SUPPORTING INFORMATION APPENDIX

Synthesis of 4-thiouridine from uridine

4-Thiouridine was synthesized from commercially available uridine following literature procedures (Fig S7) (1-3). Characterization data matched the literature reports (3).

HITS-CLIP and PAR-CLIP

The CLIP-seq protocol is based on (4) and (5).

CLIP buffers (used in the protocol described below)

1X CBB buffer

25 mM Tris HCl pH 8.0
150 mM NaCl
10 mM β-mercaptoethanol
0.1% NP-40
1 mM imidazole
1 mM MgAcetate
2 mM $CaCl_2$

1X CEB buffer

25 mM Tris HCl pH 8.0
150 mM NaCl
10 mM β-mercaptoethanol
0.1% NP-40
1 mM imidazole
1 mM $MgCl_2$
2 mM EGTA

1X High salt

50 mM Tris HCl 7.4
1 M NaCl
1 mM EDTA
1% NP-40

0.1 % SDS

0.5% NaDeoxyCholate

PNK buffer 1

20 mM Tris HCl 7.4

10 mM $MgCl_2$

0.2 % Tween-20

PNK buffer 2

50 mM TrisHCl pH 7.4

10 mM $MgCl_2$

0.5% NP-40

PNK+

50 mM TrisHCl 7.4

20 mM EGTA

0.5% NP-40

5X PNK pH 6.5 (aliquot)

350 mM TrisHCl pH 6.5

50 mM $MgCl_2$

25 mM DTT

1X PK Buffer

100 mM TrisHCl pH 7.4

50 mM NaCl

10 mM EDTA

Strains, growth and lysis. WT Puf2p was TAP-tagged at its genomic locus (strain obtained from Open Biosystems). Mutant, TAP-tagged Puf2p constructs were expressed from a *CYC1* promoter on a CEN plasmid. Δpoly(N) Puf2p is a.a. 1-1016 Puf2p. PUF domain Puf2p is a.a. 464-893 Puf2p. ΔRRMΔpoly(N) Puf2p is Δ311-400 in Δpoly(N) Puf2p.

WT Puf2p cells were grown to 0.5-1.0 OD_{660} in YPAD for HITS-CLIP. WT Puf2p cells for PAR-CLIP and cells expressing mutant Puf2p were grown in synthetic media to the

same OD. Cells were frozen in 10-15 ml CBB with Roche EDTA-free protease inhibitor by dropping in liquid nitrogen. Cells were lysed by grinding for 15 min with mortar and pestle.

Purification and 3'ligation (Day 1)

1. Wash 400-800 µL/sample calmodulin-agarose beads 3x CBB, rotating for 10 min.
2. Centrifuge at low speed (3 krpm) for 5 min.
3. Add 100 U/ ml RNAsin and 1:20,000 dilution RNAse ONE.
4. Incubate in 37°C water bath for 5 min.
5. Centrifuge 5 min at 12 krpm.
6. Incubate 1 hour with calmodulin beads 4°C.
7. Wash 400-800 µL/sample Dynabeads Sheep anti-Mouse IgG (Life Technologies) 3x CEB, rotating for 10 min.
8. Wash column 2x5 ml CBB, then 2x5 ml CBB (0.2% NP-40) on column.
9. Incubate with 3 ml CEB 45 min at 4°C on rocking platform and collect in 15 ml tube. Elute further with 2x1 ml CEB and pool.
10. Incubate anti-Mouse IgG beads with eluate for 1 h at 4°C with RNAsin and protease inhibitor on rocking platform.
11. Wash with high salt (10 min at 4°C on rocking platform), transfer to new tube.
12. Wash twice with PNK buffer 1, then 10 min 37°C in:
    - 45 µL H_2O, 12 µL 5X PNK pH 6.5 buffer, 1.5 µL NEB PNK, 1.5 µL RNAsin
13. Wash 2X PNK+, 2x high salt, 2x PNK buffer 1. Transfer to a PCR striptube and step down volume.
14. 3' ligation *per* 20 µL (use ~80 µL total vol.):
    - 11.9 µL H_2O, 1.8 µL 10X buffer, 1 µL Ligase, 0.5 RNAsin, 1.5 µL L3 adapter, 4 µL PEG400.
    - Incubate 16°C rotating 3 h to overnight in strip-tube.

5' ligation (Day 2)

1. Wash 1x high salt.
2. Wash 2x PNK.
3. PNK (in Eppendorf tube):

- 2 µL 10X buffer, 1µL PNK, 2 µL 10 mM ATP, 14 µL H_2O, 1 µL RNAsin
- Incubate 10 min at 37°C. Shake 1000 RPM for 15 sec every 4 min.

4. Wash 2x PNK+, 1X high salt, 2X PNK buffer 2
5. 5' ligation (in strip-tube):
    - 2 µL 10X, 2 µL BSA, 2 µL 10 mM ATP, 2 µL 5' linker 20 µM, 11.5 µL H_2O, 0.5 µL T4 RNA ligase, 0.5 µL RNAsin
    - 16°C rotating 2 h to overnight.
6. Gel and transfer:
    - Add ~10 µL PNK+, ~20 µL 4X NuPAGE loading dye, ~4 µL NuPAGE reducing agent.
    - Heat for 10 min at 70°C. Do not chill.
    - Load on NuPAGE gel. Run according to manufacturer's instructions.
    - Transfer to nitrocellulose membrane 1 h at 100 V, using the NuPAGE transfer buffer.
7. Proteinase K digestion:
    - Pre-incubate >200 µL 1X proteinase K buffer with 1.3 mg/ml proteinase K at 37°C for 14 min.
    - Add 200 µL pre-incubated buffer to membrane slice. Incubate 30 min to 4 h at 37°C.
    - Add 200 µL PK+7M urea solution. Incubate 30 min 37°C.
    - Prepare phase-lock tube 25 sec, 14 kRCF.
8. Combine PK extraction with 400 µL phenol-chloroform in phase-lock tube.
9. Mix 5 min at 30°C, 1100 RPM.
10. Spin 5 min, 13 kRPM. Transfer to new tube.
11. Precipitate with 0.7 µL GlycoBlue (Life Technologies), 40 µL 3 M NaOAc, 1 ml 100% EtOH.
12. Place -20°C overnight.

Reverse transcription (Day 3)

1. Spin down pellet 30 min at 4°C.
2. Wash pellet 3x 500 µL 80% ethanol.

- Large, white pellets are too salty to make cDNA. In general, pellets can be washed down to the blue core of GlycoBlue.

3. Combine 8 µL RNA and 2 µL RTP primer 5 µM.
4. Heat 65°C 5 min. Chill and quick spin.
5. Add 1 µL 10 mM dNTPs, 2 µL H_2O, 1 µL 0.1 M DTT, 4 µL 5X RT buffer, 1 µL RNAsin, 1 µL SSIII.
6. Incubate 50°C for 50 min, 55°C for 10 min, 90°C for 5 min. Hold at 4°C.

PCR

1. Combine 10 µL 5X Phusion buffer, 1 µL RP1 primer (10 µM), 1 µL index primer (10 µM), 1.25 dNTPs (10 mM), 32.25 µL water, 4°C RT product and 0.5 µL Phusion.
2. Run 30 sec 98°C, then 26 cycles of 10 sec 98°C, 30 sec 60°C, 15 sec 72°C. Finish with 10 min at 72°C and hold at 4°C.
    - Initial tests are done with 1-2 µL RT product and 18, 26 and 35 cycles to identify the lowest cycle number that gives a smear.
3. Clean-up with a PCR clean-up kit before high-throughput sequencing.

