HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2015 February 01.

Published in final edited form as:

Nat Struct Mol Biol. 2014 August ; 21(8): 732–738. doi:10.1038/nsmb.2847.

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

Zachary T. Campbell, **Cary T. Valley**, and **Marvin Wickens**

Department of Biochemistry, University of Wisconsin-Madison Madison, WI 53706-1554, USA

Abstract

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, as they bind specific single-stranded RNA sequences using short repeated modules, each contributing three amino acids that contact an RNA base. Here, we identify 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 design 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.

INTRODUCTION

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 (7-10 nucleotide) sequence elements [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 or TRM (Fig 1A). TRMs contact RNA bases through a combination of edge-on and stacking interactions [13].

Corresponding author: Marvin Wickens Phone 608-263-0858 FAX 608-262-9108 wickens@biochem.wisc.edu.

AUTHOR CONTIBUTIONS

Z.T.C. preformed the experiments with assistance from CTV on Fig. 4 and Supplemental Supplementary Fig. 6. ZTC wrote the manuscript with input from MW and CTV.

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 have analyzed binding to a single RNA sequence and only at the targeted base, leading to unanticipated consequences. 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 we determined allow inference of the specificity of entire PUF proteins based solely on 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, see Methods). Fourteen of the most common TRMs at each repeat were selected for further analysis. In parallel, we examined the specificity of three artificial TRMs previously reported to preferentially bind cytosine, and eight novel TRM combinations of our own design [16,17].

We chose the *C. elegans* PUF protein, FBF-2, as a scaffold. Its specificity had been analyzed biochemically, structurally, and through the use of compensatory mutations [4,9,15,18,19] (Fig. 1a). Importantly, we reasoned that since FBF-2 is less than 20% identical to human PUM1 and PUM2, it was unlikely to elicit 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, mutations were introduced 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). SEQRS yields a proxy for binding affinity, in which the number of reads for a specific sequence is correlated with its affinity measured *in vitro* [4]. In our experiments, a DNA library encoding a random 20-mer region was transcribed to generate a random pool of RNAs. A sufficient quantity of RNA to cover all possible 20-mer sequences is incubated with recombinant proteins. The pool was then incubated with purified GST-tagged recombinant protein immobilized on magnetic resin to enable capture of the RNA protein complex. After repeated washing, bound RNAs were thermally eluted, and converted into double-stranded DNA using reverse transcription followed by PCR. The RNA was reverse transcribed into DNA using a primer complementary to the constant region. The single stranded DNA was amplified using a

primer set that re-introduces the T7 promoter. This enrichment procedure, analogous to SELEX, was repeated for five cycles prior to 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 using hierarchical clustering (Fig. 2B). 230 preferentially enriched unique sequences for each individual TRM were used to identify binding similarities. The heat map (Fig. 2B) was used to define three clusters (A, B, and C), specific for U, A or G at position +2, 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 uridine at position +2; however, TRMs varied considerably in their degree of non-target enrichment, as shown by degeneracy in the sequence logos, and revealed comprehensively in SSLs (Fig. 2C) [4,24]. SSLs represent binding data as a series of concentric rings. 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, non-target 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 synthetic TRMs (0.24, Supplementary Fig. 1C).

De novo designed TRMs provide a means both to diversify and to improve RNA specificity, and reveal complex interactions among TRM residues. TRMs CQ–F and CE–Y were more specific for adenosine than any natural TRM (Supplementary Fig. 1A and 1C). 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 adenosine or guanine. Taken together, our TRM design data suggest that while 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 naturally occurring TRMs.

In some instances, new bases were accommodated as a result of relaxed specificity. For example, while switches to cytosine specificity were not observed, several TRMs tolerated cytosine, yielding more than 5% of reads with that base at +2 (Supplementary Fig. 2A). However, cytosine enrichment paralleled that of the other three “non-targeted” bases

suggestive of 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, while 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 adenosine 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 likely due to 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 (see Discussion).

The TRM code enables identification of naturally occurring RNA binding sites. Using only the TRM data, we predicted the specificity 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 $P < .01$). 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 these data 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). The average specificity of natural TRMs was calculated on a repeat-by-repeat basis, and depicted 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, consistent with the high conservation of both RNA and protein identities in this region (Fig. 3C)[19]. We propose that this provides an evolutionary starting point for the evolution of novel 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 it did bind its wild-type site (Supplemental Fig 6). The engineered PUM2 proteins, designed using the TRM code, were tested against the same set of RNAs. We first introduced the best U-specific TRM (NQ–H) into repeat seven (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 guanine recognition (SE–H), maintaining the U-specific Repeat 7. The resulting nine mutant proteins bound their novel cognate elements and not the wild-type sequence. This result was observed 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 using 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 to three of the four RNA bases.

