Campbell, C. T. Valley, M. Wickens, A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* **21**, 732-738 (2014).
112. Z. T. Campbell *et al.*, Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep* **1**, 570-581 (2012).
113. D. F. Porter, Y. Y. Koh, B. VanVeller, R. T. Raines, M. Wickens, Target selection by natural and redesigned PUF proteins. *Proc Natl Acad Sci U S A* **112**, 15868-15873 (2015).
114. A. C. Goldstrohm, M. Wickens, Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* **9**, 337-344 (2008).
115. W. Olivas, R. Parker, The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J* **19**, 6602-6611 (2000).
116. J. S. Jackson, Jr., S. S. Houshmandi, F. Lopez Leban, W. M. Olivas, Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA* **10**, 1625-1636 (2004).
117. S. S. Houshmandi, W. M. Olivas, 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).

118. D. Lee *et al.*, PUF3 acceleration of deadenylation in vivo can operate independently of CCR4 activity, possibly involving effects on the PAB1-mRNP structure. *J Mol Biol* **399**, 562-575 (2010).
119. Y. Saint-Georges *et al.*, Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One* **3**, e2293 (2008).
120. A. C. Goldstrohm, B. A. Hook, D. J. Seay, M. Wickens, PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol* **13**, 533-539 (2006).
121. A. C. Goldstrohm, D. J. Seay, B. A. Hook, M. Wickens, PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J Biol Chem* **282**, 109-114 (2007).
122. J. J. Chritton, M. Wickens, Translational repression by PUF proteins in vitro. *RNA* **16**, 1217-1225 (2010).
123. N. H. Blewett, A. C. Goldstrohm, A Eukaryotic Translation Initiation Factor 4E-Binding Protein Promotes mRNA Decapping and Is Required for PUF Repression. *Molecular and Cellular Biology* **32**, 4181-4194 (2012).
124. L. J. Garcia-Rodriguez, A. C. Gay, L. A. Pon, 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).
125. M. Chatenay-Lapointe, G. S. Shadel, Repression of mitochondrial translation, respiration and a metabolic cycle-regulated gene, SLF1, by the yeast Pumilio-family protein Puf3p. *PLoS One* **6**, e20441 (2011).
126. E. Eliyahu *et al.*, Tom20 Mediates Localization of mRNAs to Mitochondria in a Translation-Dependent Manner. *Molecular and Cellular Biology* **30**, 284-294 (2010).
127. C. J. Kershaw *et al.*, Integrated multi-omics analyses reveal the pleiotropic nature of the control of gene expression by Puf3p. *Sci Rep-Uk* **5**, (2015).

CHAPTER 2

The nucleic acid binding domain and translational repression activity of a *Xenopus* terminal uridylyl transferase

Christopher P. Lapointe¹ and Marvin Wickens^{1,2}

This chapter was published:

Lapointe CP and Wickens M. 2013. *Journal of Biological Chemistry*. Jul 12;288(28):20723-33. doi: 10.1074/jbc.M113.455451. PMID: 23709223.

I performed all experiments and contributed all data. I wrote the paper with Marv.

¹Integrated Program in Biochemistry, University of Wisconsin, Madison, WI 53706

²Department of Biochemistry, University of Wisconsin, Madison, WI, 53706

ABSTRACT

Terminal uridylyl transferases (TUTs) catalyze the addition of uridines to the 3' ends of RNAs and are implicated in the regulation of both messenger RNAs and microRNAs. To better understand how TUTs add uridines to RNAs, we focused on a putative TUT from *Xenopus laevis*, XTUT7. We determined that XTUT7 catalyzed the addition of uridines to RNAs. Mutational analysis revealed that a truncated XTUT7 enzyme, which contained solely the nucleotidyl transferase and poly(A) polymerase-associated domains, was sufficient for catalytic activity. XTUT7 activity decreased upon removal of the CCHC zinc finger domains and a short segment of basic amino acids (the basic region). Furthermore, the basic region bound nucleic acids *in vitro*. We also demonstrated that XTUT7 repressed translation of a polyadenylated RNA, to which it added a distinct number of uridines. We generated a predicted structure of XTUT7's catalytic core that indicated histidine 1269 was likely important for uridine specificity. Indeed, mutation of histidine 1269 broadened XTUT7's nucleotide specificity and abolished XTUT7-dependent translational repression. Our data reveal key aspects of how XTUT7 adds uridines to RNAs, highlight the role of the basic region, illustrate that XTUT7 can repress translation, and identify an amino acid important for uridine specificity.

INTRODUCTION

Addition of non-templated uridines to the 3' ends of RNAs is an emerging form of RNA control that can influence RNA stability and processing (1-4). Terminal uridylation regulates both messenger RNAs (mRNAs)¹ and small, non-coding RNAs. Uridylation of specific mRNAs in *Schizosaccharomyces pombe* and mammalian cells promotes removal of the 5' cap and subsequent degradation (5-11). Uridylation of microRNA precursors (pre-miRNAs) can either block or promote processing depending on the cellular context (12-18). In nematodes, plants and algae, uridylation destabilizes microRNAs (miRNAs) and/or small-interfering RNAs (siRNAs) (19-22). Mammalian miRNAs are also uridylated (23-26). Despite the apparent pervasiveness of RNA uridylation, little is known about the enzymes that add uridines to RNAs.

Terminal uridylyl transferases (TUTs; aka poly(U) polymerases) add uridines to RNAs (27-29). TUTs are non-canonical ribonucleotidyl transferases (rNTases) of the DNA polymerase- β superfamily, which contains enzymes that add nucleotides to a variety of substrates, including RNAs (28). The nucleotide specificity of a particular rNTase is difficult to predict by amino acid sequence and must be experimentally derived, as determinants for specificity remain unclear. Additionally, TUTs possess several conserved domains: the nucleotidyl transferase domain (NTD), the poly(A) polymerase-associated domain (PAPD), and the nucleotide recognition motif (NRM) (3,28). NTDs contain the conserved catalytic motif characteristic of rNTases and the catalytic triad of acidic residues, typically aspartates (28). PAPDs encode an NRM, which mediates nucleotide specificity by contacting the base in the active site (28,30-33).

TUT7 orthologs are poly(U)-adding enzymes implicated in the regulation of let-7 miRNA biogenesis, a family of miRNAs critical during development and oncogenesis (6,16,17,29,34). Uridylation of let-7 precursor (pre-let-7) RNAs can either block or promote processing, depending on cell type. In mammalian stem cells, the paralogs TUT7 and TUT4 add several uridines to pre-let-7 after recruitment by the RNA-binding protein LIN28 (12,14-16). Uridylation blocks processing of pre-let-7 into mature miRNAs, as well as destabilizes pre-let-7 RNAs. In mammalian somatic cells, however, TUT7 acts independent of LIN28 and adds a single uridine to a subset of pre-let-7 RNAs (17). Monouridylation of these pre-miRNAs creates an optimal 3' end for downstream processing into mature miRNAs.

To further understand TUT7-dependent RNA uridylation, we identified and focused on *Xenopus* TUT7 (XTUT7), as it may have key roles in the oocyte and/or embryo. We sought to better understand how XTUT7 adds uridines to RNAs and its potential role in mRNA regulation. We utilized *Xenopus* oocytes due to their experimental advantages, primarily their ability to efficiently translate microinjected RNAs. With this approach, we identified XTUT7 domains important for catalytic activity, illustrated that XTUT7 can repress translation of a polyadenylated RNA, and pinpointed an important residue for uridine specificity. Our experiments also revealed a key role for a small region of basic amino acids that binds nucleic acids.

RESULTS

XTUT7 domain structure-A single *XTUT7* ortholog is encoded by the *Xenopus* genome. Putative *XTUT7* cDNAs were cloned from both *X. laevis* and *X. tropicalis* oocytes and are collectively referred to as *XTUT7*. *XTUT7* cDNA encodes conserved domains typical of rNTases (28). In particular, the domains include two NTDs, two PAPDs, and an NRM encoded in each PAPD (Fig. 1A). *XTUT7* also possesses a C2H2 zinc finger domain, three CCHC zinc finger domains, and a short, arginine-rich segment of basic amino acids (the basic region or BR). The amino acids that span the C-terminal NTD and PAPD are 84% identical and 95% similar between *Xenopus* and *Homo sapiens* *TUT7*. Overall, the proteins are 57% identical and 77% similar. *XTUT7* includes the conserved motif characteristic of known rNTase active sites, which contains two of its three putative catalytic aspartates (Fig. 1B) (28).

XTUT7 and its orthologs encode two distinct NTDs and PAPDs. *XTUT7* orthologs have canonical NTDs in their C-terminal halves and cryptic NTDs (NTD*) in their N-terminal halves. The NTD contains the three canonical catalytic aspartates (Fig. 1C). In contrast, the NTD* contains aspartate to asparagine and aspartate to lysine substitutions of aspartates two and three, respectively. *XTUT7* orthologs also have distinct NRMs encoded in their N- and C-terminal PAPDs. The NRMs encoded in their C-terminal PAPDs contain highly conserved type-2 NRMs characteristic of rNTases, which includes an invariant histidine (Fig. 1D) (28). The NRMs encoded in their cryptic, N-terminal PAPDs (PAPD*) are more divergent among species and contain an arginine in place of the histidine (Fig. 1D).

XTUT7 is a poly(U)-adding enzyme-To test whether XTUT7 can catalyze rNTase activity, XTUT7 was tethered to reporter RNA using MS2 coat protein and MS2 binding sites (Fig. 2A). mRNAs encoding portions of XTUT7 fused to three HA tags and MS2 coat protein (3HA/MS2) were microinjected into *X. laevis* oocytes to directly translate the proteins. Radiolabeled RNA substrates that contained three MS2 binding sites were then microinjected and analyzed by polyacrylamide gel electrophoresis. Diagrams of the proteins tested are shown in Fig. 2B.

XTUT7 proteins that contained a wild-type NTD added nucleotides to the end of the reporter RNA. Tethered XTUT7-FL and the control poly(A) polymerase GLD2 (47,48) yielded a heterogeneous mix of slower migrating RNA relative to the reporter alone control (Fig. 2C). This indicated that the enzymes added long nucleotide tails to the reporter RNA. Substituting catalytic aspartates with alanines in the NTD of XTUT7 (DADA) or GLD2 (D242A) prevented extension of the reporter RNA. Mutant and wild-type enzymes were expressed comparably; therefore, differences in activity were not due to differences in expression levels. Truncated *X. laevis* and *tropicalis* XTUT7 proteins (XTUT7-C and XT-TUT7-C, respectively) that lacked the NTD* and PAPD* were as active as the full-length protein, and again inactivated by mutation of catalytic aspartates. A construct of *X. laevis* TUT4 (XTUT4-C) that lacked its NTD* and PAPD* extended the reporter RNA much like XTUT7-C (Fig. 2D). XTUT7's C-terminal half is therefore sufficient to add nucleotides to RNAs.

To identify the nucleotide(s) added by XTUT7, RNAs extended by XTUT7 were assayed by RT-PCR using oligo-(dT), -(dA), -(dC), or -(dG) as the RT primer (Fig. 2E). XTUT7-C and XTUT7-FL samples yielded RT-PCR products solely in oligo-(dA) primed

reactions and only when the catalytic aspartates were present (Fig. 2F). Therefore, XTUT7 added uridines to the reporter RNA. Conversely, the control poly(A) polymerase GLD-2 yielded products only with an oligo-(dT) primer, which indicated that the enzyme added adenosines to the reporter RNA (47,48). Sequencing of cloned XTUT7-C RT-PCR products confirmed that uridines had been added (Fig. 2G). Thus, XTUT7 is a poly(U)-adding enzyme.

XTUT7 extends RNAs independent of MS2 tethering- Two lines of evidence demonstrate that XTUT7 extends RNAs independent of MS2 tethering. First, XTUT7-FL and XTUT7-C extended an RNA that lacked MS2 binding sites by ~30-50 nucleotides, and this activity was eliminated by the mutation of catalytic aspartates (Fig. 3A). Second, an XTUT7-C construct that lacked both MS2 coat protein and the 3HA tag extended a reporter RNA that contained three MS2 binding sites by up to 50 nucleotides (Fig. 3B, last lane). Tethered XTUT7-C extended the same RNA by ~200 nucleotides on average. As expected, both tethered and untethered XTUT7-C added uridines to reporter RNA (Fig. 3C). Therefore, the C-terminal half of XTUT7 uridylates RNAs independent of MS2 tethering, and XTUT7's activity is increased when tethered.

*The BR and CCHC zinc finger domains mediate XTUT7's tethering-independent activity-*To examine the role of the BR and CCHC zinc finger domains in XTUT7, mutant enzymes were constructed in the context of the C-terminal half of the protein (Fig. 4A), which possesses the same rNTase activities as the full-length protein (see Fig. 2). The mutant enzymes were first assayed on a radiolabeled reporter RNA that contained MS2 binding sites. Deletion of a single zinc finger ($\Delta Z1$), all three zinc fingers ($\Delta Z123$), or the BR (ΔBR) yielded nucleotide tail lengths similar to XTUT7-C (Fig. 4B). The XTUT7

enzyme that lacked both the zinc fingers and BR ($\Delta Z123\Delta BR$) added many fewer nucleotides than wild-type XTUT7-C when expressed at comparable levels (Fig. 4D). Consequently, the BR and zinc finger domains likely act redundantly to contribute to XTUT7's catalytic activity. However, the diminished activity also could result from a population of misfolded enzyme.

To further examine the role of the BR and CCHC zinc finger domains, the XTUT7 mutants were assayed on a reporter RNA that lacked MS2 binding sites. The mutant XTUT7 enzyme that lacked the zinc fingers and BR ($\Delta Z123\Delta BR$) was inactive on the RNA without binding sites (Fig. 4C). Mutant XTUT7 enzymes that lacked the BR (ΔBR) or the zinc fingers ($\Delta Z123$) were less active than wild-type XTUT7-C when expressed at a comparable level (Fig. 4D). In addition, tethered XTUT7 mutants retained uridine specificity (Fig. 4E). Thus, the BR, as well as the CCHC zinc fingers, mediates XTUT7's tethering-independent uridylation activity.

The BR is a conserved domain that may bind nucleic acids. The BR and CCHC zinc finger domains of XTUT7 are conserved among XTUT7 orthologs, including *H. sapiens* TUT7 (Fig. 4F). The BR resembles arginine-rich motifs (ARMs) found in viral RNA-binding proteins, such as HIV REV and TAT, as both the BR and ARMs are composed primarily of arginine (Fig. 4G) (49-51). In contrast, a recently identified basic stretch of amino acids in PAPD5, a poly(A) polymerase related to XTUT7, is composed primarily of lysine (52,53). Intriguingly, the ARMs in REV and TAT, as well as the basic stretch in PAPD5, directly bind RNA (49-52).

The BR binds nucleic acids-To test whether the BR directly binds nucleic acids, wild-type (BR-WT) and mutant BR (BR-R1-6A) segments were fused to an MBP-His₆ tag

(Fig. 5A), purified (Fig. 5B), and tested using electrophoretic mobility shift assays. BR-WT bound an RNA that contained three MS2 binding sites in a concentration-dependent manner, with an apparent K_d of 40 ± 5 nM (Fig. 5C&E). BR-WT also bound a ssDNA substrate of an equivalent sequence to the RNA substrate, with an apparent K_d of 70 ± 5 nM (Fig. 5D&E). At the highest protein concentrations, the protein-nucleic acid complexes migrated progressively slower, which may indicate that multiple copies of BR-WT can bind the same nucleic acid molecule. A mutant BR in which arginines 1-6 had been changed to alanine (BR-R1-6A) bound the RNA and ssDNA substrates poorly, with estimated apparent K_d s of greater than 400 and 350 nM, respectively (Fig. 5E). Protein-RNA complexes were not observed with MBP-His₆ backbone alone on either substrate (Fig. 5E). Therefore, the BR directly binds nucleic acids and requires conserved arginines for optimal binding.

XTUT7 represses a polyadenylated RNA-To examine XTUT7's rNTase activity on polyadenylated RNA, we tethered XTUT7 to an RNA with a poly(A)₃₉ tail. XTUT7-FL and XTUT7-C extended the polyadenylated reporter RNA by a distinct number of nucleotides, which on average was 60 ± 10 nucleotides (Fig. 6A). RT-PCR of the polyadenylated RNAs extended by XTUT7-C yielded products with both oligo-(dT) and -(dA) primers, while the substrate yielded a single band in the oligo-(dT) lane (Fig. 6B). XTUT7 therefore added a poly(U) tail of discrete length to an RNA substrate with 39 adenosines.

To assess the effect of XTUT7 on translation, XTUT7 was tethered to a poly(A)₃₉ firefly luciferase mRNA (Fig. 6C). As a control, we co-injected *Renilla* luciferase mRNA that lacked MS2 binding sites and a poly(A) tail. Tethered XTUT7 specifically reduced firefly luciferase activity nearly 3-fold (p -value < 0.001), much like the characterized

translational repressor Xp54 (Fig. 6D, top and middle panels) (54). XTUT7-dependent repression was specific to firefly luciferase, as all proteins tested had no significant effect on *Renilla* luciferase luminescence (Fig. 6D, bottom panel). XTUT7-C-DADA, which lacks catalytic activity, yielded firefly luciferase levels similar to that of the control GLD2-D242A and was expressed similarly to XTUT7-C. All proteins tested had no significant effect on firefly or *Renilla* luciferase mRNA levels as determined by qPCR. Thus, XTUT7 represses translation of polyadenylated reporter mRNA by adding uridines to the RNA.

Histidine 1269 is important for XTUT7's uridine specificity and repression activity-

To visualize the likely structure of the active site region, the three-dimensional structure of XTUT7's core catalytic domains was predicted using the I-TASSER server (55-57). This analysis yielded ligand-free and ligand-bound homology models of XTUT7 (C-score = 1.02, expected RMSD = 4.3 ± 2.9 Å) (Fig. 7A). In the model, putative catalytic aspartates 1041, 1043, and 1102 are adjacent to the triphosphate moiety of UTP (Fig. 7A). Tyrosine 1154 appears to participate in a stacking interaction with uracil. Histidine 1269, contained in XTUT7's NRM, is predicted to contact a carbonyl oxygen in UTP. As expected, the predicted XTUT7 structure aligns well to a structure of the *S. pombe* poly(U)-adding enzyme CID1 (alignment RMSD = 1.2 Å, sequence identity to XTUT7 is 32%) (Fig. 7B) (30). Intriguingly, a hydrogen bond that is observed in CID1 between His336 and UTP is predicted in XTUT7 (His1269).

We reasoned that His1269 might be important for nucleotide specificity due to its proximity to UTP. Therefore, we substituted His1269 with leucine in the context of XTUT7's C-terminal half (XTUT7-H1269L), since this substitution ablates a potential hydrogen bond to UTP (Fig. 7C). Indeed, tethered XTUT7-H1269L added cytosines, as

well as uridines, to RNA (Fig. 7D). XTUT7-H1269L-dependent tails were ~ 20% cytosine, as compared to ~ 3% with the wild-type enzyme (p-values < 0.005). Both XTUT7-H1269L and XTUT7-C rarely added guanosines or adenosines to RNA (< 3 and 2%, respectively). Thus, His1269 is important for the uridine specificity of XTUT7.

To test whether altered uridine specificity affected XTUT7's catalytic activity, XTUT7-H1269L was tethered to a polyadenylated reporter RNA. XTUT7-H1269L added a heterogeneous length tail to an RNA with a poly(A)₃₉ tail, rather than the discrete ~60 nucleotide tail added by XTUT7-C (Fig. 7E). The tail added by XTUT7-H1269L was between ~50-150 nucleotides in length, which was shorter than the tail added by the wild-type enzyme to RNA that lacked a poly(A) tail. Accordingly, incorporation of non-uridine residues by XTUT7 prevents formation of the discrete length tail on the poly(A)₃₉ reporter RNA.

To determine the effect of XTUT7-H1269L on translational repression, XTUT7-H1269L was assayed using poly(A)₃₉ firefly luciferase mRNA. XTUT7-H1269L not only prevented translational repression, but instead activated it ~3-fold (p-value < 0.05) (Fig. 7F). This increase in firefly luciferase activity was less than the increase yielded by the poly(A) polymerase GLD2. Firefly and *Renilla* luciferase mRNA levels were not significantly affected by any protein tested. Thus, the H1269L substitution alleviated XTUT7-dependent translational repression, which could result from either the mutant enzyme's relaxed nucleotide specificity or its addition of a heterogeneous length tail to poly(A)₃₉ firefly luciferase mRNA.

DISCUSSION

The C-terminal half of XTUT7 is sufficient for rNTase activity-Our studies demonstrate that XTUT7 is a poly(U)-adding enzyme and identify key domains and residues important for activity. Collectively, XTUT7's NTD and PAPD were sufficient for activity and therefore represent core catalytic domains. Similarly, mutations in the NTD of full-length XTUT7 abolished rNTase activity. The BR and CCHC zinc finger domains flanking the catalytic core likely enable efficient XTUT7-dependent uridylation, as their removal decreased XTUT7 rNTase activity. In contrast, the conserved NTD* and PAPD* of XTUT7 lacked activity and were dispensable for it in the context of the full-length protein. The analogous domains in XTUT4 were also dispensable for catalytic activity. Thus, the C-terminal half of XTUT7 is sufficient for rNTase activity.

Recent work on the mammalian TUT7 ortholog, TUT4, suggested that its NTD* was necessary for rNTase activity, and therefore contrasts with our finding that the analogous domain of XTUT7 was dispensable (16). The NTD* of TUT4 was required for the enzyme to uridylate synthetic pre-let-7 RNA both in the presence or absence of LIN28, which is thought to recruit TUT4 to its let-7 RNA target. Despite the high degree of similarity between *Xenopus* and human TUTs (77% similar), their individual domains could function differently. However, we note that in our assays, XTUT7 is tethered to RNA, which makes the assay more sensitive. Our assays are independent of LIN28, as nearly all of the XTUT7 proteins we tested lacked the C2H2 zinc finger domain needed for LIN28-dependent uridylation, likely through protein-protein contacts (16). Although the NTD* and PAPD* of XTUT7 orthologs are dispensable for catalytic activity, they nonetheless may have critical roles *in vivo*. For example, they may mediate protein-protein interactions, as

suggested by LIN28-pre-let-7-TUT4 experiments (16). Indeed, TUT4 segments that contain its NTD* and PAPD* promote cell proliferation independent of catalytic activity (58).

The XTUT7 basic region-XTUT7 contains nucleic acid binding domains, including the arginine-rich BR. XTUT7 possesses tethering-independent rNTase activity redundantly mediated by its BR and CCHC zinc finger domains. These domains are likely required for efficient catalytic activity when XTUT7 is tethered to RNA. Together, these findings suggest that the BR and at least one of the CCHC zinc finger domains bind RNA. Indeed, we show that the BR binds both RNA and ssDNA *in vitro*. Given the modest preference of the BR for binding RNA and that rNTases lack catalytic activity on DNA substrates (59), we suggest that the BR is an RNA-binding domain *in vivo*.

XTUT7's BR resembles RNA-binding domains present in certain viral proteins, such as the ARM found in HIV REV (49,60). ARMs are flexible RNA-binding domains that typically confer specificity for particular RNAs by recognizing RNA sequences and/or structures (61,62). For example, the ARM in REV specifically recognizes and binds its RNA target (63-66). Critical arginines in the ARM make base-specific contacts with the RNA and are necessary for binding. Similarly, XTUT7's BR requires highly conserved arginines for optimal RNA-binding activity.

BRs are present in other rNTases. Human PAPD5, a non-canonical poly(A) polymerase, binds a subset of RNAs likely through a small, lysine-rich stretch of amino acids (52). PAPD5's basic stretch is also required for efficient catalytic activity, much like the BR in XTUT7. A search for similar BRs in human rNTases reveals that five of the seven enzymes contain characterized or putative BRs, including TUT7, TUT4, and

PAPD5. Thus, BRs appear to be a common feature of rNTases that are likely utilized to bind RNA substrates and/or facilitate catalytic activity.

We speculate that the BR and CCHC zinc fingers facilitate TUT7 binding to particular RNAs *in vivo*. TUT7 orthologs uridylylate pre-let-7 in the absence of LIN28 both *in vitro* and *in vivo* and this activity requires pre-let-7's stem (16,17,67). We therefore propose that the BR of TUT7, likely in cooperation with the zinc fingers, binds the accessible region of the pre-let-7 stem.

XTUT7 homology model and nucleotide specificity-We generated a homology model of the three-dimensional structure of XTUT7's catalytic core that identified an amino acid important for uridine specificity. Not surprisingly, the predicted structure of XTUT7 is similar to those of other poly(U)-adding enzymes, particularly CID1, and predicted that a histidine would be important for uridine specificity (30). The analogous histidine in CID1 is required for optimal uridine specificity *in vitro* (30-32). Indeed, substituting His1269 with leucine broadened XTUT7's nucleotide specificity *in vivo* so that it added both uridines and cytosines. These data suggest that a histidine-UTP contact is a critical determinant for XTUT7 uridine specificity and likely represents a common mechanism of uridine recognition among XTUT7 orthologs. Direct determination of the XTUT7 structure is needed to test this rigorously.

XTUT7 and translational control-XTUT7 can repress translation and may represent a new class of translational repressor proteins. XTUT7 repressed translation of a polyadenylated reporter mRNA without affecting mRNA stability. We propose that the U-tail added by XTUT7 binds poly(A). The poly(A)-poly(U) hybrid may block recognition of the poly(A) tail by poly(A)-dependent factors, such as poly(A)-binding protein. This would

mask the effects of the poly(A) tail, including its ability to stimulate translation. The presence of the A-U duplex is consistent with the observation that an XTUT7 mutant that added cytosines no longer repressed translation. Furthermore, the mutant enzyme also produced a heterogeneous length tail, while the wild-type enzyme added a discrete number of uridines to an RNA with 39 adenosines. We infer that the newly formed A-U hybrid prevents further catalysis, implying a novel mechanism that terminates poly(U) synthesis. While RNAs with repressive A-U hybrid tails have yet to be discovered *en masse* in cells, they may well exist. For instance, the mammalian poly(U)-adding enzymes TUT7 and TUT4 associate with polyadenylated RNAs, and polyadenylated mRNAs in *S. pombe* are uridylated *in vivo* (6,8,68,69). Thus, XTUT7 may have unanticipated roles in the regulation of mRNAs, in addition to its activities in miRNA control.

EXPERIMENTAL PROCEDURES

MS2-fusion protein plasmids-The pCS2+3HA:MS2, pCS2+3HA:MS2:Xp54, and pCS2+3HA:MS2:GLD2-D242A plasmids were previously described (35). Newly constructed MS2 fusion plasmids, and the primers and restriction sites used for their construction, are listed in Supplementary Table 1. All MS2-fusion proteins were designed to be: *N-terminus* - three hemagglutinin (3HA) tags - MS2 coat protein - protein to be tested - *C-terminus*. XTUT7 (also known as ZCCHC6) and XTUT4 (also known as ZCCHC11) cDNAs were cloned from both *Xenopus laevis* (XL) and *Xenopus tropicalis* (XT) stage VI oocytes. cDNAs corresponding to XTUT7-FL (XT), XTUT7-C (XL), and XTUT4-C (XL) were deposited to GenBank with accession numbers KC493151, KC493152, and KC493153, respectively. Mutations and deletions in XTUT7 were inserted by site directed mutagenesis using *Pfu*-Ultra DNA polymerase (Agilent) and mutated/deleted residues are listed in Supplementary Table 1. Specific amino acids (i.e. H1269) discussed in the results and discussion are referenced by their location in XTUT7-FL. Domain predictions of XTUT7 proteins were completed using the InterProScan Sequence Search Tool (36,37) and Pfam (38) on XTUT7-FL.

Multiple sequence alignments- XTUT7 sequence homologs were identified by reciprocal best BLAST (NCBI): *Ciona intestinalis* (GI#: 198429697), *Strongylocentrotus purpuratus* (115933324), *Bos taurus* (329664700), *Canis lupus familiaris* (73946401), *Macaca mulatta* (109112038), *Mus musculus* (259016375), *Rattus norvegicus* (293354419), *Hydra magnipapillata* (221116335), *Monodelphis domestica* (334332807), *Ornithorhynchus anatinus* (345314193), *Danio rerio* (326668285), *Homo sapiens* (297307111), *Caenorhabditis elegans* (17554128), and *Amphimedon queenslandica*

(340382961). Sequence logos were derived from MUSCLE (39) sequence alignments of the putative XTUT7 orthologs using WebLogo (40).

Reporter RNA plasmids-The pLG-MS2 (firefly luciferase), pLG-MS2+A39 (polyadenylated firefly luciferase), pSP65-ren (*Renilla* luciferase), pLGMS2-luc (RNA with three MS2 binding sites), pLGMS2+A39-luc (RNA with three MS2 binding sites and a poly(A)₃₉ tail), pLG:FBE-ACAmut (RNA that lacked MS2 binding sites), and pLG:FBE-ACAmut+A39 (RNA with a poly(A)₃₉ tail that lacked MS2 binding sites) plasmids have been described (41-44).

In vitro transcriptions-RNAs were *in vitro* transcribed from restriction digested plasmids using either the AmpliScribe SP6 High Yield Transcription or T7-Flash Transcription kits (Epicentre). RNAs encoded in pCS2+3HA:MS2 (NotI, SP6), pSP65-ren (Sall, SP6), pLG-MS2+A39 (BamHI, T7), pLGMS2+A39-luc (BamHI, T7), pLG-MS2 (BglII, T7) and pLGMS2-luc (BglII, T7) plasmids were prepared with the indicated reagents. All reactions included m⁷G(5')ppp(5')G cap analog (NEB). In some cases, [32 P]UTP was included in order to radiolabel the RNA.

Oocyte injections and RNA analysis-Oocyte injections were performed as described (41,44). Oocytes were collected after overnight incubation (~16 hrs). Total RNA was extracted from 10 oocytes using TRI reagent (Sigma). Total RNA from 3 oocytes was separated on denaturing 6% polyacrylamide gels and analyzed by phosphor imaging. Densitometric analyses were completed using ImageQuant software (GE Healthcare Life Sciences).

Luciferase assays and Western blotting-Dual Luciferase Assays (Promega) and Western blotting were performed as described (35,43). Student's two-tailed t-tests were used to calculate all p-values.

RT-PCR assays-Total RNA was treated with 4 units TURBO-DNase (Life Technologies) for 1 hr at 37°C, and then purified using the GeneJET RNA Purification Kit (Fermentas). 1 µg of total RNA was reverse transcribed using ImPromII reverse transcriptase (Promega) and 1 µM of oligo -(dA)₁₈, -(dT)₁₈, -(dC)₂₀, or -(dG)₂₀ for the RT primers, as indicated. cDNA was PCR amplified using a firefly luciferase-specific forward primer (GCGTTAATCAGAGAGGCGAATTATGTG) and the corresponding RT primer. For qRT-PCR assays, 100 ng of total RNA were reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) and a random hexamer primer, and 5% of the cDNA was amplified using the Perfeta qPCR FastMix UNG Low ROX kit (Quanta Biosciences). Firefly luciferase levels were compared to *Renilla* luciferase and β-actin mRNA levels.

Tail sequencing assays-The tail sequencing assay was performed essentially as described (45) with the following modifications. The P1 anchor primer (AATATTCACCTTGATCTGAAGC) was 5' phosphorylated using PNK enzyme (Promega) and 3'-blocked with cordycepin (Sigma-Aldrich) using TdT enzyme (NEB). 400 ng of modified P1 primer were pre-annealed with 400 ng of P1' (GCTTCAGATCAAGGTGAATATTA AAAA), and ligated to 1-2 µg of total RNA using T4 RNA Ligase (Fermentas) at 37°C for 1 hr. Reverse transcription reactions were completed using 1 µM P1' oligo and SuperScript III reverse transcriptase (Invitrogen). Two rounds of nested PCR amplification were performed using forward primer 1 (GCGTTAATCAGAGAGGCGAATTATGTG) and forward primer 2

(ACCTCTCTCTCTCTCAGGGCTGATTACTAG). P1' was the reverse primer in both reactions. PCR products from the second PCR were TOPO-TA cloned (Invitrogen) and sequenced. 10 tails added by XTUT7-C (227 bases total) and 14 tails added by XTUT7-H1269L (751 bases) were independently cloned and sequenced.

Protein purifications-Amino acids 453-540 of XTUT7-C (KC493152) were fused to an N-terminal maltose-binding protein (MBP) tag and a C-terminal 6-histidine (His₆) tag by cloning into the XbaI restriction site of the previously described pHMTC plasmid (46). Residues of interest were mutated by site-directed mutagenesis (Supp. Table 1). Plasmids were transformed into BL21-CodonPlus(DE3)-RIL cells and grown in LB+ampicillin/chloramphenicol media at 37°C until OD₆₀₀ ~ 0.6. Protein expression was induced with 0.5 mM IPTG for 3 hrs at 37°C. Cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.02% (v/v) Tween-20), and lysed by incubation with 0.5 mg/mL lysozyme, followed by a freeze-thaw cycle. Cleared lysates were incubated with pre-washed sepharose-amylose resin (NEB) for ~2 hrs at 4°C. The resin was washed 2X in lysis buffer and 1X in wash buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl 0.02% (v/v) Tween-20). MBP-fusion proteins were eluted 3X in elution buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.02% (v/v) Tween-20, 50 mM maltose). Eluted proteins dialyzed overnight in 4L of lysis buffer and were concentrated using Vivaspin 30K MWCO columns (Sartorius). Protein concentrations were measured by Bradford assays.

Electrophoretic mobility shift assays-RNA that contained three MS2 binding sites (transcribed from pLGMS2-luc) was 5' end-labeled using PNK enzyme (Promega) and [³²P]ATP. ~3.5 fmol of end-labeled RNA were incubated with purified MBP-fusion proteins (0-1 μM) in 45 mM Tris-HCl pH 8.0, 90 mM NaCl, 0.02% (v/v) Tween-20, and 20

units RNasin (Promega) for 30 minutes on ice. 3X loading buffer (6% glycerol, 0.06% bromophenol blue) was added to each reaction and then loaded onto 5% polyacrylamide/TBE gels (BioRad). Gels were run for ~1 hr at 100V at 4°C, dried and exposed to storage phosphor screens overnight. The percent of RNA bound was calculated using ImageQuant software (GE Healthcare Life Sciences). Apparent K_d values were determined using Graphpad Prism 5 software and the "One site--Specific binding with Hill slope" equation. All values were reported with associated standard error.

XTUT7 structure prediction-The I-TASSER server was used to generate the predicted structure of amino acids 91-422 of XTUT7-C (KC493152) (46-48). Specific amino acids discussed in the text are referred to by their position in the context of XTUT7-FL (KC493151). PyMOL was used to visualize the model, create images and perform structural alignments.

REFERENCES

1. Wickens, M., and Kwak, J. E. (2008) Molecular biology. A tail tale for U. *Science* **319**, 1344-1345
2. Norbury, C. J. (2010) 3' Uridylation and the regulation of RNA function in the cytoplasm. *Biochem Soc Trans* **38**, 1150-1153
3. Schmidt, M. J., and Norbury, C. J. (2010) Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *WIREs RNA* **1**, 142-151
4. Kim, Y. K., Heo, I., and Kim, V. N. (2010) Modifications of small RNAs and their associated proteins. *Cell* **143**, 703-709
5. Song, M. G., and Kiledjian, M. (2007) 3' Terminal oligo U-tract-mediated stimulation of decapping. *RNA* **13**, 2356-2365
6. Rissland, O. S., Mikulasova, A., and Norbury, C. J. (2007) Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Mol Cell Biol* **27**, 3612-3624
7. Mullen, T. E., and Marzluff, W. F. (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev* **22**, 50-65
8. Rissland, O. S., and Norbury, C. J. (2009) Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat Struct Mol Biol* **16**, 616-623
9. Schmidt, M. J., West, S., and Norbury, C. J. (2011) The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* **17**, 39-44

10. Su, W., Slepencov, S. V., Slevin, M. K., Lyons, S. M., Ziemniak, M., Kowalska, J., Darzynkiewicz, E., Jemielity, J., Marzluff, W. F., and Rhoads, R. E. (2013) mRNAs containing the histone 3' stem-loop are degraded primarily by decapping mediated by oligouridylation of the 3' end. *RNA* **19**, 1-16
11. Hoefig, K. P., Rath, N., Heinz, G. A., Wolf, C., Dameris, J., Schepers, A., Kremmer, E., Ansel, K. M., and Heissmeyer, V. (2013) Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay. *Nat Struct Mol Biol* **20**, 73-81
12. Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V. N. (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**, 276-284
13. Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S., and Miska, E. A. (2009) LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat Struct Mol Biol* **16**, 1016-1020
14. Hagan, J. P., Piskounova, E., and Gregory, R. I. (2009) Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* **16**, 1021-1025
15. Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., Yeom, K. H., Han, J., and Kim, V. N. (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696-708

16. Thornton, J. E., Chang, H. M., Piskounova, E., and Gregory, R. I. (2012) Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA* **18**, 1875-1885
17. Heo, I., Ha, M., Lim, J., Yoon, M. J., Park, J. E., Kwon, S. C., Chang, H., and Kim, V. N. (2012) Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. *Cell* **151**, 521-532
18. Newman, M. A., Mani, V., and Hammond, S. M. (2011) Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA* **17**, 1795-1803
19. van Wolfswinkel, J. C., Claycomb, J. M., Batista, P. J., Mello, C. C., Berezikov, E., and Ketting, R. F. (2009) CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell* **139**, 135-148
20. Ibrahim, F., Rymarquis, L. A., Kim, E. J., Becker, J., Balassa, E., Green, P. J., and Cerutti, H. (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *Proc Natl Acad Sci U S A* **107**, 3906-3911
21. Ren, G., Chen, X., and Yu, B. (2012) Uridylation of miRNAs by hen1 suppressor1 in *Arabidopsis*. *Curr Biol* **22**, 695-700
22. Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T. T., Meyers, B. C., Mo, B., and Chen, X. (2012) The *Arabidopsis* nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr Biol* **22**, 689-694

23. Jones, M. R., Quinton, L. J., Blahna, M. T., Neilson, J. R., Fu, S., Ivanov, A. R., Wolf, D. A., and Mizgerd, J. P. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat Cell Biol* **11**, 1157-1163
24. Burroughs, A. M., Ando, Y., de Hoon, M. J., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y., and Daub, C. O. (2010) A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res* **20**, 1398-1410
25. Wyman, S. K., Knouf, E. C., Parkin, R. K., Fritz, B. R., Lin, D. W., Dennis, L. M., Krouse, M. A., Webster, P. J., and Tewari, M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res* **21**, 1450-1461
26. Jones, M. R., Blahna, M. T., Kozlowski, E., Matsuura, K. Y., Ferrari, J. D., Morris, S. A., Powers, J. T., Daley, G. Q., Quinton, L. J., and Mizgerd, J. P. (2012) Zcchc11 Uridylates Mature miRNAs to Enhance Neonatal IGF-1 Expression, Growth, and Survival. *PLoS Genet* **8**, e1003105
27. Aphasizhev, R. (2005) RNA uridylyltransferases. *Cell Mol Life Sci* **62**, 2194-2203
28. Martin, G., and Keller, W. (2007) RNA-specific ribonucleotidyl transferases. *RNA* **13**, 1834-1849
29. Kwak, J. E., and Wickens, M. (2007) A family of poly(U) polymerases. *RNA* **13**, 860-867

30. Yates, L. A., Fleurdepine, S., Rissland, O. S., De Colibus, L., Harlos, K., Norbury, C. J., and Gilbert, R. J. (2012) Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat Struct Mol Biol* **19**, 782-787
31. Lunde, B. M., Magler, I., and Meinhart, A. (2012) Crystal structures of the Cid1 poly(U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res* **40**, 9815-9824
32. Munoz-Tello, P., Gabus, C., and Thore, S. (2012) Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure* **20**, 977-986
33. Martin, G., Doublet, S., and Keller, W. (2008) Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. *Biochim Biophys Acta* **1779**, 206-216
34. Roush, S., and Slack, F. J. (2008) The let-7 family of microRNAs. *Trends Cell Biol* **18**, 505-516
35. Cooke, A., Prigge, A., and Wickens, M. (2010) Translational repression by deadenylases. *J Biol Chem* **285**, 28506-28513
36. Zdobnov, E. M., and Apweiler, R. (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848
37. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**, W695-699
38. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy,

- S. R., and Bateman, A. (2010) The Pfam protein families database. *Nucleic Acids Res* **38**, D211-222
39. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797
40. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190
41. Gray, N. K., Coller, J. M., Dickson, K. S., and Wickens, M. (2000) Multiple portions of poly(A)-binding protein stimulate translation in vivo. *EMBO J* **19**, 4723-4733
42. Chritton, J. J., and Wickens, M. (2010) Translational repression by PUF proteins in vitro. *RNA* **16**, 1217-1225
43. Cooke, A., Prigge, A., Opperman, L., and Wickens, M. (2011) Targeted translational regulation using the PUF protein family scaffold. *Proc Natl Acad Sci U S A* **108**, 15870-15875
44. Kwak, J. E., Wang, L., Ballantyne, S., Kimble, J., and Wickens, M. (2004) Mammalian GLD-2 homologs are poly(A) polymerases. *Proc Natl Acad Sci U S A* **101**, 4407-4412
45. Charlesworth, A., Cox, L. L., and MacNicol, A. M. (2004) Cytoplasmic polyadenylation element (CPE)- and CPE-binding protein (CPEB)-independent mechanisms regulate early class maternal mRNA translational activation in *Xenopus* oocytes. *J Biol Chem* **279**, 17650-17659

46. Hook, B. A., Goldstrohm, A. C., Seay, D. J., and Wickens, M. (2007) Two yeast PUF proteins negatively regulate a single mRNA. *J Biol Chem* **282**, 15430-15438
47. Wang, L., Eckmann, C. R., Kadyk, L. C., Wickens, M., and Kimble, J. (2002) A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature* **419**, 312-316
48. Rouhana, L., Wang, L., Buter, N., Kwak, J. E., Schiltz, C. A., Gonzalez, T., Kelley, A. E., Landry, C. F., and Wickens, M. (2005) Vertebrate GLD2 poly(A) polymerases in the germline and the brain. *RNA* **11**, 1117-1130
49. Lazinski, D., Grzadzielska, E., and Das, A. (1989) Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59**, 207-218
50. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., and Crothers, D. M. (1990) Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **249**, 1281-1285
51. Calnan, B., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. (1991) Arginine-mediated RNA recognition: the arginine fork. *Science* **252**, 1167-1171
52. Rammelt, C., Bilen, B., Zavolan, M., and Keller, W. (2011) PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* **17**, 1737-1746
53. Berndt, H., Harnisch, C., Rammelt, C., Stohr, N., Zirkel, A., Dohm, J. C., Himmelbauer, H., Tavanez, J. P., Huttelmaier, S., and Wahle, E. (2012) Maturation

- of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* **18**, 958-972
54. Minshall, N., Thom, G., and Standart, N. (2001) A conserved role of a DEAD box helicase in mRNA masking. *RNA* **7**, 1728-1742
 55. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40
 56. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738
 57. Roy, A., Yang, J., and Zhang, Y. (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* **40**, W471-477
 58. Blahna, M. T., Jones, M. R., Quinton, L. J., Matsuura, K. Y., and Mizgerd, J. P. (2011) Terminal uridyltransferase enzyme Zcchc11 promotes cell proliferation independent of its uridyltransferase activity. *J Biol Chem* **286**, 42381-42389
 59. Read, R. L., Martinho, R. G., Wang, S. W., Carr, A. M., and Norbury, C. J. (2002) Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. *Proc Natl Acad Sci U S A* **99**, 12079-12084
 60. Tan, R., Chen, L., Buettner, J. A., Hudson, D., and Frankel, A. D. (1993) RNA recognition by an isolated alpha helix. *Cell* **73**, 1031-1040
 61. Weiss, M. A., and Narayana, N. (1998) RNA recognition by arginine-rich peptide motifs. *Biopolymers* **48**, 167-180

62. Bayer, T. S., Booth, L. N., Knudsen, S. M., and Ellington, A. D. (2005) Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* **11**, 1848-1857
63. Daugherty, M. D., Liu, B., and Frankel, A. D. (2010) Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat Struct Mol Biol* **17**, 1337-1342
64. Olsen, H. S., Nelbock, P., Cochrane, A. W., and Rosen, C. A. (1990) Secondary structure is the major determinant for interaction of HIV rev protein with RNA. *Science* **247**, 845-848
65. Daly, T. J., Cook, K. S., Gray, G. S., Maione, T. E., and Rusche, J. R. (1989) Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature* **342**, 816-819
66. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**, 1547-1551
67. Yeom, K. H., Heo, I., Lee, J., Hohng, S., Kim, V. N., and Joo, C. (2011) Single-molecule approach to immunoprecipitated protein complexes: insights into miRNA uridylation. *EMBO Rep* **12**, 690-696
68. Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., and Landthaler, M. (2012) The mRNA-

bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**, 674-690

69. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., and Hentze, M. W. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406

ACKNOWLEDGEMENTS

We thank members of the Wickens laboratory for helpful discussions, suggestions and critical reading of the manuscript. We are grateful to Phil Anderson and Scott Kennedy for advice on experiments. We also thank Laura Vanderploeg and the Biochemistry Media Laboratory for help with figures. C.P.L. was supported by Genentech Foundation and Wharton Predoctoral Fellowships. M.W. was supported by NIH grants GM31892 and GM50942.

FOOTNOTES

¹The abbreviations used are: mRNA, messenger RNA; pre-miRNA, precursor microRNA; miRNA, microRNA; siRNA, small-interfering RNA; TUT, terminal uridylyl transferase; rNTase, ribonucleotidyl transferase; NTD, nucleotidyl transferase domain; PAPD, poly(A) polymerase-associated domain; NRM, nucleotide recognition motif; TUT7, either mammalian (species indicated) or generic TUT7 enzymes; XTUT7, *Xenopus* TUT7; XL, *Xenopus laevis*; XT, *Xenopus tropicalis*; BR, basic region; NTD*, cryptic nucleotidyl transferase domain; PAPD*, cryptic poly(A) polymerase-associated domain; ZF, zinc finger domain; ARM, arginine-rich motif.

Figure 1

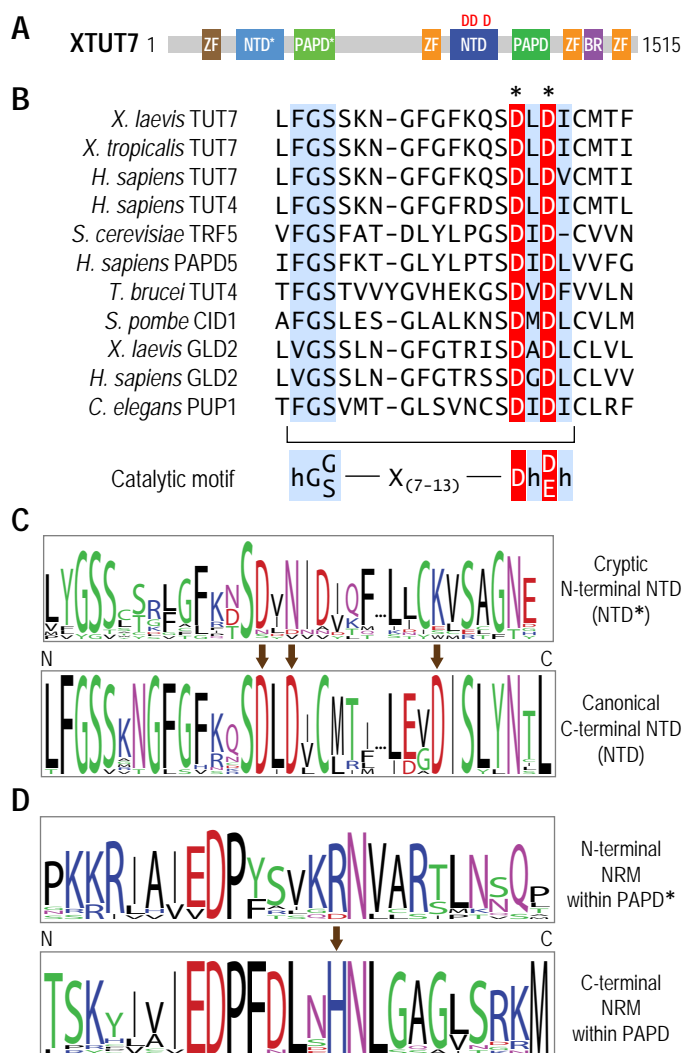


Figure 1. XTUT7 domain structure. **(A)** Diagram representing XTUT7 and its predicted protein domains (amino acids indicated). ZF (brown), C2H2 zinc finger domain (227-252); NTD*, cryptic NTD (295-437); PAPD*, cryptic PAPD (526-575); NTD, nucleotidyl transferase domain (998-1146); PAPD, poly(A) polymerase-associated domain (1215-1268); ZF (orange), CCHC zinc finger domains (946-962, 1327-1343, 1432-1448); BR, basic region (1344-1361). Putative catalytic aspartates are indicated. **(B)** Multiple sequence alignment of the characteristic catalytic motifs contained in the NTD of XTUT7 and other previously identified nucleotidyl transferases. The consensus rNTase catalytic motif - hG[G/S]X₇₋₁₃Dh[D/E]h - is shown below the sequence alignment, where h indicates any hydrophobic amino acid (28). **(C)** Sequence logos representing multiple sequence alignments of the putative catalytic motifs contained in the NTD and NTD* of XTUT7 orthologs. The arrows indicate analogous positions in the NTD and NTD*, and highlight amino acid substitutions of putative catalytic aspartates in the NTD*. **(D)** Sequence logos representing multiple sequence alignments of the nucleotide recognition motifs (NRM) encoded in the PAPDs and PAPD*s of XTUT7 orthologs. The arrow indicates an invariant histidine in the C-terminal NRM.

Figure 2

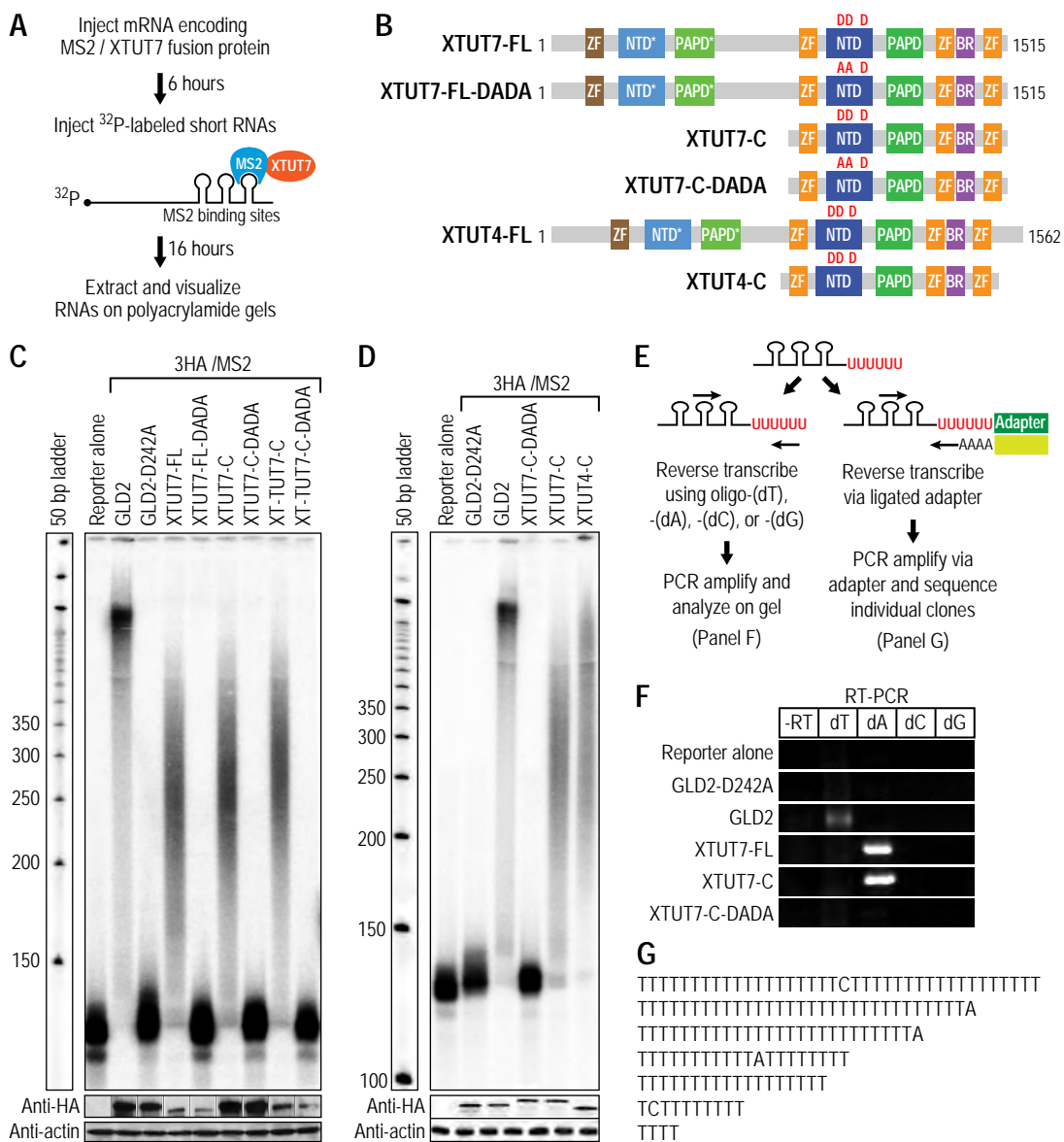


Figure 2. XTUT7 is a poly(U)-adding enzyme. **(A)** mRNAs encoding MS2 coat protein fused to XTUT7 were microinjected into oocytes. Following a 6 hour incubation to allow translation of the fusion proteins, ^{32}P -labeled RNA that contained MS2 binding sites was injected. MS2 coat protein binds the MS2 binding sites thus tethering XTUT7 to the reporter RNA. For clarity, only one binding event is depicted. After 16 hours, RNAs were extracted, analyzed on polyacrylamide gels and visualized by phosphor imaging. **(B)** Diagrams representing proteins tested in subsequent panels. The D's denote catalytic aspartates, while the A's denote aspartate to alanine substitutions. The XTUT4-FL diagram represents the domain structure of full-length XTUT4, which is a paralog to XTUT7. XTUT4-C represents the construct tested in panel D, which is analogous to XTUT7-C. **(C)** Extracted RNAs were separated on denaturing polyacrylamide gels. All proteins assayed were 3HA/ MS2 fusions. The reporter alone sample serves as a negative control for extension, while the GLD2 sample is a positive control for extension. GLD2-D242A is a mutant GLD2 containing an alanine in place of a catalytic aspartate. 3HA/MS2/XTUT7 fusion proteins tested are depicted in panel B. XTUT7-FL and XTUT7-C are derived from *X. tropicalis* and XTUT7-C is derived from *X. laevis*. XTUT7-DADA proteins contain aspartate to alanine substitutions as indicated in panel B. The bottom panels indicate protein levels as determined by Western blotting for HA-tagged fusion proteins and actin. **(D)** The indicated 3HA/MS2 fusion proteins were assayed as in panel C. All proteins tested here and in subsequent figures were derived from *X. laevis*, except when noted. The bottom panels indicate protein levels as determined by Western blotting for HA-tagged fusion proteins and actin. **(E)** Schematic of the assays used to determine the nucleotide(s) added by XTUT7. First, RNAs were assayed by RT-PCR (left

half, for panel F). Second, RNAs were assayed by a tail sequencing assay (right half, for panel G). **(F)** RT-PCR analysis for the indicated fusion proteins. All samples were analyzed on the same gel but were separated for clarity. RT primers are indicated. The dT lane indicates the addition of adenosines, the dA lane indicates addition of uridines, etc. **(G)** Seven representative, independently cloned sequences that illustrate the nucleotides added by XTUT7-C are shown.

Figure 3

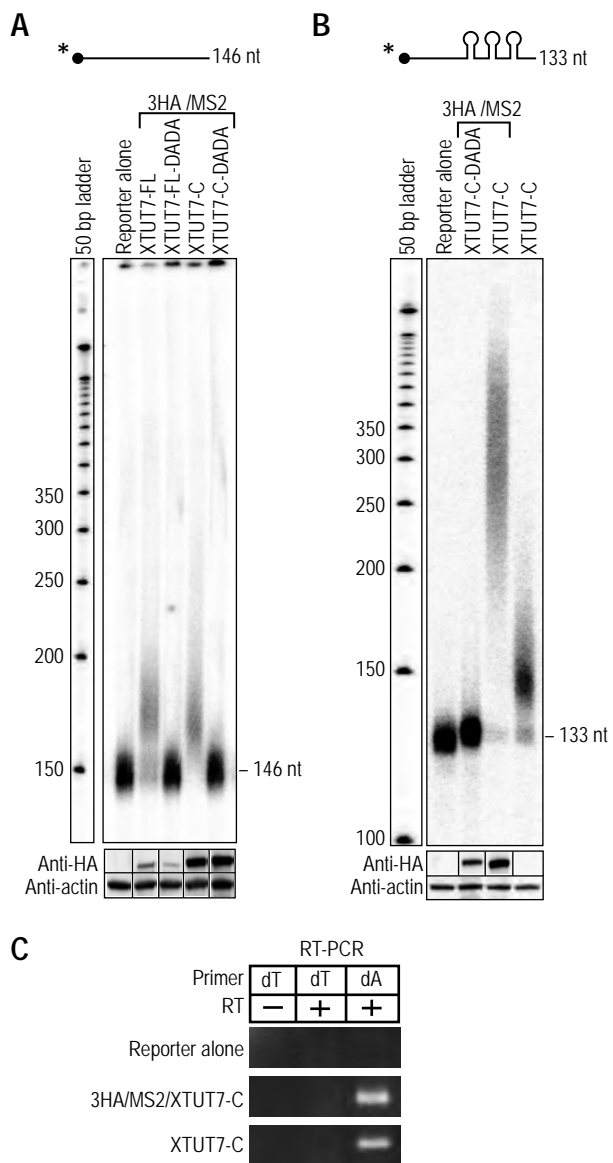


Figure 3. XTUT7 extends RNA independent of MS2 tethering. **(A)** The indicated 3HA/MS2/XTUT7 fusion proteins were assayed on a radiolabeled reporter RNA that lacked MS2 binding sites. The size of the unmodified reporter RNA is 146 nucleotides. Diagrams of the XTUT7 proteins tested are depicted in Figure 2B. The bottom panels indicate protein levels as determined by Western blotting for HA-tagged fusion proteins and actin. **(B)** The indicated XTUT7 proteins were assayed on the radiolabeled reporter RNA with three MS2 binding sites. The size of the unmodified reporter RNA is 133 nucleotides. The last XTUT7 protein tested lacks both the MS2 fusion protein and the 3HA tag. The bottom panels indicate protein levels as determined by Western blotting for HA-tagged fusion proteins and actin. **(C)** RT-PCR assays were performed as in Figure 2F for the indicated proteins. All samples were analyzed on the same gel but were separated for clarity. RT primers are indicated. The dT lane indicates the addition of adenosines and the dA lane indicates addition of uridines.

Figure 4.

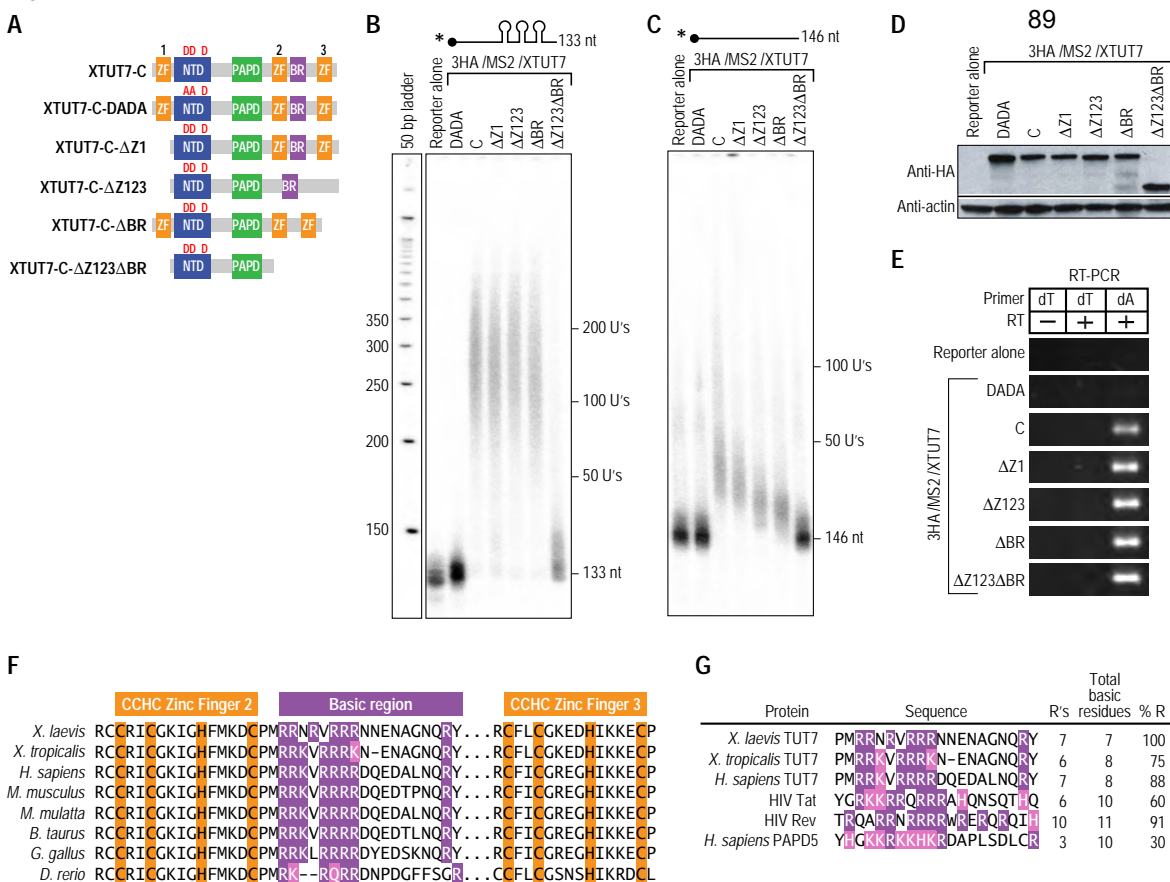


Figure 4. The BR and CCHC zinc finger domains mediate XTUT7's tethering-independent rNTase activity. **(A)** Diagram of XTUT7 mutants tested. CCHC zinc fingers 1, 2, and 3 are indicated **(B&C)**. Indicated 3HA/MS2/XTUT7 fusion proteins were assayed in parallel on radiolabeled reporter RNA with (panel B) or without (panel C) three MS2 binding sites as in Fig. 2C. The sizes of the unmodified reporter RNAs are 133 and 146 nucleotides, respectively. All proteins tested in panels B and C were assayed in parallel and analyzed on the same gel, which was subsequently split into two panels for clarity. **(D)** Protein levels for all samples tested in panels B and C as determined by Western blotting for HA-fusion proteins and actin. **(E)** RT-PCR assays were performed as in Figure 2. All samples were analyzed on the same gel but were separated for clarity. RT primers are indicated. The dT lane indicates the addition of adenosines and the dA lane indicates addition of uridines. **(F)** Multiple sequence alignment of the BR-containing regions from the indicated XTUT7 orthologs. Critical residues for defining CCHC zinc finger domains are highlighted in orange. The BR was defined by the above deletion experiments. Conserved arginines are highlighted in purple. The "..." in the alignments indicates a break in sequence. **(F)** BRs from XTUT7 orthologs were compared to arginine rich motifs (ARMs) from viral RNA-binding proteins and a basic stretch from PAPD5. Arginines (purple); lysines and histidines (pink). The percent arginine composition of each region was also calculated as a comparison.