Yeast three-hybrid assays

The yeast three-hybrid assay was performed as described previously (6). Sequences are described in Tables S5-S6. The compensatory mutant experiment was performed with pGADT7-Puf2p (1-1016) (Δpoly(N)) as WT specificity Puf2p. R1 SNE Puf2p was pGADT7-Puf2p (1-1016) R1 NTQ to SNE. The sequences from *PMP2* and *ARF1* used in the three-hybrid are described in Tables S5 and S6.

Sequence analysis

The phylogenetic tree in Fig 1B was constructed using COBALT (7). TRMs were extracted from an alignment from (8) using custom scripts.

Data analysis

Duplicate removal and mapping. Fastq-mcf was run on the raw fastq files to remove the adapter and remove duplicate reads, using the first 35 bases (fastq-mcf adapters.fa

-D 35). Because the first five bases are a random adapter, reads that map to the same place but differ in their adapter are retained. Because all experiments are collapsed to their unique reads at the very first step, further processing is faster, and samples are immediately comparable. This method does not control for sequencing error, a problem with highly over-sequenced samples. We also wrote scripts that filter mapped reads using information saved from the random barcode, but this method was slower and more complex than simply filtering the raw fastq files with fastq-mcf.

Reads were mapped to the EF4 genome using bowtie2 (bowtie2 –x genome –U fastq –phred33 –local). Poor quality reads, and reads mapping to multiple places in the genome, were removed by filtering with samtools view –q 7. PAR-CLIP reads are heavily UV-damaged, so a relatively generous cutoff is applied.

Peak finding. The CLIP-seq peak caller written for this work is found at github.com/dfporter.

We reasoned that an experimenter generally has a lower bar for what they would consider a peak. As a result, only such regions of the genome need be investigated. Regions of at least 10 reads raw coverage were extracted as a first step.

For each region, the highest point is found, and the peak is extended until coverage drops to 20% of the peak height. Overlapping peaks are merged. Peaks were assigned to the closest gene. We took only the highest peak per gene.

A Poisson p value was calculated for each peak. This was done by binning the targeted gene and counting the number of reads in the given CLIP-seq experiment falling in each bin. Reads were placed according to their 5’ position, so they were not counted in more than one bin. The bin size was 50 bp. Introns were discarded. The average number of reads in a bin was taken to be λ, the single Poisson parameter, from which is obtained the p value as the chance of obtaining a peak as high as that observed in the peak in a

bin. This number is multiplied by the number of bins in the gene to give the used p value.

In the case of ribosomal loci, signal is modeled in the 1 kbp region around the peak, rather than by gene. Originally, the software applied this method across the genome, but the local method had the undesirable effect that many tRNAs were called in smaller datasets using the local method.

RNA-seq or a negative IP was used to calculate a negative binomial p value. This is also done by binning the assigned gene and placing the 5' ends of the control dataset into bins. We discarded bins with zero reads. The negative binomial requires two parameters, which were fit using R. Resulting p values were multiplied by the number of bins in the given gene to obtain the peak's p value. An important note is that our programs treated control datasets exactly the same as experimental datasets, except that there was no removal of random 5-mer adapters if none were present, and that this included the removal of duplicate reads from fastq files when the input was RNA-seq. We noticed this odd decision increased enrichment of the binding site >2% in HITS-CLIP (and did not change the PAR-CLIP enrichment). Since enrichment of the binding site was our metric for correct program performance, we left the oddity in place. An effect of this decision is that the abundances of very abundant genes are under-estimated, while other genes are over-estimated (due to the change in the normalization factor based on dataset size). NB p values are therefore over-estimates for low abundance genes and under-estimates for abundant genes, which apparently balances out to a more accurate picture of binding. This may be due to NB and Poisson comparisons being more valuable (in maximizing motif enrichment) at opposite ends of the abundance spectrum. Our Poisson p value cutoff is very low (10^{-6} for the low stringency cutoff), meaning that all peaks are peaks by the conventional standard of being a region of unlikely deep CLIP-seq coverage, and that less than one false target is expected by this metric.

HITS-CLIP samples were controlled using RNA-seq of wild-type yeast cells grown to log phase in synthetic complete media, at 2% glucose, using data from (9) (specifically GSM1299413). PAR-CLIP samples were controlled using RNA-seq of cells grown to log phase in synthetic media in the presence of 4-thiouridine, using data from (10) (specifically GSM1070246).

We used a height of 10 reads, a Poisson p value of 10^{-6}, and a NB p value of 10^{-4} for the low cutoff and a height of 20 reads, a Poisson p value of 10^{-7}, and a NB p value of 10^{-8} for the high cutoff.

For the R1 SNE mutant, which overall provided the largest dataset, there was no decrease in site enrichment between high and low stringency cutoffs (Table S4, Fig S2), indicating the low cutoff is more accurate. We therefore used the low cutoff for the R1 SNE mutant. A low cutoff was also applied when combining replicates for GO analysis.

Prediction of mRNA targets

Binding location within mRNA. Histograms in Fig 4 are fit to a Gaussian kernel for easier visualization. When identifying the position of maximum motif density, both UAAU and UAAG 3'UTR motifs are counted twice to improve the fit. When determining the accuracy of the site, true peak locations were defined as the center of the peak range.

Identification of factors important for binding. We ranked Δpoly(N) Puf2p and R1 SNE targets by height, and took the top 200 targets as the set of frequent targets. Presence in this set was defined as the dependent variable. Additional gene expression data was obtained from (10). mRNA localization and ribosome profiling data were obtained from (11) and (12). Only genes with data in both (10) and (11) were included. The random forests model (13) was built using Δpoly(N) Puf2p (on the site UAAU) and then applied to R1 SNE Puf2p. Performance of the prediction was evaluated with the ROCR package (14).

RNA-seq

RNA was extracted from log phase *S. cerevisiae* in minimal media according to published protocols (15). Stranded libraries were prepared for 100 bp single read sequencing on a HiSeq 2000 (Illumina) using poly(A) selection and the TruSeq v3 kit (Illumina). Analysis utilized 4 biological replicates of Δ*puf2* BY4742 (Open Biosystems) cells expressing R1 SNE Puf2p, 3 replicates of cells expressing Δpoly(N) Puf2p (referred to as WT Puf2p in this context), and 4 replicates of cells transformed with an empty vector. Puf2p constructs were expressed from a *CYC1* promoter on a CEN plasmid. After mapping to the EF4 genome, reads were assigned to genes using HTSeq (16), and differential expression was determined using DESeq2 (17) with an adjusted p value cutoff of 0.05. DESeq2 was run with the default settings, except at 0.05 FDR, as described in the DESeq2 manual vignette. DESeq2 calculates adjusted p values by the Wald test, followed by the Benjamini and Hochberg method to correct for multiple comparisons. The p values in tables S9-S11 are the Benjamini and Hochberg adjusted p values. Sensitivity for mild effects on the most frequent 100 targets is increased because the most frequent 100 targets are mostly abundant RNAs. We compared multiple MAPQ cutoffs, and found that applying high filters removed the vast majority of reads from strong WT Puf2p targets that exist as duplicate genes or near-duplicates, including *PMP2*, *PMP1*, *PHO11*, *PHO12*, *HXT6* and *HXT7*, due to nucleotide similarity between paralogs. Dataset 3 includes both the results without filtering and with a MAPQ >= 20 filter; we assume the filtered results are necessary for other investigators, if harder to discuss in a single paragraph of the main text. All conclusions in the main text are true for both, except the significance of R1 SNE's repression of the WT network changes from p value 0.077 to 0.02 with MAPQ filtering, crossing the 0.05 threshold. We chose to present results for the unfiltered data, as the inclusion of the expected *PMP1/*2 was more important for accuracy than the issues associated with allowing a read to have multiple alignments. In either case, the effect of R1 SNE on the WT network is roughly half its effect on its own network, consistent with retaining roughly half the WT targets (by median reduction of the top 100 targets, filtered: cognate network: 5.4%, non-cognate network: 3.3%; unfiltered: cognate network: 7.3%, non-cognate network: 3.7%).

qRT-PCR

RNA was extracted from log phase Δ*puf2* BY4742 *S. cerevisiae* in minimal media using the same vector constructs and RNA extraction used for RNA-seq (15). RNAs were reverse transcribed using oligo(dT) primers. Amplification was performed using Taqman probes (Applied Biosystems) according to the manufacturer's instructions. We performed 3 technical replicates of 6 or 7 biological replicates, and analyzed using the $\Delta\Delta C_q$ method. The spreadsheet provided by (18) was used to estimate abundances. P values were determined by t-test.