To determine whether addition of 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 altered TRM. Binding for each pair was assayed 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, consistent with our prior work, which showed 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’UTR of human cyclin B1 mRNA. Cyclin B1 is a critical regulator of the cell cycle responsible for entry into mitosis and exit from G2 [26,27]. We altered repeat 3 of FBF-2 so that it should now bind a sequence in the 3'UTR of cyclin B1 mRNA, in which position 7 is non-consensus (UGUGUUUU). We refer to this protein as a “neo-PUF” (Fig. 5A). The *in vitro* consensus differs slightly from the consensus derived from yeast three-hybrid studies, UGUANNAU [18]. 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 and 5B). 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, 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. +7U

sequences were enriched by the neo-PUF, while the wild-type protein enriched +7A, regardless of the larger sequence context. Enrichment oscillates 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 applicable 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 RT-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 (Supplemental 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, demonstrating that the PUF scaffold was functionally inert (Supplemental Fig. 8). Increased Cyclin B1 protein abundance was confirmed by immunofluorescence spectroscopy, in which the fraction of Cyclin B1 positive cells elevated by approximately 500 percent (Fig. 6B).

Cyclin B1 over-expression renders specific cancer cell lines hypersensitive to anti-mitotic chemotherapeutic drugs [26,27]. To determine whether the neo-activator enhanced sensitivity to such drugs, we performed viability assays following treatment with paclitaxel or vinblastine (Fig. 6C). Expression of the neo-activator reduced viability following exposure to either drug, in both U2OS and HeLa cells. The effect was similar to that achieved by transfecting the Cyclin B1 gene. Similarly, the neo-activator produced anticipated outcomes of Cyclin B1 over-expression including increased growth rate and delayed re-entry into the cell cycle following arrest (Supplementary Fig. 7C and 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 binding preferences 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 using the code. Our analysis of PUFs from an evolutionarily distant organism suggests high conservation of PUF

binding motifs. Second, designed TRMs can be as specific as their natural counterparts. The collection of novel TRMs we report effectively doubles the repertoire of known TRM combinations. Third, while 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 the N-terminal repeats. Similarly, mutations in C-terminal repeats or the bases they recognize affect binding more profoundly than those elsewhere [19]. Finally, certain TRMs broaden RNA base specificity, and may be mistaken for a switch, underscoring the importance of unbiased validation.

What are the limits to manipulation of RNA recognition? As our selection experiments focused on a single repeat required for RNA binding, we performed additional candidate tests which suggest the experiments at repeat seven are generally applicable at different sites and in a different scaffold (Fig. 4). However, structural analyses have shown 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 seven has the greatest specificity of the combinations reported here. Local differences in curvature can cause base-flipping, and complicate predictions of specificity in previously unexamined proteins, as we suggest is 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 despite the fact that these repeats always recognize UGU in natural binding sites. Successful manipulations of specificity in this region suggests 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 likely 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 using 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

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

homologues of Pumilio and as a result are predisposed towards a common set of TRM combinations. All of the PUF proteins were detected by homology to *S. cerevisiae* PUFs. TRM combinations were inferred based on 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 residues 121-632, and the *D. dictyostelium* proteins PufA 450-792 and PufB 690-1036 [18]. TRM mutants were generated using site directed mutagenesis as described [33,34]. Recombinant fusion proteins were purified as described using 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, 1 mM DTT) containing 20% glycerol prior to 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 using the AmpliScribe T7-Flash Transcription Kit (Epicentre). The reaction was treated with RNase free DNase to remove residual DNA and purified using 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 units of RNase inhibitor (Promega) in 8-sample strip tubes. The samples were incubated for 30 minutes at 25°C prior to 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 minutes and then cooled on ice. A 5 μl aliquot of the sample was added to a 10μl 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 prior to sequencing. Samples were purified using the PCR Clean-Up columns (Fermentas). Approximately equal amounts of barcoded DNA were combined based on individual concentrations determined by UV-Vis spectroscopy (Thermo). After pooling samples, 3 pmoles of DNA were sequenced on an Illumina HiSeq 2000 instrument. Data were analyzed as described [4,35].