Figure 5

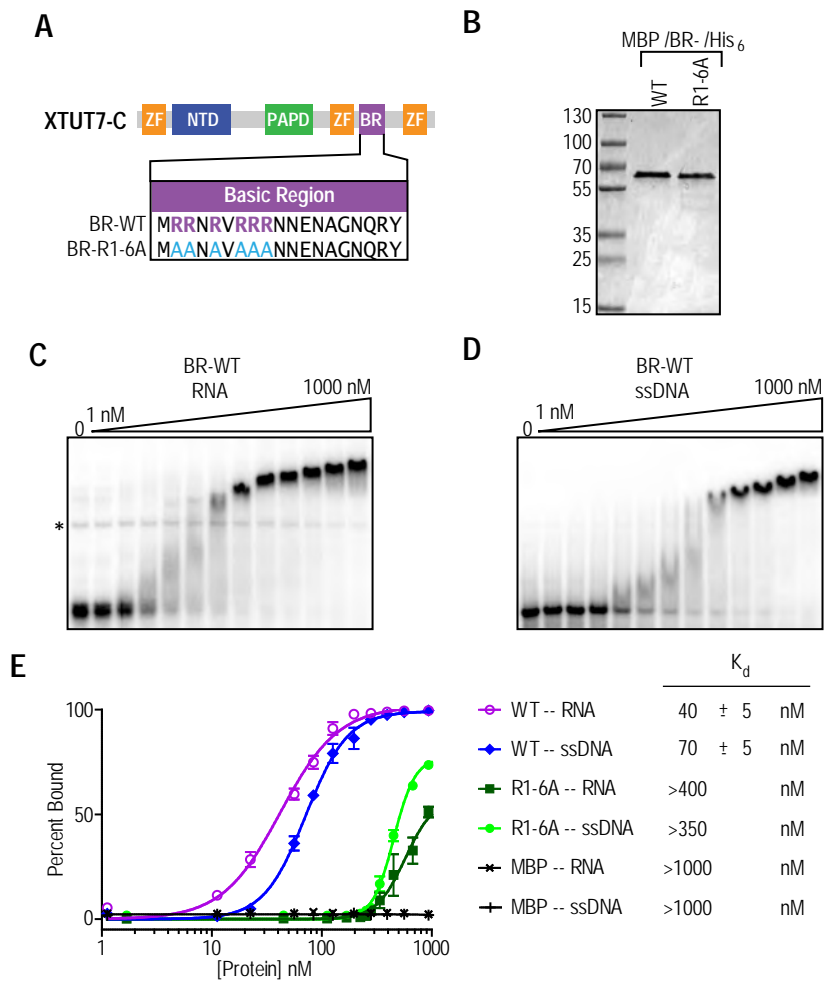


Figure 5. The BR binds nucleic acids *in vitro*. **(A)** Diagram of XTUT7's BR. MBP/BR/His₆ fusion proteins were recombinantly expressed, purified and tested by EMSAs. The R1-6A mutant protein was created by making the indicated amino acid substitutions in the BR. **(B)** Protein gel stained with coomassie blue. MW ladder (kDa) is indicated. The expected MW of the proteins was ~ 57 kDa. **(C)** EMSA using the wild-type BR fusion protein and a radiolabeled RNA containing three MS2 binding sites. Protein concentrations ranged from 1 nM to 1 μ M, as indicated above the gels. The * indicates an RNA artifact that is present even in the absence of proteins. **(D)** EMSA using the wild-type BR fusion protein and a radiolabeled ssDNA of equivalent sequence to the RNA substrate. Protein concentrations ranged from 1 nM to 1 μ M, as indicated above the gels. **(E)** The average percent of nucleic acid substrate bound at each protein concentration was calculated and plotted using non-linear regression analysis from three experiments. The apparent K_d for fusion protein and nucleic acid substrate pairs are indicated to the right of the plot and reported with associated standard error.

Figure 6

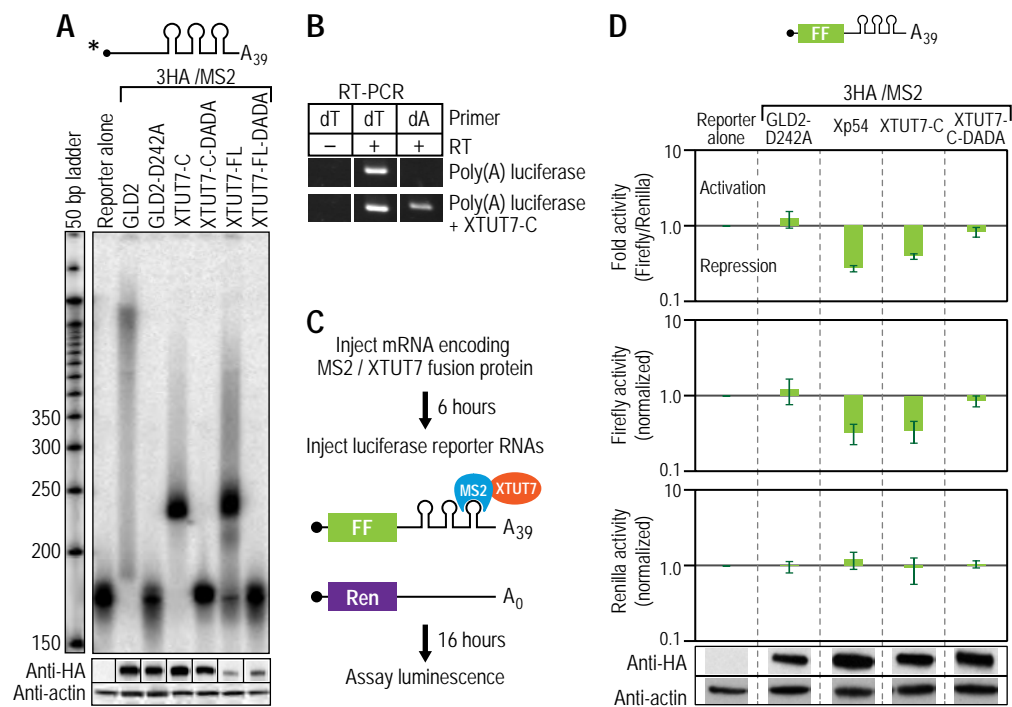
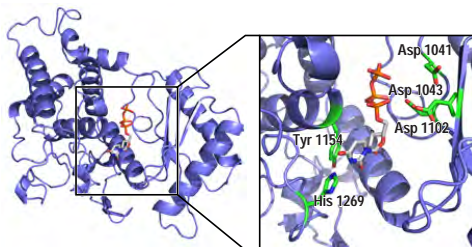
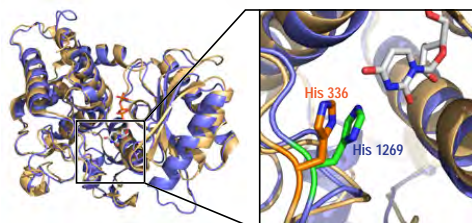
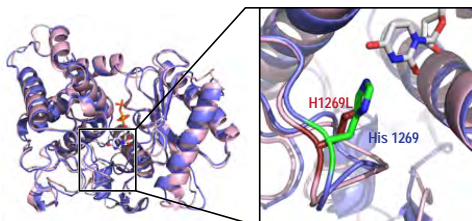


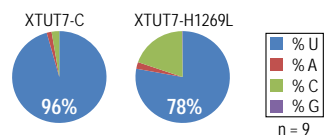
Figure 6. XTUT7 represses a polyadenylated RNA. **(A)** 3HA/MS2 fusion proteins were assayed as in Fig. 2C, except the reporter RNA contained poly(A)₃₉ on the 3' end. The bottom panels indicate protein levels as determined by Western blotting for HA-tagged fusion proteins and actin. **(B)** Samples were assayed by RT-PCR as in Fig. 2F, except the RNA contained poly(A)₃₉ on the 3' end. All samples were analyzed on the same gel but isolated for clarity. **(C)** Schematic of the assay used to determine the effect of XTUT7 on translation. RNAs were injected as in Fig. 2A, except two non-radiolabeled reporter RNAs were co-injected. The first contained the firefly luciferase open reading frame (FF) upstream of three MS2 binding sites and a poly(A)₃₉ tail. The second contained the *Renilla* luciferase open reading frame (Ren) that lacked MS2 binding sites and a poly(A) tail. After 16 hours, luciferase levels were determined. **(D)** Relative luciferase levels in oocytes that express the indicated fusion proteins were determined. Luciferase levels were normalized to reporter alone samples (no fusion protein) in each panel. Error bars represent the standard deviation from three experiments. The bottom and middle panels represent the average relative *Renilla* or firefly luciferase levels, respectively. The top panel represents the average firefly/*Renilla* luciferase levels. Protein levels from a representative experiment are depicted below the luciferase data, and all bands are from the same Western blot but were separated for clarity.

Figure 7

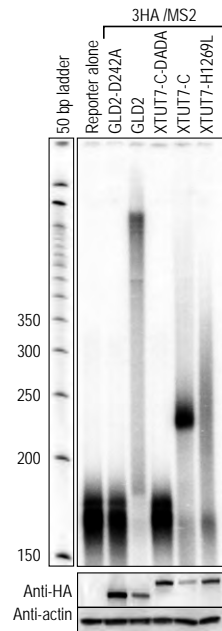
A XTUT7/UTP homology model

B XTUT7/CID1 overlay
RMSD = 1.2 ÅC XTUT7/XTUT7-H1269L homology model overlay
RMSD = 0.96 Å

D
TTTTCTTTTTCTTTTTCTATTCTTTTTATTCTTTCTTTTTCTTTTTCTTCTCTTTTTCT
TTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTT
ATCTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTT
ATCTTTTTTTTTCTTTTTTTTTTTTTTTTTCTTTTTCTTTTTCTTTTTCTCT
ATTTTTATTTTTTTTTCTTTCTTTTTCTTTCT
CTTCTTTCTTTCTTTCTTTTTCTTTCT



E



F

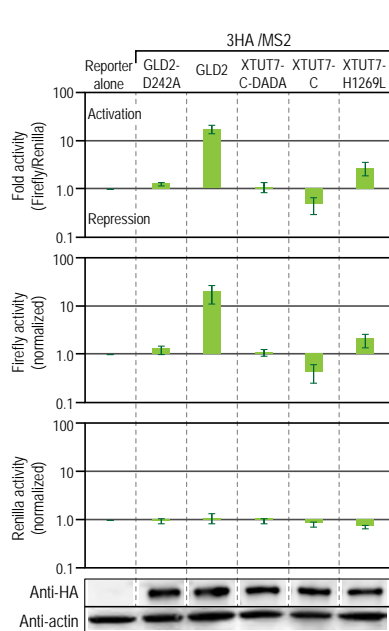


Figure 7. Histidine 1269 is important for XTUT7's uridine specificity and repression activity. **(A)** The predicted structure of XTUT7's core catalytic domains bound to UTP (XTUT7/UTP) as generated by the I-TASSER server (55-57). The XTUT7 homology model had a C-score of 1.02, an expected TM-score of 0.85 ± 0.08 , and an expected RMSD of 4.3 ± 2.9 Å. Asp1041, Asp1043, Asp1102, Tyr1154, and His1269 are highlighted in green. The UTP molecule is gray. **(B)** Alignment of the predicted XTUT7 structure (blue) with a described *S. pombe* CID1 structure (gold/orange, PDB ID 4E8F) (RMSD = 1.2 Å) (30). Analogous histidines and their position relative to UTP are indicated. **(C)** Alignment of the predicted wild-type XTUT7 structure (blue) to the predicted XTUT7-H1269L mutant structure (pink) (RMSD = 0.96 Å). The XTUT7-H1269L homology model had a C-score of 0.99, an expected TM-score of 0.85 ± 0.08 , and an expected RMSD of 4.4 ± 2.9 Å. The position of the histidine and the H1269L substitution relative to UTP are indicated. **(D)** The nucleotides added by XTUT7-H1269L were determined as in Fig. 2G. Six representative, independently cloned sequences that illustrate the nucleotides added by XTUT7-H1269L are shown. The average percent nucleotide composition of nine representative cloned tails added by XTUT7-C and XTUT7-H1269L are indicated. **(E)** Indicated proteins were assayed using the poly(A)₃₉ radiolabeled reporter RNA as in Fig. 6A. The bottom panels indicate protein levels as determined by Western blotting for HA-fusion proteins and actin. **(F)** Relative luciferase levels were determined as in Fig. 6D. Error bars represent the standard deviation from three experiments. The bottom and middle panels represent the average relative *Renilla* or firefly luciferase levels, respectively. The top panel represents the average firefly/*Renilla* luciferase levels. Protein

levels from a representative experiment are depicted below the luciferase data. All bands are from the same Western blot but were separated for clarity.

Chapter 3

Protein-RNA networks revealed through covalent RNA marks

Christopher P. Lapointe¹, Daniel Wilinski^{1,2}, Harriet A. J. Saunders¹, and Marvin
Wickens¹

This chapter is published:

Lapointe CP, Wilinski D, Saunders HAJ, and Wickens M. 2015. Protein-RNA networks revealed through covalent RNA marks. *Nature Methods*. **12**, 1163-1170.
doi:10.1038/nmeth.3651. PMID: 26524240.

¹ Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

² Present address: Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

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

ABSTRACT

Protein-RNA networks are ubiquitous and central in biological control. We present an approach, termed “RNA Tagging,” that identifies protein-RNA interactions *in vivo* by analyzing purified cellular RNA, without protein purification or crosslinking. 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 using high-throughput sequencing. We used this approach to identify hundreds of RNAs bound by a *Saccharomyces cerevisiae* PUF protein, Puf3p. The method revealed that while RNA-binding proteins productively bind specific RNAs to control their function, they also “sample” RNAs without exerting a regulatory effect. We exploited the method to uncover hundreds of new and likely regulated targets for a protein without canonical RNA-binding domains, Bfr1p. The RNA Tagging approach is well-suited to detect and analyze protein-RNA networks *in vivo*.

INTRODUCTION

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

Over the last decade, powerful RNA immunoprecipitation-based approaches have made it possible to identify RNAs bound by a specific protein¹. In RNA immunoprecipitation (RIP), RNA-binding proteins are immunopurified from cell lysates, and associated RNAs are identified by microarray or deep sequencing^{2,3}. UV-crosslinking prior to immunoprecipitation (CLIP) covalently links interacting proteins and RNAs, which facilitates their purification⁴⁻⁷. CLIP also employs a 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 non-physiological interactions *in vitro*⁸⁻¹⁰. In CLIP, UV-crosslinking is relatively inefficient or requires nucleotide analogs to enhance efficiency^{6,11,12}. CLIP also requires numerous enzymatic steps. Moreover, since transient interactions are permanently captured by crosslinking, biologically meaningful interactions are difficult to distinguish from those that are not¹⁰.

We sought a method to identify global protein-RNA interactions *in vivo*, in which interactions were unambiguous and must have occurred inside the cell. The approach we report here, termed “RNA Tagging”, is independent of protein purification, crosslinking, or radioactive-labeling steps. We use the approach to identify RNAs bound by two *Saccharomyces 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 own^{13,14}. 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 using targeted or high-throughput sequencing assays, facilitated by a reverse-transcription step that is selective for uridylated 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”, 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* mRNA^{15,17}. We grew strains that expressed 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 U's to endogenous, wild-type *PHD1* mRNA, a known target²², 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/I nucleotides to serve as a 3' adapter (G/I-tailing)²³ (**Fig. 1b**). Inosines were included to reduce the stability of potential G-quadruplexes²⁴. 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 SPRI beads. DNA libraries were paired-end sequenced on an Illumina HiSeq 2500 instrument.

Tagged RNAs were identified 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 U-tags of various lengths were accurately detected, we sequenced synthetic DNA libraries with known numbers of uridines. The libraries contained the adapter sequences, a poly(A)₁₂ tail, and variable length U-tags (**Supplementary Fig. 2a**). 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 to a single location in the yeast genome (“uniquely mapped”). We detected just over one million Tagged RNAs, which corresponded to approximately 175,000 Tagged RNAs Per Million uniquely mapped reads (“TRPM”). Tagged RNAs had U-tags that ranged from 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**).

As assessed by RNA Tagging, Puf3p bound hundreds of RNAs *in vivo*. Of the RNAs detected with 3' uridines in the PUF3-PUP strain, 476 mRNAs were enriched above background in two biological replicates and were termed “Puf3p targets” (see Online Methods) (**Fig. 2b**). The number of TRPM detected for each Puf3p target was highly reproducible ($\rho = 0.93$, $P = 0$) (**Fig. 2c**). TRPM was moderately correlated with the mean U-tag length ($\rho = 0.5$, $P = 0$) and not correlated with RNA abundance (**Supplementary Fig. 3a,b**). The number of U's in the U-tag was weakly and inversely correlated with RNA abundance ($\rho = -0.37$, $P = 0$) (**Supplementary Fig. 3c**). The set of RNA Tagging targets

significantly overlapped with those identified by RIP-chip¹⁵ and PAR-CLIP²⁵ (hypergeometric tests, all $P < 2.2 \times 10^{-16}$) (**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 Puf3-binding elements in their 3' UTRs (**Fig. 2f**). Importantly, Puf3p tagged approximately 70% (170/246) 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, while the PAR-CLIP targets yielded a more degenerate element (**Fig. 2f** and **Supplementary Fig. 4**).

The above data demonstrate that RNA Tagging globally identifies protein-RNA interactions *in vivo*. The approach reproducibly identified over four hundred 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 affinities 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 employed a two-dimensional ranking of Puf3p targets uniquely enabled by the RNA Tagging approach. Targets have two attributes - the number of Tagged RNAs detected and the number of U's added. Based on these two parameters, we hierarchically clustered Puf3p targets by the number of Tagged RNAs detected at increasing U-tag lengths. Clustering results were visualized by a heat map, with the highest ranked target at the top (**Fig. 3a**). As expected, target rank was strongly correlated to TRPM ($\rho = -0.91$, $P = 0$) and U-tag length ($\rho = -0.75$, $P = 0$) (**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 (see Online Methods). Class A Puf3p targets, which consist 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 vivo*^{16,27}. In contrast, Class C was the lowest ranked group, and these targets had the least 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' UTRs^{15,22,28,29}. 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 Tagged RNAs and the number of U's detected on target RNAs were uncorrelated with the distance from the binding element to the 3' terminus of the transcripts (**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 co-variation of target rank (and hence classes) 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 as well as *PUF3*-dependent stability, while Class C targets would exhibit the least.

Puf3p target classes correlated with localized translation at mitochondria. We mined published data that identified mRNAs¹⁸ and proteins³² 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**). Enrichments 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 profiling³³. 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 while 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 correlated with sensitivity to deletion of *PUF3*. We next 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 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. Importantly, Class C targets were hardly enriched

for the effects of *PUF3* on either their abundance or stability. All specific mRNAs previously shown to be stabilized in a *puf3Δ* strain were Class A or B targets, which independently corroborated our meta-analysis of the global experiments²⁷ (**Supplementary Fig. 9**).

The correlation between Puf3p target classes and known Puf3p biological functions, as well as with 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 they were bound; yet, they were largely unregulated. We refer to this behavior as “sampling.” We define the term “sampling” to mean that the protein bound to RNA sufficiently long to tag it, but insufficiently long to exert its regulatory effect – likely too brief to recruit effector proteins or allow them to act. On average, the RNAs that are sampled are more abundant, which may 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 pathway^{35,36} and is localized to the endoplasmic reticulum (ER) under normal conditions^{37,38} and P-bodies after stress³⁹. Bfr1p was also found associated with over a thousand mRNAs by RIP-chip²⁹. Intriguingly, its reported mRNA targets were not enriched for those with a role in the secretory pathway.

RNA Tagging with BFR1-PUP identified more than a thousand 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 snoRNAs (snR11 and snR31) were detected above background in three biological replicates and were termed “Bfr1p targets” (see Online Methods) (**Fig. 4b**). TRPMs were reproducibly detected across replicates (all pair-wise $\rho \geq 0.84$) (**Fig. 4c**). TRPM, U-tag length, and RNA abundance were all largely uncorrelated (**Supplementary Fig. 10**). Approximately 30% of the targets were previously identified by RIP-chip²⁹, which represents a significant overlap (hypergeometric test, $P < 2.2 \times 10^{-16}$) (**Fig. 4d**). Unlike Puf3p, 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 while RIP-chip targets were at most weakly enriched (**Fig. 4e**). Deeper dissection revealed 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 ribosomal proteins (**Supplementary Fig. 11**). In contrast, mRNAs uniquely identified by RIP-chip were enriched for ribosome biogenesis and the processing of ncRNAs.

Bfr1p binds mRNAs translated at the ER

To more closely examine Bfr1p targets, we performed a two-dimensional analysis with Bfr1p targets as we had with Puf3p. Bfr1p targets were grouped into four classes,

Classes A to D, with Class A again containing the highest ranked targets (**Fig. 5a**). Target rank was strongly correlated with TRPM ($\rho = -0.87$, $P = 0$), while target rank was weakly correlated with the average number of U's in the U-tag and RNA abundance (**Supplementary Fig. 12**). The weak correlation between target rank and the number of U's in the U-tag indicated that in this case, unlike that of Puf3p, 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 secreted⁴⁰, predicted to have a transmembrane domain, and localized to the ER³² (hypergeometric tests, 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 mitochondria³² (**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-bodies⁴¹ (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^{37,38}, its presence on polysomes³⁷, 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 experiment⁴².

Bfr1p targets were highly enriched for abundant, ER-translated mRNAs. In comparison to 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 complex-dependent and SEC complex-independent translocation events (Kolmogorov-Smirnov tests, all $P < 2.2 \times 10^{-16}$) (**Fig. 5g,h**). Class A Bfr1p targets were also most enriched for abundant mRNAs, and the enrichment progressively decreased across classes (**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* (see 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 non-productive binding events *in vivo* since the number of uridines added by the poly(U) polymerase likely is a direct reflection of the time the protein is bound to the RNA. In organisms with endogenous enzymes that add and remove uridines, endogenous uridylated mRNAs are sufficiently stable to be detected⁴³⁻⁴⁶ 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, enrichments missed in earlier RIP-chip studies. Additionally, Bfr1p is part of a large protein complex³⁷ and is required for the localization of mRNAs to P-bodies³⁹ and the bud tip³⁶. 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 were difficult to examine. For example, it may be possible to detect RNAs both directly and indirectly associated with a protein of interest, aided by using a poly(U) polymerase with its own intrinsic but weak RNA-binding activity¹³. 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^{47,48}. 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 valuable 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 in this regard. Regardless, the versatility of RNA Tagging should enable approaches to unexplored problems in RNA biology in living cells.

ONLINE METHODS

Yeast Strains

All *Saccharomyces cerevisiae* strains were constructed in BY4742 yeast (*MAT α* ; *his3 Δ 1*; *leu2 Δ 0*; *lys2 Δ 0*; *ura3 Δ 0*). To construct RNA Tagging chimeras, the DNA sequence for the open reading frame (ORF) of *Caenorhabditis elegans pup-2* followed by a stop codon and the *URA3* marker, including its native promoter and terminator sequences, was inserted in-frame at the 3' end of *PUF3* and *BFR1* using standard yeast transformation techniques. The BFR1-PUP2 strains also contained a 3-HA epitope tag on the C-terminus of the fusion protein. Catalytically inactive PUP2 strains (PUP2mut strains) had Asp185Ala and Asp187Ala substitutions in the PUP-2 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 RGSH₆ epitope tag fused to the C-terminus of Phd1p, and either the wild-type or mutant *PHD1* 3' UTRs, which had substitutions that disrupted known Puf5p-binding elements (UGUAGUUA to ACAAGUUA, and UGUAACAUUA to ACAACAUUA). Cells were selected on 5-FOA containing plates. Integration of the epitope tag and 3' UTRs at the endogenous *PHD1* locus was confirmed by sequencing. The *pup-2* ORF and a 3-HA epitope tag were then inserted in-frame at the 3' end of *PUF5* as above in both the wild-type and mutant *PHD1* strains.

Yeast growth and total RNA Isolation

All strains were grown by inoculating 5 mL YPAD cultures with the indicated frozen yeast strains or freshly streaked colonies, and incubating at 30°C and 180 rpm. After ~

24 hours, 25 mL YPAD cultures were seeded at $A_{660} \sim 0.0002$ and grown at 30°C and 180 rpm until A_{660} 0.5-0.8. Yeast were harvested by centrifugation for 10 minutes at 3,000 rpm at 4°C, and the pellets were washed once with 40 mL of ice-cold water. Cells were resuspended in 500 μ L RNA ISO Buffer (0.2M Tris-HCl pH 7.5, 0.5M NaCl, 0.01M EDTA, 1% SDS). Then, ~ 200 μ L of acid washed beads and 500 μ L of Phenol:Chloroform:Isoamyl alcohol (25:24:1) (PCA) were added. Cells were lysed by vortexing for 20 sec followed by 20 sec on ice ten times. Samples were then separated from the beads, split evenly into two tubes, and 375 μ L of RNA ISO Buffer and 375 μ L of PCA were added to each tube. Samples were mixed by gently shaking and were separated by centrifugation for 15 minutes at 15,000 rpm at 4°C. The aqueous layer was removed (~ 500 μ L) and further extracted by two additional extractions (PCA followed by chloroform). Following 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 hour. Total RNA was pelleted by centrifugation for 30 minutes at 15,000 rpm at 4°C. Pellets were washed 1X with $\sim 70\%$ ethanol, and resuspended in 43 μ L of water. Separate tubes for each sample were then recombined, and treated with 8 Units of TURBO DNase (Life Technologies) for 1 hour at 37°C. Total RNA was purified using the GeneJet RNA Purification kit (Thermo Fisher Scientific) and eluted in 30 μ L of water. RNA samples were stored at -80°C.

Targeted RNA Tagging RT-PCR assays

Terminator treatment: To deplete rRNA, 2 μ g of total RNA were treated with 2 Units of Terminator enzyme (Epicentre) for 60 minutes at 30°C. The reactions were subsequently

purified using 1.8 volumes of room temperature RNA Clean XP beads (Agencourt) and the standard protocol. rRNA-depleted RNA was eluted in 12 μL of water.

G/I-tailing: Terminator-treated samples were G/I-tailed by using 1,200 Units of yeast poly(A) polymerase (PAP) (Affymetrix), 0.5 mM GTP, 0.15 mM ITP, and incubated at 37°C for 90 minutes. Samples were diluted to 100 μL with water and G/I-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 hour. Samples were pelleted by centrifugation for 30 minutes at 15,000 rpm at 4°C. Pellets were washed once in ~70% ethanol, and resuspended in 10 μL of water.

Selective reverse transcription: G/I-tailed samples were selectively reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) under nearly standard conditions. The G/I-tailed samples were split equally (typically 3 μL) across all RT reactions. 3 μL of samples were added to 1 μL of 1 μM U-select primer (GCCTTGGCACCCGAGAATTCCACCCCCCCCCAAA), 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 μM oligo-(dT)₄₂ (TT) in place of the U-select primer. A master mix of 4 μL of 5X reaction buffer, 1 μL of 100 mM DTT and 1 μL of 40 U per μL RNase Inhibitor per reaction was prepared separately. The primer-RNA mixes and the master mix were incubated at 65°C for 5 minutes followed by 5 minutes 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), 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 minutes followed by 5 minutes 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: GACAGCATCCGGGTTGTATT. The *HSP10* specific reverse primer was: TTTTCCTGTCATACATAATGGCC. *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: ACAAGAACAAGAAAACCACGC and AAGATGCATGTATCCCGCTC. 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 sec, 3) 50°C for 30 sec, 4) 72°C for 90 sec, 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 using 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 minute G/I-tailing reaction, an additional 1,200 Units of yeast PAP was added to the reactions and incubated for an additional 30 minutes 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 minutes 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: 60 μL of cDNA was added to 10 μL of 10X Klenow Buffer (500 mM Tris-HCl pH 7.5, 100 mM MgCl_2 , 10 mM DTT, 0.5 mg per mL BSA), 12 μL of water, 5 μL of 10 mM dNTPs, 10 μL of 10 μM 2nd strand synthesis primer (GTTTCAGAGTTCTACAGTCCGACGATCNNNNN), and 3 μL of 5 U per μL Exo- Klenow

DNA Polymerase (Life Technologies). Reactions were incubated at 37°C for 30 minutes, 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 GoTaq polymerase (Promega). 5 µL of cDNA was added to 8.33 µL of 2X GoTaqGreen master mix, 2 µL of water, 0.67 µL of 10 µM RP1 primer (AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA), and 0.67 µL of 10 µM barcoded primer (CAAGCAGAAGACGGCATAACGAGATXXXXXXGTGACTGGAGTTCCTTGGCACCCGA GAATTCCA). Standard Illumina barcodes were inserted at the XXXXXX position in the primer. The PCR cycle was: 1) 94°C for 2 min, 2) 94°C for 10 sec, 3) 40°C for 2 min, 4) 72°C for 1 min, 5) Repeat 2-4 once, 6) 94°C for 10 sec, 7) 55°C for 30 sec 8) 72°C for 1 min, 9) Repeat 6-8 7X, 10) 94°C for 15 sec, 11) 55°C for 30 sec, 12) 72°C for 1 min, 13) Repeat 10-12 14 times, 14) 72°C for 5 min, 15) Hold at 4°C. To scale up, ~9 individual reactions were completed for each sample and were pooled prior to cleanup. PCR samples were size-selected twice using the RNA Clean XP beads at a 0.8:1 (bead:reaction) ratio. Samples were eluted in ~ 20 µL of water.

Synthetic U-tag libraries

Preparation: Synthesized oligos were purchased (Integrated DNA Technologies) and their sequences were:
 CCTTGGCACCCGAGAATTCCACCCCCCCCCAAA(A)TTTTTTTTTTTTTGATCGTCGGA

CTGTAGAACTCTGAAC. At the **(A)** 0,2,4,6,8,10, and 12 adenosines were inserted to create various length U-tag standards. Synthetic libraries were amplified using GoTaq polymerase, 0.8 μ M RP1 oligo, 0.8 μ M barcoded primer, and 2 nM of oligo as template. Otherwise, the PCR conditions were the same as for the transcriptome-wide library preparations. Completed reactions were run on a 6% acrylamide TBE-Urea gel, and the bands corresponding to the libraries were excised from the gel (~180 bps). The gel slices were crushed in 200 μ L of water, flash frozen, incubated at 37°C and 1,000 rpm for 1 hour, flash frozen again, incubated at 37°C and 1,000 rpm for 1 hour, and separated using a filter column. The libraries were ethanol precipitated with GlycoBlue as the co-precipitant. Pellets were washed once in 70% ethanol and resuspended in 10-20 μ L of water.

Analysis: Raw FASTQ files of the sequenced libraries were analyzed two different ways. First, the number of uridines in the U-tag of every read in each of the libraries was determined. Using this data, the mean U-tag length (in nucleotides) and associated standard deviation was calculated for each of the synthetic libraries. Second, the base composition at each position for every read in the libraries was determined. These calculations were then used to determine how often a single uridine residue was detected in the A0 library (no U-tag encoded in the synthesized oligo), which served as the background rate referenced below.

High-throughput sequencing

Samples were sequenced on an Illumina HiSeq 2500 instrument to obtain 50 base pair paired-end read data sets. Throughout, the first sequencing read, which covers the 5' end of the sequenced DNA fragment, is termed “Read 1”, and the second sequencing read, which covers the 3' end of the sequenced DNA fragment, is termed “Read 2”. Raw data was deposited at the NCBI Sequence Read Archive (accession: SRP063022).

FASTQ file manipulations and alignments

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 (TTTGGGGGGGGGTGGAATTCTCGGGTGCCAAGG) and the poly(A) tail sequence (AAAAAAAAA) were removed from Read 1's using FASTA/Q Clipper [*fastx_clipper -a sequence -l 15 -n -l -v input -o output -Q 34*]. Any Read 1's that were shorter than 15 nucleotides after removal of either sequence were discarded. Reads 1's were then aligned to the *S. cerevisiae* genome (version R64-1-1) using bowtie⁴⁹ with the following parameters: a seed length (-l) of 25 nucleotides, no more than 2 mismatches (-n), and only a single reportable alignment (-m) in the genome [*bowtie -t genome input output -l 25 -m 1 -S --sam -p 3 -n 2*]. Reads that aligned to more than one location were discarded.

Read 2: The 5' adapter sequence (GATCGTCGGACTGTAGAACTCTGAAC) was removed from Read 2's using FASTA/Q Clipper and the same parameters as above. The last six nucleotides of the resulting Read 2's, which represent the random hexamer sequence from the 2nd strand synthesis step, were then removed using FASTA/Q

Trimmer [*fastx_trimmer -t 6 -i input -o output -Q 34*]. The resulting Read 2's 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 3 adenosines followed any number of uridines at their 3' end (A-U tail sequences) were identified using regular expression searches in Perl. Read 2's 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's 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 nucleotides and the *-v* alignment mode was used to exclude reads with 3 or more mismatches.

Definition of Tagged RNAs

A Tagged RNA was defined as a DNA fragment with sequence that aligned uniquely to the yeast genome and contained at least 8 adenosines followed by at least 1 uridine at their 3' end that were not encoded by any adapter sequence or the genome. Typically, Read 1 identified the genomic location of a Tagged RNA while 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, Tagged RNAs per million uniquely mapped reads). Where indicated, TRPM enrichment was calculated as a ratio of TRPMs obtained in strains with and without the relevant RBP-PUP chimera.

Reproducible RNA Tagging targets

In order to be identified as a target, genes with Tagged RNAs had to pass three criteria. First, the number of TRPMs detected for a particular gene must be at least 10-fold greater than the number of TRPMs detected for that gene in the non-tagging control sample. Second, the number of TRPMs detected for a particular gene must 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**) (see above for synthetic libraries). Thus, the error rate was defined as the number of TRPM detected by error per gene $[0.03 * (\text{total \# of TRPM}) / (\text{total \# of genes with TRPM})]$. Third, a gene must have passed both of the above criteria in all of the biological replicates. See **Supplementary Data 3** and **4** for comprehensive target lists of Puf3p and Bfr1p, respectively. See **Supplementary Data 5** for sequencing results of a control strain (BY4742) without any tagging chimeras.

Hierarchical clustering

Tagged RNAs per million uniquely mapped reads (TRPM) for each target were calculated across U-tag lengths of 1-10 uridines for each sample. TRPMs for biological replicates were then averaged (mean). Each U-tag length encompassed all TRPMs with at least the indicated number of uridines. Prior to clustering, the data was sorted from most to least TRPMs detected with at least 1 U in the U-tag. The data sets were \log_2 -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 if it was distinct from directly adjacent putative classes. As an example, the enrichment of putative Class A targets for a given observation (e.g. RNAs with increased abundance in $\Delta PUF3$) was compared to the enrichment in putative Class B targets. If the enrichments of putative Class A and 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 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 R^2 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 (ρ) and Pearson's (r) correlation coefficients and their associated P -values were determined using the `rcorr` 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 if the observed overlap between two datasets was significant.

The total population size was defined as 6,607 genes, except for the following analyses: mRNA localization to mitochondria (6,256 genes), proteins with predicted transmembrane domains (TMHMM analyses, 6,713 genes), and yeast GFP protein 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(oneway_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 was randomized 100,000 times using the `sample` function.

Venn diagrams

Proportional Venn diagrams were generated using Biovenn⁵⁰ and then redrawn for publication.

MEME and directed motif searches

To be as inclusive as possible, 3' UTRs were defined as the longest isoform for a particular gene previously observed⁵¹ or, if not previously defined, as 200 bases. MEME²⁶ 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, the binding elements present in each class, as determined using all of the Puf3p targets, were combined to generate the indicated motifs in Figure 3a. Unbiased MEME analyses were conducted as 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 identified²⁵ but shortened here for consistency. In all cases, motifs were prepared for publication using WebLogo 3⁵². The total number of genes with the C[AUC]UGUA[AUC]AUA 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.

PBE location 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 8 adenosines (with or without a U-tag) were included. Using this information, the most detected isoform for particular mRNAs, the lengths of the 3' UTRs, and the position of the PBE relative to the stop codon and 3' termini were determined. Genes with undetected 3' termini, and genes with negative or very large (>1,000 nucleotides) distances to 3' termini were excluded from the analyses. For Supplementary Figure 6d&e, the mean number of Tagged RNAs, number of U's 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 6 U's were not analyzed since our deep sequencing did not yield 3' termini for those mRNAs.

GO analyses

All GO analyses were completed using Yeast Mine from the Saccharomyces Genome Database (yeastmine.yeastgenome.org). All parameters were set to default (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 (TMD), the sequences of all proteins (6,713 proteins, including dubious proteins) were downloaded from the Saccharomyces Genome Database. The sequences were then analyzed using the TMHMM 2.0 server⁵³. Proteins with at least 1 predicted TMD were counted as positives.

RNA-seq

RNA isolation: Total RNA was isolated from *S. cerevisiae* (BY4742) cells using standard methods. 50 ml of cells with A_{660} 0.5-0.8 were collected by centrifugation at 3,200 rpm at 4°C, washed once with cold water, and snap frozen in liquid N₂. The tubes were vortexed for 30 seconds then incubated on ice for 30 seconds, which was repeated six times. The supernatant was removed, extracted with 1 mL of PCA, and ethanol precipitated. RNA pellets were resuspended in 50 µL water.

Library preparations: 2 µg of RNA were used as input. Samples were depleted 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 water. Libraries were sequenced on an Illumina HiSeq 2000 to get 50 base pair 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 (see **Supplementary Data 6**).

Accession code for sequencing data

NCBI Sequence Read Archive accession: SRP063022.

ACKNOWLEDGEMENTS

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) and P. Anderson (University of Wisconsin-Madison) and their labs, and the 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. Vanderploeg 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 (GM50942), and by Wharton and Biochemistry Scholar Fellowships to C.P.L.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests. C.P.L. and M.W. have filed a patent that encompasses the RNA Tagging approach.

Figure 1

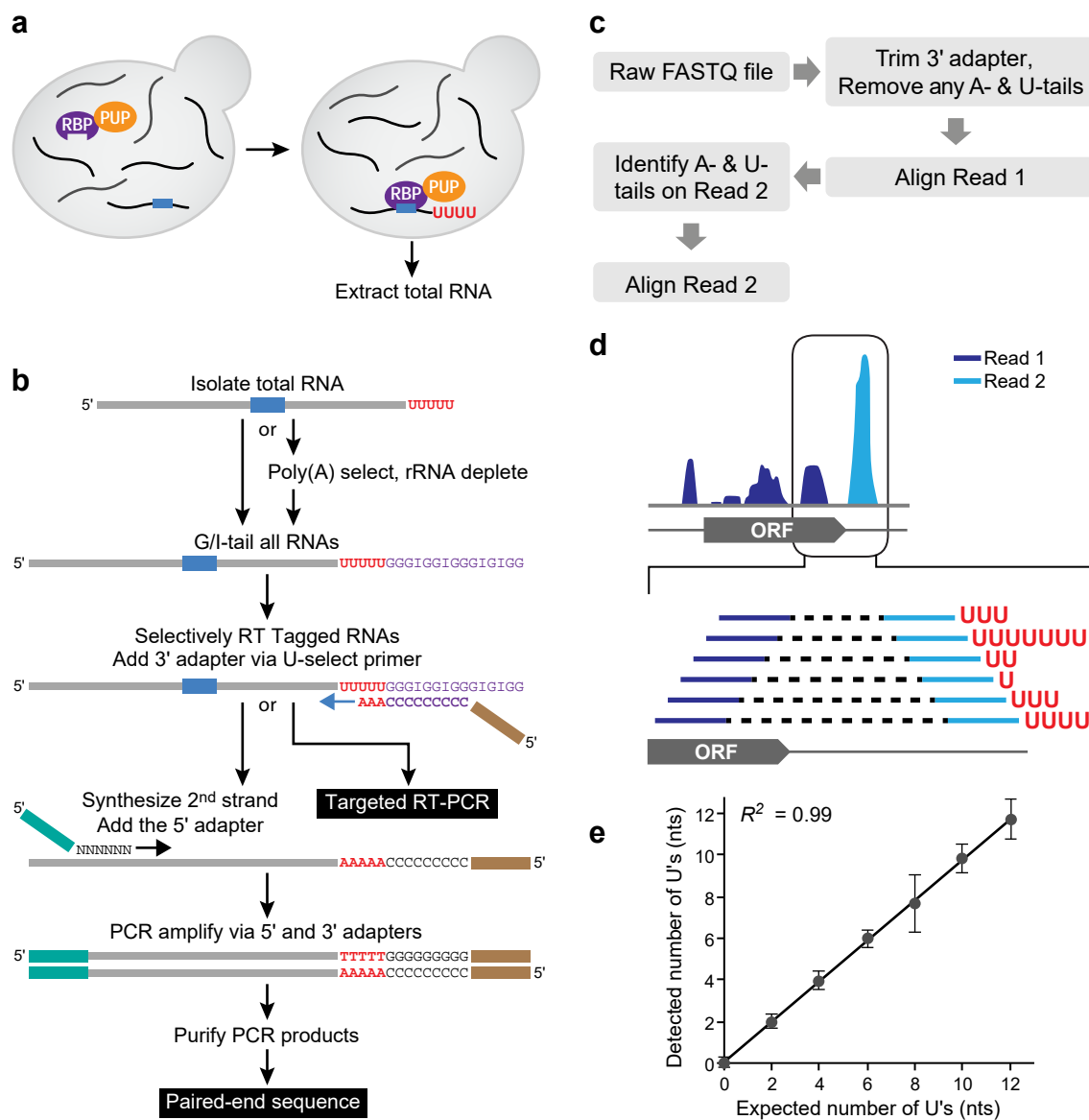


Figure 1. The RNA Tagging approach. **a)** Strategy. RBP, RNA-binding protein. PUP, poly(U) polymerase. **b)** Schematic of targeted RT-PCR and transcriptome-wide RNA Tagging assays. RNAs are tailed with a combination of guanosines (G) and inosines (I) (purple). The U-select primer contained the Illumina 3' adapter sequence (brown), nine cytosines (purple) that base pair with the G/I tail, and three adenosines (red) that select for uridines at the 3' end of the mRNA. **c)** Computational identification of Tagged RNAs. A-tails refers to the poly(A) tail and U-tails refers to 3' terminal uridines, which were often in the U-tag. **d)** Schematic showing Tagged RNAs aligned to a representative gene. ORF, open reading frame. **e)** Plot of the mean U-tag length detected by high-throughput sequencing of synthetic DNA libraries that contained U-tags of 0, 2, 4, 6, 8, 10, and 12 nucleotides. At least 50,000 reads were detected for each library (>1 million total reads). The R^2 value ($R^2 = 0.99$, $n = 7$) was determined by linear regression analysis, and error bars represent standard deviation.

Figure 2

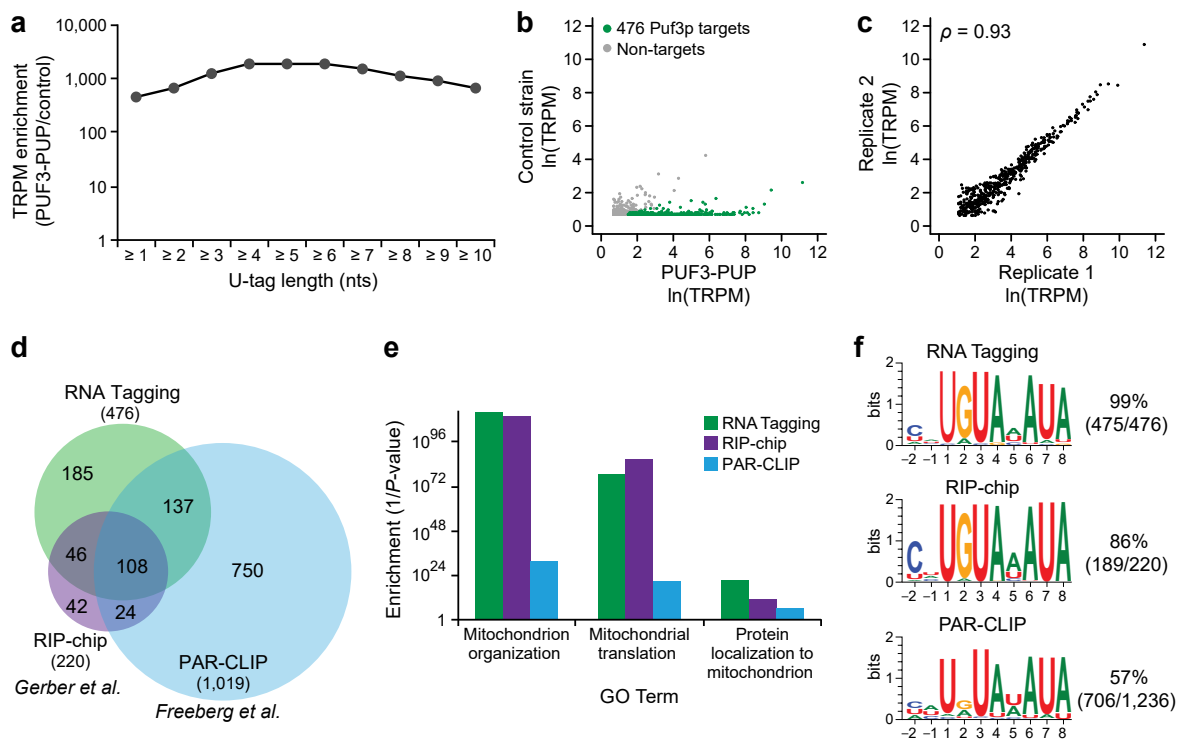


Figure 2. RNA Tagging identified transcriptome-wide Puf3p targets. **a)** Enrichment of Tagged RNAs detected across different U-tag lengths in PUF3-PUP yeast relative to a control yeast strain (BY4742). TRPM, Tagged RNAs per million uniquely mapped reads. **b)** Scatter plot of Tagged RNAs detected in the PUF3-PUP strain relative to the control strain (BY4742). Puf3p target mRNAs (see Online Methods) are colored green; non-targets are grey. **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 ($\rho = 0.93$, $P = 0$, $n = 476$). **d)** Proportional Venn diagram depicting the overlap between Puf3p targets identified by RNA Tagging versus those identified by other approaches^{15,25}. **e)** Plot of selected Go Term enrichments ($1/P$ -value) of Puf3p targets identified by RNA Tagging, RIP-chip¹⁵, and PAR-CLIP²⁵. For simplicity, only 3 biological process terms are shown (see **Supplementary Data 1** for all enriched terms). **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. The numbers indicate the fraction of 3' UTRs in each set that contributed to the motif.

Figure 3

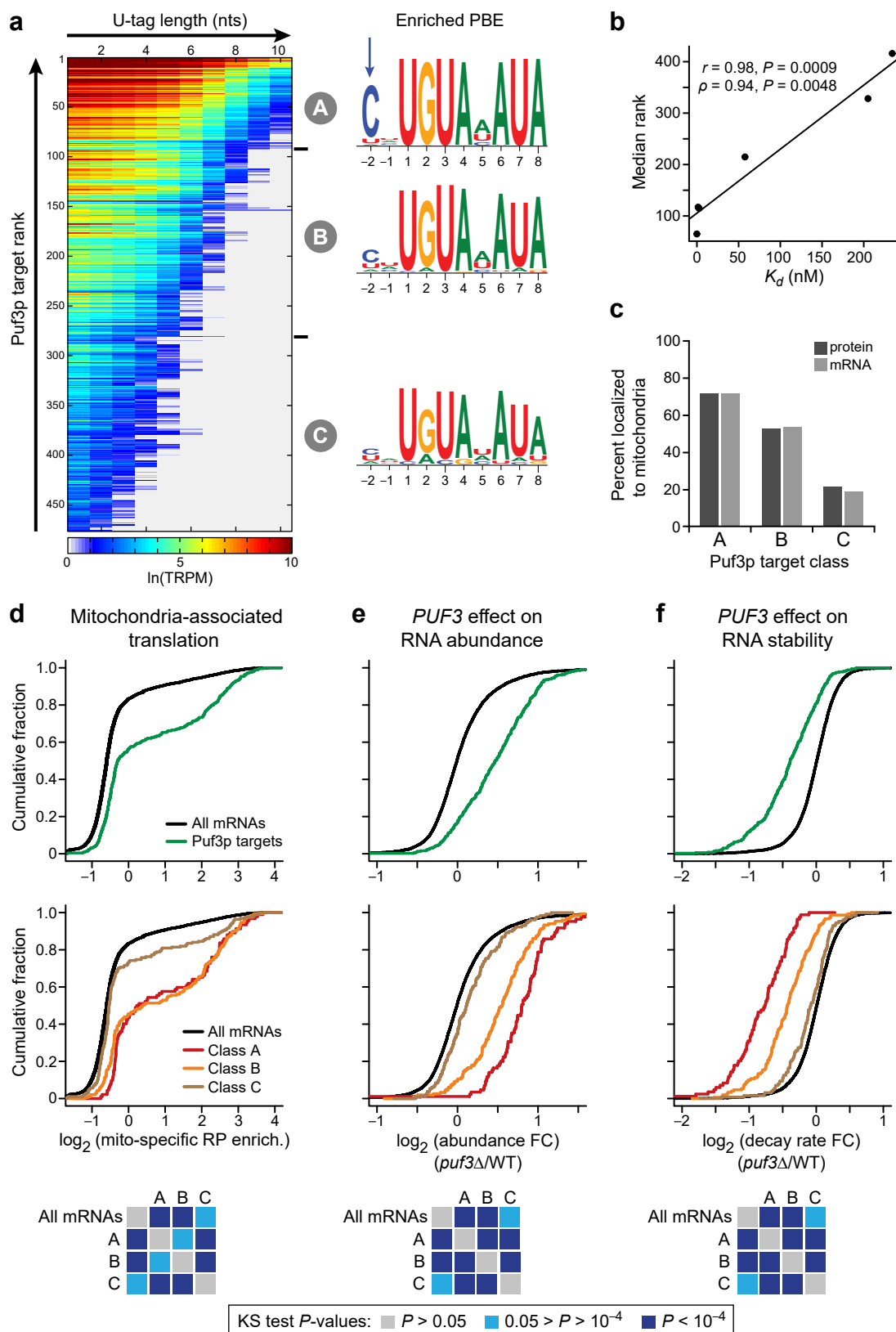


Figure 3. Puf3p target classes correlated with *in vitro* binding affinity and *in vivo* regulation. **a)** Heat map of clustered Puf3p targets, with Classes A (92 targets), B (189), and C (195) indicated. Each row in the heat map is an individual Puf3p target, and the colors indicate the number of TRPM detected with U-tags of at least the indicated number of uridines (columns). The highest ranked target is at the top of the heat map, and the lowest ranked target is at the bottom. The binding elements enriched in each of the Puf3p target classes are indicated. TRPM, Tagged RNAs per million uniquely mapped reads. PBE, Puf3p-binding element. **b)** Plot of the median rank of Puf3p targets that contain six distinct binding elements relative to the published *in vitro* binding affinity (K_d) of purified Puf3p for the same sequences¹⁶. Pearson's (r) and Spearman's (ρ) correlation coefficients and associated P -values (P) are indicated ($r = 0.98$, $P = 0.0009$; $\rho = 0.94$, $P = 0.0048$; $n = 6$). **c)** Enrichment of Puf3p target classes for mRNAs and proteins localized to mitochondria. Mitochondria-localized mRNAs and proteins were obtained from published experiments^{18,32}. **d-f)** Empirical cumulative distributions were plotted for all Puf3p targets (top) and the three Puf3p target classes (middle) relative to all mRNAs for the following attributes: enrichment for mRNAs bound by ribosomes at mitochondria³³ (all mRNAs, $n = 6,094$; Class A, $n = 92$; Class B, $n = 189$; Class C, $n = 194$) (**d**), as well as change in mRNA abundance³⁴ (all mRNAs, $n = 4,305$; Class A, $n = 85$; Class B, $n = 151$; Class C, $n = 130$) (**e**) and stability³⁴ (all mRNAs, $n = 4,228$; Class A, $n = 84$; Class B, $n = 150$; Class C, $n = 128$) (**f**) in *puf3* Δ relative to wild-type. The P -values from Kolmogorov-Smirnov (KS) tests comparing the different distributions are indicated (bottom).

Figure 4

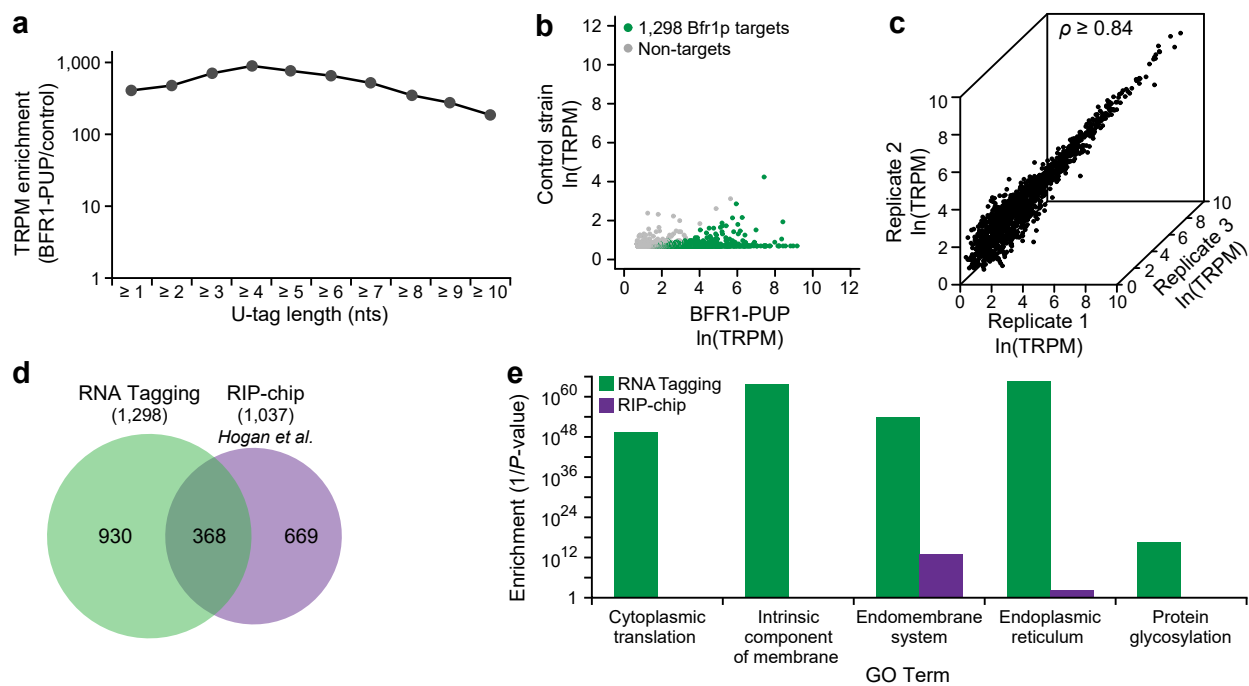


Figure 4. RNA Tagging identified transcriptome-wide Bfr1p targets. **a)** Enrichment of Tagged RNAs detected across different length U-tags in BFR1-PUP yeast relative to a control yeast strain (BY4742). TRPM, Tagged RNAs per million uniquely mapped reads. **b)** Tagged RNAs detected in the BFR1-PUP strain relative to the control strain (BY4742). Bfr1p target mRNAs (see Online Methods) are colored green while non-targets are grey. **c)** The number of Tagged RNAs detected for the 1,298 Bfr1p targets in three biological replicates. Spearman's correlation coefficient (ρ) is indicated (all pair-wise $\rho \geq 0.84$, $P = 0$, $n = 1,298$). **d)** Proportional Venn diagram depicting the overlap between Bfr1p targets identified by RNA Tagging versus published RIP-chip targets²⁹. **e)** Selected Go Term enrichments ($1/P$ -value) of Bfr1p targets identified by RNA Tagging and RIP-chip (see **Supplementary Data 2** for all enriched terms).

Figure 5

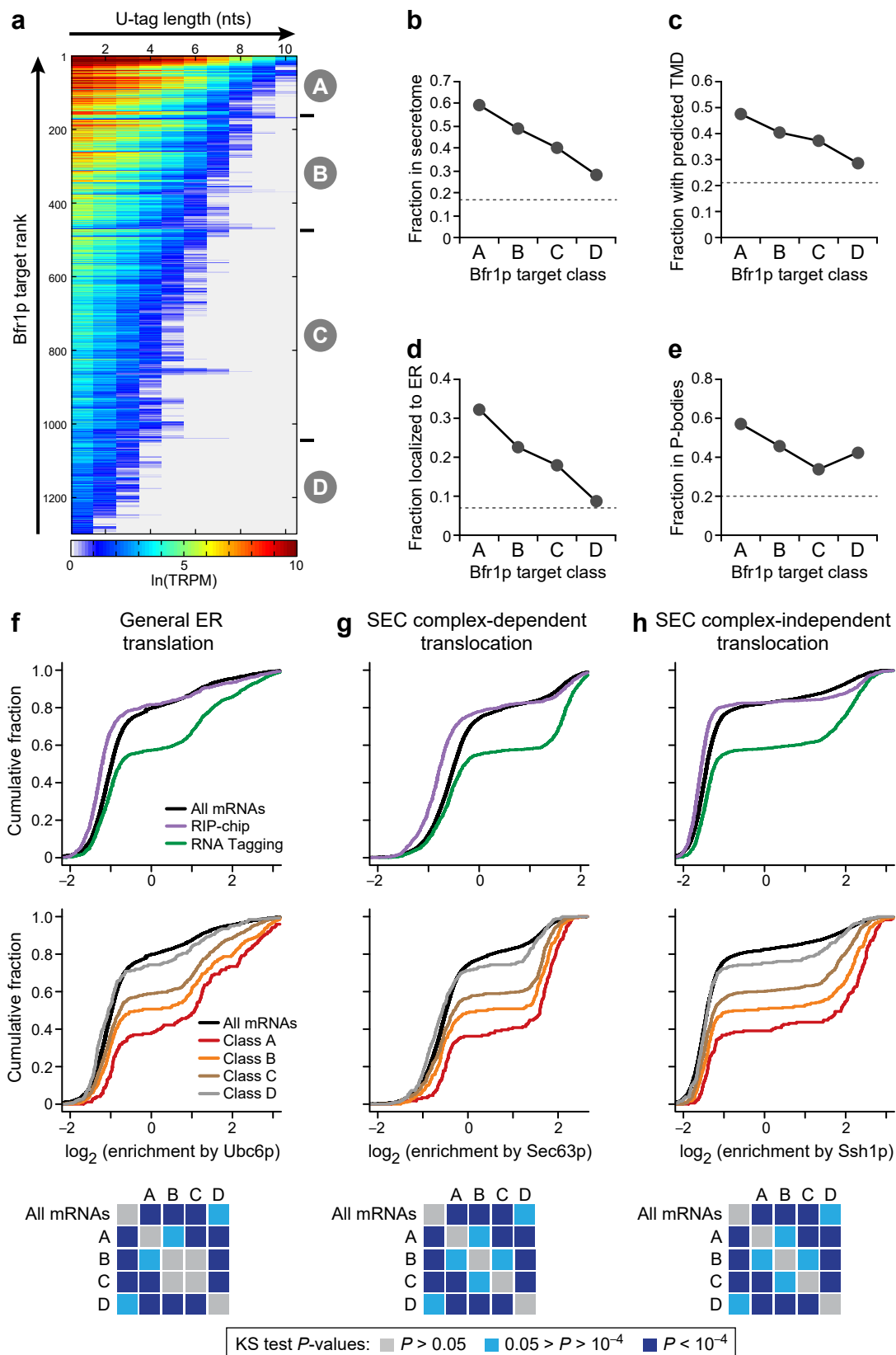
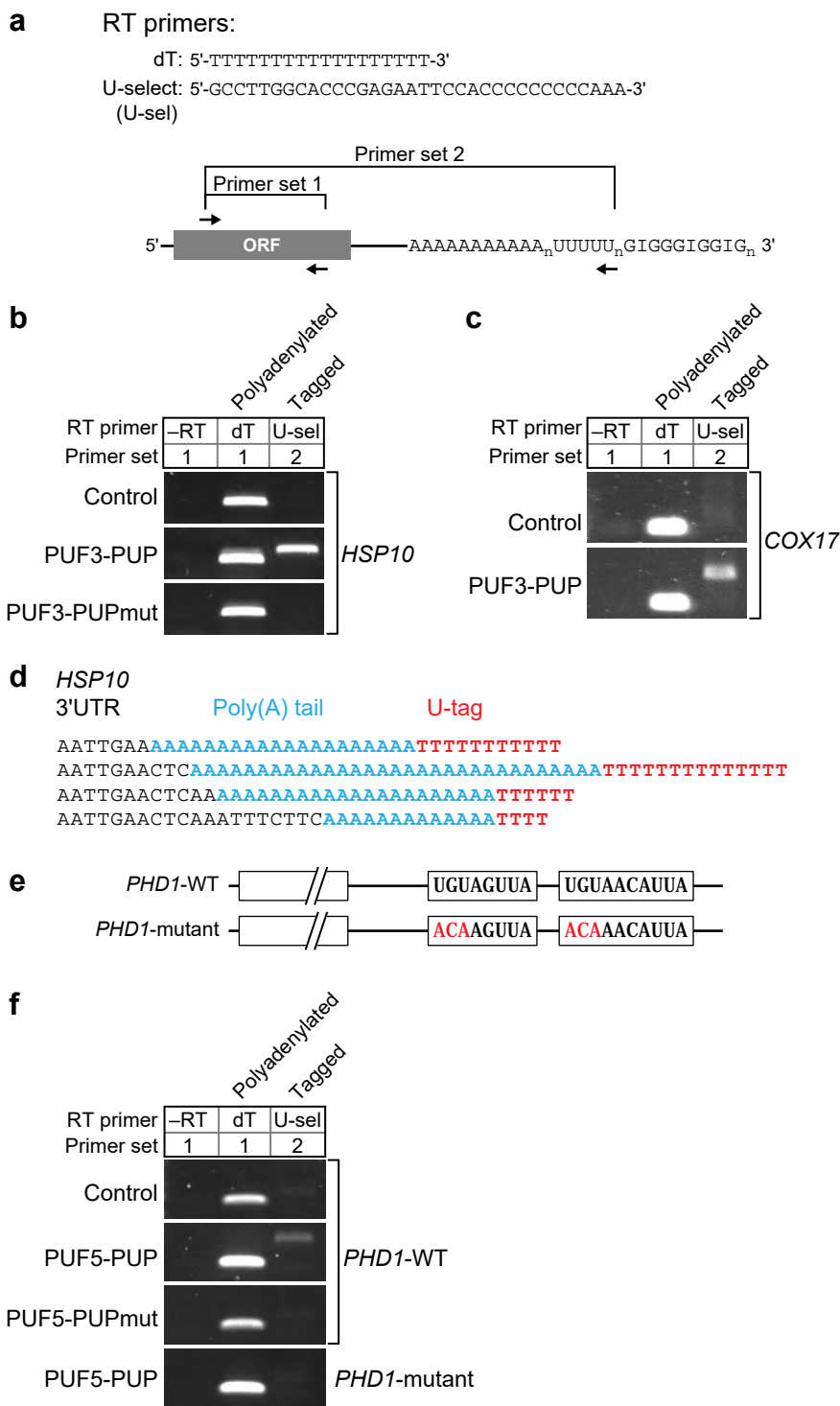


Figure 5. Bfr1p target classes correlated with membrane functions. **a)** Heat map of clustered Bfr1p targets, with Classes A (174 targets), B (297), C (566), and D (261) indicated. Each row in the heat map is an individual Bfr1p target, and the colors indicate the number of TRPM detected with U-tags of at least the indicated number of uridines (columns). The highest ranked target is at the top of the heat map, and the lowest ranked target is at the bottom. TRPM, Tagged RNAs per million uniquely mapped reads. **(b-e)** Enrichments of Bfr1p target classes for mRNAs encoding proteins found in the secretome⁴⁰ **(b)**, with predicted transmembrane domains (TMD) **(c)**, localized to the endoplasmic reticulum³² (ER) **(d)**, and mRNAs found in P-bodies⁴¹ **(e)**. The grey, dotted line represents the enrichment of all mRNAs for the given attribute. **(f-h)** Empirical cumulative distributions were plotted for the indicated target sets (top) and the four Bfr1p target classes (middle) relative to all mRNAs for the following attributes: enrichment for mRNAs bound by ribosomes generally at the ER (all mRNAs, $n = 5,935$; Class A, $n = 173$; Class B, $n = 296$; Class C, $n = 561$; Class D, $n = 261$) **(f)**, at the SEC complex (all mRNAs, $n = 5,974$; Class A, $n = 174$; Class B, $n = 297$; Class C, $n = 560$; Class D, $n = 261$) **(g)**, and at the SSH1 translocon complex (all mRNAs, $n = 5,785$; Class A, $n = 174$; Class B, $n = 297$; Class C, $n = 561$; Class D, $n = 260$) **(h)**, obtained from published ER-specific ribosome profiling (RP) experiments⁴². The P -values from Kolmogorov-Smirnov (KS) tests comparing the different distributions are indicated (bottom).

