RESOURCE

nature structural & molecular biology

A protein-RNA specificity code enables targeted activation of an endogenous human transcript

Zachary T Campbell, Cary T Valley & Marvin Wickens

Programmable protein scaffolds that target DNA are invaluable tools for genome engineering and designer control of transcription. RNA manipulation provides broad new opportunities for control, including changes in translation. PUF proteins are an attractive platform for that purpose because they bind specific single-stranded RNA sequences by using short repeated modules, each contributing three amino acids that contact an RNA base. Here, we identified the specificities of natural and designed combinations of those three amino acids, using a large randomized RNA library. The resulting specificity code reveals the RNA binding preferences of natural proteins and enables the design of new specificities. Using the code and a translational activation domain, we designed a protein that targets endogenous cyclin B1 mRNA in human cells, increasing sensitivity to chemotherapeutic drugs. Our study provides a guide for rational design of engineered mRNA control, including translational stimulation.

Comprehensive analysis of specificity among modular DNA-binding proteins, including TAL effectors and zinc-finger transcription factors, has led to powerful tools for genome engineering and manipulation of transcription[1,2]. PUF (named for Pumilio and fem-3–binding factor) proteins provide an attractive scaffold with which to target RNAs instead and enable access to new events, including the processing, transport, localization, translation and decay of messenger and non-coding RNAs. PUF proteins regulate mRNA expression through physical association with short sequence elements (of 7–10 nt)[3–8]. A single RNA base is discriminated by a bundle of three α-helices, typically present in eight tandem 'PUF repeats' arranged in a semi-crescent (**Fig. 1a**)[9–12]. The identity of the targeted base is dictated by the combination of three amino acid residues, referred to here as a tripartite recognition motif (TRM) (**Fig. 1a**). TRMs contact RNA bases through a combination of edge-on and stacking interactions[13].

The modular nature of RNA recognition by PUFs suggests that their specificity might be systematically modified[14]. Indeed, mutant proteins with altered specificity have been created, as demonstrated first in human Pumilio[10]. However, many mutants behave unexpectedly, with either lost or broadened specificity *in vivo*[13,15]. Previous attempts to understand the consequences of PUF mutations analyzed binding to a single RNA sequence only at the targeted base. To identify TRMs that confer the greatest specificity, we used an unbiased high-throughput-sequencing approach that provides a proxy for biochemical affinity *in vitro*[4]. We determined the specificities of 25 natural and engineered PUF variants in binding reactions initially containing 4^{20} unique RNAs. The TRM specificities that we determined allow inference of the specificity of entire PUF proteins solely on the basis of primary sequence and predict known PUF binding sites *in vivo*. We used the code to design an artificial translation factor that specifically elevates translation of cyclin B1 mRNA in human cells.

RESULTS

Experimental design: selection of TRMs and scaffold

To determine which TRMs commonly occur in nature, we scored the prevalence of TRMs at each PUF repeat in 94 PUF proteins (**Fig. 1b** and Online Methods) and selected 14 of the most common TRMs at each repeat for further analysis. In parallel, we examined the specificity of three artificial TRMs previously reported to preferentially bind C nucleotides and eight new TRM combinations of our own design[16,17].

We chose the *Caenorhabditis elegans* PUF protein FBF-2 as a scaffold. Its specificity has previously been analyzed biochemically, structurally and through the use of compensatory mutations[4,9,15,18,19] (**Fig. 1a**). Importantly, we reasoned that because FBF-2 is less than 20% identical to human PUM1 and PUM2, it would have a low likelihood of eliciting regulation on its own in mammalian cells—an essential feature of a neutral tethering device. Furthermore, the potential for recognition of flanking bases via manipulation of a small pocket might provide opportunities to extend recognition sites[9,20,21].

The RNA-recognition patterns of TRMs

To analyze TRM specificities, we introduced mutations into the seventh repeat of FBF-2, which binds the +2 RNA base. We determined the specificity of 25 TRMs, using an unbiased approach, termed SEQRS, that combines *in vitro* selection, high-throughput sequencing of RNA and sequence-specificity landscapes (SSLs)[4] (**Fig. 2a**). In SEQRS, the number of reads for a specific sequence is correlated with its affinity measured *in vitro*[4]. In our experiments, we transcribed a DNA library encoding random 20-mer regions to generate a random pool of RNAs. We then incubated the pool, containing a sufficient quantity of RNA to cover all possible 20-mer sequences, with purified GST-tagged recombinant protein immobilized on magnetic resin to enable capture of

Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin, USA. Correspondence should be addressed to M.W. (wickens@biochem.wisc.edu).

Received 24 January; accepted 30 May; published online 29 June 2014; doi:10.1038/nsmb.2847

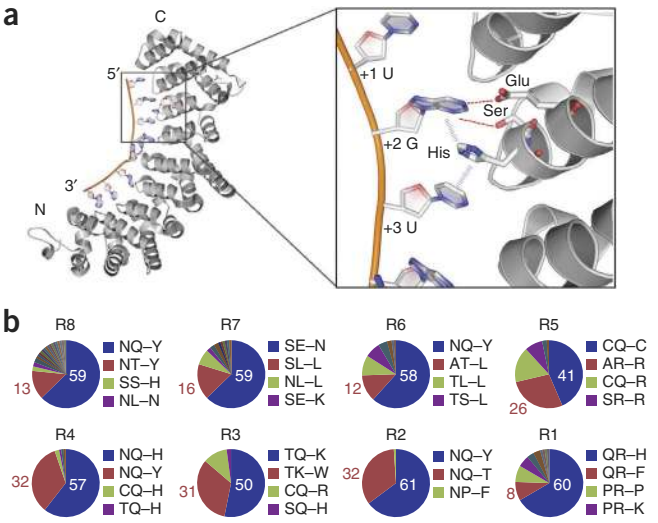


Figure 1 RNA recognition by the PUF proteins. (**a**) The structure of *C. elegans* FBF-2 bound to RNA[9]. RNA recognition is modular; each PUF repeat contributes an RNA-recognition helix. Inset, three amino acid residues (referred to as a tripartite recognition motif or TRM) form edge-on contacts (red lines) and stacking interactions (blue lines) with RNA bases. (Not all atoms are shown.) By convention, the two edge-on residues are given in the same order in which they are found in the primary sequence and are followed by a dash and the stacking residue; SE–H is shown here. (**b**) Abundance of natural TRMs inferred from sequence alignment. Pie charts represent TRM enrichment as a function of PUF repeat (R8–R1) on the basis of 94 proteins.

RNA–protein complexes. After repeated washing, we thermally eluted bound RNAs and converted them into double-stranded DNA with reverse transcription, using a primer complementary to the constant region, and subsequent PCR amplification, using a primer set that reintroduces the T7 promoter. We repeated this enrichment procedure, analogous to SELEX, for five cycles before multiplexed deep sequencing[22,23].

We systematically quantified the specificity of each TRM mutant for all possible 10-mer sequences. To identify the similarities in binding preferences for high-affinity sites, we analyzed the data with hierarchical clustering (**Fig. 2b**) and used 230 preferentially enriched unique sequences for each individual TRM to identify binding similarities. We used the heat map (**Fig. 2b**) to define three clusters specific for U, A or G at position +2 (clusters A, B and C, respectively). Changing the boundaries of the clusters increased degeneracy at position +2. To identify variations between TRMs in the same cluster, we generated sequence logos corresponding to TRMs present in cluster A. All TRMs in cluster A preferentially bound U at position +2; however, TRMs varied considerably in their degree of nontarget enrichment, as shown by degeneracy in the sequence logos and revealed comprehensively in SSLs, which represent binding data as a series of concentric rings, with outward rings containing sequences with increasing numbers of mismatches (**Fig. 2c**)[4,24]. Comparison of the NQ–T and CR–Y TRMs revealed substantial peaks in outward rings of the landscape with CR–Y, which are much reduced in NQ–T; NR–Y is intermediate. These data demonstrate the broadened specificity of the CR–Y and NR–Y relative to NQ–T.

Specificity at the targeted base and elsewhere

To directly compare the specificity of each TRM, we calculated the enrichment of all four bases at RNA position +2 by searching the data for all permutations of the FBF-2 binding element (**Fig. 2d**)[4]. TRMs with much broadened specificity, such as SL–H and TQ–R, were apparent. Relatively modest changes in a single edge-on residue, such as TR–Y to AR–Y, resulted in altered specificity. Similarly, a non-conservative change in the stacking residue, such as TQ–R and TQ–W, altered specificity. To rank the precision of TRMs for preferred bases, we calculated specificity-coefficient values (**Supplementary Fig. 1a**). These values incorporate enrichment at the targeted site as compared to the flanking nontarget bases. G- and U-binding TRMs were more selective than A-binding TRMs (**Supplementary Fig. 1b**). The specificities of natural TRMs (0.37) were slightly greater on average than those of synthetic TRMs (0.24, **Supplementary Fig. 1c**).

De novo–designed TRMs provide a means both to diversify and to improve RNA specificity, and they reveal complex interactions among TRM residues. TRMs CQ–F and CE–Y were more specific than any natural TRM for A (**Supplementary Fig. 1a**,**c**). C and Q as edge-on residues appear to be a common feature among both natural and synthetic A-specific TRMs. However, stacking residues can determine whether certain edge-on pairs (such as C and E) specify recognition of A or G. Taken together, our TRM design data suggest that although the stacking residue does not make hydrogen-bonding interactions to the base, cation-π and van der Waals contacts have a profound influence on specificity. We conclude that *de novo* design of TRM variants provides a means to discover binding arrangements that are more specific than those in naturally occurring TRMs.

In some instances, new bases were accommodated as a result of relaxed specificity. For example, although we did not observe switches to C specificity, several TRMs tolerated C, yielding more than 5% of reads with that base at +2 (**Supplementary Fig. 2a**). However, C enrichment paralleled that of the other two nontargeted bases, suggesting broadened specificity (**Supplementary Fig. 2b**). The identities of stacking residues affected specificity at adjacent bases differentially (**Supplementary Fig. 3a**,**b**). For example, asparagine broadened specificity at position +3 but not at +1, whereas phenylalanine behaved in an opposite fashion. Finally, basic and polar uncharged residues in edge-on positions also appeared to broaden specificity immediately upstream of the targeted site, at position +1 (**Supplementary Fig. 3c**,**d**).

TRM substitutions affected bases flanking the targeted nucleotide (**Fig. 2c**). To quantify these effects, we calculated enrichment values for flanking bases (**Supplementary Fig. 4**). These effects can be substantial. Two of the TRMs (TQ–R and SQ–R) displayed deviations of >40% from wild-type sequence preferences at flanking sites. Many TRMs increased accommodation of A binding by repeat 8, one nucleotide away from the targeted base (**Supplementary Fig. 4c**).

Prediction and the distribution of specificity in nature

The TRM specificity code provides RNA binding preferences for the majority of naturally occurring TRMs (**Fig. 1b**). We used these data to predict the specificities of two PUF proteins from the slime mold *Dictyostelium discoideum* (**Supplementary Fig. 5**) and compared the predicted consensus elements to experimentally determined motifs from SEQRS (**Fig. 3a**). The *in silico* predictions correlated well. For example, a single cysteine-to-threonine mutation in repeat 3 of PufA versus PufB altered specificity from A to U, as predicted from the code. An 'extra' nucleotide is present at position 5 of the DdPufA site. This is probably because of base flipping, in which a base is extruded from the binding surface of the protein[9,19,25]. Sites of base flipping are not yet predictable computationally (described below).