Yeast three-hybrid assays

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

Immunoprecipitations 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 100U per ml RNase inhibitor and protease inhibitors (Roche)). Cells were frozen at -80C and subsequently thawed at 25°C with rotating for 10 minutes. 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 without protease inhibitors). RNA samples were purified using the RNeasy Mini Kit (Qiagen). Purified RNA (~100 ng) was reverse transcribed using 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 using the Gibson cloning method [37]. pCDNA3.1 was modified to include a 9 x MYC tag cloned into the XmnI site. Subsequent inserts were introduced into the PacI site. The FBF-2 inserts comprised residues 160-600, Cyclin B1 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 *Saccharomyces cerevisiae* PAB1p RRMs 1-3 to the C-terminus of FBF-2.

Cell culture and transfections

U2OS and HeLa cells were kind gifts of Drs. Shigeki Miyamoto (UW-Madison) and Ron Raines (UW-Madison) respectively. Cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS. Transfection experiments were conducted in 6 well plates one day post seeding. Based on 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 diluting both components into 50μl Opti-MEM I followed by mixing and a half hour incubation (Invitrogen). The resulting mixture was added to cells for 24 hours at 37 °C.

Western blots

Cells were harvested 24 hours post-transfection. Cell pellets were clairified and then boiled in 6x SDS-PAGE loading buffer for 5 minutes. Electorphoresis was conducted on 4-15% gradient SDS-PAGE gels prior to transfer to nitrocellulose paper. Antibodies for Cyclin B1 (Santa Cruz GSN1), Myc (Sigma M4439), and Actin (MP-Biomedicals 691002) were obtained from commercial sources. Confirmation for each antibody is available from the manufacturer. GSN1, Myc, and Actin antibodies were diluted to a working concentration of 1:5,000, 1:10,000, and 1:25,000 respectively. HRP-labeled secondary antibodies were visualized using ECL reagent (Pierce) on an ImageQuant LAS 4000 (GE Healthcare). Gel bands were quantified using Imagequant TL (GE Healthcare).

Immunofluorescence

Transfected U2OS cells were fixed with 4% formaldehyde in 6 well μ-slides (Ibidi) 1× PBS for 30 minutes 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 minutes and washed with ~200 μl PBS three times. Cells were blocked with 1% BSA in PBS for 10 minutes, 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 300nM DAPI for 10 minutes. The cells were again washed with PBS three times and incubated with FITC-conjugated goat anti-mouse IgG for 30 minutes. Cells were washed with 1× PBS three times and visualized in 0.5 % pphenylenediamine (Sigma) in 20 mM Tris, pH 8.8 with 90 % glycerol. Cells were quantified using ImageJ and color was applied using Adobe Photoshop CS3. PHH3 staining was done with an anti Ser-10 histone H3 antibody (sc-8656-R, Santa Cruz Biotechnology). Validation for sc-8656-R is available from Santa Cruz Biotechnology.

Drug sensitivity assays

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

Cell growth assays

Transfected cells were assayed using the CellTiter-Glo system according to the manufacturer's instructions (Promega).

Statistics and data presentation

All reported p-values were determined using a Student's t-test. Error bars represent standard deviation 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 web site: http://www2.biochem.wisc.edu/zcampbell/supplement/

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank L. Vanderploeg for help with figures, M. Preston and S. Waghray for careful editing. This work was supported by National Institutes of Health grant GM050942 to M.W.