Reporter assays for regulation

HIS3. The HIS3 repression assay was performed as described previously (19). BY4742 cells lacking both the *PUF2* and *HIS3* genes (Open Biosystems) were transformed with vectors expressing the indicated protein or RNA. The *HIS3* gene was expressed on a vector with the tested 3'UTR. 3-AT denotes 3-Amino-1,2,4-triazole, a *HIS3* inhibitor added to increase stringency. Converting all UAAUs to UAAGs created the *PMP2 mut* construct.

β-galactosidase. Reporter constructs comprised β-galactosidase under the *CYC1* promoter, followed by the tested 3'UTR. All expression levels were obtained from six biological replicates, each measured at four stages of growth in log phase. Protein levels were determined by taking the expression value at OD660 = 0.45 estimated from the measured OD660 values. All measurements were taken in Δ*puf2* BY4742 cells. β-galactosidase levels were first normalized to β-galactosidase expression in Δ*puf2* BY4742 cells bearing an empty expression vector.

REFERENCES

1. Shi L, Zhang G, & Pan F (2008) $Fe_2(SO_4)_3 \cdot xH_2O$-catalyzed per-*O*-acetylation of sugars compatible with acid-labile protecting groups adopted in carbohydrate chemistry. *Tetrahedron* 64(11):2572-2575.
2. Polshettiwar V & Kaushik MP (2006) Alumina encapsulated phosphorus pentasulfide (P_4S_{10}/Al_2O_3) mediated efficient thionation of long chain amides. *Tetrahedron Lett* 47(14):2315-2317.
3. El-Tayeb A, Qi A, Nicholas RA, & Müller CE (2011) Structural modifications of UMP, UDP, and UTP leading to subtype-selective agonists for P2Y2, P2Y4, and P2Y6 receptors. *J Med Chem* 54(8):2878-2890.
4. Licatalosi DD*, et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456(7221):464-469.
5. Konig J*, et al.* (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* 17(7):909-915.
6. Yosefzon Y*, et al.* (2011) Divergent RNA binding specificity of yeast Puf2p. *RNA* 17(8):1479-1488.
7. Papadopoulos JS & Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073-1079.
8. Huerta-Cepas J, Capella-Gutiérrez S, Pryszcz LP, Marcet-Houben M, & Gabaldón T (2014) PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res* 42(D1):D897-D902.
9. Dang W*, et al.* (2014) Inactivation of yeast Isw2 chromatin remodeling enzyme mimics longevity effect of calorie restriction via induction of genotoxic stress response. *Cell Metab* 19(6):952-966.
10. Freeberg MA*, et al.* (2013) Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. *Genome Biol* 14(2):R13.
11. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, & Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159(3):635-646.
12. McManus CJ, May GE, Spealman P, & Shteyman A (2014) Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res* 24(3):422-430.
13. Breiman L (2001) Random forests. *Mach Learn* 45(1):5-32.
14. Sing T, Sander O, Beerenwinkel N, & Lengauer T (2005) ROCR: visualizing classifier performance in R. *Bioinformatics* 21(20):3940-3941.
15. Wilhelm BT, Marguerat S, Goodhead I, & Bahler J (2010) Defining transcribed regions using RNA-seq. *Nat. Protocols* 5(2):255-266.
16. Anders S, Pyl PT, & Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166-169.
17. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):1-21.
18. Willems E, Leyns L, & Vandesompele J (2008) Standardization of real-time PCR gene expression data from independent biological replicates. *Anal Biochem* 379(1):127-129.
19. Hook BA, Goldstrohm AC, Seay DJ, & Wickens M (2007) Two yeast PUF proteins negatively regulate a single mRNA. *J Biol Chem* 282(21):15430-15438.

SUPPLEMENTAL FIGURE LEGENDS

Fig S1 Diagram of analysis pipeline.

Fig S2 Correlation between NB p value and metrics for motif enrichment and peak geometry. Genes are ranked by NB p value. A low stringency cutoff is applied here. Not all experiments have the same number of targets, so lines terminate at different ranks. Ribbons represent standard error.

Fig S3 A Fraction of RNA targets of the given type, for each CLIP-seq experiment. **B** Fraction of mRNA peaks in each region of the RNA.

Fig S4 Puf2p mutants have reduced discrimination between RNAs. Results of RIP, followed by qRT-PCR for top Puf2p target *ZEO1* normalized to actin, *ACT1* (3 biological replicates, each with 3 technical replicates).

Fig S5 A Example western blot of Puf2p mutants and actin expression in lysate. **B** Levels of Puf2p mutant proteins by western blot, normalized to actin protein expression levels in the same cells, and then normalized by replicate to Δpoly(N) Puf2p. Squares represent replicates. The PUF domain of Puf2p is highly stabilized relative to other Puf2p constructs.

Fig S6 Relative enrichment of CLIP-seq signal over RNA-seq signal at dual motifs in the given location.

Fig S7 Scheme 1 New synthesis of 4SU.

Fig S8 Binding to the overlapping UAAUAAU site in *PMP2* requires a downstream two nucleotides for a full cognate motif of UAAUAAUUA. All peaks with significant binding ($p < 0.05$, t-test) compared to the empty vector control are marked with an asterisk.

Fig S9 Puf2p represses β-galactosidase reporters through binding its cognate RNA element in 3'UTRs. Reporter constructs comprised *CYC1* promoter-driven β-galactosidase followed by a 3'UTR from either *PMP2* or *ARF1*. A mutant *PMP2* 3'UTR was constructed in which UAAUs were converted to UAAG, and a mutant *ARF1* 3'UTR was constructed in which UAAGs were converted to UAAU. Tested 3'UTRs are diagramed on the left. When WT Puf2p is expressed, both *PMP2* and *ARF1* 3'UTRs are repressed when the 3'UTR contains UAAU sites, and not in the case of UAAG sites. When R1 SNE Puf2p is expressed, there is an indirect activation of the reporter, an effect also observed when no 3'UTR is present. However, the expression of R1 SNE Puf2p represses the UAAG-containing *PMP2* mutant, and not the UAAU form. The effect of R1 SNE Puf2p on UAAG-containing *ARF1* RNA was not significantly different than its effect on the UAAU-containing *ARF1* RNA. This may be caused by the low number of UAAG sites (two) and insufficient sensitivity at the given replicate number. Asterisks mark cases in which β-galactosidase activity for the 3'UTR bearing the cognate site (i.e., UAAU for WT Puf2p) is significantly different from the non-cognate 3'UTR (UAAG for WT Puf2p) at p value < 0.05 by t-test. All expression levels were obtained from six biological replicates (except one n=5 set), each measured at four stages of growth in log phase, and β-galactosidase expression levels from the resulting curve at OD660 0.45 were used for comparison.