The TRM code enables identification of naturally occurring RNA-binding sites. Using only the TRM data, we predicted the specificity

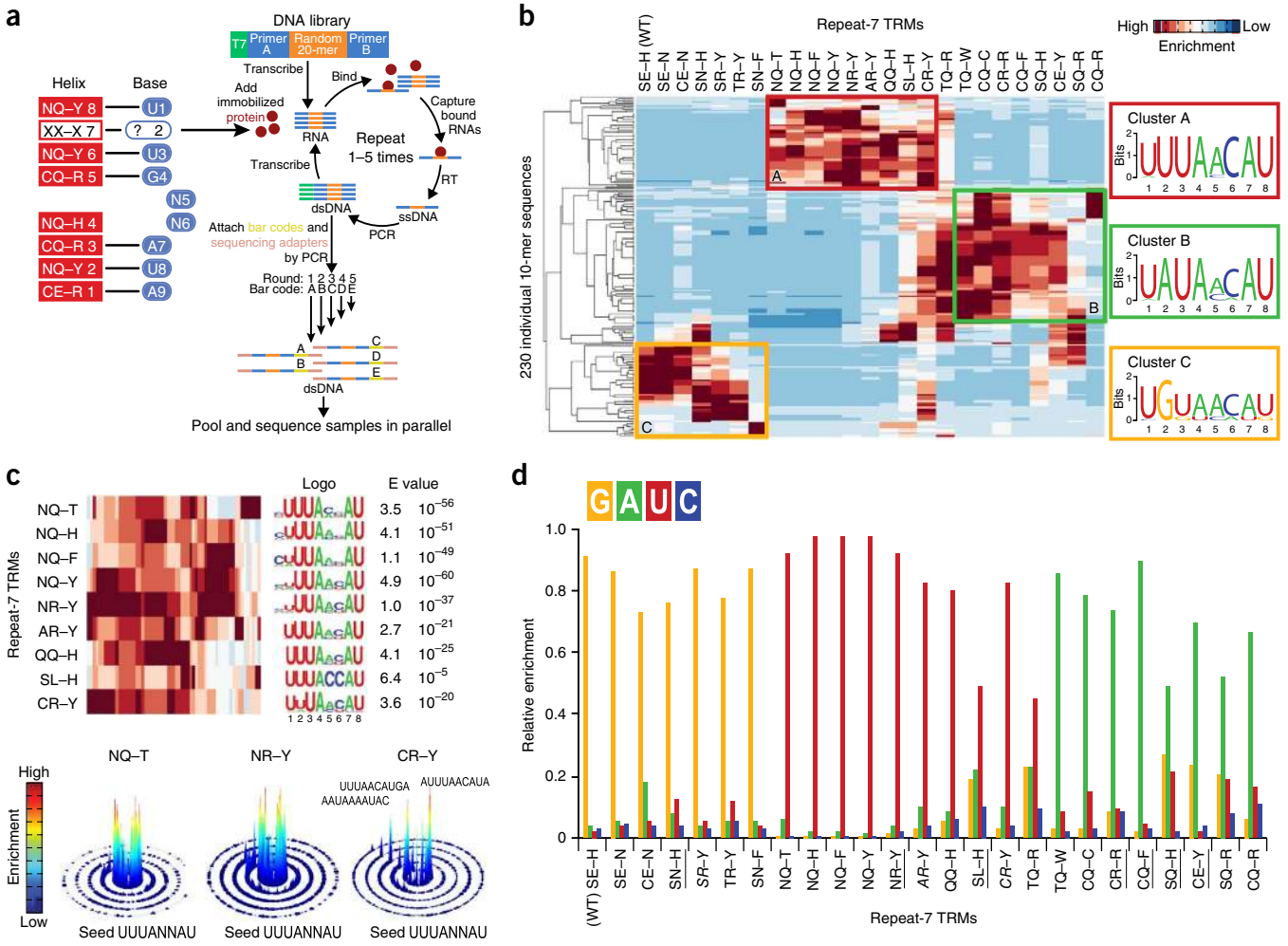


Figure 2 A quantitative TRM recognition code. (**a**) Experimental overview. TRM substitutions are introduced in PUF repeat 7 of FBF-2. The predicted site of variation is base +2 in the RNA sequence. TRM mutants are analyzed with the SEQRS technique (main text). RT, reverse transcription; ssDNA and dsDNA, single-stranded and double-stranded DNA, respectively. (**b**) Hierarchical clustering revealing three classes of TRM binding specificity. Left, clustering of similar binding profiles for each mutant (*x* axis) according to enrichment values for highly enriched 10-mer sequences identified for each TRM (*y* axis). For each TRM, data were normalized to the maximum enrichment value. WT, wild type. Right, representative motifs generated from three clusters identified empirically. (**c**) Top, sequence logos for members of cluster A, showing a common specificity consistent with the results from clustering at left. E value, expectation value. Bottom, SSLs, as developed by the Ansari laboratory[24], for three representative TRMs, showing nonequivalent differences in overall specificity. The innermost ring contains sequences perfectly matched to a given seed motif, and subsequent rings contain sequences with increasing numbers of mismatches from the seed motif. The z axis (height) corresponds to the number of reads of a particular sequence, normalized to the starting library. The color scheme indicates a given threshold of enrichment for a particular sequence. (**d**) Enrichment for G (yellow), U (red), A (green) and C (blue) at position +2 of the PUF binding element. TRMs previously described as preferential C-binders, SR–Y, AR–Y and CR–Y, are italicized[16,17]. The remaining synthetic combinations are underlined.

of human PUM-2 (**Fig. 3b**), allowing degeneracy in TRMs that had exhibited low specificity. The TRM-derived model correctly identified genuine sites of occupancy *in vivo* with similar levels of sensitivity to experimentally derived consensus binding elements (Wilcoxon-Mann-Whitney rank-sum test $P < 0.01$). We conclude that TRM data appear to provide a useful tool for the prediction of specificity.

TRM repeats and RNA bases have been subjected to extensive mutagenesis in prior work for Puf3p, FBF-2 and Puf4p[19]. We compared data from this study to TRM specificities to determine how the specificity of TRMs at a given repeat compares to the average tolerance of RNA or TRM substitutions (**Fig. 3c**). We calculated the average specificity of natural TRMs on a repeat-by-repeat basis and depicted them as a plot of specificity coefficients as a function of repeat. Among naturally occurring PUF proteins, C-terminal repeats contain TRMs with the highest specificity, consistently with the high conservation of both RNA and protein identities in this region (**Fig. 3c**)[19]. We propose that this provides a starting point for the evolution of new target specificity through variation of specificity in N-terminal repeats.

Design of new specificity

To determine how broadly the TRM code applies, we examined mutations in different repeats and scaffolds (**Fig. 4**). We prepared ten RNA variants that contained one to six mutations in the RNA sequence that binds human PUM2. We engineered protein variants designed to bind these RNAs in a PUM2 scaffold, which had not been used to derive the TRM code. Wild-type PUM2 protein did not bind detectably to any of the mutant RNAs in yeast three-hybrid assays, though

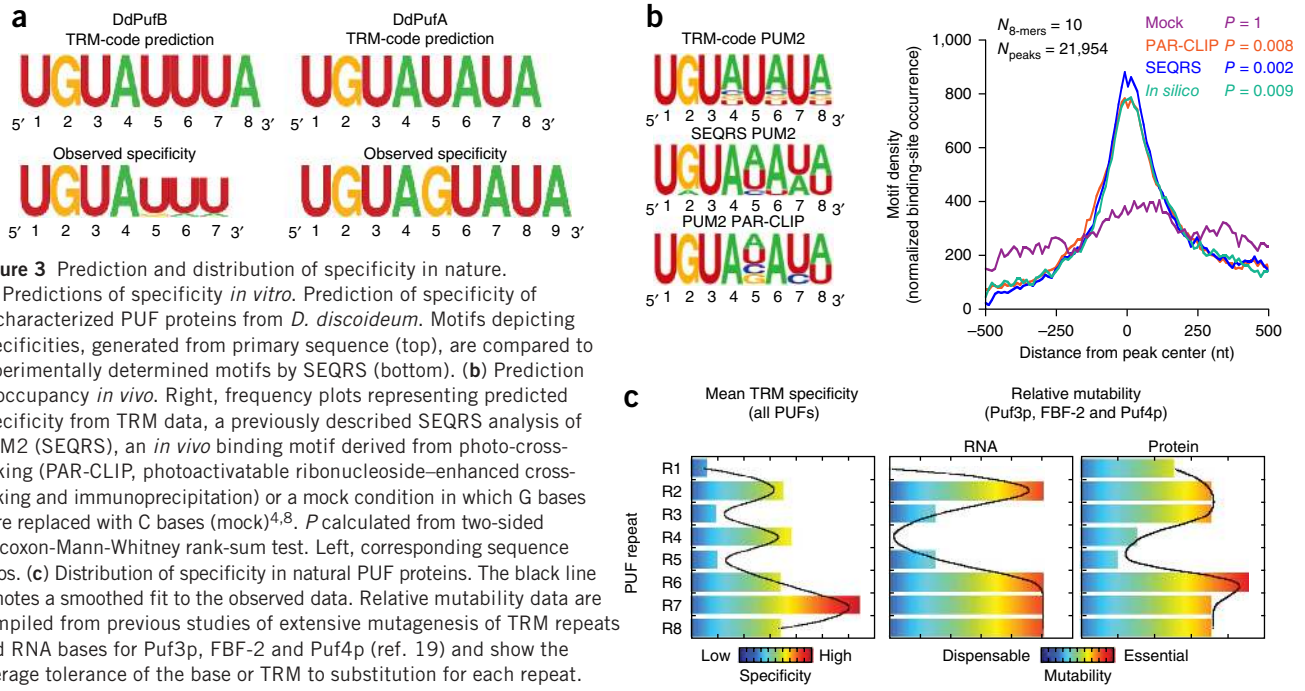


Figure 3 Prediction and distribution of specificity in nature. (**a**) Predictions of specificity *in vitro*. Prediction of specificity of uncharacterized PUF proteins from *D. discoideum*. Motifs depicting specificities, generated from primary sequence (top), are compared to experimentally determined motifs by SEQRS (bottom). (**b**) Prediction of occupancy *in vivo*. Right, frequency plots representing predicted specificity from TRM data, a previously described SEQRS analysis of PUM2 (SEQRS), an *in vivo* binding motif derived from photo-cross-linking (PAR-CLIP, photoactivatable ribonucleoside–enhanced cross-linking and immunoprecipitation) or a mock condition in which G bases were replaced with C bases (mock)[4,8]. *P* calculated from two-sided Wilcoxon-Mann-Whitney rank-sum test. Left, corresponding sequence logos. (**c**) Distribution of specificity in natural PUF proteins. The black line denotes a smoothed fit to the observed data. Relative mutability data are compiled from previous studies of extensive mutagenesis of TRM repeats and RNA bases for Puf3p, FBF-2 and Puf4p (ref. 19) and show the average tolerance of the base or TRM to substitution for each repeat.

it did bind its wild-type site (**Supplementary Fig. 6**). We tested the engineered PUM2 proteins, designed with the TRM code, against the same set of RNAs. We first introduced the best U-specific TRM (NQ–H) into repeat 7 (**Fig. 4a**). This mutant protein bound the U-containing site and not the wild-type sequence. We then designed a series of additional substitutions, using the most specific TRM for G recognition (SE–H), maintaining the U-specific repeat 7. The resulting nine mutant proteins bound their new cognate elements and not the wild-type sequence. We observed this result with as many as six mutations in the RNA elements. Similarly, in the FBF-2 scaffold, we tested binding of two double-TRM-mutant proteins and one single-TRM-mutant protein, each designed with the TRM code (**Fig. 4b**). These bound their targeted RNAs and not the wild-type site (**Fig. 4b**). We conclude that the TRM data are applicable to different scaffolds and repeats and that they enable tailored recognition of three of the four RNA bases.