REFERENCES AND NOTES

1. Proudfoot C, McPherson AL, Kolb AF, Stark WM. Zinc finger recombinases with adaptable DNA sequence specificity. PLoS One. 2011; 6:e19537. [PubMed: 21559340]
2. Wood AJ, et al. Targeted genome editing across species using ZFNs and TALENs. Science. 2011; 333:307. [PubMed: 21700836]
3. Wickens M, Bernstein DS, Kimble J, Parker R. A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet. 2002; 18:150–7. [PubMed: 11858839]
4. Campbell ZT, et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. Cell Rep. 2012; 1:570–81. [PubMed: 22708079]
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. 2008; 3:e3164. [PubMed: 18776931]
6. Gerber AP, Herschlag D, Brown PO. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol. 2004; 2:E79. [PubMed: 15024427]
7. Kershner AM, Kimble J. Genome-wide analysis of mRNA targets for Caenorhabditis elegans FBF, a conserved stem cell regulator. Proc Natl Acad Sci U S A. 2010; 107:3936–41. [PubMed: 20142496]
8. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010; 141:129–41. [PubMed: 20371350]
9. Wang Y, Opperman L, Wickens M, Hall TM. Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. Proc Natl Acad Sci U S A. 2009; 106:20186–91. [PubMed: 19901328]
10. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. Cell. 2002; 110:501–12. [PubMed: 12202039]
11. Wang X, Zamore PD, Hall TM. Crystal structure of a Pumilio homology domain. Mol Cell. 2001; 7:855–65. [PubMed: 11336708]
12. Zhu D, Stumpf CR, Krahn JM, Wickens M, Hall TM. A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. Proc Natl Acad Sci U S A. 2009; 106:20192–7. [PubMed: 19918084]
13. Koh YY, et al. Stacking interactions in PUF-RNA complexes. RNA. 2011; 17:718–27. [PubMed: 21372189]
14. Wang Y, Wang Z, Tanaka Hall TM. Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. FEBS J. 2013; 280:3755–67. [PubMed: 23731364]
15. Opperman L, Hook B, DeFino M, Bernstein DS, Wickens M. A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. Nat Struct Mol Biol. 2005; 12:945–51. [PubMed: 16244662]
16. Dong S, et al. Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. J Biol Chem. 2011; 286:26732–42. [PubMed: 21653694]
17. Filipovska A, Razif MF, Nygard KK, Rackham O. A universal code for RNA recognition by PUF proteins. Nat Chem Biol. 2011; 7:425–7. [PubMed: 21572425]
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. 2005; 11:447–58. [PubMed: 15769874]
19. Valley CT, et al. Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. Proc Natl Acad Sci U S A. 2012; 109:6054–9. [PubMed: 22467831]
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. 2012; 287:6949–57. [PubMed: 22205700]
21. Stumpf CR, Kimble J, Wickens M. A Caenorhabditis elegans PUF protein family with distinct RNA binding specificity. RNA. 2008; 14:1550–7. [PubMed: 18579869]

22. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriphage T4 DNA polymerase. Science. 1990; 3:505–510. [PubMed: 2200121]
23. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990; 346:818–22. [PubMed: 1697402]
24. Carlson CD, et al. Specificity landscapes of DNA binding molecules elucidate biological function. Proc Natl Acad Sci U S A. 2010; 107:4544–9. [PubMed: 20176964]
25. Miller MT, Higgin JJ, Hall TM. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol. 2008; 15:397–402. [PubMed: 18327269]
26. Brito DA, Rieder CL. 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. 2009; 66:437–47. [PubMed: 18792104]
27. Russo AJ, et al. E2F-1 overexpression in U2OS cells increases cyclin B1 levels and cdc2 kinase activity and sensitizes cells to antimitotic agents. Cancer Res. 2006; 66:7253–60. [PubMed: 16849574]
28. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev. 2006; 20:515–24. [PubMed: 16510870]
29. Gray NK, Collar JM, Dickson KS, Wickens M. Multiple portions of poly(A)-binding protein stimulate translation in vivo. EMBO Journal. 2000; 19:1–11. [PubMed: 10619838]
30. Lam KN, van Bakel H, Cote AG, van der Ven A, Hughes TR. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. Nucleic Acids Res. 2011; 39:4680–90. [PubMed: 21321018]
31. Enuameh MS, 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. 2013; 23:928–40. [PubMed: 23471540]
32. Lu G, Hall TM. Alternate modes of cognate RNA recognition by human PUMILIO proteins. Structure. 2011; 19:361–7. [PubMed: 21397187]
33. Campbell ZT, Baldwin TO. Two lysine residues in the bacterial luciferase mobile loop stabilize reaction intermediates. J Biol Chem. 2009
34. Campbell ZT, Baldwin TO. Fre Is the Major Flavin Reductase Supporting Bioluminescence from Vibrio harveyi Luciferase in Escherichia coli. J Biol Chem. 2009; 284:8322–8. [PubMed: 19139094]
35. LeGendre JB, et al. RNA targets and specificity of Staufen, a double-stranded RNA-binding protein in Caenorhabditis elegans. J Biol Chem. 2013; 288:2532–45. [PubMed: 23195953]
36. Li L, et al. GLIPR1 suppresses prostate cancer development through targeted oncoprotein destruction. Cancer Res. 2011; 71:7694–704. [PubMed: 22025562]
37. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–5. [PubMed: 19363495]