Supplementary Figure 1



Supplementary Figure 1. RNA Tagging identified *in vivo* protein-RNA interactions. **a)** Schematic of the RT-PCR assay for targeted RNA Tagging. The RT primers and PCR primer sets used in panels b and c are shown. PCR primer set 1 was two gene-specific primers and primer set 2 used a gene-specific forward primer and the U-select RT primer as the reverse PCR primer. **(b,c)** PUF3-PUP tagged *HSP10* **(b)** and *COX17* **(c)** mRNAs. In each panel, gel slices were run on the same gel and were separated here for clarity. RT and PCR primers used in each column are indicated. “-RT” lanes (no reverse transcriptase) monitored genomic DNA contamination, which was minimal. “dT” lanes used the oligo(dT) primer, and illustrate that polyadenylated mRNA was present in all samples. “U-sel” lanes used the U-select primer, which detects RNAs with U-tags. The control strain (BY4742) lacked an RNA Tagging chimera. PUF3-PUP is the active RNA Tagging chimera and PUF3-PUPmut is a catalytically inactive chimera, which harbors active site mutations in the PUP (Asp185Ala, Asp187Ala). **d)** Representative Sanger sequencing results of tagged *HSP10* mRNA. The PCR product from the U-select (U-sel) lane of the PUF3-PUP sample in panel b was cloned and individual colonies were sequenced. Black text indicates genomically encoded *HSP10* 3' UTR sequence, bold blue text indicates non-genomically encoded adenosines (the poly(A) tail), and bold red text indicates non-genomically encoded thymidines, which represent the 3' U-tag added by PUF3-PUP. **e)** *PHD1* mutant alleles. The two PUF-binding elements in *PHD1* mRNA were disrupted *via* UGU to ACA substitutions in the endogenous *PHD1* locus. Active or inactive (DD185/187AA) versions of PUP-2 were fused to the endogenous copy of PUF5 (PUF5-PUP and PUF5-PUPmutant, respectively) in the wild-type and mutant *PHD1* strains. **f)**

PUF5-PUP requires its binding elements to tag *PHD1* mRNA. Lanes are as in panels b and c.

Supplementary Figure 2

a

U0: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAA*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

U2: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAATT*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

U4: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAATT*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

U6: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAATT*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

U8: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAATT*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

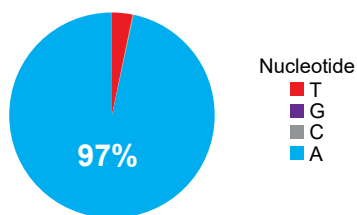
U10: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAATT*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

U12: 5'-GTTTCAGAGTTCTACAGTCCGACGATCAAAAAAAAAATT*TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG-3'

Key	Illumina 5' adapter:	GTTTCAGAGTTCTACAGTCCGACGATC
	Poly(A) tail mimic:	AAAAAAAAAAAA
	U-tag mimic:	TTTTTTTTTTTT
	U-select RT primer:	TTTGGGGGGGGTGAATTCTCGGGTGCCAAGG

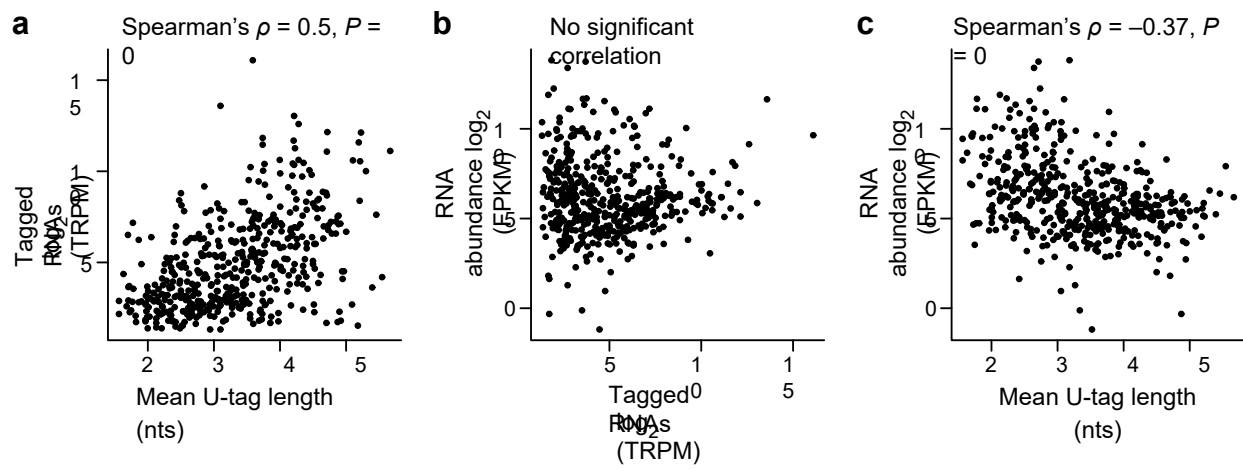
b

Percent composition of the starred (*) position in panel (a)



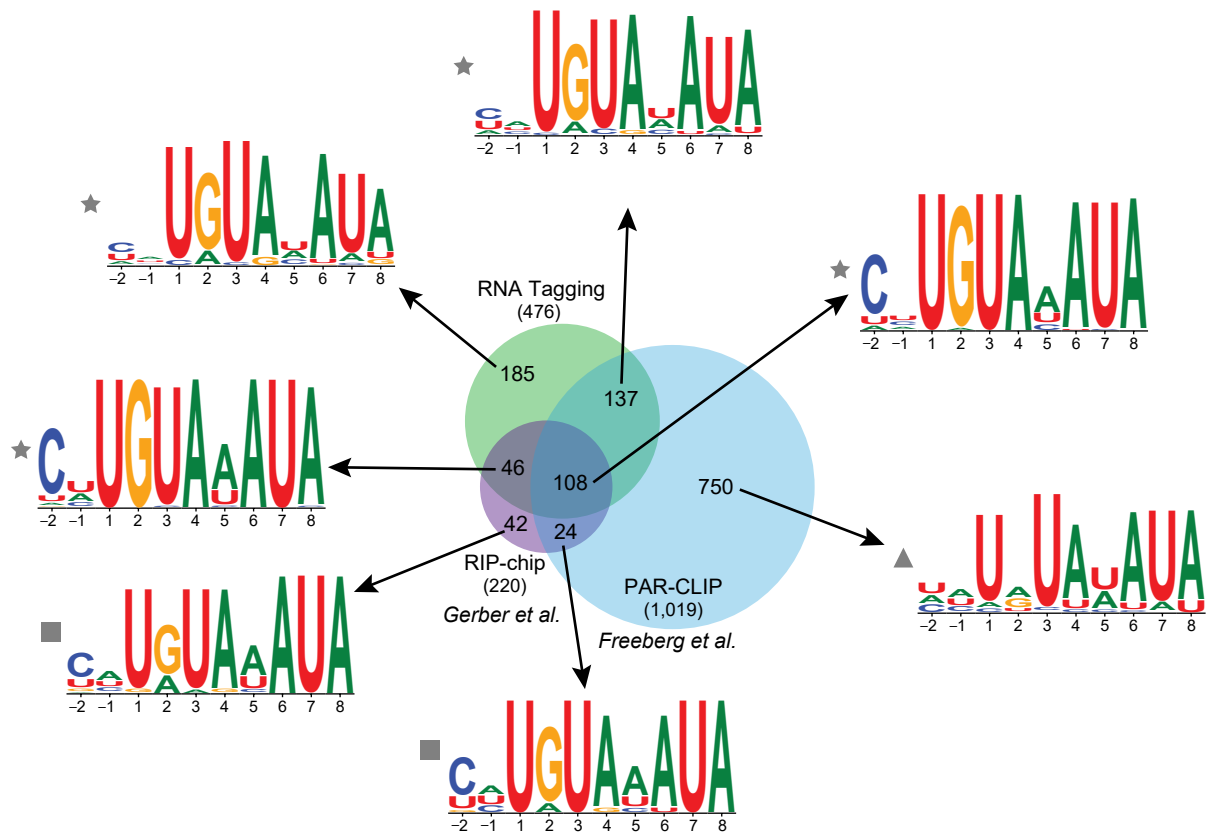
Supplementary Figure 2. Illumina sequencing accurately detected U-tags of multiple lengths. **a)** Synthetic libraries with various length U-tags, shown here as the reverse complement for clarity. The indicated libraries were paired-end sequenced on an Illumina HiSeq2500. The purple sequence represents the Illumina 5' adapter, the blue sequence represents a poly(A) tail of 12 nucleotides, the red sequence represents U-tags of multiple lengths, and the black sequence represents the U-select RT primer. The starred (*) position in the U0 library was further analyzed in panel b. **b)** Accuracy of identifying Tagged RNAs by a single non-templated uridine. The percent nucleotide composition of position 13 in Read 2 of the U0 library, which corresponds to the starred (*) position in panel a, was calculated and plotted ($n = 310,745$). The actual bases detected by sequencing were reverse complemented here for clarity.

Supplementary Figure 3



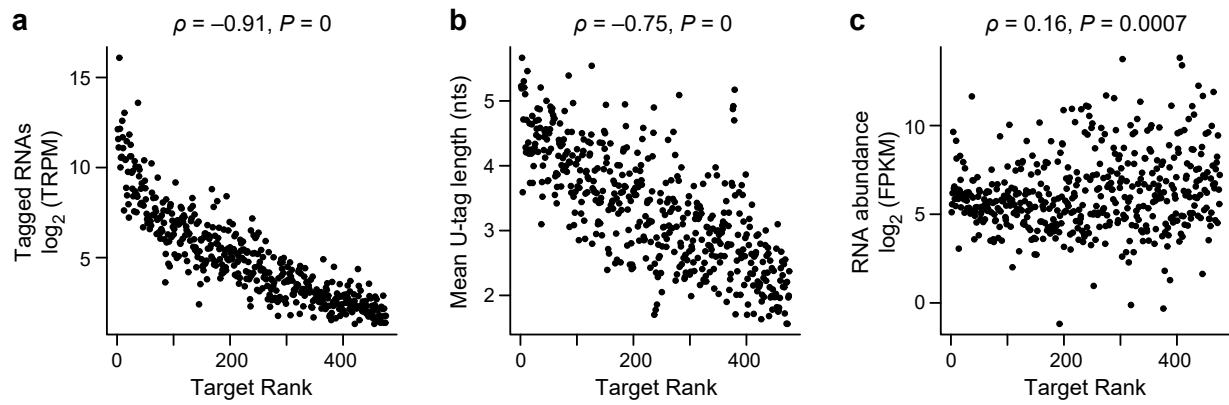
Supplementary Figure 3. Comparison of Puf3p RNA Tagging results and RNA abundance. **a)** The mean number of Tagged RNAs detected for Puf3p targets was correlated with the mean length of their U-tag ($\rho = 0.5$, $P = 0$, $n = 476$). Spearman's correlation coefficient (ρ) and associated P -value (P) are indicated. TRPM, Tagged RNAs Per Million uniquely mapped reads. **b)** The mean number of Tagged RNAs (TRPM) detected for Puf3p targets was uncorrelated with their mean abundance (Spearman correlation, $P > 0.1$). FPKM, fragments per kilobase of exon per million reads mapped. **c)** The mean length of the U-tag on Puf3p targets was weakly correlated with their mean abundance (FPKM) ($\rho = -0.37$, $P = 0$, $n = 476$). Spearman's correlation coefficient (ρ) and associated P -value (P) are indicated.

Supplementary Figure 4



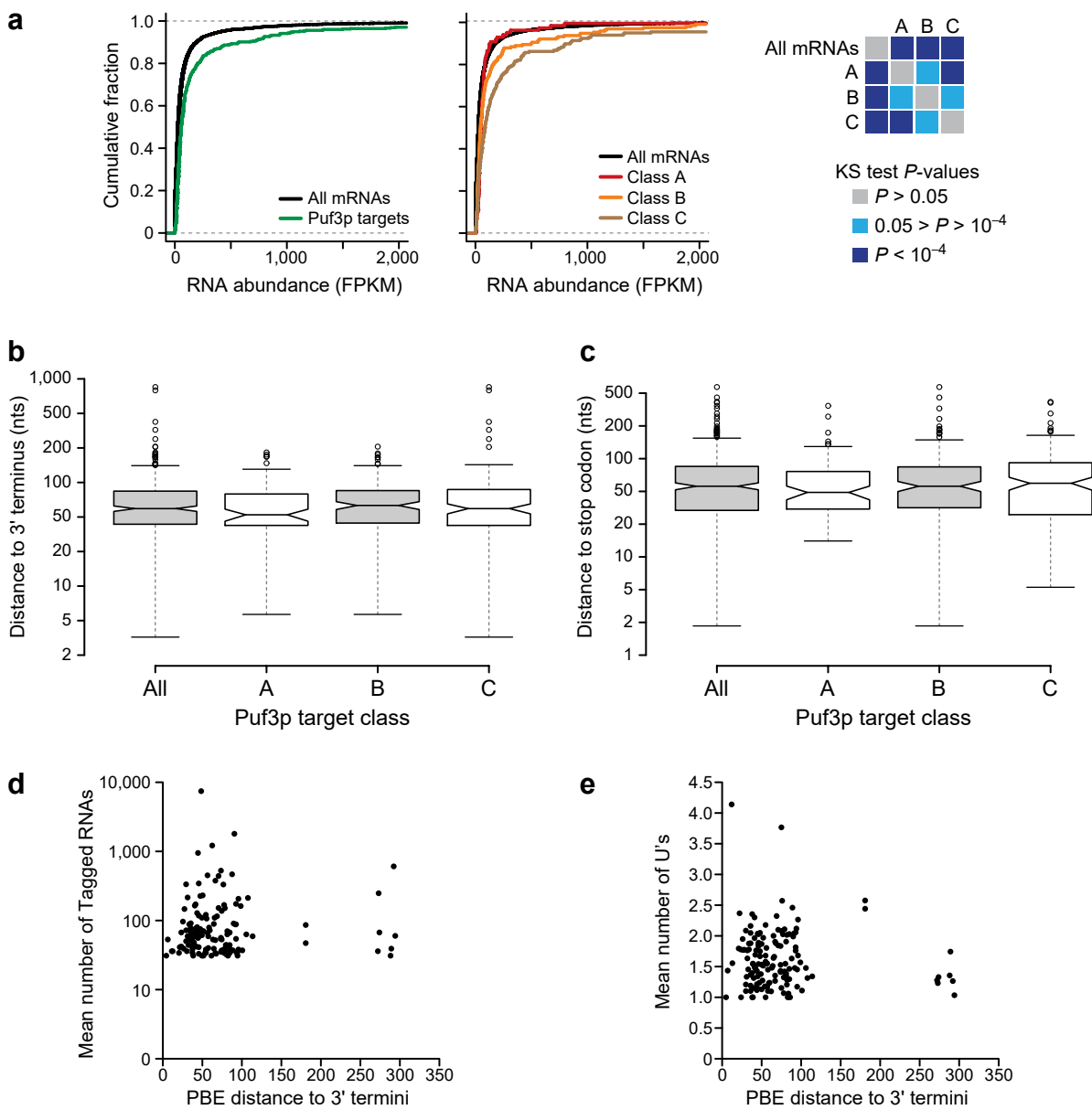
Supplementary Figure 4. Puf3p-binding element comparison between multiple methods. Proportional Venn diagram of Puf3p targets identified using RNA Tagging, RIP-chip¹⁵, and PAR-CLIP²⁵. The numbers indicate the number targets in each area of the plot. Position-weight matrices (plotted in bits) of the Puf3p-binding elements (PBEs) found in each group of targets are indicated. PBEs were derived as follows: PBEs with grey stars, MEME analysis of all RNA Tagging targets; PBEs with grey squares, MEME analysis of all RIP-chip targets; PBEs with grey triangles, PBEs in PAR-CLIP peaks.

Supplementary Figure 5



Supplementary Figure 5. Puf3p target rank was correlated with TRPM and U-tag length, but was largely uncorrelated with RNA abundance. **a)** The mean number of Tagged RNAs (TRPM) detected for Puf3p targets was correlated with their RNA Tagging rank ($\rho = -0.91$, $P = 0$, $n = 476$). Spearman's correlation coefficients (ρ) and associated P -values (P) are indicated in all panels. TRPM, Tagged RNAs Per Million uniquely mapped reads. **b)** The RNA Tagging rank of Puf3p targets was correlated with the mean length of their U-tags ($\rho = -0.75$, $P = 0$, $n = 476$). **c)** RNA Tagging rank of Puf3p targets was largely uncorrelated with their mean RNA abundance ($\rho = 0.16$, $P = 0.0007$, $n = 476$). FPKM, fragments per kilobase of exon per million reads mapped.

Supplementary Figure 6



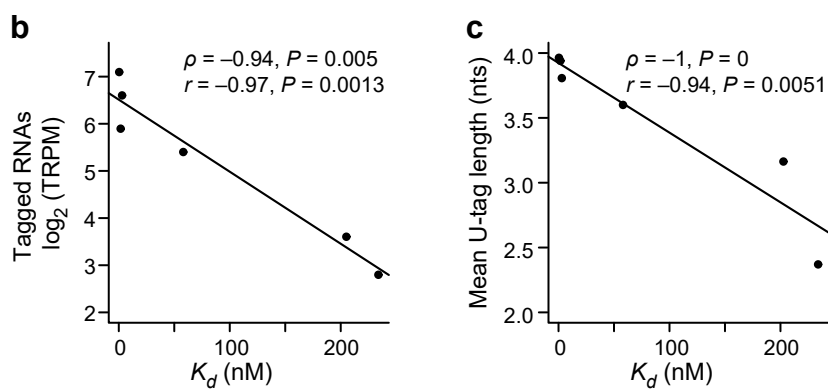
Supplementary Figure 6. Comparison of RNA abundance and the position of binding elements across Puf3p targets. **a)** Class C targets were the most abundant Puf3p targets. Empirical cumulative distributions of RNA abundance were plotted for all Puf3p targets (left) and the three Puf3p target classes (middle) relative to all mRNAs (all mRNAs, $n = 6,595$; Class A, $n = 92$; Class B, $n = 189$; Class C, $n = 195$). The P -values from Kolmogorov-Smirnov (KS) tests comparing the different distributions are indicated (right). **b,c)** Puf3p-binding elements were similarly positioned in the 3' UTRs of each class of Puf3p targets. The distance from each binding element to the 3' terminus (**b**) and the stop codon (**c**) of the target was calculated and plotted (all targets, $n = 404$; Class A, $n = 90$; Class B, $n = 169$; Class C, $n = 145$) (Tukey whiskers indicated). There were no statistical differences between any of the groups (Fisher-Pitman permutation tests, $P > 0.1$). **d,e)** The mean number of Tagged RNAs (**d**) and number of U's (**e**) detected for targets were compared to the distance from the PBE to the 3' terminus for isoforms of 64 Puf3p targets (143 distinct mRNAs) detected by at least 31 reads (24,417 reads total). No significant correlations were observed (Pearson and Spearman correlations, $P > 0.1$).

Supplementary Figure 7

a

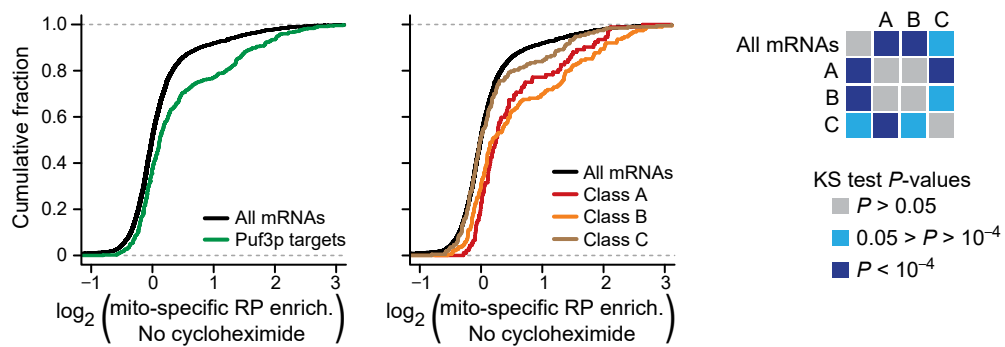
Puf3p binding elements	K_d (nM)	K_{rel}
CCUGUAAAUA	0.31	1
CCUGUAUUAUA	1.6	5
CUUGUAUUAUA	3	10
ACUGUAAAUA	58	187
UUUGUAUUAUA	205	661
AUUGUAUUAUA	234	755

Determined in:
Zhu *et al.* *PNAS* (2009) A 5' cytosine binding pocket in Puf3p
specifies regulation of mitochondrial mRNAs.



Supplementary Figure 7. The number of Tagged RNAs and U-tag length was correlated with *in vitro* binding affinity. **a)** Published *in vitro* binding affinity data of purified Puf3p for the six indicated RNA sequences was obtained and shown here¹⁶. **b,c)** The median number of Tagged RNAs detected (TRPM) ($r = -0.97$, $P = 0.0013$; $\rho = -0.94$, $P = 0.005$; $n = 6$) **(b)** and median U-tag length ($r = -0.94$, $P = 0.0051$; $\rho = -1$, $P = 0$; $n = 6$) **(c)** of Puf3p targets containing six distinct binding elements was calculated and compared to the published *in vitro* binding affinity (K_d) of purified Puf3p for those sequences. Pearson's (r) and Spearman's (ρ) correlation coefficients and the associated P -values (P) are indicated. TRPM, Tagged RNAs Per Million uniquely mapped reads.

Supplementary Figure 8



Supplementary Figure 8. Puf3p targets were enriched for mRNAs translated at mitochondria in the absence of cycloheximide. Published mitochondria-specific ribosome profiling (RP) data in the absence of cycloheximide was mined³³. Empirical cumulative distributions were plotted for all Puf3p targets (left) and the Puf3p target classes (middle) relative to all mRNAs (all mRNAs, $n = 5,609$; Class A, $n = 92$; Class B, $n = 188$; Class C, $n = 193$). The P -values from Kolmogorov-Smirnov (KS) tests that compared the different distributions are indicated (right).

Supplementary Figure 9

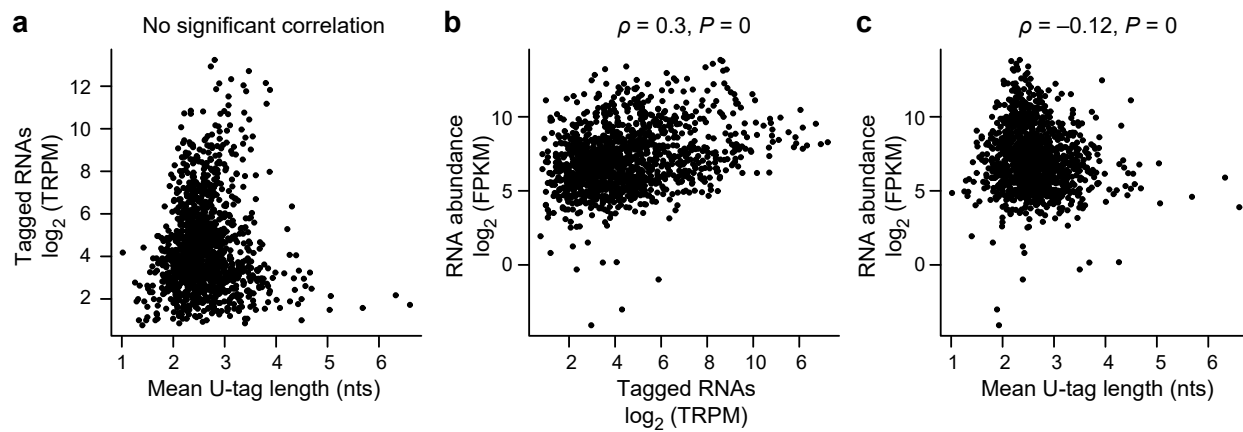
mRNA tested	WT Half-life	<i>puf3</i> Δ Half-life	Fold Change (<i>puf3</i> Δ / WT)	Puf3p Target Class	Puf3p Target Rank	PBE
YLL009C	4.0	24.3	6.1	A	5	CUUGUAUAUA
YML009C	2.9	6.7	2.3	A	11	CCUGUAAAUA
YKL087C	1.7	4.6	2.7	A	60	CCUGUAAAUA
YNR037C	4.2	8.5	2.0	A	12	CAUGUAAAUA
YOR187W	5.5	16.7	3.0	A	25	CGUGUAAAUA
YNR017W	3.1	8.0	2.6	A	47	CUUGUAUAUA
YDR347W	3.2	5.1	1.6	B	132	UCUGUAAAUA
YNL315C	3.8	10.2	2.7	B	183	CCUGUAAAUA
YHR147C	3.0	4.8	1.6	B	131	CUUGUAAAUA
YOR158W	1.7	7.2	4.2	B	125	CAUGUAUAUA
YBL090W	3.1	4.7	1.5	B	202	UUUGUAAAUA
YDR041W	3.2	3.0	0.9	C	362	CUUGUAAAUA
YDL069C	<2.0	<2.0	1.0	NA	NA	NA

** All half-lives taken from Miller, *et al. NAR* (2013)

** "NA" indicates the gene was not identified as a Puf3p target using RNA Tagging.

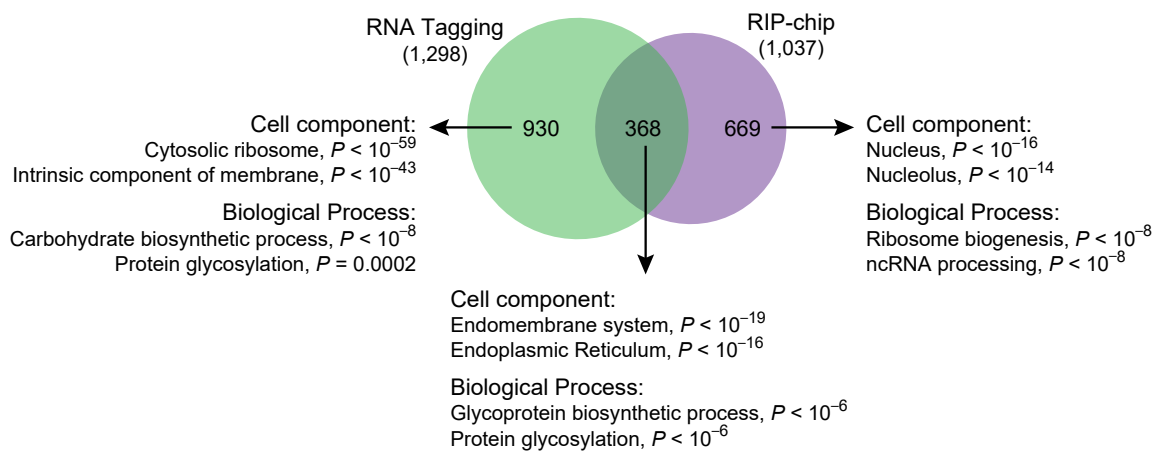
Supplementary Figure 9. mRNAs with known *PUF3*-dependent half-lives were Class A or B targets. Summary of published RNA half-lives of the indicated genes in wild-type and *puf3* Δ strains²⁷. Puf3p target class, RNA Tagging rank, and Puf3p-binding elements are indicated. “NA” indicates the gene was not identified as a Puf3p target by RNA Tagging.

Supplementary Figure 10



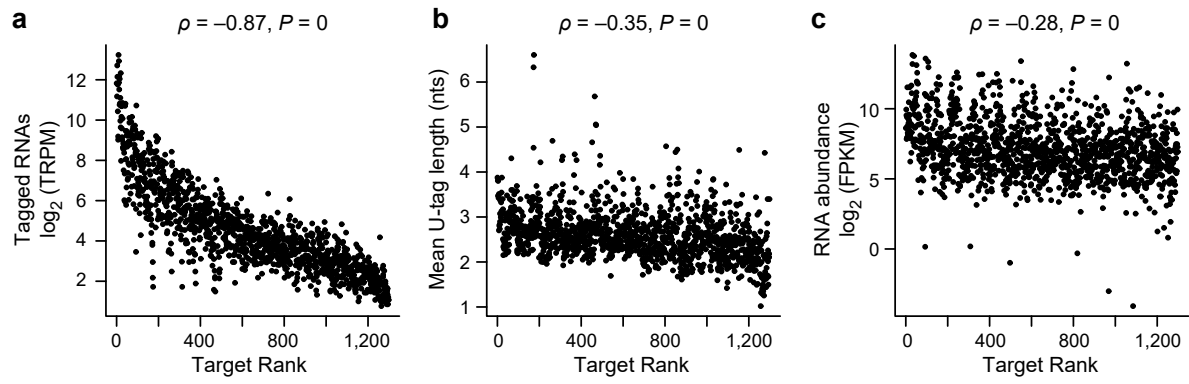
Supplementary Figure 10. Comparison of Bfr1p RNA Tagging results and RNA abundance. **a)** The mean number of Tagged RNAs (TRPM) detected and the mean length of their U-tag were uncorrelated (Spearman correlation, $P > 0.1$). TRPM, Tagged RNAs Per Million uniquely mapped reads. **b)** The mean number of Tagged RNAs (TRPM) detected for Bfr1p targets was weakly correlated with their mean abundance ($\rho = 0.3$, $P = 0$; $n = 1,298$). FPKM, fragments per kilobase of exon per million reads mapped. Spearman's correlation coefficient (ρ) and associated P -value (P) are indicated. **c)** The mean length of the U-tag on Bfr1p targets was largely uncorrelated with their mean abundance ($\rho = -0.12$, $P = 0$; $n = 1,298$). Spearman's correlation coefficient (ρ) and associated P -value (P) are indicated.

Supplementary Figure 11



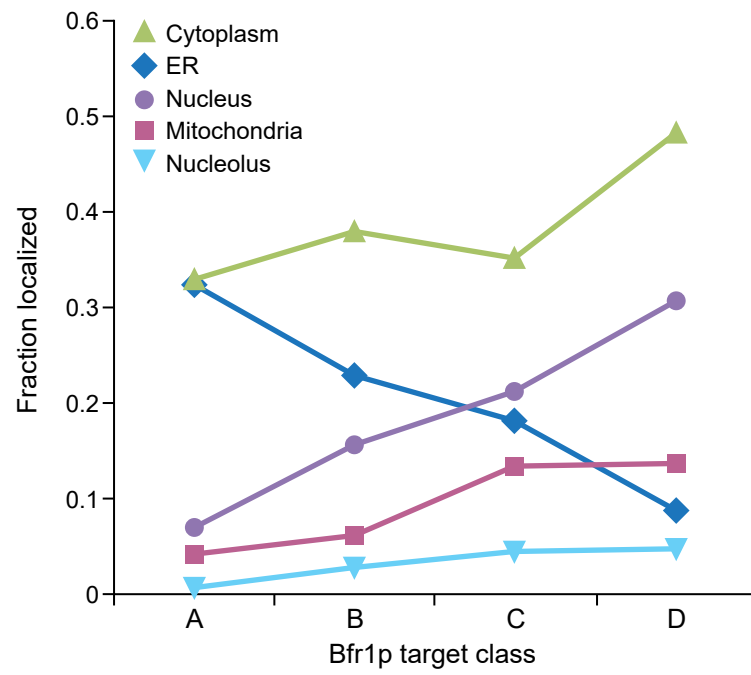
Supplementary Figure 11. Bfr1p targets identified by both RNA Tagging and RIP-chip²⁹ were enriched for membrane-associated functions. Proportional Venn diagram of Bfr1p targets identified using RNA Tagging and RIP-chip²⁹. GO analyses were performed on the three groups and enrichments for representative terms from Biological Process and Cellular Component ontologies are indicated (see **Supplementary Data 2** for complete lists).

Supplementary Figure 12



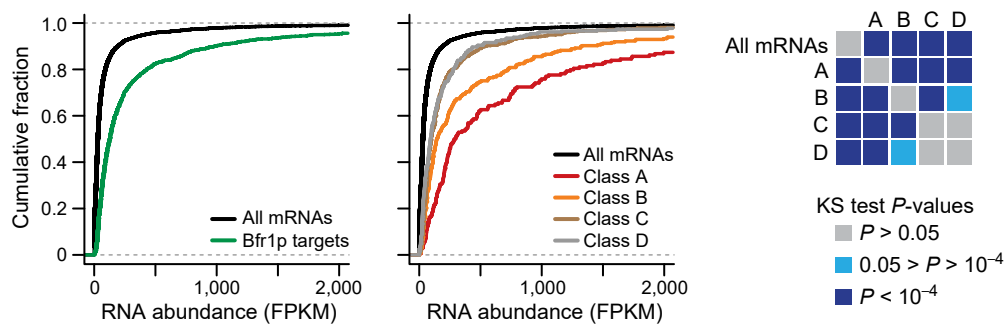
Supplementary Figure 12. Bfr1p target rank was correlated with TRPM and was very weakly correlated with U-tag length and RNA abundance. **a)** The mean number of Tagged RNAs (TRPM) detected for Bfr1p targets was correlated with their RNA Tagging rank ($\rho = -0.87$, $P = 0$; $n = 1,298$). Spearman's correlation coefficients (ρ) and associated P -values (P) are indicated in all panels. TRPM, Tagged RNAs Per Million uniquely mapped reads. **b)** The RNA Tagging rank of Bfr1p targets was weakly correlated with the mean length of their U-tags ($\rho = -0.35$, $P = 0$; $n = 1,298$). **c)** RNA Tagging rank of Bfr1p targets was weakly correlated with their mean RNA abundance ($\rho = -0.28$, $P = 0$; $n = 1,298$). FPKM, fragments per kilobase of exon per million reads mapped.

Supplementary Figure 13



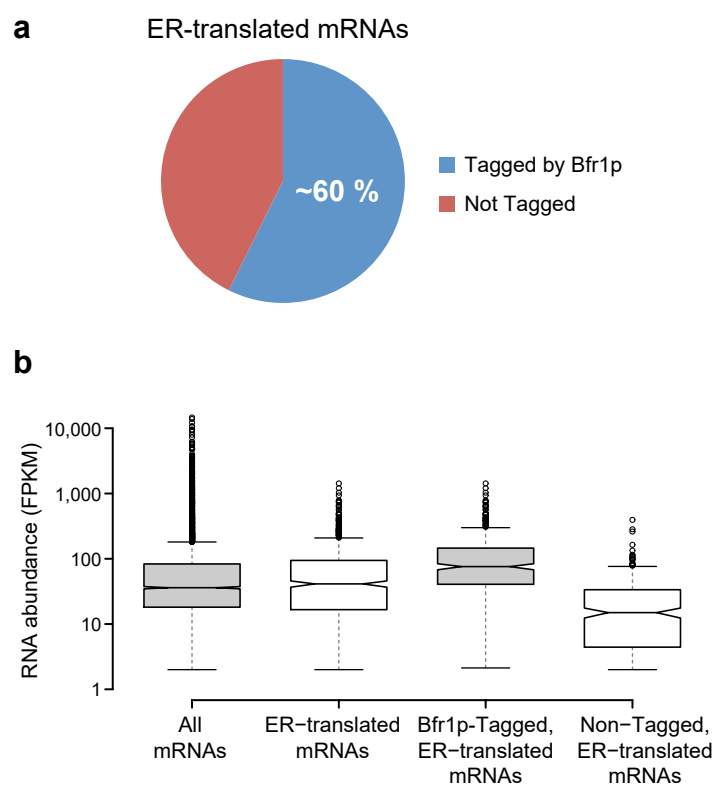
Supplementary Figure 13. Bfr1p target class was correlated with protein localization to the ER. The fraction of each class of Bfr1p targets that are localized to the cytoplasm, endoplasmic reticulum (ER), nucleus, mitochondria, and nucleolus, obtained from the yeast GFP database³², was plotted. Classes A-C of Bfr1p targets were highly enriched for ER-localized proteins (hypergeometric tests, $P < 1 \times 10^{-16}$), and the enrichment progressively decreased from Class A to D targets. No other significant enrichments were observed (hypergeometric tests, $P > 0.01$).

Supplementary Figure 14



Supplementary Figure 14. Bfr1p targets were highly enriched for abundant mRNAs. Empirical cumulative distributions of RNA abundance (FPKM) were plotted for all Bfr1p targets (left) and the Bfr1p target classes (middle) relative to all mRNAs (all mRNAs, $n = 6,595$; Class A, $n = 174$; Class B, $n = 297$; Class C, $n = 564$; Class D, $n = 261$). The P -values from Kolmogorov-Smirnov (KS) tests comparing the different distributions are indicated (right). Class A Bfr1p targets were most enriched for abundant RNAs and the enrichment progressively decreased to Class C and D targets. FPKM, fragments per kilobase of exon per million reads mapped.

Supplementary Figure 15



Supplementary Figure 15. Bfr1p bound abundant, ER-translated mRNAs. **a)** Plot of the fraction of ER-translated mRNAs (> 2-fold enrichment, $n = 736$), obtained from a published ER-specific ribosome profiling experiment⁴² ($\log_2(\text{ubc6.7mchx})$ enrichment), that were tagged by Bfr1p (422 mRNAs). **b)** Plots of the RNA abundance (FPKM) of the indicated groups of mRNAs (Tukey whiskers indicated). Of the mRNAs specifically translated at the ER, those tagged by Bfr1p were significantly more abundant than those not tagged by Bfr1p (Fisher-Pitman permutation test, $P < 10^{-6}$). FPKM, fragments per kilobase of exon per million reads mapped.

REFERENCES

- 1 McHugh, C. A., Russell, P. & Guttman, M. Methods for comprehensive experimental identification of RNA-protein interactions. *Genome biology* **15**, 203, doi:10.1186/gb4152 (2014).
- 2 Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A* **97**, 14085-14090, doi:10.1073/pnas.97.26.14085 (2000).
- 3 Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**, 939-953, doi:10.1016/j.molcel.2010.12.011 (2010).
- 4 Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215, doi:10.1126/science.1090095 (2003).
- 5 Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469, doi:10.1038/nature07488 (2008).
- 6 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141, doi:10.1016/j.cell.2010.03.009 (2010).
- 7 Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-915, doi:10.1038/nsmb.1838 (2010).
- 8 Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692-1694, doi:10.1261/rna.7151404 (2004).

- 9 Riley, K. J., Yario, T. A. & Steitz, J. A. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA* **18**, 1581-1585, doi:10.1261/rna.034934.112 (2012).
- 10 Riley, K. J. & Steitz, J. A. The "Observer Effect" in genome-wide surveys of protein-RNA interactions. *Mol Cell* **49**, 601-604, doi:10.1016/j.molcel.2013.01.030 (2013).
- 11 Darnell, R. B. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* **1**, 266-286, doi:10.1002/wrna.31 (2010).
- 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, doi:10.1111/j.1751-1097.2007.00179.x (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, doi:10.1074/jbc.M113.455451 (2013).
- 14 Kim, B. *et al.* TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *EMBO J* **34**, 1801-1815, doi:10.15252/embj.201590931 (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 biology* **2**, E79, doi:10.1371/journal.pbio.0020079 (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 U S A* **106**, 20192-20197, doi:10.1073/pnas.0812079106 (2009).

- 17 Olivas, W. & Parker, R. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J* **19**, 6602-6611, doi:10.1093/emboj/19.23.6602 (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, doi:10.1371/journal.pone.0002293 (2008).
- 19 Gadir, N., Haim-Vilmovsky, L., Kraut-Cohen, J. & Gerst, J. E. Localization of mRNAs coding for mitochondrial proteins in the yeast *Saccharomyces cerevisiae*. *RNA* **17**, 1551-1565, doi:10.1261/rna.2621111 (2011).
- 20 Chatenay-Lapointe, M. & Shadel, G. S. Repression of mitochondrial translation, respiration and a metabolic cycle-regulated gene, SLF1, by the yeast Pumilio-family protein Puf3p. *PLoS One* **6**, e20441, doi:10.1371/journal.pone.0020441 (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, doi:10.1083/jcb.200606054 (2007).
- 22 Wilinski, D. *et al.* RNA regulatory networks diversified through curvature of the PUF protein scaffold. *Nat Commun* **6**, doi:10.1038/ncomms9213 (2015).
- 23 Kusov, Y. Y., Shatirishvili, 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**, E57-57 (2001).

- 24 Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res* **36**, 5482-5515, doi:10.1093/nar/gkn517 (2008).
- 25 Freeberg, M. A. *et al.* Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. *Genome biology* **14**, R13, doi:10.1186/gb-2013-14-2-r13 (2013).
- 26 Bailey, T. E., C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology August*, 28-36 (1994).
- 27 Miller, M. A., Russo, J., Fischer, A. D., Lopez Leban, F. A. & Olivas, W. M. Carbon source-dependent alteration of Puf3p activity mediates rapid changes in the stabilities of mRNAs involved in mitochondrial function. *Nucleic Acids Res* **42**, 3954-3970, doi:10.1093/nar/gkt1346 (2014).
- 28 Wickens, M., Bernstein, D. S., Kimble, J. & Parker, R. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-157, doi:S0168952501026166 [pii] (2002).
- 29 Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology* **6**, e255, doi:10.1371/journal.pbio.0060255 (2008).
- 30 Jackson, J. S., Jr., Houshmandi, S. S., Lopez Leban, F. & Olivas, W. M. Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA* **10**, 1625-1636, doi:10.1261/rna.7270204 (2004).

- 31 Houshmandi, 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, doi:10.1261/rna.2168505 (2005).
- 32 Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691, doi:10.1038/nature02026 (2003).
- 33 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, doi:10.1126/science.1257522 (2014).
- 34 Sun, M. *et al.* Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell* **52**, 52-62, doi:10.1016/j.molcel.2013.09.010 (2013).
- 35 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).
- 36 Trautwein, M., Dengjel, J., Schirle, M. & Spang, A. Arf1p provides an unexpected link between COPI vesicles and mRNA in *Saccharomyces cerevisiae*. *Mol Biol Cell* **15**, 5021-5037, doi:10.1091/mbc.E04-05-0411 (2004).
- 37 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 mRNP complexes in yeast. *Nucleic Acids Res* **29**, 2567-2574 (2001).
- 38 Weidner, J., Wang, C., Prescianotto-Baschong, C., Estrada, A. F. & Spang, A. The polysome-associated proteins Scp160 and Bfr1 prevent P body formation

- under normal growth conditions. *J Cell Sci* **127**, 1992-2004, doi:10.1242/jcs.142083 (2014).
- 39 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, doi:10.1242/jcs.139055 (2014).
- 40 Ast, T., Cohen, G. & Schuldiner, M. A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell* **152**, 1134-1145, doi:10.1016/j.cell.2013.02.003 (2013).
- 41 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**, 127-133, doi:10.1038/nsmb.2468 (2013).
- 42 Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521, doi:10.1126/science.1257521 (2014).
- 43 Munoz-Tello, P., Rajappa, L., Coquille, S. & Thore, S. Polyuridylation in Eukaryotes: A 3'-End Modification Regulating RNA Life. *Biomed Res Int* **2015**, 968127, doi:10.1155/2015/968127 (2015).
- 44 Norbury, C. J. Cytoplasmic RNA: a case of the tail wagging the dog. *Nat Rev Mol Cell Biol* **14**, 643-653, doi:10.1038/nrm3645 (2013).
- 45 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, doi:10.1016/j.molcel.2014.02.007 (2014).

- 46 Newman, M. A., Mani, V. & Hammond, S. M. Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA* **17**, 1795-1803, doi:10.1261/rna.2713611 (2011).
- 47 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, doi:10.1038/nrm2838 (2010).
- 48 Miller, J. E. & Reese, J. C. Ccr4-Not complex: the control freak of eukaryotic cells. *Crit Rev Biochem Mol Biol* **47**, 315-333, doi:10.3109/10409238.2012.667214 (2012).
- 49 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-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, doi:10.1186/1471-2164-9-488 (2008).
- 51 Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033-1037, doi:10.1038/nature07728 (2009).
- 52 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (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, doi:10.1006/jmbi.2000.4315 (2001).

Chapter 4:

An integrated and multi-omic strategy reveals Puf3p-mediated regulation of the mitochondrial respiratory chain

This chapter is currently being prepared for publication and was done in collaboration with Dave Pagliarini's research group (UW-Madison & Morgridge Research Institute). I will be co-first author with Jonathan Stefely from the Pagliarini group.

ABSTRACT

Mitochondrial biogenesis is a complex process involving coordinated biosynthesis of lipids, metabolites, and proteins encoded by two genomes. mRNA-binding proteins provide post-transcriptional regulation of complex processes, but defining their targets remains challenging. Here, we develop a multi-omic approach to identify mRNAs regulated by Puf3p, an mRNA-binding protein linked to mitochondria. We integrated data from our recent proteomics and protein-RNA network studies to identify 140 high-confidence direct (cis) Puf3p protein targets and indirect (trans) effects. Nearly all characterized cis Puf3p targets are associated with mitochondrial respiratory chain biogenesis, including coenzyme Q (CoQ) biosynthesis. We demonstrate a role for Puf3p in CoQ biosynthesis via its regulation of Coq5p, which inhibited respiratory yeast growth and CoQ production when expressed at inappropriately high levels. We also connect three poorly characterized Puf3p targets to discrete respiratory chain-related functions. Together, our multi-omic analyses provide molecular insight into a program for coordinated post-transcriptional regulation of mitochondrial respiratory chain biogenesis.

INTRODUCTION

Mitochondria are complex organelles whose dysfunction is implicated in over 150 human diseases (Koopman et al., 2012; Nunnari and Suomalainen, 2012; Vafai and Mootha, 2012). The composition of mitochondria changes rapidly with alterations in the metabolic needs of the cell (Labbe et al., 2014), but the molecular networks that regulate this remodeling remain unclear. Mitochondrial biogenesis is complicated because the mitochondrial proteome is encoded by two separate genomes: the nuclear genome (nDNA) and the mitochondrial genome (mtDNA). A critical example addressed here involves the multi-protein respiratory chain complexes of oxidative phosphorylation (OxPhos), whose nDNA- and mtDNA- encoded subunits must be assembled in a coordinated fashion. A recent study demonstrated that while transcription of OxPhos mRNAs is not coordinated, the mitochondrial and cytoplasmic translation programs are synchronized (Couvillion et al., 2016); yet the molecular drivers of this coordination remain unclear.

OxPhos biogenesis is further complicated by the need to incorporate lipids and small molecules into protein complexes. Thus, OxPhos biogenesis poses an inherently multi-omic problem, likely requiring synchronized remodeling of the mitochondrial proteome, metabolome, and lipidome. An important example addressed here involves the biosynthesis of coenzyme Q (CoQ), a redox active lipid in the electron transport chain whose deficiency causes numerous human diseases (Laredj et al., 2014; Quinzii and Hirano, 2010). CoQ biosynthesis requires assembly of a multi-protein complex of enzymes (“complex Q”) in the mitochondrial matrix (Floyd et al., 2016; He et al., 2014;

Stefely et al., 2016b), production of the water-soluble CoQ headgroup precursor 4-hydroxybenzoate (4-HB) from tyrosine (Payet et al., 2016; Stefely et al., 2016a), and synthesis of the final lipid CoQ product (Tran and Clarke, 2007). How CoQ biosynthesis is regulated and coordinated with the larger process of OxPhos biogenesis remains obscure.

Cytoplasmic mRNA-binding proteins (RBPs) are great candidates to be post-transcriptional regulators of mitochondrial biogenesis, but pinpointing their molecular functions remains difficult. RBPs often bind large cohorts of mRNAs encoding functionally-related proteins, which may number in the hundreds to thousands (Hogan et al., 2008; Keene, 2007; Konig et al., 2012; Licatalosi and Darnell, 2010). RBPs impact the processing, stability, translation, and/or localization of the mRNAs they regulate (Muller-McNicoll and Neugebauer, 2013; Schoenberg and Maquat, 2012; Singh et al., 2015). However, it remains unclear which RBP-mRNA binding events ultimately lead to regulation. Strategies to reduce large transcriptome-wide interaction lists to core subsets, with a high concentration of regulatory interactions, are key to pinpointing RBP function.

A fundamental requirement in the study of RBPs is to uncover which RNAs they bind *in vivo*. Such “protein-RNA network” experiments primarily adhere to two distinct methodologies, which are epitomized by HITS-CLIP and RNA Tagging. In HITS-CLIP (high-throughput sequencing preceded by UV-crosslinking immunoprecipitation), cells or tissues are irradiated with UV-light, crosslinked RBP-RNA complexes are immunopurified from cell lysate, and RNA footprints protected by the RBP are identified via high-

throughput sequencing (Darnell, 2010; Licatalosi et al., 2008; Ule et al., 2003). HITS-CLIP therefore reveals both the identity and location of RBP binding events transcriptome-wide. In contrast, RNA Tagging fuses a poly(U) polymerase to the protein of interest, and RNAs bound by the fusion protein are covalently tagged with 3' uridines (“U-tags”) *in vivo* (Lapointe et al., 2015). U-tagged RNAs are subsequently identified via high-throughput sequencing. Given their distinct approaches, we suspected HITS-CLIP and RNA Tagging might prove complementary, and the union of the two data sets would yield a high-confidence set of mRNAs bound by a protein of interest. In parallel, there has yet to be a comprehensive comparison of the two techniques, which we are in an exceptional position to perform with *Saccharomyces cerevisiae* Puf3p.

Puf3p belongs to the “PUF” family of RNA-binding proteins, which are conserved across Eukarya and have critical roles in stem cells, gametogenesis, early development, and the nervous system (Miller and Olivas, 2011; Quenault et al., 2011; Spassov and Jurecic, 2003; Wickens et al., 2002). Puf3p contains eight tandem PUF RNA-binding domains in its C-terminal half (Olivas and Parker, 2000; Zhu et al., 2009), with potent regulatory regions housed in its low-complexity N-terminal half (Lee and Tu, 2015). Puf3p binds to RNA sequences called “Puf3p-binding elements” (PBEs) that conform to the consensus UGUANAUA and are typically present in the 3' untranslated regions of mRNAs (Gerber et al., 2004; Jackson et al., 2004; Lapointe et al., 2015; Olivas and Parker, 2000) (Wilinski et al. *in review*). For high-affinity interactions (sub-nanomolar range), Puf3p requires a cytosine one or two nucleotides upstream of the initial 5' U (Zhu et al., 2009). Once bound, Puf3p can destabilize the mRNA and/or repress its translation (Chatenay-Lapointe and

Shadel, 2011; Gerber et al., 2004; Houshmandi and Olivas, 2005; Jackson et al., 2004; Klass et al., 2013; Lee and Tu, 2015; Lee et al., 2010; Olivas and Parker, 2000; Rowe et al., 2014; Sun et al., 2013). Puf3p-mediated regulatory activity can be toggled on or off depending on the carbon source, as Puf3p increases translation of a few mRNAs immediately following a shift from a fermentable (glucose) to a non-fermentable (glycerol) carbon source (Lee and Tu, 2015; Miller et al., 2014).

Puf3p is implicated in the regulation of mitochondria. Puf3p binds to hundreds of mRNAs that represent nDNA-encoded mitochondrial proteins (Gerber et al., 2004; Hogan et al., 2015; Lapointe et al., 2015) (Wilinski et al. *in review*), which appears conserved across more than approximately 300 million years of evolution (Hogan et al., 2015; Jiang et al., 2010; Taylor and Berbee, 2006) (Wilinski et al. *in review*). In parallel to its role in the regulation of mRNA stability and translation, Puf3p is required for the localization of many mRNAs to the outer surface of mitochondria (Eliyahu et al., 2010; Gadir et al., 2011; Saint-Georges et al., 2008). Yeast that lack *puf3* (“ $\Delta puf3$ yeast”) have several mitochondria-related phenotypes, including reduced respiratory growth (Eliyahu et al., 2010; Gerber et al., 2004; Lee and Tu, 2015), increased respiratory activity during fermentation (Chatenay-Lapointe and Shadel, 2011), and mitochondrial morphology and motility defects (Garcia-Rodriguez et al., 2007). The particular molecular pathways regulated by Puf3p, however, remain a relative unknown.

Here we integrate protein-RNA networks from our recent RNA Tagging and HITS-CLIP studies (Lapointe et al., 2015) (Wilinski et al. *in review*) with proteomic, metabolomic, and

lipidomic profiles from our recent “Y3K” study (Stefely et al., 2016a). This multi-omic analysis defines a high-confidence set of molecular targets that Puf3p regulates to control OxPhos biogenesis. By showing that Puf3p regulates CoQ biosynthesis, our work also reveals a mechanism for synchronizing synthesis of proteins and lipids needed for the OxPhos machinery. Our analysis also provides a foundation for defining the biochemical functions of uncharacterized mitochondrial proteins (MXPs) regulated by Puf3p and a roadmap for future studies of any RNA-binding protein.

RESULTS

Yeast that lack *puf3* specifically upregulate mitochondrial proteins

We systematically surveyed our Y3K study (Stefely et al., 2016a), which in part examined the proteomes of 174 single-gene deletion ($\Delta gene$) yeast strains under two metabolic conditions (fermentation and respiration), to identify proteins that were primarily affected by loss of *puf3*. We identified a selective and significant increase in the abundance of the protein Cmc2p in $\Delta puf3$ yeast compared to wild type (WT) yeast and all other $\Delta gene$ strains in the study ($P < 0.05$) (**Figure 1A**). This observation suggests that *cmc2* mRNA may be bound by Puf3p, and that lack of Puf3p-mediated regulation causes increased expression of Cmc2p (**Figure 1B**). We also identified nine additional proteins that were selectively and significantly upregulated in $\Delta puf3$ yeast ($P < 0.05$) (**Figures 1C and S1**). Strikingly, all ten proteins function in mitochondria. For example, Cmc2p functions in OxPhos complex IV biogenesis (Horn et al., 2010). Nine of the ten upregulated protein phenotypes were identified under fermentation culture conditions (“fermenting yeast”) (**Figures 1A, 1C, and S1A**), while only Hsp60p was identified under the respiration culture condition (“respiring yeast”) (**Figure S1B**), which suggests that Puf3p functions primarily in fermenting yeast.

To identify additional protein abundance changes in $\Delta puf3$ yeast that do not meet the strict Y3K criteria for gene-specific phenotypes, we examined changes across the entire proteome of $\Delta puf3$ yeast compared to WT yeast. In fermenting yeast, 160 significant $\Delta puf3$ proteome changes were observed ($P < 0.05$ and fold change [FC] $> 25\%$) (**Figure 1D**). In contrast, only 24 such changes were observed in respiring yeast, so we primarily

focused our studies on the molecular functions of Puf3p in fermentation. The $\Delta puf3$ proteome alterations are likely a combination of both direct (cis) and indirect (trans) effects (**Figure 1E**). A key goal is to distinguish these two possibilities and thereby reveal the molecular basis of Puf3p function. Here, we use “cis Puf3p target protein” (“cis target”) to refer to a protein that is encoded by an mRNA directly bound Puf3p and whose abundance is significantly altered in $\Delta puf3$ yeast (FC > 25% and $P < 0.05$). In contrast, we use “trans Puf3p effect” (“trans effect”) to refer to any downstream effect at the protein, metabolite, or lipid level (e.g., a protein abundance change secondary to a cis target abundance change).

Integration of two methods yields high-confidence Puf3p target mRNAs

Rigorous definition of Puf3p-bound mRNAs is required to distinguish cis Puf3p target proteins from proteins affected in trans (**Figure 1E**). To curate a high-confidence list of Puf3p-bound mRNAs, we comprehensively analyzed and integrated data based on two independent approaches for identifying Puf3p-bound mRNAs: HITS-CLIP (Wilinski et al., *in review.*) and RNA Tagging (Lapointe et al., 2015).

A sub-set of Puf3p-bound RNAs were identified by two independent approaches. HITS-CLIP and RNA Tagging identified 467 and 476 Puf3p-bound RNAs, respectively, and 269 mRNAs were identified via both techniques (**Figure 2A**) (hypergeometric test, $P < 10^{-211}$). The signal detected by each approach for Puf3p-bound mRNAs was moderately correlated across methods (Spearman’s $\rho = 0.42$, $P < 10^{-12}$), but not with mRNA abundance ($P > 0.01$) (**Figures S2A–S2C**). An impressive 83% (224/269) of Puf3p-bound

mRNAs identified via both methods represented nDNA-encoded mitochondrial proteins (hypergeometric test, $P < 10^{-126}$) (**Figure 2B**). The mRNAs uniquely identified by a method were much less enriched. Furthermore, the 3' UTRs of mRNAs identified via both methods were highly enriched for sequences that conform to high-affinity PBEs: UGUAHAUA (**Figure 2C**). Remarkably, 81% (219/269) of the PBEs also contained cytosine one or two nucleotides upstream of the first uridine (the -1 and -2 positions), which is particularly indicative of high-affinity interactions. Substitution of cytosine with any other nucleotide at the -2 position of an otherwise consensus PBE reduces Puf3p-RNA interactions approximately 70-fold *in vitro* (Zhu et al., 2009). In contrast, RNAs uniquely identified by HITS-CLIP or RNA Tagging were enriched for more degenerate PBEs and upstream cytosines were much less frequent (79/198 and 102/207, respectively), with RNA Tagging providing increased enrichment in both instances. Thus, integration of the HITS-CLIP and RNA Tagging data yielded a high-confidence set of Puf3p-bound mRNAs, which predominately encode mitochondrial proteins and have high-affinity PBEs.

Our analyses suggested that high-affinity Puf3p-RNA interactions were preferentially identified via both approaches, and we previously demonstrated that strength of detection in RNA Tagging is correlated with binding affinity (Lapointe et al., 2015) (Lapointe et al., *in preparation*, see chapter 5). To test if strength of detection correlated with identification via both approaches, we separated Puf3p-bound RNAs identified via HITS-CLIP into groups, referred to as “classes”, and calculated the overlap among methods by class (see Methods for full details). In brief, class I genes were best detected in an approach and

class IV genes were the weakest detected. Approximately 90% of genes present in class I of RNA Tagging or HITS-CLIP data sets were identified via both approaches, and the overlap across methods progressively decreased to class IV genes (**Figure 2D**). Similar to RNA Tagging, class I HITS-CLIP genes were most enriched for high-affinity PBEs and the enrichment progressively decreased from class I to class IV genes (**Figure S2D**).

Puf3p-bound mRNAs identified via RNA Tagging and HITS-CLIP are often significantly upregulated in $\Delta puf3$ yeast. In our Y3K study, 165 proteins encoded by Puf3p-bound mRNAs identified via both methods were detected in fermenting yeast (**Figure 2E**). Of those, 91 (55%) were significantly upregulated by at least 25% in $\Delta puf3$ yeast relative to WT yeast (two-sided Student's *t*-test, $P < 0.05$) (**Figures 2E and 2F**). Three additional proteins exhibited less substantial but significantly increased expression, and the vast majority of the remaining Puf3p-bound mRNAs had increased albeit non-significant protein abundances (**Figure 2E**). In contrast, Puf3p-bound RNAs uniquely identified by a method were far less likely to be upregulated (**Figures 2F and S2E**).