To determine whether addition of a single altered TRM enhanced the specificity of multiply mutated proteins, we assayed binding of six pairs of nearly identical proteins (**Fig. 4c**). The pairs of proteins differed in a single repeat, possessing either the wild-type or the altered TRM. We assayed binding of each pair to an RNA containing the mutated nucleotide corresponding to the altered repeat being examined. In each case, the mutant TRM in a single repeat enhanced binding to the cognate site. The magnitude of the enhancement differed among repeats, consistently with our prior work showing that individual repeats vary in their contributions to overall affinity[19]. We conclude that the use of the TRM code to introduce sets of mutations in multiple repeats yields additive effects on targeting.

To explore the utility of TRM data for manipulation of mRNA expression, we engineered FBF-2 to bind a specific RNA sequence in the 3′ untranslated region (3′ UTR) of human cyclin B1 mRNA.

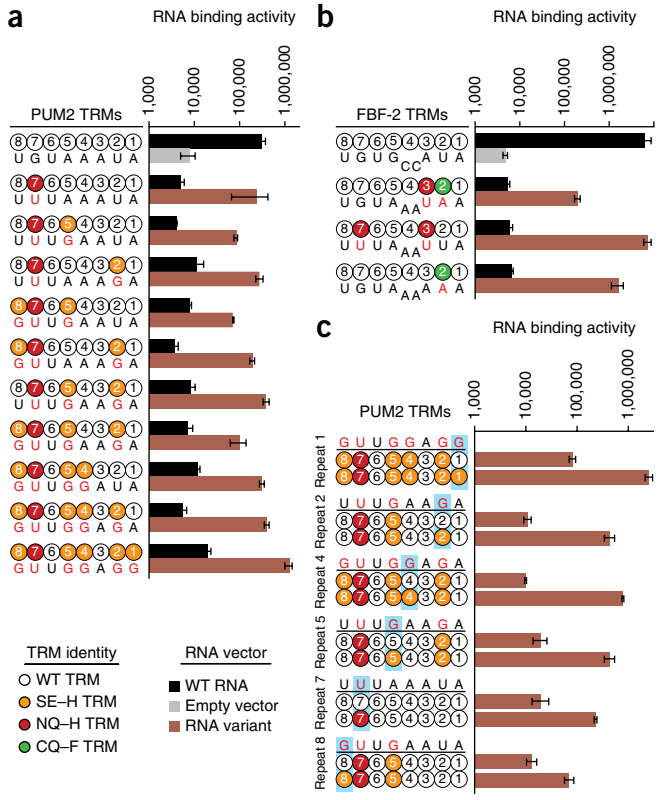


Figure 4 Modifications of PUF scaffolds according to the TRM code. TRM variants are denoted as colored circles and RNAs as colors in bar charts. RNA sequences are provided for variant sites. Red nucleotides indicate sites that differ from the wild-type sequence. Binding-activity measurements (relative β-gal) were conducted in the yeast-three hybrid system and normalized to cell density[15,18]. (**a**) Replacement of TRMs in the PUM2 scaffold to yield protein mutants with new specificity. As shown, most specificity mutants possessed binding activities comparable to that of the wild-type protein and RNA-binding element. Error bars, s.d. ($n = 3$ independent colonies). β-gal, β-galactosidase. (**b**) Mutations in FBF-2 yielding altered specificity. Error bars, s.d. ($n = 3$ independent colonies). (**c**) Additivity of specificity mutations in PUM2. Sites of comparison between TRMs are highlighted in blue. Error bars, s.d. ($n = 3$ independent colonies).

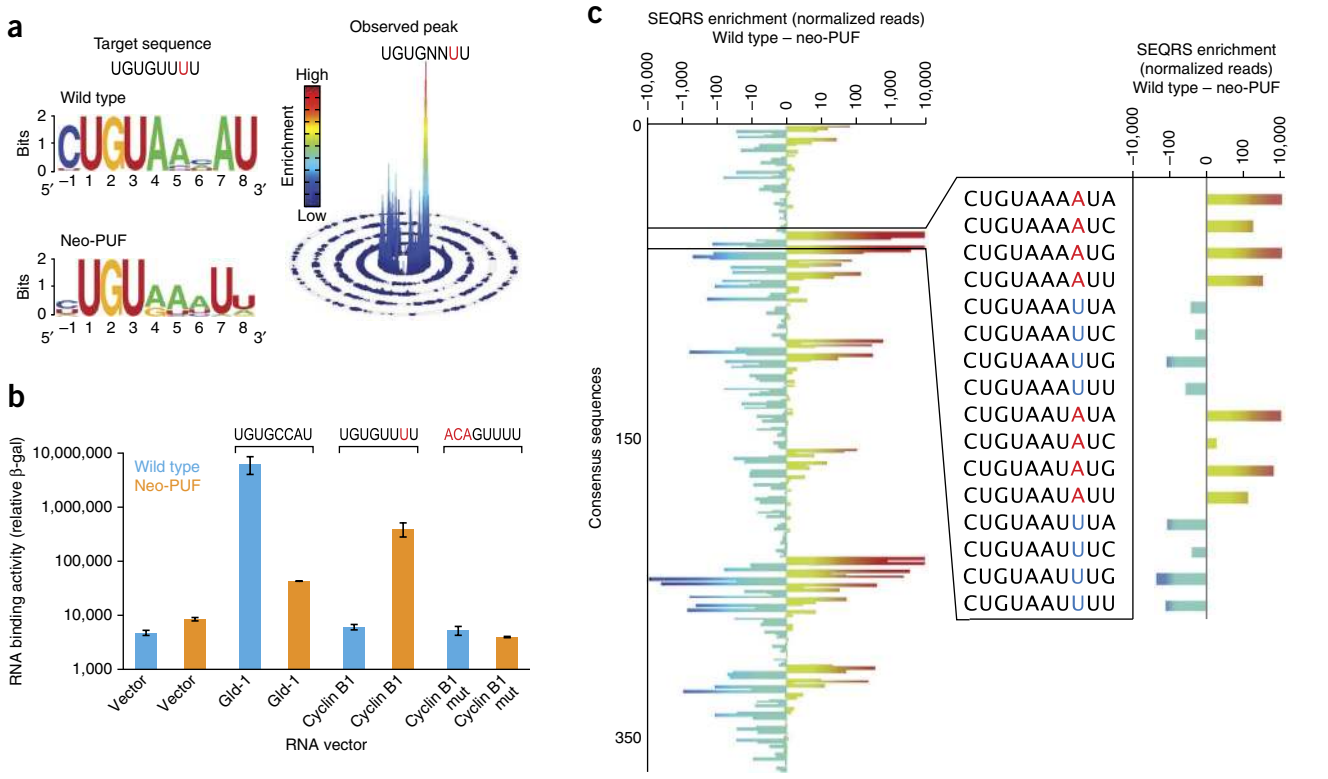


Figure 5 Engineered specificity of PUF proteins. (**a**) Left, sequence motifs of wild-type and redesigned (neo-PUF) proteins. The targeted recognition site possesses a single substitution at position seven (red) of the binding element. Right, SSL, with the major peak of enriched sequences indicated in the zero-mismatch inner ring. (**b**) Analysis of RNA binding activity in a yeast three-hybrid assay. RNA binding measurements for wild-type (blue) and neo-PUF (orange) for an empty vector, the positive-control *gld-1* RNA element and the cyclin B1–targeting element are shown. Error bars, s.d. ($n = 3$ independent colonies). Mut, mutant. (**c**) Binding enrichment values for the neo-PUF and the wild-type proteins, calculated for a consensus sequence tolerant of either A or U at position 7, HUGURWWWU, and subtracted from each other (wild type – neo-PUF). On the vertical axis are arrayed 384 sequences—a subset of the 4^{10} possible 10-mers analyzed computationally—arranged in logical order by sequence. A subset of the sequences shown on the right possess sequences altered at positions +7 and +9. The plots indicate the degree of enrichment for either the wild-type or neo-PUF proteins. Values are shaded as in **a**, with positive enrichments shaded green to red and negative enrichments shaded light blue to dark blue.

Cyclin B1 is a critical regulator of the cell cycle responsible for entry into mitosis and exit from G2 (refs. 26,27). We altered repeat 3 of FBF-2 so that it should bind a sequence in the 3′ UTR of cyclin B1 mRNA, in which position 7 is nonconsensus (UGUGUUUU). We refer to this protein as a 'neo-PUF' (**Fig. 5a**). Both SEQRS and yeast three-hybrid assays revealed that the neo-PUF now bound the desired element, with PUF repeat 3 binding U rather than A (**Fig. 5a**,**b**). To globally analyze differences in specificity between the wild type and neo-PUF, we subtracted the enrichment value of sequences obtained with the neo-PUF from those obtained with the wild-type protein; thus negative values indicate preferential binding to the neo-PUF, and positive values indicate preferential binding to the wild-type protein (**Fig. 5c**). Careful inspection of a subset of these sequences provides an example, using a region in which only the identities of position 7 and 9 vary. We found that +7 U sequences were enriched by the neo-PUF, whereas the wild-type protein enriched +7 A, regardless of the larger sequence context. Enrichment oscillated along the axis, indicating that changes at position +7 (and in the highlighted case, not at +9) dictate the enrichment of a given sequence. We conclude that the TRM data are able to accurately predict modified specificity at alternate PUF repeats.

Targeted activation of an endogenous transcript

Few tools are available to increase translation of specific endogenous mRNAs, though targeted negative control is commonplace[28]. We used the neo-PUF to stimulate translation of endogenous cyclin B1 mRNA. In stable cancer cell lines (U2OS), the neo-PUF was expressed as well as the wild-type protein (**Supplementary Fig. 7a**). The neo-PUF, but not the wild-type protein, bound endogenous cyclin B1 mRNA, as judged by RNA immunoprecipitation (RIP) followed by reverse-transcription PCR (**Supplementary Fig. 7b**). Similarly, an RNA binding–defective form of the PUF, termed RNA^{DEF}, in which H326 was replaced by alanine, did not bind cyclin B1 mRNA. This control indicates that the RNA binding activity of the PUF domain is essential for association with the cyclin B1 transcript.

To enhance translation of endogenous cyclin B1, we fused a 20-kDa segment of yeast poly(A)-binding protein (PAB) to the neo-PUF protein. This domain stimulates translation of a reporter in *Xenopus laevis* oocytes[29]. We refer to this chimera as a 'neo-activator'. The neo-activator increased cyclin B1 protein abundance by approximately 400 percent; neither the RNA binding–defective form fused to PAB (termed RNA^{DEF}-PAB) nor vector alone did so (**Supplementary Fig. 8**). The levels of protein expression of the neo-activator and the RNA binding–defective form were comparable (**Supplementary Fig. 7a**). The neo-PUF without the PAB moiety had little effect on cyclin B1 levels, thus demonstrating that the PUF scaffold was functionally inert (**Supplementary Fig. 8**). We confirmed increased cyclin B1 protein abundance by immunofluorescence spectroscopy, which indicated that the fraction of cyclin B1–positive cells increased by approximately 500 percent (**Fig. 6**).