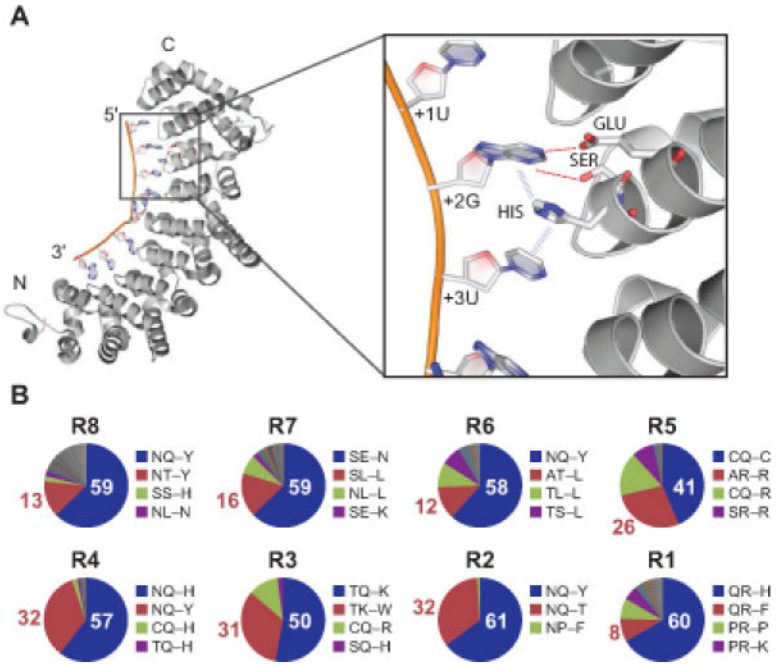

Figure 1. RNA recognition by the PUF proteins

(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 (SE), followed by a dash and the stacking residue (H). (B) Abundance of natural TRMs inferred from sequence alignment. Pie charts represent TRM enrichment as a function of PUF repeat (R8-R1) based on 94 proteins.

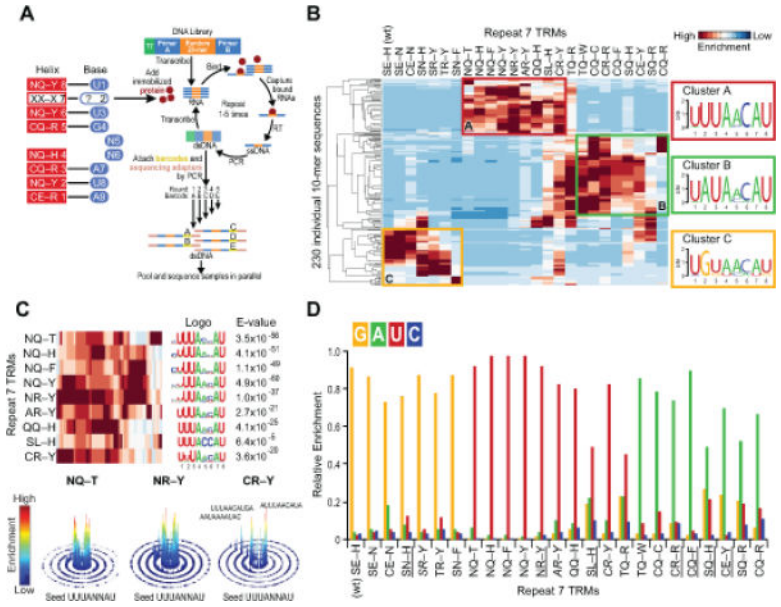

Figure 2. A quantitative TRM recognition code

(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 using the SEQRS technique (see text). (B) Hierarchical clustering reveals three classes of TRM binding specificity. Left, highly enriched 10-mer sequences for each TRM were identified (Y-axis) and the enrichment values were used to cluster similar binding profiles for each mutant (X-axis). For each TRM, the data were normalized to the maximum enrichment value. Right, three clusters were identified empirically and a representative motif was generated. (C) Sequence logos for members of cluster A reveal a common specificity consistent with the results from clustering (Top). The innermost ring contains sequences perfectly matched to a given seed motif, while subsequent rings contain increasing numbers of mismatches from that seed motif. SSLs for three representative TRMs reveal non-equivalent differences in overall specificity (Bottom). (D) Enrichment at position +2 of the PUF binding element. The relative enrichment for G (yellow), U (red), A (green), and C (blue). TRMs previously described as preferential C-binders, SR–Y, AR–Y, and CR–Y, are italicized[16,17]. The remaining synthetic combinations are underlined.