The regulatory potential of binding events in the open-reading frame (ORF) of mRNAs remains an open and important question for Puf3p, especially since all but one gene identified via an ORF peak in HITS-CLIP was absent from RNA Tagging data (**Figure S2F**). We therefore analyzed genes identified via a single peak in our HITS-CLIP experiment (451 genes) to determine the relationship between Puf3p-binding position and *puf3*-dependent protein abundance. Only two genes detected by an ORF peak (40 total) had significantly altered protein abundance in $\Delta puf3$ yeast (one increased, one

decreased), while 93 genes with a 3' UTR peak (241 total) were significantly increased (FC > 25% and $P < 0.05$, two-sided Student's t -test) (**Figure 2G**). The average PBEs present in ORF and 3' UTR peaks were highly similar (**Figure 2H**), which suggests the regulatory potential of a particular PBE is at a minimum impacted by both its quality and location.

Through the integration of our RNA Tagging and HITS-CLIP data, we have rigorously identified 269 high-confidence Puf3p-bound mRNAs. These mRNAs predominately represent nDNA-encoded mitochondrial proteins, predominately possess high-affinity PBEs, are most likely to be present near the top of one or both data sets, and are most likely to exhibit *puf3*-dependent changes in protein abundance. Thus, we henceforth refer to Puf3p-bound mRNAs identified via both HITS-CLIP and RNA Tagging as “Puf3p target mRNAs”. Our integration of two independent yet complementary approaches allowed us to focus on a substantially reduced number of mRNAs (a $\approx 60\%$ reduction) as we sought to identify specific molecular pathways regulated by Puf3p.

Puf3p targets the OxPhos biogenesis pathway

Rigorously-defined Puf3p target mRNAs enabled us to divide $\Delta puf3$ proteome changes into 91 cis Puf3p protein targets and 49 trans Puf3p effects (**Figures 1E and 3A**). In order to confidently assign cis and trans designations, we did not bin a third group of $\Delta puf3$ proteome changes whose encoding mRNAs were only detected as bound by Puf3p via a single method above (20 proteins). The majority of cis targets and trans effects detected in fermenting yeast were unaffected in respiring yeast (**Figures 3B, S3A and S3B**), which

again demonstrated that Puf3p functions are dependent on metabolic state. Cis Puf3p protein targets are highly enriched for mitochondrial proteins, especially those involved in mitochondrial translation (**Figure S3C**), whereas trans Puf3p effects with increased protein levels are enriched for oxidation-reduction processes and ATP synthesis coupled to electron transport (OxPhos) (**Figure S3D**).

Manual inspection revealed that 81 of the 87 Puf3p target proteins with known functions fit into a pathway that generates the OxPhos machinery (**Figure 3C**). This pathway includes proteins that catalyze the import, folding, and processing of nDNA-encoded mitochondrial proteins, which include many OxPhos subunits. Similarly, cis Puf3p targets included proteins that support transcription and translation of mtDNA-encoded genes, which also encode OxPhos subunits. For example, cis Puf3p targets include over half of mitochondrial ribosomal proteins (**Figures 3D and S3E**) and numerous general and gene-specific activators of mitochondrial translation (e.g., Cbp6p, Mam33p, Mba1p, and Mdm38p). Consistent with upregulated mitochondrial translation, Puf3p trans effects included increased abundance of two mtDNA-encoded proteins (Var1p and Cox2p; the only two mtDNA-encoded proteins observed in the proteomics data set). Strikingly, numerous OxPhos assembly factors were also defined as cis Puf3p targets, and, consistently, numerous OxPhos complex subunits were defined as trans effects whose abundance increased in $\Delta puf3$ yeast (**Figure 3C and S3F**). Thus, cis Puf3p target proteins group into a pathway that generates the OxPhos machinery, and, accordingly, trans Puf3p effects include many downstream OxPhos proteins.

The multi-omic nature of the Y3K data set also gave us the opportunity to investigate trans effects across multiple classes of biomolecules. For example, citrate synthase (Cit1p) and Ptc7p, a phosphatase that positively regulates Cit1p activity (Guo et al., *under review*), were significantly increased trans Puf3p effect proteins (**Figures S3G and S3H**). Consistent with an increase in citrate synthase activity, the abundance of citrate was significantly elevated in $\Delta puf3$ yeast ($P < 0.05$) (**Figure S3I**). Thus, our integrated and multi-omic analysis, which spans from mRNA to protein to metabolites, provides a high-quality map to uncover Puf3p-mediated regulation.

Puf3p regulates a promiscuous CoQ biosynthesis protein

Puf3p-regulated OxPhos biogenesis proteins included two proteins in the CoQ biosynthesis pathway (**Figure 3C**). Loss of Puf3p increased expression of both Coq5p (cis target) and Coq1p (trans effect) in fermentation (**Figure 4A**), but not in respiration (**Figure S4A**). Coq5p is a member of the multi-protein complex Q and catalyzes a C-methylation in CoQ biosynthesis (Barkovich et al., 1997; Dibrov et al., 1997; Floyd et al., 2016; He et al., 2014; Stefely et al., 2016b). Increased expression of Coq5p, such as we observed in fermenting $\Delta puf3$ yeast, may disrupt complex Q stoichiometry, organization, or function (**Figure 4B**). Thus, we hypothesized that the first two steps of CoQ biosynthesis, which are catalyzed by Coq1p and Coq2p, would be accelerated in $\Delta puf3$ yeast, while subsequent complex Q (Coq3p–Coq9p) catalyzed steps would be disrupted because of unbalanced enzyme levels (e.g., Coq5p elevation in the absence of Coq3p and Coq6p elevation) (**Figure 4C**). Consistently, fermenting $\Delta puf3$ yeast have increased abundance of polyprenylhydroxybenzoate (PPHB), an early CoQ intermediate produced

by Coq1p and Coq2p, but deficiency of demethoxy-CoQ (DMQ) and CoQ, whose production depends on complex Q (**Figure 4D**). In contrast, PPHB, DMQ, and CoQ levels were largely unaffected in respiring $\Delta puf3$ yeast (**Figures S4B and S4C**), again showing the condition-specific function of Puf3p. In follow-up studies, overexpression of Puf3p from both high- and low-copy plasmids in WT yeast suppressed production of PPHB and PPAB (the aminated analog PPHB)—striking effects because they are the inverse of those observed in $\Delta puf3$ yeast (**Figures 4E, S4D, and S4E**). Plasmid expression of Puf3p also recovered CoQ biosynthesis in fermenting $\Delta puf3$ yeast (**Figure 4E**).

To reveal how loss of Puf3p dysregulates CoQ biosynthesis at the molecular level, we examined how Coq5p overexpression impacted yeast growth and the CoQ biosynthetic pathway. Overexpression of Coq5p in WT yeast slowed fermentation growth and essentially eliminated respiratory growth (**Figure 4F**), in stark contrast to the negative controls Yjr120w and Hem25p (**Figure 4G**). Similarly, overexpression of another core complex Q member, Coq9p, inhibited respiratory growth, while overexpression of other CoQ pathway enzymes (e.g., Coq8p and Hfd1p) had limited effects (**Figures 4G and 4H**). To examine whether Coq5p overexpression inhibits respiratory yeast growth by disrupting CoQ production, we examined CoQ pathway intermediates. Importantly, Coq5p overexpression in WT yeast recapitulated the phenotype of $\Delta puf3$ yeast—deficiency of DMQ and CoQ, and elevation of PPAB and PPHB (**Figures 4I, S4F, and S4G**). Consistent with the growth effects, CoQ intermediates were most affected by overexpression of Coq5p compared to other proteins.

Coq5p physically interacts with Coq8p, and Coq8p overexpression stabilizes complex Q through an undefined molecular mechanism (Floyd et al., 2016; He et al., 2014; Stefely et al., 2016b). We thus reasoned that Coq8p overexpression might ameliorate the deleterious effects of increased Coq5p expression. Strikingly, low-copy Coq8p overexpression recovered the respiratory growth of yeast with high-copy Coq5p overexpression (**Figures 4J and 4K**). These results provide further evidence that Coq5p overexpression inhibits CoQ production by disrupting complex Q. Together, our findings demonstrate that Puf3p modulates CoQ biosynthesis by regulating the abundance of Coq5p, a potentially promiscuous protein with toxic effects when overexpressed.

Mapping Puf3p targets assists MXP functionalization

The tight functional association of the cis Puf3p protein targets in the OxPhos biogenesis pathway (93% of cis Puf3p target proteins with annotated functions) (**Figure 3C**) led us to predict that the four mitochondrial uncharacterized proteins (MXPs) that are cis Puf3p targets—Rdl2p, Ynr040w, Mpm1p, and Fmp10p (**Figure 5A**)—also function in this pathway. While the gene deletion strains for these MXPs are respiration competent (**Figure S5A**), overexpression of these MXPs in WT yeast inhibited respiratory growth (**Figure 5B**), which suggested that they may interact with proteins required for OxPhos.

To test this idea, we employed affinity enrichment mass spectrometry (AE-MS) to identify protein interaction partners for the four MXPs. We generated C-terminally FLAG tagged constructs for each MXP, transformed them into WT yeast, and cultured the yeast into a respiratory growth phase. Three of the four constructs were successfully

immunoprecipitated, and each of these significantly ($P < 0.05$) enriched a distinct interaction partner (**Figures 5C and 5D**). Rdl2p immunoprecipitated with Atp1p, an ATP synthase subunit (Takeda et al., 1986), which suggests that Rdl2p might function in OxPhos complex V assembly or function. Similarly, the Fmp10p–Qcr7p interaction suggested that Fmp10p might be important for OxPhos complex III (Crivellone et al., 1988). Interestingly, Mpm1p (Mitochondrial peculiar membrane protein 1) immunoprecipitated with Vma1p (Vacuolar membrane ATPase protein 1). While Vma1p is not annotated as an OxPhos protein, interactions between vacuoles and mitochondria are important for yeast cell metabolism (Elbaz-Alon et al., 2014; Honscher et al., 2014). We thus hypothesized that Mpm1p mediates an interaction between these organelles that is important for OxPhos biogenesis.

Examination of publicly available genetic and proteomic data sets revealed numerous interactions between Mpm1p and mitochondrial outer membrane proteins, vacuolar proteins, and proteins involved in lipid binding or metabolism (**Figure S5B**) (Costanzo et al., 2010; Gavin et al., 2002; Hoppins et al., 2011; Szklarczyk et al., 2015). Analysis of our Y3K proteomics data set revealed a striking loss of the vacuolar protein Mon1p in $\Delta mpm1$ yeast (**Figure 5E**), further supporting a functional link between Mpm1p and vacuolar biology. Interestingly, numerous perturbations of proteins involved in the metabolism of sphingolipids and glycerophospholipids, including vacuolar metabolism of phosphatidylinositol (PI) (**Figure 5F**), were also observed. These observations led us to narrow our model for Mpm1p function to a role in mediating inter-organellar transport and metabolism of lipids between vacuoles and mitochondria (**Figure 5G**). Consistently,

analysis of the Y3K lipidomics data set revealed significant ($P < 0.05$) and unique perturbations of numerous lipids in $\Delta mpm1$ yeast (**Figure 5H and 5I**). Together, these results suggest that the Mpm1p physically and functionally links mitochondria and vacuoles during OxPhos biogenesis.

DISCUSSION

Implications for genome-wide studies on RNA-binding proteins

The genome-wide approach that we employed to study Puf3p is broadly applicable to any RNA-binding protein (RBP). Despite leveraging distinct methodologies, the HITS-CLIP and RNA Tagging data sets overlapped by ~50%, which is impressive, but also suggests room for improvement in the individual methods. Importantly, overlap between the two methods was highest among the most robustly detected mRNAs. The top ends of the Puf3p data sets agree extremely well (e.g., class I genes, > 90% overlap) and are very likely to be Puf3p target mRNAs *in vivo*. Furthermore, Puf3p-RNA interactions that decreased abundance of the encoded protein were much more likely to be strongly detected mRNAs in both approaches. Thus, our analyses illustrate that RBP-mRNA interaction data sets are likely stratified, with reproducible and potential regulatory interactions concentrated near the top.

Why is there disagreement among weakly detected Puf3p-bound RNAs? Our analyses suggest that RNA Tagging detects more frequent or long-lasting Puf3p-RNA interactions because the unique RNA Tagging targets have better PBEs on average than unique CLIP targets. We speculate that this arises because RNA Tagging requires more frequent or longer interactions to deposit U-Tags, while UV-crosslinking captures transient interactions. In parallel, however, the 3' U-tags placed by RNA Tagging may affect RNA metabolism or regulation and hence Puf3p target mRNAs. HITS-CLIP, on the other hand, requires yeast to be harvested by centrifugation prior to UV irradiation, which may alter gene expression or Puf3p activity. Thus, each approach has its own limitations. The

integration of multiple approaches alleviates intrinsic biases of each technique and facilitates characterization of RBP functions.

PBE location in an mRNA impacts Puf3p-mediated regulation. RNA Tagging may be limited by the distance from the binding event to the 3' end of the RNA, since the poly(U) polymerase needs access to it to deposit U-tags. However, our previous analyses suggested there was no effect on U-tag length by PBE position in *S. cerevisiae* 3' UTRs (Lapointe et al., 2015). Our findings presented here are consistent. For example, RNA Tagging detected two mRNAs that Puf3p binds via PBEs located in their 5' UTR, and another mRNA with a binding site in its coding sequence. While HITS-CLIP identified a sizable number of mRNAs with ORF binding sites, only two led to a detectable change in protein abundance. Thus, we favor a model in which ORF binding events are transient and unlikely to lead to Puf3p-mediated regulation, perhaps because the ribosome limits residence time of Puf3p on the mRNA or a 3' UTR context is required for regulation.

Why do Puf3p target mRNAs outnumber Puf3p target proteins? We found that only a subset of Puf3p target mRNAs had increased protein levels in yeast that lack Puf3p, which may suggest that Puf3p only regulates a subset of its bound mRNAs. Alternatively, however, there are multiple possibilities for Puf3p-mediated regulation, which include effects on mRNA localization, stability, or potential unknown regulatory effects. Thus, it will prove interesting to examine our list of Puf3p target mRNAs for alternative regulatory outcomes in the future. Moreover, our approach here represents an ensemble average of interactions and regulatory effects. Puf3p function varies in response to different food

sources, and it is possible that Puf3p activity may also vary with yeast density, the cell cycle, or other environmental stimuli. Thus, the weakest Puf3p target mRNAs with unaffected protein levels may be present in a few cells at a time and missed in our bulk analyses. Single-cell analyses, which are possible with RNA Tagging or a similar approach like TRIBE (McMahon et al., 2016), will provide insight into these outstanding questions.

Finally, our strategy demonstrates the power of multi-omics for revealing RBP function. The integration of our Puf3p mRNA target list with proteomics, metabolomics, and lipidomics provided another layer of validation for the Puf3p targets, revealed the downstream biochemical effects of perturbing the Puf3p-mRNA network, and provided a comprehensive understanding of the biological system regulated by Puf3p.

Post-transcriptional regulation of OxPhos biogenesis

Puf3p represses expression of mitochondrial biogenesis factors in fermenting yeast. We have identified a set of high-confidence Puf3p target mRNAs, which predominately represent nDNA-encoded mitochondrial proteins. Similarly, nearly all Puf3p target proteins group into a pathway that generates the mitochondrial OxPhos machinery. Moreover, trans Puf3p effects included protein components of OxPhos complexes, TCA cycle metabolites, and mitochondrial lipids, such as CoQ.

Trans Puf3p effected proteins with decreased protein levels included splicing factors and epigenetic regulators, and it is tempting to speculate that $\Delta puf3$ yeast have “sensed”

mitochondrial dysfunction. Consistently, Ngg1p was a decreased trans Puf3p effect and is a component of the SLIK complex, which regulates gene expression in response to mitochondrial dysfunction (Brandl et al., 1993; Pray-Grant et al., 2002; Saleh et al., 1997).

Our integrated, multi-omic analyses provide substantial evidence that the primary function of Puf3p in fermenting yeast is to repress expression of mitochondrial biogenesis factors, as opposed to a pleiotropic regulator suggested by a previous study (Kershaw et al., 2015). We also observed few effects in respiring $\Delta puf3$ yeast, which at first glance appears contradictory to a recent study that suggested Puf3p increased translation of a few mRNAs immediately following a diauxic shift (Lee and Tu, 2015). However, our respiration culture conditions were at 4 hours post shift, and at that time point, our findings are consistent with those observed by Lee and Tu. It remains to be seen if Puf3p is a broad translational activator in the immediate time points after a diauxic shift.

New insight into mitochondrial biochemistry

Our results show that Puf3p post-transcriptionally regulates the biosynthesis of CoQ, which is required for OxPhos. A Puf3p-mediated mechanism for the suppression of CoQ biosynthesis in fermenting yeast could provide a control point to move isoprene subunits away from generating CoQ, which is not needed in fermenting yeast, and instead toward other isoprenoid molecules (e.g., sterols). Furthermore, Puf3p provides a mechanism to synchronize an increase in CoQ biosynthesis with the larger program of OxPhos biogenesis.

Overexpression of Coq5p, a *cis* Puf3p target, or Coq9p, another complex Q member, inhibited respiratory yeast growth and CoQ production, suggesting that complex Q can be disrupted by an imbalance of its subunits. Puf3p likely plays an important role in regulating coordinated expression of complex Q components during OxPhos biogenesis. Interestingly, co-overexpression of Coq8p with Coq5p recovered yeast respiratory growth, which demonstrates that Coq8p provides a quality control mechanism for stabilizing complex Q in the presence of inappropriately high levels of a complex Q subunit. Coq8p is a protein kinase-like superfamily member that likely stabilizes complex Q through either small molecule kinase activity or ATPase activity (Stefely et al., 2016b; Stefely et al., 2015), but the precise molecular mechanism for Coq8p-dependent complex Q stabilization remains to be determined. Coq5p overexpression provides a readily modifiable and tractable system in which Coq8p activity can be studied.

Our results link four MXPs to OxPhos biogenesis by uncovering their regulation as *cis* Puf3p targets. We thereby provide an important and specific biological context to study their precise biochemical functions. Furthermore, our AE-MS studies linked Fmp10p and Rdl2p to OxPhos complexes III and V, respectively, which provides a foundation for deeper biochemical investigations. Finally, our work on Mpm1p suggests that it mediates an interaction between mitochondria and vacuoles and is important for lipid metabolism and OxPhos biogenesis. Discovering Mpm1p as a protein mediator of this inter-organelle interaction provides a new biochemical target for probing mitochondria-vacuole interactions in cell biology. Collectively, our work provides a foundation for understanding the molecular basis of mitochondrial biogenesis and a generalizable strategy for future

multi-omic studies of the many RNA-binding proteins that impact human health and disease.

METHODS

KEY RESOURCES TABLE

Please see separate document for the key resources table of this manuscript.

Yeast Strains

The parental (wild type, WT) *Saccharomyces cerevisiae* strain for this study was the haploid MATalpha BY4742. Single gene deletion ($\Delta gene$) derivatives of BY4742 were obtained through the gene deletion consortium (via Thermo, Cat#YSC1054). Gene deletions were confirmed by proteomics (significant [$P < 0.05$] and selective decrease in the encoded protein) or PCR assay.

Yeast were transformed with plasmids using a standard lithium acetate protocol (Gietz et al., 1992). Briefly, BY4742 yeast were cultured in YEPD (50 mL) to a density of 2×10^7 cells/mL. Cells were pelleted and washed twice with water. For each transformation, added PEG 3350 (50% w/v, 240 μ L), lithium acetate (1 M, 36 μ L), boiled salmon sperm DNA (5 mg/mL, 50 μ L), water (30 μ L), and plasmid (4 μ L) to a pellet containing 1×10^8 cells, mixed by vortexing, and incubated (42 °C, 45 min). The transformed cells were pelleted, resuspended in water (100 μ L), and plated on selective media.

Yeast Culture Conditions

General culture procedures and media components. Yeast were stored at -80 °C as glycerol stocks and initially cultured on selective solid media plates at 30 °C. Biological

replicates were defined as separate yeast colonies after transformation and plating onto solid selective media. Cell density of liquid media cultures was determined by optical density at 600 nm (OD_{600}) as described (Hebert et al., 2013). Media components included yeast extract ('Y') (Research Products International, RPI), peptone ('P') (RPI), agar (Fisher), dextrose ('D') (RPI), glycerol ('G') (RPI), uracil drop out (Ura^-) mix (US Biological), histidine drop out (His^-) mix (US Biological), and G418 (RPI). YP and YPG solutions were sterilized by automated autoclave. G418 and dextrose were sterilized by filtration (0.22 μ m pore size, VWR) and added separately to sterile YP or YPG. YPD+G418 plates contained yeast extract (10 g/L), peptone (20 g/L), agar (15 g/L), dextrose (20 g/L), and G418 (200 mg/L). YPD media (rich media fermentation cultures) contained yeast extract (10 g/L), peptone (20 g/L), and dextrose (20 g/L). YPGD media (rich media respiration cultures) contained yeast extract (10 g/L), peptone (20 g/L), glycerol (30 g/L) and dextrose (1 g/L). Synthetic Ura^- media were sterilized by filtration (0.22 μ m pore size). Ura^- ,D media contained Ura^- mix (1.92 g/L) and dextrose (20 g/L). Ura^- ,GD media contained Ura^- mix (1.92 g/L), glycerol (30 g/L), and dextrose (1 g/L). Ura^- ,D+4HB media contained Ura^- mix (1.92 g/L), dextrose (20 g/L), and 4-HB (100 μ M). Ura^- ,GD+4HB media contained Ura^- mix (1.92 g/L), glycerol (30 g/L), dextrose (1 g/L), and 4-HB (100 μ M).

Puf3 rescue cultures. WT or $\Delta puf3$ yeast were transformed with plasmids encoding Puf3p (p423(2 μ)-HIS3-PUF3 or p413(CEN)-HIS3-PUF3) and cultured on His^- ,D plates. Starter cultures (3 mL His^- ,D+4HB) were inoculated with an individual colony of yeast and incubated (30 °C, 230 rpm, 10–15 h). For CoQ quantitation, His^- ,D+4HB (fermentation)

or His⁻,GD+4HB (respiration) media (100 mL media at ambient temperature in a sterile 250 mL Erlenmeyer flask) was inoculated with 2.5×10^6 yeast cells and incubated (30 °C, 230 rpm). Samples of the His⁻,D+4HB cultures were harvested 13 h after inoculation, a time point that corresponds to early fermentation (logarithmic) growth. Samples of His⁻,GD+4HB cultures were harvested 25 h after inoculation, a time point that corresponds to early respiration growth. For each growth condition, 1×10^8 yeast cells were pelleted by centrifugation (3,000 g, 3 min, 4 °C), the supernatant was removed, and the cell pellet was flash frozen in N_{2(l)} and stored at -80 °C prior to lipid extractions. For relative growth rate measurements, analogous cultures (initial density of 5×10^6 cells/mL) were incubated in a sterile 96 well plate with an optical, breathable cover seal (shaking at 1096 rpm). Optical density readings were obtained every 10 min.

Protein overexpression cultures. WT yeast were transformed with plasmids encoding Coq5p, Coq8p, Coq9p, Hfd1p, Yjr120w, or Hem25p (p426[2 μ]-GPD plasmids) and cultured on Ura⁻,D plates. Starter cultures (3 mL Ura⁻,D+4HB) were inoculated with an individual colony of yeast and incubated (30 °C, 230 rpm, 10–15 h). For CoQ quantitation, Ura⁻,D+4HB (fermentation) or Ura⁻,GD+4HB (respiration) media (100 mL media at ambient temperature in a sterile 250 mL Erlenmeyer flask) was inoculated with 2.5×10^6 yeast cells and incubated (30 °C, 230 rpm). Samples of the Ura⁻,D+4HB cultures were harvested 13 h after inoculation. Samples of Ura⁻,GD+4HB cultures were harvested 25 h after inoculation. For each growth condition, 1×10^8 yeast cells were pelleted by centrifugation (3,000 g, 3 min, 4 °C), the supernatant was removed, and the cell pellet was flash frozen in N_{2(l)} and stored at -80 °C prior to lipid extractions. For relative growth

rate measurements, analogous cultures (initial density of 5×10^6 cells/mL) were incubated in a sterile 96 well plate with an optical, breathable cover seal (shaking at 1096 rpm). Optical density readings were obtained every 10 min. Respiratory and fermentative growth rates were determined by fitting a linear equation to the linear growth phase and determining the slope of the line.

Cultures for FLAG IPs. BY4742 *S. cerevisiae* overexpressing C-terminally FLAG-tagged genes from p416gpd_Flag plasmids were cultured in Ura⁻,D media (30 °C, 3 mL starter cultures, ~14 h). From these starter cultures, 2.5×10^6 cells were used to inoculate Ura⁻,GD media (100 mL) (respiration culture condition). After incubating 25 hours (30 °C, 230 r.p.m), 7.3×10^8 cells were pelleted by centrifugation (3,000 g, 3 min, 4 °C), the supernatant was removed, and the cell pellet was flash frozen in N_{2(l)} and stored at -80 °C prior to IPs.

Constructs

Yeast gene constructs were generated by amplifying the *Saccharomyces cerevisiae* genes *fmp10*, *mpm1*, *rdl2*, and *ynr040w* from strain BY4742 genomic DNA with primers containing HindIII recognition sequence (forward) and Sall recognition sequence (reverse). Similarly, *hem25* and *yjr120w* were amplified with BamHI (forward) and EcoRI (reverse) primers. *Coq5* (from strain W303) was amplified with SpeI (forward) and Sall or XhoI (reverse) primers. PCR reactions contained 1× Accuprime PCR mix, 1 μM forward primer, 1 μM reverse primer, ~250 ng template, and 1× Accuprime Pfx (Invitrogen cat#12344024). After an initial 2 min denaturation at 95 °C, reactions were exposed to 5

cycles of 95 °C for 15 seconds, 55 °C for 30 seconds, and 68 °C for 2 minutes followed by 30 cycles of 95 °C for 15 seconds, 60 °C for 30 seconds, and 68 °C for 2 minutes. Amplicons were purified using a PCR purification kit (Thermo cat#K0702) and digested with the appropriate restriction enzymes and again subjected to PCR purification. Amplified genes were cloned into restriction enzyme digested yeast expression vectors (p426gpd and/or p416gpd_FLAG). The plasmid p416gpd_FLAG was generated by digesting p416gpd with XhoI and MluI and inserting a double stranded oligonucleotide containing the Flag tag nucleotide sequence and processing XhoI and MluI ends. Recombinants were confirmed by DNA sequencing. Plasmid cloning was previously reported for p426-GPD-*coq8* (Stefely et al., 2015), p426-GPD-*coq9* (Lohman et al., 2014), p426-GPD-*hfd1* (Stefely et al., 2016a), p423(2 μ)-HIS3-PUF3 (Lapointe et al., *in preparation*, see chapter 5), and p413(CEN)-HIS3-PUF3 (Lapointe et al., *in preparation*, see chapter 5).

Lipid Extractions

Frozen pellets of yeast (10^8 cells) were thawed on ice and mixed with glass beads (0.5 mm diameter, 100 μ L). CHCl₃/MeOH (2:1, v/v, 4 °C) (900 μ L) and CoQ₁₀ (10 μ L, 10 μ M, 0.1 nmol) were added and vortexed (2 \times 30 s). HCl (1 M, 200 μ L, 4 °C) was added and vortexed (2 \times 30 s). The samples were centrifuged (5,000 *g*, 2 min, 4 °C) to complete phase separation. 555 μ L of the organic phase was transferred to a clean tube and dried under Ar_(g). The organic residue was reconstituted in ACN/IPA/H₂O (65:30:5, v/v/v) (100 μ L) for LC-MS analysis.

LC-MS Lipid Analysis

LC-MS analysis was performed on an Acquity CSH C18 column held at 50 °C (100 mm × 2.1 mm × 1.7 μm particle size; Waters) using a Vanquish Binary Pump (400 μL/min flow rate; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate in ACN/H₂O (70:30, v/v) containing 250 μL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. Mobile phase B was held at 40% for 6.0 min and then increased to 60% over 3.0. Mobile phase B was further increased to 85% over 0.25 min and then to 99% for over 1.25 min. The column was then reequilibrated for 3.5 min before the next injection. Ten microliters of sample were injected by a Vanquish Split Sampler HT autosampler (Thermo Scientific). The LC system was coupled to a Q Exactive mass spectrometer by a HESI II heated ESI source kept at 325 °C (Thermo Scientific). The inlet capillary was kept at 350 °C, sheath gas was set to 25 units, and auxiliary gas to 10 units, and the spray voltage was set to 3,000 V. The MS was operated in positive and negative parallel reaction monitoring (PRM) mode acquiring scheduled, targeted PRM scans to quantify key CoQ intermediates. Phospholipids were quantified and identified using a negative dd-Top2 scanning mode.

FLAG Protein Co-Immunoprecipitation

Transformed yeast cell pellets were resuspended in Lysis Buffer (400 μL, 4 °C) [20 mM HEPES, pH 7.4, 100 mM NaCl, 10% (w/v) glycerol, 3% (w/v) digitonin (Sigma), 1 mM DTT, protease inhibitors (10 μM benzamidine, 1 μg/mL 1,10-phenanthroline and 0.5 μg/mL each of pepstatin A, chymostatin, antipain, leupeptin, aprotinin; Sigma), phosphatase inhibitors (500 μM imidazole, 250 μM NaF, 300 μM sodium molybdate, 250

μ M sodium orthovanadate, 1 mM sodium tartrate; Sigma)]. The cells were vortexed (6×30 s, with 1 min rests between beating to cool beads) and incubated (10 min, on ice) to lyse the cells. Soluble proteins were isolated by centrifuging (16,000 *g*, 10 min, 4 °C) the lysate to pellet insoluble materials and transferring the supernatant to a clean tube on ice. Anti-FLAG M2 magnetic beads were pre-equilibrated with Final Wash Buffer (20 mM HEPES, pH 7.4, 100 mM NaCl). The soluble protein mixture (equal total masses of protein, 200 μ L) was loaded onto the anti-FLAG beads and incubated (60 min, 4 °C, end-over-end agitation). The beads were washed four times with Wash Buffer (170 μ L, 4 °C) [20 mM HEPES, pH 7.4, 100 mM NaCl, 0.05% (w/v) digitonin, 10% (w/v) glycerol] and once with Final Wash Buffer (170 μ L, 4 °C). Proteins were eluted with Elution Buffer (80 μ L, ~21 °C) (20 mM HEPES, pH 7.4, 100 mM NaCl, 0.2 mg/mL 1 \times FLAG-peptide) (30 min incubation, ~21 °C, 500 rpm shaking) for LC-MS/MS analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

General

As indicated, *P*-values were calculated using an unpaired, two-tailed, Student's *t*-test. All instances where *n* replicates are reported had *n* biological replicates. *P*-values from hypergeometric tests and for Spearman correlation coefficients were calculated using the *R* software suite.

LC-MS Lipid Data Analysis

CoQ intermediate data were processed using TraceFinder 4.0 (Thermo Fisher Scientific). Discovery lipidomic data were processed using an in-house software pipeline and Compound Discoverer 2.0 (Thermo Fisher Scientific).

MEME analyses

For all analyses, the 3' untranslated regions (UTRs) was defined as the longest observed isoform for a gene (Xu et al., 2009) or 200 nts downstream of the stop codon if not previously defined. MEME was run on a local server with the command: `meme [input.txt] -oc [outputdirectory] -dna -mod zoops -evt 0.01 -nmotifs 10 -minw 8 -maxw 12 -maxsize 100000000000`.

HITS-CLIP class definition

RNAs identified via HITS-CLIP were sorted by the number of RNAs detected in their CLIP peak ("peak height") from most to least. Classes were then defined as follows: the top 10% were designated class I; 11-40% class II; 41-70% class III; and 71-100% class IV. Classes were defined to be of comparable size to the analogous RNA Tagging class to facilitate cross-method comparisons (Lapointe et al., *in preparation*, see chapter 5).

Figure 1.

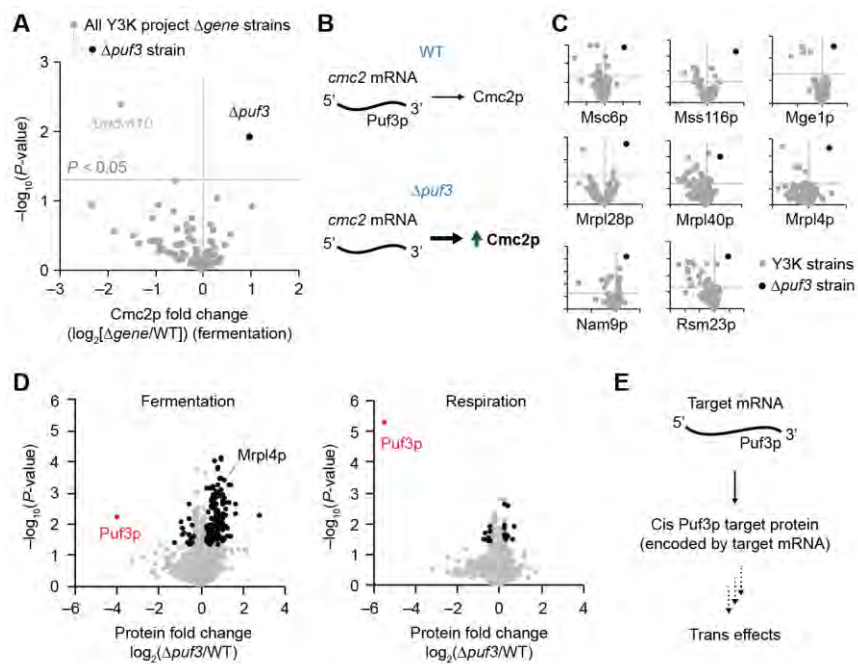


Figure 1. Yeast lacking *puf3* specifically upregulate select mitochondrial proteins

(A) Relative abundance of the protein Cmc2p (mean, $n = 3$) versus statistical significance across yeast strains (fermentation culture condition). Shown in black, the $\Delta puf3$ strain was detected as a “ $\Delta gene$ -specific phenotype” in the Y3K study.

(B) Model for how the lack of Puf3p could cause upregulation of Cmc2p.

(C) Relative abundances of the proteins shown (mean, $n = 3$) versus statistical significance across all strains in the Y3K study (as in A) (fermentation condition). See Figure S1A for expanded plots with labeled axes.

(D) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance in fermentation and respiration conditions. Proteins with fold change (FC) $> 25\%$ and $P < 0.05$ (160 and 24 proteins, respectively) are highlighted in black (two-sided Student's t -test for all panels).

(E) General model for Puf3p cis targets and Puf3p trans effects.

See also Figure S1.

Figure S1.

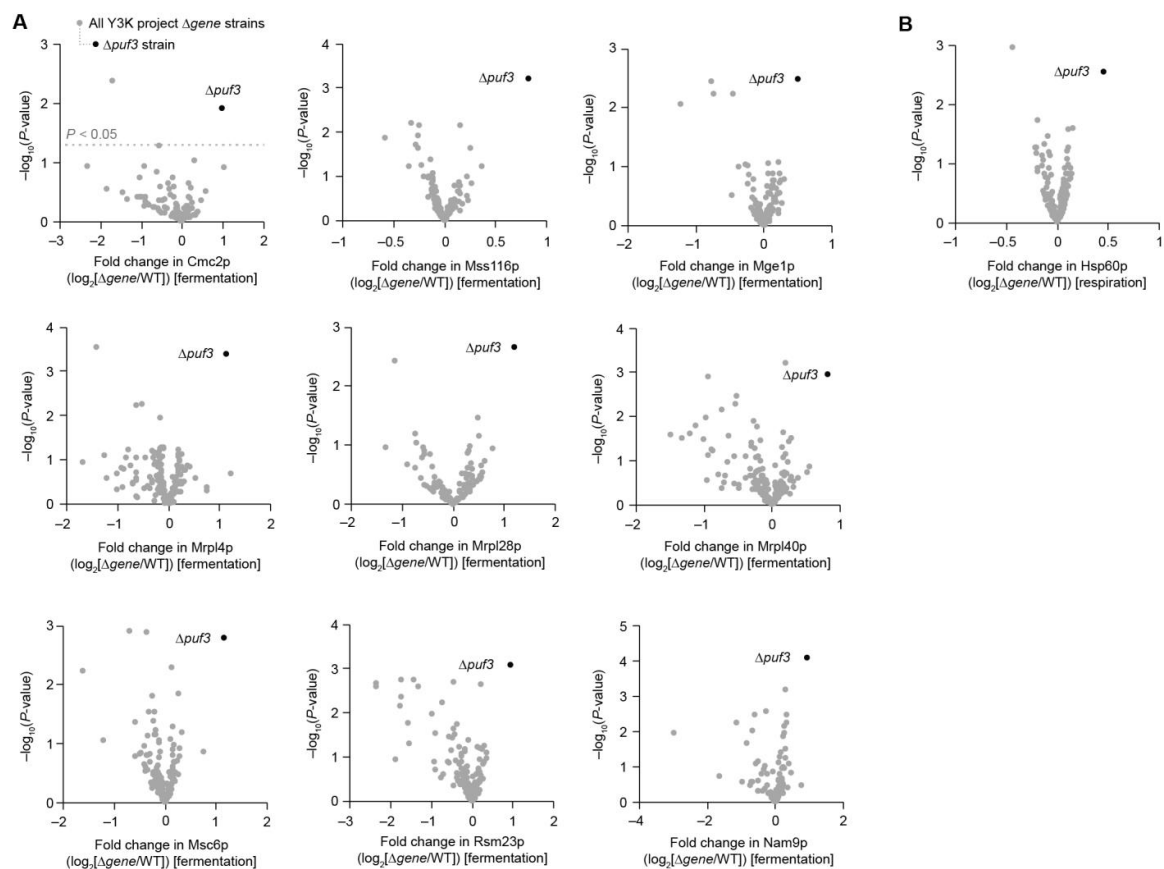


Figure S1. Yeast lacking *puf3* specifically upregulate select mitochondrial proteins

(A) Relative abundances of the proteins shown (mean, $n = 3$) versus statistical significance across all strains in the Y3K study (fermentation culture condition). Shown in black, the $\Delta puf3$ strain was detected as a “ $\Delta gene$ -specific phenotype” in the Y3K study.

(B) Relative abundance of Hsp60p (mean, $n = 3$) versus statistical significance across all strains in the Y3K study (respiration culture condition).

Figure 2.

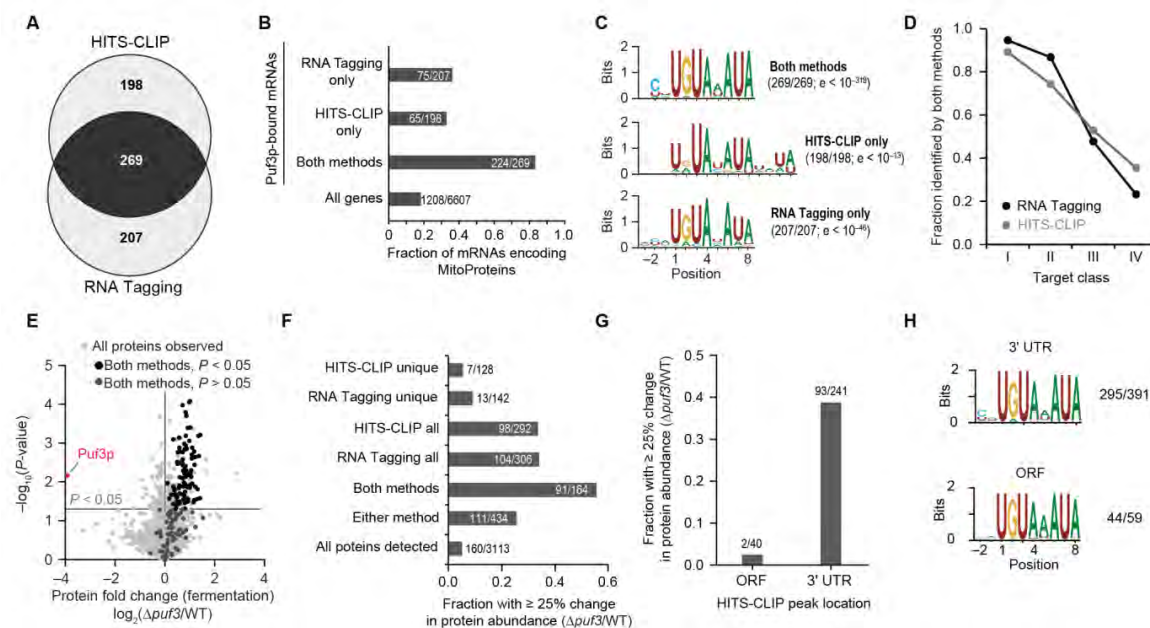


Figure 2. RNA Tagging and HITS-CLIP together identify core Puf3p-bound RNAs

(A) Overlap of RNAs identified as bound by Puf3p via RNA Tagging or HITS-CLIP.

(B) Fraction of genes that are annotated as nuclear-encoded mitochondrial proteins

(“MitoProteins”) for the indicated groups.

(C) Enriched Puf3p-binding elements identified by MEME for the indicated groups of RNAs

bound by Puf3p. Ratios indicate the proportion of 3' UTRs that contribute to each motif, and the

e values are indicated.

(D) Overlap of RNAs identified as bound by Puf3p via both RNA Tagging and HITS-CLIP versus target class for the indicated method.

(E) Relative protein abundances in Δ *puf3* yeast compared to WT (mean, $n = 3$) versus statistical significance (fermentation condition), highlighting proteins detected as Puf3p-bound mRNAs by

both RNA Tagging and HITS-CLIP with $P < 0.05$ or $P > 0.05$ (91 and 74 proteins, respectively).

P-value cutoffs are for protein abundance changes (two-sided Student's *t*-test). 104 proteins encoded by Puf3-bound mRNAs were not observed in the Y3K proteomics study.

(F) Fraction of genes with at least a 25% change in protein abundance ($P < 0.05$, two-sided

Student's *t*-test) for the indicated groups. Analyses were limited to genes with proteins detected in the Y3K proteomics study, which is the denominator of each ratio.

(G) Fraction of genes with at least a 25% change in protein abundance ($P < 0.05$, two-sided

Student's *t*-test) for Puf3p-bound RNAs identified with HITS-CLIP peaks in their open-reading

frame (ORF) or 3' untranslated region (UTR). Analyses were limited to genes detected in the

Y3K proteomics study and identified via a single HITS-CLIP peak (ORF, 40 genes; 3' UTR, 241 genes).

(H) Position-weight matrices of PBEs identified under ORF or 3' UTR HITS-CLIP peaks.

See also Figure S2.

Figure S2.

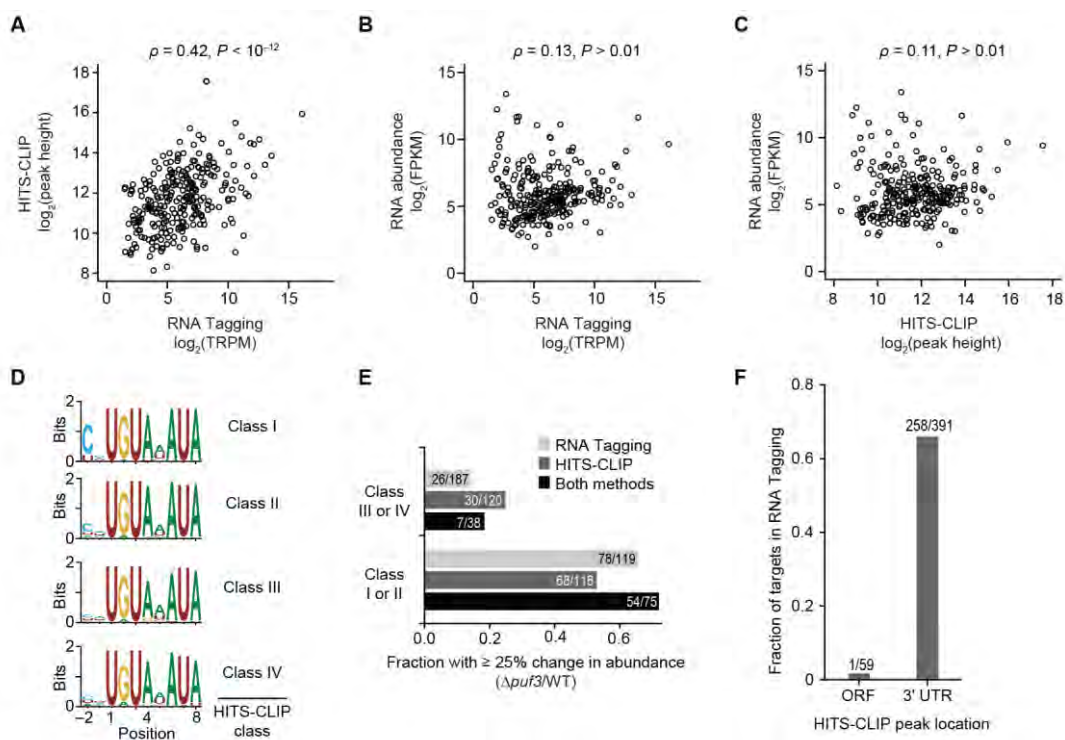


Figure S2. Comparison of RNA Tagging and HITS-CLIP with Puf3p

(A) The number of U-tagged reads per million uniquely mapped reads (TRPM) versus HITS-CLIP peak height for each Puf3p-bound RNA identified via both methods. Spearman's (ρ) correlation coefficient and associated P -value is indicated.

(B) TRPM detected in RNA Tagging versus RNA abundance (FPKM, fragments per kilobase of transcript per million mapped reads) for Puf3p-bound RNAs identified via both methods.

Spearman's (ρ) correlation coefficient and associated P -value is indicated.

(C) Peak height detected in HITS-CLIP versus RNA abundance (FPKM) for Puf3p-bound RNAs identified via both methods. Spearman's (ρ) correlation coefficient and associated P -value is indicated.

(D) Fraction of genes with at least a 25% change in protein abundance ($P < 0.05$, two-sided Student's t -test) for the indicated groups. Analyses were limited to genes with proteins detected in the Y3K proteomics study, which is the denominator of each ratio.

(E) Position-weight matrices of PBEs identified under peaks for each HITS-CLIP class.

(F) Fraction of Puf3p-bound RNAs with HITS-CLIP peaks in their open-reading frame (ORF) or 3' untranslated region (UTR) also identified by RNA Tagging. Analyses were limited to genes detected via a single HITS-CLIP peak (ORF, 59 genes; 3' UTR, 391 genes).

Figure 3.

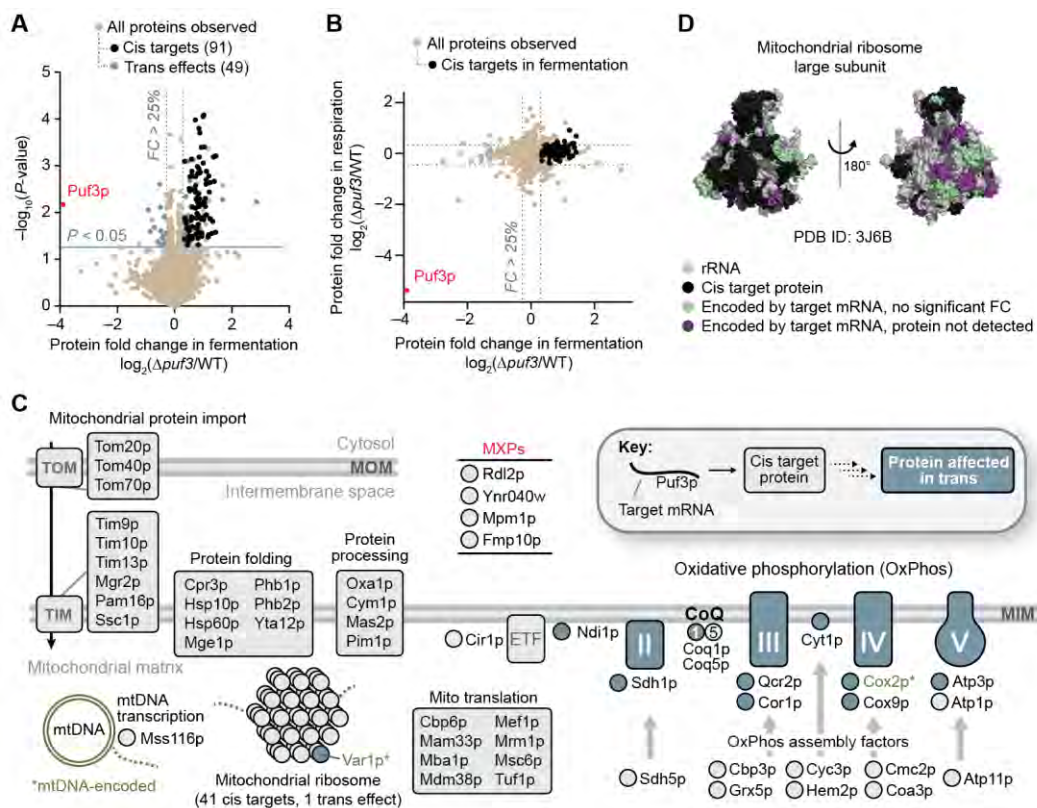


Figure 3. Puf3p targets a set of proteins that generate the mitochondrial respiratory chain

(A) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (fermentation condition), highlighting proteins with FC > 25% and $P < 0.05$ (two-sided Student's t -test) that were either identified as Puf3p targets by both RNA methods (cis targets, black) or neither RNA method (trans effects, light blue).

(B) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) in fermentation versus respiration, highlighting all fermentation Puf3p cis targets in black.

(C) Cartoon indicating 85 of 91 Puf3p cis targets that fit into a biological program that generates the mitochondrial respiratory chain. Select Puf3p trans effect proteins and 4 Puf3p-regulated MXPs are also shown. All Puf3p targets are shown in Figure S3F.

(D) Surface representation of the large subunit of the yeast mitochondrial ribosome (PDB: 3J6B) with Puf3p targets indicated.

See also Figure S3.

Figure S3.

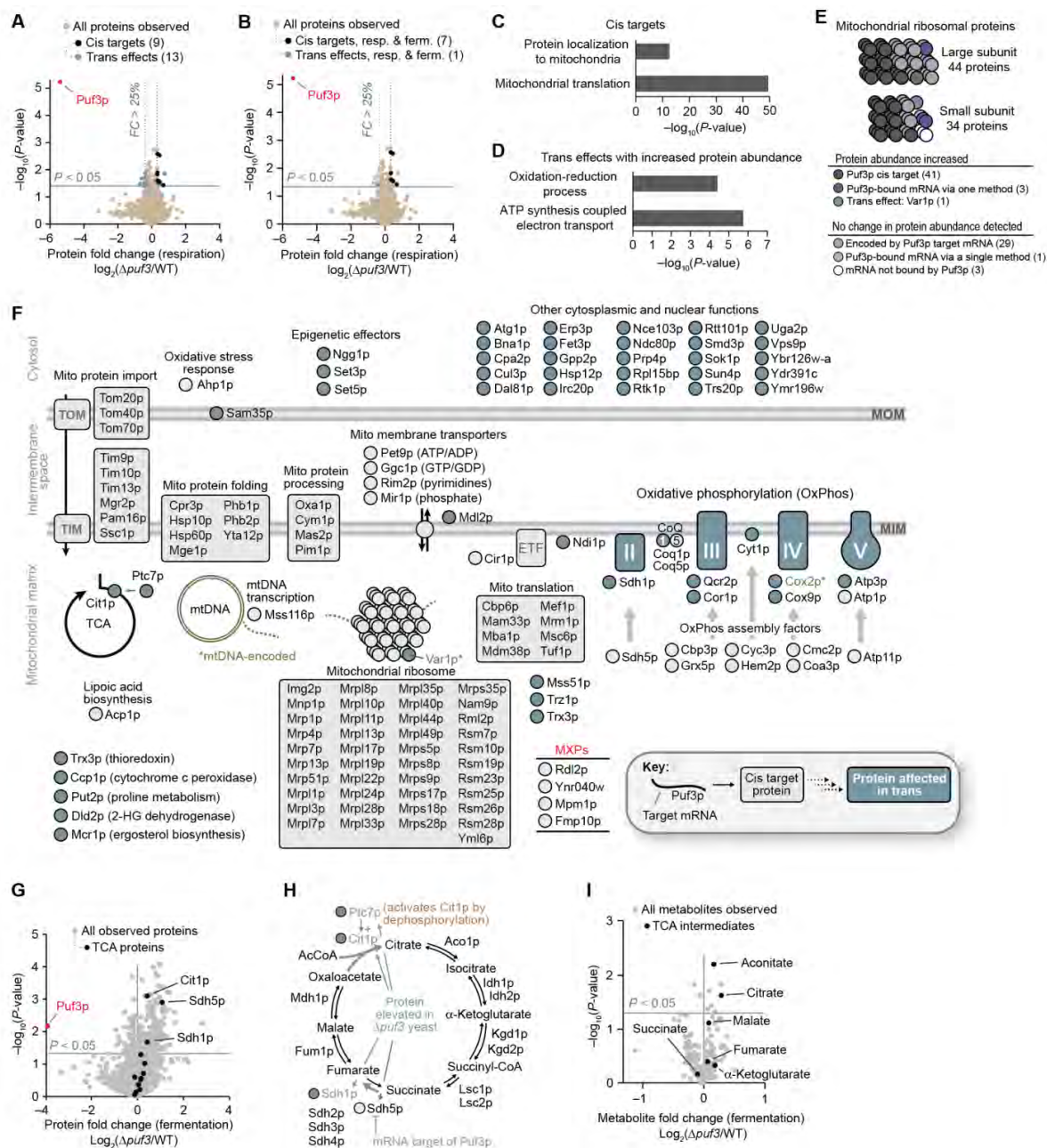


Figure S3. Puf3p targets a set of proteins that generate the mitochondrial respiratory chain

(A) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (respiration condition), highlighting proteins with FC > 25% and $P < 0.05$ (two-sided Student's t -test) that were either identified as Puf3p targets by both RNA methods (cis targets, black) or neither RNA method (trans effects, light blue).

(B) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (respiration condition), highlighting proteins with FC > 25% and $P < 0.05$ (two-sided Student's t -test) in both the fermentation and respiration data sets that are Puf3p targets by both RNA methods (cis targets, black) or neither method (trans effects, light blue).

(C) Gene Ontology (GO) terms enriched in the cis Puf3p target list.

(D) GO terms enriched in elevated Puf3p trans effect proteins.

(E) Cartoon of annotated mitochondrial ribosomal proteins with the effect by Puf3p indicated.

(F) Cartoon indicating all cis Puf3p targets and all Puf3p trans effect proteins (fermentation).

(G) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (fermentation condition), highlighting TCA proteins.

(H) Scheme of the TCA cycle highlighting proteins significantly ($P < 0.05$) elevated in $\Delta puf3$ yeast.

(I) Relative metabolite abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (fermentation condition), highlighting TCA cycle metabolites.

Figure 4.

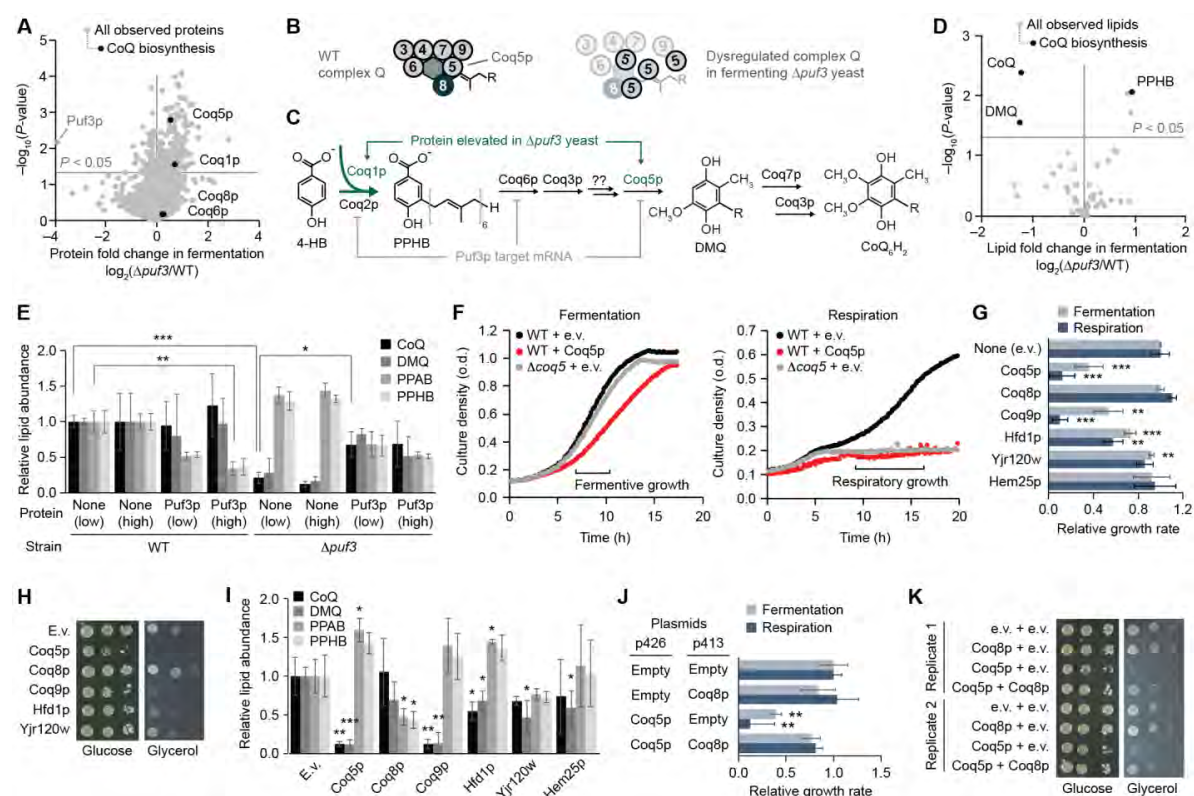


Figure 4. Puf3p regulates coenzyme Q biosynthesis enzymes

(A) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (fermentation condition), highlighting all observed CoQ biosynthesis proteins.

(B) Cartoon model for how Coq5p overexpression could dysregulate complex Q.

(C) Scheme of CoQ biosynthesis with Puf3p targets indicated. 4-HB, 4-hydroxybenzoate; PPHB, polyprenylhydroxybenzoate; DMQ, demethoxy-CoQ.

(D) Relative lipid abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (fermentation condition).

(E) Relative lipid abundances in yeast transformed with plasmids overexpressing Puf3p (high or low plasmid copy number; “none” indicates empty vector) and cultured in fermentation media. PPAB, polyprenylaminobenzoate (mean \pm SD, $n = 3$).

(F) Yeast growth curves for various strains transformed with plasmids overexpressing the plasmids shown (e.v., empty vector) and cultured in either fermentation or respiration media.

(G) Growth rates of WT yeast transformed with plasmids overexpressing the proteins shown and cultured in either fermentation or respiration media (mean \pm SD, $n = 3$).

(H) Serial dilutions of WT yeast transformed with plasmids overexpressing the proteins shown and cultured on solid media containing either glucose or glycerol.

(I) Relative lipid abundances in WT yeast transformed with plasmids overexpressing the proteins shown and cultured in respiration media (mean \pm SD, $n = 3$).

(J) Growth rates of WT yeast transformed with plasmids overexpressing the proteins shown and cultured in either fermentation or respiration media (mean \pm SD, $n = 3$).

(K) Serial dilutions of WT yeast transformed with plasmids overexpressing the proteins shown and cultured on solid media containing either glucose or glycerol.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (two-sided Student's t -test for all panels).

See also Figure S4.

Figure S4.

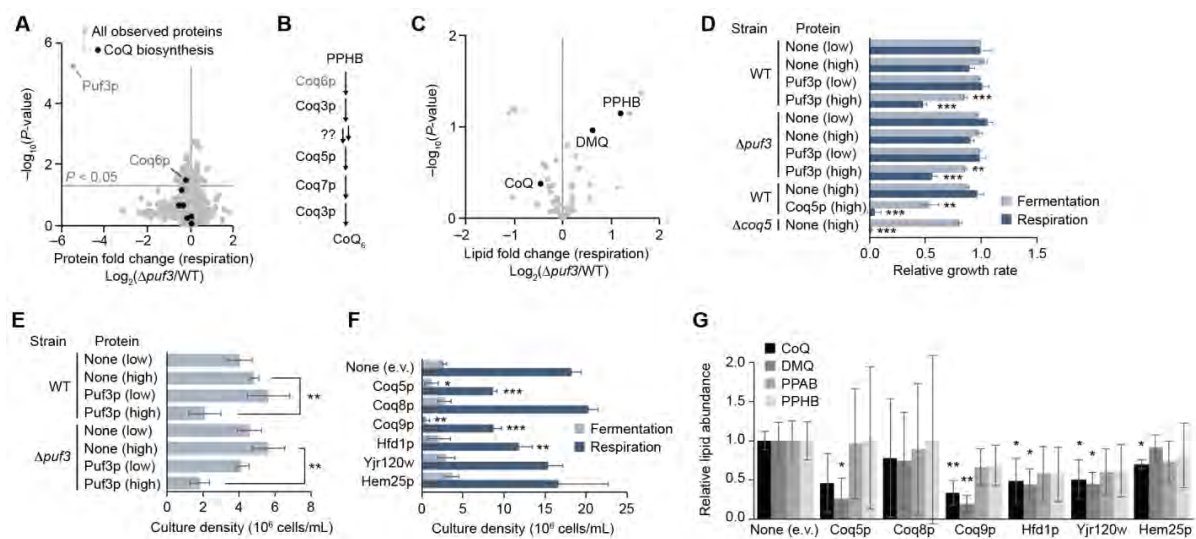


Figure S4. Puf3p regulates coenzyme Q biosynthesis enzymes

(A) Relative protein abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (respiration culture condition), highlighting all observed CoQ biosynthesis proteins.

(B) Scheme of CoQ biosynthesis.

(C) Relative lipid abundances in $\Delta puf3$ yeast compared to WT (mean, $n = 3$) versus statistical significance (respiration culture condition).

(D) Growth rates of various yeast strains transformed with plasmids overexpressing the proteins shown and cultured in either fermentation or respiration media (mean \pm SD, $n = 3$).

(E) Densities of yeast cultures at the time point of harvest for the lipid quantitation experiments depicted in Figure 4E (fermentation culture condition).