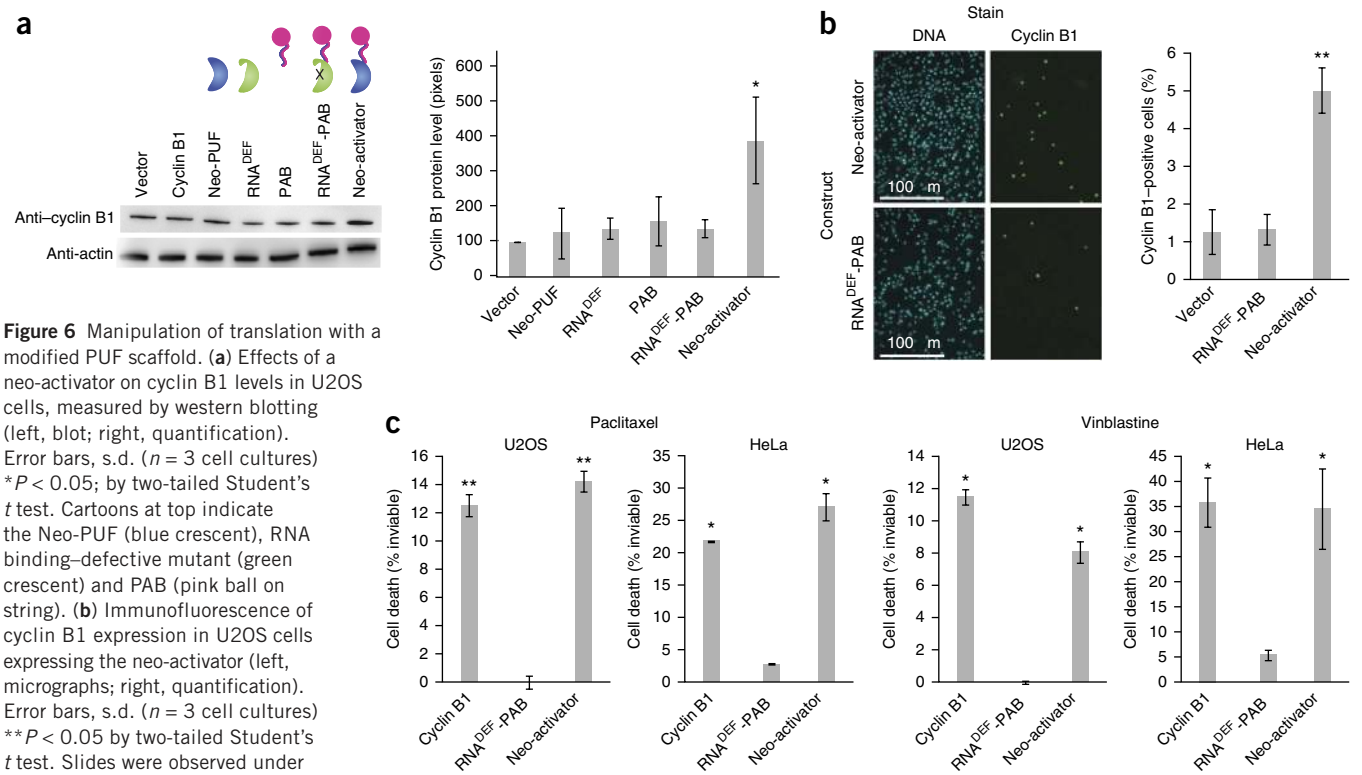


Figure 6 Manipulation of translation with a modified PUF scaffold. (**a**) Effects of a neo-activator on cyclin B1 levels in U2OS cells, measured by western blotting (left, blot; right, quantification). Error bars, s.d. ($n = 3$ cell cultures) $^*P < 0.05$; by two-tailed Student's t test. Cartoons at top indicate the Neo-PUF (blue crescent), RNA binding–defective mutant (green crescent) and PAB (pink ball on string). (**b**) Immunofluorescence of cyclin B1 expression in U2OS cells expressing the neo-activator (left, micrographs; right, quantification). Error bars, s.d. ($n = 3$ cell cultures) $^{**}P < 0.05$ by two-tailed Student's t test. Slides were observed under the same microscope with identical parameters. (**c**) Viability assays. Viability was quantified 24 h after treatment. Normalized cell death is shown for two M phase–targeting drugs: paclitaxel (left) and vinblastine (right)[27]. Cyclin B1 was overexpressed as a positive control. Error bars, s.d. ($n = 3$ cell cultures) $^*P < 0.05$; ** $P < 0.05$ by two-tailed Student's t test. Uncropped images of blots are shown in **Supplementary Figure 8**.

Cyclin B1 overexpression renders specific cancer cell lines hypersensitive to antimitotic chemotherapeutic drugs[26,27]. To determine whether the neo-activator enhanced sensitivity to such drugs, we performed viability assays after treatment with paclitaxel or vinblastine (**Fig. 6c**). Expression of the neo-activator reduced viability after exposure to either drug, in both U2OS and HeLa cells. The effect was similar to that achieved by transfection of the cyclin B1 gene. Similarly, the neo-activator produced anticipated outcomes of cyclin B1 overexpression including increased growth rate and delayed reentry into the cell cycle after arrest (**Supplementary Fig. 7c,d**)[27]. These data demonstrate that tailored post-transcriptional gene activation can be engineered in a predictable manner to generate a desired cellular phenotype.

DISCUSSION

Analysis of the preferences for binding of multiple natural and engineered proteins to a large population of RNAs reveals five principles in PUF-RNA interactions. First, the binding preferences of previously uncharacterized proteins can be deduced through the code. Our analysis of PUFs from an organism evolutionarily distant from humans suggests high conservation of PUF binding motifs. Second, designed TRMs can be as specific as their natural counterparts. The collection of new TRMs that we report effectively doubles the repertoire of known TRM combinations. Third, although the code can be used with high confidence, complete prediction of binding sites is in some cases complicated by base-flipping, in which a base is not recognized but instead is solvent exposed and serves as a 'spacer'. Similar effects in zinc-finger proteins are well documented[30,31]. Experimental determination of specificity thus is required. Fourth, specificity of C-terminal repeats on average is greater than that of N-terminal repeats. Similarly, mutations in C-terminal repeats or the bases that they recognize affect binding more profoundly than those elsewhere[19]. Finally, certain TRMs broaden RNA base specificity and may be mistaken for a switch, thus underscoring the importance of unbiased validation.

What are the limits to manipulation of RNA recognition? Because our selection experiments focused on a single repeat required for RNA binding, we performed additional candidate tests, which suggest that the experiments at repeat 7 are generally applicable at different sites and in a different scaffold (**Fig. 4**). However, structural analyses have shown that the curvature of the backbone differs substantially between members of the PUF family[9,10,14,20,32]. Subtle differences in the packing within and between PUF repeats may alter the geometry of base recognition. Selective pressure on the backbone geometry of the scaffold may explain the observation that the wild-type TRM at repeat 7 has the greatest specificity of the combinations reported here. Local differences in curvature can cause base flipping and can complicate predictions of specificity in previously unexamined proteins, as we suggest to be the case with DdPufA (**Fig. 3a**). In the future, tailored RNA recognition via the TRM code may be enhanced through the design or selection of an idealized backbone. It is striking that changes in specificity can be achieved in repeats 6, 7 and 8 even though these repeats almost always recognize UGU in natural binding sites. Successful manipulations of specificity in this region suggest that the redesigned proteins are able to escape a selective pressure to which natural PUFs are subject.

The RNA-recognition code presented here yields a quantitative assessment of TRM specificity and enhances the precision and ease of targeted RNA control. We have identified TRM combinations that are optimal, taking into account the specificity for the targeted base and minimizing effects on neighboring bases. The code enables construction of proteins that stimulate translation of a specific mRNA in

human cells. The use of proteins targeted to short recognition sites imparts recognition of target mRNAs but probably also results in unintended binding events. The use of elongated PUF scaffolds may provide a means to reduce off-target effects[17]. Neo-PUF proteins, designed on the basis of the information presented here, should enable tailored control of the many cytoplasmic events in the life of an mRNA, including its translation, stability and localization.

METHODS

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

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

ACKNOWLEDGMENTS

We thank L. Vanderploeg for help with figures and M. Preston and S. Waghray for careful editing. This work was supported by US National Institutes of Health grant GM050942 to M.W. U2OS and HeLa cells were kind gifts from S. Miyamoto and R. Raines (both at University of Wisconsin–Madison), respectively.

AUTHOR CONTRIBUTIONS

Z.T.C. and C.T.V. performed the experiments. Z.T.C., M.W. and C.T.V. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Proudfoot, C., McPherson, A.L., Kolb, A.F. & Stark, W.M. Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS ONE* **6**, e19537 (2011).
2. Wood, A.J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307 (2011).
3. 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 (2002).
4. Campbell, Z.T. *et al.* Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Reports* **1**, 570–581 (2012).
5. 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 (2008).
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. USA* **107**, 3936–3941 (2010).
8. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
9. Wang, Y., Opperman, L., Wickens, M. & Hall, T.M. Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc. Natl. Acad. Sci. USA* **106**, 20186–20191 (2009).
10. Wang, X., McLachlan, J., Zamore, P.D. & Hall, T.M. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501–512 (2002).
11. Wang, X., Zamore, P.D. & Hall, T.M. Crystal structure of a Pumilio homology domain. *Mol. Cell* **7**, 855–865 (2001).
12. Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M. & Hall, T.M. A 5′ cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc. Natl. Acad. Sci. USA* **106**, 20192–20197 (2009).
13. Koh, Y.Y. *et al.* Stacking interactions in PUF-RNA complexes. *RNA* **17**, 718–727 (2011).
14. Wang, Y., Wang, Z. & Tanaka Hall, T.M. Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. *FEBS J.* **280**, 3755–3767 (2013).
15. Opperman, L., Hook, B., DeFino, M., Bernstein, D.S. & Wickens, M. A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. *Nat. Struct. Mol. Biol.* **12**, 945–951 (2005).
16. Dong, S. *et al.* Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *J. Biol. Chem.* **286**, 26732–26742 (2011).
17. Filipovska, A., Razif, M.F., Nygard, K.K. & Rackham, O. A universal code for RNA recognition by PUF proteins. *Nat. Chem. Biol.* **7**, 425–427 (2011).
18. Bernstein, D., Hook, B., Hajarnavis, A., Opperman, L. & Wickens, M. Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. *RNA* **11**, 447–458 (2005).
19. 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. USA* **109**, 6054–6059 (2012).
20. Qiu, C. *et al.* Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. *J. Biol. Chem.* **287**, 6949–6957 (2012).
21. Stumpf, C.R., Kimble, J. & Wickens, M. A *Caenorhabditis elegans* PUF protein family with distinct RNA binding specificity. *RNA* **14**, 1550–1557 (2008).
22. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
23. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
24. Carlson, C.D. *et al.* Specificity landscapes of DNA binding molecules elucidate biological function. *Proc. Natl. Acad. Sci. USA* **107**, 4544–4549 (2010).
25. 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 (2008).
26. Brito, D.A. & Rieder, C.L. The ability to survive mitosis in the presence of microtubule poisons differs significantly between human nontransformed (RPE-1) and cancer (U2OS, HeLa) cells. *Cell Motil. Cytoskeleton* **66**, 437–447 (2009).
27. Russo, A.J. *et al.* E2F-1 overexpression in U2OS cells increases cyclin B1 levels and cdc2 kinase activity and sensitizes cells to antimitotic agents. *Cancer Res.* **66**, 7253–7260 (2006).
28. Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515–524 (2006).
29. Gray, N.K., Collar, J.M., Dickson, K.S. & Wickens, M. Multiple portions of poly(A)-binding protein stimulate translation *in vivo*. *EMBO J.* **19**, 4723–4733 (2000).
30. Lam, K.N., van Bakel, H., Cote, A.G., van der Ven, A. & Hughes, T.R. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.* **39**, 4680–4690 (2011).
31. Enuameh, M.S. *et al.* Global analysis of *Drosophila* Cys_2-His_2 zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. *Genome Res.* **23**, 928–940 (2013).
32. Lu, G. & Hall, T.M. Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure* **19**, 361–367 (2011).