Fig S10 Puf2p represses a *HIS3* reporter with a *PMP2* 3'UTR. BY4742 cells lacking both *PUF2* and *HIS3 g*enes were transformed with vectors expressing the indicated protein or RNA. In the right column, cells were grown in the absence of histidine and presence 3-aminotriazole, and therefore required the expression of the HIS3 reporter to grow. The smaller sizes of the cells in the top row indicate the HIS3 reporter is repressed only when both protein and RNA are WT.

Fig S11 Puf2p decreases the abundance of certain target RNAs. The levels of *ZEO1, MRH1, ARF1, SOD1* and *CNB1* endogenous mRNAs were determined by qRT-PCR for cells expressing no Puf2p, WT Puf2p or R1 SNE Puf2p. All experiments were performed using 6-7 biological replicates. Error bars denote 95% confidence intervals.

The table below the figure indicates binding by WT or R1 SNE Puf2p. *ZEO1* and *MRH1* are in fact targets of R1 SNE Puf2p, but so reduced in coverage that we have treated them as relative non-targets in this figure. *CNB1*, which is not bound by either WT or R1 SNE Puf2p, is included as a negative control. *SOD1* is a more complex case, as WT Puf2p binds it in the 5'UTR and R1 SNE Puf2p in the 3'UTR. Only R1 SNE Puf2p represses *SOD1*, indicating binding the 3' UTR site may be more repressive. Repression of *MRH1* has a ~0.07 p value

Fig S12 Changes in RNA abundance in response to WT Puf2p expression (top) or R1 SNE Puf2p expression (bottom), relative to cells not expressing any Puf2p. Both WT and R1 SNE Puf2p generally repress target mRNAs. Taking the top 100 RNA targets of either WT or R1 SNE Puf2p as their respective networks, WT Puf2p represses its WT network (p value < 0.05), but not the R1 SNE network (p value > 0.4). R1 SNE Puf2p represses its novel network at high significance (p value < 10^{-6}), and does not significantly repress the WT network (p value > 0.05). Notable targets, and all targets assayed by qRT-PCR, are labeled if present in the top 100 targets See Tables S7-9 for significantly altered RNA abundances and Dataset S3 for all RNA abundances

SUPPLEMENTAL TABLES

Table S1. The Puf2p family

Phyla/Subphyla	Species	Gene/Accession	Predicted RNA recognition triplets					RRMs
Ascomycota	*S. cerevisiae*	PUF1		NTQ	TWQ	NTQ	SRR	1
Ascomycota	*S. cerevisiae*	PUF2		NTQ	TWQ	NTQ	ARR	1
Ascomycota	*S. pombe*	NP_595389		NTQ	TWQ	NTQ	SRR	1
Ascomycota	*C. glabrata*	XP_447180		NTQ	TWQ	NYQ	**VRR**	1
Ascomycota	*N. crassa*	XP_962915		NTQ	TWQ	NYQ	ARR	1
Ascomycota	*L. elongisporus*	XP_001525124*		NTQ	**TYQ**	NYQ	ARR	0
Basidiomycota	*U. maydis*	XP_756415		NTQ	TWQ	NTQ	SRR	2
Basidiomycota	*C. neoformans*	AFR94783		NTQ	TWQ	**NYA**	ARR	2
Mucoromycotina	*R. delemar*	EIE81026		NTQ	TWQ	NYQ	ARR	2
Chytridiomycota	*B. dendrobatidis*	XP_006675799.1		**NIQ**	TWQ	NYQ	ARR	2
Mortierellomycotina	*M. verticillata*	KFH68401		NTQ	TWQ	NYQ	ARR	2
Non-PUF2 family:								
Ascomycota	*S. cerevisiae*	PUF5/MPT5	CRQ	NYQ	TRQ	NHQ	CCQ	0
Ascomycota	*S. cerevisiae*	PUF4	CRQ	NYQ	TRQ	NHQ	CCQ	0

Table S1. Representative proteins in the Puf2p family. We included those proteins that showed some alteration of the expected TRM pattern. * denotes a protein (in *L. elongisporus*) that has characteristics intermediate between the PUF2 and PUF5 families.

Table S2. Mapping statistics for CLIP

Technique	Protein	Replicate	Unique reads	Length (σ)	Mapped	% Mapped
CLIP-seq	WT Puf2p	1	$1 * 10^6$	44 (26.3)	$8 * 10^5$	68.6%
	WT Puf2p	2	$5 * 10^5$	54 (27.6)	$3 * 10^5$	65.5%
	Untagged cells	1	$3 * 10^5$	29 (11.2)	$2 * 10^5$	45.1%
PAR-CLIP	WT Puf2p	1	$1 * 10^6$	46 (27)	$9 * 10^5$	59.4%
	WT Puf2p	2	$8 * 10^5$	37 (10.3)	$5 * 10^5$	62.1%
CLIP-seq (mutants)	PUF domain Puf2p	1	$8 * 10^5$	37 (10.2)	$6 * 10^5$	78.0%
	PUF domain Puf2p	2	$3 * 10^5$	46 (8.4)	$2 * 10^5$	72.0%
	Δpoly(N) Puf2p	1	$2 * 10^6$	36 (10.3)	$2 * 10^6$	84.5%
	Δpoly(N) Puf2p	2	$4 * 10^5$	46 (9.8)	$3 * 10^5$	72.0%
	Δpoly(N)ΔRRM Puf2p	1	$8 * 10^5$	38 (9.9)	$4 * 10^5$	58.0%
	R1 SNE Δpoly(N) Puf2p	1	$2 * 10^6$	39 (9.4)	$2 * 10^6$	89.2%
	R1 SNE Δpoly(N) Puf2p	2	$1 * 10^6$	36 (10.5)	$1 * 10^6$	74.1%

Table S2. Mapping statistics for CLIP experiments. Unique reads and mapped reads are after duplicates are removed (using the random 5-mer adapter). Replicates were merged before peaks were called.

Table S3. DREME results

	Sequences	Motif	Positives	Negatives (/1000)	p value
HITS-CLIP WT Puf2p	1012	TAAT	770	270	4.2e-112
		TAWATA	194	40	3.5e-28
PAR-CLIP WT Puf2p	478	HHTAAT	309	230	5.5e-54
		TAHATA	141	70	6.9e-29
		TTTTHYT	129	89	6.8e-19
Δpoly(N) Puf2p	1034	TAAT	872	303	3.1e-142
		TAHATA	239	73	4.1e-24
R1 SNE Δpoly(N) Puf2p – high cutoff	1012	TAAG	878	162	7.8e-243
		TGTAHAT	107	13	1.3e-20
R1 SNE Δpoly(N) Puf2p – low cutoff	1696	TAAG	1454	167	1.4e-291
		TGTAHATA	164	15	1.1e-19
ΔRRM Δpoly(N) Puf2p	1468	HTAAT	620	188	1.3e-35
		TANATA	209	59	1.1e-11
PUF domain Puf2p	191	TAATWW	90	136	4.8e-23
Untagged cells	36	TSTTTTC	7	16	5.5e-6

Table S3. DREME results**.** A high cutoff was applied in all cases. We removed non-coding RNA before running DREME, resulting in the slightly different sequence numbers here.

Table S4. Target numbers

	High		Low	
	Targets	**% with UAAU**	**Targets**	**% with UAAU**
HITS-CLIP WT Puf2p	1131	73%	2167	66%
…Larger replicate	888	88%	1893	79%
PAR-CLIP WT Puf2p	546	66%	1025	59%
…Larger replicate	480	68%	887	64%
Δpoly(N) Puf2p	1115	81%	1786	77%
…Larger replicate	938	89%	87%	1786
ΔRRM Δpoly(N) Puf2p	1559	48%	2390	46%
PUF domain	266	59%	528	57%
R1 SNE Puf2p	1117	34% (83% UAAG)	1846	32% (83% UAAG)
Untagged cells	52	35% (21% UAAG)	85	39% (23% UAAG)

Table S4. Target numbers and proportion of peaks with UAAUs for different samples.