(F) Densities of yeast cultures at the time point of harvest for the lipid quantitation experiments depicted in Figures 4I (respiration culture condition) and S4G (fermentation culture condition).

(G) Relative lipid abundances in WT yeast transformed with plasmids overexpressing the proteins shown and cultured in fermentation media (mean \pm SD, $n = 3$).

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (two-sided Student's t -test for all panels).

Figure 5.

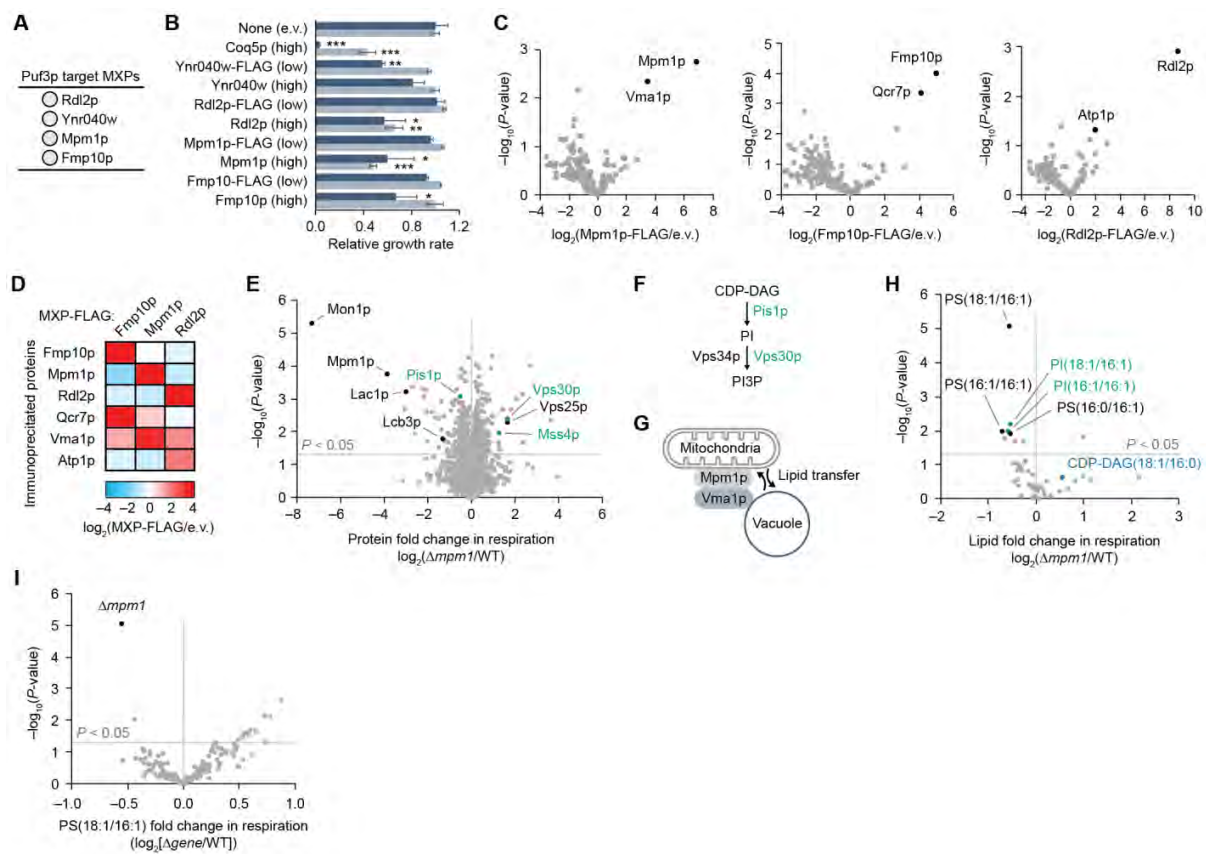


Figure 5. Functions of Puf3p-regulated mitochondrial uncharacterized proteins (MXPs)

(A) The four mitochondrial uncharacterized proteins (MXPs) that are cis Puf3p targets.

(B) Relative growth rates of WT yeast transformed with plasmids overexpressing the proteins shown and cultured in either fermentation (light blue bars) or respiration (dark blue bars) media (mean \pm SD, $n = 3$).

(C) Relative abundances of proteins co-purifying with the MXP-FLAG constructs shown compared with empty vector (e.v.) control (mean, $n = 3$) immunoprecipitated from yeast as assessed by LC-MS/MS.

(D) Heat map showing relative abundances of select proteins co-purifying with the MXP-FLAG constructs shown compared with e.v. (mean, $n = 3$) immunoprecipitated from yeast cells as assessed by LC-MS/MS.

(E) Relative protein abundances in $\Delta mpm1$ yeast compared to WT (mean, $n = 3$) versus statistical significance (respiration condition), highlighting select proteins related to lipid metabolism and vacuolar function.

(F) Scheme of phosphatidylinositol (PI) lipid metabolism.

(G) Model for how a physical interaction between Mpm1p and Vma1p could facilitate lipid transfer between vacuoles and mitochondria.

(H) Relative lipid abundances in $\Delta mpm1$ yeast compared to WT (mean, $n = 3$) versus statistical significance (respiration condition).

(I) Relative abundance of the lipid PS(18:1/16:1) (mean, $n = 3$) versus statistical significance across yeast strains (respiration culture condition) in the Y3K study.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (two-sided Student's t -test for all panels).

See also Figure S5.

Figure S5.

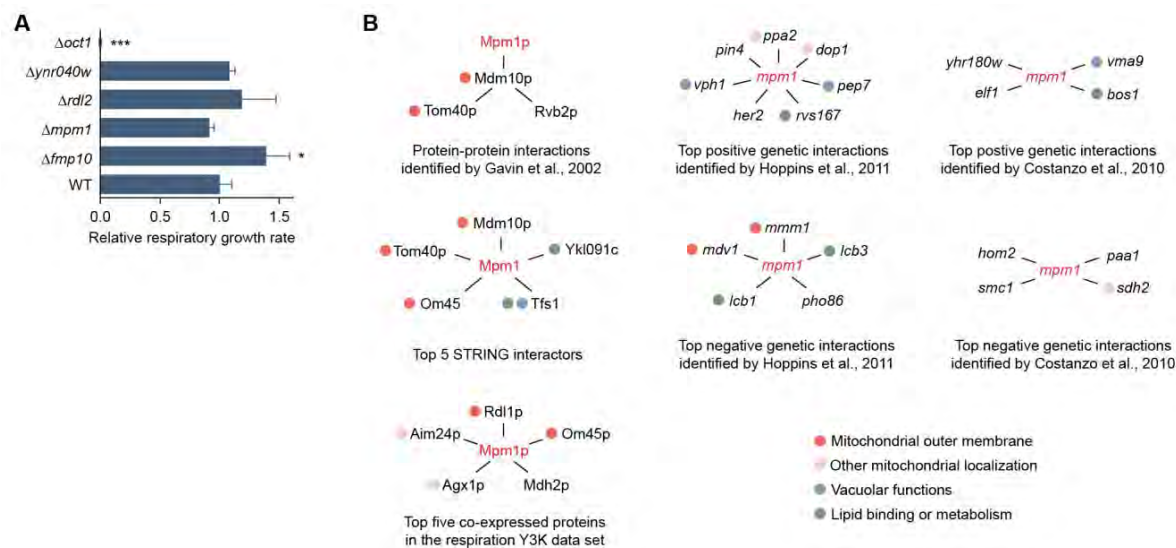


Figure S5. Functions of Puf3p-regulated mitochondrial uncharacterized proteins (MXPs)

(A) Relative growth rates of yeast strains cultured in respiration media (mean \pm SD, $n = 3$).

(B) Summary of Mpm1p genetic and proteomic interactions revealed from an examination of publicly available genetic and proteomic data sets.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (two-sided Student's t -test for all panels).

REFERENCES

- Barkovich, R.J., Shtanko, A., Shepherd, J.A., Lee, P.T., Myles, D.C., Tzagoloff, A., and Clarke, C.F. (1997). Characterization of the COQ5 gene from *Saccharomyces cerevisiae*. Evidence for a C-methyltransferase in ubiquinone biosynthesis. *J Biol Chem* 272, 9182-9188.
- Brandl, C.J., Furlanetto, A.M., Martens, J.A., and Hamilton, K.S. (1993). Characterization of Ngg1, a Novel Yeast Gene Required for Glucose Repression of Gal4p-Regulated Transcription. *Embo J* 12, 5255-5265.
- Chatenay-Lapointe, M., and Shadel, G.S. (2011). Repression of mitochondrial translation, respiration and a metabolic cycle-regulated gene, SLF1, by the yeast Pumilio-family protein Puf3p. *PLoS One* 6, e20441.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010). The genetic landscape of a cell. *Science* 327, 425-431.
- Couvillion, M.T., Soto, I.C., Shipkovenska, G., and Churchman, L.S. (2016). Synchronized mitochondrial and cytosolic translation programs. *Nature* 533, 499-503.
- Crivellone, M.D., Wu, M.A., and Tzagoloff, A. (1988). Assembly of the mitochondrial membrane system. Analysis of structural mutants of the yeast coenzyme QH₂-cytochrome c reductase complex. *J Biol Chem* 263, 14323-14333.
- Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* 1, 266-286.
- Dibrov, E., Robinson, K.M., and Lemire, B.D. (1997). The COQ5 gene encodes a yeast mitochondrial protein necessary for ubiquinone biosynthesis and the assembly of the respiratory chain. *J Biol Chem* 272, 9175-9181.
- Elbaz-Alon, Y., Rosenfeld-Gur, E., Shinder, V., Futerman, A.H., Geiger, T., and Schuldiner, M. (2014). A dynamic interface between vacuoles and mitochondria in yeast. *Dev Cell* 30, 95-102.
- Eliyahu, E., Pnueli, L., Melamed, D., Scherrer, T., Gerber, A.P., Pines, O., Rapaport, D., and Arava, Y. (2010). Tom20 Mediates Localization of mRNAs to Mitochondria in a Translation-Dependent Manner. *Mol Cell Biol* 30, 284-294.
- Floyd, B.J., Wilkerson, E.M., Veling, M.T., Minogue, C.E., Xia, C., Beebe, E.T., Wrobel, R.L., Cho, H., Kremer, L.S., Alston, C.L., *et al.* (2016). Mitochondrial Protein Interaction Mapping Identifies Regulators of Respiratory Chain Function. *Mol Cell* 63, 621-632.
- Gadir, N., Haim-Vilmovsky, L., Kraut-Cohen, J., and Gerst, J.E. (2011). Localization of mRNAs coding for mitochondrial proteins in the yeast *Saccharomyces cerevisiae*. *RNA* 17, 1551-1565.
- Garcia-Rodriguez, L.J., Gay, A.C., and Pon, L.A. (2007). 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.
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., *et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-147.
- Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS biology* 2, E79.

Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* 20, 1425.

He, C.H., Xie, L.X., Allan, C.M., Tran, U.C., and Clarke, C.F. (2014). Coenzyme Q supplementation or over-expression of the yeast Coq8 putative kinase stabilizes multi-subunit Coq polypeptide complexes in yeast coq null mutants. *Biochim Biophys Acta* 1841, 630-644.

Hebert, A.S., Merrill, A.E., Stefely, J.A., Bailey, D.J., Wenger, C.D., Westphall, M.S., Pagliarini, D.J., and Coon, J.J. (2013). Amine-reactive neutron-encoded labels for highly plexed proteomic quantitation. *Mol Cell Proteomics* 12, 3360-3369.

Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D., and Brown, P.O. (2008). Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. *Plos Biol* 6, 2297-2313.

Hogan, G.J., Brown, P.O., and Herschlag, D. (2015). Evolutionary Conservation and Diversification of Puf RNA Binding Proteins and Their mRNA Targets. *Plos Biol* 13.

Honscher, C., Mari, M., Auffarth, K., Bohnert, M., Griffith, J., Geerts, W., van der Laan, M., Cabrera, M., Reggiori, F., and Ungermann, C. (2014). Cellular metabolism regulates contact sites between vacuoles and mitochondria. *Dev Cell* 30, 86-94.

Hoppins, S., Collins, S.R., Cassidy-Stone, A., Hummel, E., Devay, R.M., Lackner, L.L., Westermann, B., Schuldiner, M., Weissman, J.S., and Nunnari, J. (2011). A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J Cell Biol* 195, 323-340.

Horn, D., Zhou, W., Trevisson, E., Al-Ali, H., Harris, T.K., Salviati, L., and Barrientos, A. (2010). The conserved mitochondrial twin Cx9C protein Cmc2 Is a Cmc1 homologue essential for cytochrome c oxidase biogenesis. *J Biol Chem* 285, 15088-15099.

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

Jackson, J.S., Jr., Houshmandi, S.S., Lopez Leban, F., and Olivas, W.M. (2004). Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA* 10, 1625-1636.

Jiang, H.F., Guan, W.J., and Gu, Z.L. (2010). Tinkering Evolution of Post-Transcriptional RNA Regulons: Puf3p in Fungi as an Example. *Plos Genet* 6.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* 8, 533-543.

Kershaw, C.J., Costello, J.L., Talavera, D., Rowe, W., Castelli, L.M., Sims, P.F.G., Grant, C.M., Ashe, M.P., Hubbard, S.J., and Pavitt, G.D. (2015). Integrated multi-omics analyses reveal the pleiotropic nature of the control of gene expression by Puf3p. *Sci Rep-Uk* 5.

Klass, D.M., Scheibe, M., Butter, F., Hogan, G.J., Mann, M., and Brown, P.O. (2013). Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Res* 23, 1028-1038.

Konig, J., Zarnack, K., Luscombe, N.M., and Ule, J. (2012). Protein-RNA interactions: new genomic technologies and perspectives (vol 13, pg 77, 2012). *Nat Rev Genet* 13, 221-221.

Koopman, W.J., Willems, P.H., and Smeitink, J.A. (2012). Monogenic mitochondrial disorders. *N Engl J Med* 366, 1132-1141.

Labbe, K., Murley, A., and Nunnari, J. (2014). Determinants and functions of mitochondrial behavior. *Annu Rev Cell Dev Biol* 30, 357-391.

Lapointe, C.P., Wilinski, D., Saunders, H.A., and Wickens, M. (2015). Protein-RNA networks revealed through covalent RNA marks. *Nat Methods* 12, 1163-1170.

Laredj, L.N., Licitra, F., and Puccio, H.M. (2014). The molecular genetics of coenzyme Q biosynthesis in health and disease. *Biochimie* 100, 78-87.

Lee, C.D., and Tu, B.P. (2015). Glucose-Regulated Phosphorylation of the PUF Protein Puf3 Regulates the Translational Fate of Its Bound mRNAs and Association with RNA Granules. *Cell Rep* 11, 1638-1650.

Lee, D., Ohn, T., Chiang, Y.C., Quigley, G., Yao, G., Liu, Y., and Denis, C.L. (2010). PUF3 acceleration of deadenylation in vivo can operate independently of CCR4 activity, possibly involving effects on the PAB1-mRNP structure. *J Mol Biol* 399, 562-575.

Licatalosi, D.D., and Darnell, R.B. (2010). APPLICATIONS OF NEXT-GENERATION SEQUENCING RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* 11, 75-87.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., *et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464-469.

Lohman, D.C., Forouhar, F., Beebe, E.T., Stefely, M.S., Minogue, C.E., Ulbrich, A., Stefely, J.A., Sukumar, S., Luna-Sanchez, M., Jochem, A., *et al.* (2014). Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis. *Proc Natl Acad Sci U S A* 111, E4697-4705.

McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W.F., and Rosbash, M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 165, 742-753.

Miller, M.A., and Olivas, W.M. (2011). Roles of Puf proteins in mRNA degradation and translation. *Wires Rna* 2, 471-492.

Miller, M.A., Russo, J., Fischer, A.D., Lopez Leban, F.A., and Olivas, W.M. (2014). Carbon source-dependent alteration of Puf3p activity mediates rapid changes in the stabilities of mRNAs involved in mitochondrial function. *Nucleic Acids Res* 42, 3954-3970.

Muller-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.

Nunnari, J., and Suomalainen, A. (2012). Mitochondria: in sickness and in health. *Cell* 148, 1145-1159.

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

Payet, L.A., Leroux, M., Willison, J.C., Kihara, A., Pelosi, L., and Pierrel, F. (2016). Mechanistic Details of Early Steps in Coenzyme Q Biosynthesis Pathway in Yeast. *Cell Chem Biol* 23, 1241-1250.

Pray-Grant, M.G., Schieltz, D., McMahon, S.J., Wood, J.M., Kennedy, E.L., Cook, R.G., Workman, J.L., Yates, J.R., and Grant, P.A. (2002). The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. *Mol Cell Biol* 22, 8774-8786.

Quenault, T., Lithgow, T., and Traven, A. (2011). PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol* 21, 104-112.

- Quinzii, C.M., and Hirano, M. (2010). Coenzyme Q and mitochondrial disease. *Dev Disabil Res Rev* 16, 183-188.
- Rowe, W., Kershaw, C.J., Castelli, L.M., Costello, J.L., Ashe, M.P., Grant, C.M., Sims, P.F., Pavitt, G.D., and Hubbard, S.J. (2014). Puf3p induces translational repression of genes linked to oxidative stress. *Nucleic Acids Res* 42, 1026-1041.
- Saint-Georges, Y., Garcia, M., Delaveau, T., Jourden, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F., and Jacq, C. (2008). Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One* 3, e2293.
- Saleh, A., Lang, V., Cook, R., and Brandl, C.J. (1997). Identification of native complexes containing the yeast coactivator/repressor proteins NGG1/ADA3 and ADA2. *Journal of Biological Chemistry* 272, 5571-5578.
- Schoenberg, D.R., and Maquat, L.E. (2012). Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* 13, 246-259.
- Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* 84, 325-354.
- Spasov, D.S., and Jurecic, R. (2003). The PUF family of RNA-binding proteins: Does evolutionarily conserved structure equal conserved function? *Iubmb Life* 55, 359-366.
- Stefely, J.A., Kwiecien, N.W., Freiberger, E.C., Richards, A.L., Jochem, A., Rush, M.J., Ulbrich, A., Robinson, K.P., Hutchins, P.D., Veling, M.T., *et al.* (2016a). Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. *Nat Biotechnol.*
- Stefely, J.A., Licitra, F., Laredj, L., Reidenbach, A.G., Kemmerer, Z.A., Grangeray, A., Jaeg-Ehret, T., Minogue, C.E., Ulbrich, A., Hutchins, P.D., *et al.* (2016b). Cerebellar Ataxia and Coenzyme Q Deficiency through Loss of Unorthodox Kinase Activity. *Mol Cell* 63, 608-620.
- Stefely, J.A., Reidenbach, A.G., Ulbrich, A., Oruganty, K., Floyd, B.J., Jochem, A., Saunders, J.M., Johnson, I.E., Minogue, C.E., Wrobel, R.L., *et al.* (2015). Mitochondrial ADCK3 employs an atypical protein kinase-like fold to enable coenzyme Q biosynthesis. *Mol Cell* 57, 83-94.
- Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell* 52, 52-62.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., *et al.* (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43, D447-452.
- Takeda, M., Chen, W.J., Saltzgaber, J., and Douglas, M.G. (1986). Nuclear genes encoding the yeast mitochondrial ATPase complex. Analysis of ATP1 coding the F1-ATPase alpha-subunit and its assembly. *J Biol Chem* 261, 15126-15133.
- Taylor, J.W., and Berbee, M.L. (2006). Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* 98, 838-849.
- Tran, U.C., and Clarke, C.F. (2007). Endogenous synthesis of coenzyme Q in eukaryotes. *Mitochondrion* 7 Suppl, S62-71.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212-1215.

Vafai, S.B., and Mootha, V.K. (2012). Mitochondrial disorders as windows into an ancient organelle. *Nature* 491, 374-383.

Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* 18, 150-157.

Xu, Z.Y., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-U1037.

Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M., and Hall, T.M. (2009). A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc Natl Acad Sci U S A* 106, 20192-20197.

CHAPTER 5:**Principles that control the architecture of a protein-RNA network**

Christopher P. Lapointe¹, Melanie A. Preston¹, Daniel Wilinski^{1,2}, Harriet A. J. Saunders¹, Zachary T. Campbell^{1,3}, and Marvin Wickens¹

¹ Department of Biochemistry, University of Wisconsin-Madison, Madison, WI.

² Present address: Life Sciences Institute, University of Michigan, Ann Arbor

³ Present address: Department of Biological Sciences, University of Texas at Dallas,
Richardson, TX

This chapter is currently being prepared for publication, in which I am the first author.

ABSTRACT

Protein-RNA networks comprised of multiple proteins and the RNAs they bind are widespread and critical in biological control. Relatively little is known, however, about how such networks are coordinated. To uncover principles that guide their formation and balance, we examined the network controlled by three *S. cerevisiae* proteins – Puf3p, Puf4p, and Puf5p – as they share similar structures, and bind similar yet distinct RNA sequences. Using RNA Tagging in tandem with meta-analyses, we constructed a map of the network of RNAs controlled by Puf3p, Puf4p, and Puf5p. We identified four primary sub-networks in the “PUF supra-network”: the Puf3p, Puf4p, and Puf5p sub-networks, and a sub-network controlled by both Puf4p and Puf5p. Individual sub-networks are balanced via an interplay between abundance and relative binding affinities. Removal of a node in the PUF supra-network (e.g. Puf4p or Puf5p) resulted in large-scale rewiring of individual sub-networks. Our findings reveal key determinants that control the architecture of protein-RNA networks encompassing multiple proteins.

INTRODUCTION

Proteins and RNAs form highly interconnected networks of interactions that underlie a wide-range of processes and whose dysfunction cause disease^{1,2}. Humans possess more than 1,500 RNA-binding proteins and 20,000 protein-coding genes, which can be alternatively spliced to create multiple, distinct isoforms^{3,4}. Single RNA-binding proteins (RBPs) often bind to hundreds of individual RNAs, and a single RNA molecule can be bound by multiple proteins simultaneously, whose fate is dictated by the particular combination of bound proteins⁵. Despite this, little is known about how multiple proteins work together to regulate RNAs *in vivo*. We sought to uncover principles that govern the formation and balance of an RNA regulatory network composed of multiple RBPs.

PUF proteins are a versatile family of mRNA regulators. They are conserved throughout Eukarya and implicated in the regulation of early development, stem cells, differentiation, the nervous system, and cancer⁶⁻⁸. Individual PUF proteins can bind hundreds of mRNAs through specific sequences present in their 3' untranslated regions (UTRs)⁹⁻²⁰. PUF proteins recruit other proteins to control mRNA stability, translation, and localization²¹⁻²⁶.

In *Saccharomyces cerevisiae*, three PUF proteins – Puf3p, Puf4p, and Puf5p – share very similar structures and bind to similar yet distinct RNA sequences. These “canonical” PUF proteins possess the traditional eight PUF repeats, which fold into a stereotypical crescent shape^{18,27-29}. Each protein binds to RNA sequences characterized by a 5' UGUA followed by a downstream 3' UA. Despite the similarity, each protein prefers distinct length binding elements dictated by the number of nucleotides between the 5' and 3' features, which we

refer to as “spacer nucleotides”. Puf3p preferentially binds sequence elements eight nucleotides in length (“8BE”, two spacer nts)^{9,17,27}. In contrast, Puf4p preferentially binds sequence elements nine nucleotides in length (“9BE”, three spacer nts)^{9,23,28-30}, and Puf5p binds to 9BEs as well as sequence elements ten nucleotides in length (“10BE”, four spacer nts)^{9,18,28,30}. Thus, canonical yeast PUF proteins are a great model to study how related proteins are integrated into RNA regulatory networks at the cellular level.

We sought to understand how the individual protein-RNA “sub-networks” controlled by Puf3p, Puf4p, and Puf5p are integrated into a larger “supra-network” *in vivo*. We combined molecular, genetic, and bioinformatic approaches, including RNA Tagging, to define the sub-networks controlled by each proteins and subsequently used a meta-analysis to determine the “PUF supra-network”. Enabled by our map, we demonstrate how the sub-networks are balanced and coordinated *in vivo*.

RESULTS

We employed RNA Tagging, which we recently developed¹⁷, to identify RNAs bound by Puf4p and Puf5p. RNA Tagging exploits a poly(U) polymerase (PUP) to covalently “U-tag” the RNAs bound by a protein of interest *in vivo*. The U-tagged RNAs are then identified via high-throughput sequencing (**Fig. 1A**). We constructed “PUF4-PUP” and “PUF5-PUP” strains of *S. cerevisiae*, in which the open-reading frame of *C. elegans* PUP-2 was fused to the 3′ end of *PUF4* or *PUF5* at their endogenous genomic loci, respectively. We cultured these strains to mid-log phase, lysed the cells under denaturing conditions, isolated total RNA, prepared high-throughput sequencing libraries, and sequenced DNA libraries on an Illumina platform to obtain paired-end reads. Following sequencing, we identified U-tagged RNAs present in each strain, which we defined as RNAs that contained at least eight adenosines (the poly(A) tail) followed by at least one uridine (the U-tag) not encoded in the genome.

The Puf4p sub-network

In PUF4-PUP yeast, we identified 507 mRNAs that were reproducibly U-tagged, which we hereafter refer to as “Puf4p targets” (**Supplementary Fig. 1A**). To determine whether Puf4p targets were enriched for Puf4p- binding elements, we searched the 3′ UTRs of Puf4p targets for enriched sequence elements using the algorithm Multiple EM for Motif Elicitation (MEME)³¹. Puf4p targets were highly enriched for a nine nucleotide long sequence element characterized by a 5′ UGUA and degenerate 3′ UA, which is consistent with the expected 9BE (**Supplementary Fig. 1B**)^{9,28,30}. By Gene Ontology (GO) analyses, Puf4p targets are enriched for mRNAs encoding proteins involved in the

processing of rRNA and ribosome biogenesis, similar to a previous report (**Supplementary Fig. 1C**)⁹. Consistently, Puf4p targets were significantly enriched for mRNAs encoding proteins that localize to the nucleolus (**Supplementary Fig. 1D**). Thus, Puf4p binds mRNAs enriched for Puf4p-binding elements that are implicated in the regulation of ribosome biogenesis.

To fully exploit our RNA Tagging data, we separated Puf4p targets into four groups, which we call “classes”. RNA Tagging provides two attributes for every Puf4p target: the number of U-tagged RNAs detected per million uniquely mapped reads (TRPM) and the number of uridines in the U-tag on every tagged RNA. To leverage this, we used k-means clustering to separate Puf4p targets into eight groups based on the number of TRPM detected at increasing U-tag lengths (from at least one U to at least eight U's) (see Methods). We sorted the k-means groups from longest to shortest U-tags, plotted the results on a heat map, and assigned groups of mRNAs to classes (**Fig. 1B**). Class I targets were detected by the most TRPMs and many U-tags of up to seven or eight U's. In contrast, Class IV targets were detected by the fewest TRPMs and rarely had U-tags longer than one or two U's.

Puf4 target class correlated with enrichment for high-affinity Puf4p-binding elements. Nearly all class I targets possessed consensus 9BEs, and the 9BE progressively degenerated from class I to class IV targets (**Fig. 1B**). Consistently, Puf4p preferentially binds 9BEs rather than 8BEs or 10BEs *in vitro* via a global SEQRS analysis¹⁸, which determines the relative binding affinity of a Puf4p for millions of sequences

simultaneously³⁰. By re-examining the Puf4p SEQRS data, we found that class I Puf4p targets were most enriched for high-affinity 9BEs, and the enrichment progressively decreased from class I to class IV targets (**Fig. 1C**). We hypothesized that the degeneracy at the 3' end of low-affinity Puf4p-binding elements was the result of a variable number of “spacer” nucleotides between the 5' UGUA and the 3' UA. Indeed, class I binding elements were almost entirely composed of consensus 9BEs (3 spacer nts), and class IV binding elements were more likely to include 8BEs (2 spacer nts) or 10BEs (4 spacer nts) (**Supplementary Fig. 2A**). We again mined the Puf4p SEQRS data and determined that 9BEs were best enriched by Puf4p *in vitro*, with weaker enrichments observed for 8BEs and 10BEs (**Supplementary Fig. 2B**).

Puf4p target class correlates with Puf4p-dependent regulation and biological function. To determine whether target classes correlated with known mechanisms of Puf4p-dependent regulation, we mined published data that determined the change in stability of mRNAs genome-wide with and without *PUF4*³². Class I Puf4p targets had the largest increase in RNA stability in the absence of *PUF4* (the mRNAs have slower decay rates in a $\Delta puf4$ strain relative to a wild-type strain), and the enrichment progressively decreased from class I to class IV targets (**Fig. 1D**). Similarly, class I targets were also most enriched for ribosome biogenesis related functions and proteins that localize to the nucleolus, and enrichment also progressively decreased from class I to class IV targets (**Supplementary Fig. 2C-D**).

Puf4p target class correlated with evolutionary conservation of its binding element in particular targets. We analyzed the conservation of 9BEs present in Puf4p targets with orthologues present in 16 species of budding yeast, which represent more than 400 million years of evolution³³. This yielded a “conservation score”, with 16 indicating the 9BE was present in all 16 budding yeasts. 9BEs present in class I Puf4p targets had a median conservation score of 11, while 9BEs present in all mRNAs had a median conservation score of 5 (**Fig. 1E**). 9BE conservation progressively decreased from class I to the very modestly conserved class IV targets. These findings strongly suggest that class I Puf4p targets are the most highly conserved.

Together, our findings suggest a model for Puf4p binding and regulation *in vivo*. In comparison to class I and II targets, class III and IV Puf4p targets were much less enriched for high-affinity 9BEs, ribosome biogenesis-related functions, nucleolar localization, *PUF4*-dependent regulation, and conserved Puf4p-binding elements.

These data are consistent with the proposed “two-handed model” for Puf4p binding to RNA²⁸: Puf4p likely samples many sequence elements with a 5' UGUA, using its C-terminal PUF domains, as it searches for an optimally positioned 3' UA with its N-terminal PUF domains. In this model, only mRNAs bound well by both “hands” of Puf4p lead to interactions long enough to enable Puf4p-dependent regulation.

The Puf5p sub-network

In PUF5-PUP yeast, we detected 916 RNAs that were reproducibly U-tagged, which we refer to as “Puf5p targets” (**Supplementary Fig. 3A**). The vast majority of Puf5p targets

were mRNAs (914), though two non-coding RNAs were also detected (TLC1 and a tRNA-Asn). Puf5p RNA targets were enriched for a 5' UGUA motif in their 3' regulatory regions (603/916, $P < 10^{-16}$). Directed searches of 3' UTRs for PUF-binding elements revealed that Puf5p mRNA targets were most enriched for 9BEs and 10BEs (**Fig. 2A**), consistent with a previous study¹⁸. Puf5p targets largely had a single 9BE or 10BE (498 mRNAs), rather both simultaneously (35 mRNAs) (**Fig. 2B**). Puf5p targets were enriched for a broad range of biological functions, including cytoplasmic translation, ribosome biogenesis, chromosome organization, and transcription (**Supplementary Fig. 3B**). Thus, Puf5p binds mRNAs enriched for Puf5p-binding elements that are implicated in fundamental cellular processes.

Puf5p target class correlated with enrichment for 10BEs. Using the same strategy as with Puf4p, we separated Puf5p targets into four classes based on the number of U-tagged reads detected and the length of their U-tags (**Fig. 2C**). Class I Puf5p targets were most enriched for 10BEs, and the enrichment progressively decreased from class I to class IV Puf5p targets (**Fig. 2D**). Class I Puf5p targets were also modestly enriched for conserved 10BEs (**Supplementary Fig. 4A**) and *PUF5*-dependent changes in RNA stability ($P < 0.00001$) (**Supplementary Fig. 4B**). In contrast, class II Puf5p targets were most enriched for 9BEs (**Fig. 2D**), which are bound well by both Puf5p and Puf4p^{18,30}. Class II Puf5p targets were also most enriched for ribosome biogenesis and chromatin-related functions (**Supplementary Fig. 4C**).

In comparison to Puf3p¹⁷ and Puf4p (this study), Puf5p target classes are less correlated with enrichments for Puf5p-binding elements, biological functions, and a known mechanism of *PUF5*-dependent regulation. Thus to ensure PUF5-PUP specificity for Puf5p-bound mRNAs, we tested whether PUF5-PUP required a Puf5p-binding element to deposit U-tags on a class I target. We selected *PHD1* mRNA as the class I target because it was also strongly detected as a Puf5p target using HITS-CLIP¹⁸, which localized its PBEs. We replaced the endogenous copy of *PHD1* mRNA with a mutant version that lacked the PBEs (UGU to ACA substitutions) and analyzed the mutant *PHD1* strain via RNA Tagging¹⁷. Importantly, zero U-tagged *PHD1* mRNAs were detected in the mutant *PHD1* strain while we detected about 51 TRPM for the wild-type allele (**Fig. 2E**). Thus, PUF5-PUP requires Puf5p-binding elements to tag mRNAs.

Puf5p RNA Tagging data displays only weak correlations between target class and *PUF5*-dependent regulation and enrichment for Puf5p-binding elements. Intriguingly, 9BE enrichment peaked in class II rather than class I Puf5p targets, despite Puf5p having similar in vitro binding preferences for both 9BEs and 10BEs. Thus, we hypothesized that Puf4p and Puf5p bind many of the same mRNAs, particularly those with 9BEs. Such a model would explain our finding that class II Puf5p targets were most enriched for 9BEs, sequences it may compete with Puf4p to bind, while class I targets were most enriched for its unique binding element, the 10BE.

The PUF supra-network

Canonical PUF proteins and their RNA targets form a large network of interactions, which we call the “PUF supra-network”. To identify mRNAs bound by multiple PUF proteins, we collectively analyzed our RNA Tagging data for each of the canonical PUF proteins: Puf3p, Puf4p, and Puf5p. We reanalyzed our published Puf3p RNA Tagging data using the same approaches as done for Puf4p and Puf5p¹⁷. Consistent with prior analyses, Puf3p target class was highly correlated with enrichment for high-affinity and highly-conserved Puf3p-binding elements, and PUF3-dependent regulation (**Supplementary Fig. 5**). We next constructed a network map of all RNAs tagged by at least one canonical PUF protein (**Fig. 3**). Puf3p, Puf4p, and Puf5p collectively U-tagged 1,417 RNAs, thereby encompassing about 20% of the yeast transcriptome. GO analyses revealed that the PUF supra-network is involved in the regulation of mitochondria, reflective of the Puf3p sub-network, and the regulation of ribosome biogenesis and chromatin related functions, reflective of Puf4p and Puf5p sub-networks (**Fig. 3**).

The Puf4p and Puf5p sub-networks are highly interconnected. In addition to the 85 RNAs U-tagged by all three proteins, 222 mRNAs were U-tagged by both Puf4p and Puf5p (**Fig. 3**). Therefore, approximately 60% of the Puf4p sub-network is included in the Puf5p sub-network. We hypothesized that many of the shared Puf4p and Puf5p targets would be strongly U-tagged by both Puf4p and Puf5p since each protein binds with high affinity to 9BEs^{9,18,23,28-30}. Indeed, 82 RNAs were class I or II targets for both Puf4p and Puf5p (red squares, **Fig. 3**). This represents 27% of their shared targets, is similar to the number of RNAs strongly U-tagged by individual proteins, and is in stark contrast to the number of

targets shared with Puf3p (**Fig. 3 & Supplementary Fig. 6A**). Overlap between Puf4p and Puf5p targets was maximal among class I targets and progressively decreased from class I to class IV targets (**Supplementary Fig. 6B-C**). Consistent with a potential co-regulatory role, yeast that lack both Puf4p and Puf5p had enhanced sensitivities to several drugs in comparison to wild-type yeast and yeast that lacked either protein alone (**Supplementary Fig. 7**).

We hypothesized that PUF proteins selected their RNA targets based on the presence of their preferred binding elements. Indeed, mRNAs uniquely bound by Puf3p, Puf4p, or Puf5p were most enriched for their preferred binding elements (**Fig. 3 & Supplementary Fig. 6D**). Intriguingly, however, mRNAs U-tagged by both Puf4p and Puf5p were most enriched for 9BEs and only very weakly enriched for 10BEs (**Fig. 3 & Supplementary Fig. 6D**), which suggested that shared Puf4p and Puf5p targets primarily possessed 9BEs. Consistently, mRNAs present in class I or II for both Puf4p and Puf5p (“class I-II shared targets”, 82 RNAs) primarily possessed a single PBE in their 3' UTR (57 mRNAs) (**Supplementary Fig. 6E**), most of which were highly-conserved (49/57) (**Supplementary Fig. 6F**).

Our analyses above illustrated that many Puf4p targets with 9BEs are not bound by Puf5p, even though they possess its high-affinity binding element (**Fig. 3**). To determine whether relative protein and RNA abundances could explain our observation, we examined the relative expression levels of Puf4p, Puf5p, and the mRNAs they bind. Via western blot analyses, Puf4p is 3-9 fold higher expressed than Puf5p in the RNA Tagging

strains (**Figs. 4A-B**), which agreed well with the relative expression level of the endogenous proteins^{34,35}. Thus we hypothesize that the low expression of Puf5p in comparison to Puf4p excludes Puf5p from many Puf4p targets, even though they possess high-affinity binding elements. Consistently, mRNAs present in class I-II of both Puf4p and Puf5p were 2-fold more abundant than unique class I-II Puf4p or Puf5p target mRNAs (Student's two-sided *t*-test, $P < 10^{-6}$, Fisher-Pitman permutation test, $P < 10^{-15}$) (**Fig. 4C**). The increased expression of those mRNAs may allow Puf5p access to them even in the presence of Puf4p. Similarly, class I-II shared Puf4p-Puf5p targets that solely possessed a 10BE, which is only weakly bound by Puf4p, or lacked any PBE were greater than 4-fold more abundant than class I-II shared targets with a 9BE (Student's two-sided *t*-test, $P < 0.001$, Fisher-Pitman permutation test, $P < 0.001$) (**Fig. 4D**).

We therefore suggest that an interplay between mRNA abundance, protein abundance, and relative binding-affinities underlies the entire PUF supra-network. The Puf3p and Puf5p sub-networks are most enriched for mRNAs with their unique binding elements. Puf5p appears excluded from relatively rare mRNAs with 9BEs, as effect that is likely compounded by its reduced expression relative to Puf4p. mRNAs bound by both Puf4p and Puf5p, most often with a single PUF-binding element (9BE) in their 3' UTR, are more abundant, which likely allows both proteins to bind. Moreover, increased RNA abundance appears to provide Puf4p access to mRNAs with relatively poor binding sites, particularly among the class I-II targets it binds with Puf5p.

Rewiring of the Puf4p & Puf5p sub-networks

We hypothesized that Puf4p and Puf5p compete to bind RNAs *in vivo*. To test this hypothesis, we performed RNA Tagging in a yeast strain that expressed PUF4-PUP and lacked *PUF5* (“Puf4p; Δ *puf5*”), and in a strain that expressed PUF5-PUP and lacked *PUF4* (“Puf5p; Δ *puf4*”). Expression of the individual fusion proteins was largely unchanged relative to fusion protein in a wild-type background (**Supplementary Fig. 8**).

The Puf4p sub-network expands in the absence of *PUF5*. We detected 1,365 U-tagged mRNAs and four non-coding RNAs in Puf4p; Δ *puf5* yeast, which we refer to as the Puf4p; Δ *puf5* sub-network (**Supplementary Fig. 9A**). In the presence of Puf5p, only 507 RNAs were U-tagged by Puf4p (see **Fig. 1**). Nearly all Puf4p targets were also Puf4p; Δ *puf5* targets (“retained Puf4p targets”), as a remarkable 98% of RNAs U-tagged by PUF4-PUP in wild-type yeast were also U-tagged by PUF4-PUP in yeast that lacked *PUF5* (**Fig. 5A**). The few Puf4p targets absent from the Puf4p; Δ *puf5* sub-network were nearly all class IV Puf4p targets (**Fig. 5B**). Retained Puf4p targets often were present in better classes in the Puf4p; Δ *puf5* sub-network (**Fig. 5C & Supplementary Fig. 9B**).

The expanded Puf4p; Δ *puf5* sub-network included many Puf5p targets. The Puf4p; Δ *puf5* sub-network gained 322 Puf5p targets absent from the Puf4p sub-network (“gained Puf5p targets”) (**Fig. 5A**). Thus, the Puf4p; Δ *puf5* sub-network included 68% of all Puf5p targets (624 RNAs). Inclusion of Puf5p targets in the Puf4p; Δ *puf5* sub-network was correlated with Puf5p target class (**Fig. 5B**). For example, 89% of class I Puf5p targets were Puf4p; Δ *puf5* targets, a marked increase from the 54% of class I Puf5p targets that were

also Puf4p targets (see **Supplementary Figs. 6B-C**). Gained Puf5p targets (322 mRNAs) and Puf5p targets absent from Puf4p; Δ *puf5* sub-network (“lost Puf5p targets”, 292 mRNAs) were similarly enriched for 8BEs and 10BEs, while 9BEs were modestly more enriched in gained Puf5p targets (**Fig. 5D**). In contrast, gained Puf5p targets were significantly higher expressed at the mRNA level than lost Puf5p targets (Student’s two-sided *t*-test, $P < 10^{-14}$, Fisher-Pitman permutation test, $P < 10^{-15}$) (**Fig. 5E**). Nearly all RNAs in the Puf4p; Δ *puf5* sub-network that were not Puf4p or Puf5p targets (551 RNAs) were present in class III or IV (**Supplementary Fig. 9C**) and were only very weakly enriched for PBEs (**Fig. 5D**), which suggests they are merely sampled by the fusion protein.

In the reciprocal experiment, we found that the Puf5p sub-network contracted in the absence of *PUF4*. We detected 464 U-tagged mRNAs and two non-coding RNAs (TLC1 and tK(UUU)D) in Puf5p; Δ *puf4* yeast, which we refer to as the Puf5p; Δ *puf4* sub-network (**Supplementary Fig. 10**). Surprisingly, the Puf5p; Δ *puf4* sub-network only included 50% of mRNAs (438) present in the Puf5p sub-network (**Fig. 6A**), which we refer to as “retained Puf5p targets”. The Puf5p; Δ *puf4* sub-network was enriched for many of the same biological functions as the native Puf5p sub-network (**Fig. 6B**), while mRNAs present in the Puf5p sub-network but absent in the Puf5p; Δ *puf4* sub-network (“lost Puf5p targets”, 478 RNAs) were much less enriched. Retention of Puf5p targets in the Puf5p; Δ *puf4* sub-network was correlated with Puf5p target class (**Fig. 6C**). Similarly, RNAs bound by both Puf4p and Puf5p were more likely to be retained in the Puf5p; Δ *puf4* sub-network than mRNAs uniquely bound by Puf5p (**Fig. 6D**). Retained Puf5p targets

(438 RNAs) were modestly yet significantly more abundant than lost Puf5p targets (Student's two-sided *t*-test, $P < 0.01$, Fisher-Pitman permutation test, $P < 0.01$) (**Fig. 6E**), consistent with retention of relatively high-expressed shared Puf4p-Puf5p targets. Retained and lost Puf5p targets were similarly enriched for 9BEs and 10BEs (**Fig. 6F**).

Together, our findings support a model in which Puf4p and Puf5p compete with each other to bind RNAs *in vivo*. In wild-type yeast, Puf4p primarily binds to RNAs with 9BEs, its high-affinity binding element, and samples many RNAs with 10BEs, which it binds with much lower affinity than Puf5p. Thus, Puf4p is predominately occluded from RNAs uniquely bound by Puf5p due to their relative binding affinities for 10BEs. Puf4p is, however, able to sample mRNAs with relatively low-affinity 10BEs, perhaps due to its increased expression relative to Puf5p. When Puf5p is absent, Puf4p binds those RNAs despite their relatively low-affinity binding elements. In contrast, Puf5p is occluded from relatively low expressed mRNAs that contain 9BEs, a sequence it binds with high affinity, by Puf4p. Loss of Puf4p competitor in Puf5p; Δ puf4 yeast substantially increased the number of potential binding sites available to Puf5p, effectively diluting Puf5p protein, and with the result that Puf5p retained only its best targets.

Discussion

Overlapping regulatory networks are ubiquitous, and in particular, members of families of RBPs with related binding specificities are common^{3,36}. Using RNA Tagging, we have probed the nature of these higher-level protein-RNA networks to reveal key principles that control their architecture.

The yeast canonical PUF supra-network is composed of four major sub-networks. We demonstrate that Puf3p, Puf4p, and Puf5p each bind their own set of mRNAs, consistent with a previous study⁹. Importantly, our data also establish that Puf4p and Puf5p form a fourth node in the PUF regulatory network. Puf4p and Puf5p bound to many of the same mRNAs with a single 9BE, which were often class I or II targets of both proteins. Moreover, yeast that lack both Puf4p and Puf5p have enhanced phenotypes in comparison to yeast that lack either protein, and a recent report observed similar effects on the destabilization of a single shared target³⁷. Since shared Puf4p-Puf5p mRNA targets only possess a single high-affinity binding site, we suspect that Puf4p and Puf5p may bind to individual mRNA molecules sequentially, or they may bind to separate pools of mRNA molecules. A dual-tagging experiment – in which Puf4p is fused to a PUP and Puf5p is fused to ADAR, as done in the TRIBE approach³⁸ – would provide insight into this outstanding question and provide insight into the dynamics of multiple RBPs and the RNAs they bind.

Relative binding affinities and abundances play major roles in the formation, balance, and function of the PUF regulatory network. For example, Puf4p and Puf5p bind to 9BEs with

high-affinity, yet Puf5p is occluded from many mRNAs with 9BEs by Puf4p, which is more abundant. Thus, altering the relative abundance of proteins with similar binding affinities can determine which mRNAs are bound by which proteins. Puf5p, in contrast, largely outcompetes Puf4p to bind mRNAs with 10BEs, most likely through its greater intrinsic affinity for the sequence. Similarly, Puf3p and Puf4p are expressed at about equal levels, yet both proteins have very distinct targets due to their preference for particular sequences.

Our data also illustrate that the interplay between relative abundance and binding affinity is dynamic and that it can be modulated to change the architecture of protein-RNA networks. When Puf5p is absent, Puf4p bound to many more Puf5p targets, which most often had low-affinity Puf4p-binding elements (e.g. 10BEs). The relatively high expression of Puf4p likely enabled binding to many the new sites available to it upon loss of *PUF5*, and also likely explains how *PUF4* partially compensates for loss of *PUF5* in our growth assays. In contrast, the relatively low amount of Puf5p was likely insufficient to account for the large increase in potential binding sites, thus effectively diluting the protein and restricting detection to its strongest targets. The impact of such drastic, and more subtle, changes on protein-RNA networks will be an important avenue of research in the future, particularly in the context of evolution given that genomic duplication and deletion events have the potential to rapidly alter the balance of a network. Similarly, post-translational modifications can alter the activity of RBPs and may represent a dynamic mechanism to alter the architecture of protein-RNA networks and their function.

Competition is likely a conserved and general principle of RNA regulatory networks. Our findings with yeast Puf4p and Puf5p are consistent with previous reports that demonstrated human and mouse splicing factors compete with each other to bind splice sites^{39,40}. Similarly, related RBPs in mammals and nematodes also bind to many of the same mRNAs^{19,41-43}. Together, the studies suggest that an intricate balance between individual proteins and their RNA targets – including their abundances and relative binding affinities – establishes the architecture and dynamics of RNA regulatory networks.

METHODS

Yeast strains. All strains were constructed in BY4742 yeast (*MAT α ;his3 Δ 1;leu2 Δ 0;lys2 Δ 0;ura3 Δ 0*) as previously described¹⁷. Briefly, we inserted the *C. elegans* pup-2 open-reading frame followed by a stop codon, the URA3 marker with its native promoter and terminator, and a 3-HA epitope tag in-frame at the 3' end of *PUF4* and *PUF5*. The mutant PHD1 strain that lacked Puf5p-binding elements is described¹⁷.

Yeast growth: Cultures were grown as described¹⁷. Briefly, a single colony of each yeast strain was inoculated in 5 mL of yeast extract-peptone-dextrose plus adenine (YPAD) media and incubated at 30 °C with 180 r.p.m. shaking for \approx 24 hours. Saturated cultures were used to seed 25 mL fresh YPAD at $A_{660} \approx 0.0002$, which were grown at 30 °C with 180 r.p.m. shaking until $A_{660} \approx 0.5-0.8$.

RNA Tagging library preparations. Total RNA isolations and sequencing library preparations were done as previously described¹⁷.

High-throughput sequencing and raw data processing. Paired-end sequencing reads were obtained from Illumina sequencing platforms. FASTQ files were processed and aligned to the *S. cerevisiae* genome (version R64-1-1) as previously described¹⁷.

Definition of U-Tagged RNAs. As previously described¹⁷, U-tagged RNAs are defined as DNA fragments that end with at least eight adenosines followed by at least one 3' terminal thymidine (representing the U-tag) not encoded by any adapter or genomic

sequence. Read 1 typically contained sequence that matched to particular genomic regions, which allowed identification of the gene. Read 2 most often identified the A-U tail sequence. The number of U-Tagged RNAs per million uniquely mapped reads (TRPMs) for every gene was calculated and used to normalize data across samples.

Reproducible RNA Tagging targets. Targets of proteins were determined as previously described¹⁷. Briefly, genes were called targets if they met three criteria: they were detected by at least tenfold more TRPM in a tagging strain relative to a control non-tagging strain (e.g. PUF4-PUP yeast versus BY4742); the number of TRPM detected must have been above the error rate for falsely detecting U-tagged RNAs (3%); and, both of the previous criteria must have been met in all biological replicates.

Clustering analysis and class definition. TRPM values for each target were calculated for U-tags of at least 1, 2, 3, 4, 5, 6, 7, and 8 uridines in length. TRPM values were averaged (mean) across biological replicates. The order of targets was then randomized, all TRPM values were \log_2 transformed, and separated into eight groups via k-means clustering (1,000 iterations, Euclidean distance) using Gene Cluster 3.0 software. K-means groups were then sorted and ranked from longest to shortest U-tags. Heat maps were generated using MatLab (v2014a).

Classes were formed according to U-tag length. Class I targets were defined as the two groups (k-means ranked groups 1 and 2) of targets with the longest U-tags, typically including the majority of targets with U-tags up to seven or eight uridines in length. Class

II targets were defined as the two groups (groups 3 and 4) with the next longest U-tags, typically including the majority of targets with U-tags up to five or six uridines. Class III was defined as groups 5 and 6, and class IV was defined as the two groups (groups 7 and 8) with the shortest U-tags, typically only one or two uridines in length.

Network map and GO analyses. The map of the PUF regulatory network was generated using Cytoscape⁴⁴. Gene Ontology (GO) analyses were performed using YeastMine from the *Saccharomyces* Genome database (<http://yeastmine.yeastgenome.org>) using the default settings (Holm-Bonferroni correction).

Motif and directed motif analyses. Enriched sequence elements were identified using MEME as previously described¹⁷. In all analyses, 3' UTRs were defined as the longest observed isoform for a given gene⁴⁵, or 200 nucleotides downstream of the stop codon if not previously defined. For directed PBE searches, perl regular expression searches were used to identify: 8BEs, TGTA[ATC][ATC]TA; 9BEs, TGTA[ATC][ATC][ATC]TA; 10BEs, TGTA[ATC][ATC][ATC][ATC]TA; 11BEs, TGTA[ATC][ATC][ATC][ATC][ATC]TA; and 12BEs, TGTA[ATC][ATC][ATC][ATC][ATC][ATC]TA.

RNA abundance analyses. The number of mRNA molecules present in a cell was estimated in the following way. We previously performed an RNAseq experiment on a wild-type yeast strain (BY4742)¹⁷ in which we obtained an FPKM value for every gene (FPKM, fragments per kilobase of transcript per million mapped reads). We summed the FPKM values for every gene (811,639 total) and divided that number by 36,000, which

we used as a rough estimate for the number of mRNA molecule present in a cell^{46,47}, to obtain the estimated number of mRNAs present in a cell for each gene. In each comparison, estimated mRNA molecules were log₂-transformed, and median abundances of different groups were compared via two-tailed Student's *t*-tests and Fisher-Pitman permutation tests.

Mined datasets. Global changes in RNA stability for all genes in *puf4*Δ and *puf5*Δ mutants relative to wild-type yeast were obtained from Sun and colleagues³². Puf3p RNA Tagging data, and wild-type yeast RNAseq data were recently published by our group¹⁷. Puf4p SEQRS data was recently published by our group³⁰. Protein sub-cellular localizations were obtained from Huh and colleagues⁴⁸.

Yeast Plate assays: Single colonies of the indicated deletion strains were grown to saturation in 5 mL YPAD. 10-fold serial-dilutions of each yeast strain were plated (YPAD plates) from ≈ 10,000 cells to 10 cells (5 μL drops). Yeast were grown at 30 °C and briefly removed to take pictures approximately every 12 hours for 6 days.

Conservation analysis: The 16 *Saccharomycotina* yeast species used to determine PBE conservation scores were chosen based on previously determined orthology⁴⁹. Sequences of 300 bases downstream of the translation termination codon were obtained from FungiDB⁵⁰. Each 3' UTR sequence was probed for putative binding elements using a custom perl script, which is available upon request. The script determines log-likelihood scores for each *k*-mer (8-10 nt) based on canonical PUF binding elements^{9,17,18}. Each

RNA Tagging or HITS-CLIP target was assigned a “conservation score” defined as the number of orthologous genes with a PUF binding element (positive log-likelihood value).

ACKNOWLEDGEMENTS

We thank members of the Wickens and Kimble labs for helpful comments and suggestions on experiments and data. We appreciate feedback on the manuscript from Hugo Medina, Trish Hoang, Brian Carrick, and Camila Lopez-Anido. We thank J. Kimble (University of Wisconsin-Madison) for use of a computational server, and L. Vanderploeg of the Biochemistry Media Lab for help with the figures. Our research was supported by the US National Institutes of Health (GM50942), and by Wharton and Biochemistry Scholar fellowships to C.P.L.

Figure 1.

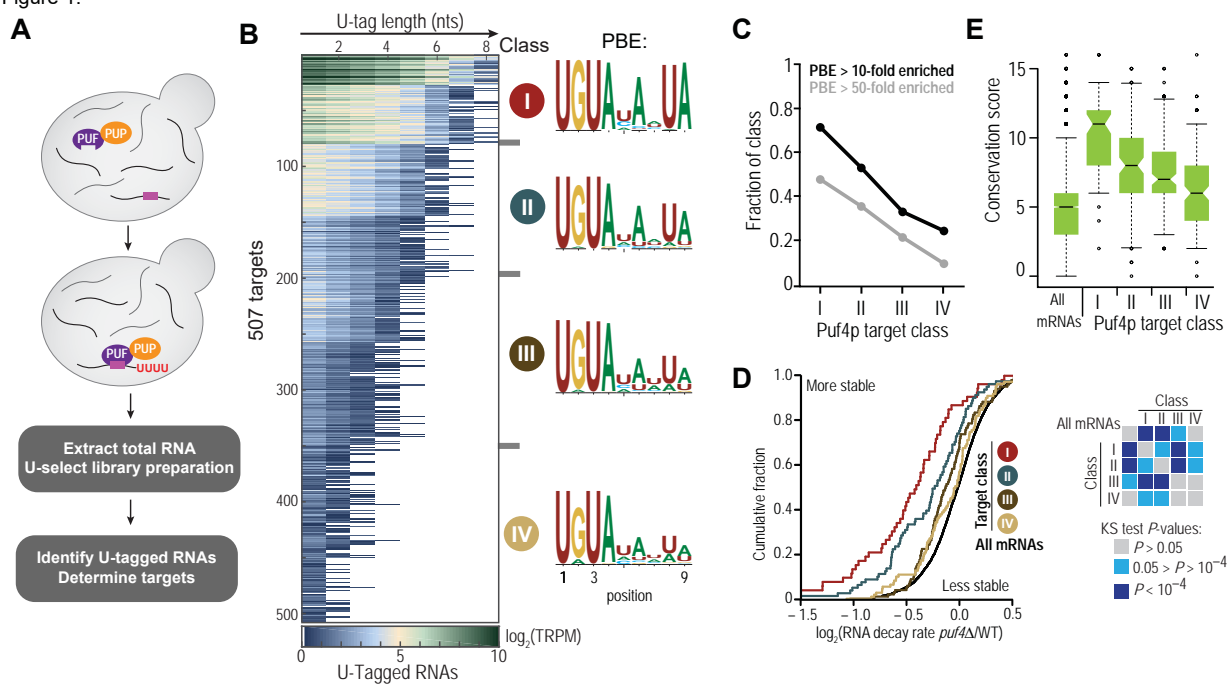


Figure 1. (a) Schematic of RNA Tagging. **(b)** Heat map displaying results of the k-means clustering analysis of all 507 Puf4p targets. Each row represents a single Puf4p target. Columns refer to the length of the U-tag detected on reads for each gene, from at least 1 uridine (leftmost column) to at least 8 uridines (rightmost). Puf4p target classes are indicated (I, II, III, & IV). The average PBE enriched in each class of targets is also indicated, with the y-axis in bits. **(c)** Plot showing the fraction of each Puf4p target class with PBEs enriched at least 10-fold (black) or 50-fold (gray) above background in SEQRs. **(d)** Empirical cumulative distribution of Puf4p target classes and all mRNAs for mRNA decay rate fold change ($puf4\Delta/wildtype$) in yeast with and without *PUF4* mined from published data³². Kolmogorov-Smirnov (KS) test *P*-values for pair-wise comparisons are indicated. **(e)** Box plot of 9BE conservation scores for all mRNAs and Puf4p target classes.

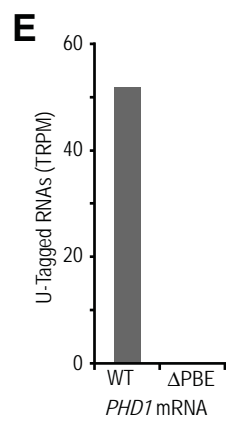
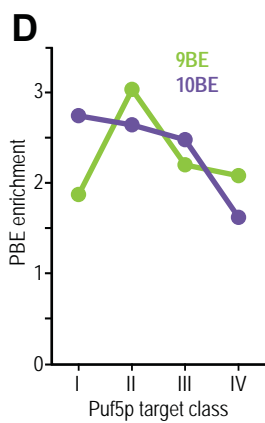
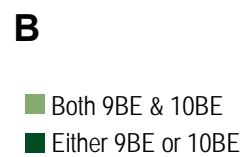
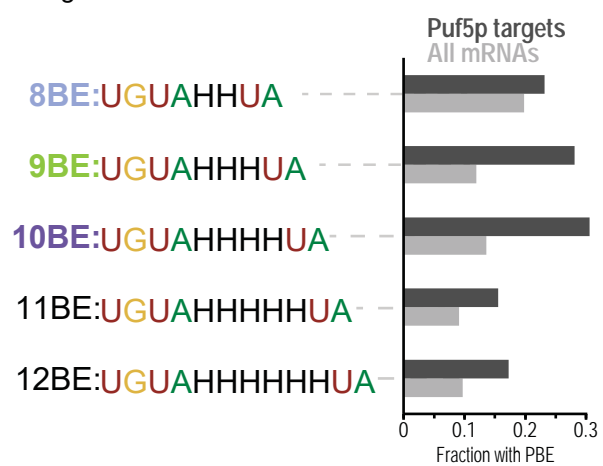
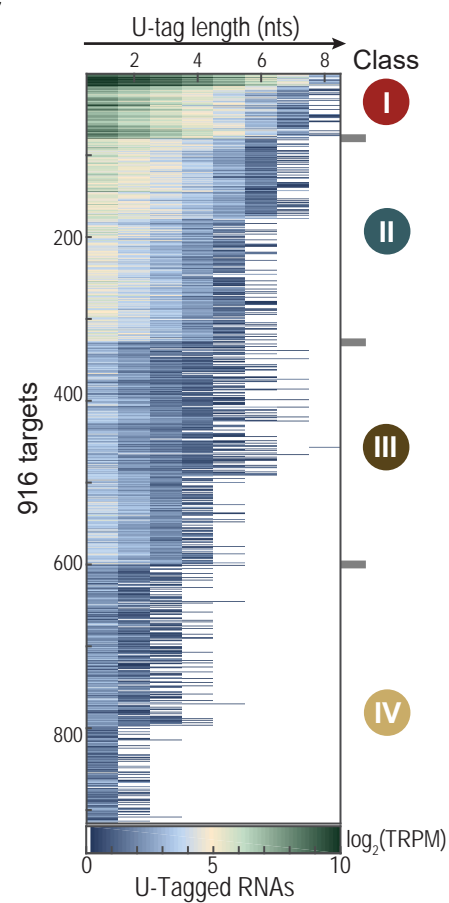
A Figure 2.**C**

Figure 2. (a) Plot of the fraction of Puf5p targets and all mRNAs with the indicated PBE. (b) Pie chart illustrating the number of Puf5p targets with both a 9BE and 10BE (light green), or either a 9BE or 10BE (dark green) in their 3' UTR. (c) Heat map displaying results of the k-means clustering analysis of all 916 Puf5p targets. Each row represents a single Puf5p target. Columns refer to the length of the U-tag detected on reads for each gene, from at least 1 uridine (leftmost column) to at least 8 uridines (rightmost). Puf5p target classes are indicated (I, II, III, & IV). (d) Plot showing the enrichment of each class of Puf5p targets relative to all mRNAs for 9BEs and 10BEs. (e) Plot of the number of U-Tagged mRNAs detected for wild-type *PHD1* mRNA (WT) and *PHD1* mRNA that lacked Puf5p-binding elements (Δ PBE).

Figure 3.