ONLINE METHODS

TRM alignments. Six *C. elegans*, one human and 89 fungal PUF proteins were used to generate a library of eukaryotic TRM combinations. The disproportionate use of fungal PUFs was important, given the substantial divergence in the fungal lineage. Many of the PUF proteins are direct homologs of Pumilio and as a result are predisposed toward a common set of TRM combinations. All of the PUF proteins were detected by homology to *Schizosaccharomyces cerevisiae* PUFs. TRM combinations were inferred on the basis of manual comparisons of multiple sequence alignments containing *S. cerevisiae* Puf4p and Puf3p, whose structures have been determined experimentally[12,25].

Mutagenesis and protein purification. The GST-fusion constructs used in the present study include *C. elegans* FBF-2(121–632) and the *D. discoideum* proteins PufA(450–792) and PufB(690–1036)[18]. TRM mutants were generated by site-directed mutagenesis as described[33,34]. Recombinant fusion proteins were purified as described with high-capacity magnetic GST-agarose beads (Sigma-Aldrich). Protein aliquots were stored in SEQRS buffer (50 mM HEPES, pH 7.4, 2 mM EDTA, 150 mM NaCl, 0.1% NP-40, and 1 mM DTT) containing 20% glycerol before flash freezing and storage at −80 °C.

***In vitro* selection.** The SEQRS protocol was conducted as described with minor modifications[4]. The initial library was transcribed from 1 µg of input dsDNA with the AmpliScribe T7-Flash Transcription Kit (Epicentre). The reaction was treated with RNase-free DNase to remove residual DNA and was purified with the GeneJET RNA Purification Kit (Fermentas). 150 ng of the purified RNA library was added to RNA-binding proteins containing ~50–100 nM of fusion protein. The total volume in the binding reactions was 100 µl in SEQRS buffer containing 200 ng yeast tRNA competitor and 0.1 U of RNase inhibitor (Promega) in eight-sample strip tubes. The samples were incubated for 30 min at 25 °C before capture of the protein–RNA complex. The binding reaction was aspirated, and the beads were washed four times with 200 µl of ice-cold SEQRS buffer. After the final wash step, the resin was resuspended in elution buffer (1 mM Tris, pH 8.0) containing 10 pmoles of the reverse-transcription primer. Samples were heated to 65 °C for 10 min and then cooled on ice. A 5-µl aliquot of the sample was added to a 10-µl ImProm-II reverse-transcription reaction (Promega). The ssDNA product was used as a template for PCR. The efficiency of the selection process was monitored by agarose gel electrophoresis.

High-throughput sequencing. The purity of each sample was determined by electrophoresis before sequencing. Samples were purified with PCR Clean-Up columns (Fermentas). Approximately equal amounts of bar-coded DNA were combined on the basis of individual concentrations determined by UV-vis spectroscopy (Thermo). After pooling of samples, 3 pmoles of DNA were sequenced on an Illumina HiSeq 2000 instrument. Data were analyzed as previously described[4,35].

Yeast three-hybrid assays. RNA-binding assays were conducted as previously described, with minor adjustments[15,18]. Luminescence data were collected with the β-Glo reagent (Promega) and measured with a 96-well synergy-2 plate reader (BioTek).

Immunoprecipitation and RT-PCR. Cells were washed three times in tissue culture–grade PBS (Gibco) and suspended in 0.5 ml of cold lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% NP-40, 100 U RNase inhibitor per ml and protease inhibitors (Roche)). Cells were frozen at −80C and subsequently thawed at 25 °C with rotating for 10 min. Lysate was clarified by centrifugation and transferred to a new tube containing 25 µl of Anti-Myc-tag mAb–Magnetic beads (MBL). Validation of the antibody is available through MBL. Binding occurred over a half-hour period of continuous end-over-end nutation at 4 °C. The beads were washed repeatedly with 3 ml of wash buffer (same as lysis buffer but without protease inhibitors). RNA samples were purified with the RNeasy Mini Kit (Qiagen). Purified RNA (~100 ng) was reverse transcribed with ImProm-II reverse-transcription reactions and random hexamers (Promega). Validated gene-specific primer sets used for amplification have been previously described[36].

Vectors. All inserts were cloned with the Gibson cloning method[37]. pCDNA3.1 was modified to include a 9× myc tag cloned into the XmnI site. Subsequent inserts were introduced into the PacI site. The FBF-2 inserts comprised residues 160–600, and cyclin B1 comprised residues 1–433 (full length). The PAB fragment (RRMs 1–3) sufficient for stimulation of translation was previously described[29]. Neo-activator constructs were generated via ligation of *S. cerevisiae* PAB1p RRMs 1–3 to the C terminus of FBF-2.

Cell culture and transfections. U2OS and HeLa cells were kind gifts from S. Miyamoto and R. Raines (both at University of Wisconsin–Madison), respectively. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS. Transfection experiments were conducted in six-well plates 1 d after seeding. On the basis of titrations of DNA concentration, we found optimal transfection efficiency with a ratio of 2 µg of pcDNA vectors and 3 µl of Lipofectamine 2000 (Invitrogen). Transfections were carried out by dilution of both components into 50 µl Opti-MEM I and subsequent mixing and half-hour incubation (Invitrogen). The resulting mixture was added to cells for 24 h at 37 °C.

Western blots. Cells were harvested 24 h after transfection. Cell pellets were clarified and then boiled in 6× SDS-PAGE loading buffer for 5 min. Electrophoresis was conducted on 4–15% gradient SDS-PAGE gels before transfer to nitrocellulose paper. Antibodies for cyclin B1 (Santa Cruz, GSN1), myc (Sigma, M4439) and actin (MP-Biomedicals, 691002) were obtained from commercial sources. Validation for each antibody is available from the manufacturers. Antibodies to GSN1, myc and actin were diluted to a working concentration of 1:5,000, 1:10,000 and 1:25,000, respectively. HRP-labeled secondary antibodies (KPL) were diluted 1:20,000 and were visualized with ECL reagent (Pierce) on an ImageQuant LAS 4000 (GE Healthcare). Gel bands were quantified with ImageQuant TL (GE Healthcare).

Immunofluorescence. Transfected U2OS cells were fixed with 4% formaldehyde in six-well microslides (Ibidi) in 1× PBS for 30 min at 25 °C and subsequently washed with ~200 µl PBS three times. Cells were then permeabilized with 0.2% Triton X-100 in PBS for 10 min and washed with ~200 µl PBS three times. Cells were blocked with 1% BSA in PBS for 10 min, washed with PBS three times and then incubated with primary antibody for 1 h at 25 °C. Cells were washed with PBS three times and then stained with 300 nM DAPI for 10 min. The cells were again washed with PBS three times and incubated with donkey FITC-conjugated goat anti-mouse IgG (Jackson) diluted 1:200 for 30 min. Cells were washed with 1× PBS three times and visualized in 0.5% *p*-phenylenediamine (Sigma) in 20 mM Tris, pH 8.8, with 90% glycerol. Cells were quantified with ImageJ, and color was applied with Adobe Photoshop CS3. PHH3 staining was done with an anti–Ser10 histone H3 antibody (sc-8656-R, Santa Cruz). Validation for sc-8656-R is available from Santa Cruz Biotechnology. The antibodies to cyclin B1 (Santa Cruz, GSN1) and Ser10 histone H3 (Santa Cruz, sc-8656-R) were diluted 1:100.

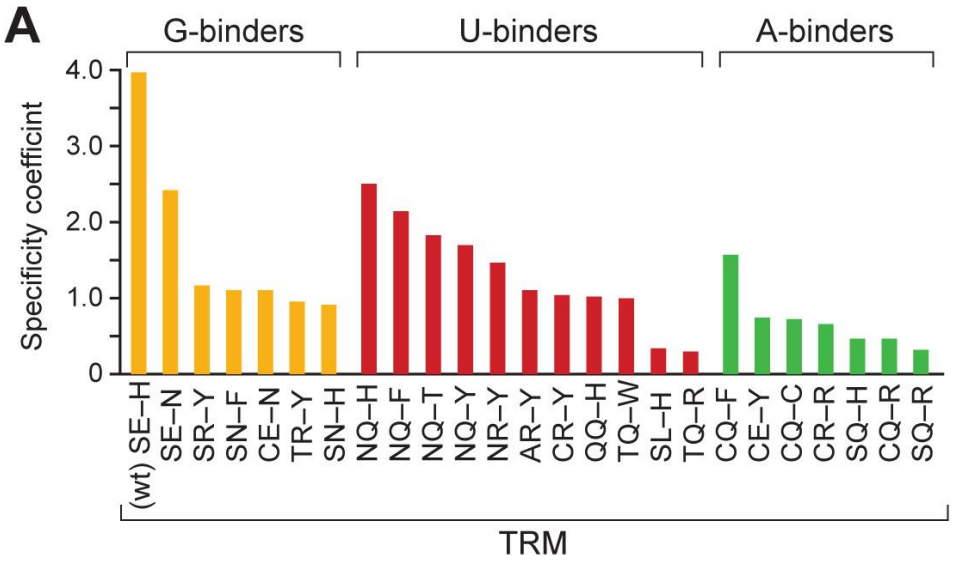
Drug sensitivity assays. Transfected cells were seeded into 96-well plates and exposed to 10 nM of either paclitaxel (Sigma), vinblastine (Sigma) or vehicle control DMSO (Sigma). After 3 h, drugs were removed, and the cells were allowed to recover for 24 h. Then viability was assessed with CellTiter-Glo reagent per the protocol provided by the manufacturer (Promega). Luminescence was recorded with a 96-well synergy-2 plate reader.

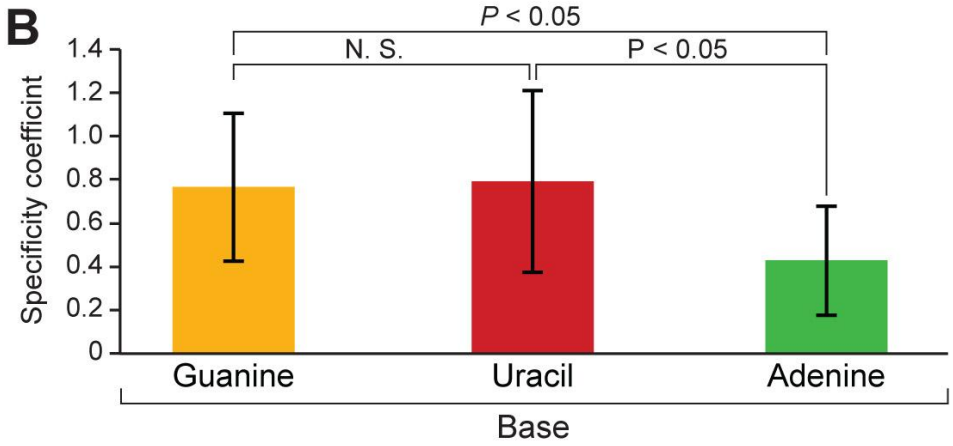
Cell growth assays. Transfected cells were assayed with the CellTiter-Glo system according to the manufacturer's instructions (Promega).