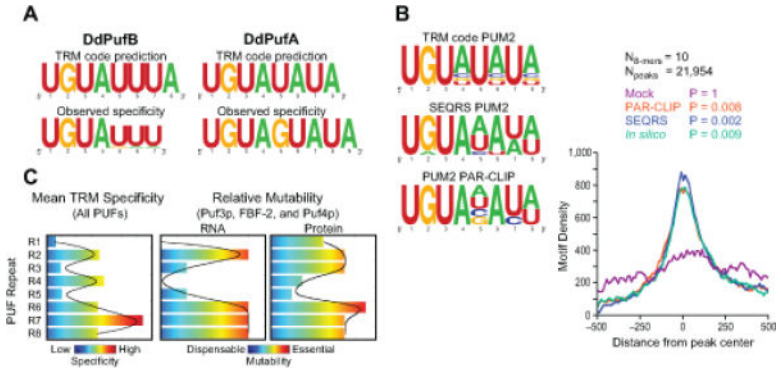

Figure 3. Prediction and distribution of specificity in nature

(A) Predictions of specificity *in vitro*. Prediction of specificity of uncharacterized PUF proteins from *Dictyostelium dictyostelium*. Motifs depicting specificities were generated from primary sequence (predicted) and compared to experimentally determined motifs by SEQRS (observed). (B) Prediction of and occupancy *in vivo*. Frequency plots represent predicted specificity from TRM data, a previously described SEQRS analysis of PUM2, an *in vivo* binding motif derived from photo-crosslinking, or a mock where G bases were replaced with C bases [4,8]. (C) The distribution of specificity in natural PUF proteins. The black line denotes a smoothed fit to the observed data. Both TRM repeats and RNA bases have been subjected to extensive mutagenesis in prior work for Puf3p, FBF-2, and Puf4p [19]. That data is compiled in the right panel ("Relative mutability"), and reveals the average tolerance of the base or TRM to substitution for each repeat.

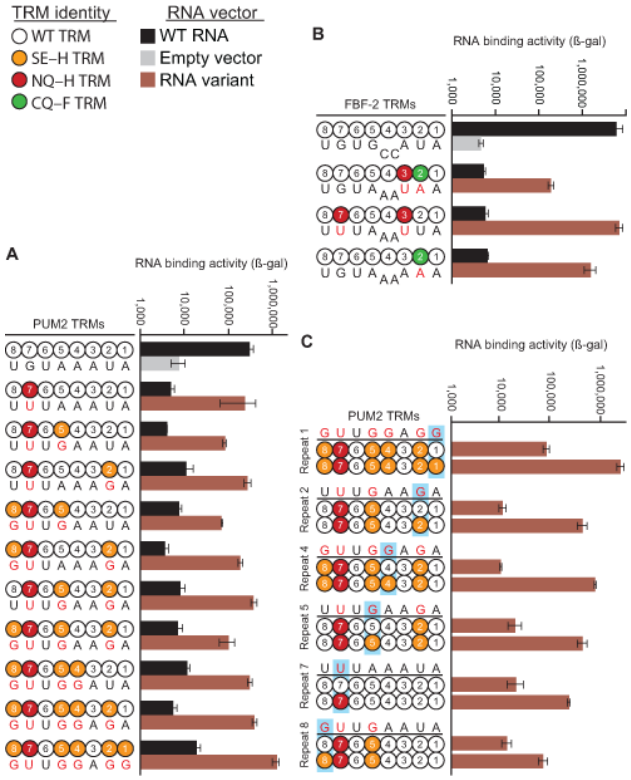

Figure 4. Modifications of PUF scaffolds using the TRM code

TRM variants are denoted as colored circles and RNAs as colors in bar charts 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. (A) Replacement of TRMs in the PUM2 scaffold yields protein mutants with novel specificity. 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). (B) Mutations in FBF-2 yield 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).