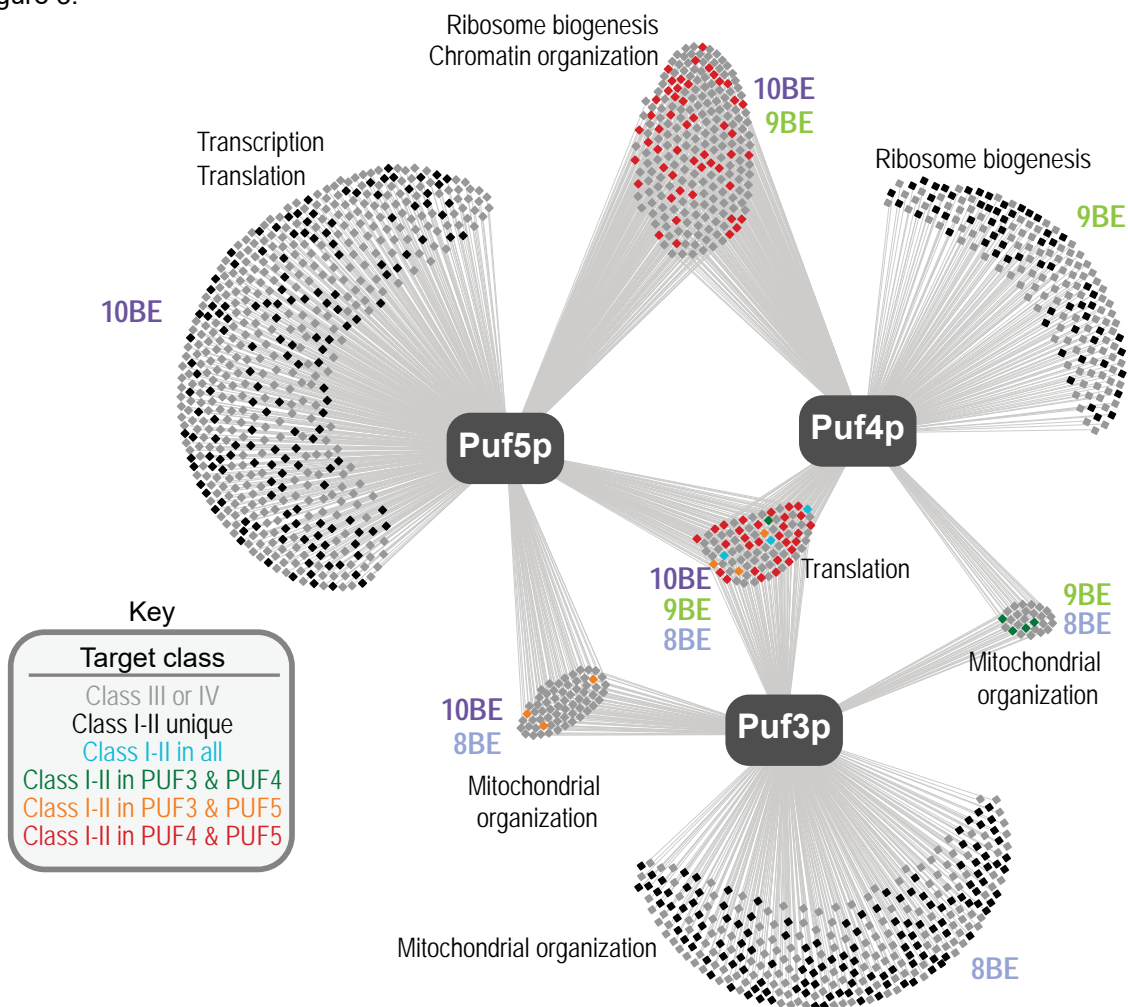


Figure 3. (a) Map of the PUF regulatory network. Each box represents a single gene and lines illustrate if it was U-Tagged by a given PUF protein. The Key indicates how genes were colored. PBEs enriched above background and broad Gene Ontology enrichments in different groups of targets are indicated.

Figure 4.

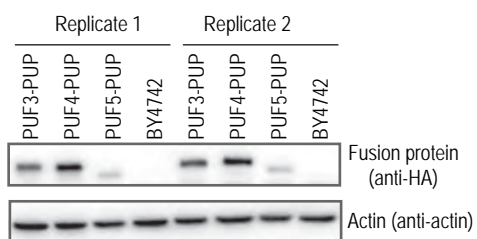
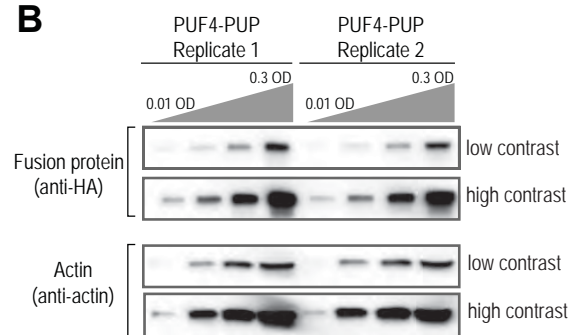
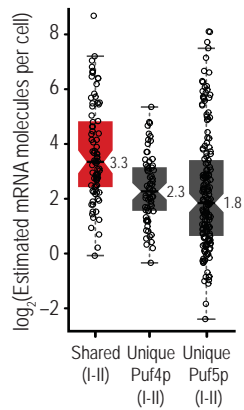
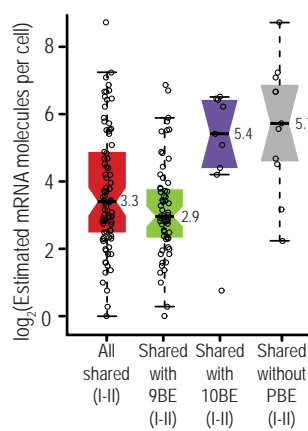
A**B****C****D**

Figure 4. (a) Western blot depicting relative protein levels of the indicated strains. PUF4-PUP and PUF5-PUP contained 3-HA epitope tags on their C-termini. Actin was used as the loading control. **(b)** Western blot depicting PUF4-PUP protein levels following 3-fold serial dilutions of cell extract. High- and low- contrast images of the same blots are shown. **(c)** Boxplot illustrating the estimated number of mRNA molecules present in each yeast cell for the indicated groups of mRNAs. The individual data points for each gene and medians are overlaid on the boxplot. “Unique Puf5p (I-II)” refers to RNAs that were uniquely U-tagged by PUF5-PUP and were present in class I or II. “Unique Puf4p (I-II)” refers to RNAs that were uniquely U-tagged by PUF4-PUP and were present in class I or II. “Shared (I-II)” refers to mRNAs U-tagged by both PUF4-PUP and PUF5-PUP and were present in class I or II of both data sets. **(d)** Boxplot illustrating the estimated number of mRNA molecules present in each yeast cell for the indicated groups of mRNAs. The individual data points for each gene and medians are overlaid on the boxplot. “Shared (I-II)” is defined as mRNAs U-tagged by both PUF4-PUP and PUF5-PUP and were present in class I or II of both data sets.

Figure 5.

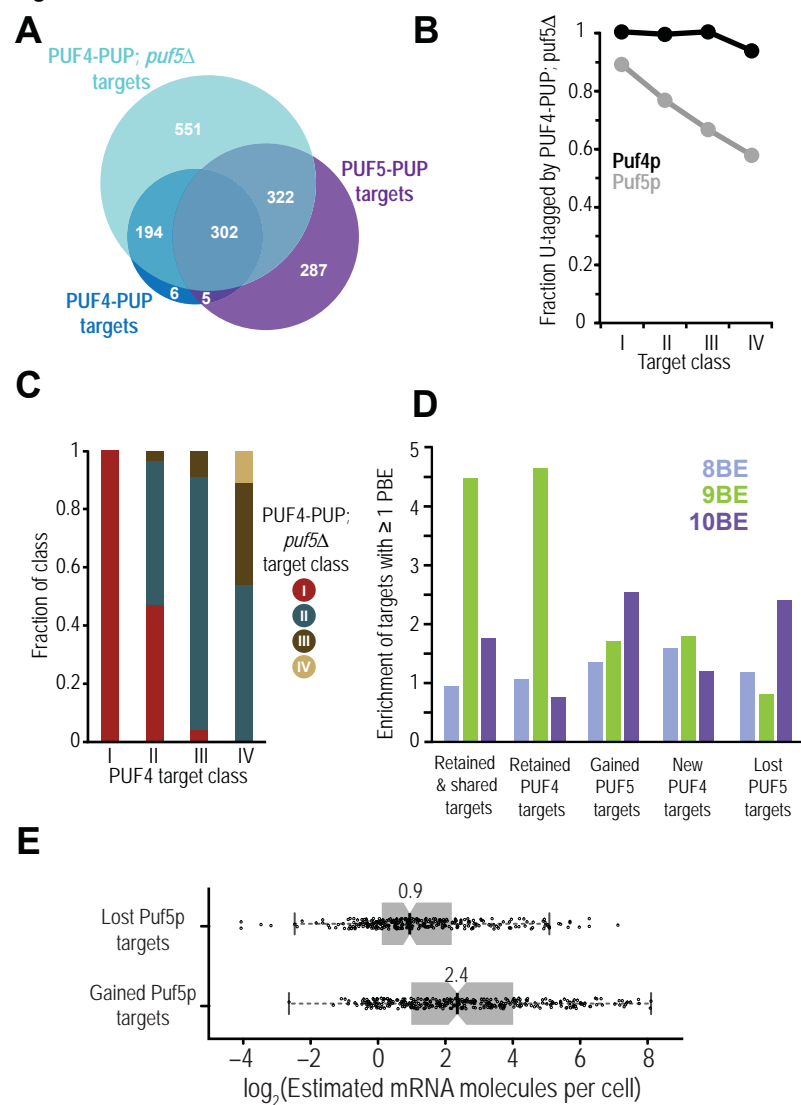


Figure 5. (a) Proportional Venn diagram illustrating overlap among the indicated targets. **(b)** Plot of the fraction of each class of Puf4p (black) and Puf5p (gray) targets that were U-Tagged by PUF4-PUP in the absence of *PUF5* (PUF4-PUP;*puf5* Δ targets). **(c)** Plot of the fraction of each Puf4p target class that was present in the indicated PUF4-PUP;*puf5* Δ target class. For example, all class I Puf4p targets were class I PUF4-PUP;*puf5* Δ targets, and ~45% class II Puf4p targets improved to class I PUF4-PUP;*puf5* Δ targets while ~55% class II Puf4p targets remained class II PUF4-PUP;*puf5* Δ targets. **(d)** Enrichment of the indicated groups of genes for 8BEs, 9BEs, and 10BEs relative to all mRNAs. “Retained and shared targets” (302) were RNAs U-tagged by PUF4-PUP, PUF5-PUP, and PUF4-PUP;*puf5* Δ . “Retained Puf4p targets” (194) were RNAs U-tagged by PUF4-PUP and PUF4-PUP;*puf5* Δ . “Gained Puf5p targets” (322) were RNAs U-tagged by PUF5-PUP and PUF4-PUP;*puf5* Δ . “New Puf4p targets” (551) were RNAs U-tagged solely by PUF4-PUP;*puf5* Δ . “Lost Puf5p targets” were RNAs U-tagged solely by PUF5-PUP. **(e)** Boxplot illustrating the estimated number of mRNA molecules present in each yeast cell for the indicated groups of mRNAs, defined in panel d. The individual data points for each gene and medians are overlaid on the boxplot.

Figure 6.

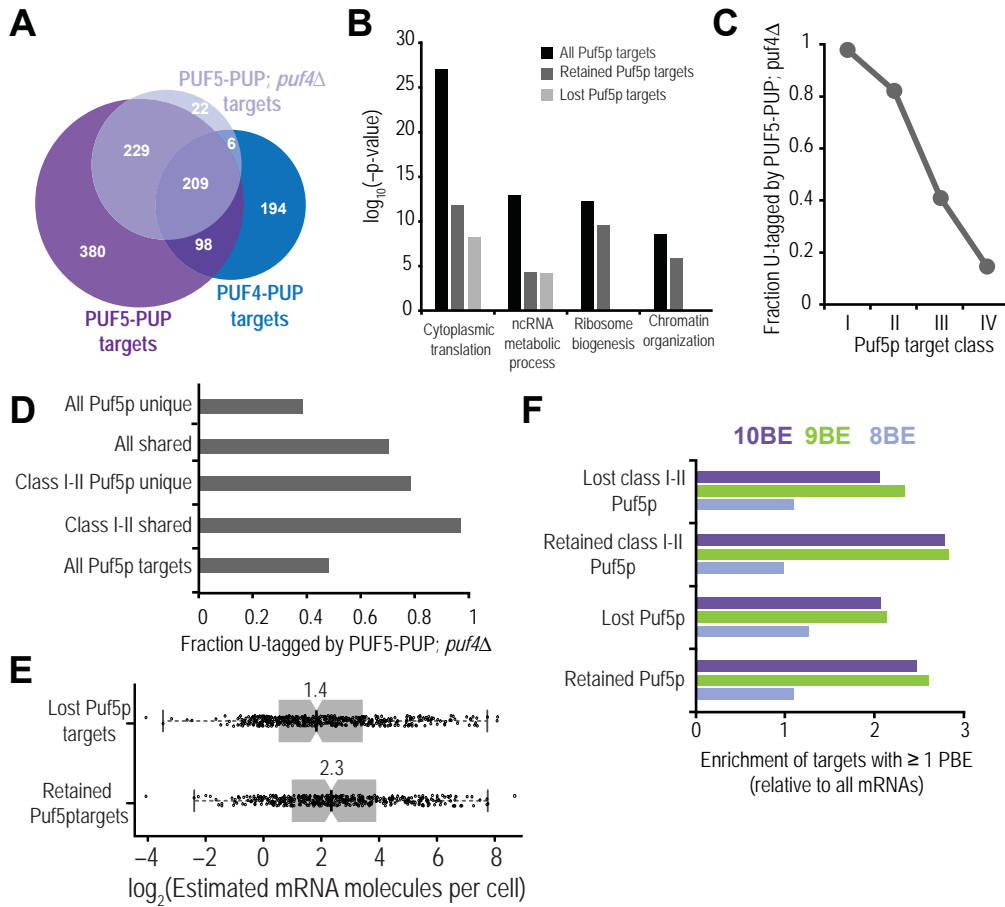
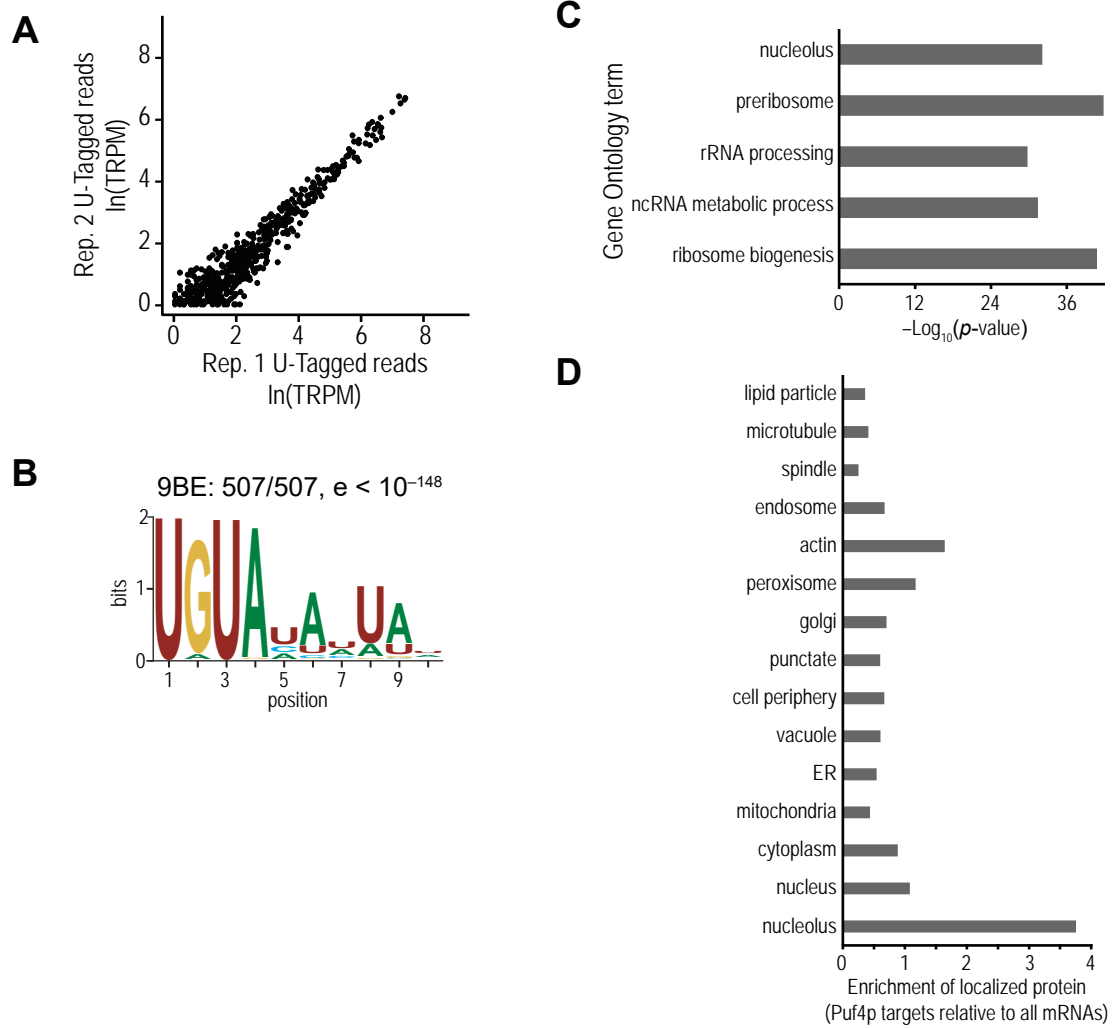
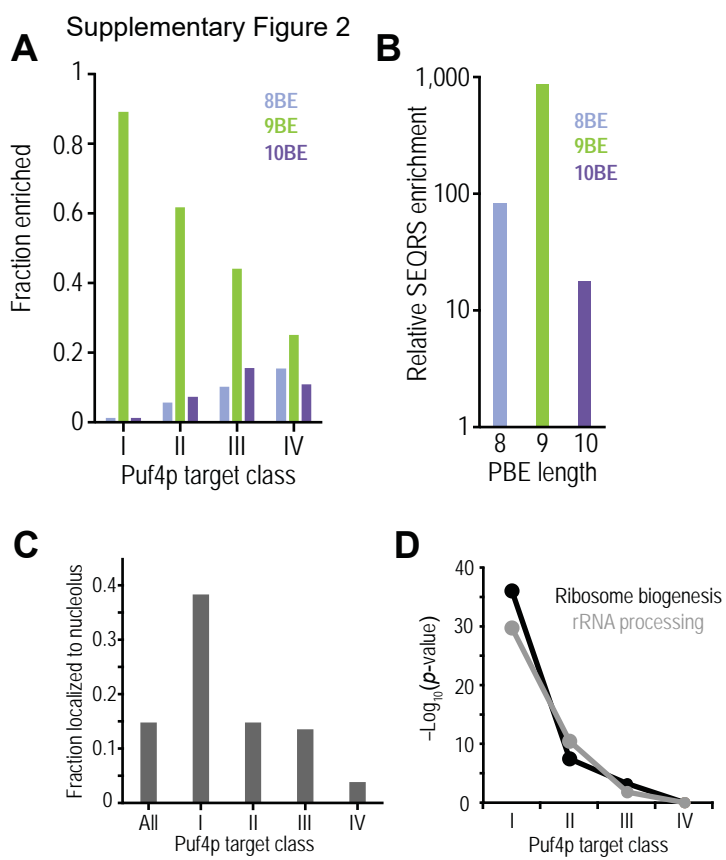


Figure 6. (a) Proportional Venn diagram illustrating overlap among the indicated targets. **(b)** Bar chart illustrating the *P*-values for the indicated Gene Ontology terms (biological process) for the indicated groups of genes. “Retained Puf5p targets” (438) were RNAs U-tagged by both PUF5-PUP and PUF5-PUP;*puf4* Δ . “Lost Puf5p targets” (478) were RNAs U-tagged by PUF5-PUP but not PUF5-PUP;*puf4* Δ . **(c)** Plot of the fraction of each class of Puf5p targets that were U-Tagged by PUF5-PUP in the absence of *PUF4* (PUF5-PUP;*puf4* Δ targets). **(d)** Plot of the fraction of the indicated groups of genes that were U-Tagged by PUF5-PUP in the absence of *PUF4* (PUF5-PUP;*puf4* Δ targets). “Shared” refers to RNAs that were U-tagged by both PUF4-PUP and PUF5-PUP. **(e)** Boxplot illustrating the estimated number of mRNA molecules present in each yeast cell for the indicated groups of mRNAs, defined in panel b. The individual data points for each gene and medians are overlaid on the boxplot. **(f)** Enrichment of the indicated groups of genes (essentially as defined in panel b) for 8BEs, 9BEs, and 10BEs relative to all mRNAs.

Supplementary Figure 1

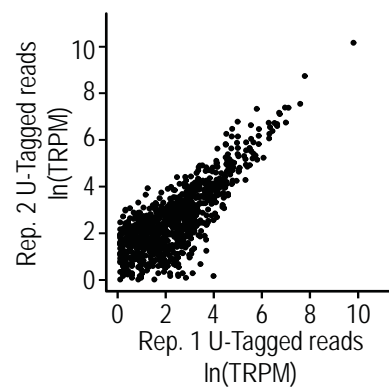
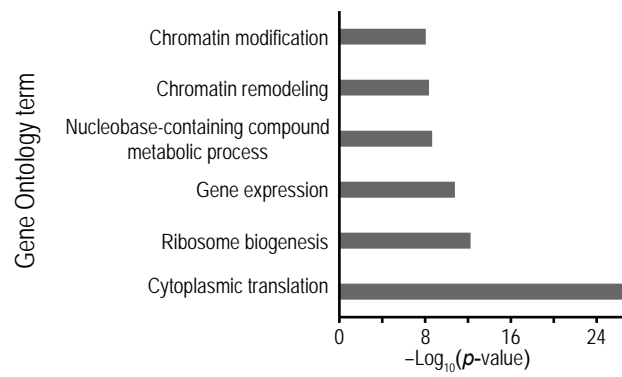


Supplementary Figure 1. (a) Scatter plot of the number of U-Tagged RNAs detected for each Puf4p target (507) in two biological replicates of PUF4-PUP yeast. TRPM, Tagged RNAs per million uniquely mapped reads. (b) Logo of the enriched sequence motif present in the 3' UTRs of Puf4p targets identified via MEME. (c) Bar chart illustrating the *P*-values for the indicated Gene Ontology terms enriched in Puf4p targets. (d) Plot of the enrichment of Puf4p targets for mRNAs encoding proteins that localize to the indicated subcellular locations, mined from GFP localization data⁴⁸.



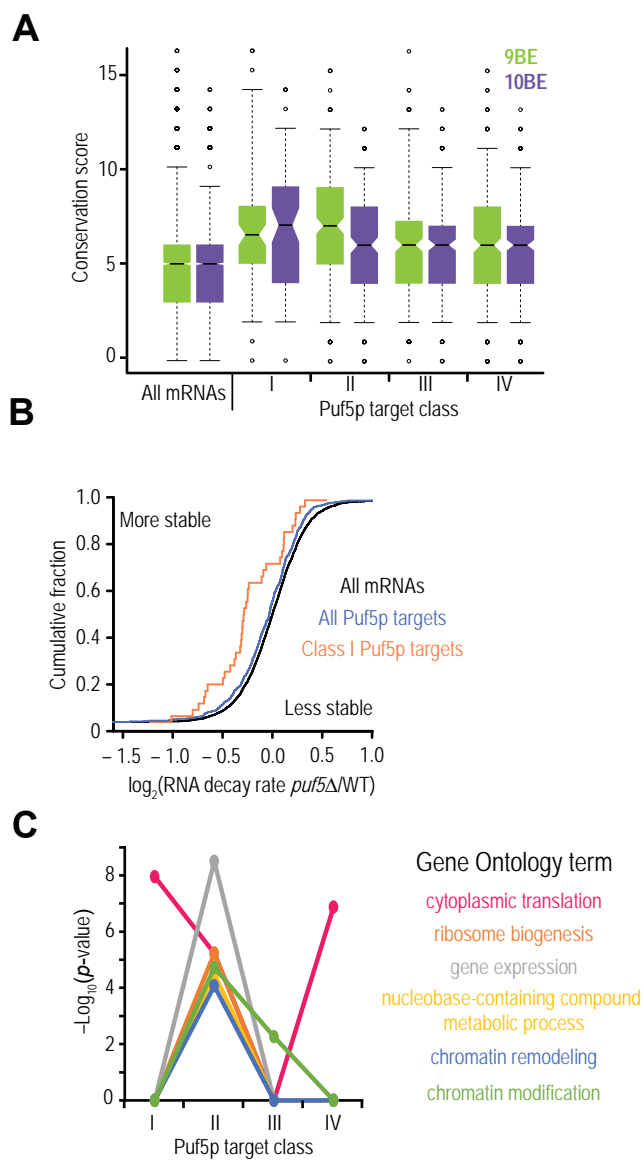
Supplementary Figure 2. (a) Plot of the enrichment of each class of Puf4p targets for 8BEs, 9BEs, and 10BEs. **(b)** Plot of SEQRS enrichment of Puf4p binding to 8BE, 9BE, and 10BEs *in vitro*. Enrichment for each PBE was calculated relative to randomers of the same length. **(c)** Plot of the enrichment of the indicated Puf4p targets for mRNAs encoding proteins that localize to the nucleolus, mined from GFP localization data⁴⁸. **(d)** Bar chart illustrating the *P*-values for the indicated Gene Ontology terms enriched in each class of Puf4p targets.

Supplementary Figure 3

A**B**

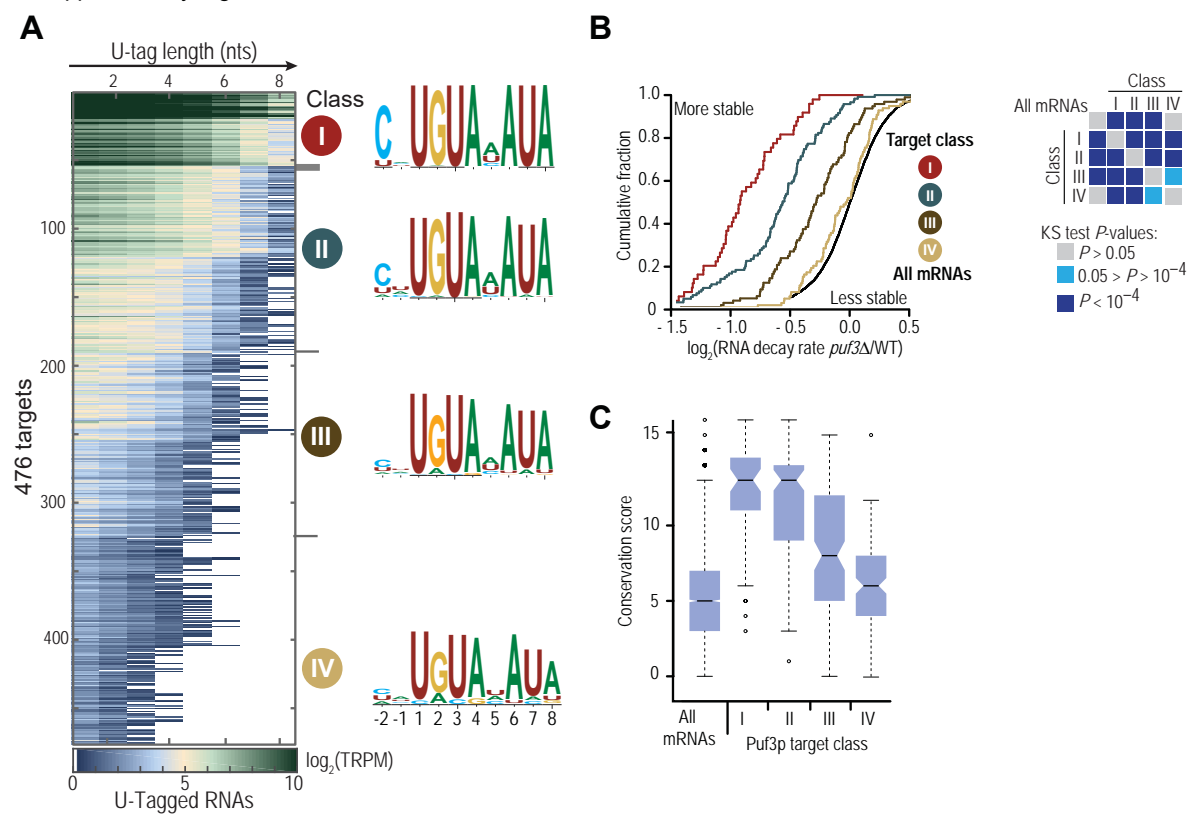
Supplementary Figure 3. (a) Scatter plot of the number of U-Tagged RNAs detected for each Puf5p target (916) in two biological replicates of PUF5-PUP yeast. TRPM, Tagged RNAs per million uniquely mapped reads. **(b)** Bar chart illustrating the *P*-values for the indicated Gene Ontology terms enriched in Puf5p targets.

Supplementary Figure 4



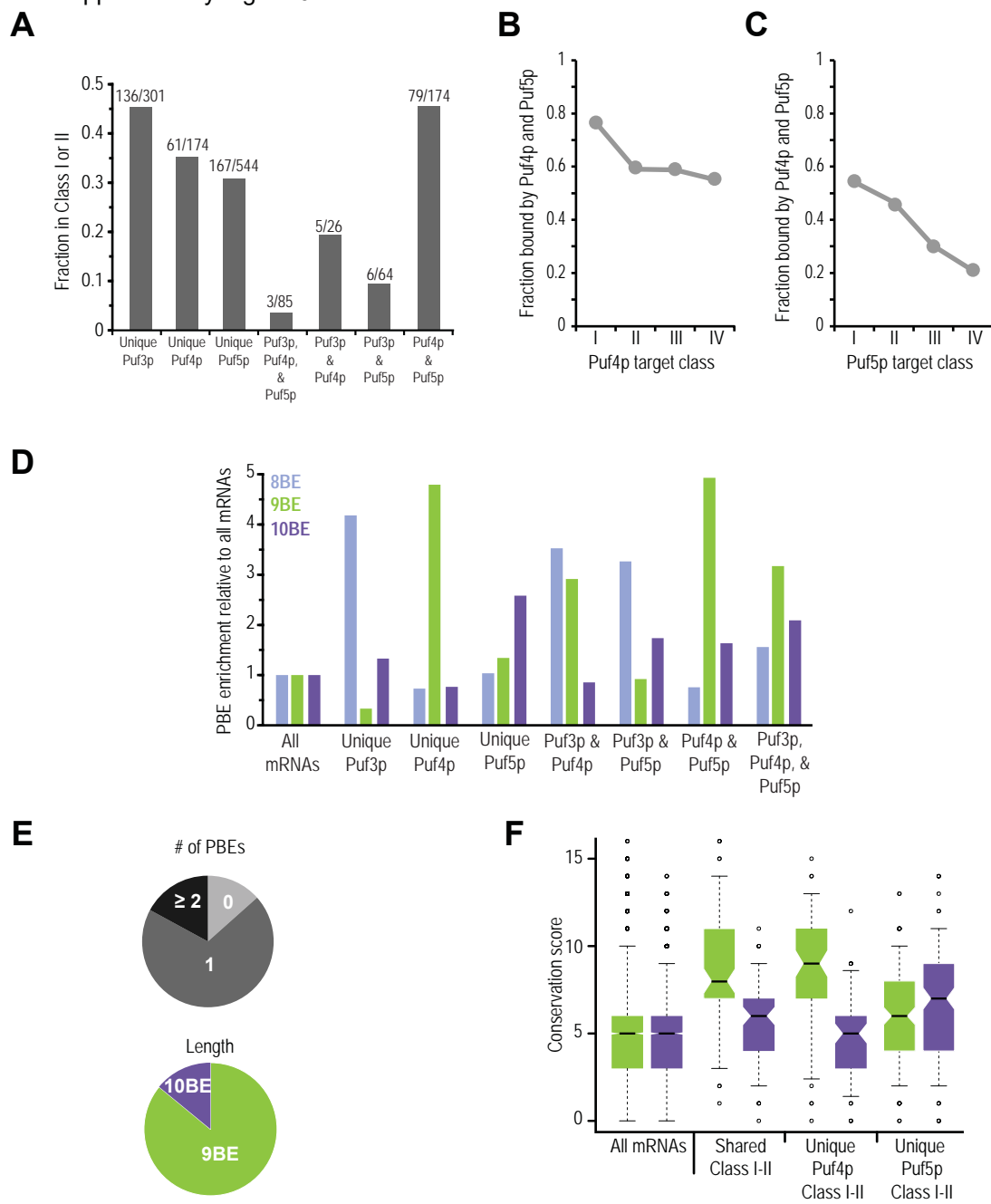
Supplementary Figure 4. (a) Box plot of 9BE and 10BE conservation scores for all mRNAs and Puf5p target classes. **(b)** Empirical cumulative distribution of all mRNAs (black), all Puf5p targets (blue), and class I Puf5p targets (orange) for mRNA decay rate fold change (*puf5* Δ /wildtype) in yeast with and without *PUF5* mined from published data³². **(c)** Bar chart illustrating the *P*-values for the indicated Gene Ontology terms enriched in each class of Puf5p targets.

Supplementary Figure 5



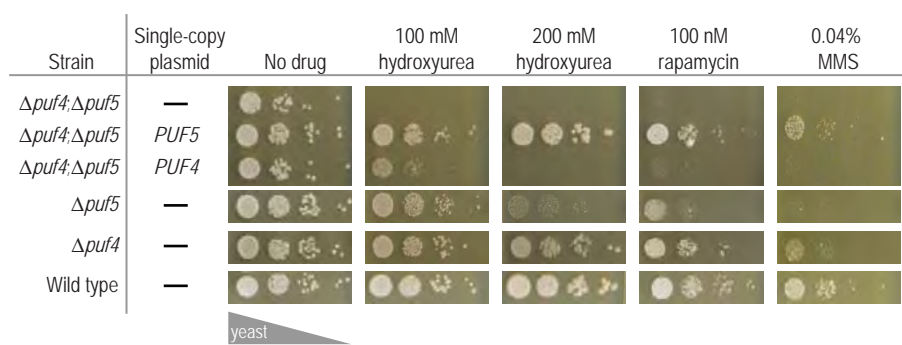
Supplementary Figure 5. (a) Heat map displaying results of the k-means clustering analysis of all 476 Puf3p targets. Each row represents a single Puf3p target. Columns refer to the length of the U-tag detected on reads for each gene, from at least 1 uridine (leftmost column) to at least 8 uridines (rightmost). Puf3p target classes are indicated (I, II, III, & IV). The average PBE enriched in each class of targets is also indicated, with the y-axis in bits. Puf3p RNA Tagging data was reanalyzed here from published data¹⁷. **(b)** Empirical cumulative distribution of all mRNAs (black), and class I (red), class II (teal), class III (brown), and class IV (gold) Puf3p targets for mRNA decay rate fold change (puf3Δ/wildtype) in yeast with and without *PUF3* mined from published data³². Kolmogorov-Smirnov (KS) test *P*-values for pair-wise comparisons are indicated. **(c)** Box plot of 8BE conservation scores for all mRNAs and Puf3p target classes.

Supplementary Figure 6



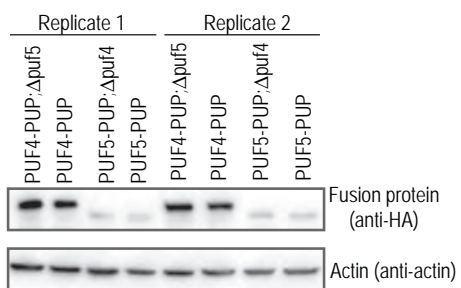
Supplementary Figure 6. (a) Plot of the fraction of the indicated groups for targets present in class I or II. For example, ~45% of genes U-Tagged solely by Puf3p (Puf3p only) were present in class I or II of Puf3p, and ~45% of genes U-tagged by both Puf4p and Puf5p were present in class I or II of both data sets. **(b)** Plot of the fraction of each class of Puf4p targets that were U-tagged by both PUF4-PUP and PUF5-PUP. For example, ~80% of class I Puf4p targets were also bound by Puf5p. **(c)** Plot of the fraction of each class of Puf5p targets that were U-tagged by both PUF4-PUP and PUF5-PUP. For example, ~55% of class I Puf5p targets were also bound by Puf4p. **(d)** Enrichment of the indicated groups for 8BEs, 9BEs, and 10BEs relative to all mRNAs. **(e)** Pie charts showing the distribution of PBEs by number (top) and length (bottom) in genes present in class I or II of both Puf4p and Puf5p targets (red genes in panel a). **(f)** Box plot of 9BE and 10BE conservation scores for all mRNAs and the indicated groups of PUF targets. “Shared Class I-II” indicates the genes were present in class I or II of both Puf4p and Puf5p targets.

Supplementary Figure 7



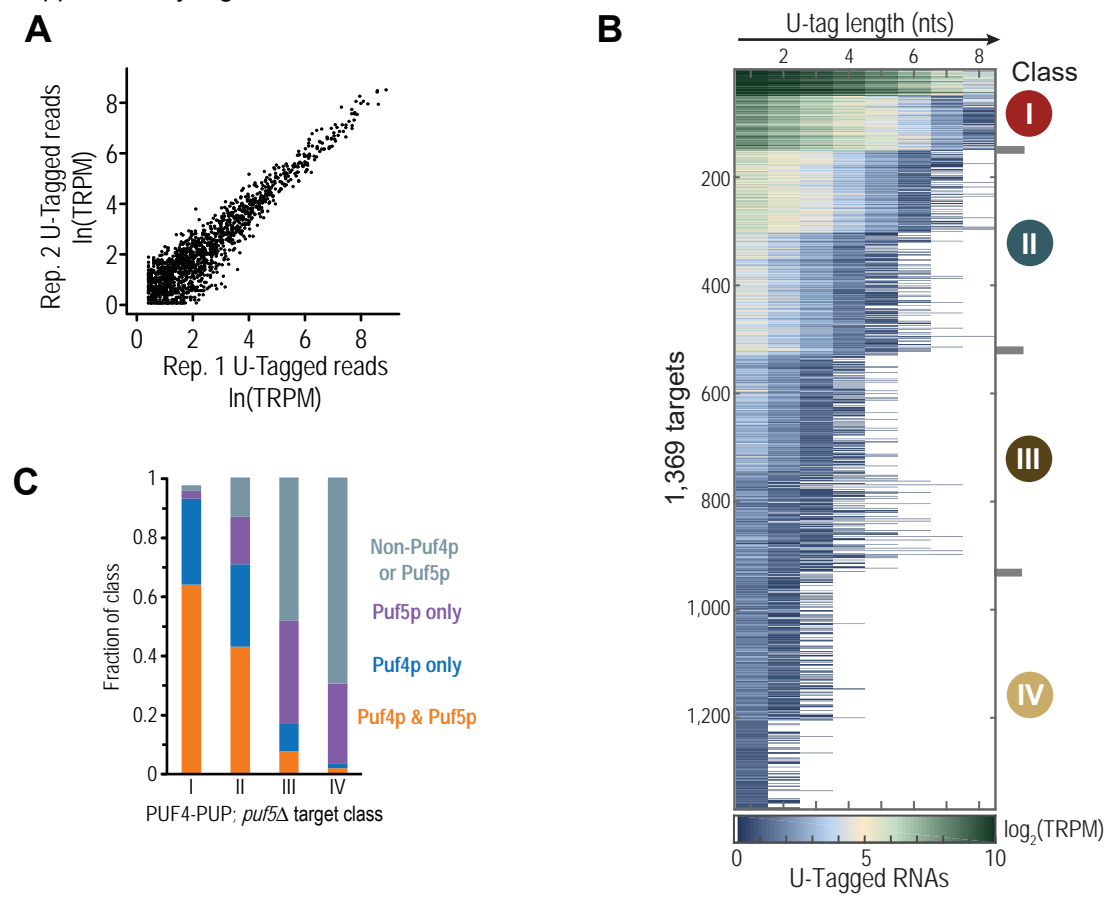
Supplementary Figure 7. Pictures of yeast growth assays of the indicated strains in the presence of the indicated drugs. Yeast were spotted via 10-fold serial dilutions (from X to Y cells). In panel *a*, the “Single-copy plasmid” column indicates if a given strain also expressed *PUF4* or *PUF5* via a single-copy plasmid (CEN).

Supplementary Figure 8



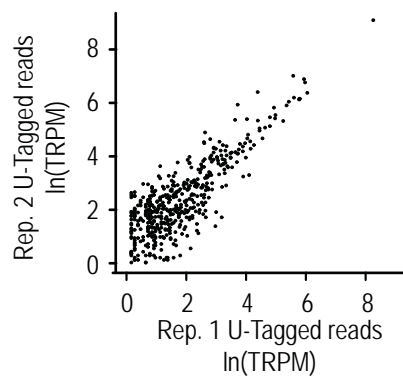
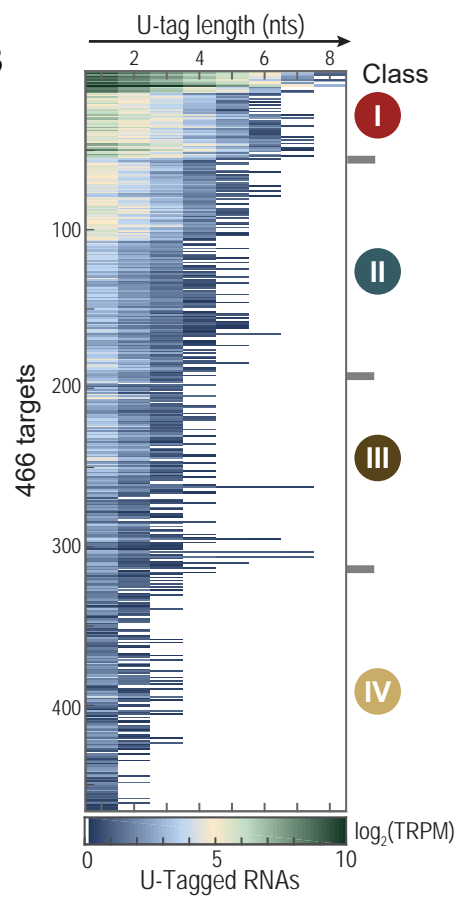
Supplementary Figure 8. Western blot that depicts relative protein abundance of PUF4-PUP and PUF5-PUP in the indicated deletion strains. Extract corresponding to 0.3 OD of yeast was loaded in each lane, and two biological replicates are shown.

Supplementary Figure 9



Supplementary Figure 9. (a) Scatter plot of the number of U-Tagged RNAs detected for each PUF4-PUP;*puf5* Δ target (1,369) in two biological replicates. TRPM, Tagged RNAs per million uniquely mapped reads. (b) Heat map displaying results of the k-means clustering analysis of all 1,369 PUF4-PUP;*puf5* Δ targets. Each row represents a single target. Columns refer to the length of the U-tag detected on reads for each gene, from at least 1 uridine (leftmost column) to at least 8 uridines (rightmost). PUF4-PUP;*puf5* Δ target classes are indicated (I, II, III, & IV). (c) Plot of the fraction of each PUF4-PUP;*puf5* Δ target class that were targets of both Puf4p and Puf5p (orange), Puf4p (blue), Puf5p (purple), or neither of the PUF proteins (gray).

Supplementary Figure 10

A**B**

Supplementary Figure 10. (a) Scatter plot of the number of U-Tagged RNAs detected for each PUF5-PUP;puf4 Δ target (466) in two biological replicates. TRPM, Tagged RNAs per million uniquely mapped reads. **(b)** Heat map displaying results of the k-means clustering analysis of all 466 PUF5-PUP;puf4 Δ targets. Each row represents a single target. Columns refer to the length of the U-tag detected on reads for each gene, from at least 1 uridine (leftmost column) to at least 8 uridines (rightmost). PUF5-PUP;puf4 Δ target classes are indicated (I, II, III, & IV).

REFERENCES

- 1 Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-543, doi:10.1038/nrg2111 (2007).
- 2 Lukong, K. E., Chang, K. W., Khandjian, E. W. & Richard, S. RNA-binding proteins in human genetic disease. *Trends Genet* **24**, 416-425, doi:10.1016/j.tig.2008.05.004 (2008).
- 3 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845, doi:10.1038/nrg3813 (2014).
- 4 Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336, doi:10.1146/annurev.biochem.72.121801.161720 (2003).
- 5 Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* **84**, 325-354, doi:10.1146/annurev-biochem-080111-092106 (2015).
- 6 Wickens, M., Bernstein, D. S., Kimble, J. & Parker, R. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-157, doi:S0168952501026166 [pii] (2002).
- 7 Quenault, T., Lithgow, T. & Traven, A. PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol* **21**, 104-112, doi:10.1016/j.tcb.2010.09.013 (2011).
- 8 Spassov, D. S. & Jurecic, R. The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life* **55**, 359-366, doi:10.1080/15216540310001603093 (2003).
- 9 Gerber, A. P., Herschlag, D. & Brown, P. O. Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS biology* **2**, E79, doi:10.1371/journal.pbio.0020079 (2004).
- 10 Gerber, A. P., Luschnig, S., Krasnow, M. A., Brown, P. O. & Herschlag, D. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in

- Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **103**, 4487-4492, doi:10.1073/pnas.0509260103 (2006).
- 11 Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology* **6**, e255, doi:10.1371/journal.pbio.0060255 (2008).
- 12 Galgano, A. *et al.* Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS One* **3**, e3164, doi:10.1371/journal.pone.0003164 (2008).
- 13 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141, doi:10.1016/j.cell.2010.03.009 (2010).
- 14 Morris, A. R., Mukherjee, N. & Keene, J. D. Ribonomic analysis of human Pum1 reveals cis-trans conservation across species despite evolution of diverse mRNA target sets. *Mol Cell Biol* **28**, 4093-4103, doi:10.1128/MCB.00155-08 (2008).
- 15 Chen, D. *et al.* Pumilio 1 suppresses multiple activators of p53 to safeguard spermatogenesis. *Curr Biol* **22**, 420-425, doi:10.1016/j.cub.2012.01.039 (2012).
- 16 Kershner, A. M. & Kimble, J. Genome-wide analysis of mRNA targets for *Caenorhabditis elegans* FBF, a conserved stem cell regulator. *Proc Natl Acad Sci U S A* **107**, 3936-3941, doi:10.1073/pnas.1000495107 (2010).
- 17 Lapointe, C. P., Wilinski, D., Saunders, H. A. & Wickens, M. Protein-RNA networks revealed through covalent RNA marks. *Nat Methods* **12**, 1163-1170, doi:10.1038/nmeth.3651 (2015).
- 18 Wilinski, D. *et al.* RNA regulatory networks diversified through curvature of the PUF protein scaffold. *Nat Commun* **6**, doi:10.1038/ncomms9213 (2015).
- 19 Prasad, A. *et al.* The PUF binding landscape in metazoan germ cells. *RNA* **22**, 1026-1043, doi:10.1261/rna.055871.116 (2016).

- 20 Porter, D. F., Koh, Y. Y., VanVeller, B., Raines, R. T. & Wickens, M. Target selection by natural and redesigned PUF proteins. *Proc Natl Acad Sci U S A* **112**, 15868-15873, doi:10.1073/pnas.1508501112 (2015).
- 21 Goldstrohm, A. C., Hook, B. A., Seay, D. J. & Wickens, M. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol* **13**, 533-539, doi:10.1038/nsmb1100 (2006).
- 22 Goldstrohm, A. C., Seay, D. J., Hook, B. A. & Wickens, M. PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J Biol Chem* **282**, 109-114, doi:10.1074/jbc.M609413200 (2007).
- 23 Hook, B. A., Goldstrohm, A. C., Seay, D. J. & Wickens, M. Two yeast PUF proteins negatively regulate a single mRNA. *J Biol Chem* **282**, 15430-15438, doi:10.1074/jbc.M611253200 (2007).
- 24 Olivas, W. & Parker, R. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J* **19**, 6602-6611, doi:10.1093/emboj/19.23.6602 (2000).
- 25 Houshmandi, 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, doi:10.1261/rna.2168505 (2005).
- 26 Lee, D. *et al.* PUF3 acceleration of deadenylation in vivo can operate independently of CCR4 activity, possibly involving effects on the PAB1-mRNP structure. *J Mol Biol* **399**, 562-575, doi:10.1016/j.jmb.2010.04.034 (2010).
- 27 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 U S A* **106**, 20192-20197, doi:10.1073/pnas.0812079106 (2009).
- 28 Valley, C. T. *et al.* Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc Natl Acad Sci U S A* **109**, 6054-6059, doi:10.1073/pnas.1200521109 (2012).

- 29 Miller, M. T., Higgin, J. J. & Hall, T. M. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol* **15**, 397-402, doi:10.1038/nsmb.1390 (2008).
- 30 Campbell, Z. T. *et al.* Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep* **1**, 570-581, doi:10.1016/j.celrep.2012.04.003 (2012).
- 31 Bailey, T. L. Discovering novel sequence motifs with MEME. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2 4, doi:10.1002/0471250953.bi0204s00 (2002).
- 32 Sun, M. *et al.* Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell* **52**, 52-62, doi:10.1016/j.molcel.2013.09.010 (2013).
- 33 Taylor, J. W. & Berbee, M. L. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* **98**, 838-849 (2006).
- 34 Hebert, A. S. *et al.* The one hour yeast proteome. *Mol Cell Proteomics* **13**, 339-347, doi:10.1074/mcp.M113.034769 (2014).
- 35 Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319-324, doi:10.1038/nmeth.2834 (2014).
- 36 Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172-177, doi:10.1038/nature12311 (2013).
- 37 Russo, J. & Olivas, W. M. Conditional regulation of Puf1p, Puf4p, and Puf5p activity alters YHB1 mRNA stability for a rapid response to toxic nitric oxide stress in yeast. *Mol Biol Cell* **26**, 1015-1029, doi:10.1091/mbc.E14-10-1452 (2015).
- 38 McMahon, A. C. *et al.* TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* **165**, 742-753, doi:10.1016/j.cell.2016.03.007 (2016).

- 39 Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453-466, doi:10.1016/j.cell.2012.12.023 (2013).
- 40 Pandit, S. *et al.* Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell* **50**, 223-235, doi:10.1016/j.molcel.2013.03.001 (2013).
- 41 Wang, E. T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**, 710-724, doi:10.1016/j.cell.2012.06.041 (2012).
- 42 Han, H. *et al.* MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**, 241-245, doi:10.1038/nature12270 (2013).
- 43 Ascano, M., Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382-386, doi:10.1038/nature11737 (2012).
- 44 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 45 Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033-1037, doi:10.1038/nature07728 (2009).
- 46 Miura, F. *et al.* Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics* **9**, 574, doi:10.1186/1471-2164-9-574 (2008).
- 47 Siwiak, M. & Zielenkiewicz, P. A comprehensive, quantitative, and genome-wide model of translation. *PLoS Comput Biol* **6**, e1000865, doi:10.1371/journal.pcbi.1000865 (2010).
- 48 Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691, doi:10.1038/nature02026 (2003).

- 49 Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54-U36, doi:10.1038/nature06107 (2007).
- 50 Stajich, J. E. *et al.* FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Research* **40**, D675-D681, doi:10.1093/nar/gkr918 (2012).

CHAPTER 6: Perspectives

Parts of this chapter will be included in an opinion essay written by myself and Dr. Marv Wickens.

My graduate research is just a small rock, a pebble really, circling a single star in an expanding universe comprised of poly(U) polymerases, RNA Tagging, PUF proteins, and protein-RNA networks. I already have discussed many pertinent points directly related to each of those topics in previous chapters. In the coming chapter, I will focus on new adaptations and uses for RNA Tagging, and several open questions related to the canonical PUF regulatory network. I will end with some thoughts on the genomic-age of RNA Biology and challenges that the field faces.

The future of RNA Tagging

RNA Tagging is a useful approach to study RNA-binding proteins from *S. cerevisiae*. I have used the approach to identify RNAs bound by four proteins: Puf3p, Puf4p, Puf5p, and Bfr1p. I employed the approach to identify RNAs bound by each protein, which pointed to their biological function. In the case of Bfr1p, our data clarified inconsistencies in the literature, and support a role for the protein in the localization of mRNAs to the ER. In parallel, RNA Tagging revealed that Puf4p and Puf5p likely co-regulate many RNAs, which was missed by a previous approach. It also helped us uncover that the Puf4p-Puf5p sub-network in the PUF regulatory network is intricately balanced through an interplay of protein and RNA relative abundances.

The current implementation of RNA Tagging represents only the first step in its maturation. To reach full potential, RNA Tagging must be adapted to other systems and organisms. Using genome engineering technologies, RNA Tagging fusion proteins can be expressed in specific tissues, cells, or cell sub-types, and the U-tagged RNAs identified *in vitro*. However, in such studies endogenous U-adding and removing enzymes may complicate experiments. Thus, knockdown of the endogenous uridylation machinery

and degradation pathways may be needed. Nevertheless, genetic approaches like RNA Tagging are poised to identify RNAs bound by RBPs in specific cells in live animals. Indeed, the Rosbash group at Brandeis independently developed an approach called TRIBE that adheres to the same principle as RNA Tagging – covalently mark bound RNAs *in vivo* (1). They used TRIBE to identify RNAs bound by particular proteins from a few neurons in fruit flies, nicely demonstrating the power of such approaches.

Time-resolved and single-cell experiments with RNA Tagging will also open new experimental doors. The development of an inducible expression system for the RNA Tagging fusion protein will enable the protein to be expressed at specific time points and for specific durations of time. Such a system will allow the dynamics of RBPs to be probed, and might yield insight into which RNAs are first bound by RBPs, which RNAs are bound at particular time points, and how interactions change with time or condition. Furthermore, RNA Tagging should enable identification of RNAs bound by RBPs in single-cells, which is currently inaccessible to other approaches. Single-cell experiments would provide insight into the variability of protein-RNA interactions from cell-to-cell, a relative unknown.

The development of new U-tagging enzymes and other tagging enzymes will also be beneficial. As is, RNA Tagging detects RNAs that are directly bound by the RBP of interest, as *C. elegans* PUP-2 lacks RNA-binding domains of its own (2-4). However, RNAs bound by multiple proteins co-localize, such as in RNA granules (5). Thus, employing a poly(U) polymerase such as TUT7, which has its own RNA-binding domains, may facilitate detection of co-localized RNAs (3). In parallel, enzymes with unusual nucleotide specificities, such as for C or G, would mitigate complications from endogenous uridylation machinery. The added tags would also enable multiple RBPs to

be probed at once, which would facilitate identification of RNAs bound by both proteins and may yield insight into the dynamics of their binding. Similarly, an engineered, split tagging enzyme would enable identification of RNAs that are simultaneously bound by two proteins, a question that is recalcitrant to current technologies and approaches.

The canonical PUF regulatory network

In Chapters 4 and 5, I demonstrated that the canonical yeast regulatory network is composed of four major subnetworks. The Puf3p sub-network represses mitochondrial biogenesis during fermentation (Chapter 4), and is distinct from Puf4p (Chapter 5) and Puf5p sub-networks (Chapter 5 & Appendix 1). In contrast, the Puf4p and Puf5p sub-networks are highly interconnected and involved in the regulation of ribosome biogenesis and chromatin (Chapter 5). My findings raise intriguing questions that warrant further investigation, a few of which I discuss below.

In our collaboration with Jon Stefely and the Pagliarini lab, we demonstrated that Puf3p represses mitochondrial biogenesis during fermentation (Chapter 4). Puf3p bound and subsequently reduced protein levels for the majority of mitochondrial ribosomal proteins and many electron transport chain biogenesis factors. Four hours after a diauxic shift (glucose to glycerol), however, loss of *PUF3* resulted in very few changes to the proteome, metabolome, and lipidome. Thus, a primary regulatory role of Puf3p appears to be repression of mitochondrial biogenesis factors. However, a recent study found that Puf3p was required for optimal translation of a few mRNAs in the earliest stages after a diauxic shift (1 and 2 hour time points) (6). The precise role of Puf3p during and after the diauxic shift remains to be determined, especially in light of directed studies that suggest Puf3p remains bound to a few mRNAs during respiration (7). Regardless, our data

illustrate that a primary function of Puf3p is to act as a post-transcriptional “gatekeeper” of mitochondrial biogenesis factors during fermentation.

Both Puf4p and Puf5p bind to a large cohort of mRNAs primarily implicated in the regulation of ribosome biogenesis and chromatin related functions (Chapter 5). Why are the mRNAs bound by both proteins? Great question, with a short answer: I don't know. Allowed room to speculate, regulation by multiple proteins likely enables dynamic regulation of mRNAs bound by both PUF proteins and thereby allows yeast greater potential to respond to stimuli and stress. While the RNA-binding domains of Puf4p and Puf5p are highly similar, the N- and C-terminal regions of the proteins are dissimilar, suggesting that they may have unique protein partners or post-translational modifications that impact activity and regulation. Since *PUF4* partially compensates for loss of *PUF5*, my data also suggest that while the two proteins have some overlapping functions, they also likely possess distinct regulatory mechanisms. Previous studies have found that Puf5p can both repress translation and destabilize mRNAs, while Puf4p destabilizes mRNAs without translation repression activity (8-10), supporting my hypothesis. However, overexpression of *PUF4* did improve its compensation for loss of *PUF5*, so it is possible that partial compensation is simply explained by a lack of Puf4p to regulate both its own and Puf5p's targets.

The dynamics of the canonical PUF regulatory network is an intriguing frontier ripe for exploration (**Figure 1**). For example, my finding that the Puf4p and Puf5p sub-networks rewire in response to loss of either protein illustrates that the PUF regulatory network is highly plastic. Similarly, the regulatory activities of Puf3p, Puf4p, and Puf5p are modulated by carbon source and other environmental stimuli (6, 7, 11). Thus, it will

be interesting to map and analyze the PUF network in different carbon sources, in different conditions, and at different time points. These analyses could shed light onto the biological relevance of high-affinity—low-affinity crosstalk among PUF proteins, and why many RNAs are co-regulated by Puf4p and Puf5p.

BROADER PERSPECTIVES

The generation of large numbers of interaction datasets for RNA-binding proteins provides a vast resource for understanding their function *in vivo*. However, it also raises a general question: what do we mean when we refer to an RNA as a “target” of an RNA-binding protein? Is a “target” any RNA that interacts with the protein, or should that label be reserved for RNAs that are regulated by the protein? Approaches to globally identify RNAs bound by proteins often identify hundreds, if not thousands, of interactions. Out of the sea of interactors, it is often difficult to identify biologically relevant interactions (*i.e.* those that lead to detectable regulation), as they can be obfuscated by transient interactions with unclear relevance. As we delve ever further into global interaction studies, particularly those that involve multiple RNA-binding proteins and the RNAs they bind, these questions will become increasingly important (12-16). We hope what follows will serve as a small seed for collective discussion on how to better highlight biologically relevant protein-RNA interactions.

The semi-arbitrary nature of genomic-age studies

Let us start with a hypothetical example – the protein “DOG”. DOG co-purifies 1,000 RNAs and directed studies have demonstrated that DOG destabilizes a few RNAs.

Thus, we would expect RNAs regulated by DOG protein to be at higher levels in cells without the protein than in cells with it. To globally measure such effects, we often turn to RNAseq, which will provide a measure of the steady-state abundance of every RNA in both conditions (with and without DOG protein). We can determine the effect of DOG on every expressed RNA by calculating a fold change between the two conditions, after which we'll be left with a list of fold-changes for every RNA detected in the cells.

Most of us have been there. We all have spent hours, days, weeks, months, or perhaps years staring at lists trying to set cutoffs. In our example, we have to decide on two for every RNA: the threshold for being co-purified by DOG, and the threshold for "regulation". For co-purification, we often consider how many reads were detected for a given RNA or other metrics of co-purification quality, which often results in a p-value for each gene. There are many great algorithms available to help us with this process, but at the end of the day we have to interpret the data. We have to decide if we only consider RNAs with $P < 0.01$ detected by 100 or more reads, or if we should consider any RNA with $P < 0.05$? Or is our cutoff too generous? Should we only consider RNAs with $P < 0.0001$ and detected by 500 reads?

We face the same questions for the RNAseq experiment. We know DOG destabilizes a few RNAs, so we expect many RNAs to increase in abundance in the cells without the protein. We see plenty of examples where that occurs. But, should a 25% increase be our threshold, or should we only look at RNAs that at least double in abundance? We also see many examples of RNAs that decrease in abundance. Are these direct or indirect effects? To find out, we return to our interaction list to see if any

were co-purified by DOG. But which interaction list do we use? The list with stringent cutoffs or the less-stringent one?

I concede that I am being a bit facetious at this point, but I wonder if this whole situation is. The process we went through with our example underlies every published high-throughput experiment. We are forced to make semi-arbitrary thresholds for interaction and regulation in order to publish. Data are often shrouded in cloaks of mystery and presented as concrete, with phrases like, “DOG has 1,000 targets”. We’re guilty of it (4, 17-19), and this type of language is pervasive throughout the literature. What we really mean, however, is that DOG interacted with 1,000 RNAs that we could detect via one particular method from our favorite cells/tissue/organism/whatever in our lab during two or three experiments (maybe even done in parallel on the same days).

We need to reconcile the differences between fantasy and reality. One simple way to begin is to be cognizant of the language we use to describe data, which returns us to our question of what is a “target”. Let’s start with a review of the theory behind the methods used to identify RNAs that interact with particular proteins *in vivo*.

Method theory

Biochemical purification is a powerful and highly successful strategy to characterize RNA-binding proteins. Protein components of heterogenous nuclear ribonucleoprotein (hnRNP) particles – the original RNA-binding proteins – were isolated from cell lysates via centrifugation and subsequently characterized via gel electrophoresis (20). Cells were often irradiated with UV light, prior to purification, to covalently crosslink the protein and RNA components, which enabled highly stringent purification conditions.

With highly purified hnRNP proteins in hand, monoclonal antibodies were soon developed and enabled immunopurification of the proteins and hnRNP particles, giving rise to the first RNA immunoprecipitation (RIP) studies.

Similar strategies prevail today (14, 16, 21). RIPs have been combined with northern blots and RT-PCR to identify candidate RNAs that co-purify with particular proteins. With the rise of the genomic age, RIPs now have been combined with microarrays and high-throughput sequencing to enable global detection of RNAs that co-purify with given proteins. More recently, UV crosslinking of protein-RNA interactions has been reintroduced in crosslinking followed by immunoprecipitation (CLIP) approaches, such as HITS-CLIP, PAR-CLIP, iCLIP, eCLIP, and irCLIP (22-27). These techniques exploit the crosslink between the protein and RNA to stringently purify the complexes from crude cell lysate. The crosslink also enables partial RNase digestions to pinpoint where the protein binds in a given RNA.