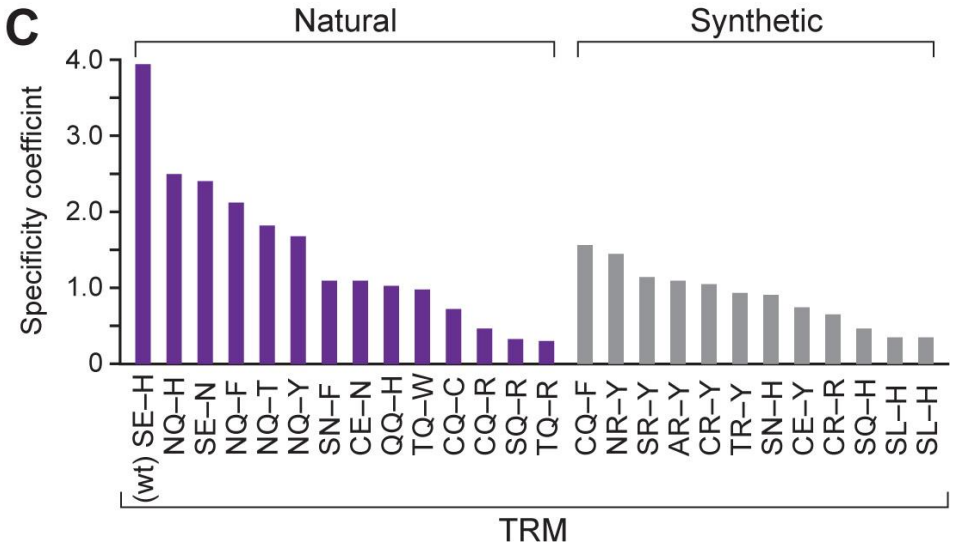
Statistics and data presentation. All reported *P* values were determined with a Student's *t* test. Error bars represent s.d. values from biological replicates unless noted otherwise. Values plotted as central points in bar graphs represent mean values.

Sequencing data. All data can be accessed online through the following website: http://www2.biochem.wisc.edu/zcampbell/supplement/.

33. Campbell, Z.T. & Baldwin, T.O. Two lysine residues in the bacterial luciferase mobile loop stabilize reaction intermediates. *J. Biol. Chem.* **284**, 32827–32834 (2009).
34. Campbell, Z.T. & Baldwin, T.O. Fre is the major flavin reductase supporting bioluminescence from *Vibrio harveyi* luciferase in *Escherichia coli*. *J. Biol. Chem.* **284**, 8322–8328 (2009).
35. LeGendre, J.B. *et al.* RNA targets and specificity of Staufen, a double-stranded RNA-binding protein in *Caenorhabditis elegans*. *J. Biol. Chem.* **288**, 2532–2545 (2013).
36. Li, L. *et al.* GLIPR1 suppresses prostate cancer development through targeted oncoprotein destruction. *Cancer Res.* **71**, 7694–7704 (2011).
37. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).


A
G-binders
U-binders
A-binders
Specificity coefficint
4.0
3.0
2.0
1.0
0
(wt) SE–H
SE–N
SR–Y
SN–F
CE–N
TR–Y
SN–H
NQ–H
NQ–F
NQ–T
NQ–Y
NR–Y
AR–Y
CR–Y
QQ–H
TQ–W
SL–H
TQ–R
CQ–F
CE–Y
CQ–C
CR–R
SQ–H
CQ–R
SQ–R
TRM

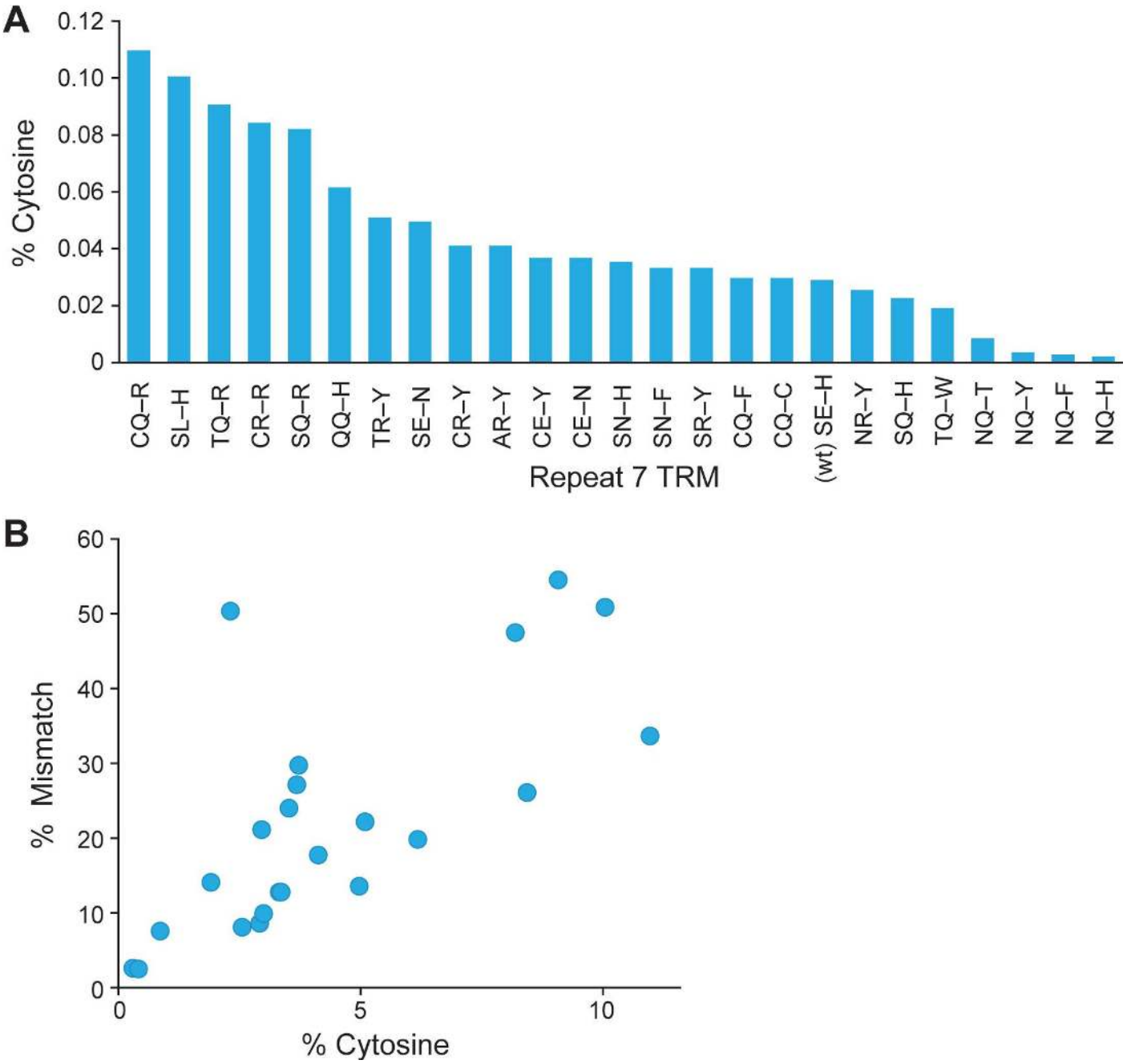

B
P < 0.05
N. S.
P < 0.05
Specificity coefficint
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
Guanine
Uracil
Adenine
Base


C
Natural
Synthetic
Specificity coefficint
4.0
3.0
2.0
1.0
0
(wt) SE–H
NQ–H
SE–N
NQ–F
NQ–T
NQ–Y
SN–F
CE–N
QQ–H
TQ–W
CQ–C
CQ–R
SQ–R
TQ–R
CQ–F
NR–Y
SR–Y
AR–Y
CR–Y
TR–Y
SN–H
CE–Y
CR–R
SQ–H
SL–H
SL–H
TRM

Supplementary Figure 1

Specificity coefficients and generality of the TRM code.

(A) The ranking of TRM specificity is shown based on the percent enrichment of the dominant base at position +2 multiplied by the inverse sum of non-WT identities at the +1 and +3 bases. The specificity coefficient thus includes both the preference for the new base at the targeted site as well as the “off-target” effects at flanking positions. Higher specificity coefficients indicate greater selectivity for a switch in identity only at the “targeted” base (in this case, +2). (B) The mean specificities for G, U, and A were calculated and compared using a two-tailed Student's T-test (error bars represent s.d.). Exclusion of the wild-type TRM from the G-specific group does not eliminate the level of significance when comparing G and A TRMs. (C) The specificities of natural and designed TRMs were compared, as in Panel (A). Natural proteins are slightly more specific (average 0.37 versus designed average 0.24).

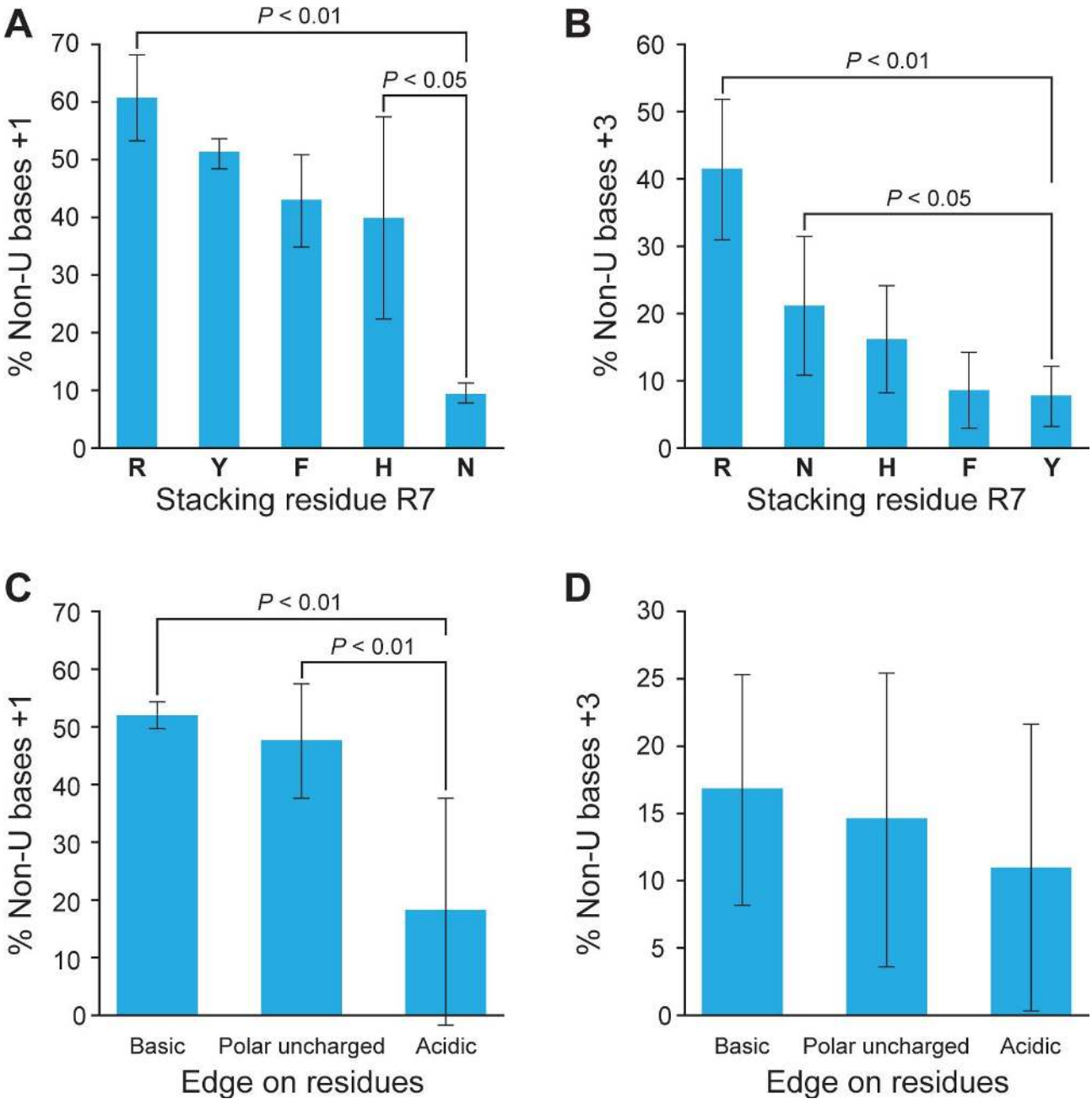


Supplementary Figure 2

Analysis of cytosine binding.