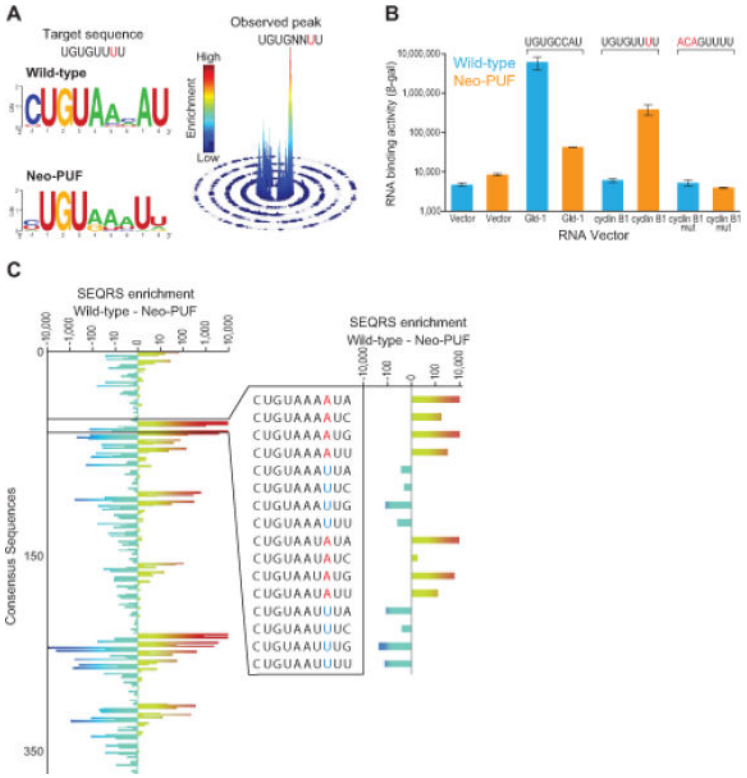

Figure 5. Engineered specificity of PUF proteins

(A) Sequence motifs of wild-type (WT) and redesigned (neo-PUF) proteins. The targeted recognition site possesses a single substitution at position seven of the binding element. (B) Analysis of RNA binding activity in a yeast three-hybrid assay. RNA binding for wild-type (blue) and neo-PUF (orange) measurements for an empty vector, positive control *gld-1* RNA element, and the *cyclin B1* targeting element are shown. Error bars, s.d. (n = 3 independent colonies). (C) Binding enrichment values for the neo-PUF and the wild-type proteins were calculated for the *in vitro* consensus sequence HUGURWWWU and subtracted (Wt-Neo) [4]. 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, posses sequences altered at positions +7 and +9. The plots indicate the degree of enrichment either for 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 to dark blue.

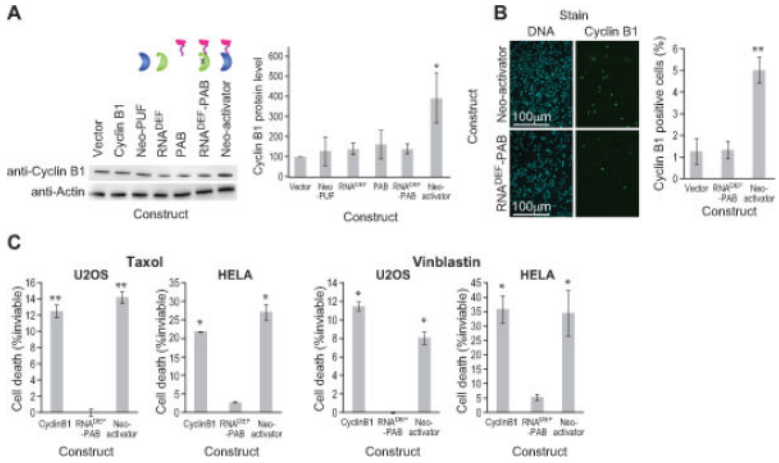

Figure 6. Manipulation of translation using a modified PUF scaffold

(A) Effects of a neo-activator on Cyclin B1 levels in U2OS cells measured by western blotting. (right - quantification). Error bars, s.d. (n = 3 cell cultures) * $P < 0.05$, ** $P < 0.05$, by two-tailed Student's t test. (B) Immunofluorescence of Cyclin B1 expression in U2OS cells expressing the Neo-activator (right - quantification). Error bars, s.d. (n = 3 cell cultures) * $P < 0.05$, ** $P < 0.05$ by two-tailed Student's t test. Slides were observed under the same microscope using identical parameters (scale bar = 100 μm). (C) Viability assays. Viability was quantified 24 hours post-treatment. Normalized cell death is shown for two M-phase targeting drugs [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.