Table S5. *PMP2* RNAs

RNA name	Sequence of insert
PMP2 WT	ATTTCTAATAATTAATTAATTTATCCT
PMP2*	ATTTCTAATAATTAATACATTTATCCT
U1G	ATTTCGAATAATTAATACATTTATCCT
A2G	ATTTCTGATAATTAATACATTTATCCT
A3G	ATTTCTAGTAATTAATACATTTATCCT
A4U	ATTTCTAAGAATTAATACATTTATCCT
U8G	ATTTCTAATAATGAATACATTTATCCT
A9G	ATTTCTAATAATTGATACATTTATCCT
A10G	ATTTCTAATAATTAGTACATTTATCCT
U11G	ATTTCTAATAATTAAGACATTTATCCT
U1G U8G	ATTTCGAATAATGAATACATTTATCCT
A2G A9G	ATTTCTGATAATTGATACATTTATCCT
A3G A10G	ATTTCTAGTAATTAGTACATTTATCCT
U4G U11G	ATTTCTAAGAATTAAGACATTTATCCT

Table S5. *PMP2* RNAs assayed by yeast three-hybrid. Compensatory G mutations are in red.

Table S6. *ARF1* RNAs

RNA	Sequence
ARF1 WT	CATAAAGATTAAGAACTTAAGAGGCCA
ARF1*	CAAAAACATTAAGAACTTAAGAGGCCA
Monomeric UAAU	CAAAAACATTAATAACTACAGAGGCCA
Monomeric UAAG	CAAAAACATTAAGAACTACAGAGGCCA
Dual TAAT	CAAAAACATTAATAACTTAATAGGCCA
TAATAAT	CAAAAACATTAATAATTACAGAGGCCA
TAATAAG	CAAAAACATTAATAAGTACAGAGGCCA
TAAGAAG	CAAAAACATTAAGAAGTACAGAGGCCA
TAAG + overlapping TAATAAG	CAAAAACATTAATAAGTTAAGAGGCCA
TAAG + overlapping TAAGAAG	CAAAAACATTAAGAAGTTAAGAGGCCA
No sites	CAAAAACATACAGAACTACAGAGGCCA

Table S6. *ARF1* yeast three-hybrid RNAs, with functional elements underlined. The *ARF1* WT sequence is derived the 3'UTR of *ARF1*. A TAAAG upstream element in *ARF1* (in red) was mutated to AAAAC in the *ARF1** construct and all other RNAs.

Table S7. mRNA abundance changes in response to WT Puf2p expression

ID	Standard name	Gene Name	Fold change (log2)	p value (adjusted)	WT rank	R1 SNE rank
YHR215W	PHO12	PHOsphate metabolism	-7.39E-01	1.36E-13	280	
YPL095C	EEB1	Ethyl Ester Biosynthesis	1.07E+00	3.59E-07		
YNL031C	HHT2	Histone H Three	-3.39E-01	4.41E-06	117	109
YEL017C-A	PMP2	Plasma Membrane Proteolipid	-4.67E-01	9.76E-06	4	385
YGL008C	PMA1	Plasma Membrane ATPase	-3.30E-01	1.31E-04	1	18
YGR035C	YGR035C	NA	8.54E-01	1.31E-04		
YNL220W	ADE12	ADEnine requiring	3.62E-01	6.18E-04		424
YLR044C	PDC1	Pyruvate DeCarboxylase	-3.78E-01	9.03E-04		46
YAR071W	PHO11	PHOsphate metabolism	-6.14E-01	1.49E-03	445	
YAR018C	KIN3	protein KINase	4.74E-01	5.12E-03		
YBL085W	BOI1	Bem1 (One) Interacting protein	-3.19E-01	1.75E-02	15	856
YHR146W	CRP1	Cruciform DNA-Recognizing Protein	-3.01E-01	2.94E-02	17	221
YBR082C	UBC4	UBiquitin-Conjugating	2.29E-01	2.94E-02		48
YNL058C	YNL058C	NA	4.85E-01	3.56E-02		
YOR202W	HIS3	HIStidine	-4.38E-01	4.31E-02		

Table S7. Genes with significant mRNA abundance changes in response to WT Puf2p expression, compared with cells lacking Puf2p.

Table S8. mRNA abundance changes in response to R1 SNE Puf2p expression

ID	Standard name	Gene Name	Fold change (log2)	p value (adjusted)	WT rank	R1 SNE rank
YDL192W	ARF1	ADP-Ribosylation Factor	-4.15E-01	2.19E-14		3
YPR074C	TKL1	TransKetoLase	-2.34E-01	2.55E-03		439
YGR185C	TYS1	TYrosyl-tRNA Synthetase	-3.57E-01	4.94E-03		255
YDL015C	TSC13	Temperature-sensitive Suppressors of Csg2 mutants	-1.89E-01	1.00E-02		
YPL184C	MRN1	Multicopy supressor of rsc nhp6	-1.83E-01	1.33E-02	249	498
YLR274W	MCM5	MiniChromosome Maintenance	2.31E-01	3.03E-02		
YPL242C	IQG1	IQGAP-related protein	2.95E-01	3.03E-02		592
YIL123W	SIM1	Start Independent of Mitosis	-3.15E-01	3.03E-02		47
YMR122W-A	NCW1	Novel Cell Wall protein	2.83E-01	3.03E-02	42	317
YDR226W	ADK1	ADenylate Kinase	-2.17E-01	3.03E-02	327	
YGL022W	STT3	STaurosporine and Temperature sensitive	-1.69E-01	3.03E-02		
YKL104C	GFA1	Glutamine:Fructose-6-phosphate Amidotransferase	1.94E-01	3.10E-02	628	394
YJR064W	CCT5	Chaperonin Containing TCP-1	-2.36E-01	3.10E-02		
YMR205C	PFK2	PhosphoFructoKinase	-2.52E-01	3.66E-02	173	162

Table S8. Genes with significant mRNA abundance changes in response to R1 SNE Puf2p expression, compared with cells lacking Puf2p.

Table S9. The effect of Puf2p redesign

ID	Standard name	Gene Name	Fold change (log2)	p value (adjusted)	WT rank	R1 SNE rank
YPL095C	EEB1	Ethyl Ester Biosynthesis	9.77E-01	2.72E-08		
YNL190W	YNL190W		-2.80E-01	4.32E-07	83	417
YBR082C	UBC4	UBiquitin-Conjugating	4.62E-01	1.39E-04		48
YGR138C	TPO2	Transporter of POlyamines	-6.59E-01	9.35E-04	36	479
YOR270C	VPH1	Vacuolar pH	2.02E-01	1.32E-03		
YCL055W	KAR4	KARyogamy	-3.04E-01	5.62E-03		
YPR028W	YOP1	YIP One Partner	3.69E-01	5.62E-03		
YPL179W	PPQ1	Protein Phosphatase Q	-2.15E-01	5.62E-03		
YDL192W	ARF1	ADP-Ribosylation Factor	4.46E-01	5.62E-03		3
YBL085W	BOI1	Bem1 (One) Interacting protein	-2.64E-01	1.61E-02	15	856
YBR159W	IFA38		2.22E-01	1.61E-02		
YEL017C-A	PMP2	Plasma Membrane Proteolipid	-3.29E-01	1.61E-02	4	385
YPL187W	MF(ALPHA)1	Mating Factor ALPHA	2.72E-01	3.98E-02		13