(A) Ranking of TRMs by percentage of binding elements with C at position 2. TRMs previously described as preferential C-binders are SR–Y, AR–Y, and CR–Y in the context of human PUM (Dong, S. et al. Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF)

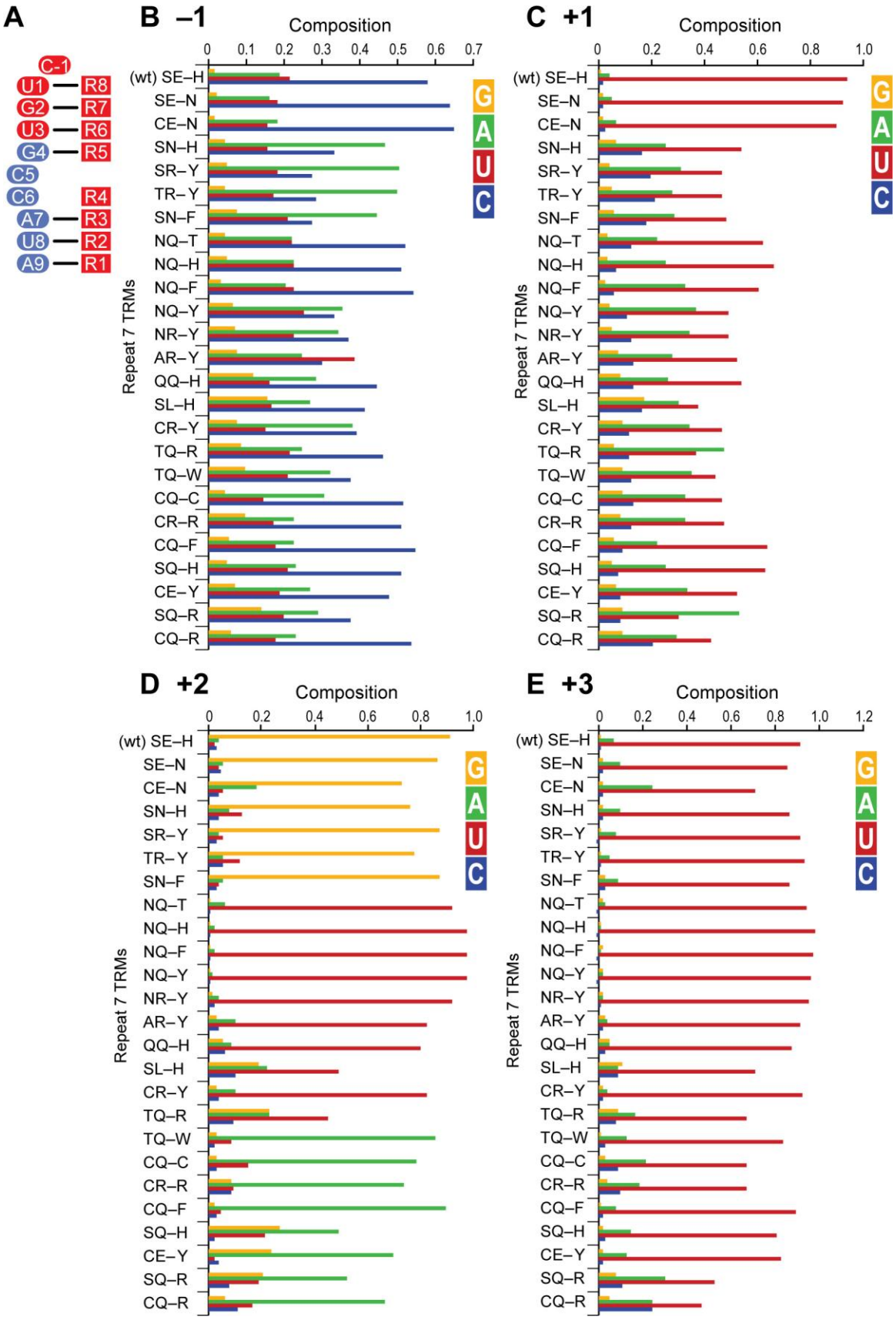
RNA-binding domains. J. Biol. Chem. 286, 26732–26742 (2011). and Filipovska, A., Razif, M.F., Nygard, K.K. & Rackham, O. A universal code for RNA recognition by PUF proteins. Nat. Chem. Biol. 7, 425–427 (2011)) (B) Correlation between specificity and C-binding preferences. The relative degree of non-specific binding was defined as the sum of all non-consensus bases for a TRM at position two and was plotted against the frequency of cytosine at position 2.



Supplementary Figure 3

Relationships between TRM composition and degeneracy at flanking sites.

Prevalence of mismatches at the flanking nucleotides +1 (A) and +3 (B). Prevalence of mismatches at the +1 (C) and +3 (D) positions and physiochemical properties of edge on-residues. Significance values were determined by two-tailed Student's t-test. Error bars, s.d. (n = 3 independent colonies).


A
C-1
U1
R8
G2
R7
U3
R6
G4
R5
C5
C6
R4
A7
R3
U8
R2
A9
R1
B −1
Composition
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Repeat 7 TRMs
(wt) SE–H
SE–N
CE–N
SN–H
SR–Y
TR–Y
SN–F
NQ–T
NQ–H
NQ–F
NQ–Y
NR–Y
AR–Y
QQ–H
SL–H
CR–Y
TQ–R
TQ–W
CQ–C
CR–R
CQ–F
SQ–H
CE–Y
SQ–R
CQ–R
G
A
U
C
C +1
Composition
0 0.2 0.4 0.6 0.8 1.0
Repeat 7 TRMs
(wt) SE–H
SE–N
CE–N
SN–H
SR–Y
TR–Y
SN–F
NQ–T
NQ–H
NQ–F
NQ–Y
NR–Y
AR–Y
QQ–H
SL–H
CR–Y
TQ–R
TQ–W
CQ–C
CR–R
CQ–F
SQ–H
CE–Y
SQ–R
CQ–R
G
A
U
C
D +2
Composition
0 0.2 0.4 0.6 0.8 1.0
Repeat 7 TRMs
(wt) SE–H
SE–N
CE–N
SN–H
SR–Y
TR–Y
SN–F
NQ–T
NQ–H
NQ–F
NQ–Y
NR–Y
AR–Y
QQ–H
SL–H
CR–Y
TQ–R
TQ–W
CQ–C
CR–R
CQ–F
SQ–H
CE–Y
SQ–R
CQ–R
G
A
U
C
E +3
Composition
0 0.2 0.4 0.6 0.8 1.0 1.2
Repeat 7 TRMs
(wt) SE–H
SE–N
CE–N
SN–H
SR–Y
TR–Y
SN–F
NQ–T
NQ–H
NQ–F
NQ–Y
NR–Y
AR–Y
QQ–H
SL–H
CR–Y
TQ–R
TQ–W
CQ–C
CR–R
CQ–F
SQ–H
CE–Y
SQ–R
CQ–R
G
A
U
C

Supplementary Figure 4

Effects of TRM substitutions on flanking bases.

(A) The binding arrangement of a canonical binding element is shown as a schematic. The four bases examined in the figure are shown in red ovals opposite the PUF repeats (boxes). Quantification across all enriched binding elements with single substitutions at the -1,+1,+2, and +3 bases are shown in panels B-E, respectively. G, A, U, and C are shown in yellow, green, red, and blue, respectively.

A

```
PUM2        QDQHGSRFIQQKLERATPAERQMVFNEILQAAYQLMTDVFGNYVIQKFFEFGSLDQKLAL
DdA         KDQVGSRIIQQKIENASLEEKQLVFDEVVVAVHSLMTDVFGNYVLQKFFEHGTTEQKRIL
DdB         RDQHGCRFLQKKLEESDLQLTEIIFKEVCDYMLELMTDPFGNYLCQKLLEHCNDRQRLTI
            :** *.*::*:*:*.:     :::*.*:     .**** ****: **::*. .   *:  :
                     R1                                  R2

PUM2        ATRIRGHVLPLALQMYGCRVIQKALESISSDQQVISEMVKELDGHVLKCVKDQNGNHVVQ
DdA         ADKLSGHILSLTLQMYGCRVIQKAIESIELDKQIL--LIGELNGHIVQCVTDQNGNHVIQ
DdB         IEKVGTDIVRISMNMHGTRAVQKMIEYLTTPEQIQ-LIKRSLKDSVVPLIQDLNGNHVIQ
              ::  .:: ::::*:* *.:** :* :   :*:   :  .*.. ::  : * *****:*
                          R3                                        R4

PUM2        KCIECVQPQSLQFIIDAFKG--QVFVLSTHPYGCRVIQRILEHCTAEQTLPILEELHQHT
DdA         KCIEKIPTHLIQFIIDSFHG--HIYQLATHPYGCRVIQRILEHCAEKQVAPILDELMRCA
DdB         KCLNKLSPQDNQFIYDAVSSDGSCIAVATHRHGCCVLQRCIDHASESQKLQLIQEVIANS
            **:: : .:  *** *:. .      ::** :** *:** ::*.: .*   :::*:   :
                                              R5

PUM2        EQLVQDQYGNYVIQHVLEHGRPEDKSKIVSEIRGKVLALSQHKFASNVVEKCVTHASRAE
DdA         VSLVQDQYGNYVIQHVLENGTPRDKSAIVCKLQGQIYNLSQHKFASNVIEKCVQHGCTAE
DdB         LVLVQDPYGNYVVQYVLDLPFQGLATEMAKRFVGHVPILATQKFSSNVVEKCLHVADAIT
              **** *****:*:**:       : :. .: *::  *: :**:***:***:  .
                       R6                                  R7

PUM2        RALLIDEVCCQNDGPHS--ALYTMMKDQYANYVVQKMIDMAEPAQRKIIMHKIRPHITTL
DdA         RILIINEILGDANSPNSSNVLLKILKDPYANYVIQKILDIVEPAQRDMIINRIQPFVPTL
DdB         RGYLIQEVIDYDN-------LLHLLQDPYANYVIQTSLTISEPHQHTKLVEAIRPHLPLL
            *  :*:*:    :       *  :::* *****:*. : : ** *:  ::. *:*.:. *
                                          R8

PUM2        RKYTYGKHILAKLEKYY-------
DdA         KKVTPGKHIISRIEKYSANNN---
DdB         KNTPYGKRIQNKIIKEGRDTYNFN
            :: . **:*  :: *
```