Recently, two alternative approaches, TRIBE and RNA Tagging, were independently developed to detect *in vivo* protein-RNA interactions (1, 4). Both approaches adhere to an identical premise: covalently mark RNAs that are bound by particular proteins *in vivo*, and subsequently identify marked RNAs *in vitro* via high-throughput sequencing. In targets of RNA-binding proteins identified by editing (TRIBE), the catalytic domain of an RNA editing enzyme called ADAR is fused to the RBP of interest (1). Bound RNAs are covalently marked *in vivo* by adenosine to inosine conversions, which are identified via high-throughput sequencing as A-to-G SNPs. RNA Tagging (disclaimer!: developed by us) instead employs an enzyme called PUP-2 to covalently mark RNAs bound by the RBP with a 3' terminal "U-tag", which is detected by

high-throughput sequencing (4). The two approaches shift from the paradigm of biochemical purification to a more genetic-based strategy, which has proven useful to the DNA community (28, 29).

No matter the technique or strategy, global approaches often identify hundreds or thousands of RNAs that interact with particular proteins. While entire sets of RNA interactors for particular proteins have often provided tremendous insight into their biological function, do we really believe that every RNA identified is meaningful? On the other hand, do we really believe the absence of an RNA is proof that it does not interact with the protein?

Our best guess is that few, if any, researchers believe every single RNA in an interaction list is relevant in a biological context. Despite their clear power, CLIP-seq approaches rely on UV crosslinking. They thus require large amounts of cells that often must be removed from their native context to capture interactions at a single time point, which likely include very weak or transient interactors. A major breakthrough of the TRIBE and RNA Tagging approaches was their amenability to small amounts of cells, demonstrated especially well in live animals by TRIBE. However, the tagging enzymes may retain intrinsic biases for particular RNA substrates and the bound RNAs are covalently modified with A-to-I conversions or 3' uridine tails, which have unclear effects on RNA.

Thus, no approach is perfect, and the specific biological question will dictate the appropriate technique to use. The primary objective for undertaking any of these experiments is to identify the RNAs that are regulated by our favorite protein(s). We all

acknowledge that there are inherent biases in every available technique that give rise to false-negative and false-positive RNA interactors. We therefore suspect that biochemical and genetic approaches will ultimately prove complementary, to compensate for their intrinsic limitations. We all also must work together to both develop and adopt strategies that highlight biologically meaningful interactions from entire interaction datasets.

What is a “target”?

We finally can return to our question of what is a target. As used now, the term target typically refers to any RNA identified in a genome-wide approach like CLIP-seq. However, as we talked about above, the cutoff for being co-purified versus not is semi-arbitrary, and the methods to identify protein-RNA interactions are imperfect. Experiments are thus wrought by false-positive and false negative interactions, whose balance can be tweaked through adjustments to applied thresholds.

To us, the term target should be reserved for high-confidence protein-RNA interactions. We propose that a “target” should meet the following criteria: the RNA 1) interacts with a protein *in vivo*; and, 2) the interaction is detected by both biochemical and genetic strategies (*i.e.* HITS-CLIP and RNA Tagging), and/or directly leads to regulation of the RNA. We have found that such stringent approaches as above efficiently highlight protein-RNA interactions that impact cellular function (see Chapter 4), by simply excluding many weak or transient interactions with unclear relevance. Returning to our example from above, DOG co-purified 1,000 RNAs via a single technique. Thus in our experiment, DOG binds 1,000 RNA, which we refer to as “DOG-bound RNAs”. Loss of DOG protein

affected the abundance of 200 DOG-bound RNAs; thus, they are now referred to as “DOG targets”. Targets are therefore subsets of all RNAs bound by a particular protein.

We believe that the ‘bound’ versus ‘target’ distinction will provide a more descriptive means to discuss genome-wide experiments, and better highlight regulatory interactions. Proteins encounter many different RNAs in their lifetime, all with a range of potential binding sites. It is not surprising then, especially from a biochemical standpoint, that a protein will stochastically bind to whichever RNA it encounters first. We fully expect a subset of bound RNAs to be stochastic, biological noise. Thus, a simple approach to indicate that a bound RNA is regulated or a very high-confidence interaction will improve communication and ultimately our understanding of RNA biology.

Our definition of target introduces some flexibility in their experimental definition, but our definition still presents several challenges. To call a bound RNA a target, the RNA must either be detected by complementary strategies or found to be directly regulated by the protein. First, what we mean by complementary strategies is that the interaction was detected by biochemical purification (*i.e.* CLIP) and genetic-based strategies (*i.e.* RNA Tagging). This requires a lab to either be proficient in two techniques, which presents both technical and intellectual hurdles, or collaborate with a lab proficient in the complementary technique, which presents logistical challenges. However, in our hands, integrating HITS-CLIP and RNA Tagging datasets proved powerful (see Chapter 4). Such integrated analyses allowed us to rigorously define Puf3p targets, which represented only about 30% of all Puf3p-bound RNAs. Our stringent target list subsequently allowed us to trace direct and indirect Puf3p effects throughout the yeast proteome, thus mapping distinct molecular pathways directly and indirectly regulated by Puf3p. Our approach

revealed novel roles for Puf3p in the regulation of coenzyme Q biogenesis and more general lipid metabolism, and helped reveal the function of uncharacterized mitochondrial proteins.

In lieu of identification by multiple approaches, an RNA may be defined as a target if it is directly regulated by the protein. Thus, a second major challenge is how to define and detect 'regulation'. RNAs are bound by batteries of proteins throughout their lifetimes, which can splice, cap, cleave, polyadenylate, export, localize, translate, modify, stabilize, or degrade bound RNAs (30). Thus, identifying the regulatory effect of a given protein is often challenging, since there are so many potential outcomes. However, it must be addressed to yield the greatest benefit from genome-wide interaction studies. Let our metaphorical call to arms join those of others whom have proposed identical approaches (12-14). With known molecular functions in hand, it is possible to combine interaction studies (*i.e.* CLIP) with appropriate functional studies (*i.e.* RNAseq in knockdown or knockout cells or animals). Interaction and functional studies synergize with each other to illustrate which RNAs bound by the protein are regulated, thus yielding the greatest insight into function.

Another challenge is the detection of regulation and its dynamic nature. Saying an RNA is regulated by a protein is simple in theory but challenging in practice. Functional studies most often result in a continuum of regulation, from none or very little, to a lot. Our research lives would be much simpler if regulation was a binary function, but we are instead left with the struggle of setting thresholds for regulation, as we attempted to illustrate above with DOG. This process will likely become more difficult, as ever smaller changes are accurately and precisely measured by constantly improved instrumentation.

Increased instrument sensitivity and accuracy will also make it possible to test the biological relevance of more transient interactions, which may have more of a “fine-tuning” role. In parallel, the targets of an RBP are likely dynamic. RNAs bound by proteins likely vary across cell type, tissue, organism, and evolution, and similar dynamics likely occur for the targets of RBPs.

Moving forward

As we proceed, let us all not get lost in an abyss of data. The improvements to CLIP made in the recent past by eCLIP (26) and irCLIP (27) have made it increasingly easy to identify RNAs bound by proteins. TRIBE and RNA Tagging also offer straightforward approaches to identify protein-RNA interactions, and each of the techniques are easily multiplexed for many proteins or conditions. Before leaping into the chasm, let us all pledge to make time to rigorously define targets and perform functional, mechanistic, and bioinformatic studies to pinpoint which interactions are biologically relevant. High-confidence targets of an RNA-binding protein will become increasingly important as we undertake studies that dissect RNA regulatory networks formed by multiple RNA-binding proteins.

Ultimately, we envision that how targets are identified, defined, and used will vary based on the goal of an experiment and the protein in question. The initial goal of an experiment on an uncharacterized RNA-binding protein will likely be very different than with a protein like FMRP or TDP-43. For the uncharacterized protein, more stringent thresholds of interaction and/or regulation can be used to limit the number of genes to consider. Fewer numbers of targets to consider will prove beneficial in the identification

of its primary function(s). For well-studied proteins like FMRP, it will likely prove beneficial to use less stringent thresholds, which may reveal subtleties of its regulation. The knowledge-base for well-studied proteins is sufficient so that researchers can better wade through the increased number of targets to identify more elusive targets and regulatory roles.

Nevertheless, it is remarkable where we stand today. We have the ability, in just a few short months of time, to identify global targets of RNA-binding proteins and combine that data with genome-wide functional studies to identify its targets and better understand its biology.

Figure 1.

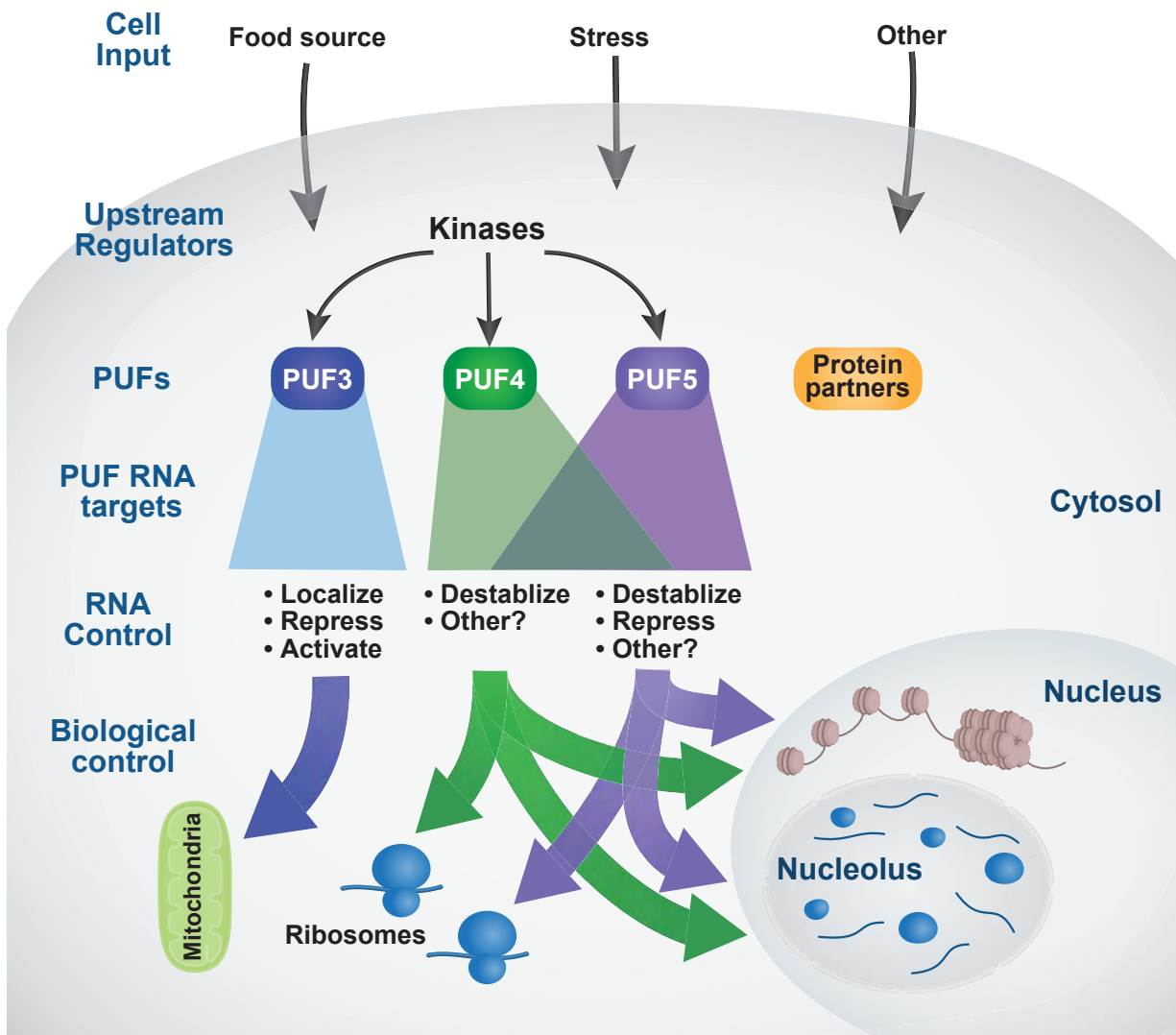


Figure 1. Cartoon schematic of the canonical PUF regulatory network present in *S. cerevisiae*. Many levels of the network remain to be explored. For example, how is the network impacted by: extracellular cues, such as carbon source and cell stressors; and kinases and phosphatases that act on Puf3p, Puf4p, and puf5p. In parallel, how do the RNAs they bind change in different conditions and how does phosphorylation affect PUF-mediated regulation?

References

1. A. C. McMahon *et al.*, TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* **165**, 742-753 (2016).
2. J. E. Kwak, M. Wickens, A family of poly(U) polymerases. *RNA* **13**, 860-867 (2007).
3. C. P. Lapointe, M. Wickens, The nucleic acid-binding domain and translational repression activity of a *Xenopus* terminal uridylyl transferase. *J Biol Chem* **288**, 20723-20733 (2013).
4. C. P. Lapointe, D. Wilinski, H. A. Saunders, M. Wickens, Protein-RNA networks revealed through covalent RNA marks. *Nat Methods* **12**, 1163-1170 (2015).
5. C. J. Decker, R. Parker, P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation. *Csh Perspect Biol* **4**, (2012).
6. C. D. Lee, B. P. Tu, Glucose-Regulated Phosphorylation of the PUF Protein Puf3 Regulates the Translational Fate of Its Bound mRNAs and Association with RNA Granules. *Cell Rep* **11**, 1638-1650 (2015).
7. M. A. Miller, J. Russo, A. D. Fischer, F. A. Lopez Leban, W. M. Olivas, Carbon source-dependent alteration of Puf3p activity mediates rapid changes in the stabilities of mRNAs involved in mitochondrial function. *Nucleic Acids Res* **42**, 3954-3970 (2014).
8. A. C. Goldstrohm, B. A. Hook, D. J. Seay, M. Wickens, PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol* **13**, 533-539 (2006).
9. A. C. Goldstrohm, D. J. Seay, B. A. Hook, M. Wickens, PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J Biol Chem* **282**, 109-114 (2007).
10. N. H. Blewett, A. C. Goldstrohm, A Eukaryotic Translation Initiation Factor 4E-Binding Protein Promotes mRNA Decapping and Is Required for PUF Repression. *Molecular and Cellular Biology* **32**, 4181-4194 (2012).
11. J. Russo, W. M. Olivas, Conditional regulation of Puf1p, Puf4p, and Puf5p activity alters YHB1 mRNA stability for a rapid response to toxic nitric oxide stress in yeast. *Mol Biol Cell* **26**, 1015-1029 (2015).
12. J. D. Keene, RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-543 (2007).
13. D. D. Licatalosi, R. B. Darnell, RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11**, 75-87 (2010).
14. J. Konig, K. Zarnack, N. M. Luscombe, J. Ule, Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* **13**, 77-83 (2011).
15. S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845 (2014).
16. E. L. Van Nostrand, S. C. Huelga, G. W. Yeo, Experimental and Computational Considerations in the Study of RNA-Binding Protein-RNA Interactions. *Adv Exp Med Biol* **907**, 1-28 (2016).

17. D. Wilinski *et al.*, RNA regulatory networks diversified through curvature of the PUF protein scaffold. *Nat Commun* **6**, (2015).
18. D. F. Porter, Y. Y. Koh, B. VanVeller, R. T. Raines, M. Wickens, Target selection by natural and redesigned PUF proteins. *Proc Natl Acad Sci U S A* **112**, 15868-15873 (2015).
19. A. Prasad *et al.*, The PUF binding landscape in metazoan germ cells. *RNA* **22**, 1026-1043 (2016).
20. G. Dreyfuss, M. S. Swanson, S. Pinol-Roma, Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* **13**, 86-91 (1988).
21. C. A. McHugh, P. Russell, M. Guttman, Methods for comprehensive experimental identification of RNA-protein interactions. *Genome biology* **15**, 203 (2014).
22. J. Ule *et al.*, CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215 (2003).
23. D. D. Licatalosi *et al.*, HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469 (2008).
24. M. Hafner *et al.*, Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).
25. J. Konig *et al.*, iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-915 (2010).
26. E. L. Van Nostrand *et al.*, Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-514 (2016).
27. B. J. Zarnegar *et al.*, irCLIP platform for efficient characterization of protein-RNA interactions. *Nat Methods* **13**, 489-492 (2016).
28. B. van Steensel, S. Henikoff, Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat Biotechnol* **18**, 424-428 (2000).
29. M. R. Hass *et al.*, SpDamID: Marking DNA Bound by Protein Complexes Identifies Notch-Dimer Responsive Enhancers. *Molecular Cell* **59**, 685-697 (2015).
30. G. Singh, G. Pratt, G. W. Yeo, M. J. Moore, The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem* **84**, 325-354 (2015).

APPENDIX 1

**RNA regulatory networks diversified
through curvature of the PUF protein scaffold**

Daniel Wilinski,^{1,3} Chen Qiu,^{2,3} **Christopher P. Lapointe**¹, Markus Nevil,¹

Zachary T. Campbell,¹ Traci M. Tanaka Hall² and Marvin Wickens^{1,✉}

This chapter was published:

Wilinski D*, Qiu C*, **Lapointe CP**, Nevil M, Campbell ZT, Hall T, and Wickens M. 2015.
Nature Communications. 6:8213 doi: 10.1038/ncomms9213. * Contributed equally

The primary findings of this paper are: 1) *S. cerevisiae* Puf5p binds to an extended binding element of 8, 9, 10, 11, or 12 nucleotides in length *in vivo*; 2) the different length sequences are accommodated by the 8 PUF repeats in Puf5p by the flexible exclusion of nucleotides away from the surface of the protein; and 3) binding length specificity of particular PUF proteins is plastic across evolution.

I collaborated with Dr. Daniel Wilinski and Markus Nevil to demonstrate primary finding # 3 above. We designed and tested chimeric proteins between *S. cerevisiae* and *S. pombe* Puf3. Using our chimeras, we illustrated that the region responsible for *S. pombe* Puf3 broadened specificity is housed in a single PUF repeat, which we used to engineer broadened specificity in *S. cerevisiae* Puf3p. These experiments and data are presented in Figure 6.

¹ Department of Biochemistry, University of Wisconsin, Madison, WI 53706

² Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

³ Authors contributed equally

✉ Correspondence: wickens@biochem.wisc.edu

ABSTRACT

Proteins bind and control mRNAs, directing their localization, translation, and stability. Members of the PUF family of RNA-binding proteins control multiple mRNAs in a single cell, and play key roles in development, stem cell maintenance and memory formation. Here we identified the mRNA targets of a *S. cerevisiae* PUF protein, Puf5p, by UV-crosslinking-affinity purification and high-throughput sequencing (HITS-CLIP). The binding sites recognized by Puf5p are diverse, with variable spacer lengths between two specific sequences. Each length of site correlates with a distinct biological function. Crystal structures of Puf5p-RNA complexes reveal that the protein scaffold presents an exceptionally flat and extended interaction surface relative to other PUF proteins. In complexes with RNAs of different lengths, the protein is unchanged. A single PUF protein repeat is sufficient to induce broadening of specificity. Changes in protein architecture, such as alterations in curvature, may lead to evolution of mRNA regulatory networks.

INTRODUCTION

RNA-binding proteins control an mRNA's life, including its translation, movement, and destruction. These events underlie diverse biological processes, ranging from early development to memory formation. Regulatory proteins bind simultaneously to short RNA sequences, typically in 3' untranslated regions (3'UTRs), and to protein effectors that determine the RNA's fate. The RNA binding specificities of the proteins determine which mRNAs are controlled, while effectors determine the outcomes.

RNA regulatory networks, in which a single RNA binding protein controls multiple mRNAs, are widespread¹. For example, Cytoplasmic Polyadenylation Element Binding protein, an RNA recognition motif-containing protein, binds to and regulates many mRNAs that participate in the regulation of embryonic cell cycles², while Nova co-regulates multiple mRNAs with roles in alternative polyadenylation and splicing³. As a result, RNA-binding proteins integrate post-transcriptional controls, as DNA-binding proteins coordinate transcriptional regulation. To understand RNA regulatory circuits in molecular terms, we need to know which mRNAs are controlled, how they are recognized, and how the networks change during evolution.

PUF proteins are exemplary mRNA regulators^{4,5}. They bind to the 3'UTRs of many mRNAs and do so through single-stranded RNA binding elements. For example, Puf3p of yeast binds nuclear-encoded mRNAs with roles in mitochondria⁶. Similarly, PUF proteins in *C. elegans*, *Drosophila* and humans control an overlapping battery of mRNAs with established roles in stem cells⁷. The RNA-binding specificities of most PUF proteins are defined in part by three amino acids (tripartite recognition motifs, or TRMs) in each of eight tandemly reiterated PUF repeats⁸. TRMs, annotated as XY-Z, recognize specific

bases through edge-on (residues X and Y) and stacking interactions (residue Z)⁹. Specificity also can be achieved through the requirement for a base that does not contact the protein, but is solvent-exposed^{8,9}. RNA-immunoprecipitation and microarray (“RIP-Chip”) studies suggested that *Saccharomyces cerevisiae* Puf5p binds ~200 mRNAs that contain 10 nt long binding elements⁶. Genetic analysis has implicated Puf5p in multiple cellular functions, including lifespan¹⁰, cell wall integrity¹¹, chromatin structure¹⁰, and mating type switching¹², consistent with the view that it participates in the control of diverse groups of mRNAs.

In this work, we used UV-crosslinking and high-throughput sequencing to define ~1000 high-confidence RNA targets of Puf5p *in vivo*. These targets possess unexpected diversity of binding element lengths, with the same RNA sequence features at the two ends but varying numbers of nucleotides in between. The lengths of sites correlate with the biological functions of the targets. The crystal structures of Puf5p – RNA complexes revealed that the RNAs assume altered conformations to accommodate a fixed protein architecture. The plasticity in binding element length is driven by the flattened curvature of the PUF protein scaffold. The findings suggest ways in which alterations in protein curvature result in new specificities and enable the evolution of new RNA networks.

RESULTS

Identification of Puf5p RNA targets

Using *in vivo* UV crosslinking *and* high-throughput sequencing (HITS-CLIP^{13,14}), we identified more than 1,000 mRNAs to which Puf5p binds in *S. cerevisiae*, representing 16% of the yeast transcriptome (Fig. 1a). The strain analyzed contained a *PUF5* gene fused to a tandem affinity purification (TAP) tag. The tagged gene was integrated by homologous recombination into the *PUF5* locus. Cells were irradiated during mid-log phase (Fig. 1a i). After lysis and mild RNase treatment (Fig. 1a ii), Puf5p was stringently purified through tandem affinity steps (Fig. 1a iii) and SDS PAGE electrophoresis. The purification of cross-linked complexes was effective, as evidenced by Western blotting (Fig. 1b). Complexes whose RNA components had been ³²P-end-labeled exhibited heterogenous, slower mobilities than Puf5p alone (Fig. 1c). To identify RNAs bound to Puf5p, adaptors were ligated, the protein digested, and the RNAs converted to cDNAs that were analyzed by high throughput sequencing (Fig. 1a iv). The adaptors contained random bar-codes, so that PCR duplication events could be discarded.

Upon aligning the sequence reads to the yeast genome, we found the majority of peaks were within 3'UTRs of mRNAs and that the set of target mRNAs were distinct from, but overlapped, those of other yeast PUF proteins. We obtained 16,300,145 and 11,100,468 reads from two biological replicates. Of these, 616,401 and 491,532 (6% and 7%) mapped to unique locations in the *S. cerevisiae* genome, after filtering by quality score and removing PCR duplicates (Fig. 1a v; Supplementary Fig. 1a). The functional enrichment of targets detected was only minimally affected by changing the filtering methods we used (Supplementary Fig. 1b). A total of 1,439 peaks were identified,

representing a total of 1,190 RNAs (Supplementary Data Set 1). Of these, 1,043 (88%) mapped to mRNAs. The remaining 12% mapped to non-coding RNAs, including snoRNA, snRNA, ncRNA, and tRNAs (Supplementary Fig. 2). Of the peaks in mRNAs, most resided in 3'UTRs (68%), but a fraction mapped to ORFs (28%) or 5'UTRs (4%) (Fig. 1d, see Supplementary Data Set 1 for complete gene list). Approximately 7% of Puf4p targets defined by RIP-chip overlap with those of Puf5p (Supplementary Fig. 3b). The overlap with Puf3p targets is 30% by CLIP analysis and 2% by RIP-chip (Supplementary Fig. 3c, d). The 1,043 Puf5p targets represent 16% of yeast mRNAs, a five-fold increase relative to the 206 targets detected in earlier RIP-Chip studies⁶ (Supplementary Fig. 3a) (see Discussion). We conclude that Puf5p is a broad regulator of a distinct set of mRNAs in *S. cerevisiae*.

Puf5p binding elements range in length from 8 to 12 nt

We developed stringent criteria to select a set of 1,043 high-confidence targets that we used to identify RNA sequence elements bound by Puf5p. We first defined significant peaks as an enrichment of independent reads in a specific genic region ($\text{modFDR} < 0.01$)¹⁶ (Supplementary Fig. 1a). To identify high-confidence targets, we required that a peak contain 1-nt or 2-nt deletions in multiple reads (a strong indicator that these RNAs had been cross-linked to Puf5p¹⁷) and a minimum of 10 reads per peak. In addition, these criteria had to be satisfied in both biological replicates (Supplementary Fig. 1a). Normalized peak heights at specific loci were reproducible between the biological replicates (Pearson correlation coefficient 0.90) (Fig. 1e). Previous studies validated several putative targets of Puf5p by showing they were regulated by that protein

in vivo^{18,19}. Among the best characterized are *SMX2* and *HO* mRNAs^{20,21}, which are used here as examples (Fig. 2a). With both mRNAs, peaks lay over the previously characterized binding elements in their 3'UTRs (Fig. 2a). (For additional examples, see Supplementary Fig. 5a-c.)

We identified five classes of Puf5p binding elements ranging from 8-12 nucleotides, each comprising a 5'-UGUA tetranucleotide sequence and a 3'UA with a variable length spacer region in between (Fig. 2b). We performed an unbiased search of the complete set of high-confidence targets for over-represented sequences in peaks using MEME²². The position weight matrix we obtained consists of a 5'UGUA tetranucleotide sequence followed by a degenerate 3' end (Fig. 2b). However, we could de-convolute the complete set of 5'UGUA-containing sequences into five classes of binding elements, ranging in length from 8 to 12 nts beginning at the 5'UGUA, each with a 3'terminal UA sequence (Fig. 2b). 71% of the 1,043 targets, and 66% of the total number of peaks (1,439), contained at least one Puf5p binding element. Peaks without enriched sequences may reflect contacts that were less sequence-specific or mediated by interactions between Puf5p and RNA-bound factors. The sequences between UGUA and UA display little difference compared to the background nucleotide frequencies surrounding the sites (Supplementary Fig. 4), though adenosine was modestly enriched 1 or 2 nts upstream of the 3'terminal UA in 9 and 10 nt elements (Supplementary Fig. 4a), and guanosines were uncommon (Supplementary Fig. 4c).

The breadth of binding element lengths associated with Puf5p is unusual among PUF proteins (Fig. 2c). For example, three other PUF proteins – human PUM2¹⁴, *S. cerevisiae* Puf4p⁶ and *C. elegans* FBF-2⁷ – show a single dominant length of site *in vivo*,

measured either by CLIP methods²³ or inferred from RIP-Chip²⁴ (Fig. 2c). Essentially the same behavior was observed for each protein *in vitro*²⁵. To examine the sequence preferences of Puf5p *in vitro*, the purified protein was incubated with an RNA library in which 20 consecutive nucleotides had been randomized, generating a theoretical complexity of 4²⁰ RNA sequences²⁵. Bound RNAs were eluted, and the process repeated five times. The RNAs were analyzed by high-throughput sequencing. In this method, termed SEQRS, the number of reads obtained is a proxy for affinities measured *in vitro*²⁵. Re-analysis of data obtained with Puf5p²⁵ revealed that the number of reads for each site length yielded a pattern similar to that seen in HITS-CLIP, in that 9-nt and 10-nt sites were the most abundant (Fig. 2c). 8, 11, and 12 nt sites were less prevalent in SEQRS than *in vivo*, but above background *in vitro*. Indeed, for each protein analyzed, we observed greater binding for sub-optimal lengths in the cell than *in vitro* (Fig. 2c). Many factors affect binding *in vivo*, including protein-protein interactions and RNA accessibility.

Site length correlates with biological function

The majority of the sites bound by Puf5p consist of individual elements, in which only a single site of unambiguous length is present in the CLIP peak (Fig 3a and Supplementary Table 1). Gene Ontology (GO) analysis of the Puf5p targets suggests that each binding element length found in mRNAs correlates with a distinct biological role (Fig. 3b, Supplementary Table 2). Surprisingly, when we analyzed mRNAs with different binding site lengths separately, RNAs with 8-nt binding elements were overrepresented for mitochondrion organization (p-value 9.5e-4); 9-nt sites for ribosome biogenesis (p-value 3.6e-14); and 10-nt sites for regulation of gene expression (p value 2.9 e-6), 11-nt

sites for translation (p-value 3.6e-3) (Fig. 3b). 12 nt sites did not correlate with a specific GO term. GO analysis of the mRNAs with binding elements in ORFs or 5' UTRs, revealed that only 10 nt binding elements in ORFs were associated with a GO term, positive regulation of pseudohyphal growth (p-value 8.1E-3).

The 8-nt Puf5p elements lie in a subset of mRNAs that also were bound to Puf3p in PAR-CLIP experiments¹⁵. Puf3p associated with ~1,000 nuclear-encoded mRNAs with mitochondrial functions¹⁵. Even if the criteria selecting high-confidence Puf5p targets are relaxed – not filtering for gapped reads, for example – a very similar enrichment emerges (Supplementary Fig. 1b). Similarly, 22% of the 9 nt Puf5p elements lie in mRNAs that bind Puf4p and are enriched for genes with ribosome assembly and nucleolar functions. Of the Puf5p targets with GO annotations mitochondrion organization or ribosome biogenesis, 46% and 31% are Puf3p or Puf4p targets, respectively⁶. We suggest that the restricted specificities of Puf3p and Puf4p for 8 and 9 nt sites, respectively, underlie the correlations between Puf5p length of binding sites and their biological functions. In particular, we suggest that the broadened specificity of Puf5p enabled its recruitment to pre-existing RNA regulatory circuits in the *S. cerevisiae* lineage (see Discussion).

Alternate RNA conformations adapt to fixed protein scaffold

To understand how Puf5p accommodates a wide range of target site lengths, we determined crystal structures of the Puf5p RNA-binding domain bound to RNAs of 9-12 nucleotides (Fig. 4, Supplementary Table 3), including the recognition sequences in *SMX2* (9-nt, 2.35 Å resolution), *MFA2* (10-nt, 2.15 Å resolution), *AAT2* (11-nt, 2.5 Å resolution), and *AMN1* (12-nt, 2.8 Å resolution) mRNAs (sites of 8 nts did not crystallize).

Each RNA corresponded to a high-confidence 3'UTR binding element in HITS-CLIP (Fig. 2a and Supplementary Fig. 5). The RNAs bound with affinities that corresponded to our structural observations, in that higher affinity RNA binding sites correlated with larger numbers of protein:RNA contacts (Fig. 4, Supplementary Fig. 6, Supplementary Table 4).

Despite the varying lengths and sequences of the RNA binding sites, the overall conformation of Puf5p was unchanged in the four crystal structures (rmsd < 0.7 Å over all C α atoms or <1.1 Å over all protein atoms). The protein scaffold comprises eight α -helical repeats flanked by a short N-terminal sequence and a C-terminal helix (R8') (Fig. 4a). The C-terminal repeats 5-8 bound the 5'UGUA RNA sequence, while repeats 1 and 2 bound the UA 3' element (Fig. 4a, b). Repeats 3 and 4 lie opposite the variable central regions of the RNAs (Fig. 4c-f). While the overall architecture of Puf5p resembles that of other PUF proteins, Puf5p's repeats are more irregular in length and structure than seen in human PUM1 and *S. cerevisiae* Puf3p and Puf4p (Supplementary Fig. 7a). For example, repeats 7 and 8 in Puf5p are unusually long (64 and 72 residues versus 36 in a typical repeat) with extended α 2 and α 3 helices and inter-helix loops. The positions of the α 3 helices relative to the α 1 and α 2 helices are also more varied in Puf5p than Puf3p or Puf4p (Supplementary Fig. 7a).

Since the curvature of the Puf5p scaffold is fixed, RNAs of different lengths adopt different conformations, as described below. Recognition of the 5'UGUA and 3'UA elements by repeats 5-8 and 1-2, respectively, are identical in all structures. Differences in RNA conformation and recognition are found opposite the central repeats 3 and 4.

“5-parallel:” 9 nt site. Puf5p binds to the 9-nt SMX2 RNA site by recognizing all but the central fifth base (Fig. 4c). Bases 1-4 and 6-9 are each recognized by a PUF

repeat (Supplementary Fig. 5a). However, the fifth base, C5, lies in an atypical conformation, in which the plane of the base is parallel to the axis of the protein, within van der Waals bonding distance of the side chain of Cys 381 in repeat 5 (Fig. 4c). The ribose rings of C5 and U6 adopt C2'-endo conformations to accommodate positioning base C5.

“8-flipped:” 10 nt site. Puf5p binds the 10-nt *MFA2* site similarly to the 9-nt *SMX2* site, but an additional base is accommodated by turning the 8th base away from the RNA-binding surface opposite repeat 3 (Fig. 4d). The positions for all but the 8th base overlap with the 9 nt *SMX2* RNA, and the protein:RNA recognition pattern is similar, though the 7th base is a uracil in *MFA2* and an adenine in *SMX2* (Fig. 4b, Supplementary Fig. 5a).

“5-stacked:” 11 nt site. Puf5p appears to recognize only the 5'-UGUA conserved element and two additional 3' bases of the 11-nt *AAT2* site (Fig. 4e, Supplementary Fig. 5b), consistent with weaker binding of Puf5p to this site than 9- or 10-nt sites (12- or 2-fold weaker binding, respectively, Supplementary Table 4). A 2.5 Å crystal structure of Puf5p:*AAT2* reveals electron density for bases 1-5 and for two 3' bases bound to Puf5p repeats 1 and 2. In contrast to the parallel orientation of base 5 in 9- and 10-nt sites, base A5 of *AAT2* stacks directly with base A4 and forms a van der Waals contact with the side chain of Cys381 in Puf5p repeat 5 (Fig. 4e). We will refer to this conformation as 5-stacked. Using the consensus sequences as a guide, we modeled the 3' bases as the conserved U10 and A11 bases and did not model bases 6-9. However, alternate conformations of the RNA are possible, including a conformation similar to that of the 12-nt *AMN1* site.

“Triple-stacked:” 12 nt site. Puf5p binds to the longer 12-nt *AMN1* site with a distinct RNA conformation. Unlike the conformations of the shorter length binding sites, bases A4, A5, and C6 stack directly with each other opposite repeat 5. Residues in Repeat 5 (Cys381 and Lys385) contact bases A5 and C6 (Fig. 4f). Puf5p repeat 4 does not interact with an RNA base using its edge-interacting residues, but base U7 is bound to repeat 3 (Fig. 4f). Electron density was observed for bases 1-7 and two 3' bases bound to Puf5p repeats 1 and 2. We modeled the 3' bases as the conserved U11 and A12 bases, as we did for the 11-nt *AAT2* site, and bases 8-10 were not included in the model.

Curvature as a determinant of specificity

The flatter RNA-binding surface of Puf5p contributes to its specificity by creating a more extended RNA-binding surface. Puf5p possesses the least curved RNA-binding surface observed among PUF proteins to date (Supplementary Fig. 7b) and binds to the longest RNA target sequences identified thus far. Puf3p preferentially binds 8-nt sites and exhibits the greatest curvature among the yeast PUFs (Supplementary Fig. 7b); this reflects the regular spacing of RNA-binding helices, which matches the spacing of bases in an extended RNA chain²⁶. Puf4p, which binds 9-nt binding sites, is intermediate in curvature, between Puf3p and Puf5p (Supplementary Fig. 7b).

Extension of the Puf5p RNA-binding surface is produced by the structural arrangements in repeats 4 and 5 and corresponds to the variability in Puf5p target sequence length relative to other PUF proteins. The largest repeat-to-repeat angle in Puf5p is centered about repeat 5 (Supplementary Fig. 7c). Repeat 5 also lacks a large side chain capable of stacking with RNA bases and lies opposite several of the atypical

RNA conformations (5-parallel, 5-stacked, and triple-stacked). The flatness combined with a protein surface lacking specificity allows “extra” RNA nucleotides, needed to span the distance between repeats with base specificity, to assume different conformations. These extra nucleotides may not contact the protein, but instead stack with one another or lie parallel to the RNA-binding surface.

Evolution of binding specificity across Ascomycota

To examine the evolution of the broad specificity of Puf5p, we probed the RNA binding preferences of Puf5p proteins from representative species across Phylum Ascomycota. This group includes the budding yeasts, filamentous fungi, and fission yeasts (Fig. 5a). We used the yeast three-hybrid assay to measure the affinities of Puf5p orthologues from six different species – *S. cerevisiae*, *Saccharomyces bayanus*, *Eremothecium gossypii*, *Candida albicans*, *Neurospora crassa*, and *Schizosaccharomyces pombe* (Fig. 5a). These proteins were identified as orthologues using SYNERGY, which relies on the species tree, sequence similarity and synteny²⁷. Their binding preferences versus length of site were evaluated using a set of RNAs 8-12 nt in length, conforming to the sequence UGUA(A)₂₋₅UA using the yeast three-hybrid system²⁸. All the RNAs thus maintained the 5'UGUA and 3'UA critical for *S. cerevisiae* Puf5p interaction and contained a single 3'UA element to define the target length unambiguously. In the three-hybrid assay, the level of expression of a reporter gene (LacZ) is a proxy for the affinity of the interaction²⁹.

Puf5p proteins across Ascomycota exhibited broad binding specificities. Puf5 proteins bound similarly to sites of 9, 10, 11 and 12 nts (Fig. 5b). The more restricted

specificities of Puf4p for 9 nt sites (Fig. 5c), and Puf3p for 8 nt sites (Fig. 5d), also were conserved across the entire Phylum, with the exception of *S. pombe* Puf3. This protein bound a broad range of site lengths, unlike its orthologues in other species that showed preference for 8-nt sites.

The broadened specificity of *S. pombe* Puf3 appears to have arisen exclusively in the fission yeast lineage, which enabled us to probe how that broadening arose during evolution. We reasoned that the broadening was not due to the identity of the RNA-interacting TRMs, as the residues are identical among all the Puf3p orthologues (with the exception of Repeat 3 in *N. crassa*, with a Gln to Arg substitution). To identify the key regions of the proteins that confer specificity, we prepared chimeras in which segments of the *S. cerevisiae* and *S. pombe* proteins were exchanged (Fig. 6a). The specificity profile – broad or narrow – was conferred by PUF Repeats 6-8. A chimeric protein possessing Repeats 6-8 from *S. pombe* exhibited broad specificity, while a chimera with Repeats 6-8 of the *S. cerevisiae* protein had narrow specificity (Fig. 6a,b). The protein sequences in Repeat 6 contain a divergent region among Puf3p orthologues (Supplementary Fig. 8). Indeed, substitution of *S. pombe* Repeat 6 alone into an *S. cerevisiae* scaffold was sufficient to confer the broad specificity profile (Fig. 6b).

DISCUSSION

Puf5p is a broad regulator of RNAs in *S. cerevisiae*, binding to more than 1,000 RNA targets, constituting ~16% of the transcriptome. 71% of these targets possess recognizable binding elements beginning with a 5' UGU sequence, which range in length from 8 to 12 nucleotides. The variations in length are accommodated by conformational adaptations of the RNA onto a fixed protein scaffold. The wide range of mRNA target site lengths is consistent with prior studies that linked Puf5p to a spectrum of functions, including cell wall integrity¹¹ and chromatin structure¹⁰.

The biological functions of target mRNAs are correlated with the length of binding elements they possess. How does this correlation arise? We propose that the correlation is imposed by other RNA-binding proteins that recognize the same binding elements, and whose specificity is much more restricted than Puf5p (Fig. 5b-d). For example, Puf3p binds 8-nt sites that are largely in mRNAs with mitochondria-related functions, while Puf4p binds 9-nt binding elements in mRNAs with roles in ribosomal biogenesis and assembly⁶.

Two PUF proteins that bind the same site could do so sequentially, competitively or cooperatively. Genetic studies demonstrate that Puf4p and Puf5p redundantly control the decay rate of common targets³⁰. In the absence of one of the proteins, the other is sufficient. However, for other common targets, the actions of two PUF proteins may be sequential. For example, *MRPL8* mRNA is a target of both Puf5p and Puf3p, possesses a single binding element, and is localized to the mitochondrial periphery in a Puf3p-dependent fashion³¹. Puf5p could exchange with Puf3p, facilitating repression (Puf5p) en route to localization to mitochondria (Puf3p).

While 71% of Puf5p targets possess discernible binding elements, 29% do not. RNAs without binding elements may associate with Puf5p indirectly, perhaps through a protein to which it and Puf5p are bound. Cross-linking to RNAs without sites could also be driven by their high concentrations in specific subcellular compartments (such as P-bodies), in which proteins and RNAs are present at high concentrations, and low complexity, Q/N-rich regions present in Puf3p, Puf4p, and Puf5p proteins that could facilitate aggregation³².

RNAs of different lengths adopt a broad range of conformations when bound to Puf5p. The flatter, extended scaffold of Puf5p, combined with its specificity for 5' and 3' sequences, imposes the requirement for these RNA conformational variations and permits recognition of 8-12-nt length RNAs. The elegance of this arrangement is that very similar sets of atomic contacts between amino acids and RNA bases are maintained in the different complexes, despite the range of RNA lengths they possess. For example, 18 of the 21 edge-on contacts made between Puf3p and its RNA target are also made in Puf5p bound to a 10-nt length site. In an analogous manner, β -catenin maintains a fixed scaffold to recognize peptides from different ligands (reviewed in ³³). Its central α -helical Armadillo (ARM) repeats interact with conserved sequence elements in an extended peptide while N- and C-terminal ARM repeats bind elements unique to that ligand. The changes in repeat-to-repeat arrangement at the junctions between the central ARM repeats and N- or C-terminal repeats seem to mark the regions with different protein-binding functions. In the same fashion, changes in curvature at specific repeat junctions in PUF proteins correlate with specialization in RNA-binding specificity.

The fact that a single repeat can broaden or narrow specificity (Fig. 6) suggests that this sort of change may be common in evolution (Fig. 6b). The sixth PUF repeat of Puf3p determines whether that protein binds 8 nt sites (*S. cerevisiae*) or accommodates 8, 9 or 10 nt sites (*S. pombe*). *S. cerevisiae* Repeat 6, which induces narrow specificity, contains additional residues relative to the same region of the *S. pombe* protein, which, although not near the RNA-binding residues, may alter the structure with corresponding effects on specificity (Supplementary Fig. 7a and 8).

From an evolutionary perspective, the broadening of Puf5p's specificity enabled new regulatory inputs into existing RNA circuits. In this view, the ability of Puf5p to recognize a wide array of target lengths arose after ancestral proteins (e.g., Puf3p) already regulated batteries of RNAs with related functions and conserved lengths of sites. Recruitment of Puf5p to these same targets, enabled by its flatter curvature, then provided new regulatory inputs and/or redundancy into that same circuit. For example, Puf5p binds regulatory kinases³⁴, whose input could be brought to bear on a pre-existing circuit. We suggest that curvature of the scaffold is critical in defining the RNAs that are controlled. Acquisition of new RNA specificities by alterations of the protein's architecture suggests ways in which new RNA circuits are established, expanded and contracted during evolution.

METHODS

HITS-CLIP

The strain we used harbored an integrated TAP tag sequence at the C-terminus of the endogenous *puf5* locus (GE YSC1178-202231131). CLIP methods were adapted from Wolf et al.³⁵. Modifications from the published protocols included disrupting the cells in the presence of liquid nitrogen, grinding with a mortar and pestle, and using 400 μ l calmodulin Sepharose beads (GE 17-0529-01) and IgG beads (Life Technologies 11202D). Following isolation of RNA, methods published previously were used to prepare libraries for high-throughput sequencing³⁶, with the exception of the Illumina TruSeq small RNA adaptor and PCR primer sequences. Detailed methods have been deposited in Protocol Exchange.

Western Blot

50 μ l IgG beads were removed from CLIP samples then incubated in 30 μ l LDS sample buffer (Life Technologies NP0007). The whole reaction was run on a Novex 6% TBE gel then transferred to PDVF membrane (Millipore IPVH00010). The membrane was probed with TAP Tag Polyclonal Antibody (1:10,000) (Pierce:CAB1001) primary antibody followed by goat anti mouse secondary antibody (1:10,000) (KPL:074-1506).

Informatic pipeline

Pre-processing. FASTQ files were uploaded to the Galaxy server³⁷ and groomed (FASTQ Groomer)³⁸. Adaptor sequences were then trimmed using Clip discarding sequences that contained the 5' adaptor or were too short after 3' adaptor clipping. The data were then filtered based on quality score using Filter FASTQ with a minimum length

of 15 bases and a minimum quality score of 20. The 5' adaptor included a 3' random bar code that was used to remove PCR duplicates by discarding any read with a perfect duplicate.

Mapping and defining peaks. The filtered reads were mapped to the *S. cerevisiae* genome using Bowtie2³⁹ (bowtie2 -x /Scgenome -q filename.fastq -S filename.sam -5 5 -N 1 -p 8). The .sam files were used to create .bam and indexed .bam files using samtools for visualization of the data in Artemis Genome browser. Peaks were defined using Pyicoteo¹⁶ (python pyicoclip filename.sam -f filename.pk --region Sc.bed --stranded). The .bed file required for Pyicoteo was downloaded from the Saccharomyces Genome Database (SGD)⁴⁰. Next, the the duplicate peaks were removed from the .pk file. Using the Pyicoteo defined summit, each peak was assigned to a genomic feature using the features table from the SGD. Sequences 200 bases upstream of the ORF and 300 bases downstream of the ORF were used as 5'UTRs and 3'UTRs, respectively and then added to the SGD features table. The number of gapped reads for each peak was defined. Kurtosis was calculated for each peak using the peak profile defined by Pyicoteo. 25 bases of genomic sequence flanking each peak summit was retrieved to define binding elements in two ways. MEME was used as an unbiased search and direct searches were used for known binding elements.

Filtering peaks. The biological replicates were combined into one list based on: 1. Each peak had a summit within 10 bases in both replicates, 2. Each peak contained a gapped read in both replicates, and 3. Each peak had a height greater than 10 reads (third quartile) in both replicates.

Gene Ontology. Functional enrichment was performed using Gene Ontology (GO)

analysis. Gene lists were uploaded to YeastMine where the p-value was calculated using the Hypergeometric distribution test (whole genome as background) and multiple test corrected using Holm-Bonferroni⁴¹.

Protein purification

The RNA-binding domain of yeast Puf5p (residues 201-600) was subcloned into the pSMX vector with an N-terminal His₆-SUMO tag⁴². *E. coli* cells BL21 Star (DE3) carrying the Puf5p plasmid were grown in Terrific Broth media to OD₆₀₀ = ~0.8 and then protein expression was induced with 0.4 mM IPTG for 20 hours at 18 °C. The cell pellet was resuspended in lysis buffer containing 20 mM Tris, pH 8.0; 0.5 M NaCl; 20 mM imidazole; 5% (v/v) glycerol and 0.1% (v/v) β-mercaptoethanol and sonicated on ice. The lysate was cleared by centrifugation and loaded onto a Ni-chelating gravity column (Thermo Scientific). His-SUMO-tagged Puf5p was eluted with a buffer containing 20 mM Tris, pH 8.0; 50 mM NaCl; 0.2 M imidazole and 1 mM DTT. Ulp1 protease was added to remove the His₆-SUMO tag, and the protein solution was loaded onto a Hi-Trap heparin column (GE Healthcare) and eluted with a gradient from 0-1 M NaCl in buffer containing 20 mM Tris, pH 8.0, 1 mM DTT. The fractions containing Puf5p were pooled and concentrated by Amicon filters and loaded onto a Superdex 200 16/60 column equilibrated in 20 mM HEPES, pH 7.4; 0.15 M NaCl and 2 mM DTT. The Puf5p peak fractions were pooled and concentrated in column buffer for crystallization and RNA-binding assays.

Protein-RNA crystallization

RNAs were purchased from Thermo Scientific. Puf5p (4 mg/ml) was mixed with each of the four different RNAs at a protein:RNA molar ratio of 1:1.2 and incubated on ice for 1

hour. Crystals were obtained at 20 °C by hanging drop vapor diffusion, mixing 1 µl Puf5p-RNA complex with 1 µl reservoir solution of 15-20% (w/v) PEG 3350 and 0.1 M citrate Bis-Tris propane (CBTP), pH 7.6. Microseeding was performed to grow larger single crystals. Crystals were cryo-protected in crystallization solution supplemented with 15% (v/v) glycerol and flash frozen in liquid nitrogen. For phasing, a Puf5p:SMX RNA complex crystal was soaked in 17.5% (w/v) PEG 3350, 0.5 M KI, 0.1 M CBTP, and 15% (v/v) glycerol for 5 minutes and then flash frozen.

X-ray data collection

X-ray data for structures of the 9-nt, 10-nt and 12-nt RNA complexes were collected at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory. Data for the 11-nt RNA complex and the iodide-soaked crystal were collected at the NIEHS in-house facility equipped with a Rigaku 007HF rotating anode generator and a Saturn 92 charge-coupled device (CCD) area detector system. All data were processed using HKL2000⁴³.

Structure determination

The crystal structure of a Puf5p:SMX2 RNA complex (space group $P2_12_12$) was determined by combining molecular replacement (MR) with iodide-single-wavelength anomalous diffraction (SAD)) phasing. The Phenix software suite was used throughout the process of structure determination⁴⁴. The anomalous signal of the SAD data extended only to 5.0 Å, and MR or SAD alone failed to solve the structure. A truncated Puf4p structure (PDB: 3BX2) containing repeats 4-8 (residues 684-887) was used as the MR search model. Following MR, AutoSol identified eight iodide sites with FOM of 0.33.

Running AutoBuild after MR-SAD phasing produced a model with $R_{\text{free}} = 44\%$. The model was further improved to $R_{\text{free}} = 38\%$ by using the EMBL-Hamburg Auto-Rickshaw web server⁴⁵. Electron density for the 9-nt RNA was clearly visible. Iterative cycles of manual model building in Coot⁴⁶ and refinement with Phenix led to the final model with $R_{\text{free}} = 28\%$ (Supplementary Table 3).

Crystals of the 10-nt, 11-nt and 12-nt RNA complexes and some crystals of the 9-nt *SMX2* RNA complex belonged to space group $P6_122$, although all crystals were grown in the same conditions as the $P2_12_12$ *SMX2* crystals. These structures were determined by MR using the Puf5p coordinates from the initial Puf5p:*SMX2* structure as the search model. Data and refinement statistics are shown in Supplementary Table 3. All models show good geometry according to MolProbity⁴⁷: 95-98% of the residues are in favored regions of the Ramachandran plot, and there are no outliers.

Electrophoretic mobility shift assays

RNA oligonucleotides were radiolabeled using ^{32}P - γ -ATP and T4 polynucleotide kinase (New England Biolabs) following manufacturer instructions. Serially-diluted Puf5p was mixed with 100 pM labeled RNA in buffer containing 10 mM HEPES, pH 7.4; 50 mM NaCl; 1 mM EDTA; 0.1 mg/ml BSA; 0.01% (v/v) Tween 20 and 0.1 mg/ml yeast tRNA. After overnight incubation at 4 °C, 4 μl loading dye (15% v/v Ficoll 400 and 0.01% bromophenol blue) was added to each 20- μl reaction prior to gel loading. 10% Novex TBE gels (Invitrogen) were run at 100 V at 4 °C for 30 minutes to resolve the samples. The gels were dried and exposed to storage phosphor screens. The screens were scanned using a Molecular Dynamics Typhoon phosphorimaging system (GE Healthcare). The band

intensities were analyzed with ImageQuant. K_d values were calculated with GraphPad Prism by fitting the data assuming one-site specific binding and a Hill coefficient of 1. ~93% of Puf5p was active, as determined using the method described in reference⁴⁸. The reported K_d values were not adjusted.

Yeast three-hybrid assays

Each orthologous PUF RNA-binding domain was cloned into activation domain–protein fusion plasmid, pGADT7²⁸. Oligonucleotides representing each RNA sequence were ordered from IDT and cloned into the Hybrid RNA plasmid, p3HR2²⁸. All experiments were conducted in the *Saccharomyces cerevisiae* strain YBZ-1 (MATa, ura3-52, leu2-3, -112, his3-200, trp1-1, ade2, LYS2::*(LexAop)*-HIS3, URA3::*(lexAop)*-lacZ, and LexA-MS2 MS2 coat (N55K)). Strains were Lithium Acetate transformed with appropriate combinations of plasmids and plated on synthetic dextrose media lacking uracil and leucine. Single colonies were selected and allowed to grow to stationary phase, then diluted and grown for about 4 hours. Optical density₆₆₀ (OD) for each culture was measured then 50 μ l of culture was added to 50 μ l of Beta-Glo (Promgea E4720) then incubated for 1 hr in the dark. Luminescence was measured by microplate reader (BioTech Synergy 4). Raw luminescence was normalized to OD₆₆₀, and each biological replicate (n=3) was then averaged and standard deviation was calculated.

REFERENCES

1. Keene, J.D. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-43 (2007).
2. Richter, J.D. CPEB: a life in translation. *Trends Biochem Sci* **32**, 279-85 (2007).
3. Darnell, R.B. RNA protein interaction in neurons. *Annu Rev Neurosci* **36**, 243-70 (2013).
4. Quenault, T., Lithgow, T. & Traven, A. PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol* **21**, 104-12 (2011).
5. Wickens, M., Bernstein, D.S., Kimble, J. & Parker, R. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-7 (2002).
6. 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).
7. Kershner, A.M. & Kimble, J. Genome-wide analysis of mRNA targets for *Caenorhabditis elegans* FBF, a conserved stem cell regulator. *Proc Natl Acad Sci U S A* **107**, 3936-41 (2010).
8. Campbell, Z.T., Valley, C.T. & Wickens, M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* **21**, 732-8 (2014).
9. Hall, T.M. Expanding the RNA-recognition code of PUF proteins. *Nat Struct Mol Biol* **21**, 653-5 (2014).

10. Kennedy, B.K. et al. Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in *S. cerevisiae*. *Cell* **89**, 381-91 (1997).
11. Kaeberlein, M. & Guarente, L. *Saccharomyces cerevisiae* MPT5 and SSD1 function in parallel pathways to promote cell wall integrity. *Genetics* **160**, 83-95 (2002).
12. Tadauchi, T., Matsumoto, K., Herskowitz, I. & Irie, K. Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. *EMBO J* **20**, 552-61 (2001).
13. Ule, J. et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-5 (2003).
14. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
15. Freeberg, M.A. et al. Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. *Genome Biol* **14**, R13 (2013).
16. Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M. & Eyras, E. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* **27**, 3333-40 (2011).
17. Zhang, C. & Darnell, R.B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **29**, 607-14 (2011).
18. Seay, D., Hook, B., Evans, K. & Wickens, M. A three-hybrid screen identifies mRNAs controlled by a regulatory protein. *RNA* **12**, 1594-600 (2006).

19. Prinz, S., Aldridge, C., Ramsey, S.A., Taylor, R.J. & Galitski, T. Control of signaling in a MAP-kinase pathway by an RNA-binding protein. *PLoS One* **2**, e249 (2007).
20. Valley, C.T. et al. Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc Natl Acad Sci U S A* **109**, 6054-9 (2012).
21. Chritton, J.J. & Wickens, M. Translational repression by PUF proteins in vitro. *RNA* **16**, 1217-25 (2010).
22. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
23. Licatalosi, D.D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-9 (2008).
24. Tenenbaum, S.A., Carson, C.C., Lager, P.J. & Keene, J.D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A* **97**, 14085-90 (2000).
25. Campbell, Z.T. et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep* **1**, 570-81 (2012).
26. 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 U S A* **106**, 20192-7 (2009).
27. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54-61 (2007).

28. Koh, Y.Y. & Wickens, M. Determining the RNA specificity and targets of RNA-binding proteins using a three-hybrid system. *Methods enzymol* **539**, 163-81 (2014).
29. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. RNA-protein interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *RNA* **11**, 227-33 (2005).
30. Ulbricht, R.J. & Olivas, W.M. Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *RNA* **14**, 246-62 (2008).
31. 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).
32. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-67 (2012).
33. Reichen, C., Hansen, S. & Pluckthun, A. Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *J Struct Biol* **185**, 147-62 (2014).
34. Chen, T. & Kurjan, J. *Saccharomyces cerevisiae* Mpt5p interacts with Sst2p and plays roles in pheromone sensitivity and recovery from pheromone arrest. *Mol Cell Biol* **17**, 3429-39 (1997).
35. Wolf, J.J. et al. Feed-forward regulation of a cell fate determinant by an RNA-binding protein generates asymmetry in yeast. *Genetics* **185**, 513-22 (2010).
36. Green, M.R. & Sambrook, J. *Molecular cloning: a laboratory manual*, (Cold Spring Harbor Laboratory Press New York, 2012).

37. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86 (2010).
38. Blankenberg, D. et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-5 (2010).
39. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
40. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-5 (2012).
41. Balakrishnan, R. et al. YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database (Oxford)* **2012**, bar062 (2012).
42. Mossessova, E. & Lima, C.D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **5**, 865-76 (2000).
43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data. *Methods enzymol* **276**, 307-326 (1997).
44. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
45. Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. & Tucker, P.A. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool

- for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr* **61**, 449-57 (2005).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
 47. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83 (2007).
 48. Cheong, C.G. & Hall, T.M. Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A* **103**, 13635-9 (2006).
 49. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
 50. SenGupta, D.J. et al. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci U S A* **93**, 8496-501 (1996).
 51. Goldstrohm, A.C., Seay, D.J., Hook, B.A. & Wickens, M. PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J Biol Chem* **282**, 109-14 (2007).
 52. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004).

ACKNOWLEDGEMENTS

We thank Natascha Buter for help with experiments, and colleagues in the Wickens lab and NIEHS for comments during the work and discussions of the manuscript. We are very grateful to Drs. Robert Darnell and Aldo Mele for helpful suggestions in troubleshooting HITS-CLIP and Dr. Fabio Parmeggiani for help analyzing the β -catenin repeats. We thank Dr. Lars Pedersen and the staff of the Southeast Regional Collaborative Access Team beamlines for assistance with X-ray data collection. Work in the Wickens lab is supported by NIH (GM50942). ZKW was supported by an NIH Postdoctoral Fellowship. The work also was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences (TMTH). The Advanced Photon Source used for this study was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38.

CONTRIBUTIONS

D.W., C.Q. and M.N. performed the experiments. D.W, C.Q., C.P.L., Z.T.C, T.M.T.H., and M.W. analyzed the data. D.W., C.Q., T.M.T.H., and M.W. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Figure 1

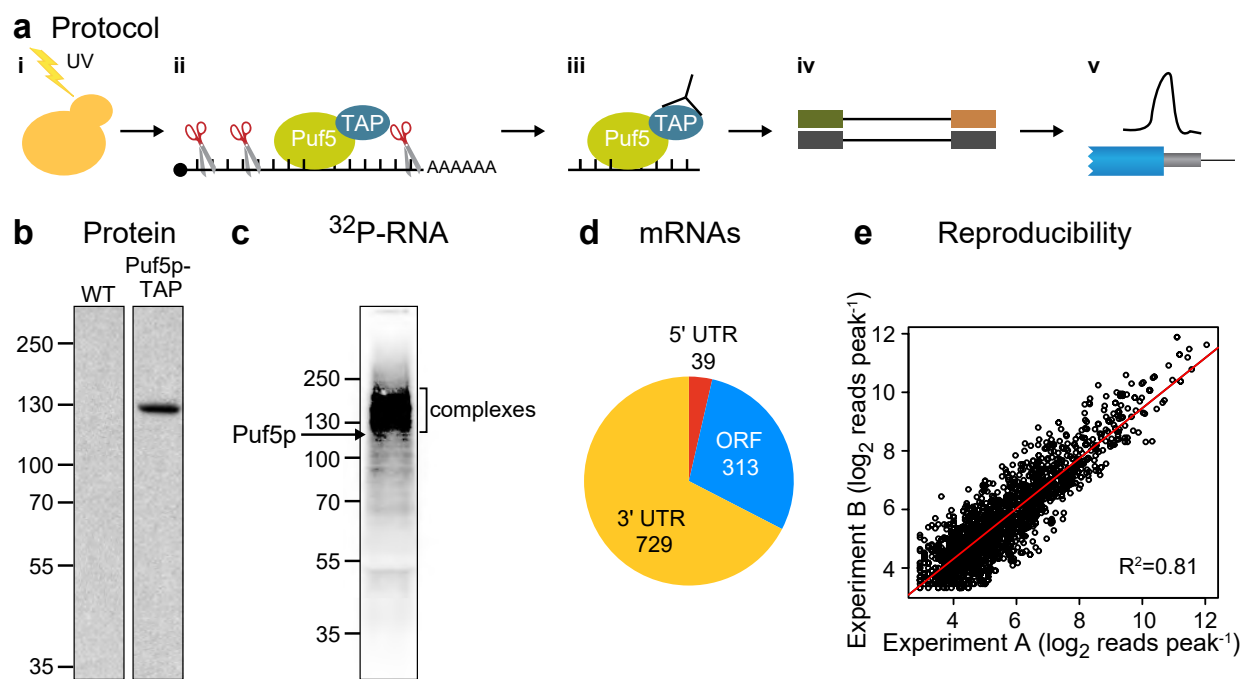


Figure 1. Puf5p HITS-CLIP. a. Summary of HITS-CLIP protocol. (i). *S. cerevisiae* cells were isolated and UV irradiated (254 nm wavelength). (ii) Cell lysate was subjected to gentle RNase A digestion and then (iii) TAP-tagged Puf5p was affinity purified sequentially with calmodulin and IgG resins. (iv) RNA adaptors were ligated to RNA fragments. (v) Libraries were PCR amplified and high-throughput sequenced. b. Western blot of WT and epitope-tagged Puf5p. c. Autoradiogram of ³²P-labelled RNA crosslinked to Puf5p. Complexes migrate higher than protein alone. d. Pie chart of Puf5p CLIP peaks found in mRNA regions. e. Reproducibility of peaks from each biological replicate. Normalized log₂ reads/peak for the two experiments are plotted.