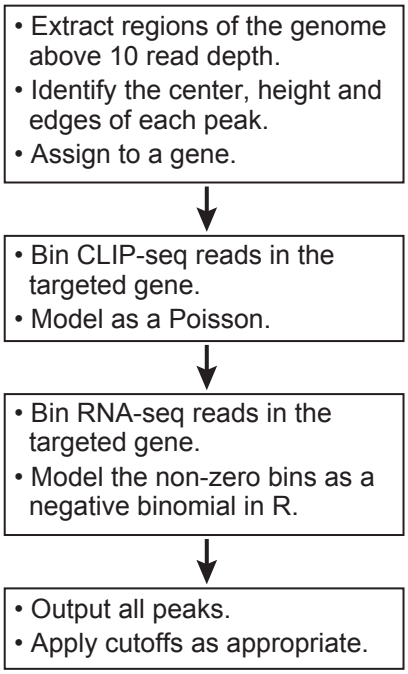
Table S9. Genes with significant mRNA abundance changes in response to R1 SNE Puf2p expression, compared with cells expressing WT Puf2p.

SUPPLEMENTARY DATASETS

Dataset S1. [In excel file.] RNA targets by CLIP-seq for Puf2p and Puf2 mutants at **high** cutoff.

Dataset S2. [In excel file.] RNA targets by CLIP-seq for Puf2p and Puf2 mutants at **low** cutoff.

Dataset S3. [In excel file.] Read counts for all genes in all RNA-seq replicates, and DESeq2-estimated relative expression levels.



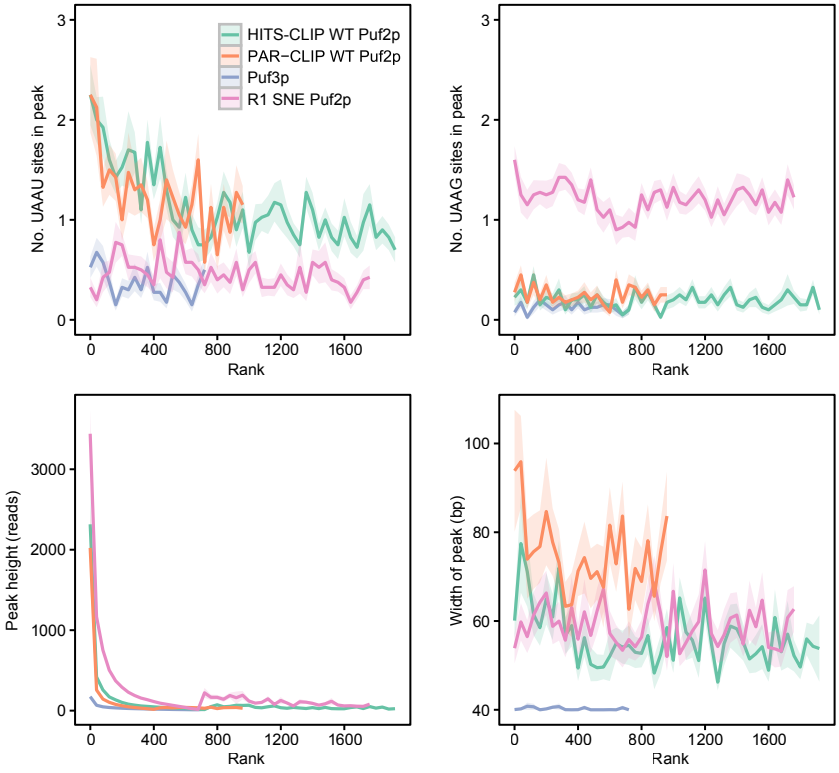
1. Identify regions of sufficient raw coverage.
While only regions with 10 read depth are initially considered as potential peaks, the final peak height cutoff is applied at the end of the analysis pipeline. Only the highest peak per gene is considered.

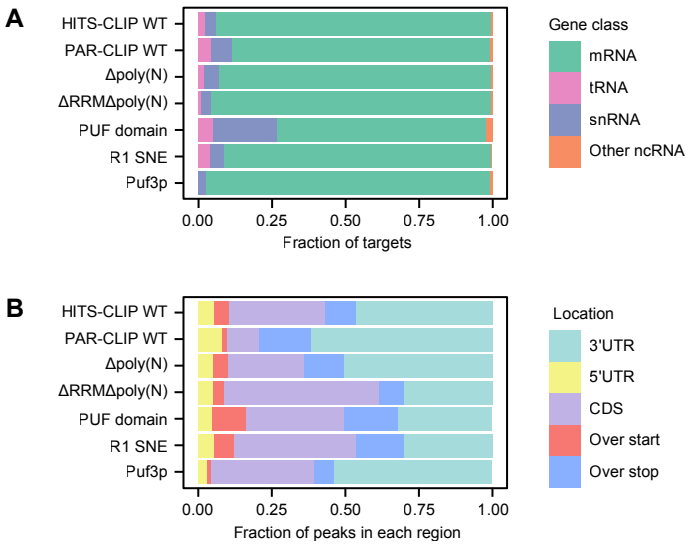
2. Identify regions that are peaks.
We assume that a CLIP-seq peak should be enriched relative to CLIP-seq signal across the entire gene.

3. Identify regions that are enriched.
We assume that CLIP-seq peaks should be enriched relative to RNA-seq signal in the targeted gene.

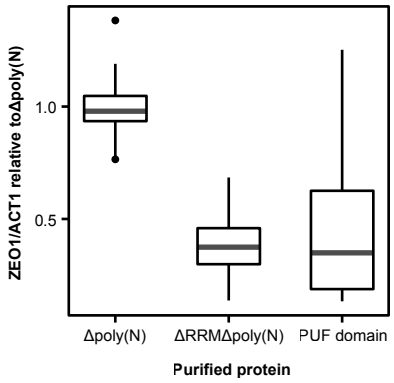
Statistical cutoffs are applied at this point before further analysis.