B

TRM identities

	R8	R7	R6	R5	R4	R3	R2	R1
Dd PufA	NQ/Y	SE/N	NQ/Y	CQ/R	NQ/H	CQ/R	NQ/Y	SQ/R
Dd PufB	NQ/Y	SE/N	NQ/Y	CQ/C	NQ/H	TQ/R	NQ/Y	CQ/R

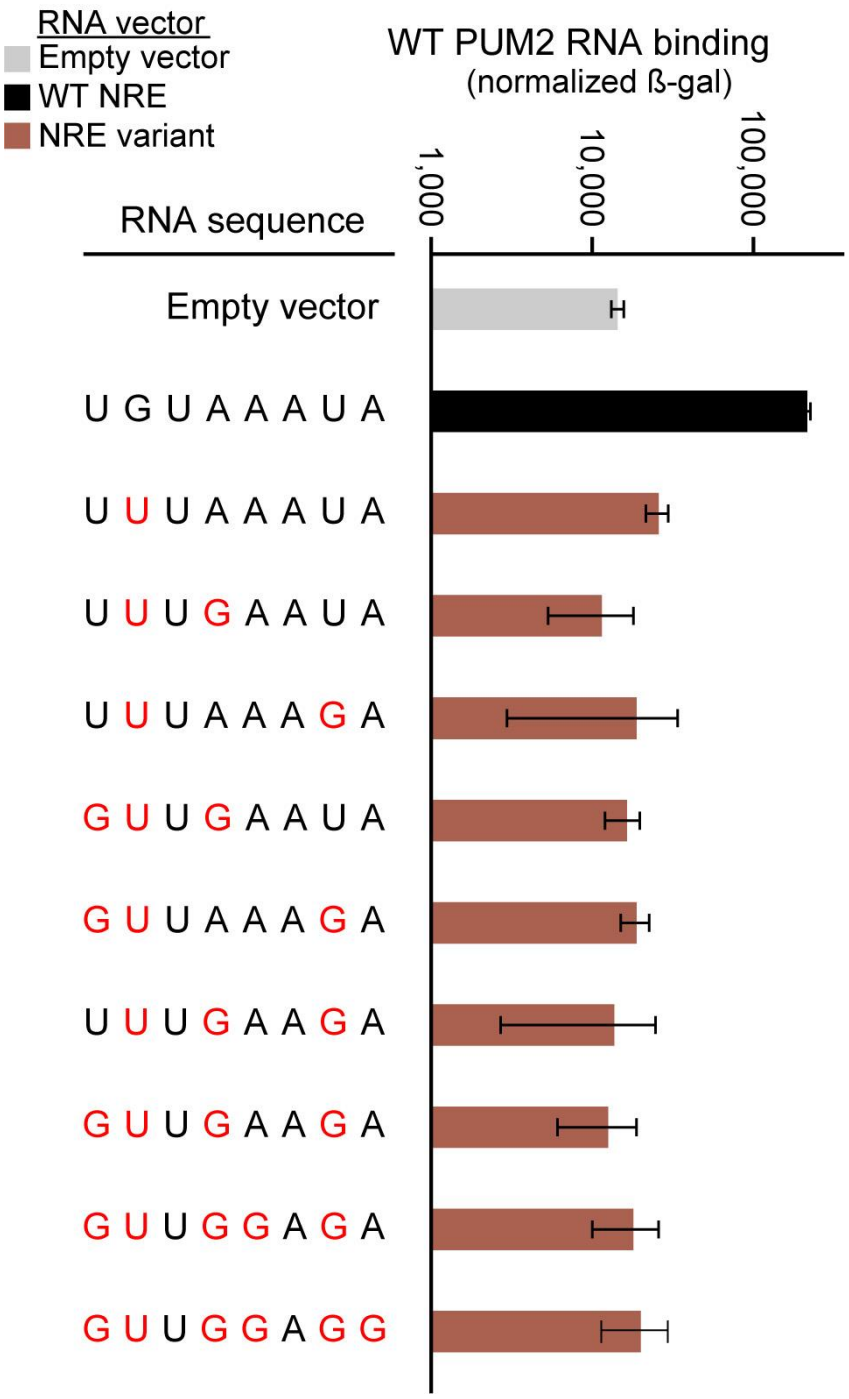
RNA-binding prediction from TRM code

Dd PufA	U	G	U	A	U	A	U	A
Dd PufB	U	G	U	A	U	U	U	A

Supplementary Figure 5

Identification of *Dictyostelium* TRMs.

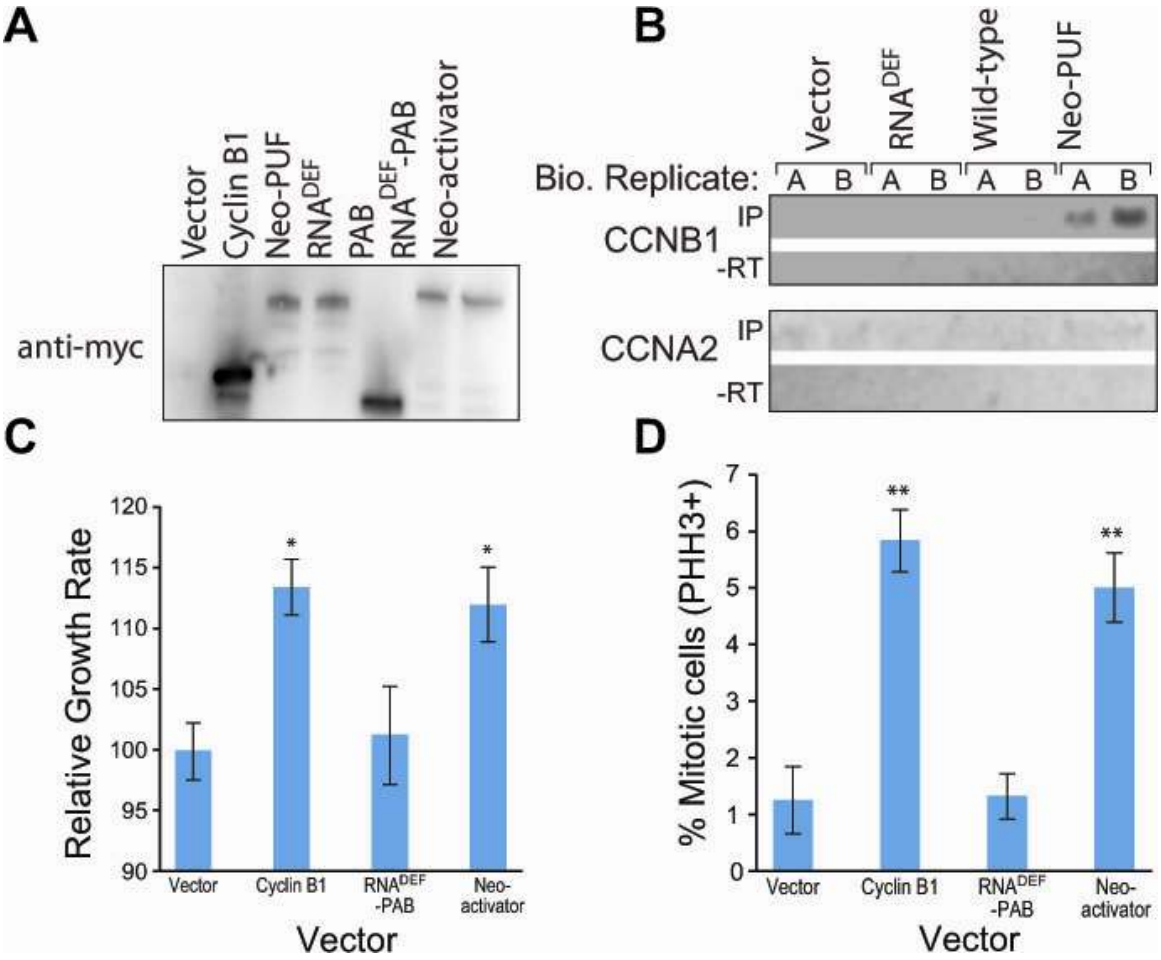
(A) The PUF proteins from Dictyostelium were aligned to PUM2 for identification of TRM combinations (red). (B) Predictions for RNA binding derived from TRM code reported in Fig. 2.



Supplementary Figure 6

Wild-type PUM2 binding to various sequence variants.

RNA mutants are shaded in brown according to the key. RNA sequences are provided for variant sites. Red nucleotides indicate sites that differ from the wild-type sequence. Binding activity measurements were conducted in the yeast-three hybrid system. Error bars are s.d. from three individual colonies.

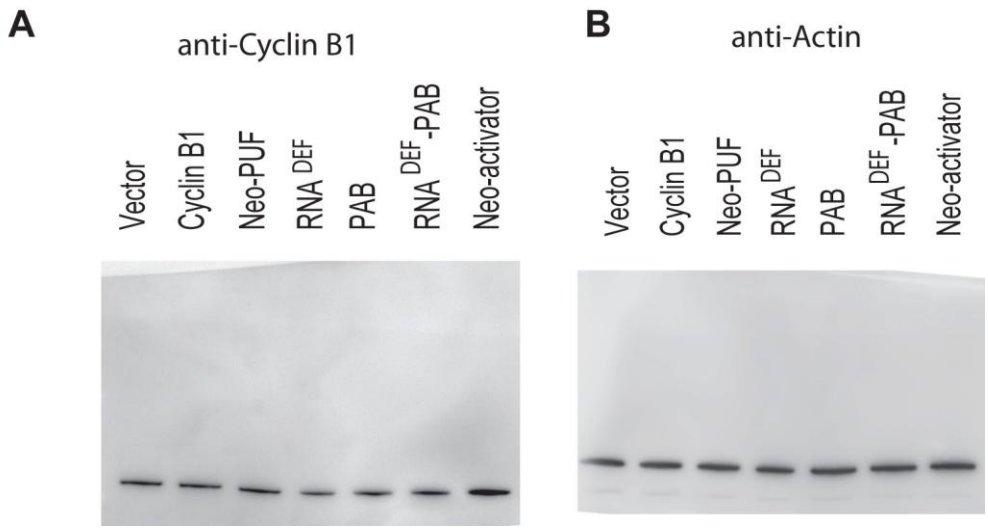


Supplementary Figure 7

Properties of the Neo-PUF and Neo-activator.

(A) Expression of myc-tagged transfected proteins in U2OS cells probed by immunoblot analysis. Acronyms are as follows - Cyclin B1 refers to full length Cyclin B1, neo-PUF refers to FBF-2 bearing the repeat 3 NQ–H substitution, RNA^{DEF} is the neo-PUF containing an RNA binding defective allele of FBF-2 with the H326A substitution, PAB is PAB1p RRMs 1-3, RNA^{DEF}-PAB is the neo-PUF containing the H326A mutation fused to PAB, and neo-activator is FBF-2 bearing the repeat 3 NQ–H substitution characterized in the main text fused to PAB. (B) RT-PCR analysis of neo-PUF binding to either to the *cyclin B1* transcript (CCNB1) or a negative control (CCNA2) *in vivo*. U2OS cells were transfected with equivalent amounts of the indicated vectors and transcripts were isolated by RNA

immuno- precipitation and RT-PCR. (C) Growth rate (inferred from measurement of metabolic flux) for U2OS cells transformed with either Cyclin B1 or the neo-activator was increased relative to vector alone or an RNA binding defective PUF–PAB chimera. Significance values were determined by two-tailed Student's t-test (* denotes $P < 0.05$, ** denotes $P < 0.01$). Error bars are s.d. from three cell cultures. (D) U2OS cells were treated with taxol for 12 hours to induce arrest. The drug was removed and the cells were allowed to recover in fresh media for four hours. The relative percentage of cells remaining in mitosis was determined by immunostaining for phosopho histone H3. Significance values were determined by two-tailed Student's t-test (* denotes $P < 0.05$, ** denotes $P < 0.01$). Error bars are s.d. from three cell cultures.



Supplementary Figure 8

Uncropped blot images.

(A-B) Related to Fig. 6