Figure 2

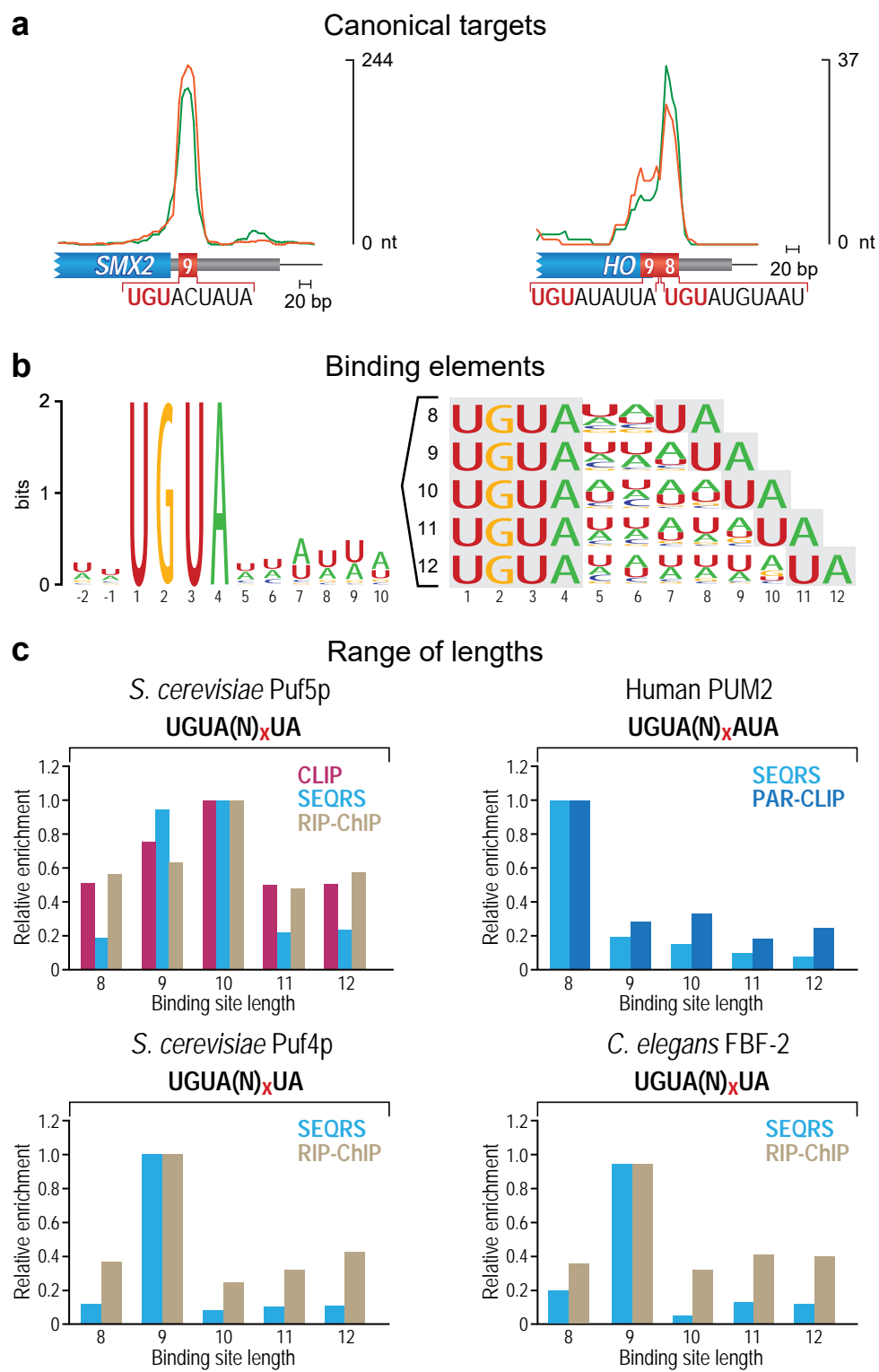
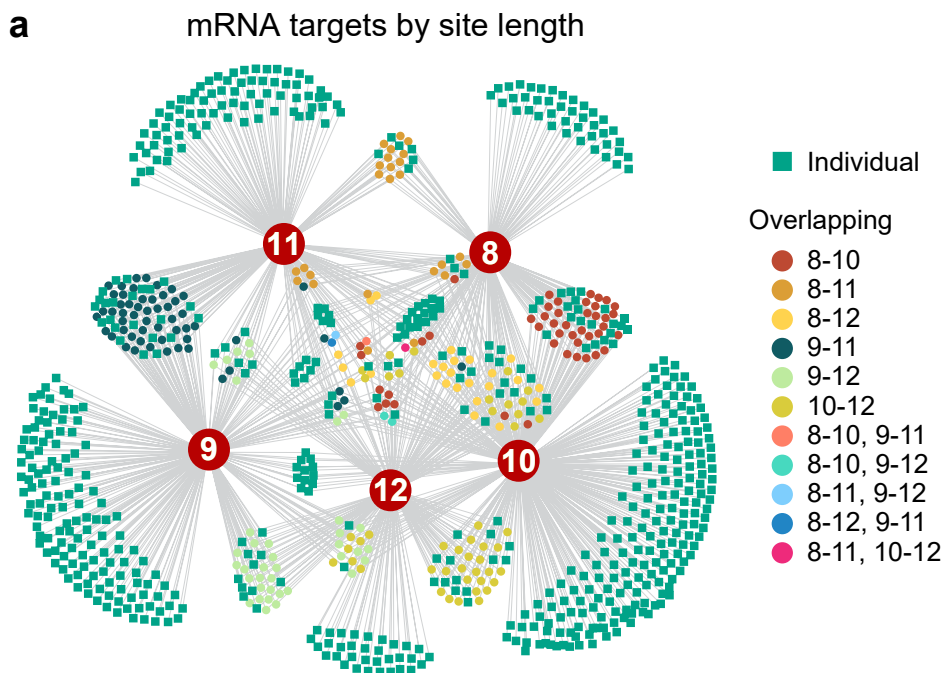


Figure 2. Puf5p binding element exhibits flexibility. a. Puf5p interaction peaks map to previously characterized Puf5p binding sites. Orange and green lines represent reads mapped for each replicate. *SMX2* has a peak over the 9 nt binding element. *HO* has a broad peak over two binding elements: a 9 nt lower affinity site and a 8/10 nt higher affinity site. b. Binding elements identified in high-confidence Puf5p target mRNAs. The MEME-derived logo is shown on the left, which was deconvoluted into 5 binding elements of 8-12 nts in length. c. Distribution of binding element lengths for 4 PUF proteins representing 3 species. Results from CLIP (red), SEQRS²⁵ (light blue), RIP-Chip^{6,7} (green), and PAR-CLIP¹⁴ (dark blue) experiments are compared, where available, and shown as enrichment relative to the predominant length for each protein, which is set to 1. The consensus RNA sequence element for each protein is shown, where N is A, C, G or U.

Figure 3



b Length correlated with biological function

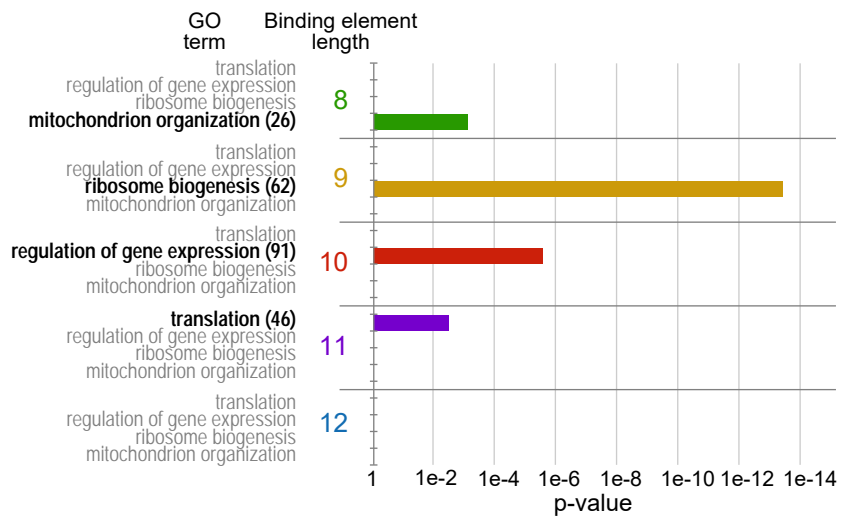


Figure 3. Binding element lengths correlate with biological functions. a. Network representation of Puf5p mRNA targets visualized in Cytoscape 3.0.1⁴⁹. Binding elements were defined as UGUN_(x)UA within 25 bp of the peak summit. Large red hubs indicate the length of the binding element. Each node (small circle or square) represents one HITS-CLIP peak in an mRNA. Green square nodes represent mRNAs containing an individual binding element of either 8, 9, 10, 11 or 12 nts; nodes with only one binding element (edge) are placed at the outer periphery of the diagram. Green square nodes with two lines indicate that the HITS-CLIP peak contained two non-overlapping binding elements. Most binding elements were unambiguously of a single length. In a minority of elements, two different lengths of binding elements co-reside in a single sequence. Circles represent these mRNAs with “overlapping” binding elements: for example, “8-10” means a single site of the sequence UGUNNUAUA, which possesses both 8 and 10 nt elements depending on the 3'UA used, and either sequence may be used *in vivo*, and “8-10, 9-11” means that two distinct overlapping sites are present under the peak. The key to the right is a color code for each combination of overlapping binding element lengths (nt). The numbers of mRNAs containing overlapping binding elements are provided in Supplementary Table 1. b. Gene Ontology term enrichment for mRNAs belonging to each length of binding element using SGD YeastMine⁴¹. GO terms that are significantly overrepresented in the gene list are bolded for each binding element length. Numbers of genes in the most enriched GO terms are in parentheses.

Figure 4

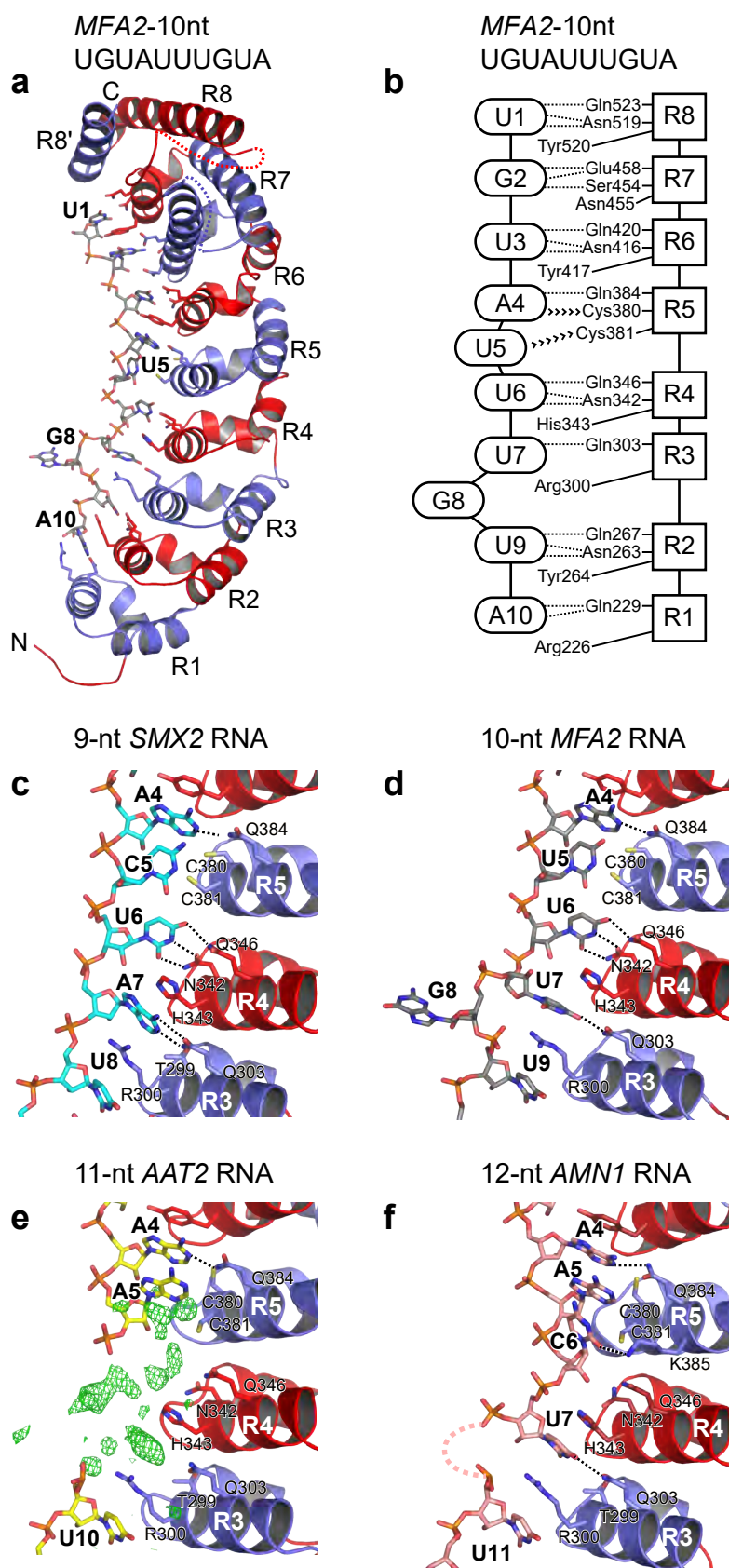


Figure 4. Crystal structures of Puf5p in complex with representative 9-12 nt target mRNAs. a. Crystal structure of Puf5p in complex with 10-nt *MFA2* RNA. Puf5p is shown as a ribbon diagram with PUM repeats colored alternately blue and red. Two disordered loops are indicated with dotted lines. RNA interacting residues and *MFA2* RNA are shown as stick models with atoms colored by element (carbon, grey; nitrogen, blue; oxygen, red; sulfur, yellow; phosphorus, orange). b. Schematic diagram of interactions between Puf5p repeats (rectangles) and *MFA2* RNA bases (ovals). Hydrogen bonds are indicated by dotted lines and van der Waals contacts are indicated by >>>>. c-f. Interactions between repeats 3-5 of Puf5p and 9-nt *SMX2* RNA (c), 10-nt *MFA2* RNA (d), 11-nt *AAT2* RNA (e) or 12-nt *AMN1* RNA (f). Hydrogen bonds are indicated with dotted lines. Discontinuous F_o-F_c electron density (contoured at 3σ) following base A5 in the 11-nt *AAT2* RNA is shown in panel e.

Figure 5

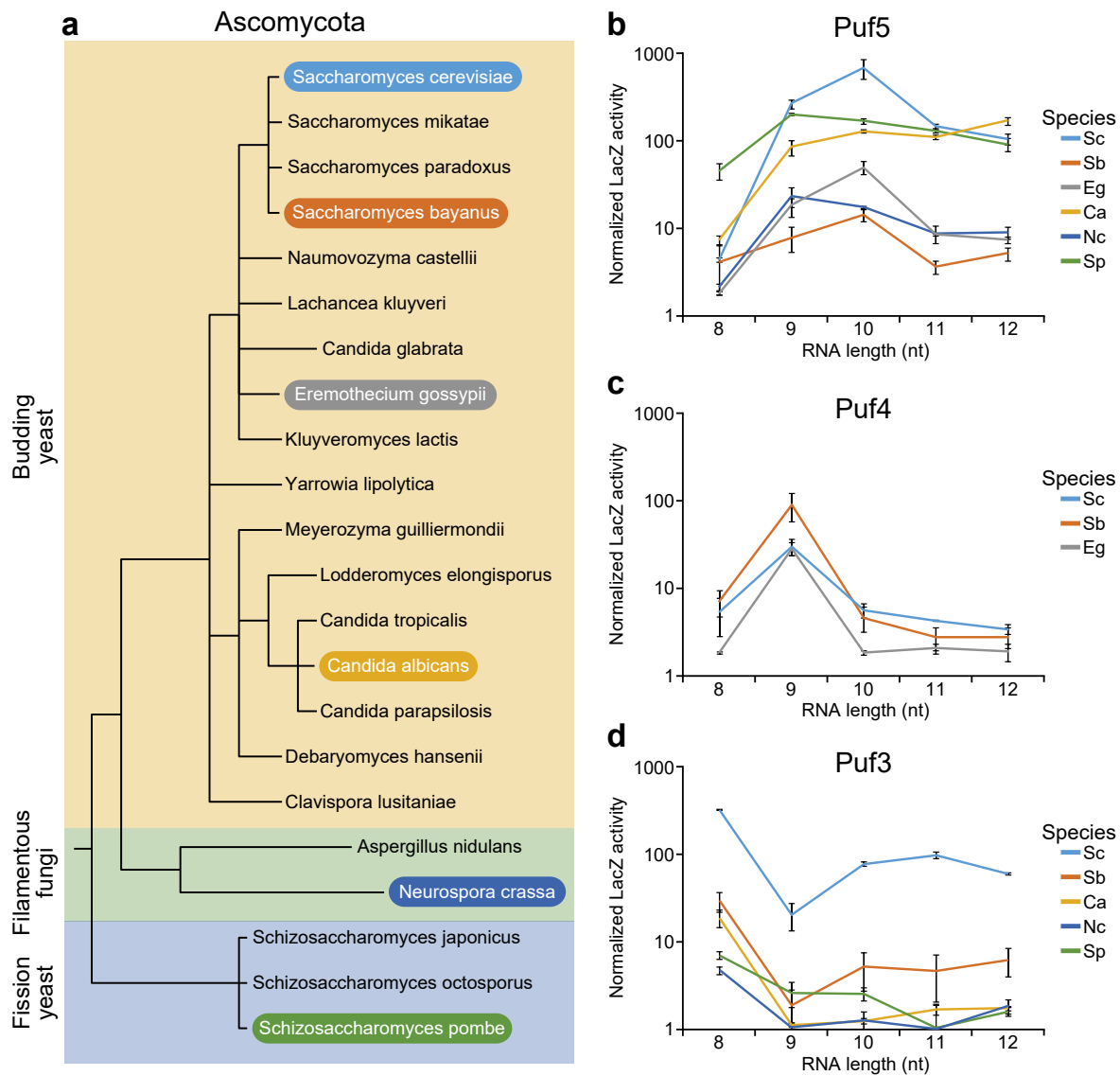


Figure 5. Binding element preference across Phylum Ascomycota. a. Phylogeny of Ascomycota fungi. Each subphylum is indicated: budding yeast-yellow, filamentous fungi – green, and the fission yeast – blue. Species used for RNA-binding studies below are highlighted. b-d RNA binding element length preferences for Puf5p, Puf4p, and Puf3p orthologues. RNA binding was assayed using the yeast three-hybrid system⁵⁰. RNAs tested were 8 (UGUAAUA), 9 (UGUAAAUA), 10 (UGUAAAAUA), 11 (UGUAAAAAUA), or 12 (UGUAAAAAAUA) nts in length. Raw luminescence values per cell for each biological replicate (n=3) were averaged then normalized to controls where the 5'UGU sequence was mutated to ACA. Error bars represent standard deviation. The full set of six species could not be tested for all five binding element lengths. *N. crassa* does not have a Puf4 ortholog; *C. albicans* and *S. pombe* Puf4p failed to specifically bind any RNA; *E. gossypii* Puf4p clones could not be obtained.

Figure 6

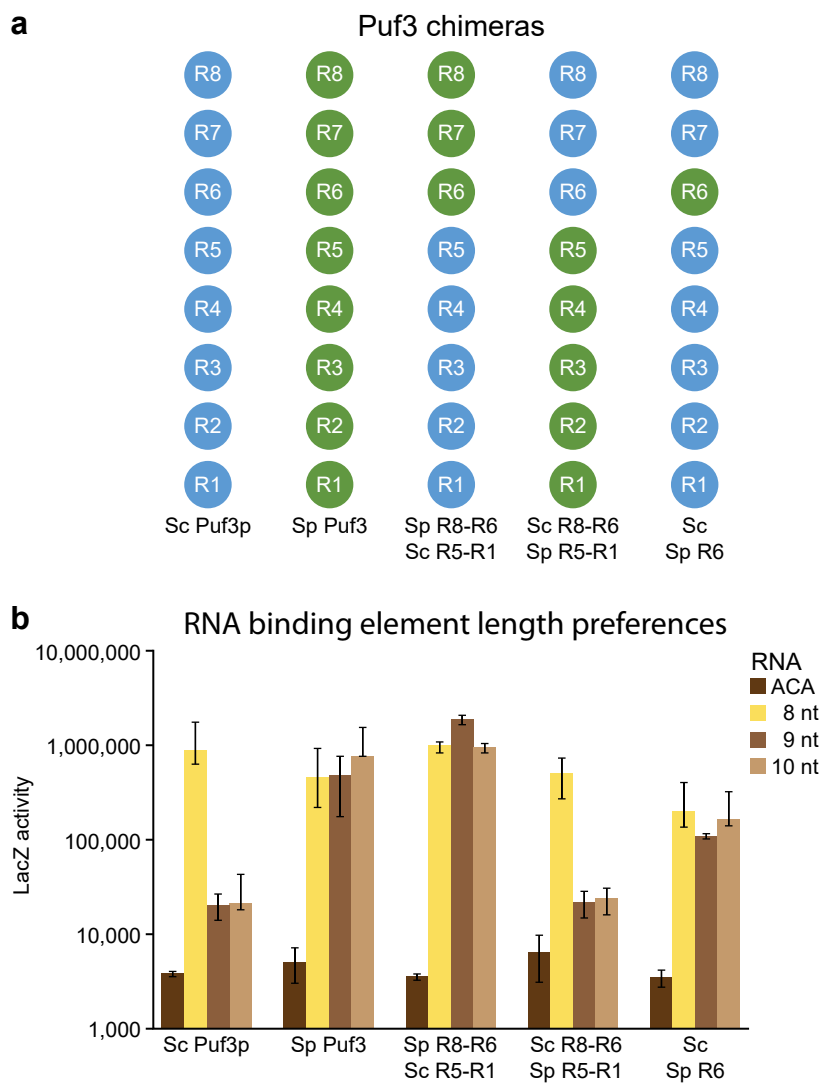
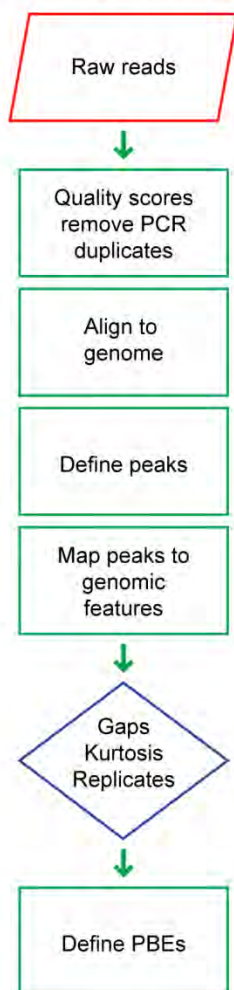


Figure 6. Evolution of Puf3p broadened RNA specificity. a. Schematic representations of chimeras tested in b. PUM repeats are represented by circles: *S. cerevisiae*, blue; *S. pombe*, green. b. Broadened *S. pombe* Puf3 specificity is linked to Repeat 6. RNA binding element length preferences for the chimeric proteins were assayed using the yeast 3-hybrid system with 8, 9 or 10 nt RNAs as in Figure 5. Raw luminescence values per cell for each biological replicate (n=3) were averaged then normalized to an acaAAAUA mutant negative control, which depresses binding more than 100-fold⁵¹. Error bars represent standard deviation.

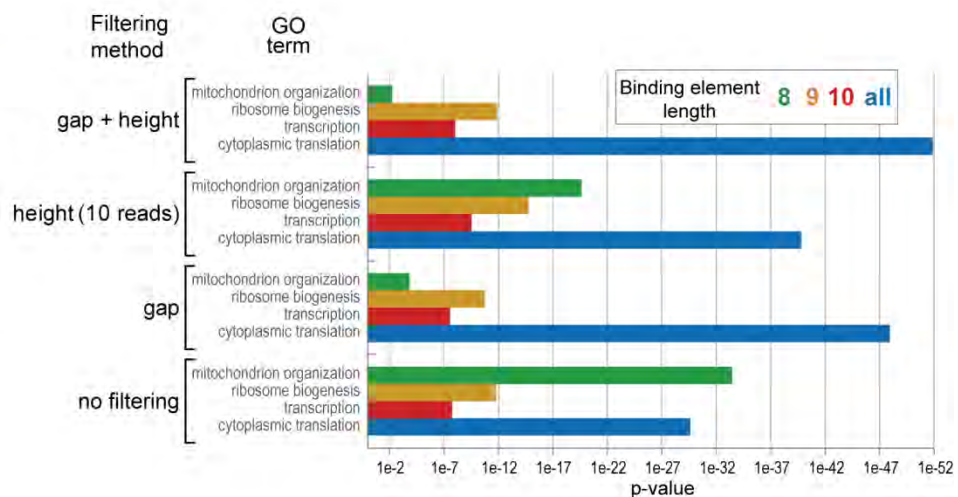
Supplementary Figure 1.

a Data analysis



b

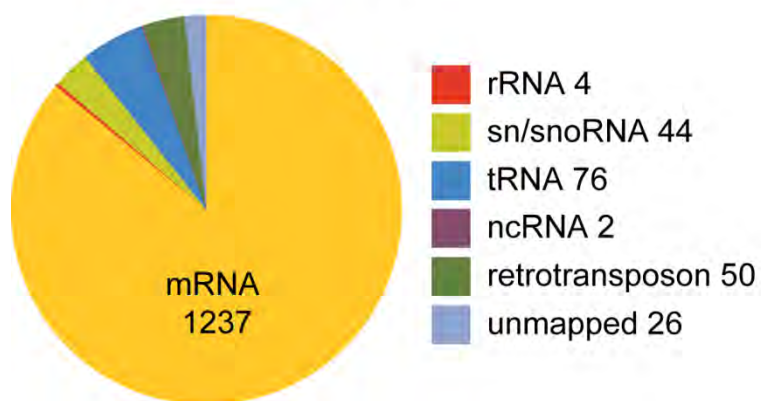
Comparison of filtering methods



Supplementary Figure 1. HITS-CLIP data analysis pipeline. a. Flow chart of the data analysis. b. GO analysis of target RNA list using various criteria for filtering peaks using SGD YeastMine⁴¹.

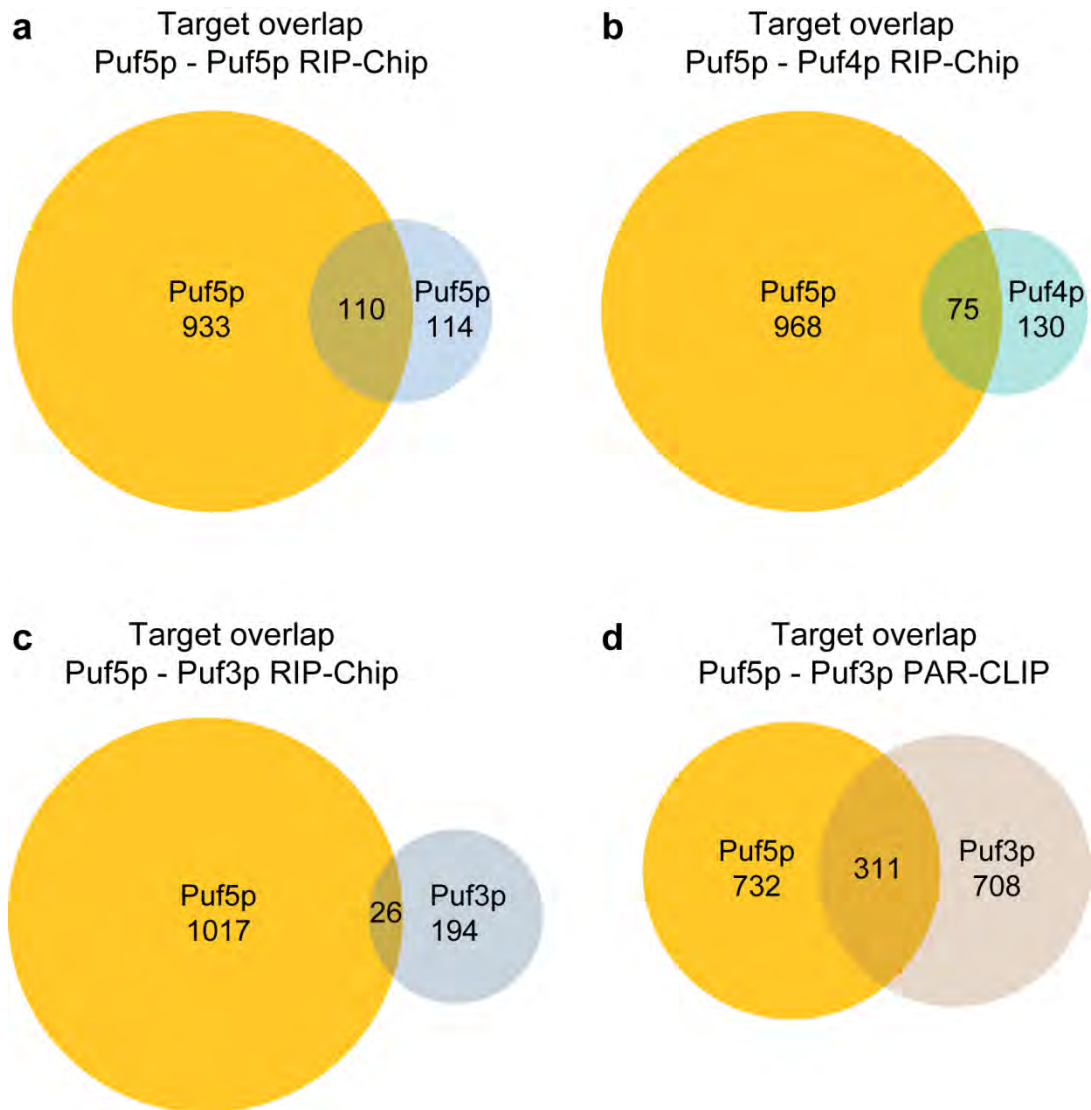
Supplementary Figure 2.

Distribution of CLIP peaks



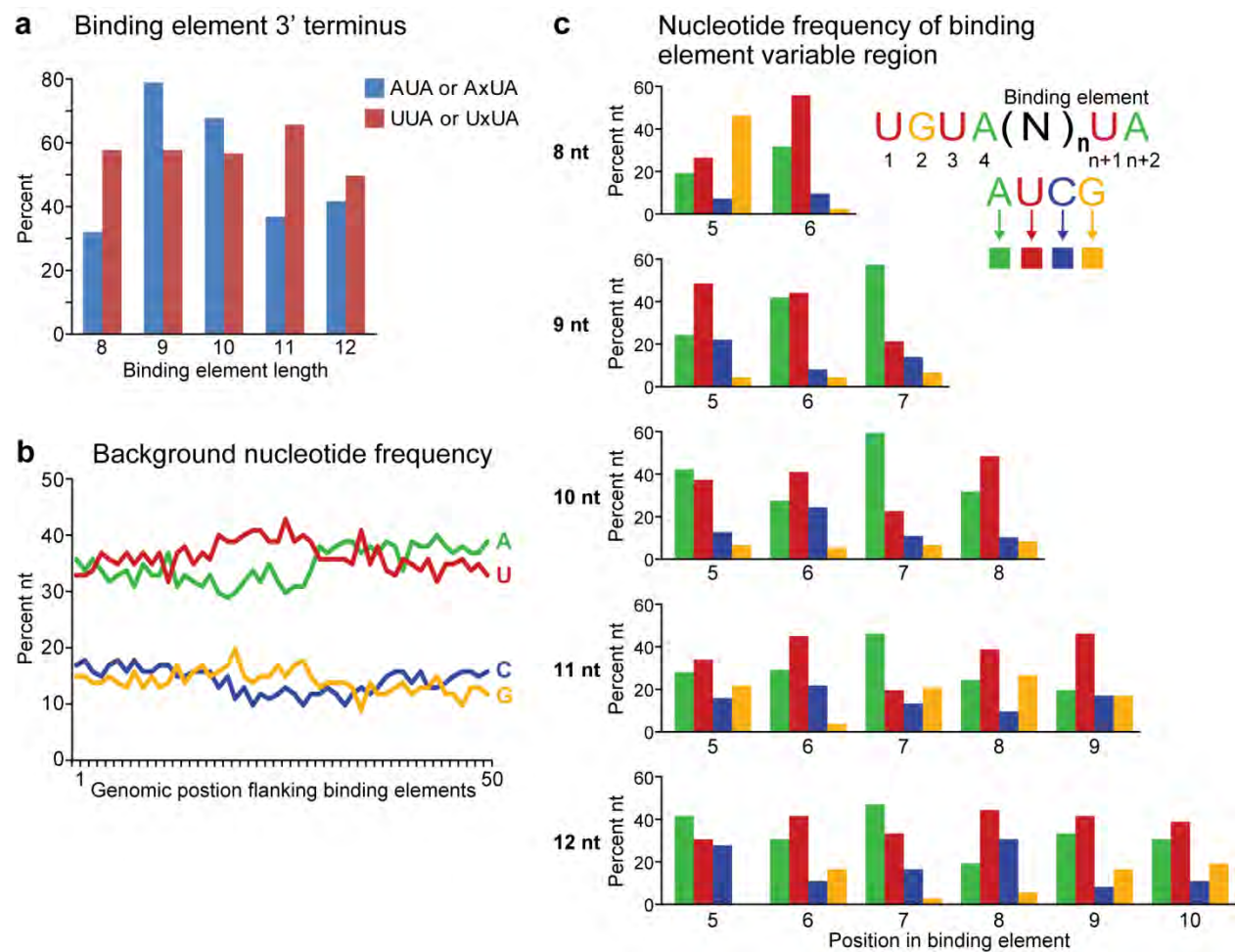
Supplementary Figure 2. Puf5p binds predominantly mRNAs. Distribution of Puf5p HITS-CLIP peaks in RNA types. Numbers of peaks in each category are indicated.

Supplementary Figure 3.



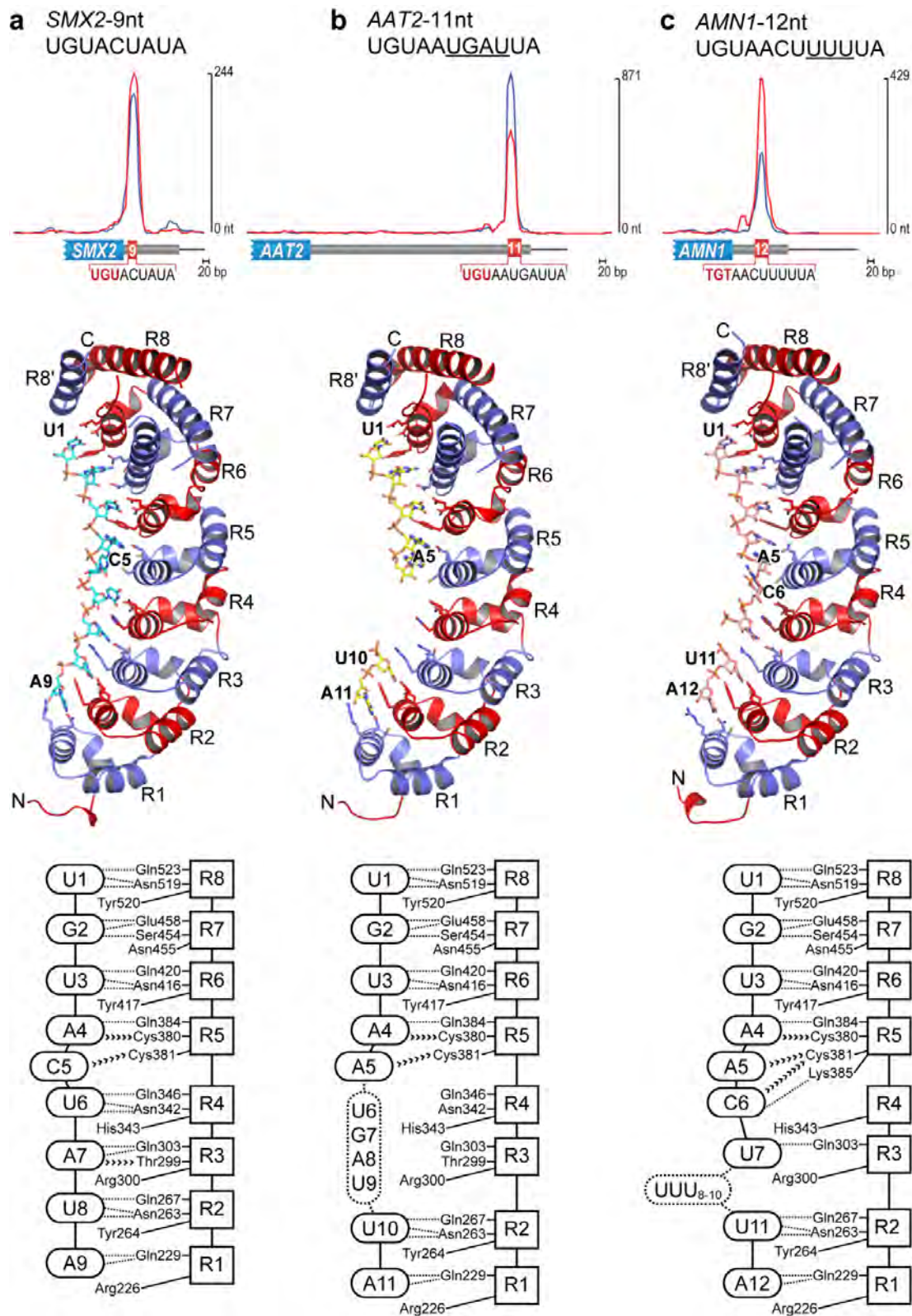
Supplementary Figure 3. Overlap of Puf5p HITS-CLIP targets with Puf3p and Puf4p targets. a. mRNA overlap between Puf5p HITS-CLIP and RIP-Chip⁶. b. mRNA overlap between Puf5p HITS-CLIP and Puf4p RIP-Chip⁶. c. mRNA overlap between Puf5p HITS-CLIP and Puf3p RIP-Chip⁶. d. mRNA overlap between Puf5p HITS-CLIP and Puf3p PAR-CLIP¹⁵. Numbers of mRNAs in each subset are indicated.

Supplementary Figure 4.



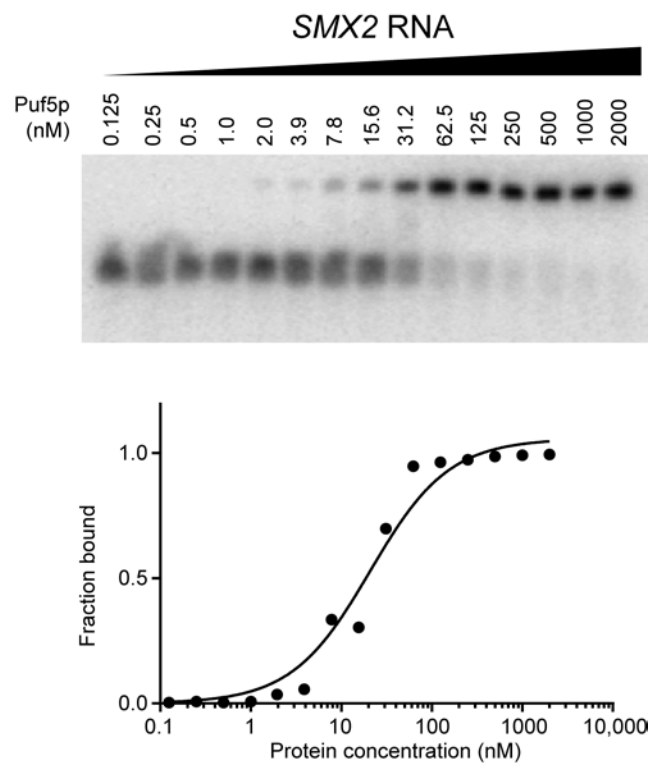
Supplementary Figure 4. Nucleotide composition of Puf5p binding elements. a. Enrichment for A's (blue), or U's (red) at the 3' terminus of each length binding element. b. The frequency of each nucleotide flanking CLIP-defined peaks: A's (green), U's (red), C's (blue), and G's (orange). c. Positional analysis of intervening nucleotides (excluding the fixed 5'-UGUA tetranucleotide sequence and the 3'UA sequences) for each length binding element.

Supplementary Figure 5.



Supplementary Figure 5. Crystal structures of Puf5p in complex with 9-nt *SMX2*, 11-nt *AAT2*, and 12-nt *AMN1* RNAs. HITS-CLIP peaks (a), ribbon drawings (b), and schematic diagrams of protein:RNA interactions (c) are shown.

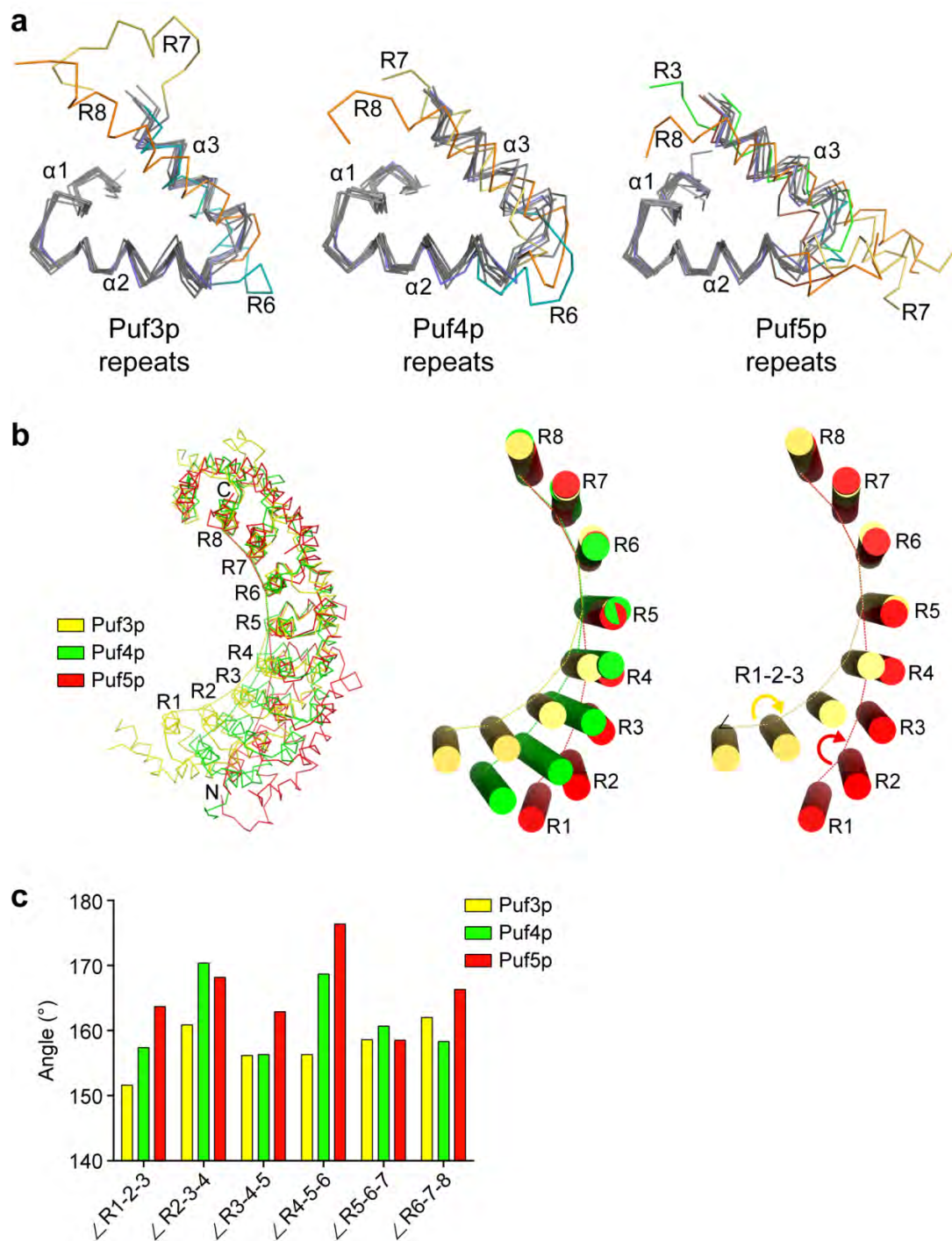
Supplementary Figure 6.



Supplementary Figure 6. Electrophoretic mobility shift assay (EMSA) of Puf5p.

Representative EMSA of Pufp5 with *SMX2* binding element (UGUACUAUA) RNA (top) and data analysis (bottom) are shown. All binding assays were performed in triplicate and the mean K_d and standard error of the mean are reported in Supplementary Table 4.

Supplementary Figure 7.



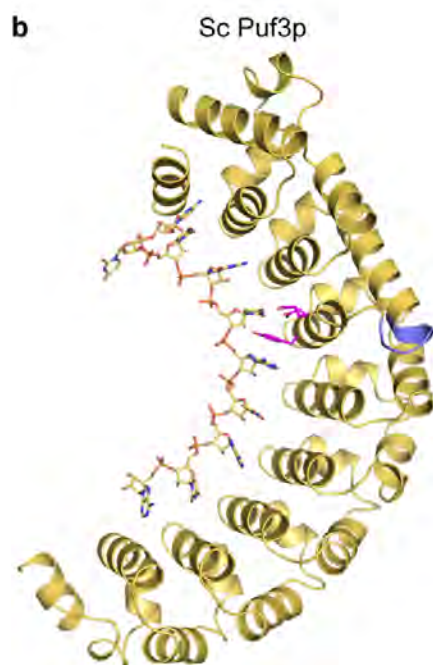
Supplementary Figure 7. Structural differences in yeast PUF proteins Puf3p, Puf4p, and Puf5p. a. Superposition of PUM repeats in Puf3p, Puf4p, and Puf5p. C α traces of the eight repeats from each protein are superimposed with human PUM1 repeat 1 (blue). Divergent structures following helix α 2 are colored teal (R6), yellow (R7), and orange (R8) for all proteins and green (R3) for Puf5p. b. Superposition of RNA-binding helices in Puf3p (yellow), Puf4p (green) and Puf5p (red). Repeats 5-8 of each structure were aligned in the superposition. C α traces are shown in the first, and the α 2 helices only are shown as cylinders in the second and third set of traces. Dotted lines connecting the C α atoms of base-stacking residues are shown, and \angle R1-2-3 is indicated. c. Repeat-to-repeat angles in Puf3p, Puf4p and Puf5p. The angles formed by lines between the C α atoms of stacking residues in sets of three successive repeats for Puf3p, Puf4p and Puf5p are plotted.

Supplementary Figure 8.

a

Puf3 Repeat 6

	Helix	Helix	Helix
Scer YLL013C	IPYLIQDQYGN Y VI Q YVLQQDQ---FTNKEMVDIKQE--IIETVANN		
Smik smik1210-g1.1	IPYLIQDQYGN Y VI Q YILQQNQ---FTNKEMVDVKQE--IIETVANN		
Spar spar52-g2.1	IPYLIQDQYGN Y VI Q YILQQDQ---FTNKEMVDIKQE--IIETVANN		
Sbay sbayc566-g3.1	IPYLIQDQYGN Y VI Q YILQQDQ---FTNKEMVDVKQE--IVETVADN		
Ncas Scas688.28	IPYLIQDQYGN Y VI Q HILEQQDNNPNVSQEMMNTKQE--IVNIVSQN		
Sklu SAKL0G13002g	IPYLIQDQYGN Y VI Q HILQHGG--EHTNIHIGSTKQN--IVDIVSKS		
Cgla CAGL0D05544g	IPYLIQDQYGN Y VI Q HILQHGSVDNLA S EHMRVIKQE--IINN V ADN		
Egos AAL152W	IPYLVQDQYGN Y VI Q HILQHGG-DN P AENHID--KSKQDIVDTISKT		
Klac KLLA0F15477g	IPFLIQDQYGN Y VI Q HILQHGT-ED T SSHI G MS-KQN--IIDIRKN		
Ylip YALI0E12001g	AYHLIQDQYGN Y VI Q HVLEQGA---PDDKEA--MMLVIKQH		
Mgui PGUG05371.1	IFYLIQDQYGN Y VM Q HILERGS---SKDREA--ILEVVLGS		
Ctro CTRG03880.3	LYYLILDQYGN Y VI Q HILENGT---PEEKEP--ILEIVLGS		
Calb orf19.1795	LYYLILDQYGN Y VI Q HILENGT---QEEKEP--ILEIVLGS		
Cpar CPAG05281	LGFLITHKFG N YVI Q ACL E NQ---LREQD--IFTTVVCK		
Dhan DEHA2C07128g	IFYLIQDQYGN Y VM Q H T LERGN---PEDREE--ILKIVLGS		
Clus CLUG04502	LFYLIQDQYGN Y VI Q HIL E RGT---PSEKEE--IFEVAFSS		
Anid AN6587	APRLIEDQYGN Y VI Q H I IQSGE---EEDRSF--MIEMVKQK		
Ncra NCU06511	AHTLITDAYGN Y VA Q H I IEAGK---PEDRAR--MIAAVMSQ		
Sjap SJAG00686	ILHLAQDQYGN Y VI Q HLMK K GS---PSEQRE--IVEVVLGN		
Soct SOCG02114	SLQLTQGQYGN Y V Q HILKEGS---EKDKKF--VFNLIAKN		
Spom SPAC1687.22c	ILKLTQDQYGN Y V Q HILRTGS---ESDKKY--IFDLMIDH		



Supplementary Figure 8. Sequence analysis of Puf3p Repeat 6. a. Sequence alignment of Puf3p Repeat 6 from Ascomycota based on a MUSCLE multiple sequence alignment⁵². Positions of α helices in the crystal structure of *S. cerevisiae* Puf3p are indicated above the sequence alignment. Magenta highlighted residues represent the RNA-binding motif in *S. cerevisiae* Repeat 6 that is conserved in all species shown. The blue box indicates the additional residues at the interface of α helices 2 and 3 in Repeat 6. b. Ribbon drawing of the crystal structure of *S. cerevisiae* Puf3p in complex with a COX17 binding element. RNA-binding motif residues of Repeat 6 are colored magenta and the additional loop residues are colored blue.

Supplementary Table 1 – Number of peaks containing binding elements

Single binding elements		Non-overlapping binding elements		Overlapping binding elements	
Binding element length (nt)	peaks	Binding element length (nt)	peaks	Binding element length (nt)	peaks
8	39	9,11	30	9-11	59
9	125	8,10	28	10-12	50
10	201	9,10	27	8-10	49
11	77	10,11	22	9-12	40
12	36	9,12	16	8-11	28
Total	478	10,12	15	8-12	28
		11,12	8	8-10-12	15
		8,11	7	8-10, 9-12	2
		9,11,12	7	8-10, 9-11	1
		8,9,10	6	8-11, 9-12	1
		9,10,11	6	8-12, 9-11	1
		8,10,11	5	8-11, 10-12	1
		8,10,12	4	Total	275
		8,12	4		
		8,9	4		
		8,9,10,12	4		
		9,10,12	4		
		10,11,12	2		
		8,9,11	2		
		Total	201		
				Total	954

Supplementary Table 1. Number of peaks containing binding elements. Left columns show the numbers of peaks possess only one binding element. The middle and right columns contain the number of peaks that possess non-overlapping or overlapping binding elements, respectively.

Supplementary Table 2 – Significantly enriched Gene Ontology terms for each binding element length and all targets.

Binding element length	Enriched GO term	p-value (Hypergeometric distribution test Holm-Bonferroni corrected)
8	154 genes used only UGUA set (all 8's)	
	mitochondrion organization	9.50E-04
	protein targeting to mitochondrion	3.90E-03
	mitochondrial transmembrane transport	5.82E-03
	protein localization to mitochondrion	1.14E-02
	establishment of protein localization to mitochondrion	1.14E-02
	protein import into mitochondrial matrix	2.03E-02
9	250 genes total only UGUA set (all 9's)	
	ribonucleoprotein complex biogenesis	2.83E-15
	ribosome biogenesis	3.59E-14
	ncRNA metabolic process	1.76E-06
	rRNA processing	1.79E-06
	rRNA metabolic process	2.18E-06
	cellular component biogenesis	1.21E-05
	ncRNA processing	1.69E-05
	RNA metabolic process	1.50E-04
	RNA processing	4.15E-04
	ribosomal large subunit biogenesis	5.72E-04
	nucleocytoplasmic transport	1.19E-03
	nuclear transport	1.33E-03
	maturation of 5.8S rRNA	2.45E-03
	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2.45E-03
	maturation of LSU-rRNA	4.42E-03
	nucleic acid metabolic process	5.53E-03
	nuclear export	6.22E-03
	nucleobase-containing compound metabolic process	6.43E-03
	gene expression	6.82E-03
	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	7.75E-03
ribosomal small subunit biogenesis	1.80E-02	
cellular component organization or biogenesis	2.94E-02	
cellular aromatic compound metabolic process	3.66E-02	
10	353 genes total only UGUA set (all 10's)	

regulation of macromolecule biosynthetic process	3.03E-07
regulation of biosynthetic process	3.48E-07
regulation of cellular macromolecule biosynthetic process	4.97E-07
regulation of cellular biosynthetic process	7.85E-07
regulation of macromolecule metabolic process	1.80E-06
regulation of primary metabolic process	2.54E-06
regulation of gene expression	2.85E-06
transcription, DNA-templated	3.64E-06
nucleic acid-templated transcription	3.89E-06
RNA biosynthetic process	5.41E-06
regulation of cellular metabolic process	7.37E-06
regulation of metabolic process	1.46E-05
regulation of cellular process	1.66E-04
regulation of biological process	1.69E-04
biological regulation	2.83E-04
regulation of nucleobase-containing compound metabolic process	5.97E-04
regulation of nitrogen compound metabolic process	6.05E-04
regulation of transcription, DNA-templated	8.28E-04
regulation of nucleic acid-templated transcription	8.78E-04
regulation of RNA biosynthetic process	8.78E-04
RNA metabolic process	9.58E-04
regulation of RNA metabolic process	1.07E-03
nucleobase-containing compound biosynthetic process	1.54E-03
organic cyclic compound biosynthetic process	1.69E-03
transcription from RNA polymerase II promoter	1.90E-03
aromatic compound biosynthetic process	2.59E-03
heterocycle biosynthetic process	3.31E-03
chromatin modification	4.66E-03
cellular macromolecule biosynthetic process	6.60E-03
cellular nitrogen compound biosynthetic process	7.26E-03
chromatin organization	1.04E-02
macromolecule biosynthetic process	1.04E-02
chromatin remodeling	2.16E-02
gene expression	2.47E-02
nucleic acid metabolic process	3.56E-02
cellular biosynthetic process	4.27E-02
nucleobase-containing compound metabolic process	4.66E-02
cellular aromatic compound metabolic process	4.75E-02

11	185 genes total only UGUA set (all 11's)	
	translation	3.60E-03
	gene expression	4.12E-03
	cytoplasmic translation	1.73E-02
<hr/>		
12	169 genes total only UGUA set (all 11's)	
	none	
<hr/>		
All	1043 genes total	
	cytoplasmic translation	8.04E-52
	ribonucleoprotein complex biogenesis	6.25E-11
	ribosome biogenesis	2.44E-10
	organic substance biosynthetic process	3.52E-10
	biosynthetic process	3.37E-09
	cellular biosynthetic process	3.30E-08
	gene expression	5.06E-08
	cellular macromolecule biosynthetic process	9.34E-08
	macromolecule biosynthetic process	1.86E-07
	translation	1.83E-06
	nucleic acid transport	4.84E-06
	nucleobase-containing compound transport	6.56E-06
	ribosomal small subunit biogenesis	7.98E-06
	RNA transport	8.37E-06
	establishment of RNA localization	8.37E-06
	ribosome assembly	1.68E-05
	ncRNA metabolic process	2.49E-05
	RNA localization	3.52E-05
	ribonucleoprotein complex assembly	6.73E-05
	nuclear transport	8.92E-05
	ribonucleoprotein complex subunit organization	1.15E-04
	rRNA export from nucleus	1.51E-04
	rRNA transport	1.51E-04
	nucleocytoplasmic transport	1.84E-04
	transcription, DNA-templated	2.29E-04
	nucleic acid-templated transcription	2.54E-04
	RNA biosynthetic process	2.55E-04
	cellular component organization or biogenesis	3.39E-04
	ribonucleoprotein complex export from nucleus	4.14E-04
	regulation of gene expression	4.19E-04
cellular component biogenesis	6.49E-04	
maturation of SSU-rRNA	6.63E-04	

	ribosomal large subunit biogenesis	6.64E-04
	ribonucleoprotein complex localization	7.30E-04
	regulation of cellular macromolecule biosynthetic process	7.46E-04
	nuclear export	7.57E-04
	regulation of macromolecule biosynthetic process	8.16E-04
	macromolecular complex subunit organization	1.04E-03
	regulation of biosynthetic process	1.83E-03
	chromatin remodeling	1.98E-03
	regulation of cellular biosynthetic process	2.65E-03
	ribosomal large subunit assembly	3.50E-03
	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	4.02E-03
	RNA export from nucleus	4.18E-03
	chromatin organization	4.36E-03
	transcription from RNA polymerase I promoter	4.67E-03
	chromatin modification	5.09E-03
	RNA metabolic process	5.21E-03
	regulation of macromolecule metabolic process	5.31E-03
	regulation of translation	5.87E-03
	posttranscriptional regulation of gene expression	7.42E-03
	ncRNA transcription	7.99E-03
	pyruvate metabolic process	8.07E-03
	nucleobase-containing compound biosynthetic process	8.31E-03
	regulation of primary metabolic process	1.35E-02
	rRNA-containing ribonucleoprotein complex export from nucleus	1.45E-02
	rRNA processing	1.51E-02
	organic cyclic compound biosynthetic process	1.54E-02
	rRNA metabolic process	1.99E-02
	nucleobase-containing compound metabolic process	2.08E-02
	nucleobase-containing small molecule metabolic process	2.57E-02
	regulation of translational elongation	3.48E-02
	organelle assembly	3.96E-02
	regulation of cellular metabolic process	4.79E-02
	cytosolic ribosome	2.25E-54
	cytosolic part	6.05E-53
	ribosomal subunit	8.61E-35
	ribosome	1.24E-34
	cytosolic large ribosomal subunit	4.10E-30

ribonucleoprotein	3.73E-
cytoplasmic small ribosomal	2.42E-
intracellular	2.04E-
membrane-bounded	2.09E-
intracellular non-membrane-bounded	2.09E-
large ribosomal	1.86E-
small ribosomal	3.67E-
ribosome	3.68E-
cytosol	3.21E-
intracellular	6.08E-
ribosome, large subunit	6.29E-
ribosome	6.73E-
extracellular	6.27E-
region	2.54E-
cell	2.54E-
cell	2.55E-
fungal type cell	5.47E-
intracellular organelle	4.80E-
cell	3.31E-
ribosomal encapsulating	3.31E-
ribosome	3.34E-
nuclear	6.36E-
DNA-directed RNA polymerase I	9.33E-
SWI/SNF superfamily-type	1.00E-
complex	2.50E-
DNA-directed RNA polymerase	2.73E-
DNA polymerase	2.73E-
cytoplasmic	2.66E-
nuclear DNA-directed RNA polymerase	2.28E-
complex	02

Supplementary Table 2. GO terms enrichments. GO terms for all Puf5p targets and those with specific length binding elements.

Supplementary Table 3 - Data collection and refinement statistics

RNA	5BE9 - SMX2 (A)	5BE9 – SMX2 (B)	5BE10-MFA2	5BE11 – AAT2	5BE12 – AMN1	
Data collection						
Space group	P2 ₁ 2 ₁ 2	P6 ₁ 22	P6 ₁ 22	P6 ₁ 22	P6 ₁ 22	
Cell dimensions	a, b, c (Å)	94.22, 100.79, 49.44	106.93, 106.93, 167.35	106.75, 106.75, 167.68	107.36, 107.36, 167.23	105.85, 105.85, 168.40
	α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
	Resolution (Å)	50-2.70 (2.75-2.70)	50-2.35 (2.39-2.35)	50-2.15 (2.19-2.15)	50.0-2.50 (2.54-2.50)	50.0-2.80 (2.85-2.80)
R_{sym}	0.049 (0.239)	0.098 (0.404)	0.105 (0.565)	0.051 (0.439)	0.094 (0.568)	
I / σ	38.0 (6.2)	22.9 (3.8)	22.9 (2.2)	18.5 (2.0)	20.0 (2.0)	
Completeness (%)	95.9 (73.9)	93.0 (83.2)	99.3 (94.6)	98.2 (96.9)	93.7 (91.7)	
Redundancy	6.8 (5.6)	7.3 (4.9)	10.8 (6.3)	5.0 (5.0)	6.9 (4.2)	
Refinement						
Resolution (Å)	42.7 - 2.70	45.0 - 2.35	40.5 - 2.15	32.4 - 2.50	38.5 - 2.80	
No. reflections	12,790	22,610	31,174	19,884	13,471	
$R_{\text{work}} / R_{\text{free}}$	24.1 / 28.5	19.2 / 23.6	17.9 / 21.7	23.4 / 27.4	20.1 / 25.9	
No. atoms						
Protein	2925	2952	2990	2931	2945	
RNA	186	186	230	150	190	
Water	0	145	208	57	13	
B-factors						
Protein	86.0	41.8	41.4	63.4	66.4	
RNA	108.3	54.6	45.8	87.0	84.3	
Water	-	42.7	43.8	50.3	56.1	
R.m.s deviations						
Bond lengths (Å)	0.004	0.005	0.004	0.004	0.003	
Bond angles (°)	0.71	0.87	0.79	0.75	0.67	

*Values in parentheses are for highest-resolution shell.

Supplementary Table 3. Data collection and refinement statistics.

Supplementary Table 4 - Puf5p Electrophoretic Mobility Shift Assays

RNA	RNA sequence	K_d (nM)	K_{rel}
NRE (8)	UGUAUUAUA	3200 ± 270	113
SMX2 (9)	UGUACUUAUA	28.4 ± 3.8	1
MFA2 (10)	UGUAUUUGUA	179.4 ± 38.3	6.3
AAT2 (11)	UGUAAUGAUUA	345.2 ± 61.4	12.2
AMN1 (12)	UGUAACUUUUUA	106.0 ± 9.7	3.7

EMSA's were performed in triplicate and the mean $K_d \pm$ standard error of the mean are reported as well as K_d 's relative to binding to the 9 nt SMX2 RNA, which was set to $K_{rel}=1$.

Supplementary Table 4. Puf5p Electrophoretic Mobility Shift Assays (EMSAs).

EMSAs were performed in triplicate and the mean $K_d \pm$ standard error of the mean are reported as well as K_d 's relative to binding to the 9 nt *SMX2* RNA, which was set to $K_{rel}=1$.