Supp. Figure 2

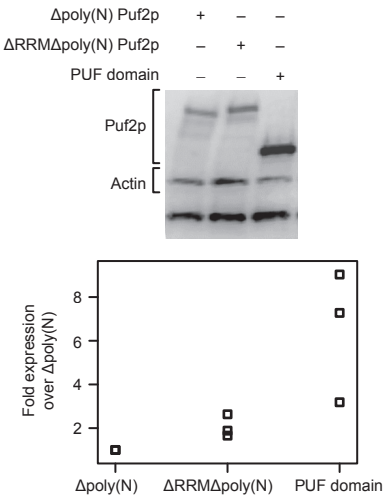


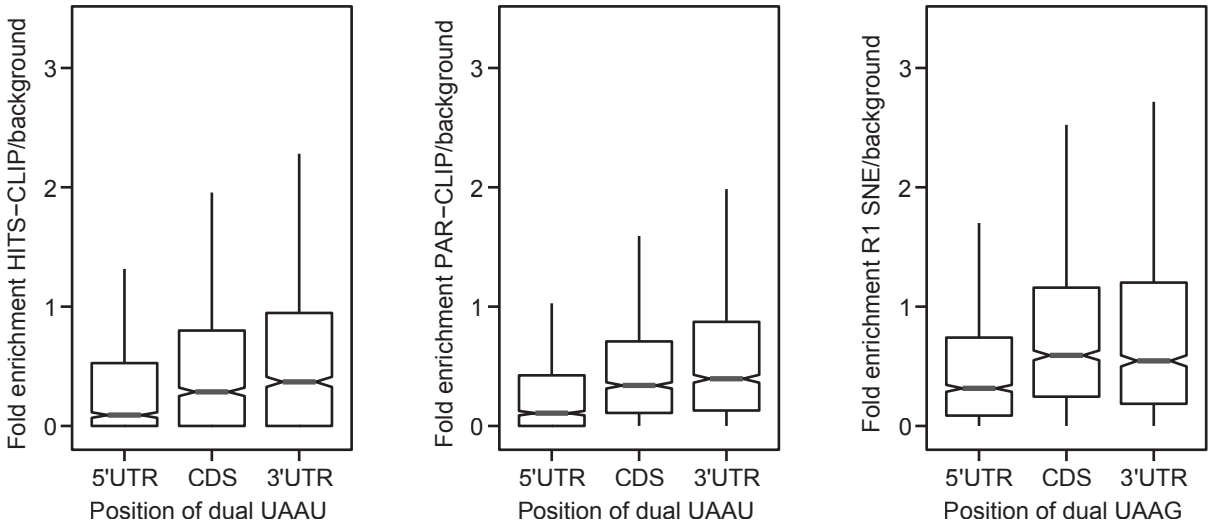

A
HITS-CLIP WT
PAR-CLIP WT
Δpoly(N)
ΔRRMΔpoly(N)
PUF domain
R1 SNE
Puf3p
0.00
0.25
0.50
0.75
1.00
Fraction of targets
Gene class
mRNA
tRNA
snRNA
Other ncRNA
B
HITS-CLIP WT
PAR-CLIP WT
Δpoly(N)
ΔRRMΔpoly(N)
PUF domain
R1 SNE
Puf3p
0.00
0.25
0.50
0.75
1.00
Fraction of peaks in each region
Location
3'UTR
5'UTR
CDS
Over start
Over stop

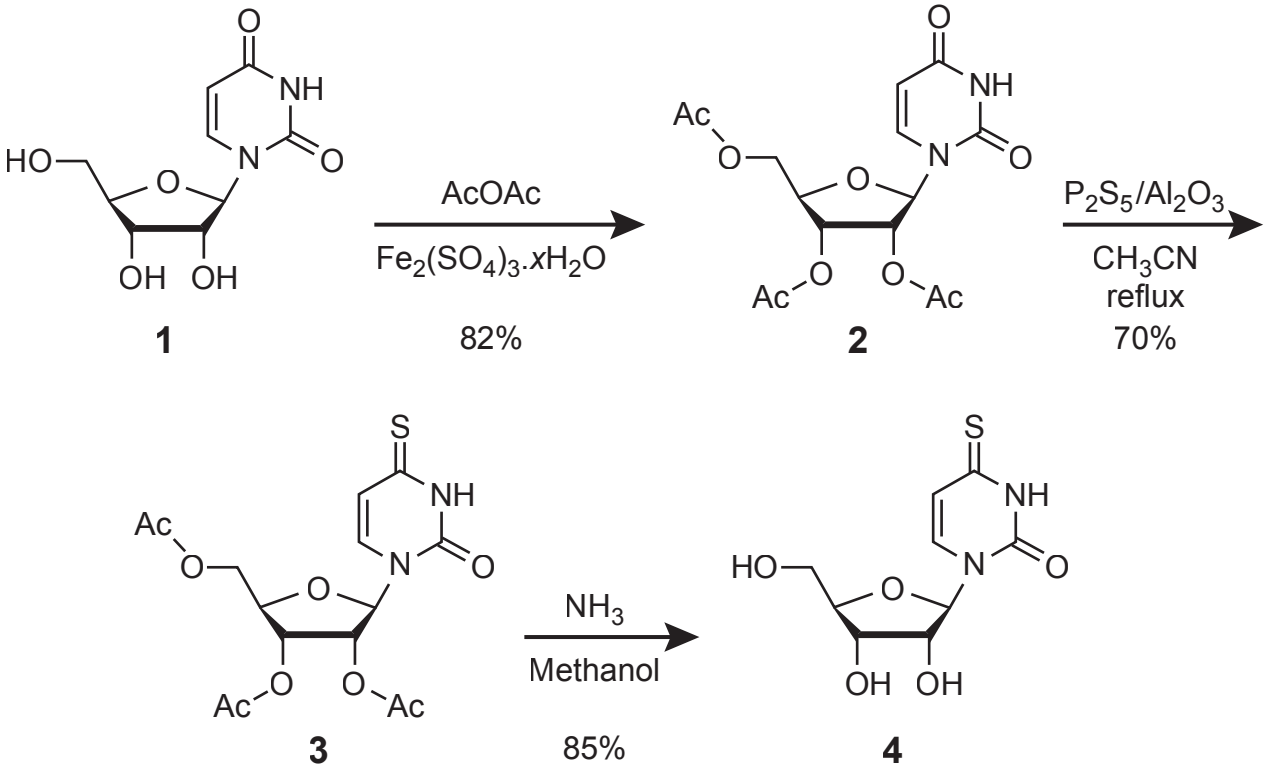
Supp. Figure 4



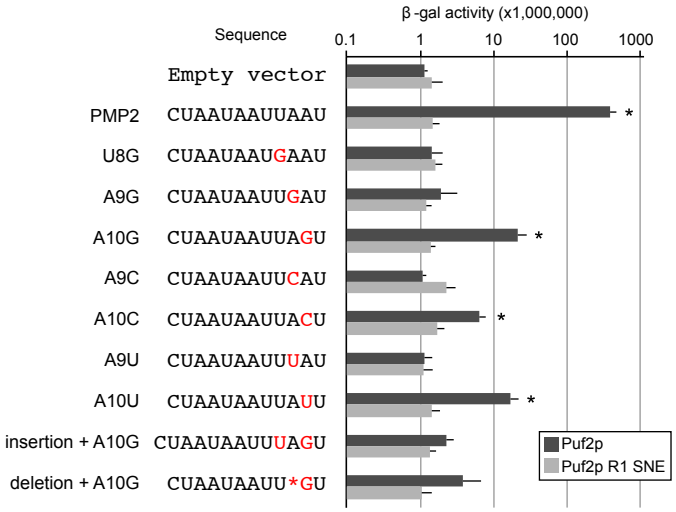
Supp. Figure 5

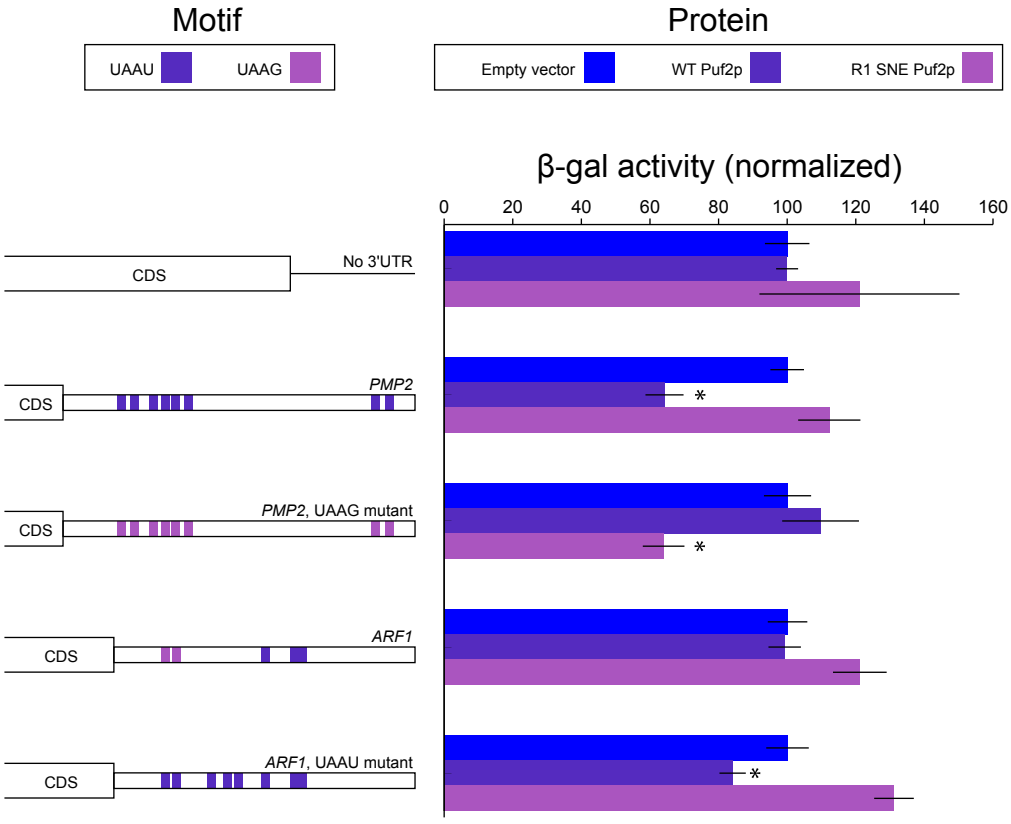



Fold enrichment HITS−CLIP/background
0
1
2
3
5'UTR
CDS
3'UTR
Position of dual UAAU
Fold enrichment PAR−CLIP/background
0
1
2
3
5'UTR
CDS
3'UTR
Position of dual UAAU
Fold enrichment R1 SNE/background
0
1
2
3
5'UTR
CDS
3'UTR
Position of dual UAAG

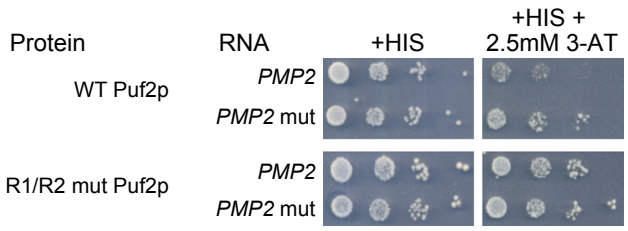

HO
O
NH
N
O
OH OH
1
AcOAc
Fe2(SO4)3.xH2O
82%
Ac
O
Ac O O Ac
2
P2S5/Al2O3
CH3CN
reflux
70%
S
Ac
3
NH3
Methanol
85%
4

Supp. Figure 8

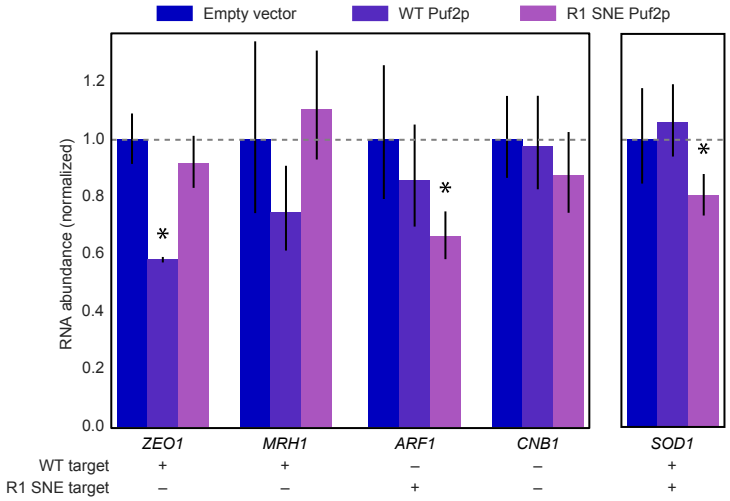


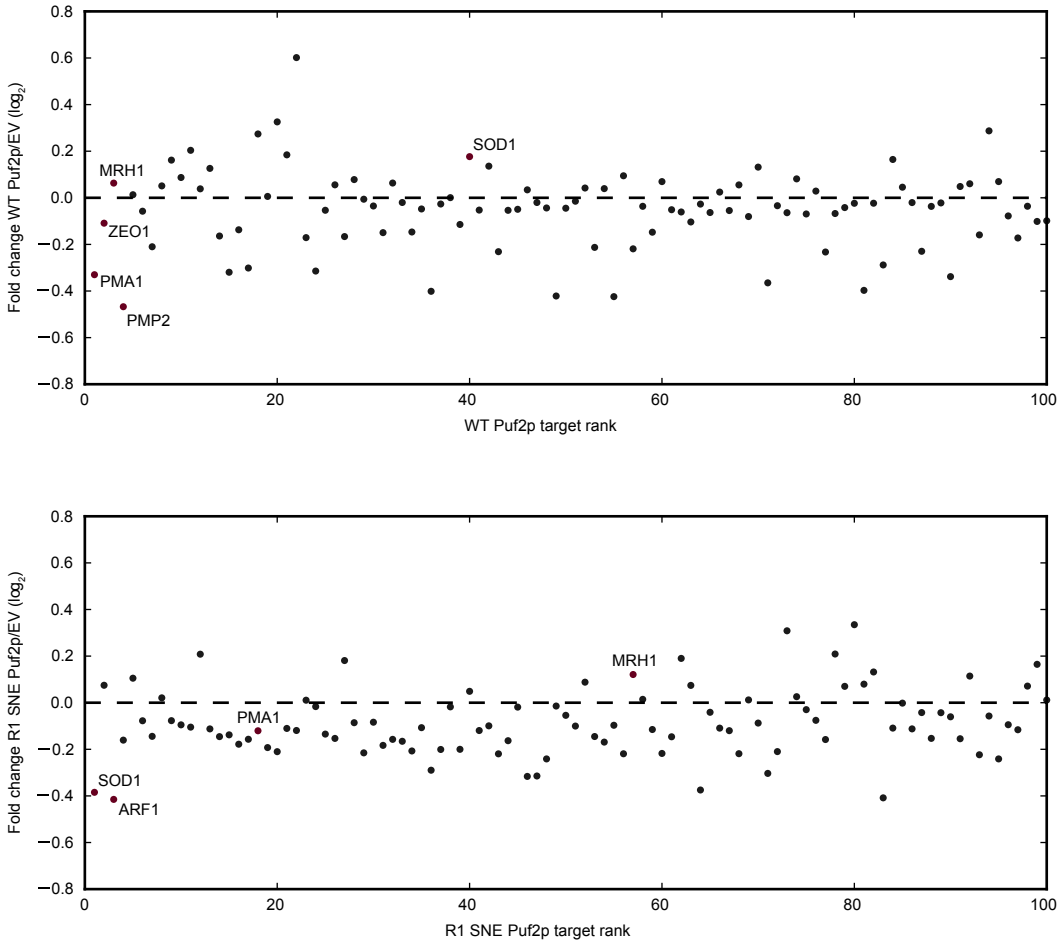

Motif
UAAU
UAAG
Protein
Empty vector
WT Puf2p
R1 SNE Puf2p
β-gal activity (normalized)
0
20
40
60
80
100
120
140
160
CDS
No 3'UTR
CDS
PMP2
CDS
PMP2, UAAG mutant
CDS
ARF1
CDS
ARF1, UAAU mutant
*
*
*

Supp. Figure 10



Supp. Figure 11




Fold change WT Puf2p/EV (log2)
WT Puf2p target rank
MRH1
SOD1
ZEO1
PMA1
PMP2
Fold change R1 SNE Puf2p/EV (log2)
R1 SNE Puf2p target rank
MRH1
PMA1
SOD1
ARF1